Original article

NipahVR: a resource of multi-targeted putative therapeutics and epitopes for the Nipah virus

Amit Kumar Gupta, Archit Kumar, Akanksha Rajput, Karambir Kaur, Showkat Ahmed Dar, Anamika Thakur, Kirti Megha and Manoj Kumar*

Virology Unit and Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39-A, Chandigarh 160036, India

*Corresponding author: Tel: 91-172-6665453; Fax, 91-172-2690585, 91-172-2690632; Email, manojk@imtech.res.in

Abstract

Nipah virus (NiV) is an emerging and priority pathogen from the *Paramyxoviridae* family with a high fatality rate. It causes various diseases such as respiratory ailments and encephalitis and poses a great threat to humans and livestock. Despite various efforts, there is no approved antiviral treatment available. Therefore, to expedite and assist the research, we have developed an integrative resource NipahVR (http://bioinfo.imtech.res.in/manojk/nipahvr/) for the multi-targeted putative therapeutics and epitopes for NiV. It is structured into different sections, i.e. genomes, codon usage, phylogenomics, molecular diagnostic primers, therapeutics (siRNAs, sgRNAs, miRNAs) and vaccine epitopes (B-cell, CTL, MHC-I and -II binders). Most decisively, potentially efficient therapeutic regimens targeting different NiV proteins and genes were anticipated and projected. We hope this computational resource would be helpful in developing combating strategies against this deadly pathogen. Database URL: http://bioinfo.imtech.res.in/manojk/nipahvr/

Introduction

Nipah virus (NiV) is a highly pathogenic virus closely related to the Hendra virus (HeV) from the genus *Henipavirus* of the family *Paramyxoviridae*. NiV is an enveloped single-stranded RNA (ssRNA), a negative-sense virus of size ∼18 250 nucleotides. It encodes nine proteins namely nucleoprotein (N), four proteins encoded by P gene (phosphoprotein (P), W, V, C protein), matrix (M) protein, fusion (F) glycoprotein, attachment glycoproteins (G) and a large polymerase (L) protein (1–3).

Different NiV proteins play a cardinal role in viral infection and disease manifestation. Viral G protein (attachment protein) first binds to the Ephrin-B2 or B3 cellular receptors found on the neuron, smooth muscles, capillaries and

arterial endothelial cells. G protein provides attachment to the host cell surface, which triggers the fusion by the F protein (4,5). Subsequently, F protein, a 546 amino acid (aa)-long type I transmembrane protein, mediates fusion of virus and host cell membranes and mediates cell entry (6). Then, viral RNA content gets synthesized and translated into the proteins. P gene encodes four gene products. A structural P protein (709 aa) is essential for genome replication encoded by unedited mRNA and localized in the cytoplasm. The three additional non-structural proteins (V, W and C) contribute towards the evasion of the innate immune response through inhibiting the different signaling pathways and are crucial for the viral infection. V and W proteins are produced by RNA editing and localized in the cytoplasm and nucleus, respectively, and the second open reading frame (ORF) generates C protein. M protein (352 aa) has a crucial role in viral budding. It also provides firmness to virion through interacting with envelope and F protein. M protein is also known to hijack cellular pathways and machines to facilitate nuclear localization. N protein (532 aa) is mainly responsible for the viral genome encapsidating. The largest NiV protein L retains all the enzymatic functions like genome replication and transcription for the viral RNA synthesis (1–3).

NiV is an emerging zoonotic virus classified as a category C priority pathogen and biosafety level-4 (BSL-4) agent that signify a rolling threat to humans and animals worldwide (7,8). It was originated and first isolated from the village 'Sungai Nipah' (7,8). It is an etiological agent of diverse diseases such as encephalitis, respiratory illness and fever (9,10). Epidemiologically, NiV transmission occurs mainly through infected animals (bats, pigs, etc.) and contaminates food consumption (11) and may spread through person to person (12). Flying foxes (fruit bats) from the genus *Pteropus* are known as a natural reservoir and host of the NiV (13). Among all *Pteropus* species, *P. giganteus* (Indian flying fox) mainly distributed in south Asian regions like Bangladesh, India and Pakistan (14,15). Other species were also found in different parts of Southeast Asia like *P. vampyrus* and *P. hypomelanus* in Malaysia and *P. lylei* in Thailand and Cambodia (16,17).

Up to now, various sporadic outbreaks were reported from different countries, mainly from South Asia, i.e. India, Bangladesh, Malaysia and Singapore, since the first incidence of Malaysia in 1998 with the high mortality rate between 40 and 75% depending on clinical manifestations (3,7,8,18,19). These are mainly endemic in India and Bangladesh (7,8,20,21). In India, the first outbreak was reported from the Siliguri, West Bengal, in 2001 with high fatalities due to NiV encephalitis (18,22). In this, the involvement of pigs as an infection mediator is not observed, and direct person-to-person transmission was reported that signify high risk to public health (22). Later in 2007, another outbreak was reported from West Bengal with 100% mortality (23). Very recently, in May 2018, the first NiV outbreak occurred in southern India in Kozhikode and Malappuram districts of Kerala. Several deaths were reported due to the unavailability of a practical solution, which is of concern to India and the world (7,8,24,25). Moreover, a study also describes the presence of NiV RNA in *Pteropus giganteus* in different Indian states signifying it as a natural reservoir in India (15). However, more surveillance studies are necessary to access the NiV outbreak risk among susceptible populations living in different geographical locations (7). Furthermore, several NiV outbreaks were also documented from Bangladesh between 2001 to 2015 (7,11,26). These are linked to many deaths due to encephalitis with neurological and respiratory complications (11,26). Moreover, different studies also provide information about NiV origin, evolution and stability over time (7,8). For example, a study shows the conservation between the isolates from Bangladesh, 2004, and India, 2007, with 99.2 and 99.8%, nucleotide and amino acid similarity, respectively (23). Similarly, another study also provides phylogenetic analysis and conservation of NiV, i.e. between 96 and 100% (25). Moreover, a recent report by Ravichandran *et al.* has also found conservancy among NiV proteins (27).

There are also efforts to combat the NiV, and different strategies (vaccines, immunotherapies, antiviral drugs) were tried to eradicate the infection (28). Distinct approaches like subunit vaccine (29), vectored vaccine and live-vectored vaccine (30) mainly utilizing the G and P proteins demonstrated to elicit an immune response (28,31). Likewise, a virus-like particle (VLP)-based vaccine is also shown protection against NiV (32). These experimental vaccines are mainly tested on animal models like a hamster, ferret, cats and pigs (28,31,32). Moreover, a subunit vaccine for use in horses has been developed based on the HeV G protein (28,31,33). Additionally, the use of monoclonal and polyclonal antibodies also showed success in treating NiV infection in animals (34). However, more studies, i.e. in vitro as well as in vivo, will be required before conducting clinical trials.

