GENOME DYNAMICS AND ADAPTATION EVENTS OF NEW WORLD ARENAVIRUSES AT THE ZOONOTIC-HUMAN INTERFACE

Thesis Submitted For the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Biotechnology

By

Himani Malhotra
11814490

Supervised By

Dr. Arvind Kumar

Co-Supervised by

Dr. Ayan Roy

LOVELY PROFESSIONAL UNIVERSITY
PUNJAB
2022

DECLARATION

I declare that the thesis entitled **"Genome dynamics and adaptation events of new world arenaviruses at the zoonotic-human interface"** has been prepared by me under the guidance of Dr. Arvind Kumar, Assistant Professor, Department of Biochemistry, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, and as per the requirement for the award of the degree of Doctor of Philosophy (Ph.D.) in Biotechnology (specialization in Bioinformatics) is entirely an authentic record of my own research work and ideas and references are duly acknowledged. It is further certified that the results incorporated in this thesis have not been submitted, in part or full, to any other university or institution for the award of any degree or diploma.

(Himani Malhotra)

Reg. No- 11814490

Department of Biotechnology

School of Bioengineering and Biosciences

Lovely Professional University

Phagwara, Punjab, India

Date:

CERTIFICATE

It is hereby certified that thesis entitled, "**Genome dynamics and adaptation events of new world arenaviruses at the zoonotic-human interface",** being submitted by **Ms. Himani Malhotra**, in Department of Biotechnology, School of Bioengineering and Biosciences, Lovely Professional University, Punjab, to award the degree of Doctor of Philosophy in Biotechnology is a record of bonafide research work carried out by **Ms. Himani Malhotra** under my supervision and guidance and has fulfilled all the requirements for the submission of the thesis. It is further certified that the results incorporated in this thesis have not been submitted, in part or full, to any other university or institution for the award of any degree or diploma.

Supervised By

Dr. Arvind Kumar

Assistant Professor

Department of Biochemistry

School of Bioengineering and Biosciences

Lovely Professional University, Punjab

Abstract

Various members of New World Arenavirus propagate high level of infection among humans. The present study was conducted to determine the genomic patterns of the Arenaviruses (JUNV, MACV, GTOV, SABV and CHPV) and analyze the convoluted procedure by which these viruses embrace to spread infection in their host. This analysis unravels the variations occurring in the codon usage patterns showing preference for AT rich codons among viral genome. Interestingly, analysis revealed compositional constraint and host selection pressure being the most important factors responsible for influencing the viral codon usage patterns. However, multivariate statistical analysis revealed hydrophobicity, aromaticity and coding sequence length as other factors affecting the protein biogenesis. The present study also analyzed the adaptation of viuses inside various hosts, revealing rapid progression of infection in humans and this knowledge can be adopted in designing vaccines against viruses with high level of accuracy. Further, molecular level phylogenetic analysis was performed, reflecting the true relationship of the out-group of viruses which can further be used to analyze substitution rates in each protein (GP, NP, Z and L) of JUNV, MACV, GTOV, SABV and CHPV. Results predicted Glycoprotein as highly conserved and subunit (GP1), the one which is involved in attachment to the host cell receptor causing virulence of viruses inside the host was selected as potential therapeutic target. Further, molecular docking, pertaining simulations were performed at binding sites of GP1of JUNV and MACV separately predicting novel ligands (MK-3207 and Dihydroergotamine) which can be further analyzed *in vitro* and *in vivo* as drugs against viruses and can be utilized as new therapeutics for JUNV and MACV.

Keywords: codon usage, host adaptation, ligands, phylogenetic analysis

Acknowledgement

Foremost I am very grateful to the God for sailing me through this journey and making this study possible.

I also take this opportunity to express my deep sense of gratitude and sincere thanks to my research supervisor, Dr.Arvind Kumar, Department of Biochemistry, Lovely professional University for his invaluable guidance, support and persistent encouragement. I am also indebted to my co-supervisor, Dr. Ayan Roy, Research associate, University of Columbia for his excellent guidance and critical suggestions all through my research work. I am truly very fortunate to have the opportunity to work with him. I found his guidance to be extremely valuable. I am thankful to the Head of the School, Dr. Neeta Raj for her guidance and support. I also show my thanks to Dr. Himanshu Singh (HOD), Dr. Vikas Kaushik and all the faculty members and staff of the Department of Bioinformatics, Lovely Professional University for their unequivocal support throughout. I am also thankful to Dr S. Banga, Dr. Atul Kumar for their guidance.

I am grateful to my friends & labmates especially Dr. Sherry Bhalla, Dr. Bhoomika Sharma, Dr. Piyush Aggarwal, Dr. Yasir Afaq, Bhupinder Singh and Gurpreet Kaur for their constructive criticism and suggestions.

I specially want to thank my husband C.A Rahul Kakkar for his support, advice and patience. Thank you for always being there in this journey. I could not accomplish this feat without you by my side. Thank you to my little daughter Radhya Kakkar and little Divyam, Dhruti and Jaishavi for always cherishing me during this time.

I am also very much grateful to my parents Bhawna Kakkar, Rajinder Kumar, Sarita Malhotra and Rajan Malhotra for their moral encouragement and constant motivation. I am also indebted to my grandparents Late Sh. Kasturi Lal, Late Sh. Om Prakash Takyar and Smt. Krishna Rani Malhotra for their love and blessings. I would like to thank my brothers Rahul Malhotra, Ankit, Sanchit, C.A Nitin Arora and my sisters C.A. Surbhi Kakkar and Chetna Malhotra for encouraging me during this period. I would like to show my gratitude to my uncle Anil Malhotra and aunty Anu Malhotra for their moral support. Lastly I would like to thank my entire family for the care that they have shown me during the period of this work.

Himani Malhotra **Date:**

Contents

List of Table

List of Figures

Index

Abbreviations

AHF	**Argentine hemorrhagic fever**
BHF	**Bolivian hemorrhagic fever**
CHPV	**Chapare virus**
CAI	**codon adaptation index**
CoA	**correspondence analysis**
CPS	**codon pair score**
ENc	**Effective number of codons**
GTOV	**Guanarito virus**
GPC	**Glycoprotein precursor**
JUNV	**Junin virus**
L	**RNA polymerase**
LCMV	**Lymphocytic choriomeningitis virus**
MACV	**Machupo virus**
NW	**New World**
NP	**Nucleoprotein**
OW	**Old World**
RCDI	**Relative codon deoptimization index**
RSCPU	**Relative synonymous codon pair usage**
SiD	**Similarity index**
tAI	**tRNA adaptation index**
VHF	**Venezuelan hemorrhagic fever**
Z	**Z-matrix protein**

I would like to dedicate this thesis to my entire family and especially to my husband C.A. Rahul Kakkar and my daughter Radhya kakkar

CHAPTER-1

INTRODUCTION

With the reformation of ecology and alteration in the climatic conditions, expansion in the number of zoonotic pathogens has been observed which leads to increase in diseases among humans (Palareti *et al*., 2016; Jones *et al*., 2008). Viruses display adverse affects among all type of life forms including bacteria, archaea, and eukaryotes (human and agricultural sector, zoonotic threats); thus study of viruses to eradicate them globally is of great concern (Karesh *et al*., 2020). The word "virus" has been derived from the Latin word-for poison (Pellett *et al.,* 2014). In the past, evolution of viruses occurred by several mechanisms; one such mechanism is "RNA world" which insists the self replication of RNAs without the help of other molecules to develop into a cellular organism. Some viruses were also originated from DNA elements like transposons (Koonin *et al.,* 2015); regressive evolution of organisms is another mechanism which led to evolution of viruses. A mutation in the genomic sequence, basically base substitution mutation is also one of the major mechanisms studied to understand the evolution of virus. Rate of mutation is higher in RNA viruses as compared to DNA viruses because of proofreading ability of DNA polymerase (Flint *et al*., 2020).

Further, transmission of viruses may occur as enzootic i.e. transmission occurred to humans from animals and little human to human transmission appeared (for example; Arenavirus); others are transmitted from human to human on a large scale once spill over developed from animal reservoir causing life threatening diseases among humans (Palareti *et al.*, 2016 ; Karesh *et al.*, 2020). Rodents and bats are examined as major natural reservoirs among many human pathogenic viruses mainly including the RNA viruses which easily adapt themselves to the new hosts (Mackenzie & Jeggo, 2020; Longdon *et al.,* 2014). Therefore, consideration of viral evolutionary pattern inside the host or natural reservoir is one of the important aspects to be considered for study of future prevalence of pandemic due to increase in number of viral infections among humans (SS *et al.,* 2012). The analysis proved that proteins showing high level of expression were found to employ or enlist low cost (in terms of protein energetic cost) amino acids, invariably in all concerned viruses. This shows the operation of theory of cost-minimization strategy in which a virus stringently minimizes the usage of energetically expensive amino acids for highly expressed protein sets.

Arenavirus belongs to *Arenaviridae* family and was approved as genus in 1971 by International Committee of Nomenclature of viruses. These groups of viruses cause diseases among humans transmitted by rodents (Shao *et al.,* 2015). Further, division of Arenaviridae family *as Mammarenavirus, Reptarenavirus* and *Hartmanivirus* genera was observed. The name of genera was being defined on the host criteria of the virus just as *Mammarenavirus,* the viral pathogens mainly infecting mammals (Hallam *et al.,* 2018; Radoshitzky *et al.*, 2015). In recent years; various species of Arenavirus infecting mammals have been described. Transmission of viruses exists vertically affecting the rate of fertility and survival of hosts with high mortality rate of 70% among rodents and 50-60% among other mammals with fever, anaemia, lethargy, skin rashes, conjunctivitis, gastrointestinal, neurological signs as other symptoms (Fedeli *et al.,* 2018).

In 1963, on the basis of availability of serological, genetic and phylogenetic data and also on the basis of antigenic properties genus *Mammarenavirus* was further subdivided into two groups Lymphocytic choriomeningitis virus (LCMV) complex or Old World (OW) Arenavirus distributed mainly in West Africa and Tacaribe complex or New World (NW) Arenavirus distributed in South America (Radoshitzky *et al*., 2015). Further, Tacaribe complex has been branched into A, B, C and D clades (Buchmeier *et al.*, 1974 ; Delgado *et al*., 2008). Clade B mainly consists of human pathogenic viruses comprising: Junin virus (JUNV), Machupo virus (MACV), Guanarito virus (GTOV), Sabia virus (SABV) and Chapare virus (CHPV). JUNV and MACV isolated in 1958 from people suffering from Argentine hemorrhagic fever (AHF) and Bolivian hemorrhagic fever (BHF), with its high mortality rate ranging from 3-15% for AHF and 5-30% for BHF (De Manzione *et al*., 1998 ; Yamasaki, 2014a ; Enria & Pinheiro, 2000).

With the increase in cases of Arenavirus and limitation in availability of treatment against this group of viruses, there is a need to conduct an in depth analysis of viruses to understand the mechanism of infection and adaptation inside the host. In, our present study we proposed Clade B of NW namely, JUNV, MACV, GTOV, SABV and CHPV as case study because of their high mortality rate emergence among humans (Sarute and Ross 2020; Rosa *et al.* 2012).Therefore, there is a need to understand the genomic pattern, evolution and also adaptation of viruses inside the

host to propose effective therapeutics and inhibitory measures against this group of viruses (Brisse & Ly, 2019a).

With the advancement in high throughput sequencing methods; analysis and sequencing of Bacteria, Archea and Eukarya genomes have been performed generating large amount of biological data. Thus this brought the urge to use computers as major tools for handling and storage of such a large amount of data, which eventually led to the emergence of a new discipline, Bioinformatics. Bioinformatics refers to the in-depth investigation of various domains, including the study of genome composition, genome expression, proteome analysis, genome and proteome engineering of numerous species (Akalin, 2006). Emergence of Bioinformatics field has provided an insight of novel perceptive into the adaptive strategies engaged by the human pathogenic viruses causing pandemic.

With the availability of sequencing data, comparative genomics has also emerged as one of the advance field in the past decade leading to functional study of various genomes and also enhances the understanding of codon usage patterns of these genomes. Due to the preference for codons among various organisms, variations in usage of codon patterns have been examined. This phenomenon of variation in codon usage patterns was proposed by Grantham and colleagues in the theory of 'genome hypothesis', which stated that the biasness of codons is usually species specific (Grantham *et al.,* 1981; Organisms, 1986; Zhou & Li, 2009).

Viruses require adapting themselves with host cellular microenvironment for survival and progression of infection inside the host (Shackelton *et al.,* 2006). This necessitates understanding the consequence of translation of viral gene mechanism in comparison to host. Information from study of biology of viral codon usage provides an insight about host adaptation, acclimatization and subsequent initiation of infection. In our present study, we destine to understand the codon usage patterns; policies for adaptation of arenaviral species, namely, JUNV, MACV, GTOV, SABV and CHPV with their specific hosts. Various parameters have been reported as efficient tools to understand and perform analysis of codon usage patterns within a gene (Sharpl & Li, 1987; Chakraborty *et al,*2017). These parameters have been found to correspond necessarily with expression levels of mRNA (Ikemura, 1985) and have

been used to anticipate sets of genes showing high level of expression in various organisms (Sharpl & Li, 1987; Reis *et al.*, 2003; Martı *et al.*, 2004; Wu *et al.*, 2005).

Molecular evolution of Arenaviruses is another important aspect that demands to be investigated to resolve the mysteries of adaptation and co-evolution in human. Phylogenetic analysis helps to predict the variation in codons of protein sequences on the basis of dN/dS. In the present study, we explained the evolutionary events of NW Arenaviruses (JUNV, MACV, GTOV, SABV, and CHPV). dN/dS parameters were estimated to predict the natural selection parameter for all the proteins and has been studied to analyse the adaptive molecular evolution and to measure the rate of sequence evolution. Value of dN/dS varies from 0 to 1. As; essentials can incorporate only minor changes in the sequence or codon patterns, most of the changes in sequence which include replacement of amino acid were eliminated, showing negative selection i.e dN/dS <=1. However, positive selection shows easy adaptation of viral as changes in amino acid sequence are considered advantageous (Roy *et al*., 2018). Using, this strategy we predicted glycoprotein precursor (GPC) as major target selected for docking to predict potential lead molecules for drug designing against Arenaviruses.

Proper identification of drug targets in any pathogenic organism is the most vital step in drug discovery process. In our study, we have selected GP1 (subunit of Glycoprotein) as drug target for further drug designing against Arenavirus (Damte *et al.*, 2013; Amineni *et al.*, 2010). Computational drug designing approaches are used to predict and evaluate drugs for various endemic (other diseases too) diseases within short duration of time (Walker, 2017). There has been an increase in study related to drug development pertaining to viral pathogenesis. However, high mutability rates and variable genome dynamics of viruses have been the major obstacles in effective drug design against the detrimental pathogens. With the increase in prevailing threat of viral diseases, there is a rising demand to design drugs for this group of viruses. Many small molecules have been developed to reduce the viral infection but no FDA approved drug has yet been designed to prevent infection caused by *Mammarenavirus*, an alarm for global health (Rathbun *et al.*, 2015).

Between 1940 and 2004, 335 infectious diseases like multi-drug-resistant tuberculosis and chloroquine-resistant malaria were reported to have spread

throughout the world's population. Also, cases of newly emerged strains of pathogens; pathogens that have only recently made their way into human populations (such as HIV-1 and the coronavirus that causes severe acute respiratory syndrome, or SARS), and pathogens that have likely historically affected humans has recently increased (like Lyme disease). With the current global COVID-19 outbreak highlights the need to better understand human pathogenic viruses at the genome level and to investigate how viruses interact with their hosts in order to stop viral outbreaks in the future (Gorbalenya *et al*., 2020). As, Arenaviruses continues to be a hazard to people, hence it is important to research these viruses and raise awareness of potential epidemics. However, there are still a lot of questions about Arenaviruses that need to be resolved. In the present study we intend to thoroughly characterize the genome and proteome of Arenaviruses for appropriate identification of potential targets for medicinal drugs, which may create a platform for effective drug advancements.

