SRI VENKATESWARA INTERNSHIP PROGRAM FOR RESEARCH IN ACADEMICS (SRI-VIPRA)

SRI-VIPRA

Project Report of 2024: SVP-2438

"**An attempt to elucidate emerging pathogens**"

IQAC

Sri Venkateswara College

University of Delhi

Benito Juarez Road, Dhaula Kuan, New Delhi

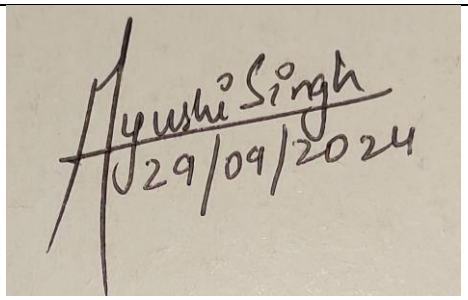
New Delhi -110021

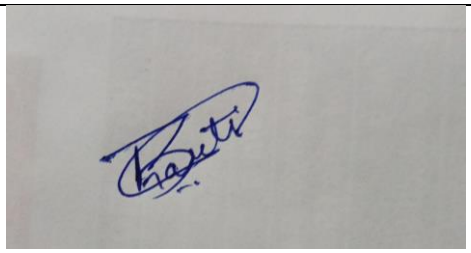
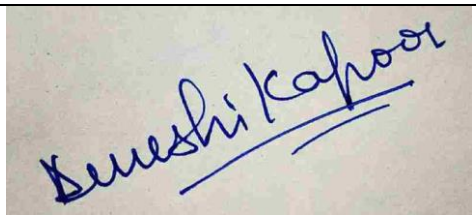
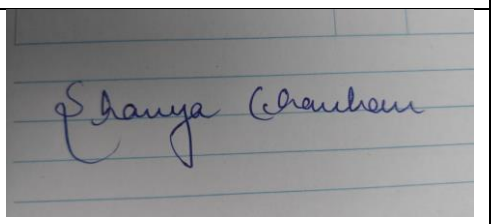
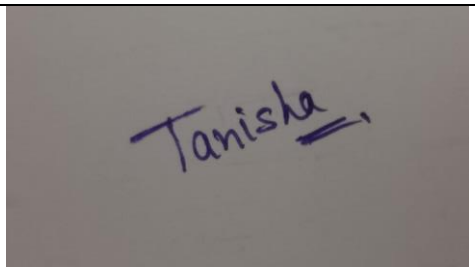
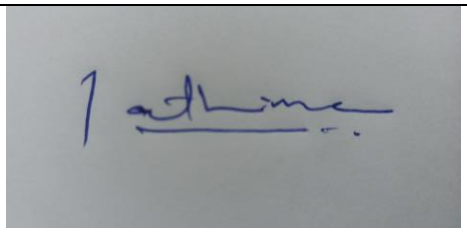
SRIVIPRA PROJECT 2024

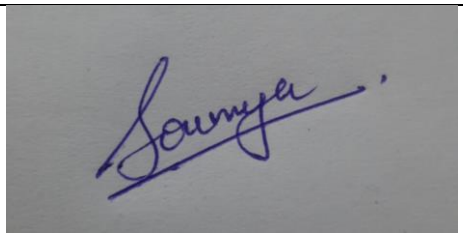
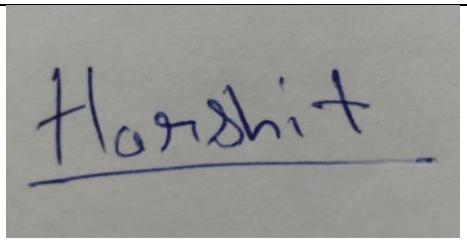
Title : An attempt to elucidate emerging pathogens

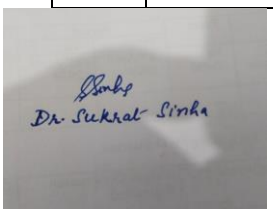
1.Name of Mentor: Dr Sukrat Sinha **Name of Department: Zoology** **Designation: Assistant Professor, Sri Venkateswara College, DU**	
2.Name of Mentor: Dr Navneet Kumar **Name of Department: Botany** **Designation: Assistant Professor, Sri Venkateswara College, DU**	

List of students under the SRIVIPRA Project

S.No	Photo	Name of the student	Roll number	Course	Signature
1		Ayushi Singh	1323045	BSc (H) Biological Sciences	

2		Bharti Kaswan	1323019	BSc (H) Biological Sciences	
3		Deveshi Kapoor	2023015	BSc (H) Zoology	
4		Shanya Chauhan	1123116	BSc (P) Life Sciences	
5		Tanisha	1123169	BSc (P) Life Sciences	
6		Sana Fathima	1323014	BSc (H) Biological Sciences	

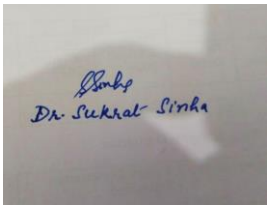
7		Soumya Tripathi	1122149	BSc (P) Life Sciences	
8		Harshit Kumar	1123123	BSc (P) Life Sciences	



Signature of Mentor

Certificate of Originality

This is to certify that the aforementioned students from Sri Venkateswara College have participated in the summer project SVP-2438 titled "**An attempt to elucidate emerging pathogens"**. The participants have carried out the research project work under my guidance and supervision from 1st July, 2024 to 30th September 2024. The work carried out is original and carried out in an online/offline/hybrid mode.



Signature of Mentor

Acknowledgements

We would like to extend our heartfelt gratitude to Dr Sukrat Sinha and Dr Navneet Kumar for their exceptional guidance and mentorship throughout our internship on emerging pathogens.

Their expertise, patience, and encouragement fostered a stimulating environment, allowing us to explore the complexities of emerging pathogens. We appreciate the trust they placed in us and the opportunities provided to contribute to ongoing research.

We are thankful for the collaborative and inclusive atmosphere, which facilitated valuable interactions among the internship cohort. This experience not only advanced our knowledge, but also instilled in us the importance of teamwork, critical thinking, and scientific inquiry.

Thank you so much for your dedication, support, and inspiration. We are grateful for the experience and look forward to applying the skills and knowledge gained during this internship in our future endeavors.

Sincerely,
Ayushi Singh
Bharti Kaswan
Deveshi Kapoor
Shanya Chauhan
Tanisha
Sana Fathima
Soumya Tripathi
Harshit Kumar

Abstract

Emerging pathogens are newly identified or previously unrecognized infectious agents that are increasing in incidence or geographic range. They can include viruses, bacteria, fungi, and parasites. Factors contributing to the emergence of these pathogens include:

- Environmental Changes: Climate change, deforestation, and urbanization can alter ecosystems and increase human exposure to new pathogens.
- Globalization: Increased travel and trade facilitate the rapid spread of infectious diseases across borders.
- Antimicrobial Resistance: Some pathogens become resistant to existing treatments, making them harder to control.
- Zoonotic Transmission: Many emerging pathogens originate in animals and can jump to humans, often due to close contact with wildlife or livestock.