Likewise, other strategies such as RNA interference (RNAi) through small interfering RNAs (siRNAs) are also used previously to inhibit N and L genes (35). Further, an anti-viral drug, ribavirin, is also used in infected persons (36) and also tested on animal experiments. However, it did not show good efficiency against the infection (36,37). Very recently, a small-molecule antiviral drug favipiravir (T-705) has shown the compelling antiviral activity in the hamster model against the henipaviruses (NiV and HeV) (38). Likewise, the potential of natural antiviral agents from

the medicinal plants can also be explored to combat viruses (39,40). Furthermore, studies also provide promising small-molecule inhibitors targeting NiV proteins (41,42). Apart from these, there are also some computational efforts to provide solutions in different ways. For example, some studies advocate the computational designing of vaccine epitopes against the specific NiV proteins (43,44). Very recently, we have also developed a quantitative structure-activity relationship (QSAR)-based prediction algorithm '*anti-Nipah*' for the identification of effective inhibitors against the NiV (45). The algorithm will predict the antiviral ability of any query compound against the NiV (45).

However, despite determinations, the Nipah Virus study is generally neglected. Currently, there is no Food and Drug Administration (FDA)-approved therapeutics or prophylactic vaccine available to treat NiV diseases in humans, and treatment is only supportive (21,28,31). Simultaneously, a broad range of Nipah hosts, its pathogenesis and the high fatality rate pose a recurring threat to humanity (46). Therefore, effective vaccines and therapeutics are an inevitable necessity of time (21). In the study, we have made efforts to provide potential therapeutic and vaccine solutions targeting all NiV proteins or genes. The resource 'NipahVR' may assist the worldwide scientific community in fighting against this lethal pathogen.

Materials and methods

Data retrieval

Complete genome sequences of the NiV were searched and collected from the NCBI database. In total, 18 complete or near-complete sequences were obtained utilizing the length-filtering criteria of size more than 10 kb and provided on the resource (Table 1). A facility with a category-wise (i.e. host, geographical area, and country) search option is applied on the web resource for the ease. Different information such as strain/isolate, host/source, length, country and geographical region was cataloged. Further, gene and protein sequences of reference NiV genome (NC_002728.1) are used for the downstream analyses, mainly diagnostic primer designing, vaccine epitope prediction and RNA-based therapeutics (i.e. siRNAs, microRNAs (miRNAs), single guide RNAs (sgRNAs)).

Codon analysis

Codon bias analysis of the complete genome sequence is performed to explore relative synonymous codon usage (RSCU) and codon frequency. Further, codon preference and context are analyzed employing the Anaconda program (47).

Phylogenomics

We have performed phylogenomic analysis to understand the phylogenetic reconstruction of NiV genomes. In the current study, we have employed complete genomes of 15 NiV that cause outbreaks in various Asian countries like India, Bangladesh, Malaysia and Singapore from 1998–2018. The genomic information was extracted from various sources like NCBI, ViPR, Viral zone and research articles (48,49). Further, the Molecular Evolutionary Genetics Analysis (MEGAv7.0) software with a Neighbor-joining method was utilized (50). The evolutionary distance was inferred through the Jukes–Cantor method, with a bootstrap test of 1000 replicates.

Diagnostic primers

In order to provide diagnostic primers, two strategies were utilized. First published literature was searched for extracting the experimentally used primer pairs for the diagnosis of the Nipah virus along with relevant information. Secondly, putative primer pairs were also designed using the PrimerDesign-M tool (51), keeping default parameters primarily. Briefly, in the region of interest option, the start and end of each genomic region were provided to design primers specific to the target gene. Further, we have chosen multiple-fragment primer design options with a flex parameter for fragment overlap option for each given genomic region. We have selected the primer length of 20 (minimum) to 25 (maximum) for each gene with the 5% detection limit. Then, complexity limit 2 was set to allow one degenerate position. Further, the maximum difference between melting temperatures (Tm) of reverse and forward primers was taken as 5°C. The window size of 10 was utilized for the investigation of dimerization, while the default dimer ratio (0.9) was chosen. Lastly, the G/C clamp option that helps to specify G or C at 3′ ends of primer was selected. This helps to promote specific binding at 3′ ends due to strong GC bases bonding.

Vaccine epitopes

For the potential epitope identification, 9-mer overlapping peptides were generated for each NiV-encoded proteins, i.e. N, P, W, V, C, M F, G and L. Further, the analyses were performed in quest to find promiscuous immune response, inducing peptides against the virus as also described previously (52). In order to have proper immune response different arms of the immune system, i.e. B-cell epitopes, T-cell epitopes and MHC binding is essential, hence considered in the study. The reliable and efficient linear B-cell epitopes of each NiV proteins were predicted using the LBtope algorithm (53), and strict criteria of 60% were

Table 1. List of 18 Nipah virus genomes

Accession	Name	Length	Host/source	Country	Date
FJ513078.1	Nipah virus isolate Ind-Nipah-07-FG from India, complete genome	18 252	*Homo sapiens*/lung tissue	India	2007
AY988601.1	Nipah virus from Bangladesh, complete genome	18 252	*Homo sapiens*	Bangladesh	2004
JN808857.1	Nipah virus isolate NIVBGD2008MANIKGONJ, complete genome	18 252	*Homo sapiens*/throat swab	Bangladesh	2008
JN808863.1	Nipah virus isolate NIVBGD2008RAJBARI, complete genome	18 252	*Homo sapiens*/throat swab	Bangladesh	2008
NC_002728.1	Nipah virus, complete genome	18 246	*Homo sapiens*	Malaysia	1999
AF212302.2	Nipah virus, complete genome	18 246	*Homo sapiens*	Malaysia	1999
AY029767.1	Nipah virus isolate UMMC1, complete genome	18 246	*Homo sapiens*/CSF of an encephalitic patient	Malaysia	1999
AJ564621.1	Nipah virus complete genome, isolate NV/MY/99/VRI-2794	18 246	*Sus scrofa* (pig)	Malaysia	1998
AY029768.1	Nipah virus isolate UMMC2, complete genome	18 246	*Homo sapiens*/throat secretion of an encephalitic patient	Malaysia	1999
AJ564622.1	Nipah virus complete genome, isolate NV/MY/99/VRI-1413	18 246	*Sus scrofa* (pig)	Malaysia	1998
AJ564623.1	Nipah virus complete genome, isolate NV/MY/99/UM-0128	18 246	*Sus scrofa* (pig)	Malaysia	1998
AJ627196.1	Nipah virus complete genome, isolate NV/MY/99/VRI-0626	18 246	*Sus scrofa* (pig)	Malaysia	1999
KY425655.1	Nipah virus isolate IRF0158, partial genome	18 214	*Homo sapiens*	Malaysia	1999
KY425646.1	Nipah virus isolate IRF0160, partial genome	18 212	*Homo sapiens*	Malaysia	1999
MH396625.1	Nipah henipavirus strain MCL-18-H-1088, complete genome	18 210	*Homo sapiens*/throat swab	India	2018
JN808864.1	Nipah virus isolate NIVBGD2010FARIDPUR, partial genome	18 167	*Homo sapiens*	Bangladesh	2010
FN869553.1	Nipah virus N gene, P gene, M gene, F gene, G gene and L gene, isolated from urine of *Pteropus vampyrus*	14 973	Urine of Pteropus vampyrus (Bat)	Malaysia	2008
AF376747.1	Nipah virus nucleocapsid protein (N), V protein (P/V/C), phosphoprotein (P/V/C), C protein (P/V/C), matrix protein (M), fusion protein (F), and glycoprotein (G) genes, complete cds	11 200	*Pteropus hypomelanus* (Bat)	Malaysia	2001