CHAPTER-2

REVIEW OF LITERATURE

Viruses depend on the host machinery for protein biogenesis, adaptation and subsequent infection. The goal of the current work is to understand how amino acid patterns, codon usage, and other factors affect the observed variances in viruses. Further, study of Arenaviruses was done in following sections:

2.1 Viruses and their classification

2.2 Arenavirus and related disorders

2.3 Structure and Genome Organization of Arenavirus

2.4 Life Cycle of Arenavirus

2.5 Conventional methods to study Arenavirus

2.6 Research methods to control the Arenaviruses

2.7 Animal models to study Arenavirus clinically

2.8 Bioinformatics as tool to explore the complexities of viruses

2.9 Viral genomics

2.10 Molecular Phylogenetic

2.11 Pharmacology

2.1 Viruses and their classification

Viruses are dominating entity in the ecosystem in regard to physical appearance and genetic diversity (Koonin *et al*., 2020) and they also display highly destructive association with all type of life forms. Availability and usage of high-throughput sequencing techniques has increased the sequencing of novel viruses extracted from humans, plants, animals, environment and has earned vast new assemblage of viruses (Aiewsakun & Simmonds, 2018). Genomes of viruses, composing either DNA or RNA, are based on the cellular systems of host cells for their replication and synthesis of viral components in genome. This makes viruses reliant on the host system, infected by them (Pellett *et al*., 2014). There are total 26 virus families which are highly pathogenic to humans and need to be studied (Siegel, 2018). The present worldwide spill over of COVID-19 brings the emergence to

understand the human pathogenic viruses at genomic level and also to study the interactions of virus with their host, so that the future viral outbreaks can be terminated (Gorbalenya *et al.*, 2020). As Arenavirus also continue to pose a threat among humans, it is essential to study these viruses and to further enhance the alertness for future epidemic. Nevertheless, many questions are still need to be answered for Arenavirus. Recently, advancement in technology for virology research has provided a wealth of information of genome, proteome and biological function of virus.

2.1.1 Family Arenaviridae

Arenaviridae family was established in 1976, viruses of this family causes chronic infections among rodents and has been recently grouped in *Bunyavirales* order (Papageorgiou *et al*., 2020). Viruses of *Arenaviridae* family represents emerging human pathogens and have sandy appearance because of the existence of host cell ribosome's in the viral particles; giving the name Arenavirus as Latin meaning of 'Arena' is sand (Gonzalez *et al*., 2007). Infections caused by Arenavirus prevail more commonly in areas of South America and South Africa and have been described to be federated with severe disturbance among humans (Shao *et al*., 2015). *Arenaviridae* mainly comprises of genera *Mammarenavirus, Reptarenavirus, Hartmanivirus* and recently expanded by addition of one more genera as *Antennavirus* infecting fishes (Maes *et al*., 2018; Radoshitzky *et al*., 2019). The name being defined on the host criteria of the virus just as *Mammarenavirus,* the viral pathogens mainly infecting mammals; *Hartmanivirus* and *Reptarenavirus* infect reptiles as hosts (Hallam *et al*., 2018). *Reptarenavirus* genus was added in 2015 with the isolation of AUNLB3 and UHV-1 (University of Helsinki virus) viruses in Boid snakes in Netherlands suffering with inclusion body disease (Bodewes *et al.*, 2013; Hallam *et al.*, 2018). Chronicle infection pervades among hosts due to Arenavirus and can scrap these viruses for lifetime.

2.1.1.1 Mammarenavirus

On the basis of similarity in geographical distribution, antigenic properties and also on phylogenetic data genus *Mammarenavirus* have been subdivided into OW Arenavirus and NW Arenavirus; further known as Lassa-LCMV serocomplex and Tacaribe serocomplex. As, OW Arenavirus comprises viruses native to rodents of

Africa and NW Arenavirus incorporate viruses native to rodents of America (Carballal *et al.*, 1987; Huang *et al.*, 2015) and were found to originated <23,000 and 41,000 years ago in these regions (Brisse & Ly, 2019b). OW and NW Arenavirus both comprises of pathogens causing fatal hemorrhagic fever among humans, however pathogens of both groups utilize different receptor for causing infection and different internalization process (McLay *et al*., 2014). Analysis performed in 2018, reported 41 viral species in acceptance with genus *Mammarenavirus* by ICTV (International Committee on Taxonomy of Viruses) (Radoshitzky *et al*., 2015) as shown in Table 2.1. In the year 1933, Armstrong and Lillie; discovered the human pathogenic LCMV and later in 1956, Tacaribe virus was isolated in Trinidad and Tabago from *Artibeus Jamaicensis trinitatis*; but this virus was found not to be pathogenic to humans (Yamasaki, 2014b).

Lassa virus and LCMV are major human pathogenic viruses belonging to OW Arenavirus group (Carballal *et al*., 1987). Lassa virus causes severe infection showing mortality of 30-60% and is a topic of great interest among researchers; however, only 20 % of people infected with this group of viruses develop Lassa fever, causing high level of fatality among infected people (Fever, 2021). On the other hand, LCMV was perceived as first member of group since 1934 (Yamasaki, 2014a). LCMV is transmitted by rodents (*Mus musculus*) and leads to severe defects in foetuses (Barton *et al*., 2002). As, human to human transmission does not occur, the increase in spread of infection depends on movement of rodents which determine the genetic diversity and geographical distribution. For example, Lassa virus which was originated in Western Africa areas mainly in Nigeria, confined there for 1000 years; but in recent years spread has been noticed in the neighbouring countries as well (Fever, 2021). In 2008, Lujo virus being isolated from patients suffering from severe illness in Zambia (Simulundu *et al*., 2016) also emerged as one of the new human pathogenic viruses belonging to LCMV complex, showing resemblance with Lassa virus (Raaben *et al*., 2018). Severe cases of infection have been observed mainly during pregnancy causing loss of foetal or maternal death.

Other human pathogen of OW Arenavirus group is Lassa virus that is endemic to large parts of Western Africa and has a high mortality rate of 30-60% causing hemorrhagic fever among humans. Recently, OW Arenavirus has been extracted from

mice, shrews and black rats in Asia leading to increase in the host and geographic areas of *Mammarenavirus* (Sarute & Ross, 2020; Zhang *et al.*, 2019*;* Lu *et al.*, 2014; Ambrosio *et al.,* 2011). Also, in the past few years there has been a increase in the number of imported cases in Europe and USA of both NW and OW Arenaviruses, initiating the demand to study these viruses (Paweska *et al.*, 2009).

2.1.1.2 New World Arenavirus

Further NW Arenavirus genus found to display high genetic diversity and on the basis of host-virus interaction, geographical location, antigenic properties, similarity and differences in amino acid sequences of proteins; the genus has been subdivided into four clades namely; A, B, C and D encompassing 18 species. NW Arenavirus is being accepted with the usage of "Arenavirus" term in correspondence for all the members of *Arenaviridae* family and genus *Mammarenavirus* (Peters, 2002; Radoshitzky *et al.,* 2015). Clade A abides five viruses; Clade C which is the smallest has only 2 viruses and both the Clades are non-pathogenic to humans as so far no case infecting humans from this group of viruses have been reported. Furthermore, Clade B comprises human pathogenic viruses JUNV isolated in 1958 (Ambrosio *et al.*, 1986), MACV in 1963 (Yamasaki, 2014b), GTOV in 1989 (Salas *et al*., 1991), SABV in 1990 (Gonzalez *et al.,*1996) and CHPV were in 2003 (Delgado *et al*., 2008) causing hemorrhagic fever in the regions of Argentina, Bolivia, Venezuela, and Brazilia among humans (Delgado *et al*., 2008). They have been listed as category A pathogens and on the biosafety level scale they have been assigned as level four (BSL-4) agents by Centre for Disease Control and Prevention (CDC) reports (Charrel & Lamballerie, 2008). Other viruses of this group encompasses Tacaribe virus, Amapari virus which have no association with humans and were isolated in 1956, 1964 in Trinidab and Brazil, respectively. Tacaribe virus was isolated from bats as LCM virus (Moraz & Kunz, 2011; Murphy *et al.,* 1970). Clade D is also known as Clade A/B include viruses Bear Canyon, Tamiami and Whitewater Arroyo viruses as shown in Table 2.1. Arenavirus with its high mortality rate ranging from 3-15% for AHF, 5-30% for BHF is of great demand among researchers (Casals, 1975 ; Ma *et al*., 2021). As knowledge related to the prevention from infections and illness caused by Arenavirus is limited, there is a necessity to study the viruses belonging to Arenavirus in detail. In, our present study we proposed Clade B of NW Arenaviruses namely,

JUNV, MACV, GTOV, SABV and CHPV as case study because of their high mortality rate emergence among humans (Coller, 2007).

2.1.1.2.1 Junin Virus

JUNV, being a threat to about 5 million people has been outcome as major pathogenic agents of NW Arenavirus. Approximately, about 1000 cases of JUNV are reported annually, making it important to study this virus. An increase in the areas affected by JUNV was observed mainly in year 1970. About 80% of the cases occur in males and is most prevalent in the people in the age group of 20-49 years (Zhang *et al*., 2019; Lu *et al*., 2014). Although the disease caused by this virus is seasonal, occurring mainly in late summer and in early winter. JUNV was isolated and characterized in 1958 with the recognition of first case in 1943 in the Pampas region of Argentina, causing disease AHF (Enria *et al.,* 2008) and later by 2000 it grew and cover almost four provinces of Argentina. Due to high mortality rate and recognition as bioterrorism agent by National Institute of Allergy and Infectious Disease (NIAID), they have been categorized as category A pathogens (Grant *et al*., 2012). JUNV being transmitted to humans by drylands vesper mouse (*Calomys musculinus*) whereas one of the other reservoir of virus is *Calomys Laucha* (Manuscript, 2008); viral infection may also occur at low frequencies in rodents *Necromys benefactus*, *Akodon azarae* and *Mus musculus*.

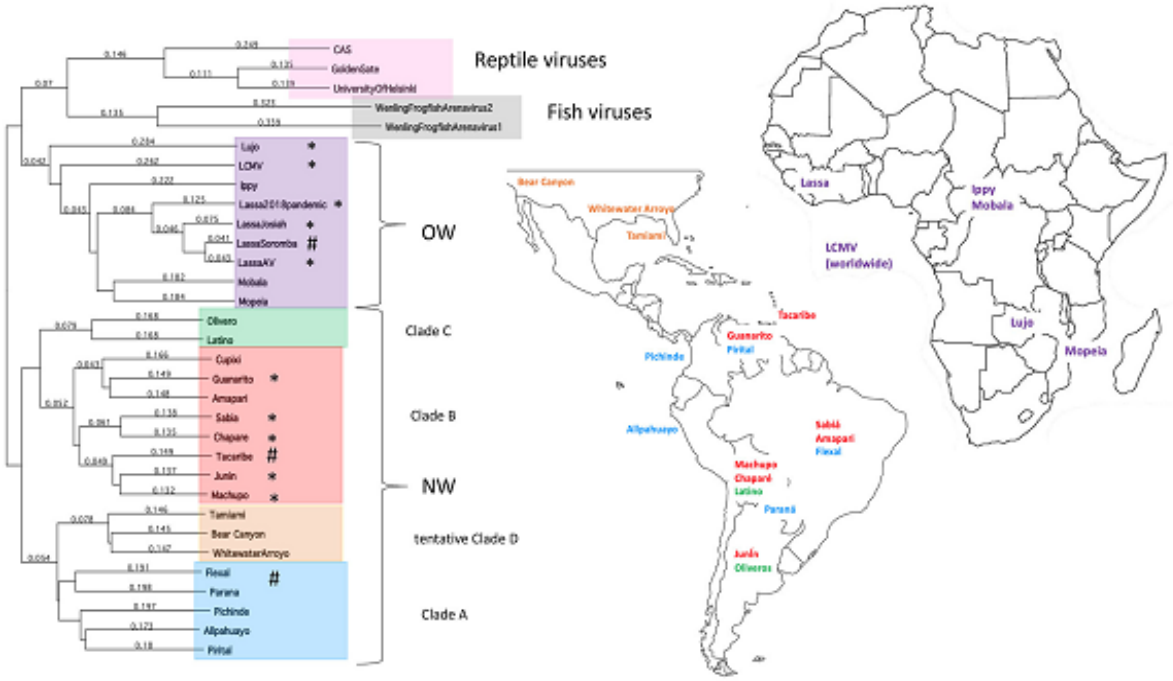


Figure 2.1: Classification of Arenavirus (Brisse & Ly, 2019b)

Transmission of infection among humans prevails either by direct contact rodents while working in fields or inhalation of dust contaminated by excreta or secreta of rodents. Symptoms such as frailty, anorectic, pain and fever persuade by incubation of 7-14 days followed by further neurological, constitutional, cardiovascular and gastrointestinal signs (Ambrosio *et al.*, 2011). Lungs being the initial site of viral infection by replication of virus; diffusion to other parenchyma organs such as vascular endothelium, kidney, myocardium and the central nervous system occurred simultaneously (Kunz, 2009). JUNV, MACV, GTOV, SABV and CHPV share a common receptor hTfR1 (human transferring receptor1) (Rojek *et al.*, 2006). Apart from pervasion of studies for identification of therapeutic facilities for prevention and cure of JUNV, no drug till date being administered.

Table 2.1: Showing classification of Arenavirus (*Mammarenavirus*) as: OW and NW Arenavirus

	Virus species name	Disease in humans	Geographic distribution	Natural host	References

OW Arenavirus	*Ippy virus*	Not reported	Central African Republic	*Arvicanthus* spp.	(Farthing, 1985)
	Lassa virus	Lassa fever	Western Africa	*Mastomys natalensis*	(Günther & Lenz, 2004)
	Lujo virus	Lujjo fever	Zambia	Unknown	(Siegel, 2018)
	Luna virus	Not reported	Southern Africa, Zambia	Unknown	(Emonet *et al.*, 2006)
	lymphocytic choriomeningitis virus	Lymphocytic choriomeningitis	Worldwide	*Mus musculus*	(McLay *et al.*, 2014)
	Mobala virus	Not reported	Central African Republic	*Praomys jacksoni*	(John & Roberts, 1985)
	Mopeia virus	Not reported	Mozambique, Zimbabwe	*Mastomys natalensis*	(Emonet *et al.*, 2006)
	Morogoro virus	Not reported	Tanzania	*Mastomys natalensis*	(Emonet *et al.*, 2006)
NW Arenavirus Clade A	*Allpahuayo virus*	Not reported	Peru	*Oecomys Bicolor* and *Oecomys Paricola*	(Sarute & Ross, 2020)
	Flexal virus	Not reported	Brazil	*Oryzomys* spp.	(Sarute & Ross, 2020)
	Paraná virus	Not reported	Paraguay	*Oryzomys buccinatus*	(Webb *et al.*, 1970)
	Pichinde virus	Not reported	Colombia	*Oryzomys albigularis*	(Polyak *et al.*, 1991)
	Pirital virus	Not reported	Venezuela	*Sigmodon alstoni*	(Sarute & Ross, 2020)
NWArenavirus Clade B	*Amapari virus*	Not reported	Brazil	*Oryzomys goeldi*, *Neacomys guinea*	(Sarute & Ross, 2020)
	Chapare virus	Chapare hemorrhagic fever	Bolivia	Unknown	(Escalera-Antezana *et al.*, 2020)
	Cupixi virus	Brazilian hemorrhagic fever	Brazil	*Hylaeamys megacephalus*	(Sarute & Ross, 2020)
	Guanarito virus	Venezuelan hemorrhagic fever	Venezuela	*Zygodontomys brevicauda*, *Oryzomys*spp.	(Gowen & Bray, 2011)

	Junín virus	Argentine hemorrhagic fever	Argentina	*Calomys musculinus, Calomylaucha , Akodon azarae*	(Grant *et al.*, 2012)
	Machupo virus	Bolivian hemorrhagic fever	Bolivia	*Calomys callosus*	(Bell *et al.*, 2015)
	Sabiá virus	Brazilian hemorrhagic fever	Brazil	Unknown	(Delgado *et al.*, 2008).
	Tacaribe virus	Not reported	Trinidad	Unknown, possibly *Artibeus* spp. Bats	(Sarute & Ross, 2020)
NW Arenavirus Clade C	*Latino virus*	Not reported	Bolivia	*Calomys callosus*	(Sarute & Ross, 2020)
	Oliveros virus	Not reported	Argentina	*Bolomys obscures*	(Sarute & Ross, 2020)
NW Arenavirus Clade D	*Bear Canyon virus*	Not reported	California, USA	*Peromyscus californicus*	(Sarute & Ross, 2020)
	Tamiami virus	Not reported	Florida, USA	*Sigmodon hispidus*	(Winn & Murphy, 1975)
	Whitewater Arroyo virus	Not reported	Southwestern USA	*Neotoma albigula*	(Sarute & Ross, 2020)

2.1.1.2.2 Machupo Virus

MACV, one of the other human pathogenic virus of Clade B of NW Arenavirus encompasses two single stranded RNA molecules causing BHF and was isolated in 1963 with the description of first case in 1959 (Bell *et al.*, 2015). MACV, an enveloped RNA virus showing high similarity with JUNV prevails mainly in the month of April-September. MACV was identified in two rural areas of Bolivia in 1959 and later in 1962-1963 increase in cases have been emerged in neighbouring areas as well, leading to the largest epidemic that extended till midst of 1964 (Bell *et al.*, 2015). In 1971, outbreak occurred in Cochamba, Bolivia, representing an extension to these areas.