Examples of emerging pathogens include:

- Monkeypox virus- It is a zoonotic virus belonging to the Orthopoxvirus genus, closely related to smallpox. First identified in 1958 among monkeys in a research facility, it was later found to cause illness in humans, with the first human case reported in 1970 in the Democratic Republic of the Congo.
- Primary Amoebic Meningoencephalitis (PAM)- It is a rare but severe infection of the brain caused by the free-living amoeba *Naegleria fowleri*. This organism is commonly found in warm freshwater environments, such as lakes and hot springs, as well as in poorly maintained swimming pools.
- Cryptococcus neoformans-It is a fungal pathogen that primarily affects individuals with weakened immune systems, such as those with HIV/AIDS. It is commonly found in the environment, particularly in soil and associated with bird droppings, especially from pigeons.
- Dengue-It is a viral infection caused by the dengue virus, transmitted primarily by Aedes mosquitoes, especially Aedes aegypti. It is prevalent in tropical and subtropical regions worldwide. Symptoms typically appear 4 to 10 days after infection and can include high fever, severe headaches, joint and muscle pain, rash, and mild bleeding.
- Zika virus-It is a mosquito-borne virus primarily transmitted by Aedes mosquitoes, particularly Aedes aegypti. It gained global attention during outbreaks in the Americas around 2015–2016. Most Zika infections are mild or asymptomatic, with common symptoms including fever, rash, joint pain, and conjunctivitis.
- COVID-19- It is caused by the novel coronavirus SARS-CoV-2, first identified in Wuhan, China, in late 2019. It primarily spreads through respiratory droplets when an infected person coughs, sneezes, or talks.

Symptoms can range from mild to severe and include fever, cough, shortness of breath, fatigue, and loss of taste or smell. Some individuals may experience severe respiratory illness or complications, particularly older adults and those with underlying health conditions.

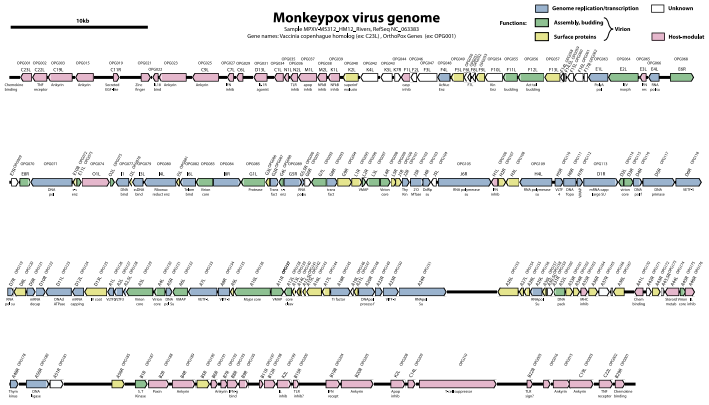
- *Helicobacter pylori*- It is a spiral-shaped, gram-negative bacterium that infects the stomach lining and is a major cause of peptic ulcers and chronic gastritis. It was discovered in 1982 by Barry Marshall and Robin Warren, who later received the Nobel Prize for their work.
- Nipah virus- It is a zoonotic virus first identified in 1999 during an outbreak in Malaysia, linked to pig farming. It is primarily transmitted from bats (especially fruit bats) to humans, but can also spread through direct contact with infected animals, contaminated food, or human-to-human transmission.

TABLE OF CONTENTS

Monkey Pox

INTRODUCTION

Monkeypox is a zoonotic virus belonging to the Orthopoxvirus genus, the same group as the smallpox virus. It primarily circulates in wild animals like rodents and primates, with occasional spillover into humans. First discovered in 1958 in laboratory monkeys, the virus caused the first human case in 1970 in the Democratic Republic of Congo. Human-to-human transmission occurs via



direct contact with bodily fluids, respiratory droplets, or lesion material, although it's less contagious than smallpox. Symptoms include fever, rash, swollen lymph nodes, and muscle aches, with the illness typically lasting 2-4 weeks. Vaccines developed for smallpox have shown cross-protection against monkeypox.

For more in-depth research, primary proteins such as BR-203, BR-209, and others are often studied to understand the virus's structure and behaviour, which helps in developing therapeutic strategies.

GENOMIC STRUCTURE

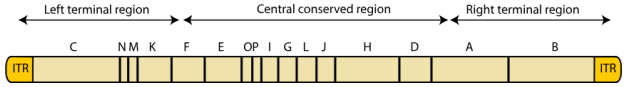
The genome of MPV appears to be typical of orthopoxviruses, including a central conserved region, more variable left and right end regions, and an ITR with tandem repeats. . The MPV-ZAI genome contains a 6379-bp ITR.

The poxvirus genome is about 200kb in size and encodes about 200 proteins. It is a linear double-stranded DNA genome with covalently closed hairpin ends (no free 3' or 5' ends). It contains 10kb inverted terminal repeats (ITR) at each end. Genes are closely packed: intergenic regions of more than 100bp are rare.

The central conserved region encodes "housekeeping" genes involved in transcription, replication and virion assembly. The genes encoded in the terminal regions vary between poxviruses and encode proteins involved in host range and

Figure 1 Genomic structure of Monkey pox virus analysed using Expasy ViralZone.

pathogenesis.



ITR= Inverted terminal repeats

BR-203 Virulence Protein

BR-203 is the virulence protein present in monkey pox virus. It is believed to have role in avoiding the apoptosis of infected lymphocytes. Central African clade of Monkey pox virus encode full protein length of 221 amino acids(aa). Western African

clade encode only the N terminal fragment of about 51 amino acids(aa) due to two base deletion which causes a frameshift. This protein is an ortholog to Myxoma virus MT-4 gene.

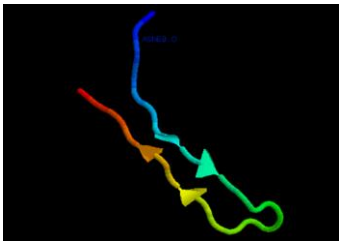


Figure 2 BR-203 protein structure analysed by using Rasmoll

COP-C3L Complement Control Protein

This protein inhibits the early steps of host complement cascade. It's present in variola and another ortholog of this protein is present in smallpox. MOPICE Monkey Pox Inhibitor Enzyme has three full short consensus repeats (SCR) due to single base selection. In one study, MOPICE bind neural C3B and C4b. C3L is said to account for the virulence of Western African Clade and Central African Clade.

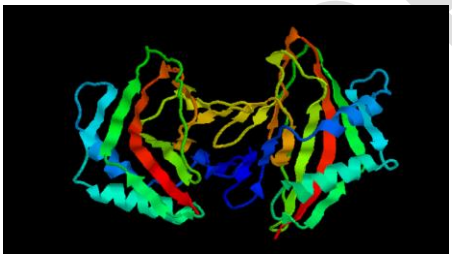


Figure 3 COP-C3L protein structure analysed by using Rasmoll.

263 amino acids(aa)

***GMQE** 0.95*

***QMEANDisCo Global:** 0.88 ± 0.05*

BR-209 Interleukin 1B binding protein

This protein functions as interleukin 1B binding protein that prevents 1L1B binding from binding to 1L1a receptor. Central African Clade contain open reading frames and it codes two fragments of BR-209; N terminal protein of 210 aa and C terminal of 126 aa. Western African clade has one base insertion near N terminal and a four-base deletion resulting in 2 frameshifts which as a result encodes central 163 aa and C terminal 132 aa.

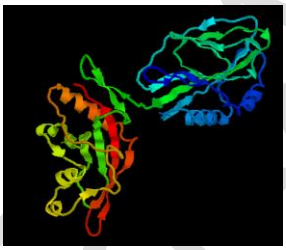


Figure 4. BR-209 structure analysed using Rasmoll.

326 amino acids(aa)

***GMQE** 0.95*

***QMEANDisCo Global:** 0.88 ± 0.05*

COP-A44L Hydroxysteroid dehydrogenase

It is predicted to encode 3B hydroxysteroid dehydrogenase to convert pregnanolone to progesterone and dihydropyran androsterone which gets converted later to androstenedione.