selected. The result is further analyzed and integrated on the web server. Further, in order to identify the efficient MHC class I-binding peptides (putative cytotoxic T lymphocytes (CTL) epitopes) from the Nipah proteins, the ProPred1 prediction server (54) was utilized, and preeminent 4% were selected. Similarly, MHC class II binders (potential T helper (Th) epitopes) were estimated using the ProPred tool (55), and the uppermost 3% peptides were recommended as promiscuous binders. Furthermore, potential CTL epitopes were also derived using the CTLPred tool (56) developed using the artificial neural network (ANN) and support vector machine (SVM) techniques through employing the combined approach with the default cut-off of 0.51 for ANN and 0.36 for SVM. The top 3 epitopes were selected

for each protein. Moreover, experimentally proven NiV epitopes were also searched.

siRNAs and miRNAs

RNA-based therapeutics could provide an alternative way to fight against the pathogens. For the designing of siRNAs against the different NiV genes, two different algorithms, i.e. a virus-specific algorithm, VIRsiRNApred (57) and DesiRm (58), were used. Further, the immunomodulatory potential of the siRNAs is deduced using the imrna program (59). For the prediction of siRNAs using VIRsiRNApred, Model-2 was utilized. It is developed on 1725 viral siRNAs employing hybrid nucleotide frequencies, binary and

thermodynamic features. Further, only efficient siRNAs having at least 55% predicted inhibition score were considered. Additionally, off-targets were also elucidated against the *Homo sapiens* (human) genome assembly GRCh37 (hg19). Correspondingly, potential siRNAs using the threshold of 0.80 were also deduced applying the DesiRm algorithm. Moreover, the imRNA program with the 'siRNA immunotoxicity' option along with the 'screen siRNA library' module is used to explore the immunomodulatory or non-immunomodulatory potential of siRNAs. SiRNAs with a score of 4.5 and above are considered as immunomodulatory, and less than the threshold is non-immunomodulatory. Moreover, we have also predicted NiV miRNAs. First, the VMir algorithm (60), which is consists of two programs VMir analyzer and VMir viewer, is used with the default settings to detect the putative precursor miRNAs (pre-miRNAs) hairpin (HP) structure. In brief, the maximum HP size of 200, minimum HP score (100) and minimum window count size of 35 are utilized. Further, these pre-miRNAs were subjected to the MatureBayes tool (61) to identify mature miRNAs.

sgRNAs and genome editing

Recently, clustered regularly interspaced short palindromic repeat-associated protein (CRISPR/Cas) system-based genome editing employing sgRNAs also shown to have an application to target a particular genomic region (62) or viral pathogen (63). For this, we have used the ge-CRISPR tool (64) to screen the NiV genes/genome on both forward and reverse strands to discover and extract the possible sgRNAs, i.e. 20 base pair upstream sequences as putative targets based on the protospacer adjacent motif (PAM) mainly 'NGG'.

Web resource development

The eventual goal is to provide the web resource of the putative therapeutic regimens and solutions from the study to assist in fighting with the deadly Nipah virus and support scientific society in therapeutic development. This platform, 'NipahVR', is hosted on the Linux environment using a LAMP (Linux, Apache HTTP Server, MySQL and PHP) open-source web development platform. The front-end of the web interface is built using the PHP, HTML, CSS and JavaScript as also accomplished earlier (52), and the back-end of the resource is complemented with MySQL for the data management.

Results and discussion

NipahVR is an integrative and systematic resource mainly dedicated towards the putative therapeutics and vaccinome
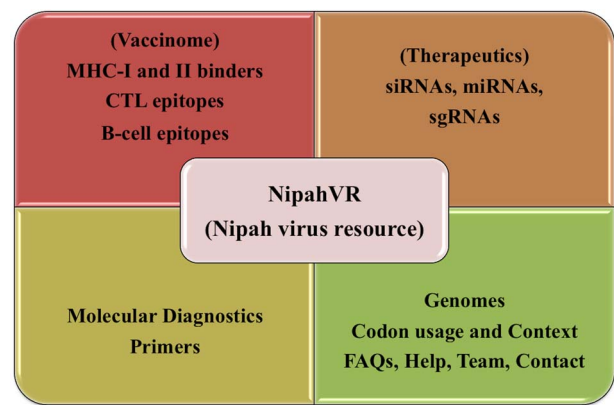
Figure 1. Overview of NipahVR resource components.

against the Nipah virus. A well-structured and dynamic web interface is developed for navigation. It is classified into different divisions like genomes, phylogenomics, molecular diagnostic primers and most importantly vaccine epitopes (B-cell, CTL, MHC-I and -II binders) and therapeutics (siRNAs, sgRNAs, miRNAs). The complete architecture of the NipahVR compendium is shown in Figure 1, demonstrating all the components.

NipahVR genomes

Genomic information of available 18 Nipah virus sequences was compiled and provided on the resource (Table 1). It is equipped with an advance genome search facility for easy navigation. Nipah genomes can be searched using different search options such as host/source (i.e. human, pig, bat), geographical region (Asia) and country (India, Bangladesh, Malaysia) with the detailed meta-information.

Codon usage and context

We have calculated the codon frequency and pattern in the genomes, which vary due to the nucleotide composition, GC percentage, expression level, etc. Additionally, codon preference is represented through histogram, where rare codons are shown in blue and black color signify preferred codons (Figure 2). The most preferred codons are AAA, AUG, UAA, and GCG, CGC, CGU is the least preferred or rare codons in the Nipah virus reference genome. Additionally, using the Anaconda software, we have also calculated codon pair residual values in the genome, indicating an association between two codons. A two-colored matrix depicting average residual values, red color denotes rare, and green color shows preferred codon pairs (Figure 3).