Isolation of virus was performed from rodents belonging to *Muridae* and *Cricetidae* families. Rodent vector *Calomys callosus* one of the major natural hosts of

MACV prevails the infection among humans by cause of inhalation of aerosolized excretion and secretions, contamination of food and water by direct or indirect contact of rodents. Recently, cases of nosocomal infection of BHF among patients have also been reported (Bell *et al.*, 2015; Pyle & Whelan, 2019). Serious outbreaks of MACV occurred in the year 1962 to 1964 in Province of ElBeni which highly populated province of northeast Bolivia infecting about thousands of people; out of which 180 died. Also, this outbreak leads to increase in the invasion of MACV to rodents present in small towns.

Further, Bell and their colleagues examined the pathogenic effect of MACV in *Cynomolgus macaque* host transmitting virus to humans. In this study, they incubated *macaque* host for 6-10 days with MACV and observed signs of depression, anorexia and a petechial skin rash. Neurological disorder was also observed. Hence in this study researchers finally predicted the usage of *Cynomolgus macaque* as model organism for study of MACV in humans (Bell *et al.*, 2015). Mostly, increase in cases of BHF was observed during late rainy season (April) and early rainy season (July). Erythema, petechiae, facial edema and shock are some of the symptoms that prevail more commonly in patients infected with JUNV or MACV (Koma *et al.*, 2016; Meyer *et al., 2*014). High fatality rate of 5-30% and clinical similarity with AHF pervades MACV as one of the important premiers that permit the admittance of virus on the NIAID list of Priority Pathogens (Meyer *et al.*, 2014).

2.1.1.2.3 GuanaritoVirus

In addition, GTOV virus, a human pathogenic virus of Clade B of Tacaribe complex group drown one-third of the infected population (Gowen & Bray, 2011; Mori, 2009; De Manzione *et al.*, 1998). In,1989 cases of virus causing VHF prevails in the plains of central Venezuela among agricultural workers; finally characterization of virus as GTOV was performed in 1992 in wild rodents from State of Portuguesa and further 29 isolates of virus were observed in humans and mouse.

With two natural hosts identified for GTOV as *Zygodontomys brevicauda* (cane rat) and *Sigmodon alstoni* (cotton rat); cane rats prevails at highest level identifying as the major host of GTOV and cotton rat was characterized as an intermediate host having infections from spillover of virus from cane rats (Fulhorst *et al.*, 2002; Peters *et al.*, 2017). Some of the symptoms reported in patients suffering

from GTOV are Pharyngitis, vomiting and diarrhoea. GTOV has drown one-third of the infected population having a fatality rate of 33% that compose categorization of virus in category A and one of the major potential bioterrorism threat among people (Gowen & Bray, 2011).

2.1.1.2.4 Sabia and Chapare viruses

SABV was first recognized in Sao Paulo, Brazil in 1990 among persons working in fields and having haemorrhagic fever symptoms. Later in 1994, three more SABV cases in humans were reported, including two cases in laboratory personnel handling SABV samples in the United States, raising concerns about the virus's potential for transmission by aerosol. Recently, in January, 2020 a new fatal case of Brazilian haemorrhagic fever (BHF) has been reported in 52-year old man and detection of virus was performed by Next-generation sequencing method (de Mello Malta *et al*., 2020). CHPV first occurred in 2003 in near Cochabamba, Bolivia, CHPV Province, causing BHF. Initial suspected symptoms of SABV are yellow fever and also this correlates well with the infection caused due to CHPV, as both had identified extensive liver necrosis (Delgado *et al*., 2008). The rodent host species for both the viruses are still unknown. On the phylogenetic analysis of complete L or S segment nucleotide and amino acid sequences; monophyletic phylogeny prevades between CHPV and SABV. Also, in 2019 researchers from the US CDC reported human to human transmission of CHPV in Bolivia (Escalera-Antezana *et al*., 2020). To control the increase in occurrence of infection by these viruses in future, we need to study these viruses at genomic level.

2.2 Arenavirus and related disorders

Transmission of Arenavirus to humans occurs mostly through chronic infections carried by rodents and bats (for Tacaribe and LCM viruses), further affecting the survival of rodents with high mortality rate of 70%. As, transmission of viruses depends on the geographical distribution of their rodents, as rodents are distributed worldwide this brings about emergence of viruses in humans also. Symptoms such as fever, anemia, lethargy, skin rashes, myalgia, anorexia, diarrhea, nausea, sore throat, malaise and lymphadenopathy were found to display as initial symptoms and conjunctivitis, gastrointestinal, neurological signs as other symptoms. Bleeding of gums can also occur in some of these cases (Gonzalez *et al.*, 2007).

Table 2.2: Showing symptoms of NW and OW Arenaviruses

Symptoms	OW Arenavirus	NW Arenavirus				
		JUNV	MACV	GTOV	SABV	CHPV
Flu-like early symptoms	Yes	Yes	Yes	-	-	-
Haemorrhage	Mild	Yes	Yes	Yes	Yes	Yes
Leukopenia	Yes	Yes	Yes	Yes	Yes	-
Thrombocytopenia	Yes	Yes	Yes	Yes	-	-
Neurological symptoms	-	Yes	-	-	-	-
Vomiting	Yes	Yes	Yes	Yes	Yes	Yes
Petechiae	Yes	Yes	Yes	Yes	Yes	-
Sensorineural hearing loss	Yes	No	No	No	-	-
Liver damage	Yes	-	-	-	-	-
Vascular damage	No	No	No	-	-	-

2.3 Structure and Genome Organization of Arenavirus

Arenavirus particles appear to be spherical to pleomorphous in shape. Arenavirus with a unique morphology have single-stranded RNA in the virion and size varies between 60- 230 nm in diameters. The genome of Arenavirus possessing negative sense single-stranded RNA encompasses two segments pertaining small (S) RNA segment of size 3.4 kb and large (L segment) of size 7.2 kb. Small segment code for 2 proteins namely; glycoprotein precursor (GPC) and NP (NP), while large segment codes for RING finger protein (Z) as well as viral L (RdRp) L. Each segment of Arenavirus has been encoded by using ambisense coding strategy showing separation of two arenaviral by intergenic region. Further, GPC was degraded into N-terminal GP1; having binding capacity with host receptor and the transmembrane GP2, indulge in viral fusion by signal peptidases and subtilisin kexin isozyme-1or site-1 protease (SKI-1/S1P) (Meyer *et al*., 2002).

Virus was attached to host cell receptor by interaction with various cellular factors is the initial step for virus to cause infection among hosts. GP1 binds to the cellular receptor and further fusion was mediated by GP2 (Fedeli *et al*., 2018). Clark and colleagues in 2018 predicted the 3D structure of GP1 protein of JUNV and MACV that binds to htf1 for interaction with host and for causing infection inside the

host cell. So, GP1 protein can be studied as one of the prominent target for drug designing against NW Arenavirus (Clark *et al*., 2018). NP also belongs to S segment and together with L and viral RNA constitutes the viral riboNPs (vRNPs) which are required for replication and transcription of genes of Arenavirus inside the host cell. Further, vRNPs interact with Z creating mature infectious virions (Martínez-Sobrido & de la Torre, 2016). Study of viral infection mechanism among humans examined that Z-protein inhibit type I IFNs (interferons) by binding precisely to RIG-I (Retinoic Acid Inducible Gene-1) and MDA-5 (Malondialdehyde) which are known to be cytosolic sensor proteins involved in stimulation of interferon by recognizing the double stranded RNA generated during replication of viral infection (Brisse *et al*., 2021). Therefore, these sensory proteins are also admitted to be part of common mechanism followed by arenaviral pathogens to cause infection (Ly & Liang, 2015).

2.4 Life Cycle of Arenavirus

Arenaviruses are internalized into the host cytoplasm by process of endocytosis through attachment with their cell-surface receptors (htf1 for clade A, B of NW Arenavirus and alpha-dystroglycan for OW, NW Arenaviruses clade C) composing infection (Rojek *et al.*, 2006; Radoshitzky *et al.*, 2007). The binding of virus to cell surface facilitates its fusion with late endosomes, which is pH-dependent. Further, releasing the viral RNA into cytoplasm where replication and transcription occur by RNPs. NP and L are the first proteins to be synthesized after infection further directing the replication and transcription of RNA (Palareti *et al.*, 2016). NP protects the viral genome from host cell and is involved in the synthesis of r-NPs.

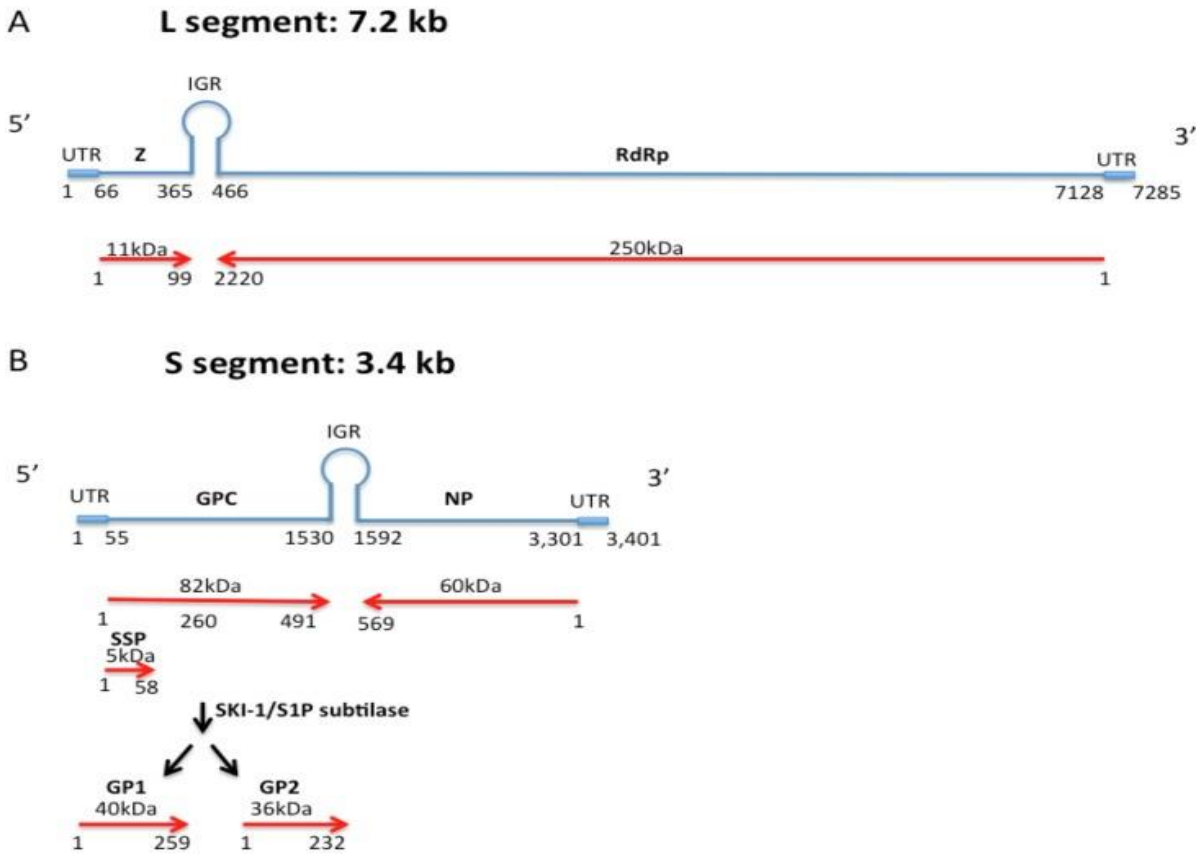


Figure 2.2: Genome organization of Arenavirus (Zapata and Salvato 2013)
UTR: untranslated region, Z: RING finger protein, IGR: intergenic region, RdRp: *RNA* derived riboNP, *GPC:* Glycoprotein precursor, NP: NP, SSP: signal peptidase.

Further, initiation of replication is performed by slippage mechanism of L and uncapped antigenomic and genomic RNAs are generated by reading transcription-termination signals through intergenic region (Lee *et al.*, 2000). Synthesis of GPC is performed in endoplasmic reticulum after completion of one round of replication, further initiating the synthesis of Z. GPC is N-glycosylated, and maturation-trafficking process from ER to cell surface is performed depending on stable signal peptide (SSP) which gets cleaved off during synthesis of GPC (López *et al.*, 2001). Finally, after synthesis of Z initiate the budding and further initiated the formation of more infectious virions (Hass *et al.*, 2004).

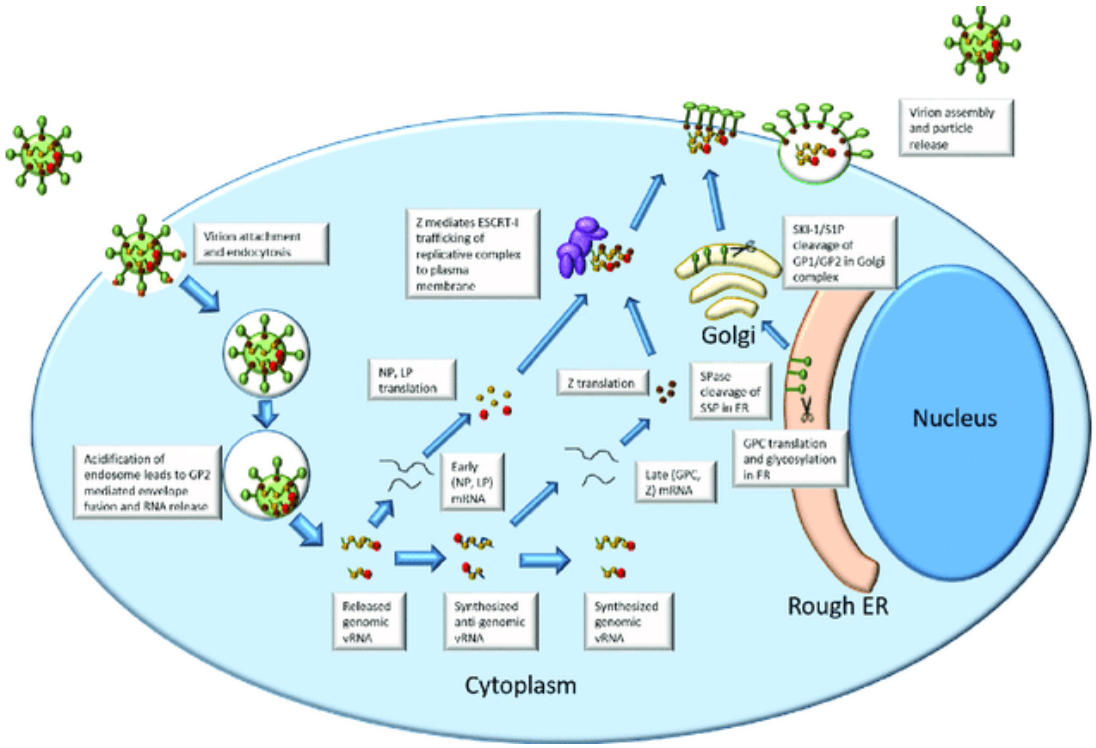


Figure 2.3: Life cycle of Arenavirus (Hallam *et al.*, 2018)

2.5 Conventional methods to study Arenavirus

Proper assessment of human pathogenic NW Arenavirus has been a challenging yet, interesting chore. Several techniques have been successfully implemented for gaining a better insight into the complex puzzles of Arenavirus. Some commonly employed approaches have been discussed below:

2.5.1 Clinical lab studies

The complement-fixation (CF) test and fluorescent antibody techniques (FAT) are used more frequently for an early diagnosis of Arenaviral infection (Casals, 1975). In a study, during the first week of arenaviral infection, the presence of leucopenia and thrombocytopenia was observed. The count of white blood cells and platelets was approximately around 1000-2000 and 5000-100,000 per mm^3, respectively. Furthermore, a mild decrease in aspartate transaminase (AST), creatine phosphokinase (CPK), and lactate dehydrogenase (LDH) was observed. Inflation in serum creatinine and urea was observed in severe cases of viral infection. The level of cerebrospinal fluid (CSF) was found to be normal even during acute illness in patients. However, patients with late neurological syndrome (LNS) showed a small

amount of increase in the cell number and antibodies present against virus (Enria *et al*., 1987).

