

# MMV-db: vaccinomics and RNA-based therapeutics database for infectious hemorrhagic fever-causing mammarenaviruses

Taimoor Khan<sup>1</sup>, Abbas Khan<sup>1</sup> and Dong-Qing Wei<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai 200240, P.R. China

<sup>2</sup>State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P.R. China

<sup>3</sup>Peng Cheng Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nanshan District, Shenzhen, Guangdong 518055, P.R. China

\*Corresponding author: Tel: +0086-13918500529; Fax: +862134204573; Email: [dqwei@sjtu.edu.cn](mailto:dqwei@sjtu.edu.cn)

Citation details: Khan, T., Khan, A., Wei, D.Q. MMV-db: vaccinomics and RNA-based therapeutics database for infectious hemorrhagic fever-causing mammarenaviruses. *Database* (2021) Vol. 2021: article ID baab063; DOI: <https://doi.org/10.1093/database/baab063>

## Abstract

The recent viral outbreaks and the current pandemic situation urges us to timely address any emerging viral infections by designing therapeutic strategies. Multi-omics and therapeutic data are of great interest to develop early remedial interventions. This work provides a therapeutic data platform (Mammarenavirus (MMV)-db) for pathogenic mammarenaviruses with potential catastrophic effects on human health around the world. The database integrates vaccinomics and RNA-based therapeutics data for seven human pathogenic MMVs associated with severe viral hemorrhagic fever and lethality in humans. Protein-specific cytotoxic T lymphocytes, B lymphocytes, helper T-cell and interferon-inducing epitopes were mapped using a cluster of immune-omics-based algorithms and tools for the seven human pathogenic viral species. Furthermore, the physiochemical and antigenic properties were also explored to guide protein-specific multi-epitope subunit vaccine for each species. Moreover, highly efficacious RNAs (small Interfering RNA (siRNA), microRNA and single guide RNA (sgRNA)) after extensive genome-based analysis with therapeutic relevance were explored. All the therapeutic RNAs were further classified and listed on the basis of predicted higher efficacy. The online platform (<http://www.mmvdb.dqweilab-sjtu.com/index.php>) contains easily accessible data sets and vaccine designs with potential utility in further computational and experimental work. Conclusively, the current study provides a baseline data platform to secure better future therapeutic interventions against the hemorrhagic fever causing mammarenaviruses.

**Database URL:** <http://www.mmvdb.dqweilab-sjtu.com/index.php>

## Introduction

The human-infecting seven mammarenaviruses associated with causing viral hemorrhagic fever are named *Lassa virus* (LASV), *Chapare virus* (CHAPV), *Lujó virus* (LUJV), *Guanarito virus* (GTOV), *Junín virus* (JUNV), *Machupo virus* (MACV) and *Sabiá virus* (SABV). Demographically, LASV and LUJV are considered indigenous to Africa, whereas CHAPV, GTOV, JUNV, MACV and SABV are common in American countries (1, 2). Mammarenavirus (mammalian arenaviruses) is an important genus of animal viruses and accommodates in the family called *Arenaviridae*. These are basically enveloped and spherical viral particles with a diameter of 50–300 nm (3, 4). The genome consists of two single-stranded (RNA) molecules known as L (large) and S (small) segments. Each genomic segment is responsible for the production of two different proteins, whereas the L segment serves as a genomic code for zinc-binding matrix protein (Z) and viral RNA-dependent RNA polymerase (L). Similarly, the S segment encodes an envelope glycoprotein precursor (GPC) and a nucleoprotein (NP) (5, 6). Additionally, two glycoprotein subunits called GP1 and GP2 of the

spike are obtained after posttranslational cleavage during GPC synthesis (7, 8). The virions then utilize GP1 subunit to facilitate cell-surface receptor binding and enter the cell through endocytosis (9, 10). The other GP2 subunit mediates pH-dependent membrane fusion followed by uncoating and releasing viral ribonucleoprotein complexes inside the cell (11).

The mechanism of viral pathogenesis exhibits severe outcomes and a high fatality rate (12, 13) correlated with mammarenaviral hemorrhagic fevers. These viruses can be transmitted through aerosol or contact with infected person (3). The virus then gains systematic entry into the host lymphoid system with undetected pneumonic symptoms (14). The prominent targeting of macrophages (15, 16) and liver damage are considered a hallmark of pathogenicity (17, 18) during these human mammarenavirus infections. Similarly, compromised immune response function with secondary bacterial infections (19, 20) and leukocyte dysfunction in polymorphonuclear cells causing leukopenia (21, 22) are also associated with mammarenaviral infections. Other abnormalities, including the defective function of

Received 17 June 2021; Revised 24 August 2021; Accepted 18 October 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

**Table 1.** Presenting genome-based accession information and other basic details of the cohort study MM species

Index	Virus name	Accession ID	Genomic length	Genome type
1	Lassa virus segment L	NC004297.1	7279 bps	linear RNA
	Lassa virus segment S	NC004296.1	3402 bps	
2	Chapare virus segment L	NC010563.1	7107 bps	linear RNA
	Chapare virus segment S	NC010562.1	3357 bps	
3	Lujo virus segment L	NC012777.1	7163 bps	linear RNA
	Lujo virus segment S	NC012776.1	3189 bps	
4	Guanarito virus segment L	NC005082.1	7081 bps	linear RNA
	Guanarito virus segment S	NC005077.1	3343 bps	
5	Junin virus segment L	NC005080.1	7114 bps	linear RNA
	Junin virus segment S	NC005081.1	3341 bps	
6	Machupo virus segment L	NC005079.1	7196 bps	linear RNA
	Machupo virus segment S	NC005078.1	3439 bps	
7	Sabia virus segment L	NC006313.1	7133 bps	linear RNA
	Sabia virus segment S	NC006317.1	3366 bps	

macrophages, depletion of the T- and B-lymphocytes, down-regulated response of primary and secondary antibodies and high interferon (IFN) concentrations (23–25), has also been linked with hemorrhagic diseases.

The therapeutic approaches adapted for hemorrhagic fevers caused by different mammarenaviruses with related symptoms (26–28) vary with pathological conditions. Till now, there have been very few effective treatment options available to combat hemorrhagic fever in clinical setups. These treatment regimens include administering an adequate dosage of neutralizing antibodies during immune serum treatment (29) with related complications of transient cerebellar-cranial nerve syndrome (30, 31). Such passive antibody therapy options are also harbored by transfusion-borne diseases and require alternate treatments (32). The current anti-mammarenaviral therapy also includes the use of ribavirin (1- $\beta$ -d-ribofuranosyl-1 H-1,2,4-triazole-3-carboxamide) with partial efficacy against some mammarenavirus infections. Meanwhile, the use of ribavirin is harbored by associated toxicity and adverse side effects, including severe anemia, thrombocytosis and birth-related defects in humans (33–35). Another candidate drug called T-705 (favipiravir) with targeted inhibition of target viral RNA synthesis and broad antiviral activity against RNA viruses (36, 37) is also used as a treatment option. Furthermore, no Food and Drug Regulatory Authority (FDA)-licensed vaccines are currently available to prevent mammarenavirus infections. The only designated live-attenuated vaccine that advanced to human clinical trials is called Candid 1 (Candidate no. 1), with efficacy against JUNV mediated infections (38, 39). Still, the continued search of potential vaccines expanded to several recombinant viruses, inactivated mammarenaviruses or alike particles (40) and other candidates tested in various animal models (41, 42) needs further evaluation with its potential therapeutic significance as a human vaccine.

For instance, the development of biological web databases for different diseases, i.e. breast cancer and cytomegaloviruses, are of great interest to researchers (43, 44). In this study, annotated data sets based on the genome and proteome analysis for seven species of human-infecting mammarenaviruses are presented. The analysis basically comprised of genome/proteome collection, immune-based epitopes prediction, vaccine designing and RNA-based therapeutics analysis. Additionally, the comprehensive therapeutic information

is curated in the form of data sets available for free access to researchers. The extensive genomic and protein-specific investigation provides putative vaccine designs and RNA therapeutics options for utility in both advanced computational and experimental research. The novel platform, with protein-specific vaccine designs for each species and shortlisted potential siRNAs, microRNAs (miRNAs), and sgRNAs with all the necessary information, will aid in future therapeutic strategies against mammarenaviruses infections.