Phylogenomics

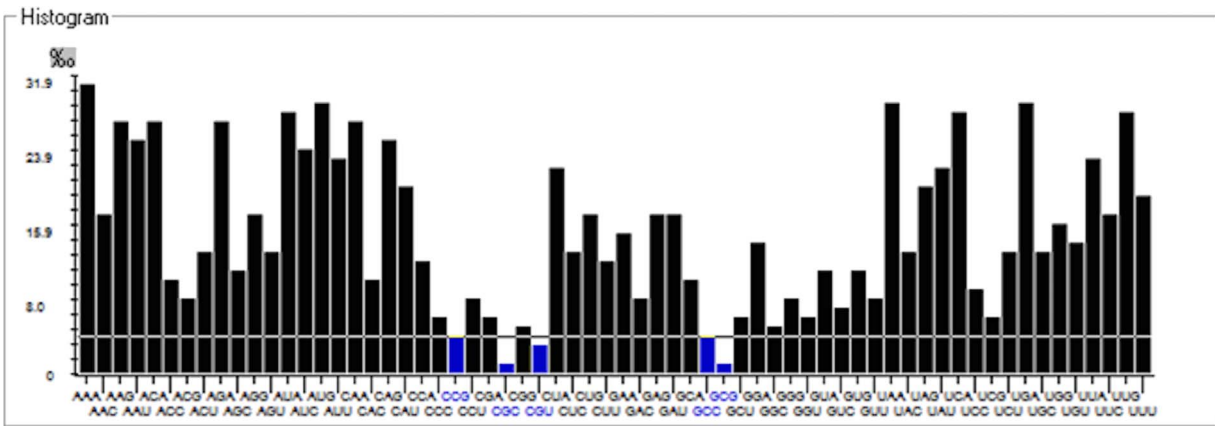The reconstructed phylogenetic tree showed the sum of branch length of 0.11. Out of 15, 07 NiV genomes from

Figure 2. Histogram showing codon distribution (rare (blue) and preferred (black)) of NiV genome.
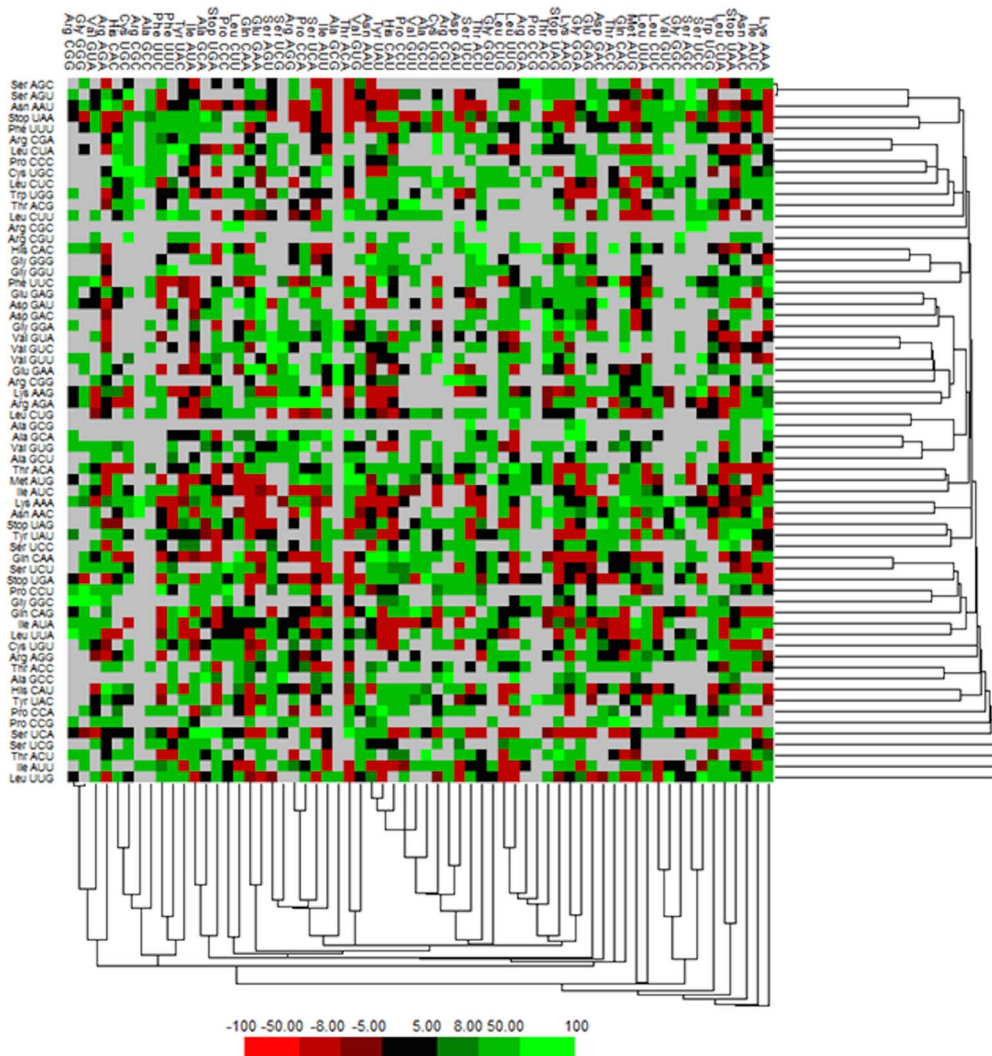

Figure 3. Matrix illustrating codon context analysis.

Malaysia outbreaks (1998–99) were clustered together with the bootstrap of 100. Further, from the 04 NiV genomes from Bangladesh outbreak, two genomes from 2004 and 2010 outbreaks were clustered together. Interestingly, the 02 NiV genomes from the 2008 epidemic of Bangladesh were grouped with the NiV of 2007, West Bengal, India. Moreover, the recent outbreak of the NiV virus in Kerala, India, was found in the Indian and Bangladesh
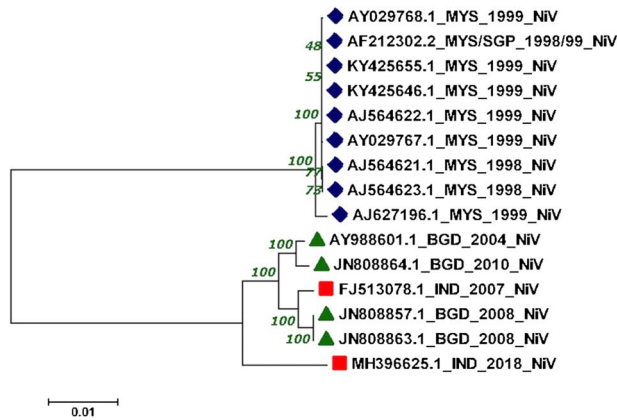
Figure 4. Phylogenetic tree showing the relationship of 15 NiV complete genomes employing Neighbor-joining method.