2.5.2 Reverse Transcriptase PCR-based assay

PCR-based assay such as RT-PCR detects viruses rapidly by diagnosing viral genome in early stages of infection. This is highly reproducible and novel method to quantify the viral RNAs in cell culture and tissues. Screening of the viruses can be done within less than 2 hours and diagnosis can be performed before onset of symptoms. This method is highly faster, specific and sensitive; although it is more complex as compare to immunological assays. With the progress in RT-PCR methods, such as the use of multiplexed oligonucleotide microarray, genetic variations of viruses can also be diagnosed (Li *et al.,* 2015; Lozano *et al*., 1995).

2.5.3 Serological detection

As Arenavirus are detected as BSL-4 agents, it is challenging to diagnose these pathogens in lab without BSL-4 facilities. Therefore, antibody detection methods, including IgG-ELISA and IgM-ELISA (enzyme-linked immunosorbent assay) and immunofluorescence assay (IFA), are generally employed using recombinant-NPs (R-NPs). R-NPs may be highly useful in detecting Ig-G and Ig-M antibodies with early and easy detection in lab conditions. Since the IFA results are highly specific owing to different fluorescence properties shown by different viruses, this method has advantage over other methods; however, it requires experienced personnel. An IFA detects the antibody bound to fixed monolayer of infected cells in serum. IgG- and IgM-ELISA are very sensitive, specific and reliable methods and they can detect recent infections, especially during convalescence. Furthermore, various novel monoclonal antibodies have been generated using R-NPs. Sandwich antigen-capture ELISAs, which employes the use of novel monoclonal antibodies (mAbs), is found to be sensitive and specific for detection of Arenavirus NPs. For example, mAb E4-2 and C11-12 have been used to detect the antigens of all the pathogenic NW Arenavirus tested, whereas mAb C6-9 detects only the JUNV antigen (Sanchez *et al.,* 1989; Ruo *et al.,* 1991; Drosten *et al*., 2003*;* Nakauchi *et al.,* 2009; Fukushi *et al*., 2012).

2.5.4 Next generation sequencing

With the advancement in high throughput sequencing methods, identification of Arenavirus pathogens can be performed with high level of sensitivity and specificity. Multiplex panel has been designed which composed of two pools having 285 and 256 primers for amplification of 46 viral species causing haemorrhagic fever. The process is also helpful in diagnosis of genomic variants of viral species which causes haemorrhagic fever. This enables the detection of viral species even with lowest concentration. Furthermore, sequencing of the amplified genome was performed by Ion Torrent Personal Genome Machine system. This method allows identification of viral species within less than 3.5 hours which is very helpful during pandemic or viral outbreak situation (Brinkmann *et al*., 2017).

2.6 Research methods to control the Arenaviruses

In the last 10 years, there has been an increase in the study of molecular biology of Arenavirus in regions of South American and West African countries.

In 1979, Candid #1vaccine, a live attenuated vaccine which was developed with the help of Argentine Ministry of Health and Social Actions was employed to prevent AHF in regions of Argentine and was developed through dilution and pseudo single burst cloning of the JUNV strain XJ44 in FRhL-2 diploid. Candid#1 vaccine examined to be effective for AHF, considered preventive to be used by immunologically impaired recipients and pregnant women (Koma *et al*., 2016). Candid#1, the vaccine discovered for JUNV because of similarity of JUNV with MACV has been examined for compatibility in MACV. Glycoprotein of Candid #1 live attenuated vaccine of JUNV strain was found to be effective in replicating MACV by using reverse genetics approaches and its was able to attenuate MACV. They have examined a stable model of mouse attenuated with recombinant MACV strain having Candid#1-glycoprotein MACV, and propagation was established to be safe and protective for MACV infection (Golden *et al*., 2017).

Further, Golden and colleagues studied MACV and identified a variant of MACV strain Carvallo termed Car91 that was attenuated in guinea pigs. Car68, one of the strains causing lethal disease in guinea pigs, was compared with Car91 at sequence level and similarity among two strains was found. Car91being non pathogenic, was analyzed interpreting that it can be used as protection agent for

GTOV, one of the other virus belonging to NW Arenavirus group (Golden *et al*., 2017).

Ribavirin (1-D ribofuranosyl.1.2.4. triazole-3-carboxamide) was also reported as antiviral drug for AHF. The results showed that there was no decrease in mortality rate of people suffering from hemorrhagic fever and many side effects were examined among patients as it causes anaemia, febrile syndrome. Although, till date no clinically approved drug is available for this group of viruses and many novel small molecules of drug candidates have been discovered but their approval as drug is still to be justified. T-705, pyrazine derivative (6-flrara – 3 hydraxy – 2 pyrazinecarbox – amide), ST-366, ST-294 and ST-193,T-705, 10-allyl-6-chloro-4-methoxy-9(10H)-acridone) and conjugated phosporo diamidate morpholino oligomers (PPMOs) have been summarized as one of the molecules for drug development against arenaviral activity (Kerber *et al*., 2015). Also, a novel compound (compound 0013) that blocks the interaction of Tsg101, one of the components of host and is mainly involved in transport in endoplasmic sorting complexes with PTAP Late (L) domain of Z of JUNV was reported having the potential to be used as antiviral drug for treatment of AHF (Lu *et al*., 2014).

Sobrido and Torre employed reverse genetics approach to understand the biology of Arenavirus to interpret host-virus interactions and to construct recombinant Arenavirus, such as recombinant tri-Segmented (r3) Arenavirus and recombinant bicistonic Arenavirus, providing insights into functioning of genes of Arenavirus. This study of Arenavirus biology may help in identification of novel drug targets and also in development of attenuated Arenavirus vaccines. Therefore, mechanism of reverse genetics may help researchers in development of more effective drugs against arenaviral pathogens (Martínez-Sobrido & de la Torre, 2016).

Arenavirus infections are generally treated by transfusion of immune plasma consisting of neutralizing antibodies in defined dose. However, this treatment method is accompanied with the challenge of maintaining stocks of immune plasma. Therefore, there is an increasing demand of other effective treatments.

Zeitlin and colleagues demonstrated the utilization of immune therapy for prevention of infection caused due to JUNV. Monoclonal antibody development was further employed, by utilizing three neutralizing murine monoclonal antibodies which

were previously defined and their expression was evaluated in guinea pig which was examined as model organism for study of AHF. These monoclonal antibodies provided high percentage of protection rate against JUNV lethality when administered 2 day after viral infection, and J199 provided 100% protection when treatment was provided after 6 days of infection with JUNV. Therefore, these predictions suggested usage of J199 as one of the effective therapy for treatment of JUNV (Zeitlin *et al*., 2016). Similarly, many monoclonal antibodies were developed for treatment of Arenavirus, but neither of them succeeded. Thus, there is an urgent demand for genomic level study of JUNV (Enria *et al*., 2008).

Meyer and colleagues used computational approaches to predict occurrence of mutations at host-parasite interaction sites and their effect in prevalence of infection caused due to MACV. In their study they have examined occurrence of mutations and substitutions at host-parasite interaction sites. With the difficulty in prediction of host-parasite interactions; in the present study they have used molecular dynamics simulations method to computationally predict the interaction between glycoprotein (GP1) of MACV and human transferrin receptor. Both were pulled away and force versus distance curve was plotted. Also mutations were introduced to differentiate wild type and mutant interactions and also their free energy was calculated. Secondly, out of 2 hydrogen bonds involved in interaction of host and virus, one was examined to be more important as compared to the other. The study also provides insight of multiple mutations by applying steering forces. Third it also predicted the interaction site important for infection and hence can be used as suppressor site (Meyer *et al*., 2014). These studies concluded that computational approaches can be used to study viruses at genomic level and to design effective drugs and vaccine for this group of viruses.

2.7 Animal models to study Arenavirus clinically

To understand the pathogenesis of viruses various animal models were infected with Arenavirus experimentally showing data that can be useful for future studies of viruses (McKee *et al*., 1987). Guinea pigs which include strain13 and Hartley guinea pig have been used as model organisms to study Arenavirus (JUNV, GTOV and Lassa). Both the strains show same level of infection in the spleen, lungs, intestine, lymph nodes and gastrointestinal tract developing initial symptoms as

leucopenia, thrombocytopenia. Further, increased in infection causes hemorrhagic fever resulting to convulsions and death. Further, no decrease in mortality was observed when guinea pigs were treated with antiviral drug ribavirin or tributylribavirin, although replication of viral replication was delayed and death also delayed (Sanchez *et al*., 1989).

Non-human primates such as AG129d mouse, African green monkey and marmosets infected with Arenavirus showed similar symptoms such as hemorrhage as those reported in case of humans infected with AHF (McKee *et al*., 1985). *Callithrix jacchus* can also be used to study Arenavirus as they develop acute hematological and neurological manifestations; also they died after 17-24 days from inoculation. However, treatment with ribavirin increases the survival rates. *Macaca mulatta* also show similar symptoms when infected with different strains of JUNV during the second week after inoculation. Further treatment with ribavirin shows prevention from viral infection but illness in central nervous system was also observed (McKee *et al*., 1988).

2.8 Bioinformatics as tool to explore the complexities of viruses

2.8.1 Origin of Bioinformatics

Bioinformatics is defined as the field which applies computational techniques to understand and store the information associated with biological systems and macromolecules (Akalin, 2006). Rapid progress in genome sequencing technologies and advances in Bioinformatics and Cheminformatics based research domains have provided a massive scope for enhancement of drug discovery technologies. Availability of genomes of both the host and concerned pathogen provides a platform for subtractive genomics-based drug target identification in concerned pathogen (Stumm *et al.,* 2002). Subtractive genomic based approach involves subtraction of host genome from the pathogen while screening the tentative targets in the pathogen.

In the year 1953 James Watson and Francis Crick determined the double helical structure of deoxyribonucleic acid (DNA). This particular event brought changes in the field of biological science and leads to the origin of discovery of modern molecular biology. The discovery of DNA structure consort vision to interpret the genetic codon and synthesis of protein. Meanwhile, in 1970s and 80s new

scientific techniques were developed which include specifically recombinant DNA technology, genetic engineering, fast sequencing methods etc.

Around the same time, the term "Bioinformatics" was coined by Ben Hesper and Paulien Hogeweg (Hesper and Hogeweg, 1970). In 1970, with the efforts of of Margaret O. Dayhoff, Walter M. Fitch, Russell F. Doolittle and others Bioinformatics which is a multi-disciplinary field began. Initially, it was referred as "the study of information processes in biotic systems" (Hesper and Hogeweg, 2021). However, in the late 1980s it was described as combination of biological science, computer science and information technology used for the analysis of molecular data, particularly involving large-scale sequencing of DNA (Luscombe *et al.*, 2001). The arrival of the INTERNET is another important milestone in the development of Bioinformatics as a full-fledged discipline. The discipline Bioinformatics serve as the combination of genomics, biotechnology and information technology, and compasses analysis and interpretation of data, modeling of biological phenomena, and progression of algorithms and statistics (Fenstermacher, 2005).

The sequencing of first viral genome of phiX174 was performed in 1977 by Sanger. In 1995 first bacterial genome *Haemophilus influenza* was sequenced, leading to the introduction of discipline 'Genomics'. The need for Bioinformatics was further accelerated when the Human Genome Project (HGP) was launched in 1990. The focus of the project was to sequence the entire genome of human. Information gleaned from the HGP is not very useful until the huge data is managed and interpreted in a proper way by the computational tools leading to the materialization of Bioinformatics. Further, storage of sequencing data were greatly supported by advancement of computational tools (Searls, 2000) and specific databases (Birney *et al*., 2002).

With the availability of complete genome sequences of different organisms lead to the development of public repositories of gene data like GenBank (Searls 2000; Benson *et al.*, 2018), EMBL ((Kanz et al., 2005), DDBJ (Mashima *et al.*, 2017), Protein Databank (PDB) (Bernstein *et al.*, 1977) and several others. After the formation of the databases, tools became available to execute various analyses. Two programs, which greatly facilitated the similarity search, were FASTA (Pearson and Lipman, 1988) and Basic local alignment search tool (BLAST) (Altschul *et al.*, 1990).

Since then; many programs have also been further developed. Accessibility of free and open source software has taken Bioinformatics and its application to all-together new heights.

2.8.2 Bioinformatics based platforms suitable for research pertaining to viruses

Several 'omics' based research disciplines i.e., genomics (for genome data), proteomics (for protein data), transcriptomic (for gene transcription data); have emerged with the enormous advancement of Bioinformatics. The 'cause and effect' exchange in biological systems are being frequently employed by mathematical models and computational simulations for proper elucidation of biological complexities. Proper blend of knowledge-driven computational simulations and data oriented Bioinformatics holds the key in appropriate realization of the riddles of human virus interactions. Meaningful analysis of high-throughput sequencing data promises to unravel the enigma of the complex associations. Specific biological databases, providing access to annotated genomic data of viral populations, like ViPR (Pickett *et al.*, 2012), RVDB (Bigot *et al.*, 2020) and ViruSurf (Canakoglu *et al.*, 2021) have opened up revenues to extract significant information.

2.9 Viral genomics

Outbreak of many viral diseases occurred in the history of humans destroying the whole community and entire population. Some of the famous outbreaks are smallpox, Spanish flu pandemic of 1918, HIV/AIDS (ongoing), H1N1 virus in 2009, Ebola virus in 2014-15 which causes epidemic in regions of Western Africa and ongoing pandemic of COVID. Investigation of the outbreak comprises eradication of infection from current population and also to prevent the reoccurrence of virus in future generation. In the past many epidemiological methods have been employed to control the outbreaks; but these methods were successful only in control of outbreaks in small regions (limited areas).

So, with the era of discovery of Bioinformatics discipline study of comparative genomics of viruses by employing various tools and software showed that each virus has its own genomic story and will help to monitor the infection caused by each virus more effectively. Genomic analysis of viruses helps us to understand the outbreaks by distinguishing isolates of same viral strain. Example: In 1992, sequencing of HIV from a patient led to understand how the virus was

transmitted which was not possible earlier. Outbreak of Ebola virus in 2014-15, also brings the emergence to study the viruses at sequence level. Recent advancements in high-throughput sequencing, computational methods and processing of data increases the efficiency to analyse the virus at genomic level and help to interpret the mechanism of infection and adaptation of virus inside the host machinery.

Further, study of genetic code and its usage preferences is one of the most important aspects of biological science. Genetic code is degenerative with 61 codons encoding for 20 standard amino acids. 18 amino acids out of 20 are encoded by more than one that is two to six synonymous codons and only methionine, tryptophan are encoded by one codon. However, study of synonymous codons has shown that alteration in synonymous codons affect the protein biogenesis which includes transcription, translation, posttranslational modifications, and co-translational modifications.

In 1981, Grantham proposed the existence of high variation in codon usage patterns within and between organisms which hypothesized to propose the phenomena of variation among codon patterns (Grantham *et al.*, 1981). Many parameters including gene expression level, availability of tRNA, codon-anticodon interaction, gene length, secondary structures of protein, compositional constraints, mRNA folding stability, natural selection for translation, aromaticity, hydrophobicity, hydrophilicity contribute to codon bias by influencing the occurrence of codon patterns in the genome (Romero *et al.*, 2003). Although mutational constraint and translational selection being admired as major aspects impressing the codon usage bias in all prokaryotic and eukaryotic organisms (Chem & Chem, 1993; Romero *et al.*, 2003; Gu *et al*., 2004). While some cases such as *Drosophilla melaongaster* (Moriyama & Powell, 1997), *Caenorhabditis elegans* (Duret, 2000), *E.coli* (Ikemura,1981) inferred natural selection and availability of tRNAs as the major factors impacting the preference of codons in highly expressed genes over the others influencing the speed and expression level for translational precision.