## Methodology

### Data collection

The whole-genome sequences (L and S segments) information used in this study were retrieved from the available online platform National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) (45), whereas the protein sequences (Z, L, G, N) of human-infecting mammarena species were collected from UniProt (<https://www.uniprot.org/>) (46). Further analysis was performed for the shortlisted human hemorrhagic fever-causing mammarena species. The accession and basic information of the genomic data set comprising of seven MMV species included in the study are given in Table 1. A list of protein sequences and accession IDs used in the study are also listed (Supplementary Sheet S1). The basic data information and sequences were then subjected to further analysis.

### Data processing

#### Epitopes prioritization

All the protein sequences of each mammarena species were initially scanned for immunogenic cytotoxic T lymphocyte (CTL) epitopes, B lymphocyte (B cell) epitopes, Helper T lymphocyte (HTL) epitopes and IFN-gamma-inducing peptides. The obtained epitope sequences for each species were further utilized to design highly immunogenic and antigenic epitopes-based *in silico* vaccines against each strain. To achieve the desired objectives, CTL epitopes for each protein of all species were predicted with the help of NetCTL 1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>) (47) and further characterized on the basis of combined score. The cut-off value used to predict CTL epitopes was set at 0.75. Similarly, B-cell epitopes prediction was carried out through ABCPred (<http://crdd.osdd.net/raghava/abcpred/>) server (48).

The predicted linear B-cell epitopes were further filtered with a defined cut-off score of 0.5 in the process. Epitope ranking was done based on the binding score: the higher the score, the higher the probability of peptide inducing an immune response. Next, HTL epitopes (15mer) were obtained from the immune epitope database (IEDB) server (<http://tools.iedb.org/mhcii/>) (49) that showed good affinity for human Major Histocompatibility Complex (MHC) molecules (HLA-DRB1\*01:02, HLA-DRB1\*01:01, HLA-DRB1\*01:04, HLA-DRB1\*01:03 and HLA-DRB1\*01:05), whereas the percentile ranking is inversely proportional to epitopes binding affinity and implies that a lower percentile rank would depict higher binding affinity (49).

Furthermore, IFN- $\gamma$ -inducing peptides were filtered among these positive MHC-II peptides by employing IFNepitope web server (<http://crdd.osdd.net/raghava/ifnepitope/>) (50). The predictions were performed using IFNepitope server. Next, to select the best combination of epitopes that passes all experimental principles, antigenic epitopes were screened among the predicted cell epitopes by using Vaxijen v2.0 (51) with a default threshold of 0.4. To discriminate between allergens and nonallergens, AllerTOP v2.0 (52) based on the k-nearest neighbors approach was used. The analyzed shortlisted peptides for each target protein with increased potential efficacy were included in further vaccine constructs.

### Vaccine designing

Computational methods are of great interest to understand the molecular mechanisms of pathogenesis, drug resistance, and the development of novel therapeutics (44, 53–57). All the predicted epitopes for each protein were ranked accordingly based on the higher binding affinity. The final vaccine candidates were composed of adjuvant CTL, HTL (IFN +ive), and B-cell epitopes joined together by AAY, GP GPG, and KK linkers (58, 59), respectively. Herein, the vaccine sequences were further stabilized with added N-terminal human beta defensin-2 (hBD-2) sequence to ensure enhanced immunogenic response (60). The vaccine construct also needed to be antigenic for eliciting the proper immune response. For this purpose, the VexiJen server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (51) was employed to predict the vaccine's antigenicity while keeping the threshold at the default 0.4. Another critical parameter, allergenicity, was predicted with the help of AllgPred server (<http://crdd.osdd.net/raghava/allgpred/>) (61) at an accuracy of around 85%. Allergenic sequence can be identified when there is a score greater than the threshold ( $>-0.4$ ). Physiochemical properties such as amino acid composition, molecular weight, theoretical pI, *in vivo* and *in vitro* half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) for experimental processing parameters were also employed to verify the vaccine. It was performed to unveil these properties for each vaccine construct by opting for an online webserver ProtParam (<https://web.expasy.org/protparam/>) (62). Furthermore, the 3D structures for all the vaccine constructs were predicted using the Robetta web server (<https://robetta.bakerlab.org/>) (63). In this procedure, the submitted sequences undergo domain-based initial recognition to forecast structure. This is followed by 3D modeling of submitted sequences depending on the type of templates available in the database. If matching templates are available, then comparative modeling is

performed; otherwise, *de novo* modeling for 3D structures is performed. Finally, to address therapeutic implications, all the developed vaccine designs were listed and included as a separate data set for all the proteins of each MMV species.

### Genome-based therapeutic RNA screening

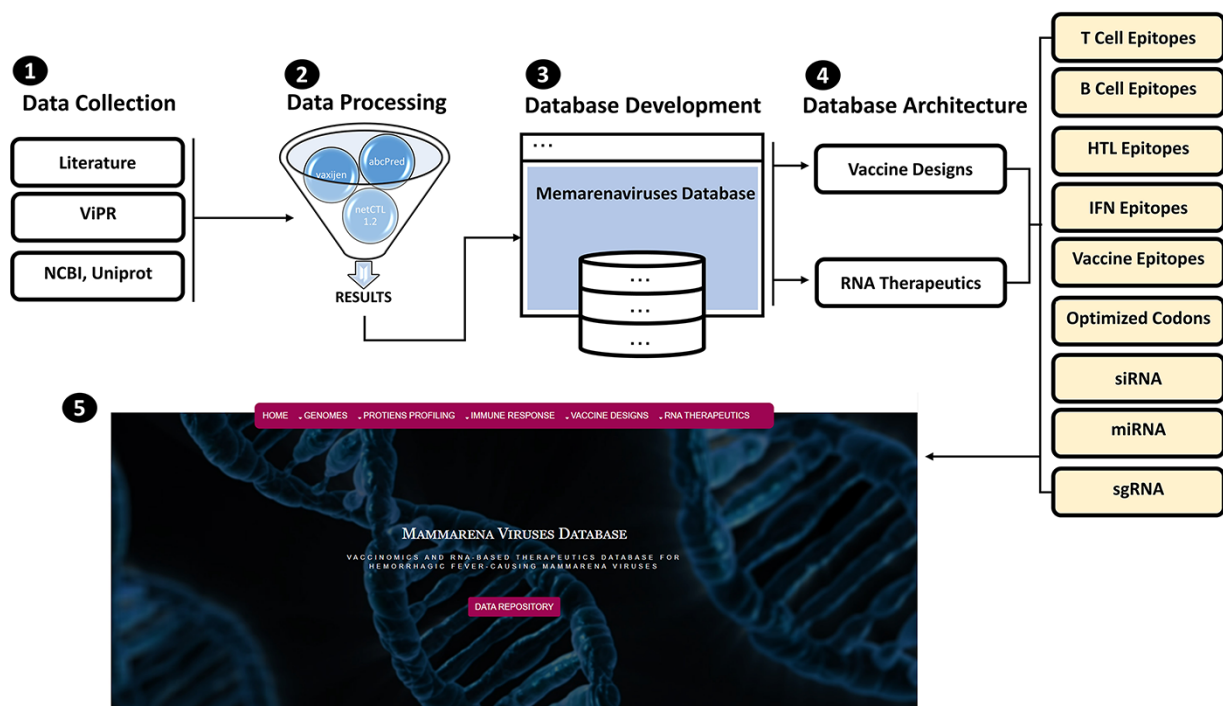
The genome sequences (L and S Segments) for each species were further analyzed to predict siRNAs against each virus. Herein, virus-specific VIRsiRNApred server was employed with utilizing model 2 (64). The model is constructed based on integrated variable features, including hybrid nucleotide frequencies, thermodynamic properties, and binary pattern of already identified 1725 viral siRNAs. siRNAs that are highly efficacious with inhibition  $\geq 50\%$  were included. To evaluate immunomodulatory (IM) impact, the imRNA tool (65) was utilized to investigate IM and non-IM siRNAs. Similarly, putative miRNAs for Mammarena (MM) viruses were also predicted using two-step approaches. First, VMir algorithm (66) was used to predict precursor miRNA (pre-miRNA) hairpins by deploying default parameters. Second, the Mature Bayes tool (67) was used to identify mature miRNAs. All possible sgRNAs for MM viruses were also predicted using the ge-CRISPR tool (68) based on the Protospacer Adjacent Motif. This algorithm scans 'NGG' motifs in both forward and reverse strands of the genome and picks up putative sgRNAs located 20 nucleotides upstream of the motif. A regression-based algorithm was run on geCRISPR predictions to predict sgRNA with an efficiency of 0% to 100%.