NiV genomes but at a distant branch. The phylogenetic reconstruction of NiV genomes is shown in Figure 4. The NiV genomes from Asian outbreaks are grouped according to their geographical location. The Malaysian outbreak of NiV (23) was closely related while the Indian and Bangladesh epidemic NiV displayed close resemblance with each other, except the recent Kerala outbreak, which is unrelated with all the NiV genomes. Our phylogenomic study showed that due to the course of time, the NiV showed significant mutations at the genomic level.

Molecular diagnostic primers

We have collected primer pairs utilized for the detection of the NiV. Detailed information about primers like the respective primer name, a sequence of primer, its orientation (forward and reverse), genomic region and study reference is provided (Table S1). Overall, 55 forward and 53 reverse primers were reported against different genes of NiV and compiled on the web server. Additionally, we

have also designed primer pairs for each gene utilizing the PrimerDesign-M tool. In total, two primer pairs for N-gene; six primer pairs for P-gene; three primer pairs for M-gene; six primer pairs for F-gene; one primer pair for G-gene and nine primer pairs for polymerase gene. Gene name, start–end and melting temperature (Tm) of primers were reported. A detailed list of all predicted primer pairs is provided in Table S2. These primers (experimental and designed) will be valuable for the detection and diagnosis of NiV.

Putative epitopes

In this study, efforts were made for the identification of potential vaccine candidates for the Nipah virus. Epitopes encompassing promising MHC I and II binders, CTL epitopes and B-cell epitopes is cataloged. 9-mer peptides were generated from the Nipah proteins. Overall, 979 MHC-I and 1628 MHC-II binding peptides were deduced and presented on the server. For both MHC-I and II binders, peptide sequence, respective protein region, MHC-alleles and counts were provided. The protein-wise number of peptides for both the MHC classes is shown in Figure 5. Likewise, 27 potential CTL epitopes with sequence, start–end and allele information were recorded (Table 2). Furthermore, 400 efficient B-cell epitopes (Figure 6) along with sequence, a b-cell confidence score is provided belonging to different Nipah proteins. Among all, for N protein EKKNNQDLK, P (W, V) protein SPEDEEPSS, for C protein LLTLFRRTK, for M protein AAYPLGVGK, for F protein SRLEDRRVR, for G protein DPLLAMDEG and for L protein KLSQNLLVT peptides are the highest-scoring and confident epitopes, which can be focused. Further, we have also found four experimentally proven linear b-cell epitopes, i.e. three for N protein (SIQTKFAP, SNRTQGE and SPSAAE) and G (NQILKPKLISYTLPVVG). However, we did not find any T-cell epitopes.
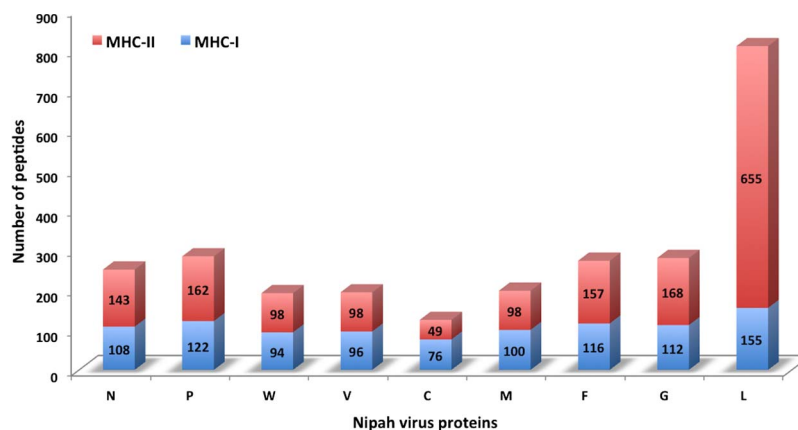
Figure 5. Chart displaying the number of MHC-I and -II binders from different proteins.

Table 2. Table showing potential 27 CTL epitopes against individual NiV proteins along with detail information

| Proteins	Epitopes	Start	End	Alleles
N	AYGLRITDM	150	158	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
N	NLRSRLAAK	478	486	HLA-A2, HLA-A*0201, HLA-A*2402, HLA-A2.1, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
N	ALNINRGYL	347	355	HLA-A2, HLA-A*0203, HLA-A*0205, HLA-A3, HLA-A*0301, HLA-A2.1, HLA-B*3501, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, HLA-B35, Mamu-A*01
P/W/V	AVPFTLRNL	190	198	HLA-B*3501, HLA-B*51, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, HLA-B35, Mamu-A*01
P	INSIKLINL	521	529	HLA-B*3501, HLA-B7, HLA-B*0702, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, HLA-B35, Mamu-A*01
P/W/V	TTGLNPTAV	183	191	HLA-A*0206, HLA-B*5301, HLA-B*0702, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
W/V	SEDPIIREL	337	345	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
C	GECLRMMEM	54	62	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
C	APVENLNKL	44	52	HLA-A2, HLA-A*0203, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
C	KLRGECLRM	51	59	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, HLA-B*2706, Mamu-A*01
M	KKVLTSGSI	142	150	HLA-A2, HLA-A3, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01, HLA-A*3301, HLA-A*6801
M	YLKIDADLS	209	217	HLA-A2.1, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
M	FRRNNAIAF	196	204	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01, HLA-A*6802
F	EAMKNADNI	136	144	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01, HLA-A*6802
F	RFALSNGVL	375	383	HLA-B*3501, HLA-B*51, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
F	ANCISVTCQ	385	393	HLA-A1, HLA-B*5102, HLA-B*5103, HLA-B*5401, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
G	KINEGLLDS	36	44	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
G	ILSAFNTVI	46	54	HLA-A11, HLA-A3, HLA-A*0301, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
G	QRIIGVGEV	247	255	HLA-A31, HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B8, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
L	QPSDDKRLS	39	47	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
L	SDSCIIHMR	194	202	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
L	VSMIEPLVL	287	295	HLA-A*0203, HLA-B*3501, HLA-B*51, HLA-B*0702, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01