Recently Roy and colleagues predicted codon and amino acid usage patterns in *Bifidobacterium* genus and anticipated that GC content is high in the genome which is around 60.48%, making the prediction that codons of GC richness are preferred over AT richness. Finally, they described from their study that evolutionary selection

is one of the important aspects governing the codon usage bias in *Bifidobacterium* genus (Roy *et al*., 2015).

The use of synonymous codon pairs does not happen by random, according to a study of bacterial and eukaryotic genomes, and such use definitely has the goal of increasing the speed and fidelity of protein synthesis (Cannarozzi *et al.*,2010; Fredrick and Ibba, 2010; Guo *et al.*, 2009; Plotkin and Kudla, 2011; Zhang *et al.*, 2013).

Therefore, in this present work we intend to explore the puzzles of codon and amino acid usage patterns among the selected members of the Arenavirus and simultaneously compare the observed patterns with that of human and rodent host. Such a comparison might provide a vivid portrait about the adaptive strategies employed by the viruses. Comparative genomics and proteomics based research have been significantly effective in inferring in-depth idea about various genomic and proteomic traits.

2.9.1 Codon usage analysis study in viruses

With the availability of sequencing data many studies have been performed on codon usage analysis of RNA viruses predicting the causes and implications of codon usage biasness. Further from this study they interpreted that codon usage bias of synonym codons either depends on mutational or transitional pressure and it also depends on chromosomal location of each gene as genes located in GC rich region, preferred GC codons over AT. This information is helpful in understanding the gene expression pattern, evolution and also for prediction of drug against viral sites which are more important in cause of disease (Jenkins & Holmes, 2003). Mutational pressure is dependent on occurrence of mononucleotides and dinucleotides in genome and their occurrence may vary. The codon at position 2-3 has more impact on codon bias as compare to 1-3 position. Some of the other factors that affect the codon usage bias of viruses are secondary RNA structure, nonspecific RNA folding, hydrophobicity, aromaticity and all these factors desired to be studied in detail while predicting variation among codon and amino acids patterns (Belalov & Lukashev, 2013). Examining various human RNA viruses which are diverse genetically and ecologically, it was reported that very low variation has been observed among RNA viruses (Jenkins & Holmes, 2003). Some cases on study of codon usage of human pathogenic viruses have been reported as Butt and colleagues examined codon usage

biasness in Zika virus and predicted that natural selection is favored over mutational for codon bias in Zika virus. Li and colleagues studied torque teno sus virus, causative agent of porcine circovirus associated disease (PCVAD).

Yao and colleagues recently studied the codon usage pattern of *Flaviridae* viruses, causing infection in humans and transmitted from mosquitos, sandflies. Overall results interpreted low level of codon bias and high level of GC content in the genome of 65 virus strains under study (Yao *et al*., 2019).

Further, the effect of codon usage and amino acid usage in Nipah virus and HIV (human immunodeficiency virus) was also examined. Various statistical parameters such as Relative Synonymous Codon Usage, Similarity Index, Correspondence analysis and evolutionary rate were examined to predict the occurrence of variable patterns among codon and amino acid. These parameters also conclude that natural selection as main cause of variation among patterns (Roy *et al*., 2017).

Kumar and Kumar inspected the codon usage bias of genome of Newcastle virus, pathogenic to both avian and non-avian species and concluded from the results that mutational selection is the major factor causing codon bias in Newcastle virus genome and also aromaticity is one of the other factors affecting the codon bias (Kumar & Kumar, 2017). Codon usage analysis was also performed on Human Pappiloma virus by Kamatani and Shirotouna to understand the main factors influencing codon patterns in virus. The results showed compositional constraint the main factor and AT rich codons occurred more frequently.

Kumar and colleagues examined the codon usage bias of Equine influenza viruses (EIVs) of H3N8 subtype and in this study, they have evaluated codon patterns of genome of Equine influenza virus in 92 strains and results have shown preference for codons ending with A/U. Many statistical analysis were also performed showing mutational and natural selection both factors affected the host virus interaction, it also predicted that these viruses are less adapted to host horse (Kumar *et al*., 2016).

As detail study of codon and amino acid usage bias prevail us with information that can be utilized to design synthetic genes and proteins with high level of expression and more translational accuracy. Present study aims to unravel the codon usage and amino acid patterns and associated factors influencing the observed

variations in JUNV, MACV, GTOV, SABV, and CHPV which belong to NW Arenavirus and are, defined as great threat for people as they can be used as bioterrorism agent.

2.9.2 Host adaptation

Availability of large amount of genomic sequence data of various pathogenic species and *Homo sapiens* which act as host for these pathogens provide us the scope to study the mechanism of host pathogen interaction. With the advancement in high throughput techniques, transcriptomic and proteomics disciplines coupled with genomics has led to the study of expression level of genes and protein products resulting in the discovery of new genes and protein products. Moreover, these techniques may hold the key to identify the hidden features of host-virus interactions Viral genomes, owing to their dependency on the availability of host machinery for protein biogenesis, persuade the requirement of exploration of codon usage patterns of various hosts that they infect (Roy *et al.,* 2017; Shackelton *et al.,* 2006). This helps us in understanding the mechanism of viral gene expression in comparison to host. Knowledge of codon usage biology in viruses provides an insight into host adaptation, acclimatization and subsequent onset of infection.

Recently, *in silico* studies has been performed in Zika, Nipah, HIV, influenza A virus to understand the adaptation mechanism of viral genes on the basis of codon usage bias in *Homo sapiens* (Roy *et al.,* 2017; Butt *et al*., 2016; Schrauwen & Fouchier, 2014). Similarly, Lee and colleagues studied the *Lactobacillus salivarius*, bacterium species present in animal gut and is one of the promising gut probiotic bacterium. This analysis explained that 56 protein coding genes for extracellular proteins and 124 orthologs that are related to production of exopolysaccharides are mutable and have revealed that two factors are responsible for host adaptation of bacteria one is by ability of bacteria to gain niche adhesion and by utilization of efficient nutrients. This study provides us insight of genome of *L. salivarius* and other mutualistic bacteria (Lee *et al*., 2017).

Recently, Khandia and colleagues analyzed the codon usage and adaptation of Nipah virus and also adaptation of viruses to their host. Further, factors influencing the adaptation of virus to host machinery for protein biogenesis were also examined. Host adaptation of Nipah virus was also compared with the host adaptation of

Hepavirus genus among 10 hosts. Overall very low amount of bias was observed in Nipah virus codon usage and also aromaticity, one of the factors of codon bias has no impact on genome of Nipah virus. Statistical parameters such as codons adaptation index, similarity index, tRNA adaptation index were examined to predict the host adaptation of Nipah virus inferring that natural selection is the major parameter for host adaptation as compare to mutational parameter (Khandia *et al*., 2019). In our present study, we intend to understand the codon and amino acid usage patterns and policies of adaptation of arenaviral species, namely, JUNV, MACV, GTOV, SABV and CHPV with their specific hosts.

2.10 Molecular Phylogenetic

Molecular Phylogenetic is defined as merger of molecular biology and statistics to study the evolutionary relationships among species and to predict how they emerge from one another with time. With the advancement in computer science many tools and software have been developed to understand the evolutionary relationship of humans with other species, which are the ancestors and further the results are inferred in the form of graphical tree to understand the mechanism of evolution in detail. Evolutionary study can be performed on hundreds of datasets predicting the variations occurring in functionality and structure of molecules with time. Different methods are used based on nature of data and study to be performed (Godini & Fallahi, 2019).

With the availability of whole genome sequences of various pathogens has given more opportunities to explore and compare the genomes in larger scale. Recently evolutionary study of *Mycobacterium* genus was performed showing secretary proteins evolve faster than non-secretary proteins in the pathogenic members. Similarly, evolutionary studies by maximum likelihood method were performed for H7N9 Avian virus, torque teno sus virus and Zika virus showing occurrence of adaptive evolution due to which viruses optimize their survival in different environments and adapt themselves to new environments of hosts (Butt *et al.,* 2016; Li *et al.,* 2019).

2.10.1 Phylogenetic analysis

Availability of mammoth sequencing data has made it feasible to explore the evolutionary riddles of eukaryotic and prokaryotic genomes. Phylogenetic analysis

and computational study of genomes of prokaryotes and eukaryotes can be performed to understand the evolutionary history of species.Phylogenetic analysis and computational study of genomes of prokaryotes and eukaryotes can be performed to understand the evolutionary history of species. Phylogenetic relations are being defined by predicting homologous relationships which include both orthologous (homologous sequences which are separated due to speciation) and paralogous (homologous sequences which are separated due to gene duplication) sequences. Many different computational methods have been developed for phylogenetic analysis include distance based methods and word based methods which are based on construction of tree showing relationships among samples based on the substitutions occurring in nucleotide sequences. Distance based methods are further divided into Neighbour Joining method and UPGMA (Unweighted pair group method in Arithmetic mean). Word based methods or discrete methods include Maximum Parsimony and Maximum Likelihood methods. Although distance based method is more informative as compared to word based method.

UPGMA Method: UPGMA method is used to produce mainly rooted trees for which length of edge can be viewed as times which is being measured by a molecular clock with a constant rate.

Neighbour Joining Method: In this method pairwise "distances" matrix is computed between sequences that approximate to provide insight of evolutionary distance. Clustering techniques are used to compute evolutionary distances, which include number of nucleotide or amino acid substitutions between sequences.

Maximum Parsimony Method: This method searches for the most parsimonious tree that is based on least number of evolutionary changes observed to explain differences.

Maximum Likelihood Method: This method requires a probabilistic model for the process of nucleotide substitution and creates a probability distribution on a set of trees (D'Addato *et al*., 1986).

Evolution occurs due to mutations in the molecular sequences and this variation can be due to variation in GC content and genome size. To predict evolution graphically in the form of tree we need to make estimation about substitutions occurring in the sequences and for that stoschastic models are employed. Some of the popular evolutionary models used are: Jukes-Cantor method, Kimura 2-Parameter

(K2P), Hasegawa-Kishino-Yano method and Equal-input. These models predict the probability of occurrence of a particular feature of tree.

Bootstrap method

Bootstrapping is done to validate the results of phylogenetic tree on particular datasets. Validation of results is done by performing the analysis multiple times and predicts the accuracy of the tree analyzed by using the dataset. Bootstrapping predict the confidence level of phylogenetic analysis by using statistical models.

Presently, phylogenetic analysis of various viral pathogens has been performed providing relevant information about functional and evolutionary attributes of pathogens. Example: Phylogenetic analysis of Ebola virus by using Maximum Likelihood method was performed after sequencing and during outbreak of virus. This study provides information about genetic variations in different strains of Ebola virus. Identification of mutations and spread of virus are of great interest to know the viral fitness. In 2013, Hetzel and colleagues performed experiment with viruses isolated from Boa constrictor and virus was known as UHV (University of Helsinki virus). Sequencing of UHV virus was performed and then compared with other viruses detected in boid snake AvNLB3, high sequence identity was examined among viral genes showing it belongs to Arenavirus. Phylogenetic analysis was performed with divergent Arenavirus detected in snake showing AvNLB3 and UHV have high level of similarity and are closely related (Bodewes *et al*., 2013). Phylogenetic analysis of partial NP sequences isolated from 14 rodent species in South Africa and Zimbawe were performed, predicting that isolates belongs to Mopeia virus and also Merino Walk virus was characterized as one of the novel virus (Grobbelaar *et al.,* 2021).

A well known reliable index to study selection in a populaton is dN/dS (nonsynonymous substitutions per site/synonymous substitutions per site) (Ka/Ks) for a large set of genes, comparing the ratio of occurrence of synonymous and non-synonymous substitutions within a particular region based on comparisons of related species. No selection is observed when ratio of dN/dS=1. As, the ratio of Ka/Ks <1, is widely used as an indicator of the extent of purifying selection or negative selection acting to conserve coding sequences. Ratio of dN/dS>1 suggests positive selection. Codon alignment of viral sequences is performed to identify the domains or genes which are under selection and this information was applied to compute this test.

Computation of dN/dS parameter is highly preferred to measure rate of evolution of sequences (Roy *et al.,* 2018).

Phylogenetic analysis of GTOV, causative agent of VHF was performed by Cajimat and Fulhorst. Pirital virus one of the other virus which coexist with GTOV virus have high sequence similarity with GTOV virus and also Large segment of viral genome of these two viruses have high level of structure similarity. Although, phylogenetic analysis of nucleotide sequence of Z and L proteins grouped Pirital virus with Pichinde virus, both of which cause hemorrhagic fever in regions of South America (Belalov & Lukashev, 2013).

So, phylogenetic analysis of pathogens provides information about mutational region of genome and also region which is highly conserved. Further, this will elaborate information regarding the region which needs to be studied as drug target. In present study, analysis of evolutionary fluctuations and mutability dynamics of the crucial viral segments of arenaviral might shed light into the complex strategies employed by the pathogenic viruses to evade immune responses and defense mechanisms exhibited by associated host systems. Sliding window based evolutionary profiling would also be executed to gain insight into the conserved sights under extreme evolutionary constraint.

2.11 Pharmacology

Pharmacology is the branch of science that study how the drugs are targeted inside the body, what are the physiochemical properties of drugs inside the body and what are the toxic effects of drugs on the human system. Drugs have different potency and selectivity, and variable effects of drugs with varying doses are observed more frequently. With the advancement in high throughput sequencing techniques and structure prediction methods, the availability of protein sequence and structure has increased. This information may be helpful in evaluating human receptors as targets for drug binding. Many other proteins such as enzyme, transporters and ion channels can also be targeted for drug designing (Currie, 2018a).

Interaction of the drug with receptor can be determined by examining structure, shape, and reactivity, which predicts how tightly they bind. Drugs with weaker bond interactions bind to the receptor for short duration and drugs binding with stronger bond interactions bring about drug-receptor interactions for longer-

duration. These interactions include covalent bonds, ionic interactions, hydrogen bonds, hydrophobic effects, and Van der Waals forces. Drugs which bind with high affinity requires lower doses and have high rate of association as compare to drugs which binds with low affinity requires higher doses and have high rate of dissociation (Currie, 2018a).

Computational drug designing approaches are used to predict and evaluate drugs for various endemic (other diseases too) diseases. It has reduced the time span of effective and precise drug development. Two types of drug designing approaches are predominantly used-SBDD and LBDD. SBDD methods are based on the analysis of 3-dimensional structure of targets available in PDB which include proteins or RNA. Furthermore, important sites for interaction of target protein are profiled and evaluated. This information can be used for drug designing and prediction of drugs with the key sites of interaction, thus, interrupting the pathways required for survival of the pathogens. LBDD approaches focus on ligands which are already known for a target to organize structure-activity relationship, which defines the association between physiochemical properties and activities of ligands (Methods, 2017).

Extensive progress has been made in the areas of drug development pertaining to human viral pathogens. However, high mutability rates and variable genome dynamics of viruses have been the major obstacles in effective drug design against the detrimental pathogens. Ebola virus, one of the life threatening viruses causing severe health risk across the world was also studied. For this, researchers have used computational methods to design effective drugs against Ebola virus and compound ZINC58935541 was found to be best docked. This study will help in prediction of antiviral drug against Ebola virus (Mali & Chaudhari, 2019). Recently, usage of various approaches for *in silico* drug discovery against Zika virus, belonging to genus *Flavivirus* was also reported with NS3 helicase, one of the main components of replication in Zika virus (Cox *et al*., 2015). Recently Shah and colleagues presented the work based on *in silico* drug designing of Nipah virus and results of SBDD interpreted novel bioisosteres of Favipiravir as promising drugs (Shah *et al*., 2018).

With the increase in prevailing threat of Arenavirus, there is a rising demand to design drugs for this group of viruses also. Many small molecules have been

developed to reduce the viral infection but no FDA approved drug has yet been designed to prevent infection caused by *Mammarenavirus,* an alarm for global health.

2.11.1 Pharmacokinetics

Pharmacology is further divided into pharmacodynamics and pharmacokinetics. This study provides an insight into the biological behaviour of the drug inside the body. Drug overcomes many barriers including physical, chemical or biological barriers to administer to the final site. Passive diffusion and facilitated diffusion are the two methods by which drug passes through the membrane. Drugs selected based on binding interaction with receptor were further screened on the basis of physiochemical properties like absorption, distribution, metabolism, excretion and toxicity (ADMET). These properties provide information about how the body react to drug; which refers to movement of drug inside the body and out of the body. Study of these parameters helps us to increase the efficiency and decrease the toxicity of drug therapy designed for patients. Pharmacokinetics is the study of the physiochemical properties of drugs to know the effect of drug on the human body; that is, how the body affects the drug (Fig. 2.4). It is, however, also the study of associated toxicity (Currie, 2018b).