### Development of database

The intricate process of database development was followed by using Apache HTTP (Hypertext Transfer Protocol) server v2.2.1.7 through open-source Linux, MySQL (My Structured Query Language) and PHP (Hypertext Preprocessor) to develop and deploy online the 'MMV-db' database. Front-end development and user interaction interface were designed using CSS (Cascading Style Sheets), HTML (Hypertext Markup Language), PHP, and JavaScript, which also provides searching and downloading function. For back-end development of the database WAMP (Windows, Apache, MySQL, PHP) server accompanied by scripting in environments like HTML and PHP was used. Data storage, manipulation and retrieving from the databases were managed through MySQL to confer complete control over the web contents.

### Results and discussion

MMV-db focus spanned from basic protein features profiling to advanced epitopes-based vaccine designs and RNA-based therapeutics for all human-infecting MM viruses. This database is a collective platform for a total of seven hemorrhagic fever-causing-related mammarenaviruses. The database includes multiple-features profiling, including genome and proteins sequences, vaccine designs and therapeutic RNAs information represented in different tabs of the developed online platform. The overall workflow of the strategy, including the data source utilized in the design of this database, has been given in Figure 1.



**Figure 1.** (From 1 to 5) represents the general schematic workflow and different steps followed in the development of mammarenavirus database.

### Immune-based features profiling for mammarenaviruses

The antigenic and nonantigenic proteins for each species were identified with a VaxiJen threshold scoring system (51). The server utilizes an alignment-free, covariance-based approach with a focus on the properties of amino acids (51). We choose the target organism as a virus and initiated the analysis with a sequence-based output with default criteria. The antigenicity profiling of all the proteins was performed including all the mammarenaviruses. The cut-off value of 0.4 was used as an indicative threshold to differentiate between potential viral antigenic and nonantigenic proteins (51). Proteins were further subjected to allergenicity prediction analysis. The performed allergenicity check helps to prevent any possible allergic responses in the host (69). The server allpred v. 2.0 (70) was utilized to predict the allergenicity of the proteins, whereas a score greater than the threshold ( $>0.4$ ) represents allergenic sequences [49]. The input sequence was added as a single letter amino acid code, while the selected prediction approach was an amino acid composition-based Support Vector Machines (SVM) module (70). The immune-based analysis of antigenicity and allergenicity was performed to profile each species-specific protein. The output data were arranged on the basis of obtained scores to differentiate between antigenic, nonantigenic, allergenic and nonallergenic proteins. The antigenicity and allergenicity profiles for each of the four (Z, L, G and N) specific proteins of all species are shown in Figure 2.

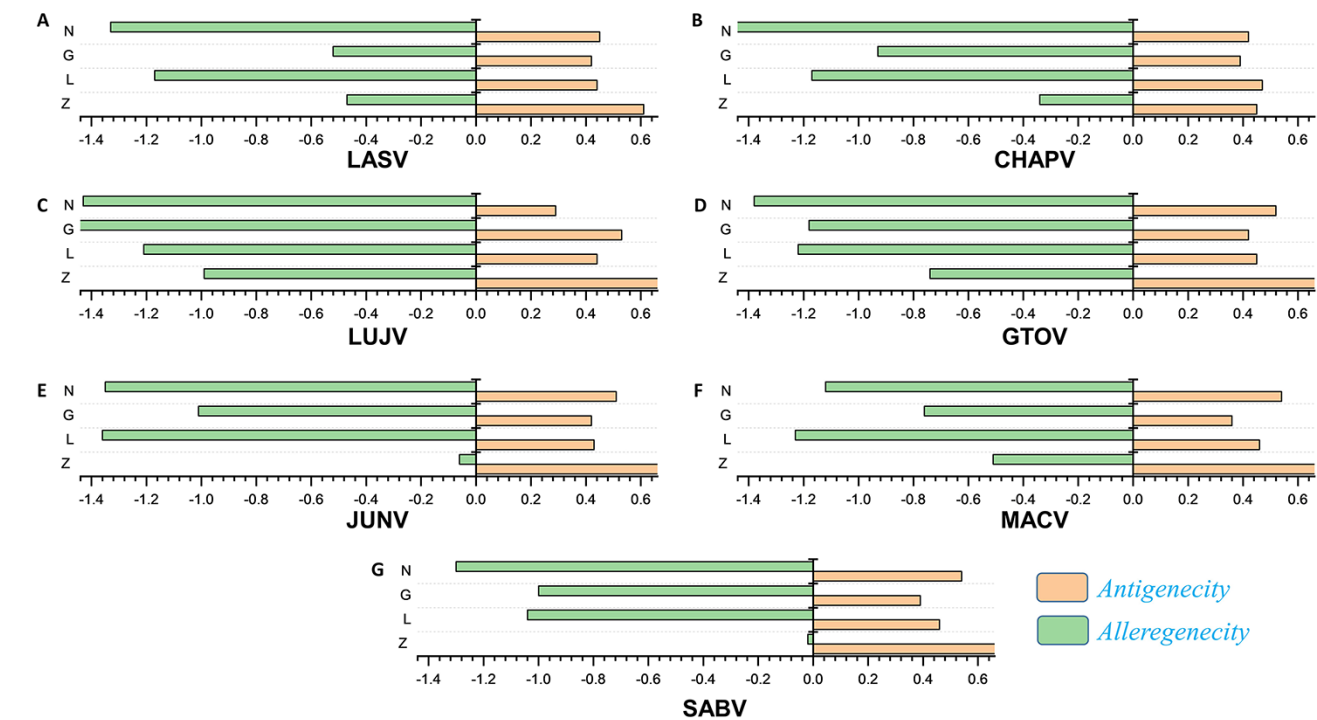
Similarly, a total of 639 CTL epitopes, 2275 B-cell epitopes, 116 746 HTL epitopes and 9945 IFN epitopes were analyzed for all the mammarenavirus species. The predicted epitopes were further classified on the basis of species-specific proteomes. The total count of whole proteome-specific T-cell, B-cell, HTL and IFN-inducing epitopes were calculated for each studied species and presented as shown in Figure 3A–D, respectively.

### Immunogenic and putative vaccine epitopes prediction

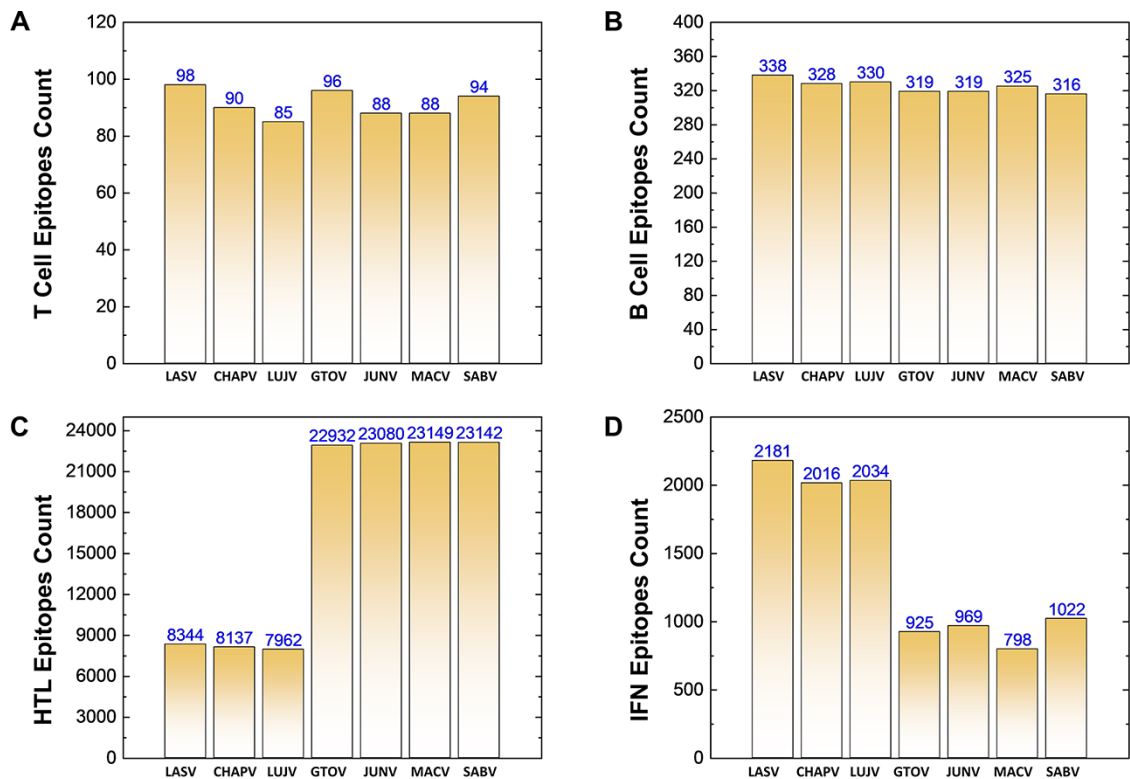
The immunogenic and potential vaccine epitopes screening was performed in a sequential manner for all the seven mammarenaviruses i.e. LASV, CHAPV, LUJV, GTOV, JUNV, MACV, and SABV. Moreover, the order of proteins in the results is presented as Z, L, G (representing GPC) and N (representing NP) for each species. The prediction of T-cell, B-cell and HTL epitopes obtained after protein sequence-based analysis are presented with the total number of each type of epitope for a specific protein in individual species (Figure 4).