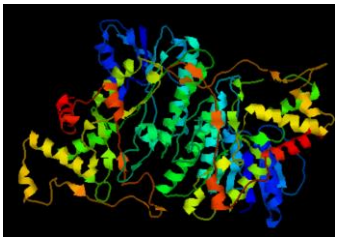


Figure 5 COP-A44L analysed with Rasmoll.

346 amino acids(aa)

***GMQE** 0.51*

***QMEANDisCo Global:** 0.52 ± 0.05*

A35R

A35R is VACV A33R ortholog in MPXV. VACV A33R protein is expressed on the surface of EEV but not IMV. A33R protein is a type II integral membrane glycoprotein with a hypothetical C-type lectin domain. In addition, the A33R protein also regulates the interaction of A36R with microtubule motor proteins. In addition, A33R is a target for neutralizing antibody responses against EEV in the presence of complement.

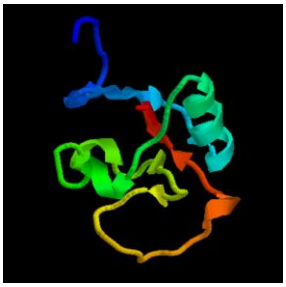


Figure 6 A35R structure analysed in Rasmoll.

***GMQE** 0.36*

***QMEANDisCo Global:** 0.79 ± 0.09*

B6R

B5R protein is essential for wrapping IMV and for EEV to induce the formation of actin tails on the host cell surface expressing A33 and A36 proteins. The EEV B5R is the major target of EEV-neutralizing antibodies after smallpox vaccination as demonstrated by antibody depletion experiments

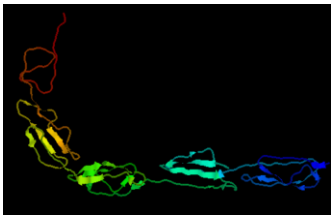


Figure 7 B6R structure analysed in Rasmoll.

***GMQE** 0.40*

***QMEANDisCo Global:** 0.56 ± 0.06*

A29L

MPXV A29L protein is a homolog of VACV A27L. The C-terminal leucine zipper structural domain (80-101 aa) of A27L protein attaches to the viral membrane, whereas the N-terminal region (21-32 aa) enables attachment to host cells via its GAGs-binding domain

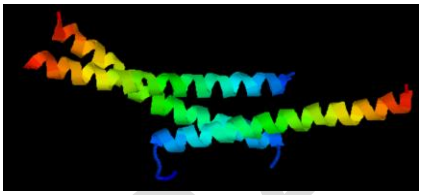


Figure 8 A29L structure analysed using Rasmoll.

***GMQE** 0.30*

***QMEANDisCo Global:** 0.67 ± 0.07*

M1R

The MPXV M1R is the homologous protein of VACV L1R. VACV L1R is located on the membrane of IMV. L1R outer domain, facing the cytoplasm in intracellular viruses, contains three intramolecular disulfide bonds. The L1R attaches to the cell surface by binding to non-GAG molecules on the cell surface and adheres to the viral membrane via a C-terminal transmembrane anchor.

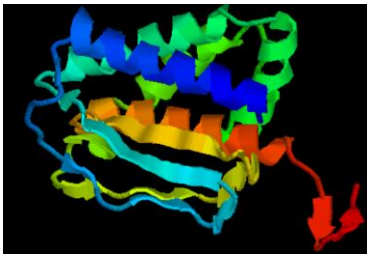


Figure 9 M1R analysed using Rasmoll.

***GMQE**0.65*

***QMEANDisCo Global:**0.86 ± 0.06*

Conclusion

Genomic and DNA analyses of the monkeypox virus, combined with protein modeling, provide valuable insights into its pathogenesis and potential therapeutic targets. Continued research in these areas is vital for developing effective treatments and vaccines, particularly in light of the ongoing global outbreaks.

Primary amoebic meningoencephalitis (PAM)

INTRODUCTION

Primary amoebic meningoencephalitis (PAM) is a rare, often fatal infection of the central nervous system caused by the amoeba *Naegleria fowleri*, which can enter through the nose during swimming in warm, contaminated fresh water. Symptoms progress rapidly from changes in smell or taste, headaches, and nausea to confusion and death. Diagnosis typically involves a spinal tap to analyze cerebrospinal fluid and possibly a brain biopsy. Treatment is challenging and usually involves a combination of medications, including miltefosine, Amphotericin B, Rifampin, and azithromycin, among others.

GENOMIC STRUCTURE

Naegleria fowleri, a member of the Heterolobosea class, causes primary ameobic meningoencephalitis, a typically fatal brain infection. It contains a single ribosomal DNA (rDNA) cistron (5.8S, 18S, and 28S rRNA genes) on closed circular extrachromosomal rDNA elements (CEREs), with approximately 4,000 copies per cell. Genome sequencing of Naegleria species confirmed that rDNA genes are absent from the nuclear genome. This report presents the complete sequence of the N. fowleri (strain LEE) CERE, highlighting the full rRNA coding sequences and identifying five putative open reading frames (ORFs) along with repeated sequences in the non-ribosomal sequence (NRS).

Genome size 29.5 Mb

Naegleriapore a

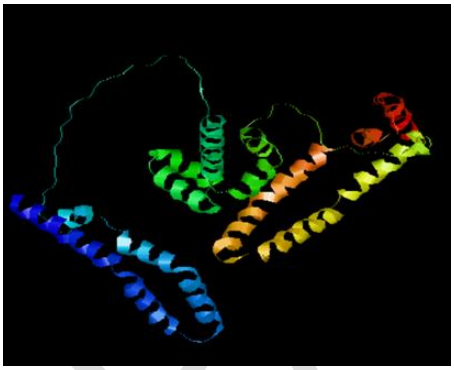


Figure 10naegleriapore a

Biounit OligoState-Monomer

METHOD -AlphaFold v2

Seq Similar - 0.61

Coverage- 1.00

GMQE - 0.74

Heat Shock Protein

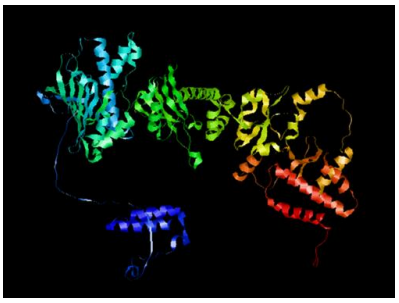


Figure 11heat shock protein

Biounit OligoState-Monomer

Method -AlphaFold v2

Seq Similarity-0.57

Coverage - 0.99

Range -2-747

GMQE-0.82

Clamodulin-dependent protein

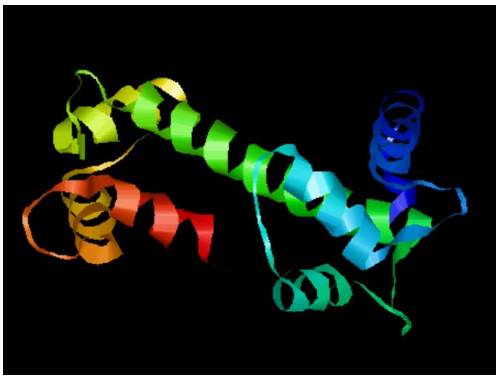


Figure 12 calmodulin-independent protein

Biounit Oligostate-Monomer

Method-AlphaFold v2

Seg Similar-0.60

Coverage-1.00

Range-1-149

GMQE-0.86

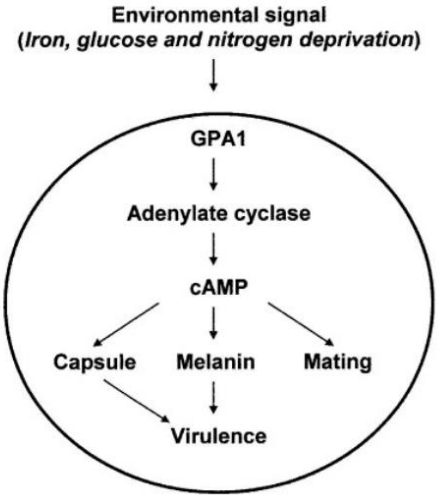
CONCLUSION

Genomic and DNA analyses of PMA, alongside protein modelling, shed light on its pathogenesis and highlight potential therapeutic targets .Ongoing research in these fields is crucial for advancing treatment strategies. However, it's also essential to explore integrative approaches that consider environmental and host factors, which may further enhance our understanding and treatment efficacy.