Pharmacodynamics study was performed to understand the alterations observed in the effect of drugs with time as opposed to concentration (Hugh & Jackson, 1994).

In the present study, virtual screening of potential ligands (small inhibitory molecules) was executed from publicly available small-molecule repositories like ZINC and PubChem. Extensive molecular docking and pertaining simulations was executed employing the probable lead molecules and the filtered viral (associated active binding sites and grooves). Subsequently, *in silico* absorption, distribution, metabolism, and excretion (ADME) and toxicity profiling were also executed to facilitate further experimental investigations and clinical trials pertaining to the proposed potential lead molecules from our analysis.

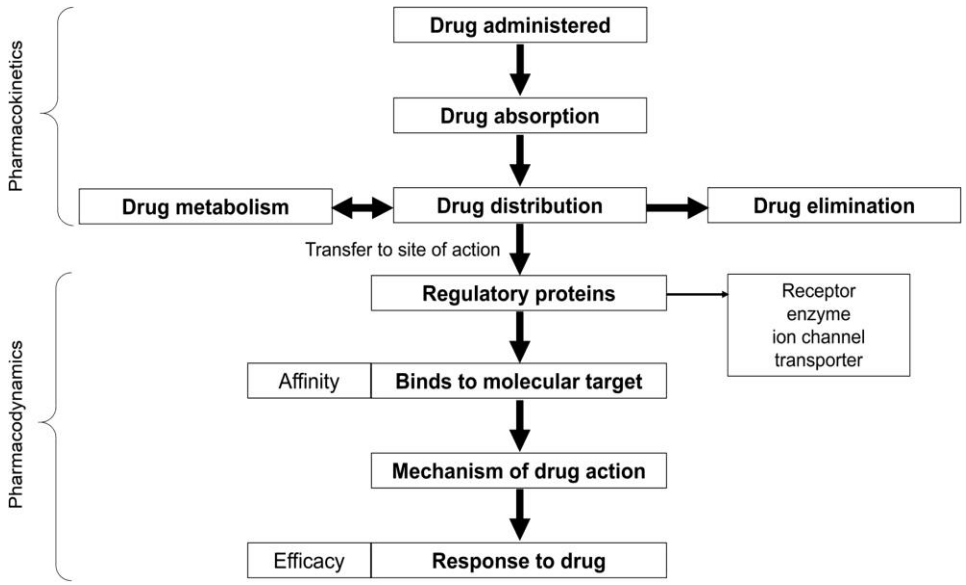


Figure 2.4: Flowchart showing mechanism of drug discovery (Currie, 2018b)

Research Gap

1. Demand for study of NW Arenaviruses; JUNV, MACV, GTOV, SABV and CHPV, at genomic level is required urgently since they are considered to be bioterrorism agents and causative agents of fatal hemorrhagic fever.
2. Viruses depend on host machinery for protein biogenesis. As no study has been performed to understand the adaptation of viruses inside the host machinery, this area needs to be studied to analyze the mechanism by which viruses cause infection inside the host cell.
3. Evolutionary dynamics of NW Arenaviruses also remains unattended and obscure.
4. Prediction of several small molecules inhibiting NW Arenaviruses has been performed yet no FDA-licensed drug exists for treatment or prevention of Arenavirus. Ribavirin is the only antiviral drug available against JUNV. Scarcity of effective drugs against the menacing Arenavirus is another domain of viral genomics that needs to be catered.

CHAPTER-3

HYPOTHESIS

1. Codon usage bias is an imperative genomic index which is frequently employed to study the molecular evolution of viruses (JUNV, MACV, GTOV, SABV and CHPV) and design effective drugs.

2. Adaptation study of NW Arenaviruses (JUNV, MACV, GTOV, SABV and CHPV) with their hosts elucidates the mechanism that viruses employ to infect host cellular machinery.

3. Pertaining evolutionary fluctuations and mutability dynamics of the crucial segments of NW arenaviral (JUNV, MACV, GTOV, SABV and CHPV) might shed light into the complex strategies employed by the pathogenic viruses to evade immune responses and defense mechanisms exhibited by associated host systems.

4. Extensive molecular docking and pertaining simulations would be executed employing the probable lead molecules against New World Arenaviruses (JUNV, MACV, GTOV, SABV and CHPV).

CHAPTER-4

RESEARCH OBJECTIVES

Objective1:

Investigating the riddles of codon and amino acid usage patterns in the New World Arenaviral pathogens (JUNV, MACV, GTOV, SABV and CHPV).

Objective2:

Profiling the influence of host machinery on the progression of disease by New World Arenaviruses (JUNV, MACV, GTOV, SABV and CHPV).

Objective3:

Exploring the evolutionary signatures of crucial New World Arenaviral (JUNV, MACV, GTOV, SABV and CHPV) components

Objective4:

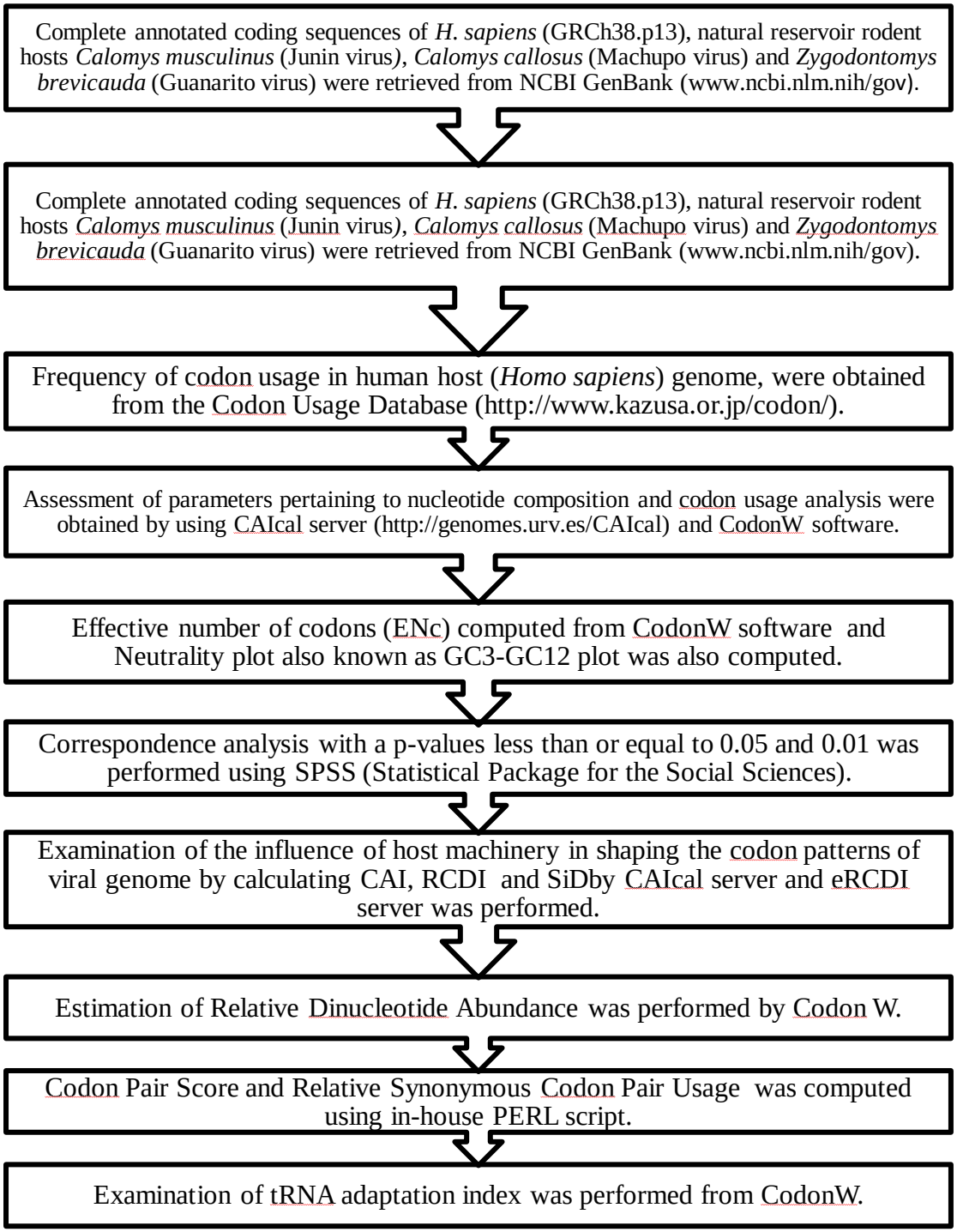
In silico drug development against New World Arenaviruses (JUNV, MACV, GTOV, SABV and CHPV) pathogens affecting human beings worldwide

CHAPTER-5

MATERIAL AND METHODS

5.1 ResearchDesign

This study was conducted in Bioinformatics lab, Department of Biotechnology, school of Bioengineering and Biosciences, Lovely Professional University, Phagwara. Framework of methods and softwares used to conduct the present research study on Arenaviruses has been shown below:



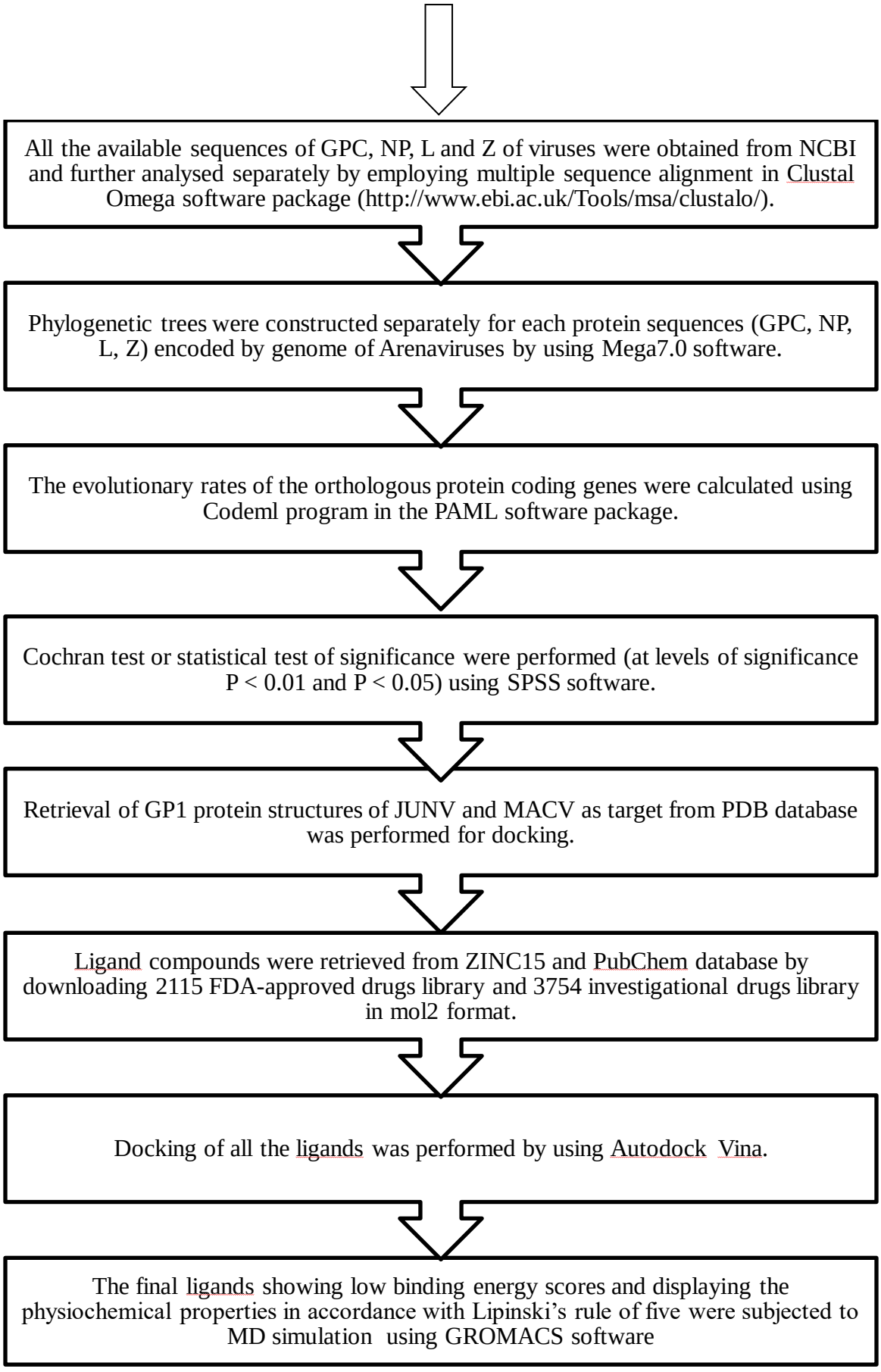


Figure 5.2: Research Design

5.2 Retrieval of coding sequences

Coding sequences of selected strains of JUNV (JUNV), MACV (MACV), GTOV (GTOV), SABV (SABV) and CHPV (CHPV) viruses were considered for extensive examination of codon and amino acid usage. Complete coding sequences and respective encoded protein products for viruses comprising of large RNA (L) segment (L and Zs) and small RNA (S) segment (GPC and NP proteins) were retrieved from NCBI (www.ncbi.nlm.nih.gov/genbank/) (Benson *et al*., 2018) and Virus Pathogen Resource (ViPR) database (Pickett *et al*., 2012). The viral strains accredited with 'Finished' status in the ViPR database (Pickett *et al*., 2012) showing completion of sequencing of genome sequence were considered for further analysis. Furthermore, sequences having unknown start and stop codons at the start and termination site were rejected from the final data file. Also, sequences displaying internal stop codons were rejected with the purpose to avert sampling lapse and stochastic variations (Wright, 1990). The final dataset comprised of 54 JUNV, 45 MACV and 45 GTOV, 5 SABV and 4 CHPV genomes and have been detailed in Table 5.1, 5.2, 5.3, 5.4 and 5.5. Human beings (*Homo sapiens*) serve as a common host to all the selected members of NW Arenavirus. Therefore, complete annotated coding sequences of *H. sapiens* (GRCh38.p13) were retrieved from NCBI GenBank (www.ncbi.nlm.nih.gov/genbank/) (Benson *et al.*, 2018) for our investigation. Coding sequences of the natural reservoir rodent hosts of JUNV (*Calomys musculinus)*, MACV (*Calomys callosus)* and GTOV (*Zygodontomys brevicauda)* were also retrieved from NCBI GenBank (Benson *et al.*, 2018) as shown in Table 5.6, 5.7and 5.8. Frequency of codon usage in human host (*Homo sapiens*) genome, were obtained from the Codon Usage Database (http://www.kazusa.or.jp/codon/) (Nakamura *et al*., 2000).