First, the prediction of potential CTL epitopes related to the four L, S, G and N was performed. For this purpose, Net CTLpan v1.2 was utilized (47), and predictions were performed using 12 different supertypes of human leucocyte antigen with the rest of the default parameters. The sequences of the peptides having a % Rank  $<1\%$  ( $<E$ ) were chosen as MHC binders according to the given selection criteria of the webserver. The protein-specific T-cell epitopes (Supplementary Sheet S2) were further classified for each species, as shown in Figure 4A. Moreover, linear B-cell epitopes also work as antigen and interact with B-cell receptors. Identification of B-cell epitopes is vital to generate a protective host antibody response. Correspondingly, total B-cell epitopes were predicted (Supplementary Sheet S3) by using ABCpred server (48) and further classified for each protein as shown in Figure 4B. Likewise, HTL epitopes based on protein sequences were also predicted using the IEDB MHC-II binding prediction tool (49). Using the selection criterion mentioned in methodology, MHC-II binders were predicted, and a collection of these epitopes were further screened for potential IFN induction property. The total number of HTL epitopes for each protein of the examined species is shown in Figure 4C. Furthermore, the IFN- $\gamma$  produced by T-helper cells



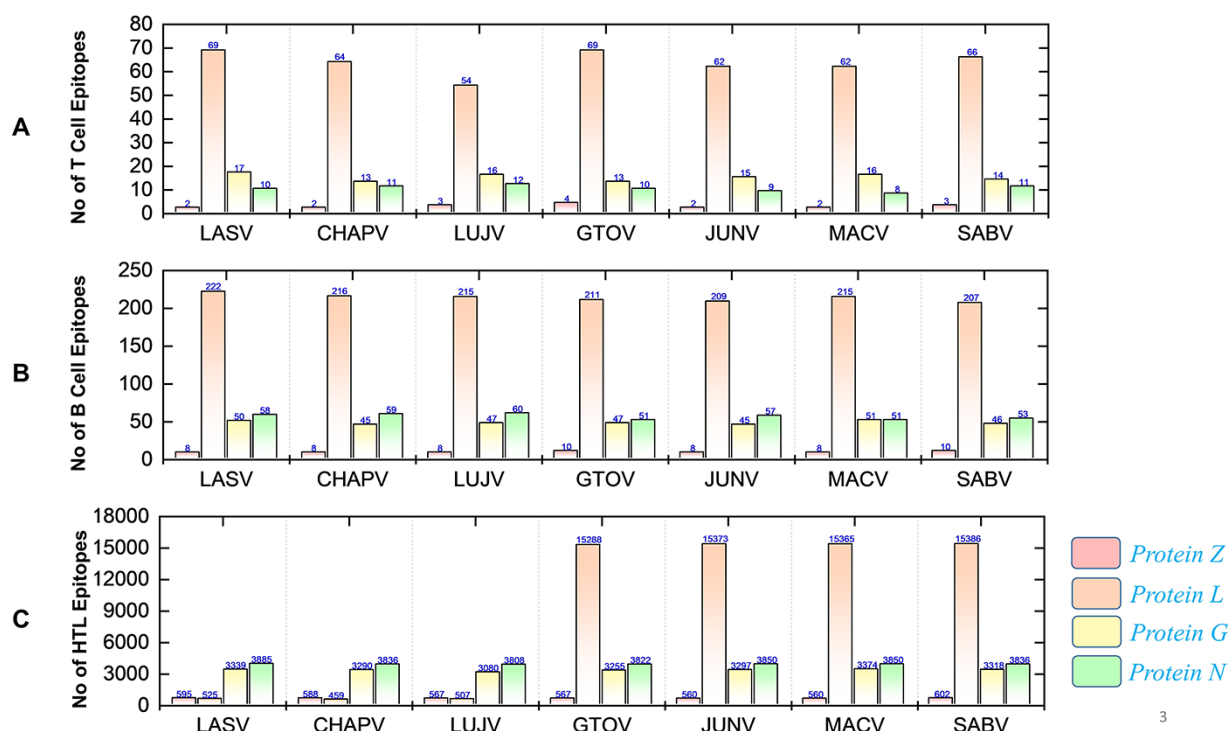


**Figure 2.** (A–G) Represents the antigenicity and allergenicity scores for whole-protein sequences in the individually studied mammarenaviruses.

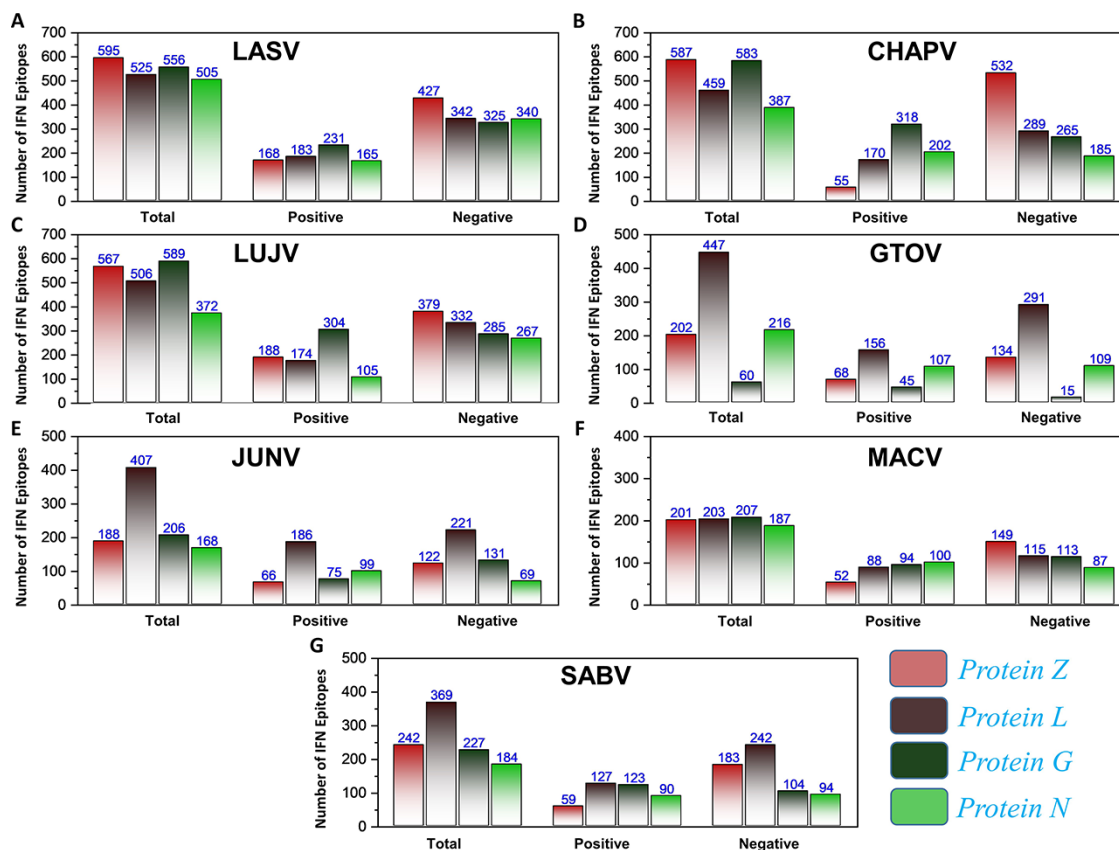


**Figure 3.** (A–D) Demonstrates the total identified number of predicted immune epitopes for each specific protein; (A) represents the total number of T-cell epitopes, (B) represents the total number of B-cell epitopes, (C) represents the total number of HTL epitopes, while (D) represents the total number of IFN epitopes predicted for each protein for the seven mammarenavirus species, respectively.