Cryptococcus neoformans

- **Introduction:** *Cryptococcus neoformans* is a pathogenic yeast found in soil and decaying wood. It primarily affects immunocompromised individuals, with HIV infection being the leading risk factor. It is acquired through the respiratory route, where the fungi are inhaled, initially affecting the lungs and potentially spreading to the central nervous system (CNS) as cryptococcal meningitis or to the bloodstream as cryptococcaemia. Its most common manifestation is **cryptococcal meningitis**, which is life-threatening. Mortality rates range from 41% to 61%, particularly in HIV-positive patients. Complications include **raised intracranial pressure**, which can lead to blindness, and acute renal impairment due to treatment toxicity.
- **Serotypes and Genome:** Most isolates of *C. neoformans* are haploid. The size of the genome is approximately 19 Mb with 14 chromosomes. *Cryptococcus neoformans* is traditionally classified into three varieties with five serotypes: var. *grubii* (serotype A), var. *neoformans* (serotype D), var. *gattii* (serotypes B and C), and serotype AD (hybrid of serotypes A and D). In an analysis of 36 North Indian isolates, most of them, i.e., 31 (87%), were found to belong to serotype A (var. *grubii*). All those strains possessed the MAT-alpha mating type.
- **Virulence:**
    - Many studies have sought to correlate individual phenotypes with virulence, the most common being capsule size. Other studies found that strains with increased capsule shedding were associated with higher patient mortality and that neurovirulence may be related to the total amount and speed of accumulation of capsule in the brain rather than the capsule size of individual cells. The disparate nature of these results likely reflects a simple association with a single phenotype driving virulence but it is likely a combination of multiple phenotypes that determines the overall virulence profile of a strain.
    - In a current study in a set of strains with high genetic similarity, hypervirulent strains were associated with larger average capsule, increased cell size variation, and increased production of microcells, released capsule, and clustered capsule. In addition to the individual phenotypes themselves being associated with various routes of pathogenesis, increased morphological heterogeneity as a whole may contribute to virulence via immune escape.
    - Several properties related to virulence have been attributed to the cryptococcal GXM, which is the capsular polysaccharide of ***C. neoformans.***

- Melanin production has also been associated with cryptococcal virulence. Several studies have shown that spontaneous or ultraviolet induced mel^- mutants had a reduced or no capacity to kill mice when compared with Mel^+ strains. The role of the *CNLAC1* gene (the structural gene encoding laccase) in the virulence of *C. neoformans* has been suggested after its cloning and genetic deletion in knockout strains. Virulence of the knockout strains was significantly reduced as compared with that of Mel^+ strains in animal models, confirming that laccase and/or enzyme products are responsible for the pigment-associated virulence demonstrated in earlier studies.
- Additional structures have been identified, but their relation with cryptococcal virulence is uncertain. For example, phospholipase activity was detected in several strains of *C. neoformans.* A superoxide dismutase activity was demonstrated in *C. neoformans*, which could be involved in the neutralization of free radicals, similar to melanin and mannitol. Proteolytic activities and some new extracellular and cell-associated enzymatic activities have recently been demonstrated in some strains of *C. neoformans.*
- Recent studies have also attributed to role of a G-protein alpha-subunit homolog (GPA1) in sensing diverse exogenous signals required for mating and virulence, through regulation of intracellular cAMP. Disruption of the gene encoding GPA1 resulted in a mutant (*gpa1*) viable but deficient in mating, in response to nitrogen starvation. Under the same conditions, melanin synthesis and capsule production were also inhibited in the *gpa1* mutant. Virulence of the *gpa1* mutant was also significantly reduced in a rabbit model of cryptococcal meningitis, and reintroduction of the wild-type *GPA1* gene complemented the *gpa1* mutant phenotype and restored mating and melanin and capsule production, as well as virulence. Interestingly, the addition of exogenous cAMP suppressed the *gpa1* mutant phenotype, restoring mating and melanin and capsule production.



Visualisation of *CNLAC1* gene through Rasmol

<table>
<tr>
<td>

- *Molecule information:*
- *No of Groups: 109*
- *No of Atoms: 883*
- *No of Bonds: 916*
- *Linear*
- *112 aa*
- *QMEAN:* 0.71 ± 0.08

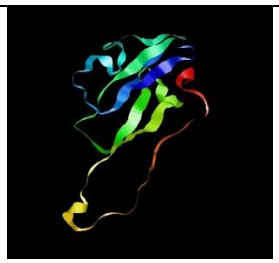
</td>
<td>


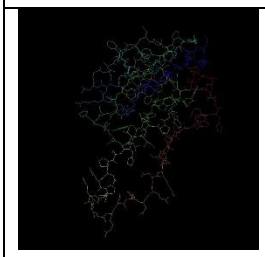
Figure 13: Display Type-Ribbon Structure
</td>
</tr>
<tr>
<td>


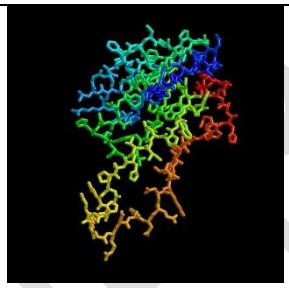
Figure 14:Display Type-Wireframe
</td>
<td>


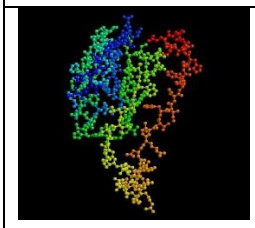
Figure 15: Display Type-Stick
</td>
</tr>
<tr>
<td>


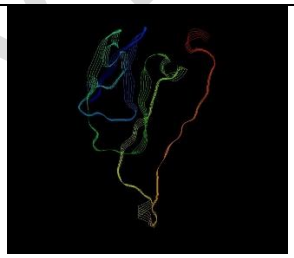
Figure 16: Display Type-Ball and Stick
</td>
<td>


Figure 17: Display Type-Strand
</td>
</tr>
</table>

Dengue

Dengue virus belongs to the genus *Flavivirus* in the family Flaviviridae. It is a positive-stranded encapsulated ribonucleic acid (RNA) virus that is composed of three structural protein genes that encode the nucleocapsid or core protein, a membrane-associated protein, an enveloped glycoprotein, and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) known for causing pathogenecity. It is transmitted mainly by the *Aedes aegypti* mosquito and also by the *Aedes albopictus* mosquito. There are four antigenetically related but distinct serotypes of the dengue virus: DENV-1, DENV-2, DENV-3, and DENV-4. Each serotype has several subtypes or genotypes. In humans, one serotype produces lifelong immunity against reinfection but only temporary and partial immunity against the other serotype. Each serotype has unique characteristics and can present with severe manifestations in a particular population depending upon its interaction with the host response.