Additionally, we have also analyzed putative epitopes from all the arms of the immune system, i.e. B-cell epitopes, MHC-I and II binders and CTL epitopes to find the common epitopes, which could be promising and can be recommended as the potential vaccine candidates (Figure 7). We have found the 24 epitopes reported to be B-cell epitope and also efficient MHC-I and II binders (Table S3); 70 epitopes belonging to both b-cell as well as MHC-I binders (Table S4.); and 109 epitopes which are b-cell as well as MHC-II binders (Table S5). Further, there are two epitopes ('ILSAFNTVI' (G protein) and 'FRRNNAIAF' (M protein)) which are characterized in all three categories, i.e. CTL
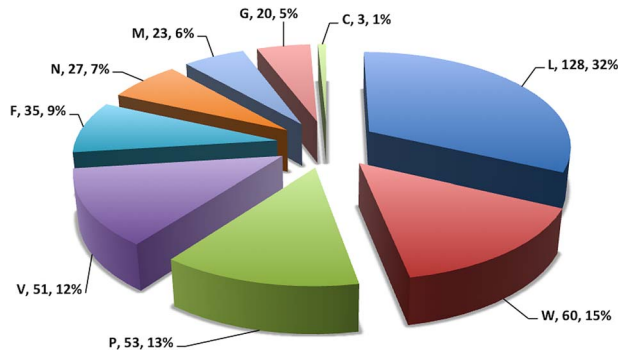
Figure 6. Pie chart showing number of putative b-cell epitopes for individual Nipah proteins.
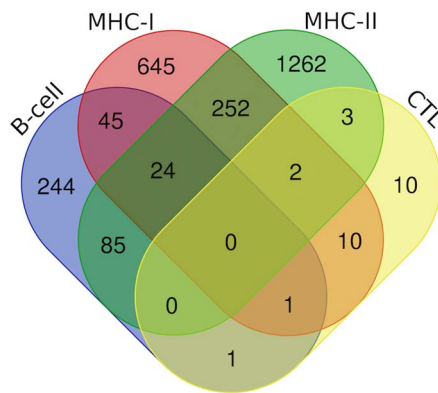
Figure 7. Venn diagram showing the number of peptides/epitopes belonging and common in diverse epitope classes.

epitopes and both MHC-I and II binders. Likewise, 13 epitopes from CTL epitopes, and MHC-I binders, 5 epitopes from CTL epitopes and MHC-II binders were catered (Table S6). Furthermore, two epitopes were found to be as B-cell as well as CTL epitope, i.e. 'QPSDDKRLS' from L protein and 'NLRSRLAAK' from N protein. Along with this, 278 peptides that are putative MHC-I and II binders were also cataloged (Table S7).

siRNAs and miRNAs

RNA-based therapeutic interventions provide another approach to counter lethal viruses through silencing the genes. In this study, we have also tried to provide a compendium of potent siRNAs against the individual NiV genes. Overall, 118 putative siRNAs with very few off-targets using the VIRsiRNApred algorithm, which is developed using the experimentally proven viral siRNAs and 441 siRNAs employing DesiRm tool, were cataloged along with inferred inhibition efficiency in percentage. Additionally, the immunomodulatory potential of these siRNAs is also predicted, which could be helpful and crucial in the development of vaccine adjuvant or RNA-based immunotherapy and therapeutics. Furthermore, the number of siRNA off-targets to the human genome is also

presented. The resource provides a complete picture of these efficient siRNAs with a small number of off-targets and detail information such as sense–antisense sequence, gene region, start–end, efficacy scores and immunomodulatory potential. The set of efficacious siRNAs, i.e. 18 using the VIRsiRNApred algorithm (Table S8) and 43 using DesiRm (Table S9) is specified.

Furthermore, we have also designed and identified the Nipah precursor and mature miRNAs. Totally, 22 precursor miRNAs (pre-miRNAs) were identified along with 44 mature Nipah-miRNAs (22 5p and 22 3p). Among these, 3 pre-miRNAs are from N, 3 from P/V/C, 1 from M, 2 from F, 3 from G and 10 belong to the L gene. Detailed information related to miRNAs, i.e. mature miRNA sequences (5p and 3p), genomic region, the precursor (hairpin) and mature miRNA location on Nipah genome, precursor (hairpin) length, GC content, score and rank is extracted and provided.

sgRNAs and genome editing

On the basis of our analysis, overall, 1412 sgRNAs from the NiV were obtained. Out of these, we have found and listed 126 sgRNAs that can act as putative targets against the virus. Apart from this, a list of the 21 most efficient sgRNAs is given in Table 3. The output of this displays sequence of sgRNA, associated PAM (NGG), information of strand (sense/antisense), start and end coordinates, AT and GC content and efficiency of each sgRNA (percentage efficiency). This information will be beneficial to predict or identify CRISPR sgRNA targets against NiV and will certainly reduce experimental time and cost.

Conclusion

Nipah virus is a priority pathogen from the *Paramyxoviridae* family and a BSL-4 agent, which causes various diseases such as encephalitis, respiratory illness and fever. Moreover, a high case fatality rate during epidemic, broad host range and lack of therapeutics or prophylactic vaccines critically demand efforts and a multidisciplinary approach to develop combat strategies against this virus. Up to now, very few computational studies focusing on the NiV are performed, and there is no such kind of therapeutic web resource available for it. In the current work, we have developed all-inclusive resource 'NipahVR' for the putative therapeutic solutions targeting individual Nipah genes and proteins. It provides a compendium of various components that includes genomics, diagnostic primers, vaccine epitopes (MHC-I and -II binders, CTL, B-cell) and therapeutics (siRNAs, miRNAs, sgRNAs). Here, based on our analysis, we are also endorsing and providing a catalog of potential vaccine epitopes and efficient siRNAs, miRNAs and sgR-

Table 3. List of 21 efficient sgRNAs targeting Nipah virus genes

Strand	PAM	Strand (+/−)	Start	End	GC%	Efficiency	Genes
TATGTATTCAGAGAGACCCG	GGG	+	346	365	45	70.34	N
ATACCAGTAATGGAGAGGAG	AGG	+	443	462	45	57.24	N
GGAGAATCTGAACAAGTTGA	GGG	+	2565	2584	40	60.28	P
GGAGTTAAAGAGCAAAACGT	TGG	+	3051	3070	40	69.49	P
ACTCCGATGCCAAAGTCCCG	AGG	+	3585	3604	60	56.47	P
ATACCCCAGGAGCTAACGAG	AGG	+	5256	5275	55	55.70	M
GTATAGCTCAACACTTCACG	TGG	−	5352	5371	45	64.14	M
GGGATCAATTAGCCCCCAGA	GGG	−	6014	6033	55	57.34	M
CTAGTCATAATACATCACAC	AGG	+	6214	6233	35	60.40	M
GTATTATGCATGAATCTGAA	CGG	−	6238	6257	30	56.11	M
CTATACTCTCTAAAAGGGAG	TGG	−	8790	8809	40	60.22	G
GGATGAGGCTAGGATCCTGA	GGG	+	12 314	12 333	55	66.82	L
GGGTATAGGGATAGACACGG	AGG	+	12 627	12 646	55	59.66	L
GTACTTTCTTCCACGGAACG	AGG	−	12 707	12 726	50	55.37	L
GCACTCTTACCAACACCCAG	AGG	−	12 841	12 860	55	76.57	L
CTATAAATCAGACAATACAG	AGG	+	13 481	13 500	30	70.77	L
GTGCCTATGAGACAAACACG	AGG	+	13 861	13 880	50	68.35	L
GGTTTATTAGATACAACTAA	AGG	+	14 817	14 836	25	56.53	L
ATACAACGTCTGTAACCACT	GGG	−	15 496	15 515	40	56.20	L
GGGGAAATTGAAAGGACTAG	TGG	+	17 099	17 118	45	55.29	L
ATATCATAAATAGGACAGCG	GGG	+	17 182	17 201	35	59.24	L

NAs. However, medicinal plant-based antivirals and other chemical compounds are not in the scope of current work. We anticipate that NipahVR will be useful and assist the wider scientific community in determining efficient antiviral candidates to fight against the Nipah and exterminate the infectivity. We will periodically update and maintain the stable functioning of the NipahVR web resource.