Table 5.1: List of strains of JUNV for codon and amino acid usage analysis

Strain Name	GenBank Accession	Segment	Protein Name
P2290_Ledesma	KR260732	S	G1, G2, hypotheticaL, GPC
P2290_Ledesma	KR260733	L	L
P3406_Ledesma	KR260730	S	G1, G2, hypotheticaL, GPC
P3406_Ledesma	KR260731	L	L
P35032	KU978799	L	L, Z
P35032	KU978800	S	G1, G2, hypotheticaL, GPC
P3766	KR260734	S	G1, G2, hypotheticaL, GPC

P3766	KR260735	L	L
Candid #1	AY819707	L	L, Z
Candid #1	FJ969442	S	G1, G2, GPC, NP
Candid-1	AY746353	S	GPC, G1, G2,NP
Candid-1	AY746354	L	L, Z
Candid1-rec	HQ126700	L	L, Z
Candid1-rec	HQ126701	S	G1, G2, GPC, NP
Candid1-wt	HQ126698	L	L, Z
Candid1-wt	HQ126699	S	G1, G2, GPC, NP
Cba IV4454	DQ272266	S	G1, G2, GPC,NP
JNM-6682	DQ531486	S	G1, G2, GPC, NP
JNM-7354	DQ531488	S	G1, G2, GPC, NP
MC2	AY216507	L	L, Z
MC2	D10072	S	G1, G2, GPC, NP
P1998	DQ854730	S	G1, G2, GPC, NP
P2031	DQ854731	S	G1, G2, GPC, NP
P2045	DQ854733	S	G1, G2, GPC, NP
P2290	DQ854736	S	G1, G2, GPC, NP
P3096	DQ854737	S	G1, G2, GPC, NP
P35302	DQ854738	S	G1, G2, GPC, NP
P3766	DQ854735	S	G1, G2, GPC, NP
P3778	DQ854734	S	G1, G2, GPC, NP
P3790	DQ854739	S	G1, G2, GPC, NP
P3790	JN714129	S	G1, G2, GPC, NP
P3790	JN714130	L	Z , L
P4036	DQ854732	S	G1, G2, GPC, NP
recXJ13	JN200801	S	G1, G2, GPC, NP
Rumero	AY619640	L	L , Z
Rumero	AY619641	S	G1, G2, GPC, NP
Rumero	JN801476	S	G1, G2, GPC, NP
Rumero	JN801477	L	Z , L
rXJ13	FJ805379	L	L, Z
rXJ13	FJ805380	S	G1, G2, GPC, NP
XJ#44	DQ489718	L	L, Z
XJ#44	GQ121040	S	G1, G2, GPC,NP
XJ13	AY358022	L	L, Z
XJ13	AY358023	S	G1, G2, GPC, NP
XJ13	FJ805377	L	L, Z
XJ13	FJ805378	S	G1, G2, GPC, NP
XJ17	JF799977	L	L, Z
XJ17	JF799981	S	GPC, N, G1, G2
XJ34	JF799978	L	L, Z
XJ34	JF799982	S	GPC, N, G1, G2
XJ39	JF799979	L	L, Z
XJ39	JF799983	S	GPC, N, G1, G2
XJ48	JF799980	L	L, Z

XJ48	JF799984	S	GPC, N, G1, G2

Table 5.2: List of strains of MACV for codon and amino acid analysis

Strain Name	Sequence length	Segment	Protein name
Carvallo	AY358021	L	L, Z
Carvallo	AY129248	S	GPC, NP
Mallele	JN794586	S	GPC, NP
Carvallo	JN794584	S	GPC, NP
Carvallo	JN794583	L	L, Z
Mallele	JN794585	L	L, Z
Carvallo	KM198593	L	L, Z
Carvallo	KM198592	S	GPC, NP
Chicava	AY624354	L	L, Z
Mallele	AY619645	S	G1, G2, GPC, NP
9530537	AY571959	S	GPC, NP
Carvallo	AY619642	L	L, Z
9301012	AY924205	S	GPC, NP
MARU 249121	AY924208	S	GPC, NP
Carvallo	AF485260	S	GPC, NP
Chicava	AY624355	S	G1, G2, GPC, NP
Carvallo	AY619643	S	G1, G2, GPC, NP
200002427	AY924204	S	GPC, NP
9430084	AY924203	S	GPC, NP
MARU 222688	AY924207	S	GPC, NP
Carvallo	AY216511	L	L, Z
Mallele	AY619644	L	L, Z
Chicava	AY924202	S	GPC, NP
Carvallo	AY571904	S	GPC, NP
MARU 216606	AY924206	S	GPC, NP
SPB201004275	KU978805	L	L, hypothetical
SPB201004275	KU978804	S	hypothetical, GPC
OBT2102	KU978798	S	hypothetical, GPC
Calomys 9301012	KU978797	S	hypothetical, GPC
Calomys 14795	KU978796	S	hypothetical, GPC
9530537	KU978795	S	hypothetical, GPC
9430069	KU978794	S	hypothetical, GPC
FSB3270	KU978792	S	hypothetical, GPC
Calomys 9301012	KU978790	L	L, hypothetical
Calomys 14795	KU978789	L	L, hypothetical
9530537	KU978788	L	L, hypothetical
9430069	KU978787	L	L, hypothetical
Calomys 221600	KU978786	L	L, hypothetical

930060_Chicava	KU978785	L	L, hypothetical
FSB3270	KU978784	L	L, hypothetical
Calomys 221600	KU978783	S	hypothetical, GPC
930060_Chicava	KU978793	S	hypothetical, GPC
OBT2102	KU978791	L	L, hypothetical
Carvallo	NC_005078*	S	GPC, NP
Carvallo	NC_005079*	L	L, Z

Table 5.3: List of strains of GTOV for codon and amino acid analysis

Strain name	**GenBank accession number**	**Segment**	**Protein Name**
260456	KR260720	S	hypothetical, GPC
260456	KR260721	L	L
308573	KR260718	S	hypothetical, GPC
308573	KR260719	L	L
329330	KR260722	S	hypothetical, GPC
329330	KR260723	L	L
362273	KR260728	S	hypothetical, GPC
362273	KR260729	L	L
CFF106	KR260724	S	hypothetical, GPC
CFF106	KR260725	L	L
CVH-950801	KU746279	L	L, Z
CVH-950801	KU746280	S	hypothetical, GPC
CVH-960104	KU746281	S	hypothetical, GPC
CVH-960104	KU746282	L	L, Z
CVH-960201	KU746283	L	L, Z
CVH-960201	KU746284	S	hypothetical, GPC
INH-95551	NC_005077*	S	GPC, NP
INH-95551	NC_005082*	L	L, Z
S-26764	KR260726	S	hypothetical, GPC
S-26764	KR260727	L	L
VHF1608	KU059748	L	L, Z
VHF1608	KU059749	S	hypothetical, GPC
VHF1750	KU746273	S	hypothetical, GPC

VHF1750	KU746274	L	L, Z
VHF1986	KU746275	L	L, Z
VHF1986	KU746276	S	hypothetical, GPC
VHF3900	KU746277	S	hypothetical, GPC
VHF3900	KU746278	L	L, Z
VHF4016	KU746285	L	L, Z
VHF4016	KU746286	S	hypothetical, GPC
VHF4696	KU746287	S	hypothetical, GPC
VHF4696	KU746288	L	L, Z
AV 97021119	AY573922	S	GPC, NP
CVH-950801	AY572557	S	GPC, NP
CVH-960101	AY497548	S	GPC, NP
CVH-960102	AY572558	S	GPC, NP
CVH-960103	AY572556	S	GPC, NP
CVH-960302	AY572555	S	GPC, NP
CVH-961104	AY572561	S	GPC, NP
INH-95551	AY129247	S	GPC, NP
INH-95551	AY216504	L	L, Z
S-56764	AY572554	S	GPC, NP
VAV-1608	AY572560	S	GPC, NP
VHF-1750	AY572559	S	GPC, NP
VHF-3990; AV 97021237	AY576604	S	GPC, NP

Table 5.4: List of strains of SABV for codon and amino acid analysis

Strain name	GenBank accession number	Segment	Protein Name
SPH114202	AY216506	L	L, Z
SPH114202	AY358026	L	L, Z
SPH114202	JN801474	S	GPC, NP
SPH114202	JN801475	L	Z, L
SPH114202	U41071	S	GPC, NP

Table 5.5: List of strains of CHPV for codon and amino acid analysis

Strain	GenBank Accession	Sequence Length	Segment	Protein Name
810419	EU260463	3357	S	GPC, NP
810419	EU260464	7107	L	L, Z
810419	NC_010562*	3357	S	GPC, NP
810419	NC_010563*	7107	L	L, Z

Table 5.6: List of selected sequences of *C. musculinus* (host of JUNV)

S.No	Accession number	Protein id
1.	EU164541.1	ABX89905.1
2.	U83817.1	AAB87172.1_1
3.	U83817.1	AAB87173.1_1
4.	U83817.1	AAB87174.1_1
5.	HM167814.1	ADM53537.1
6.	HM167815.1	ADM53538.1
7.	HM167816.1	ADM53539.1
8.	HM167817.1	ADM53540.1_1
9.	HM167822.1	ADM53545.1
10.	AF385599.1	AAL07679.1
11.	AF385600.1	AAL07680.1
12.	AF385601.1	AAL07681.1
13.	AF385602.1	AAL07682.1
14.	AF385603.1	AAL07683.1
15.	AF385604.1	AAL07684.1
16.	MF110338.1	ASV65584.1_1
17.	KX987858.1	AQY60260.1
18.	KX987859.1	AQY60261.1
19.	DQ029205.1	AAZ20175.1
20.	DQ029206.1	AAZ20176.1
21.	DQ029207.1	AAZ20177.1
22.	DQ029208.1	AAZ20178.1
23.	DQ029209.1	AAZ20179.1
24.	DQ029210.1	AAZ20180.1
25.	DQ029212.1	AAZ20182.1
26.	DQ029213.1	AAZ20183.1
27.	DQ452305.1	ABE03842.1_1
28.	DQ452306.1	ABE03843.1_1
29.	DQ452307.1	ABE03844.1_1
30.	DQ452308.1	ABE03845.1_1
31.	DQ452309.1	ABE03846.1_1
32.	DQ452310.1	ABE03847.1_1
33.	DQ452311.1	ABE03848.1_1
34.	DQ452312.1	ABE03849.1_1
35.	DQ452313.1	ABE03850.1_1
36.	DQ452314.1	ABE03851.1_1
37.	DQ452315.1	ABE03852.1_1
38.	DQ452316.1	ABE03853.1_1
39.	DQ452317.1	ABE03854.1_1
40.	DQ452318.1	ABE03855.1_1
41.	DQ452319.1	ABE03856.1_1
42.	DQ452320.1	ABE03857.1_1
43.	DQ452321.1	ABE03858.1_1

44.	DQ452322.1	ABE03859.1_1
45.	DQ452323.1	ABE03860.1_1
46.	DQ452324.1	ABE03861.1_1
47.	DQ452325.1	ABE03862.1_1
48.	DQ452326.1	ABE03863.1_1
49.	DQ452327.1	ABE03864.1_1
50.	DQ452328.1	ABE03865.1_1
51.	DQ452329.1	ABE03866.1_1
52.	DQ452330.1	ABE03867.1_1
53.	DQ452331.1	ABE03868.1_1
54.	EU255275.1	ABX54883.1_1

Table 5.7: List of selected sequences of *C. callosus* (host of MACV)

S.No	Accession number	Protein id
1	MF110335.1	ASV65581.1
2	KT965002.1	ALM98874.1
3	KT950918.1	ALL28982.1
4	KT950893.1	ALL28958.1
5	AY033188.1	AAK56441.1
6	AY033187.1	AAK56440.1
7	AY033186.1	AAK56439.1
8	AY033185.1	AAK56438.1
9	AY033184.1	AAK56437.1
10	AY033183.1	AAK56436.1
11	AY033182.1	AAK56435.1
12	AY033181.1	AAK56434.1
13	AY033180.1	AAK56433.1
14	AY033179.1	AAK56432.1
15	AY033178.1	AAK56431.1
16	AY033177.1	AAK56430.1
17	DQ447282.1	ABE60827.1
18	DQ447281.1	ABE60826.1
19	DQ447280.1	ABE60825.1
20	DQ447279.1	ABE60824.1
21	AY439002.1	AAS00504.1_1
22	AY439001.1	AAS00503.1_1
23	AY041188.1	AAK64638.1_1
24	AF159293.1	AAF62383.1
25	AY275113.1	AAQ18671.1_1
26	KF207856.1	AHZ13689.1
27	KF207855.1	AHZ13688.1
28	KF207852.1	AHZ13685.1
29	KF207851.1	AHZ13684.1
30	KF207841.1	AHZ13674.1
31	KC953163.1	AGX25609.1_1

32	U83819.1	AAB87165.1_1
33	U83819.1	AAB87164.1_1
34	U83819.1	AAB87163.1_1
35	EU164540.1	ABX89904.1

Table 5.8: List of selected sequences of *Z. brevicauda* (host of GTOV)

S.No	Accession Number	Protein Id
1.	JF492743.1	AEB19615.1
2.	JF492744.1	AEB19616.1
3.	JF492745.1	AEB19617.1
4.	JF492746.1	AEB19618.1
5.	JF492747.1	AEB19619.1
6.	JF492748.1	AEB19620.1
7.	JF492749.1	AEB19621.1
8.	JF492750.1	AEB19622.1
9.	JF492751.1	AEB19623.1
10.	JF492752.1	AEB19624.1
11.	JF492753.1	AEB19625.1
12.	JF492754.1	AEB19626.1
13.	JF492755.1	AEB19627.1
14.	EU579519.1	ACD76799.1_1
15.	EU579521.1	ACD76801.1_1
16.	HQ545533.1	ADQ87558.1_1
17.	HQ545575.1	ADQ87600.1
18.	HQ545576.1	ADQ87601.1
19.	HQ545577.1	ADQ87602.1
20.	HQ545603.1	ADQ87628.1
21.	HQ545604.1	ADQ87629.1
22.	HQ545605.1	ADQ87630.1
23.	HQ545606.1	ADQ87631.1
24.	HQ545637.1	ADQ87662.1
25.	HQ545638.1	ADQ87663.1
26.	GU397417.1	ADT70693.1_1
27.	EU340259.2	ABY87967.2_1
28.	MF110573.1	ASV65819.1_1

5.3 Assessment of parameters pertaining to nucleotide composition and codon usage analysis

Nucleotide composition properties like %A (Adenine), %G (Guanine), %C (Cytosine) and %T (Thymine); occurrence of GC (Guanine + Cytosine) at all the three positions of synonymous codons (GC1, GC2 and GC3) were computed in JUNV, MACV, GTOV, SABV and CHPV. Similarly, overall occurrence of AT

(frequency of Adenine + Cytosine) and GC (Guanine + Cytosine), A3s, G3s, C3s and T3s (nucleotide composition of Adenine, Guanine, Cytosine and Thymine at third position of synonymous codon) in JUNV, MACV, GTOV, SABV and CHPV were examined using CAIcal server (Puigbò *et al*., 2008). Further, grand average of hydropathicity (GRAVY) and aromaticity which are amino acid usage parameters were also examined by employing CAIcal server (http://genomes.urv.es/CAIcal) (Puigbò *et al*., 2008).

RSCU (Relative synonymous codon usage) one of the codon usage indices was computed by CodonW (Ver. 1.4.2) software (http://www.molbiol.ox.ac.uk/cu) (Peden, 1999) and formulated as:

$$\text{RSCU} = \frac{\text{Frequency of codon}}{\text{Expected frequency of codon (if codon usage was uniform)}}$$

RSCU calculate the preference for particular codon and non-randomness selection feature of genes (Sharpl & Li, 1987). Codons having RSCU value greater than 1.6 were acknowledged as overrepresented codons in genome and those having value less than 0.6 were examined as underrepresented codons (Gu *et al*., 2004). Accordingly, RSCU greater than 1.0 demonstrate positive codon usage biasness (Reis *et al*., 2003).

5.4 Effective number of codons

Effective number of codons (ENc) computed from CodonW software (http://www.molbiol.ox.ac.uk/cu) (Peden, 1999; Roy *et al.,* 2015) can have values from 20 to 61.Value equal to or close to 20 depicts that each amino acid has been encoded by one single codon only and there is no biasness whereas, value equal to or close to 61 shows that a particular amino acid can be encoded by more than one codon which is the case with no codon biasness. It provides information of preferential codons in the genome and codon usage patterns across the genes were investigated by computing ENc-GC3 plot (Grantham *et al*., 1981). ENc is computed from formula by Wright1990 as:

$$\text{ENc} = 2 + S + (29/S2 + (1 - S2))$$

Where S signifies frequency of GC3s.

5.5 Neutrality plot

Neutrality plot provides information about consequence of mutational constraints and natural selection on genes of viral genome. This plot is also known as

GC3-GC12 plot, where GC3 is plotted on X-axis and GC12 represented on Y-axis (Khandia *et al*., 2019). Slope value of the regression line (close to or above 1) reflects the consequence of mutational constraint only, value (close to or below 0) reflects natural selection effect also.

5.6 Correspondence analysis (CoA) of codon and amino acid usage data

Correspondence analysis (CoA) is a multivariate statistical tool computed to determine the variations occurring in intra and inter-genomic regions with reference to codon and amino acid usage patterns (Wang *et al*., 2011). CoA includes 59 orthogonal axes representing the 59 codons by exclusion of Met, Trp and 3 stop codons. Correspondence analysis with a p-values less than or equal to 0.05 and 0.01 was performed using SPSS (Statistical Package for the Social Sciences) (Frey, 2017) software to depict the changes in patterns of codon and amino acid in genome sequence. Correspondence analysis was executed on the basis of RSCU values for all the concerned genomes of JUNV, MACV, GTOV, SABV and CHPV viruses.