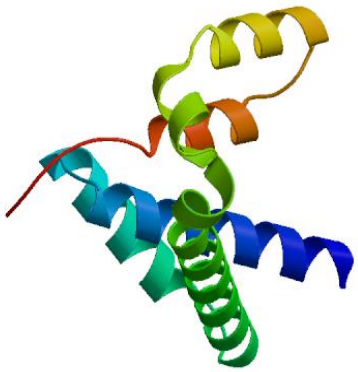
Minister of state for Health and Family Welfare reported that over 32,000 dengue cases have been reported across India in 2024 to date. In August, India has seen an almost 50% rise in the number of dengue cases reported this year, as compared to the same period in 2023. The Lancet editorial identifies “the triad of urbanisation, climate change and the movement of people and goods” as facilitating th spread of dengue and its mosquito vector.

Protein modelling is useful for understanding the structure and function of dengue proteins which can help develop treatments and vaccines.

The non- structural proteins are responsible for the pathogenesis. Some of them are-

NS2A -

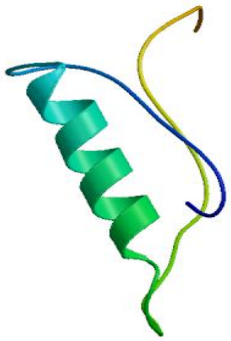
- **Viral Replication**: NS2A is essential for the replication and assembly of the dengue virus, facilitating the proper folding and maturation of other viral proteins.
- **Membrane Remodeling**: It is involved in altering cellular membranes, creating a favorable environment for viral replication and assembly within the host cell.



NS5 -

 RNA Polymerase Activity: NS5 is responsible for the replication of the viral RNA genome. Its RNA-dependent RNA polymerase function is essential for producing viral progeny, thereby facilitating infection and spread.

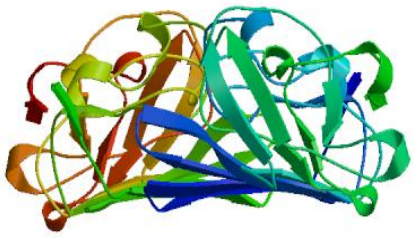
 Modulation of Host Cell Processes: NS5 can manipulate host cell functions to create a more favorable environment for viral replication. This includes altering signaling pathways and potentially affecting apoptosis.



NS1-

 Immune Evasion: NS1 helps the virus evade the host immune response by modulating the complement system and inhibiting the activation of immune cells.

 Vascular Permeability: It can increase vascular permeability, contributing to plasma leakage and hemorrhagic symptoms seen in severe dengue cases.

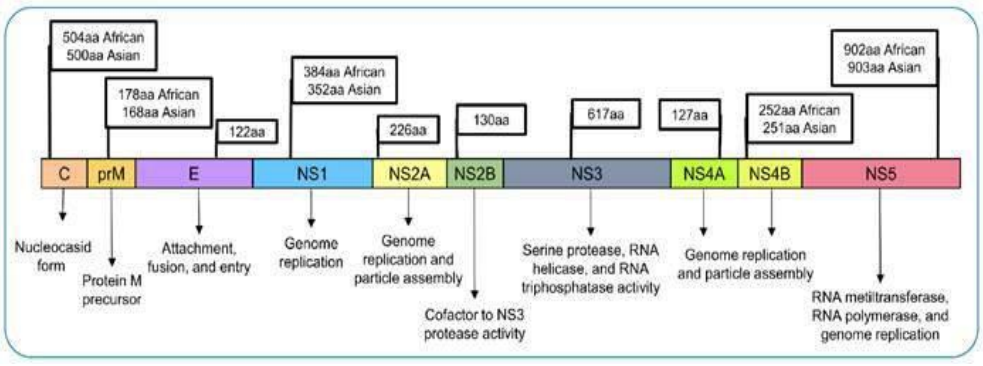


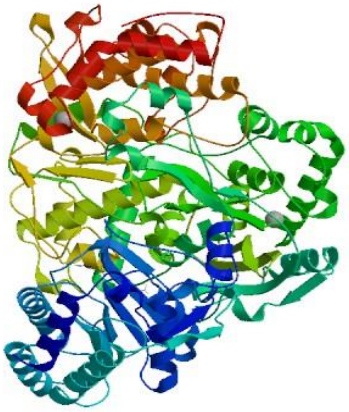
Zika Virus

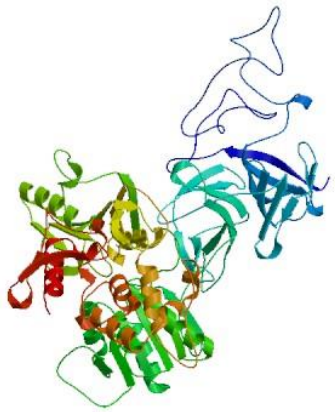
Zika virus is a mosquito-borne virus first identified in Uganda in 1947 in a Rhesus macaque monkey followed by evidence of infection and disease in humans in other African countries in the 1950s.Zika virus is transmitted primarily by ***Aedes aegypti*** mosquitoes.Zika virus is a single-stranded RNA virus of the Flaviviridae family, genus Flavivirus.The ZIKV genome spans approximately 10.8 kb in length. The Zika virus genome consists of a single-stranded RNA molecule that encodes for 3,419 amino acids, which are organized into three structural proteins (C, prM, and E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5).

The ZIKV genome consists of 10,794 nucleotides in a single-stranded positive-sense RNA . ZIKV RNA has two untranslated regions (UTRs) and a single open reading frame (ORF). Zika virus infection during pregnancy is a cause of microcephaly and other congenital malformations in the infant.Zika virus infection can also cause Guillain-Barré syndrome, neuropathy and myelitis, particularly in adults and older children.As of August 16,2024, India reported 113 confirmed cases of Zika virus, with 100 of those cases in Pune district.

Genomic structure of Zika virus



Genome Polyprotein structure Accession - XER79097 GMQE-0.00 QMEANDisCo Global:0.63 ± 0.05 Seq Identity-64.83% Amino acids- 3421	
NS5 protein NS5 is comprised of two domains, an N-terminal methyltransferase domain and an RNA-dependent RNA polymerase (RdRP) domain at the C-terminal end . It performs three essential roles in the viral life cycle: genome replication,capping and interferon suppression. Accession - 6WCZ_B GMQE-0.86 QMEANDisCo Global:0.82 ± 0.05 Seq Identity-100.00% Amino acid - 903	

Genomic Polyprotein structure Accession - XDO24423 GMQE - 0.00 QMEANDisCo Global: 0.63 ± 0.05 Seq Identity-65.42% Amino acids - 3423	

SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2

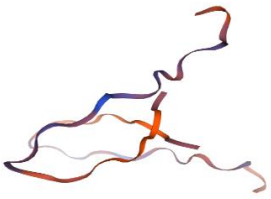
INTRODUCTION

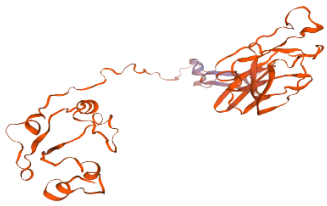
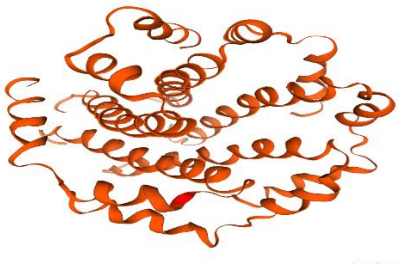
Severe acute respiratory syndrome coronavirus 2 is a strain of coronaviruses which is a group of related positive single-stranded RNA viruses that causes respiratory diseases in mammals and birds. It is thought to be of natural animal origin, by a spillover infection. SARS-CoV-2 strain is a descendant of a coronavirus that infects wild bats that spread to humans through an intermediary wildlife host likely due to the live wildlife trade at the Huanan wet market in the city of Wuhan, China where the first case was reported on December 31, 2019. The virus spreads through virus-laden fluid created in an infected person's respiratory tract. In the early phase of the infection, viral replication results in direct virus-mediated tissue damage. In the late phase, the infected host cells trigger an immune response by releasing cytokines. In severe COVID-19 illness a cytokine storm is seen. The sudden release of cytokines may cause multisystem organ failure. COVID-19 affects several organs in one's body, but mainly affects the lungs and upper and lower respiratory tracts. Over 775 million cases were reported and more than 7 million have died globally and in India over 45 million cases were reported and more than 5 lakh people were killed by the virus.