Authors' roles

This study is conceived, designed and supervised by M.K. genomic data collection and curation; A.K.G., web server development; A.K.G., codon analysis; A.K.G., vaccine epitope analysis; A.K., K.M., A.K.G., phylogenetic analysis; A.R., sgRNAs and diagnostic primers; K.K., siRNA analysis; S.D., miRNA analysis; A.T., data interpretation; A.K.G., A.K., A.R., M.K.. Manuscript writing, A.K.G., A.R., K.K., M.K.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Funding

References

1. Harcourt,B.H., Tamin,A., Ksiazek,T.G. *et al.* (2000) Molecular characterization of Nipah virus, a newly emergent paramyxovirus. *Virology*, **271**, 334–349.
2. Wang,L., Harcourt,B.H., Yu,M. *et al.* (2001) Molecular biology of Hendra and Nipah viruses. *Microbes Infect*, **3**, 279–287.
3. Chua,K.B., Bellini,W.J., Rota,P.A. *et al.* (2000) Nipah virus: a recently emergent deadly paramyxovirus. *Science*, **288**, 1432–1435.
4. Bonaparte,M.I., Dimitrov,A.S., Bossart,K.N. *et al.* (2005) Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *Proc Natl Acad Sci USA*, **102**, 10652–10657.
5. Negrete,O.A., Levroney,E.L., Aguilar,H.C. *et al.* (2005) EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. *Nature*, **436**, 401–405.
6. Tamin,A., Harcourt,B.H., Ksiazek,T.G. *et al.* (2002) Functional properties of the fusion and attachment glycoproteins of Nipah virus. *Virology*, **296**, 190–200.
7. Sharma,V., Kaushik,S., Kumar,R. *et al.* (2019) Emerging trends of Nipah virus: a review. *Rev Med Virol*, **29**, e2010.
8. Lo Presti,A., Cella,E., Giovanetti,M. *et al.* (2016) Origin and evolution of Nipah virus. *J Med Virol*, **88**, 380–388.
9. Chua,K.B., Goh,K.J., Wong,K.T. *et al.* (1999) Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *Lancet*, **354**, 1257–1259.
10. Ang,B.S.P., Lim,T.C.C. and Wang,L. (2018) Nipah virus infection. *J Clin Microbiol*, **56**.
11. Luby,S.P., Rahman,M., Hossain,M.J. *et al.* (2006) Foodborne transmission of Nipah virus, Bangladesh. *Emerg Infect Dis*, **12**, 1888–1894.

12. Luby,S.P., Gurley,E.S. and Hossain,M.J. (2009) Transmission of human infection with Nipah virus. *Clin Infect Dis*, **49**, 1743–1748.

13. Drexler,J.F., Corman,V.M., Muller,M.A. *et al.* (2012) Bats host major mammalian paramyxoviruses. *Nat Commun*, **3**, 796.

14. Epstein,J.H., Prakash,V., Smith,C.S. *et al.* (2008) Henipavirus infection in fruit bats (Pteropus giganteus), India. *Emerg Infect Dis*, **14**, 1309–1311.

15. Yadav,P.D., Raut,C.G., Shete,A.M. *et al.* (2012) Detection of Nipah virus RNA in fruit bat (Pteropus giganteus) from India. *Am J Trop Med Hyg*, **87**, 576–578.

16. Reynes,J.M., Counor,D., Ong,S. *et al.* (2005) Nipah virus in Lyle's flying foxes, Cambodia. *Emerg Infect Dis*, **11**, 1042–1047.

17. Wacharapluesadee,S., Lumlertdacha,B., Boongird,K. *et al.* (2005) Bat Nipah virus, Thailand. *Emerg Infect Dis*, **11**, 1949–1951.

18. Chadha,M.S., Comer,J.A., Lowe,L. *et al.* (2006) Nipah virus-associated encephalitis outbreak, Siliguri, India. *Emerg Infect Dis*, **12**, 235–240.

19. Ksiazek,T.G., Rota,P.A. and Rollin,P.E. (2011) A review of Nipah and Hendra viruses with an historical aside. *Virus Res*, **162**, 173–183.

20. Peterson,A.T. (2015) Mapping risk of Nipah virus transmission across Asia and across Bangladesh. *Asia Pac J Public Health*, **27**, NP824–NP832.

21. Ambat,A.S., Zubair,S.M., Prasad,N. *et al.* (2019) Nipah virus: a review on epidemiological characteristics and outbreaks to inform public health decision making. *J Infect Public Health*, **12**, 634–639.

22. Harit,A.K., Ichhpujani,R.L., Gupta,S. *et al.* (2006) Nipah/ Hendra virus outbreak in Siliguri, West Bengal, India in 2001. *Indian J Med Res*, **123**, 553–560.

23. Arankalle,V.A., Bandyopadhyay,B.T., Ramdasi,A.Y. *et al.* (2011) Genomic characterization of Nipah virus, West Bengal, India. *Emerg Infect Dis*, **17**, 907–909.

24. Arunkumar,G., Chandni,R., Mourya,D.T. *et al.* (2019) Outbreak investigation of Nipah virus disease in Kerala, India, 2018. *J Infect Dis*, **219**, 1867–1878.

25. Yadav,P.D., Shete,A.M., Kumar,G.A. *et al.* (2019) Nipah virus sequences from humans and bats during Nipah outbreak, Kerala, India, 2018. *Emerg Infect Dis*, **25**, 1003–1006.

26. Gurley,E.S., Montgomery,J.M., Hossain,M.J. *et al.* (2007) Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerg Infect Dis*, **13**, 1031–1037.

27. Ravichandran,L., Venkatesan,A. and Febin Prabhu Dass,J. (2018) Epitope-based immunoinformatics approach on RNA-dependent RNA polymerase (RdRp) protein complex of Nipah virus (NiV). *J Cell Biochem.*.

28. Satterfield,B.A., Dawes,B.E. and Milligan,G.N. (2016) Status of vaccine research and development of vaccines for Nipah virus. *Vaccine*, **34**, 2971–2975.