is vital in clearing the virus; therefore, selected HTL epitopes were subjected to shortlisting based on IFN- $\gamma$  induction. The selected HTL epitopes, as shown in Supplementary Sheet S4, were further characterized for IFN-inducing positive and negative epitopes for all proteins of each species as shown in Figure 5.



**Figure 4.** Demonstrates the total identified number of classified immune epitopes for each specific protein; (A) represents the number of specified T-cell epitopes, (B) represents the number of B-cell epitopes and (C) represents the number of HTL epitopes predicted for each protein for the seven mammarena species, respectively.



**Figure 5.** (A-G) Represents the total number of predicted IFN epitopes along with positive and negative IFNs for each protein of the seven MMVs, respectively.

Further evaluation made by subjecting epitope sequences to antigenicity and allergenicity prediction helped to identify putative vaccine candidates. After the extensive analysis, putative vaccine epitopes from all groups were shortlisted, with qualifying parameters for each protein of all species included in further vaccine designs.

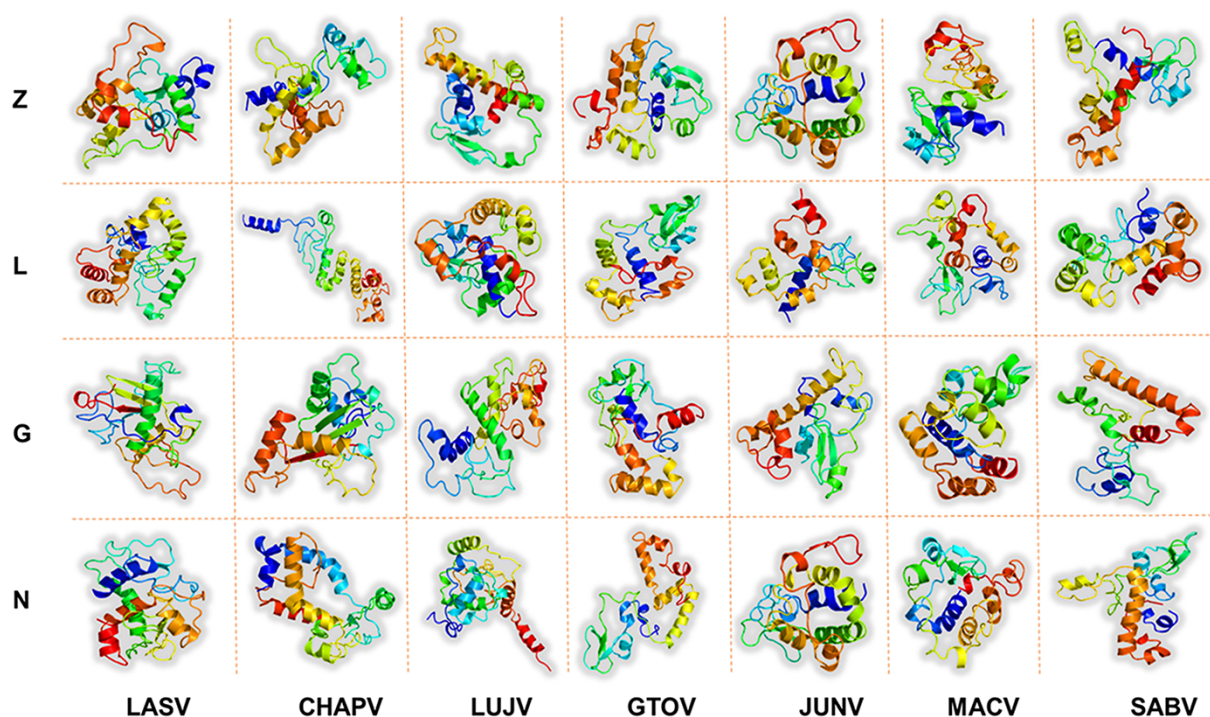
### Construction of multi-epitope vaccines

Vaccines therapy based on targeting large protein or whole organisms is considered highly effective in reducing the burden of viral infections and associated mortality (71) around the world. However, it is also harbored by a high antigenic burden of such vaccines, imprecise immunological responses and associated reactogenic reactions (72, 73). Alternatively, peptides-based vaccine designs are more appropriate options for clinical use. The feasibility of peptide vaccines is reflected by reduced production time and cost, robust immune-specific response provoking, lowered risk of antigen-induced allergies and flexibility to varied antigens (74, 75). However, the administration of peptide vaccines needs appropriate adjuvants to ensure immunogenic nature (76). This also shelters the inclusion of feebly antigenic individual peptides and enhances immunogenicity in the overall multi-epitope-based vaccine designs (77). Herein, the designed Multi-epitopes Vaccine Construct (MEVC) sequences based on shortlisted T-cell, HTL and B-cell epitopes with fulfilled selection parameters for robust immune response are shown (Supplementary Sheet S5). The topographical arrangement of the epitope-based vaccines was further equipped with appropriate linkers. Different types of linkers such as EAAK, GP GPG, AAY and KK were used to join these small peptides together and design a full-length MEVC. The linkers utilized in MEVC constructs are related to the facilitation of epitopes display and efficient induction

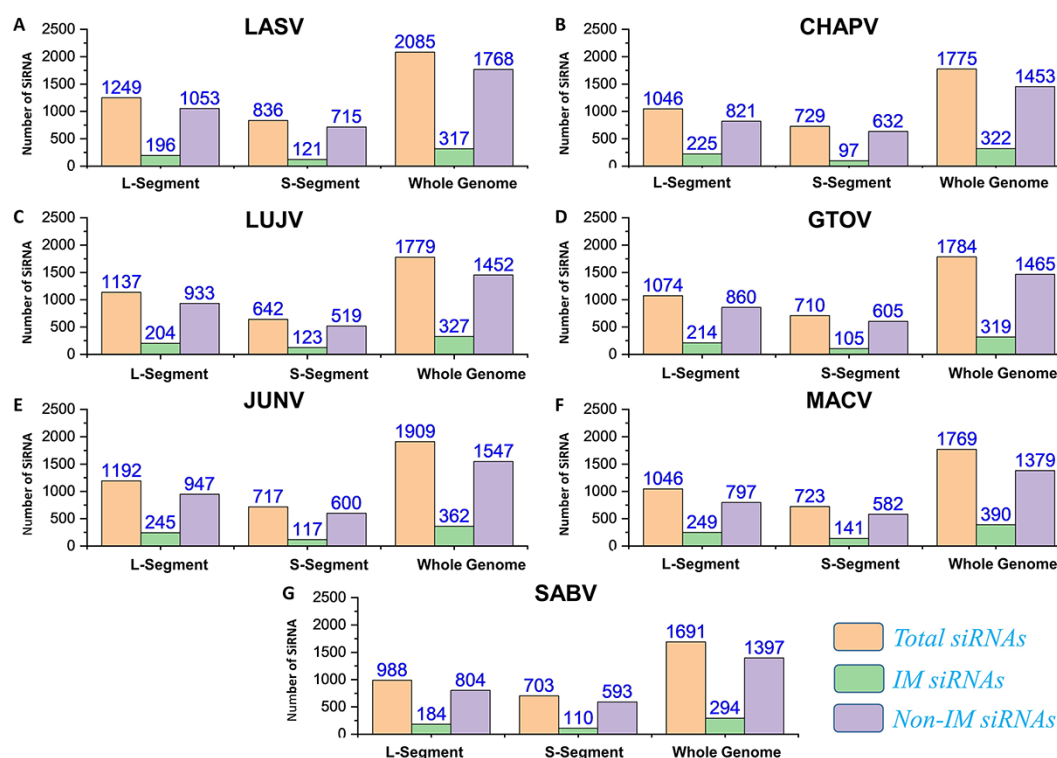
of immune response (78). Altogether, the linkers hindered epitopes folding (58), and enhanced adjuvant immunogenicity (59) was considered during MEVC designing. Thus, N-terminal of the modeled peptide vaccines was furnished with a nontoxic adjuvant called human beta defensin-2 (hBD-2) with added EAAAK linker. The adjuvant (hBD-2) has a regulated expression and also potentiates immune response against attached antigen [83]. The amino acid sequences for each MEVC were further verified for antigenic and allergenic properties (Supplementary Sheet S6) to elude any adverse immune reaction in further experimental designs. The systematic analysis was followed for all the potential vaccine construct designs for each protein of the seven mammarena species. The final MEVC designs were also subjected to codon optimization using the JCAT server (79), and explored DNA sequences are presented (Supplementary Sheet S6) for further utility in procedures of *in silico* and experimental cloning. Furthermore, the robetta-predicted 3D structures of final vaccine constructs for each protein of the seven mammarena species are shown in Figure 6.