GENOME

Coronaviruses contain a positive-sense, single-stranded RNA genome. The genome size is 29.9 kb with one chromosome. The genome size is one of the largest among RNA viruses. The genome has a 5'methylated cap and a 3'polyadenylated tail. The genome organisation for a coronavirus is 5'-leader-UTR-replicase (ORF1ab), spike(S), envelope(E), membrane(M), nucleocapsid(N)-3'UTR-poly (A) tail. The ORF1a and ORF1b , which occupy the first two-thirds of the genome encode the replicase polyprotein (pp1ab). The replicase polyprotein self cleaves to form 16 non-structural proteins (nsp1-nsp16). The later reading frames encode the four major structural proteins: spike, envelope, membrane and nucleocapsid. Interspersed between these reading frames are the reading frames for the accessory proteins. Notable variants are Alpha(B.1.1.7), Beta(B.1.351), Gamma(P.1), Delta(B.1.617.2) omicron(B.1.1.529)

PROTEINS

PROTEIN	STRUCTURE
ORF7b protein ORF7b protein, is one of the structural accessory proteins in SARS-CoV-2 and possesses a transmembrane helical domain (TMD) that is necessary for its Golgi complex localization **GMQE-0.39** **QMEANDisCo Global- 0.86**±0.12 **Seq Identity-23.08%** **Amino Acids-43**	
ORF8a ORF8 protein, a viral accessory protein that has been proposed to interfere with immune responses, consists of an N-terminal signal sequence followed by an immunoglobulin (Ig)-like fold, similar to the ORF8 Ig-like domain protein of SARS-CoV-2 **GMQE-0.31** **QMEANDisCo Global-0.58**±0.12 **Seq Identity-34.78%** **Amino Acids-39** **ORF8b** **GMQE-0.06** **QMEANDisCo Global-0.11**±0.12 **Seq Identity-19.23** **Amino Acids-84**	 

ORF10 ORF10 protein which contains eleven cytotoxic T lymphocyte (CTL) epitopes, each nine amino acids in length, and may suppress the antiviral innate immune response by degrading mitochondrial antiviral signalling protein (MAVS) through mitophagy **GMQE-0.37** **QMEANDisCo Global-0.37**±0.08 **Seq Identity-25%** **Amino Acids-38**	
SPIKE PROTEIN **The spike protein plays a key role in the receptor recognition and cell membrane fusion process.** **GMQE-0.49** **QMEANDisCo Global-0.38**±0.05 **Seq Identity-23.73** **Amino Acids-298**	
ENVELOPE PROTEIN envelope (E) protein is the envelope small membrane protein that plays a central role in virus morphogenesis and assembly, such as the E proteins from three highly pathogenic human coronaviruses. **GMQE-0.32** **QMEANDisCo Global-0.36**±0.05 **Seq Identity-25.86** **Amino Acids-298**	

MEMBRANE PROTEIN

membrane (or matrix) glycoprotein is a component of the viral envelope that plays a central role in virus morphogenesis and assembly via its interactions with other viral proteins; similar to the Membrane (M) protein of coronaviruses (CoVs) such as Middle East respiratory syndrome (MERS)-related CoV, severe acute respiratory syndrome (SARS) CoV, and SARS-CoV-2

GMQE-0.69

Seq Identity-91.23%

Amino Acids-57

Helicobacter pylori

INTRODUCTION

Helicobacter pylori belongs to the family helicobacteraceae and is a bacterial infection. It is a gram negative spiral bacterium and it causes gastritis and peptic ulcer. This bacterium was first identified in 1983 by Barry Marshall and Robin Warren as the common cause of gastric ulcer in humans. In 2005 they were awarded noble prize in physiology for this discovery. It is a RNA bacterium with a genomic size of 1.7Mb. It can be transmitted from person to person by saliva or by fecal contamination with food or water. Poor hygiene also contributes to higher helicobacter pylori prevalence. Mostly people get infected as children and parents play a important role in their transmission. Its symptoms include bloating, nausea, indigestion, loss of appetite and dark stools.

SPREAD OF INFECTION

As this bacteria grows in the mucus layer of the stomach so it infects the first part of small intestine or duodenum and it causes inflammation in the body. It uses its tail like body to move around and burrow under the stomach lining. It releases urease to neutralize the stomach acid and allowing it to survive in stomach acidic environment.

HELICOBACTER PYLORI PREVALENCE IN INDIA

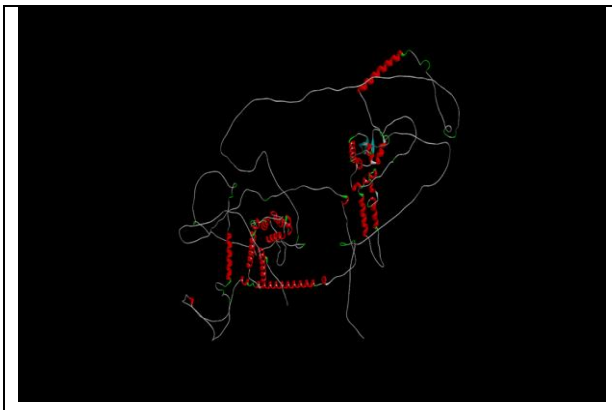
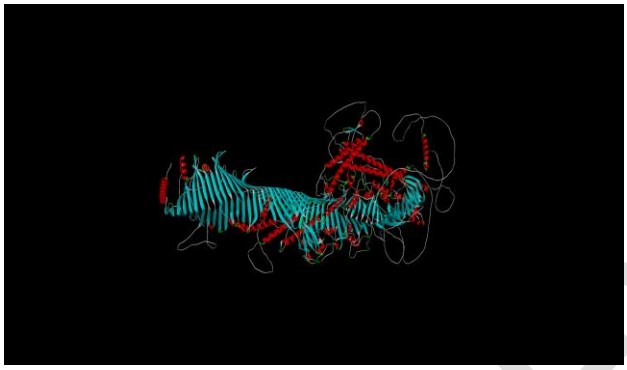
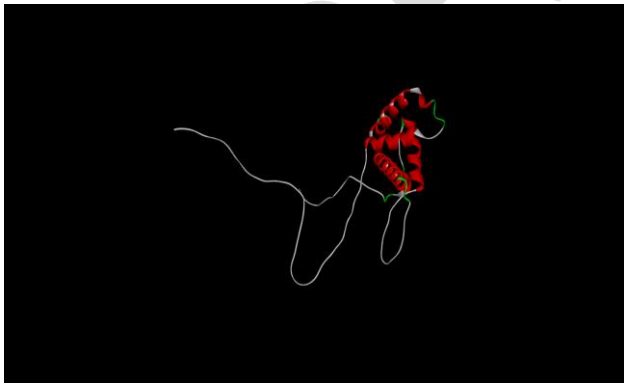
The cases of gastric cancer are low in India but the commonness of gastric cancer is extremely high in India ranging from 49.94% to 83.30%. It was noted that the helicobacter pylori and gastric cancer relation was found positive in 50% of the patients while the remaining 50% showed negative relationship.