29. Mungall,B.A., Middleton,D., Crameri,G. *et al.* (2006) Feline model of acute nipah virus infection and protection with a soluble glycoprotein-based subunit vaccine. *J Virol*, **80**, 12293–12302.

30. DeBuysscher,B.L., Scott,D., Marzi,A. *et al.* (2014) Single-dose live-attenuated Nipah virus vaccines confer complete protection by eliciting antibodies directed against surface glycoproteins. *Vaccine*, **32**, 2637–2644.

31. Broder,C.C., Weir,D.L. and Reid,P.A. (2016) Hendra virus and Nipah virus animal vaccines. *Vaccine*, **34**, 3525–3534.

32. Walpita,P., Cong,Y., Jahrling,P.B. *et al.* (2017) A VLP-based vaccine provides complete protection against Nipah virus challenge following multiple-dose or single-dose vaccination schedules in a hamster model. *NPJ Vaccines*, **2**, 21.

33. Broder,C.C., Xu,K., Nikolov,D.B. *et al.* (2013) A treatment for and vaccine against the deadly Hendra and Nipah viruses. *Antiviral Res*, **100**, 8–13.

34. Geisbert,T.W., Mire,C.E., Geisbert,J.B. *et al.* (2014) Therapeutic treatment of Nipah virus infection in nonhuman primates with a neutralizing human monoclonal antibody. *Sci Transl Med*, **6**, 242ra282.

35. Mungall,B.A., Schopman,N.C., Lambeth,L.S. *et al.* (2008) Inhibition of Henipavirus infection by RNA interference. *Antiviral Res*, **80**, 324–331.

36. Chong,H.T., Kamarulzaman,A., Tan,C.T. *et al.* (2001) Treatment of acute Nipah encephalitis with ribavirin. *Ann Neurol*, **49**, 810–813.

37. Georges-Courbot,M.C., Contamin,H., Faure,C. *et al.* (2006) Poly(I)-poly(C12U) but not ribavirin prevents death in a hamster model of Nipah virus infection. *Antimicrob Agents Chemother*, **50**, 1768–1772.

38. Dawes,B.E., Kalveram,B., Ikegami,T. *et al.* (2018) Favipiravir (T-705) protects against Nipah virus infection in the hamster model. *Sci Rep*, **8**, 7604.

39. Lin,L.T., Hsu,W.C. and Lin,C.C. (2014) Antiviral natural products and herbal medicines. *J Tradit Complement Med*, **4**, 24–35.

40. Dhama,K., Karthik,K., Khandia,R. *et al.* (2018) Medicinal and therapeutic potential of herbs and plant metabolites / extracts countering viral pathogens - current knowledge and future prospects. *Curr Drug Metab*, **19**, 236–263.

41. Niedermeier,S., Singethan,K., Rohrer,S.G. *et al.* (2009) A small-molecule inhibitor of Nipah virus envelope protein-mediated membrane fusion. *J Med Chem*, **52**, 4257–4265.

42. Hotard,A.L., He,B., Nichol,S.T. *et al.* (2017) 4′-Azidocytidine (R1479) inhibits henipaviruses and other paramyxoviruses with high potency. *Antiviral Res*, **144**, 147–152.

43. Sakib,M.S., Islam,M.R., Hasan,A.K. *et al.* (2014) Prediction of epitope-based peptides for the utility of vaccine development from fusion and glycoprotein of nipah virus using in silico approach. *Adv Bioinformatics*, **2014**, 402492.

44. Saha,C.K., Mahbub Hasan,M., Saddam Hossain,M. *et al.* (2017) In silico identification and characterization of common epitope-based peptide vaccine for Nipah and Hendra viruses. *Asian Pac J Trop Med*, **10**, 529–538.

45. Rajput,A., Kumar,A. and Kumar,M. (2019) Computational identification of inhibitors using QSAR approach against Nipah virus. *Frontiers in Pharmacology*, **10**.

46. Luby,S.P. (2013) The pandemic potential of Nipah virus. *Antiviral Res*, **100**, 38–43.

47. Moura,G., Pinheiro,M., Silva,R. *et al.* (2005) Comparative context analysis of codon pairs on an ORFeome scale. *Genome Biol*, **6**, R28.

48. Pickett,B.E., Sadat,E.L., Zhang,Y. *et al.* (2012) ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res*, **40**, D593–D598.

49. Masson,P., Hulo,C., De Castro,E. *et al.* (2013) ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Res*, **41**, D579–D583.

50. Kumar,S., Stecher,G. and Tamura,K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*, **33**, 1870–1874.

51. Yoon,H. and Leitner,T. (2015) PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. *Bioinformatics*, **31**, 1472–1474.

52. Gupta,A.K., Kaur,K., Rajput,A. *et al.* (2016) ZikaVR: an integrated Zika virus resource for genomics, proteomics, phylogenetic and therapeutic analysis. *Sci Rep*, **6**, 32713.

53. Singh,H., Ansari,H.R. and Raghava,G.P. (2013) Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS One*, **8**, e62216.

54. Singh,H. and Raghava,G.P. (2003) ProPred1: prediction of promiscuous MHC class-I binding sites. *Bioinformatics*, **19**, 1009–1014.

55. Singh,H. and Raghava,G.P. (2001) ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, **17**, 1236–1237.

56. Bhasin,M. and Raghava,G.P. (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, **22**, 3195–3204.

57. Qureshi,A., Thakur,N. and Kumar,M. (2013) VIRsiRNApred: a web server for predicting inhibition efficacy of siRNAs targeting human viruses. *J Transl Med*, **11**, 305.

58. Ahmed,F. and Raghava,G.P. (2011) Designing of highly effective complementary and mismatch siRNAs for silencing a gene. *PLoS One*, **6**, e23443.

59. Chaudhary,K., Nagpal,G., Dhanda,S.K. *et al.* (2016) Prediction of Immunomodulatory potential of an RNA sequence for designing non-toxic siRNAs and RNA-based vaccine adjuvants. *Sci Rep*, **6**, 20678.

60. Sullivan,C.S. and Grundhoff,A. (2007) Identification of viral microRNAs. *Methods Enzymol*, **427**, 3–23.

61. Gkirtzou,K., Tsamardinos,I., Tsakalides,P. *et al.* (2010) Mature-Bayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors. *PLoS One*, **5**, e11843.

62. Sander,J.D. and Joung,J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, **32**, 347–355.

63. Price,A.A., Sampson,T.R., Ratner,H.K. *et al.* (2015) Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proc Natl Acad Sci USA*, **112**, 6164–6169.

64. Kaur,K., Gupta,A.K., Rajput,A. *et al.* (2016) ge-CRISPR - an integrated pipeline for the prediction and analysis of sgRNAs genome editing efficiency for CRISPR/Cas system. *Sci Rep*, **6**, 30870.