### Analyzing physicochemical properties

The ProtParam server (62) was utilized to determine the MW and Isoelectric Point (pI) of the MEVC sequences designed against each protein, respectively. The pI values suggested a more suitable acidic nature of the candidate vaccines. Similarly, the distribution number of positively and negatively charged residues, extinction coefficient of the protein in water, *in vitro* half-life of the vaccine constructs in mammalian reticulocytes and *in vivo* in yeast and *Escherichia coli* were also listed. Furthermore, the aliphatic index and GRAVY value of the proposed vaccine were determined, which reflects the stability of the protein with temperature change. Similarly, the



**Figure 6.** Represents the refined 3D structures of MEV designs for each protein of all the studied mammarenavirus species. x-axis includes names of the MMV species, while y-axis shows the vaccine-targeted proteins.



**Figure 7.** (A–G) Represents number of total siRNAs and classified IM/non-IM siRNAs identified for genomic segments L, S and whole genome of each mammarenavirus.

obtained negative GRAVY value indicates hydrophilic nature and improved interactions of the vaccines with nearby water molecules. The physiochemical properties of each vaccine with all the details are provided in Supplementary Sheet S6.

### Genome-based prioritization of candidate therapeutic RNAs

The potential therapeutic RNAs were also identified for all the seven species of human-infecting mammarenaviruses. The genome-based analysis was performed for each species as L, S segments and whole genome. The predicted therapeutic RNAs were further characterized on the basis of efficacy scores, including siRNAs, IM/non-IM siRNAs, miRNAs and sgRNAs. The data are also presented in different tabs with the accessible necessary information in the online resource.

#### SiRNAs

Silencing viral genes by RNA interference is an excellent alternative therapeutic option (80). Regarding this, all possible efficient siRNAs targeting genes of the mammarenavirus species were investigated. The analysis performed was specified to L, S and whole-genome segments for all the studied mammarenavirus species. The analysis involved a collective number of 2085, 1775, 1779, 1784, 1909, 1769 and 1691 siRNAs for each of the species, respectively. This was achieved after deploying the VIRsiRNA algorithm (64) with identifying potent siRNAs of more than 50% predicted ability of inhibition. Following, all these siRNAs with higher efficacy scores were further subjected to IM potential analysis by using the iMRNA tool (65). Total siRNAs, immuno-modulatory (IM) siRNAs and non-IM siRNAs with predicted counts for L, S and whole-genomic segments specific to each species are

presented in Figure 7A–G. All the relevant information of the total siRNAs and their therapeutics-related IM potential scores (Supplementary Sheet S7) are also available in separate tabs of the MM-db virtual resource.

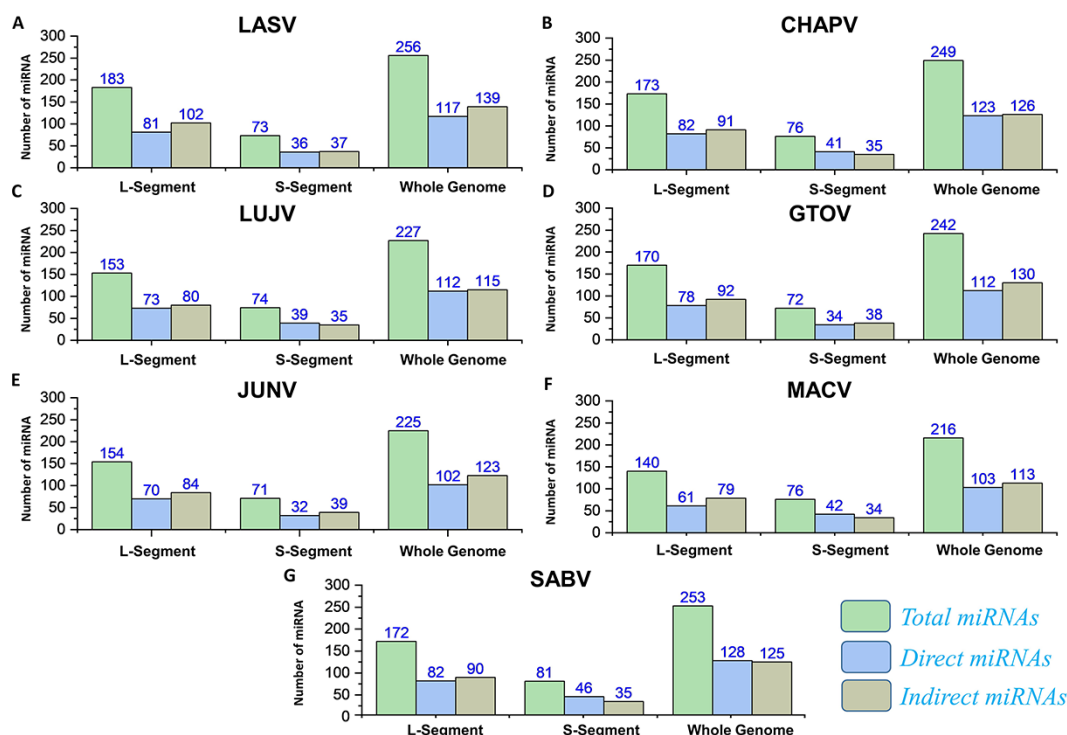
#### MicroRNAs

The applications of miRNA interference-based therapies spans from targeting a single gene to complex cellular pathways (80). In our data, all the pre-miRNAs along with mature miRNAs (3p and 5p) were also identified by deploying VMIR tool (Analyzer and Viewer) (66). Complete information of the hairpin sequence, its length, start-end location, mature miRNA sequence, orientation and scores were retrieved. This analysis revealed a total number of 256, 249, 227, 242, 225, 216 and 253 potential miRNAs for whole genome of each mammarenavirus. The mature miRNAs (3p and 5p) (Supplementary Sheet S8) obtained after mature Bayes tool analysis (67) were further classified on the basis of orientation as direct or reverse. The computed number of orientation (direct and reverse) specific and total miRNAs for each L, S and whole-genome segments of mammarenavirus species are shown in Figure 8. The resulting mature miRNAs with mentioned orientation are also available in the data sets provided on the server.

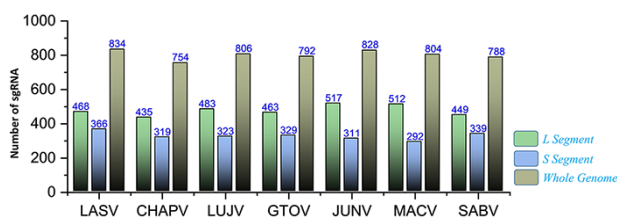
#### Putative sgRNAs

The recently advanced biological strategies also involve sgRNA-based therapeutic interventions (81). Significantly, putative efficient sgRNAs were also enlisted after utilizing the ge-CRISPR-based analysis (68). The useful information regarding sgRNA sequences in 5' to 3' direction, start and end positions, Guanine-Cytosine (GC)%, Protospacer Adjacent





**Figure 8.** (A–G) Represents the number of total miRNAs and orientation-based classified direct/indirect mature RNAs identified for L, S and whole-genome segments of each mammarenavirus.