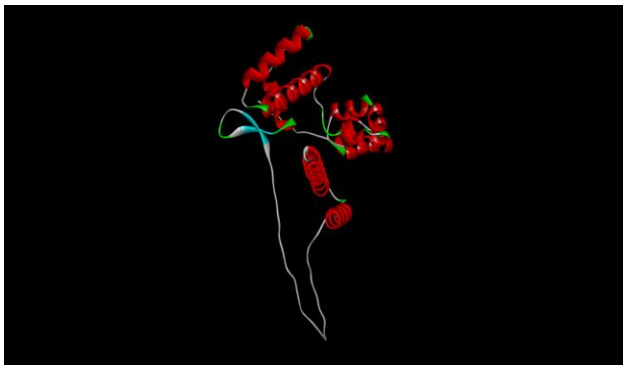
HELICOBACTER PYLORI PREVALENCE IN THE WORLD

According to the statistics from 1980 to now the global prevalence of helicobacter pylori has reduced. The most cases of helicobacter pylori has been seen in African region from the 1980s and the least cases were reported in Switzerland. The infection has decreased in the adults from 52.6% to 43.9% but is still higher in children and adolescents with 35.1%.

PROTEIN MODELLING

Non structural and structural proteins both are responsible for causing infection in humans. These proteins are:- Bab A, Sab A, Alp A, Oip A, Hop Q, Lipopolysaccharides (LPS), Cag A, Vac A. Some of these are shown below:-

	**Bab A** A blood group antigen binding adhesin that belongs to the Hop protein family. It is important for the early stages of infection, and contributes to helicobacter pylori persistence and the development of gastric disease. GMQE- 0.49 Sequential Identity- 72.42%
	**Sab A** A salic acid binding adhesin. Sab A is an outer membrane protein that helps helicobacter pylori adhere to the stomach surface, which is important for colonization. It interacts with sialylated Lewis antigens in inflamed gastric tissue. GMQE- 0.64 Sequential Identity- 100.00%
	**Alp A** These proteins are important for colonization and may bind to laminin and collagen IV in the extracellular matrix. GMQE- 0.61 Sequential Identity- 94.48%

	**Cag A** A protein that's directly associated with the development of gastric ulcer, gastric cancer, and acute gastritis. Cag A acts as a scaffold protein that interacts with multiple host signaling molecules, including SHP2 and PAR1/MARK. This perturbs host signaling pathways, leading to abnormal cell growth and destruction of epithelial cell polarity. GMQE- 0.76 Sequential Identity- 98.67%

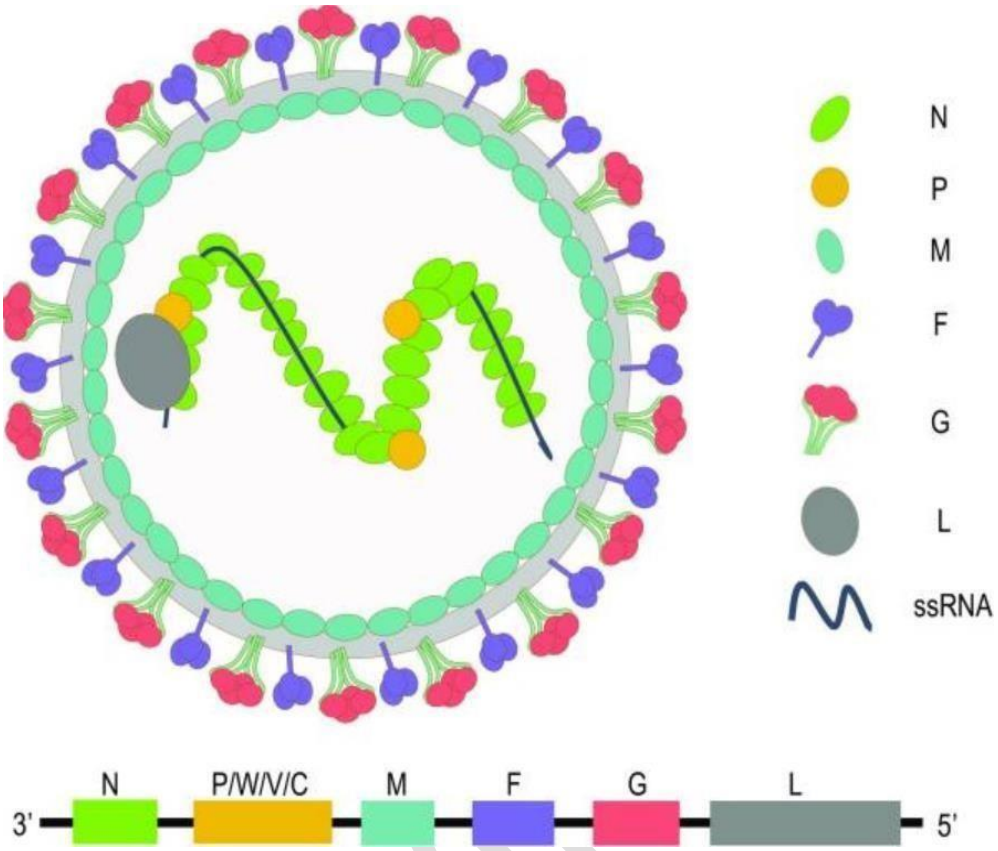
NIPAH VIRUS

INTRODUCTION

NiV was named after Kampung Sungai Nipah (Nipah River Village) in Malaysia, where it was first isolated in 1998, before its subsequent spread into Singapore via exported pigs in 1999, leading to the abattoir worker infections

In 2001, human cases of NiV infection were discovered independently in India and Bangladesh, and since then, infections have been observed annually in Bangladesh, and human-to-human transmission through direct contact with infected individuals is common. In 2014, a serious illness most probably caused by NiV was reported in several people after contact with infected horses or patients in the Philippines. NiV infection can cause fever and encephalitis in humans and a neurological and respiratory syndrome in pigs or horses. To date, over 600 human cases of NiV infection have been reported in South Asia and South-East Asia, with fatality ranging from 40% to 70%, accordingly it poses a major threat to human health

Belonging to the genus *Henipavirus* [the other pathogenic member of the genus is Hendra virus (HeV), reviewed in of the family *Paramyxoviridae*, NiV is classified as a Biosafety Level-4 (BSL-4) pathogen due to its high pathogenicity and the lack of any effective treatments or vaccines. The NiV genome consists of a negative-sense, single-stranded RNA of approximately 18.2 kb, encoding six structural proteins, nucleoprotein (N), phosphoprotein (P), matrix protein (M), fusion protein (F), attachment glycoprotein (G), and the large protein or RNA polymerase protein (L). In addition, the *P* gene encodes three nonstructural proteins by RNA editing (V and W proteins) or an alternative open reading frame (C protein)



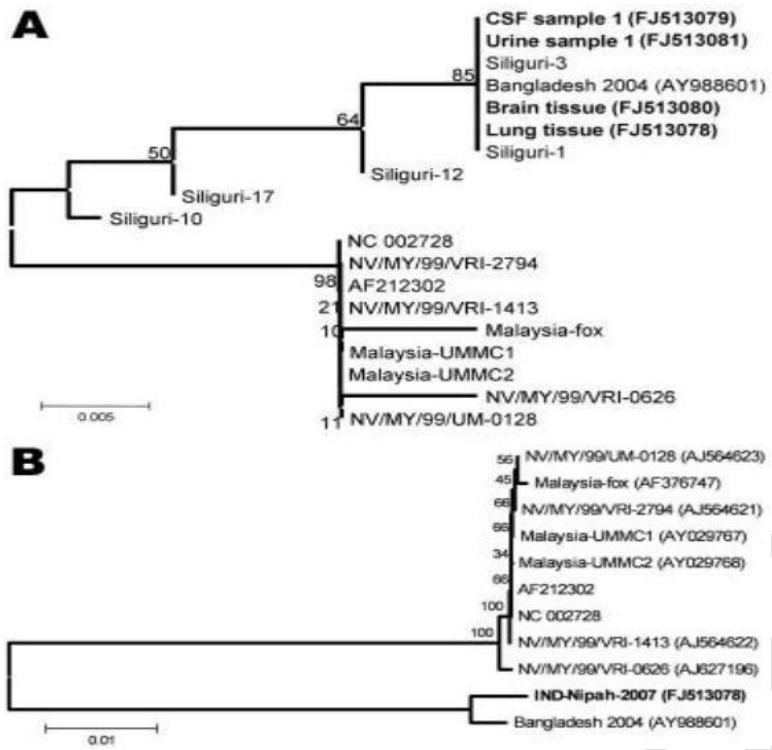
Schematic representation of the viral structure (upper panel) and genome organization (lower panel). Different genes or proteins are indicated in different colors.

NiV has a wide range of hosts, from its natural reservoir *Pteropid* bats to humans, horses, dogs, cats, cows, and pigs. Close contact with infected patients or domestic animals (e.g., pigs and horses) plays an important role in the spread of NiV. Intraspecific transmission (in bats, pigs, and horses) is also possible via saliva, urine or secretions upon high density populations of animals

GENOMIC STRUCTURE

The length of the full genome of the isolate from India was 18,252 nt. The sequence of this virus (INDNipah-07–1, GenBank accession no. FJ513078) was closer to the virus from Bangladesh, with 99.2% (151 nt substitutions) and 99.80% (17 aa substitutions) identity at nucleotide and amino acid levels respectively. Of the 151 nt substitutions, 9 occurred in the N open reading frame (ORF),11 in the phosphoprotein ORF, 8 in the matrix ORF, 11 in the fusion glycoprotein ORF, 7 in the attachment protein ORF, and 47 in the large polymerase ORF. Fifty-eight substitutions occurred in nontranslated regions at the beginning and the end of each ORF. The intergenic sequences between gene boundaries were highly conserved in the isolate from India,

compared with the isolate from Bangladesh, which showed 1 change (GAA to UAA) between the attachment protein and large polymerase genes. No change was observed in the leader and the trailer seq.



A) Phylogenetic analysis based on partial nucleocapsid (N) gene nucleotide sequences (159 nt, according to Nipah virus [NiV] Bangladesh sequence, GenBank accession no. AY988601, 168–327 nt) of the 4 NiVs sequenced during this study (**boldface**). Five sequences of the viruses from Siliguri (*8*) and from representative NiV sequences obtained from GenBank indicated by the respective accession numbers. Values at different nodes denote bootstrap support. B) Full genome–based phylogenetic analysis of the NiV sequenced from the lung tissue of a patient (**boldface**). Representative NiV sequences obtained from GenBank are indicated by the respective accession numbers. Values at different nodes denote bootstrap support. Scale bars indicate nucleotide substitutions per site.

The Table compares amino acid substitutions in the different regions of the genome of the isolate from India with those of the viruses from Bangladesh and Malaysia. Of the 17 aa substitutions, 7 were unique to the isolate from India, and 10 were similar to the isolates from Malaysia. Overall, however, the isolate from India was closer to the isolate from Bangladesh, although distinct differences were observed.

Regionwise amino acid substitutions in the Nipah virus genome*

Region and amino acid position	India	Bangladesh	Malaysia

Phosphoprotein 228 K R R 276
S G G 285 R H R 310 R
G G

Nucleocapsid protein 188 E D E
211 R Q Q

Matrix protein 13 I M M

Fusion protein 19 I M M 207 L S L 252 D
G D

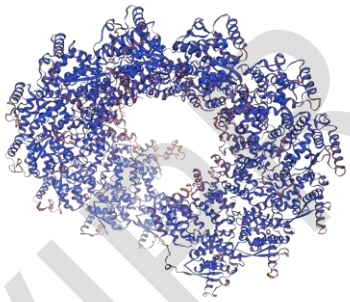
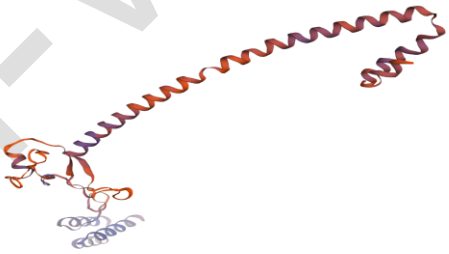
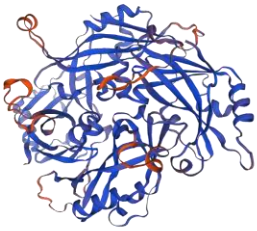
Attachment protein 304 V I I

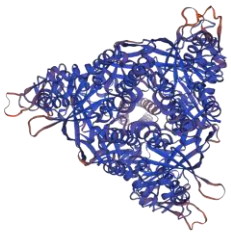
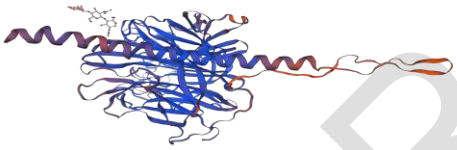
Large polymerase protein 94 I T I
112 K R K 632 N S N
639 N D N 665 T I T 1748 I V I

VIRAL PROTEIN FUNCTIONS AND STRUCTURE

Nipah virus has an approximately 18.2 kb genome encoding six structural proteins and three nonstructural proteins. The viral ribonucleocapsid (RNP) surrounded by the viral envelope consists of its genome and the N protein, which is essential for the viral life cycle as a template for RNA-dependent RNA-polymerase (RdRp), composed of polymerase L and a polymerase cofactor P. Within the RNP, N is responsible for viral genome wrapping and facilitates viral replication and transcription. The synthesis of viral mRNA is catalyzed by L and P, and the latter also inhibits interferon signaling via host STAT-1 and acts as a chaperone of N^0 (the unassembled form of N) to prevent it nonspecific binding to host RNA. The M protein contributes to viral assembly and release. G and F are two important surface glycoproteins of NiV; the former induces viral attachment to two cellular receptors, ephrin-B2 and ephrin-B3, despite a lack of hemagglutination or neuraminidase activity. And this subsequently triggers F-mediated membrane fusion between virus and host cells. The nonstructural protein C participates in the host immune response and serves as a virulence factor

The crystal structure of the P protein is a tetramer with a parallel coiled coil, while the N^0-P complex, whose binding site is located in residues 1–50 (P_{50}) of the N-terminal domain of P, is characterized by an asymmetric pea-like form composed of three heterodimers. N^0 remains an open conformation in the complex due to P-mediated inhibition of the polymerization of N

Nucleoprotein GMQE- 0.75 QMEANDisco Global- 0.81 ± 0.05 Seq. Identity- 98.68%	
Phosphoprotein GMQE- 0.38 QMEANDisco Global- 0.59 ± 0.06 Seq. Identity- 90.55%	
Matrix Protein GMQE- 0.83 QMEANDisco Global- 0.83 ± 0.05 Seq. Identity- 100.00%	

Fusion Protein GMQE- 0.76 QMEANDisco Global- 0.82 ± 0.05 Seq. Identity- 99.59%	
Attachment Glycoprotein GMQE- 0.78 QMEANDisco Global- 0.82 ± 0.05 Seq. Identity-95.68%	
Large protein GMQE-0.78 QMEANDisco Global- 0.60± 0.11 Seq. Identity-0.78