**Figure 9.** Represents the total number of sgRNAs identified for L, S and whole-genome segments respective to each of the seven mammarenavirus.

Motifs (PAM) and predicted sgRNA efficiency is also presented in Supplementary Sheet S9. The number of predicted sgRNAs for the different genomic segments of each species is shown in Figure 9. The data are also made easily available for researchers in the online platform to be utilized in the identification of CRISPR targets, including efficient sgRNAs in further experimentation designs against mammarenaviruses.

## Conclusion

Despite the advancement in the field of biology, many viral disease still suffer from the lack of proper treatment i.e. vaccines or drugs. Our investigations present 28 different multi-epitopes-based protein-specific vaccine designs with a critical role in immune response induction against seven different mammarena species. Further exploration of physicochemical properties also suggests experimental feasibility of the vaccines. However, the efficacy and safety of the highly specific MEVC candidates need further demonstration through lab

experiments and remain elusive. The developed online platform also offers RNA-based therapeutic options for further investigations against human-infecting MMVs. Altogether, MMV-db offers a novel predisposed source of advanced multi-epitopes-based vaccines and RNA-based therapeutics against MMV species pathogenic to humans.

## Supplementary data

Supplementary data are available at Database Online.

## Acknowledgements

The computations were partially performed at the Pengcheng Lab and the Center for High-Performance Computing, Shanghai Jiao Tong University.

## Funding

Dong-Qing Wei is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (Grant No. 32070662, 61832019, 32030063), the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2021ZD02).

## Conflict of interest

The authors declare no conflict of interest.

## Data Availability

All the data are given in the Supplementary materials.

## References

- Clegg, J. (2002) Molecular phylogeny of the arenaviruses. *Arenaviruses I*, 262, 1–24.
- Bowen, M.D., Peters, C.J. and Nichol, S.T. (1996) The phylogeny of New World (Tacaribe complex) arenaviruses. *Virology*, 219, 285–290.
- Fulhorst, C.F., Cajimat, M.N., Milazzo, M.L. et al. (2008) Genetic diversity between and within the arenavirus species indigenous to western Venezuela. *Virology*, 378, 205–213.
- Jay, M.T., Glaser, C. and Fulhorst, C.F. (2005) The arenaviruses. *J. Am. Vet. Med. Assoc.*, 227, 904–915.
- Buchmeier, M., Southern, P., Parekh, B. et al. (1987) Site-specific antibodies define a cleavage site conserved among arenavirus GP-C glycoproteins. *J. Virol.*, 61, 982–985.
- Lenz, O., ter Meulen, J., Klenk, H.-D. et al. (2001) The Lassa virus glycoprotein precursor GP-C is proteolytically processed by subtilase SKI-1/S1P. *Proc. Natl. Acad. Sci. USA*, 98, 12701–12705.
- Wright, K., Spiro, R., Burns, J. et al. (1990) Post-translational processing of the glycoproteins of lymphocytic choriomeningitis virus. *Virology*, 177, 175–183.
- Eichler, R., Lenz, O., Garten, W. et al. (2006) The role of single N-glycans in proteolytic processing and cell surface transport of the Lassa virus glycoprotein GP-C. *Virol. J.*, 3, 41.
- Vela, E.M., Zhang, L., Colpitts, T.M. et al. (2007) Arenavirus entry occurs through a cholesterol-dependent, non-caveolar, clathrin-mediated endocytic mechanism. *Virology*, 369, 1–11.
- Martinez, M.G., Cordo, S.M. and Candurra, N.A. (2007) Characterization of Junin arenavirus cell entry. *J. Gen. Virol.*, 88, 1776–1784.
- Meyer, B., de La Torre, J. and Southern, P. (2002) Arenaviruses: genomic RNAs, transcription, and replication. *Arenaviruses I*, 262, 139–157.
- Elsner, B., Schwarz, E., Mando, O.G. et al. (1973) Pathology of 12 fatal cases of Argentine hemorrhagic fever. *Am. J. Trop. Med. Hyg.*, 22, 229–236.
- McLeod, C., Stookey, J.L., Eddy, G.A. et al. (1976) Pathology of chronic Bolivian hemorrhagic fever in the rhesus monkey. *Am. J. Pathol.*, 84, 211–224.
- Kenyon, R., McKee, K., Jr, Zack, P. et al. (1992) Aerosol infection of rhesus macaques with Junin virus. *Intervirology*, 33, 23–31.
- Peters, C., Jahrling, P., Liu, C. et al. (1987) Experimental studies of arenaviral hemorrhagic fevers. *Arenaviruses*, 134, 5–68.
- Salas, R., Pacheco, M., Ramos, B. et al. (1991) Venezuelan haemorrhagic fever. *Lancet*, 338, 1033–1036.
- Richmond, J.K. and Baglolle, D.J. (2003) Lassa fever: epidemiology, clinical features, and social consequences. *BMJ*, 327, 1271–1275.
- Sogoba, N., Feldmann, H. and Safronetz, D. (2012) Lassa fever in West Africa: evidence for an expanded region of endemicity. *Zoonoses Public Health*, 59, 43–47.
- Child, P., MacKenzie, R., Valverde, L. et al. (1967) Bolivian hemorrhagic fever. A pathologic description. *Arch. Pathol.*, 83, 434–445.
- McKee, K.T., Mahlandt, B., Maiztegui, J.I. et al. (1985) Experimental Argentine hemorrhagic fever in rhesus macaques: viral strain-dependent clinical response. *J. Infect. Dis.*, 152, 218–221.
- Carballal, G., Rodriguez, M., Frigerio, M. et al. (1977) Junin virus infection of Guinea pigs: electron microscopic studies of peripheral blood and bone marrow. *J. Infect. Dis.*, 135, 367–373.
- Laguens, R., Gonzalez, P., Ponziñibio, C. et al. (1986) Damage of human polymorphonuclear leukocytes by Junin virus. *Med. Microbiol. Immunol.*, 175, 177–180.
- Carballal, G., Cossio, P., Laguens, R. et al. (1981) Junin virus infection of Guinea pigs: immunohistochemical and ultrastructural studies of hemopoietic tissue. *J. Infect. Dis.*, 143, 7–14.
- De Guerrero, L., Boxaca, M., Weissenbacher, M. et al. (1977) Experimental infection of the Guinea pig with Junin virus. Clinical picture, dissemination, and elimination of the virus. *Medicina*, 37, 271–278.
- Maiztegui, J., Laguens, R., Cossio, P. et al. (1975) Ultrastructural and immunohistochemical studies in five cases of Argentine hemorrhagic fever. *J. Infect. Dis.*, 132, 35–43.
- Manzione, N.D., Salas, R.A., Paredes, H. et al. (1998) Venezuelan hemorrhagic fever: clinical and epidemiological studies of 165 cases. *Clin. Infect. Dis.*, 26, 308–313.
- Delgado, S., Erickson, B.R., Agudo, R. et al. (2008) Chapare virus, a newly discovered arenavirus isolated from a fatal hemorrhagic fever case in Bolivia. *PLoS Pathog.*, 4, e1000047.
- Harrison, L.H., Halsey, N.A., McKee, K.T., Jr et al. (1999) Clinical case definitions for Argentine hemorrhagic fever. *Clin. Infect. Dis.*, 28, 1091–1094.
- Enria, D.A. and Maiztegui, J.I. (1994) Antiviral treatment of Argentine hemorrhagic fever. *Antiviral Res.*, 23, 23–31.
- Enria, D., Fernandez, N., Briggiler, A. et al. (1984) Importance of dose of neutralising antibodies in treatment of Argentine haemorrhagic fever with immune plasma. *Lancet*, 324, 255–256.
- Enria, D., de Damilano, A., Briggiler, A. et al. (1985) Late neurologic syndrome in patients with Argentinian hemorrhagic fever treated with immune plasma. *Medicina*, 45, 615–620.
- García, C.C., Sepúlveda, C.S. and Damonte, E.B. (2011) Novel therapeutic targets for arenavirus hemorrhagic fevers. *Future Virol.*, 6, 27–44.
- Enria, D.A., Briggiler, A.M. and Sánchez, Z. (2008) Treatment of Argentine hemorrhagic fever. *Antiviral Res.*, 78, 132–139.
- Snell, N.J. (2001) Ribavirin-current status of a broad spectrum antiviral agent. *Expert Opin. Pharmacother.*, 2, 1317–1324.
- Khan, S.H., Goba, A., Chu, M. et al. (2008) New opportunities for field research on the pathogenesis and treatment of Lassa fever. *Antiviral Res.*, 78, 103–115.
- Kiso, M., Takahashi, K., Sakai-Tagawa, Y. et al. (2010) T-705 (favipiravir) activity against lethal H5N1 influenza A viruses. *Proc. Natl. Acad. Sci. USA*, 107, 882–887.
- Morrey, J.D., Taro, B.S., Siddharthan, V. et al. (2008) Efficacy of orally administered T-705 pyrazine analog on lethal West Nile virus infection in rodents. *Antiviral Res.*, 80, 377–379.
- Barrera Oro, J.G. and McKee, K.T. (1991) Toward a vaccine against Argentine hemorrhagic fever. *Bull. PAHO*, 25, 118–126.
- Ölschläger, S. and Flatz, L. (2013) Vaccination strategies against highly pathogenic arenaviruses: the next steps toward clinical trials. *PLoS Pathog.*, 9, e1003212.
- Branco, L.M., Grove, J.N., Geske, F.J. et al. (2010) Lassa virus-like particles displaying all major immunological determinants as a vaccine candidate for Lassa hemorrhagic fever. *Virol. J.*, 7, 1–19.
- Bredenbeek, P.J., Molenkamp, R., Spaan, W.J. et al. (2006) A recombinant Yellow Fever 17D vaccine expressing Lassa virus glycoproteins. *Virology*, 345, 299–304.
- Pushko, P., Geisbert, J., Parker, M. et al. (2001) Individual and bivalent vaccines based on alphavirus replicons protect Guinea pigs against infection with Lassa and Ebola viruses. *J. Virol.*, 75, 11677–11685.
- Khan, T., Khan, A., Nasir, S.N. et al. (2021) CytomegalovirusDb: multi-Omics knowledge database for Cytomegaloviruses. *Comput. Biol. Med.*, 135, 104563.
- Khan, A., Khan, T., Nasir, S.N. et al. (2021) BC-TFdb: a database of transcription factor drivers in breast cancer. *Database*, 2021, 1–10.
- Jenuth, J.P. (2000) The NCBI. In: Misener, S. and Krawetz, S.A. (eds.) *Bioinformatics Methods and Protocols*. Springer, New Jersey, pp. 301–312.

46. Magrane,M. (2011) UniProt knowledgebase: a hub of integrated protein data. *Database*, 2011, 1–13.
47. Larsen,M.V., Lundegaard,C., Lamberth,K. *et al.* (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinf.*, 8, 1–12.
48. Saha,S. and Raghava,G.P.S. (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, 65, 40–48.
49. Vita,R., Mahajan,S., Overton,J.A. *et al.* (2019) The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.*, 47, D339–D343.
50. Dhanda,S.K., Vir,P. and Raghava,G.P. (2013) Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct.*, 8, 1–15.
51. Doytchinova,I.A. and Flower,D.R. (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf.*, 8, 1–7.
52. Dimitrov,I., Bangov,I., Flower,D.R. *et al.* (2014) AllerTOP v. 2—a server for in silico prediction of allergens. *J. Mol. Model.*, 20, 1–6.
53. Khan,A., Zia,T., Suleman,M. *et al.* (2021) Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: an insight from structural data. *J. Cell. Physiol.*, 236, 7045–7057.
54. Khan,A., Khan,S., Saleem,S. *et al.* (2021) Immunogenomics guided design of immunomodulatory multi-epitope subunit vaccine against the SARS-CoV-2 new variants, and its validation through in silico cloning and immune simulation. *Comput. Biol. Med.*, 133, 104420.
55. Khan,A., Ali,S.S., Khan,M.T. *et al.* (2020) Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro). *J. Biomol. Struct. Dyn.*, 39, 1–12.
56. Khan,T., Khan,A., Ali,S.S. *et al.* (2021) A computational perspective on the dynamic behaviour of recurrent drug resistance mutations in the pncA gene from Mycobacterium tuberculosis. *RSC Adv.*, 11, 2476–2486.
57. Khan,A., Heng,W., Wang,Y. *et al.* (2021) In silico and in vitro evaluation of kaempferol as a potential inhibitor of the SARS-CoV-2 main protease (3CLpro). *Phytother. Res.*, 35, 2841–2845.
58. Negahdaripour,M., Nezafat,N., Eslami,M. *et al.* (2018) Structural vaccinology considerations for in silico designing of a multi-epitope vaccine. *Infect. Genet. Evol.*, 58, 96–109.
59. Parvizpour,S., Pourseif,M.M., Razmara,J. *et al.* (2020) Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discov. Today*, 25, 1034–1042.
60. Schröder,J.-M. and Harder,J. (1999) Human beta-defensin-2. *Int. J. Biochem. Cell Biol.*, 31, 645–651.
61. Saha,S. and Raghava,G. (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.*, 34, W202–W209.
62. Garg,V.K., Avashthi,H., Tiwari,A. *et al.* (2016) MFPPi-multi FASTA ProtParam interface. *Bioinformation*, 12, 74.
63. Chivian,D., Kim,D.E., Malmström,L. *et al.* (2003) Automated prediction of CASP-5 structures using the Robetta server. *Proteins*, 53, 524–533.
64. 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, 1–12.
65. Nagpal,G., Chaudhary,K., Dhanda,S.K. *et al.* (2017) Computational prediction of the immunomodulatory potential of RNA sequences. In: Bindewald, E. and Shapiro, B.A. (eds.) *RNA Nanostructures*. Springer, New York, pp. 75–90.
66. Sullivan,C.S. and Grundhoff,A. (2007) Identification of viral microRNAs. *Meth. Enzymol.*, 427, 1–23.
67. 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.
68. 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, 1–12.
69. McKeever,T.M., Lewis,S.A., Smith,C. *et al.* (2004) Vaccination and allergic disease: a birth cohort study. *Am. J. Public Health*, 94, 985–989.
70. Sharma,N., Patiyal,S., Dhalla,A. *et al.* (2020) AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief. Bioinf.*, 22, 1–12.
71. Yen,C., Hyde,T.B., Costa,A.J. *et al.* (2015) The development of global vaccine stockpiles. *Lancet Infect. Dis.*, 15, 340–347.
72. Chen,R.T. and Hibbs,B. (1998) *Vaccine Safety: Current and Future Challenges*. SLACK Incorporated Thorofare, New Jersey.
73. Whitaker,J.A., Ovsyannikova,I.G. and Poland,G.A. (2015) Adversomics: a new paradigm for vaccine safety and design. *Expert Rev. Vaccines*, 14, 935–947.
74. Skwarczynski,M. and Toth,I. (2016) Peptide-based synthetic vaccines. *Chem. Sci.*, 7, 842–854.
75. Parmiani,G., Russo,V., Maccalli,C. *et al.* (2014) Peptide-based vaccines for cancer therapy. *Hum. Vaccin Immunother.*, 10, 3175–3178.
76. Reed,S.G., Bertholet,S., Coler,R.N. *et al.* (2009) New horizons in adjuvants for vaccine development. *Trends Immunol.*, 30, 23–32.
77. Wack,A. and Rappuoli,R. (2005) Vaccinology at the beginning of the twenty-first century. *Curr. Opin. Immunol.*, 17, 411–418.
78. Reche,P., Flower,D.R., Fridkis-Hareli,M. *et al.* (2018) *Peptide-based Immunotherapeutics and Vaccines 2017*. Hindawi, London.
79. Grote,A., Hiller,K., Scheer,M. *et al.* (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.*, 33, W526–W531.
80. Davidson,B.L. and McCray,P.B. (2011) Current prospects for RNA interference-based therapies. *Nat. Rev. Genet.*, 12, 329–340.
81. Zhou,L.-Y., Qin,Z., Zhu,Y.-H. *et al.* (2019) Current RNA-based therapeutics in clinical trials. *Curr. Gene. Ther.*, 19, 172–196.