5.7 Codon adaptation index (CAI)

"Codon adaptation index (CAI) of a gene is defined as the relative adaptation of codon usage of the concerned gene with respect to the codon usage of the highly expressed genes" (Sharpl&Li, 1987, Roy *et al.,* 2015) and were determined by using CAIcal server (http://genomes.urv.es/CAIcal) (Puigbo *et al.,* 2008). CAI values vary from 0 to 1 and estimate viral gene adaptation inside the host cellular environment by employing a collection of highly expressed reference genes. A high CAI value (around 1) for a particular gene suggests a high level of correlation in its codon usage pattern with the host and significant adaptability to the host environment (Sen *et al.,* 2008).

5.8 Relative codon deoptimization index

Relative codon deoptimization index (RCDI) analyzes the degree of acclimatization of viral genomes inside the host microcellular environment and were assessed by RCDI/eRCDI server (Puigbò *et al*., 2010). If RCDI value is low indicating better adaptation and increased translation of a viral gene segment in host system (Ramaiah *et al*., 2017). RCDI values of the JUNV, MACV, GTOV, SABV and CHPV genes were assessed in reference to associated human and rodent host employing RCDI/eRCDI server (Puigbò *et al*., 2010).

5.9 Similarity index

Similarity index estimates the magnitude of the impact of host genome in driving codon usage patterns of viruses. Similarity index values ranges from 0 to 1, value close to 1 implies a thorough consequence of host on viral codon usage (Nasrullah *et al*., 2015).

Similarity index, referred as D (A, B) has been estimated as:

$$R(A,B) = \frac{\sum_{i=1}^{i=59} a_i X b_i}{\sqrt{\sum_{i=1}^{i=59} a_i^2 X \sum_{i=1}^{i=59} b_i^2}}$$

$$D(A,B) = \frac{1 - R(A,B)}{2}$$

"where R (A,B) represents the cosine value of an included angle between A and B spatial vectors" (Nasrullah *et al*., 2015) and is an estimate of similarity between NW Arenavirus and its respective host. ai refers to the RSCU value for a particular codon in JUNV,MACV, GTOV,SABV and CHPV. bi signifies the RSCU value for the same codon in case of the associated host.

5.10 Estimation of Relative Dinucleotide Abundance

Relative Dinuclotide Abundance (P_{xy}) was analyzed using CodonW software (http://www.molbiol.ox.ac.uk/cu) (Peden, 1999). P_{xy} value greater than 1.25 depicts over-representation of dinucleotides and P_{xy} value less than 0.78 show under-representation of dinucleotides.

5.11 Computation of Codon Pair Score and Relative Synonymous Codon Pair Usage

"Relative Synonymous Codon Pair Usage (RSCPU) represented as ratio of observed frequencies to the expected frequencies of codon pairs" (Cannarrozzi *et al*., 2010). RSCPU values were computed by using an in-house BioPerl script and further RSCPU values are used to analyze the Codon pair Score (CPS) values for codon pairs of selected Arenavirus and its host human and rodents by using script. Positive CPS scores show over-representation of codon pairs, whereas, negative CPS scores depicts under-representation of codon pairs for virus and host (Cannarrozzi *et al*., 2010).

5.12 Examination of tRNA adaptation index

tRNA adaptation index (tAI) estimates usage of tRNA by the coding sequences of viral genome. "tAI defines adaptation level of coding sequence of virus with the corresponding tRNA pool of host cell by computing the presence of tRNAs for every codon of coding sequence". tAI was computed from RSCU values of codons of viral genomes (Tuller *et al*., 2011).

5.13 Generation of multilocus sequence analysis (MLSA) based phylogeny among proteins of pathogenic members of Arenaviruses

All the available sequences of GPC, NP, L and Z of JUNV, MACV, GTOV, SABV and CHPV were obtained from NCBI (Benson *et al.,* 2018) and further analysed separately by employing multiple sequence alignment in Clustal Omega software package (http://www.ebi.ac.uk/Tools/ msa/clustalo/). Subsequently, after alignment of sequences, phylogenetic tree of interest for each protein was generated by using aligned sequences as input.

Phylogenetic trees were constructed separately for each protein sequences (GPC, NP, L, Z) encoded by genome of Arenavirus (JUNV, MACV, GTOV, SABV, CHPV) and Mega7.0 software (Kumar & Kumar, 2017) was employed for the purpose. Maximum Likelihood based phylogeny method and further Jones-Taylor-Thorton model of substitution with bootstrap value of 1000 replicates were examined for generation of the respective phylogenetic trees. Then, out group was selected from phylogenetic analysis to perform substitution selection.

5.14 Assessment of evolutionary signatures

The ratio dN (number of nonsynonymous substitutions per non-synonymous site) /dS (number of synonymous substitutions per synonymous site) has been found to be an excellent estimator of the evolutionary selection constraint on a protein-coding gene. dN/dS > 1 signifies positive natural selection whereas dN/dS < 1 symbolizes purifying (refining) selection. At neutral evolutionary stage, dN/dS = 1, i.e., the rate of synonymous and non-synonymous substitutions are equal" (Roy *et al*., 2015). The evolutionary rates of the orthologous protein coding genes were calculated using Codeml program in the PAML software package (Yang, 2007) (v4.9) (http:// abacus.gene.ucl.ac.uk/software paml.html) with runmode = -2 and CodonFreq = 1.

Pairs of sequences with low level of dN/dS value were examined as highly conserved and can be consider as target for drug designing.

5.15 Statistical analysis

Cochran test and statistical test of significance were performed (at levels of significance $P < 0.01$ and $P < 0.05$) by using SPSS software (Frey, 2017).

5.16 Retrieval of protein and ligand structures

GP1 subunit of glycoprotein binds to the human transferrin receptor 1 (htf1) and causes infection among humans (Rojek *et al.,* 2006). Highly resolved X-ray diffraction crystal structure of Glycoprotein (GP1) of JUNV ((PDB ID: 5W1K; Chains: E, J, P, R; Resolution: 3.99Angstrom) (Clark *et al*., 2018) and MACV (PDB ID: 5W1M; Chains: E, J, K, L; Resolution: 3.91 Angstrom) (Clark *et al*., 2018) was available only; furthermore, GP1 structure of GTOV, SABV and CHPV was not available and prediction of structures was done on basis of Threading method by using i-Tasser software (Manuscript, 2005), but predicted results were not stable and refined and which cannot be used further for docking analysis. Therefore, we retrieved GP1 protein structures of JUNV and MACV as target from PDB database (Bernstein *et al.*, 1977) to perform docking.

5.17 Processing of viral structures

Further refinement of both structures was performed by removal of water molecules, addition of polar hydrogen and Kollaman charges in AutoDock tools (Forli *et al*., 2016). Also, grid box was selected by defining grid dimensions as centre_X=-37.414, centre_Y=-0.048, centre_Z=-85.385; size_x=126, size_y=126, size_z=126 for GP1 of JUNV and center_X=75.663, center_Y=222.274, center_Z=221.976; size_x=126, size_y=104, size_z=126 for GP1 of MACV within their active site which was concluded using CASTp (Computer Atlas of Surface Topography of Proteins) server (Binkowski *et al*., 2003). Moreover, ligand compounds were retrieved from ZINC15 database (Ren *et al*., 2020) by downloading 2115 FDA-approved drugs library and 3754 investigational drugs library in mol2 format. Also, 38 ligands (literature) were retrieved from PubChem database (Kim *et al*., 2016) and considered for further analysis. In addition, compounds prevailing mol2 structures were converted to PDBQT format structures by using Open Babel tool (O'Boyle *et al*.,

2011) and further PRODRG server (Schüttelkopf & Van Aalten, 2004) was used for energy minimization of the structures.

Table 5.9: List of ligands selected for molecular docking against GP1protein of JUNV and MACV

S.No	Molecular Formula	Name	Refrences
1	C18H23N3O3	4-Piperidin-3-yloxy-6-propan-2-yloxyquinoline-7-carboxamide	(Plewe *et al*., 2020)
2	C17H14ClNO2	3-Chloro-5-methoxy-10-allylacridine-9(10H)-one	(Sepúlveda *et al*., 2008)
3	C13H17N3O4S3	1,1-Dioxo-2-propyl-3-[(prop-2-ynylamino)methyl]thieno[3,2-e]thiazine-6-sulfonamide	(Roccatagliata *et al*., 1996)
4	C26H36N2O9	[8-Butyl-3-[(3-formamido-2-hydroxybenzoyl)amino]-2,6-dimethyl-4,9-dioxo-1,5-dioxonan-7-yl] 3-methylbutanoate	(Ortiz-Riano *et al*., 2014)
5	C12H9F3N2O2	A771726	(Sepúlveda *et al*., 2018)
6	C3H5NO	Acrylamide	(Cordo & Candurra, 2003)
7	C10H11N5O4	Adenosine dialdehyde	(Andrei & De Clercq, 1990)
8	C9H8N4O4	5-(Aziridin-1-yl)-2-nitro-4-nitrosobenzamide	(Knox *et al*., 1993)
9	C25H27N5O5	Avn-944	(Dunham *et al*., 2018)
10	C2H4N4O2	1,2-Diazenedicarboxamide	(García *et al*., 2003)
11	C28H48O6	24-Epibrassinolide	(Wachsman *et al*., 2000)
12	C16H24O4	Brefeldin A	(Candurra & Damonte, 1997)
13	C8H10N4O2	Caffeine	(Candurra and Damonte 1999)
14	C10H5F3N4O	Carbonyl cyanide p-trifluoromethoxyphenylhydrazone	(Candurra & Damonte, 1997)
15	C12H16N4O3	3-(4-Aminoimidazo[4,5-c]pyridin-1-yl)-5-(hydroxymethyl)cyclopentane-1,2-diol	(Barradas *et al*., 2011)
16	C34H36N4O6	Chlorin e6	(Guo *et al*., 2011)
17	C17H19ClN2S	Chlorpromazine	(Candurra & Damonte, 1997)
18	C6H3Cl3	1,2,4-Trichlorobenzene	(Bakke *et al*., 1992)
19	C7H8O4	3-(Dihydroxymethyl)benzene-1,2-diol	(Wilson, M.S. , Metink-Kane, 2012)
20	C12H11F3N2O2	Antiproliferative agent A771726	(Sepúlveda *et al*., 2018)
21	C2H5O6P	2-Phosphoglycolic Acid	(Gong *et al*., 1999)
22	C19H28O2	Dehydroepiandrosterone	(Acosta *et al*., 2008)
23	C48H78N7O20P3S	(25R)-24-Oxo-DHCA-CoA	(Andrei & De Clercq, 1990)
24	C11H9N5O2	DHPA	(Andrei & De Clercq, 1990)
25	C19H30O2	Epiandrosterone	(Acosta *et al*., 2008)
26	C12H24O2	Lauric acid	(Bartolotta *et al*., 2001)
27	C9H5ClN4	m-Chlorophenyl carbonylcyanide hydrazone	(Candurra & Damonte, 1997)

28	C17H20O6	mycophenolic acid	(Gong *et al*., 1999)
29	C14H14N4O2S2	NSC 4492	(Sepúlveda *et al*., 2013)
30	C17H14Br2O	NSC 14560	(Sepúlveda *et al*., 2010)
31	C10H12N5Na2O8P	NSC 20265	(García *et al*., 2000)
32	C14H18N6O4S3	NSC 71033	(Sepúlveda *et al*., 2010)
33	C9H13N3O6	Pyrazofurin	(Andrei & De Clercq, 1993)
34	C8H12N4O5	Ribavirin	(Mammarenavirus *et al*., 2020)
35	C24H25N3O	ST-294	(Bolken *et al*., 2020)
36	C5H4FN3O2	T-705	(Y Furuta, T komeno, 2017)
37	C10H20N2S4	Tetradine	(Maskin, 1996)
38	C21H24F3N3S	TFP(trifluoperazine)	(Maskin, 1996)

5.18 Molecular Docking of the ligands with target structure

Screening of downloaded structures of ligands (.PDBQT format) was performed by docking each ligand within active site of GP1 target protein of JUNV and MACV separately in AutoDock vina software (Trott & Olson, 2010) and binding energy was calculated. Ligands showing low binding energy score were selected for further analysis (Kar *et al*., 2020).

5.19 Computation of physicochemical properties and potential toxicity of the selected ligands

Physiochemical properties of best selected ligands for JUNV and MACV were computed using SwissADME (Daina *et al*., 2017) and pkCSM servers (Pires *et al*., 2015). The Pan-Assay Interference Structures (PAINS) analysis (Baell & Holloway, 2010) was also performed in SwissADME server (Daina *et al*., 2017) for each selected ligand. The toxicity parameters like mutagenicity, carcinogenecity and cytotoxicity of the selected ligands were estimated using the ProTox-II web-server (Banerjee *et al*., 2018) and further validation was performed with the vNN-ADMET server (Schyman *et al*., 2017).

5.20 Visualization of interaction between ligand and target

Furthermore, based on the binding energy and physiochemical properties; binding poses of selected structures were studied. The molecular interactions (hydrogen bonds and hydrophobic interactions) between the target proteins and compounds were studied using LigPlot+ (Laskowski and Swindells, 2011) and PLIP (protein-ligand interactor profiler) (Salentin *et al*., 2015) tools (Fig 6.10, 6.11, 6.12, 6.13).

.

5.21 Computation of Molecular Dynamics simulations

Molecular Dynamics (MD) simulations predict the stability of protein-ligand complex and allow the validation of molecular docking results (Khan *et al.,* 2020; Kumar *et al.,* 2020). The final ligands showing low binding energy scores (in terms of binding energy values) and displaying the physiochemical properties in accordance with Lipinski's rule of five for GP1 protein of JUNV and MACV were subjected to MD simulation for a timescale of 120 nanoseconds (ns) ,at a standard temperature of 300 K and pressure level of 1.013 bar (Umesh *et al*., 2021) using GROMACS software (version 2019) (Abraham *et al.,* 2015).The equilibration steps were set with constant pressure and temperature (NPT) (Umesh *et al*., 2021). The parameters like the root mean square deviations (RMSD) and the root mean square fluctuations (RMSF) were computed to analyze the conformation and structural stability of respective protein-ligand complexes (Khan *et al*., 2020).

CHAPTER-6

RESULTS AND DISCUSSION

6.1 Codon and amino acid usage analysis

Extensive variations and aberrant trends have been inspected in genes and genomes for the usage of synonymous codons. Except methionine (Met) and tryptophan (Trp) (encoded by unique codons), all other amino acids are encoded by two to six codons that are known as synonymous codons. Usage of synonymous codons is often done in distinct densities to encode the protein sequences of a given organism and also among different organisms. Various factors like genomic mutational pressure (Karlin and Mrazek, 1996), natural selection for efficient translation (Romero *et al.*, 2003), gene expression level (Sharp and Li, 1986; Biome and Bernard 1999; Romero *et al.*, 2003;), abundance of transfer RNA (tRNA) (Ohkubo *et al.,* 1987; Duret, 2000), replicational transcriptional selection (McInerney 1998; Das *et al.,* 2005; Guo and Yu 2007; Guo and Yuan 2009; Romero *et al.*, 2000) have been reported to contribute to the usage of codon among various species. With the increase in the mortality rate of JUNV, MACV, GTOV, SABV and CHPV confers the need to understand the viruses at genomic level and also to perceive adaptation of viruses in various hosts mainly humans and rodents. Present research efforts have been focused to comprehensively traverse the complex codon usage profile of NW Arenavirus and also to explore the potential factors influencing the patterns of codon and amino acid usage.

6.2 Genomic AU richness in Arenavirus

Thorough scrutiny of viral codons or investigation of nucleotide composition of genomes of NW Arenaviral concedes the richness of AU codons among the JUNV, MACV, GTOV, SABV and CHPV genomes.

Through assessment of RSCU (Relative Synonymous Codon Usage), it was inspected that codon set greater than 1.0 demonstrates positive codon usage biasness (Parvathy *et al.*, 2022). Out of all possible codon sets (excluding start and stop codons) as shown in Table 6.1, 6.2, 6.3, 6.4 and 6.5; 49.15% (29 out of 59) in JUNV, 49.15% (29 out of 59) in MACV and 47.45% (28 out of 59) in GTOV, 49.45% (29 out of 59) in SABV and 47.45 % (28 out of 59) in CHPV were preferred (RSCU greater than 1.0) codon sets respectively.

The average AU and GC contents (%) were observed to be 56.65±3.16 and 43.34±3.16 in JUNV, 56.744±3.69 and 43.25±3.69 in MACV, 57.98±2.39 and 42.011±2.39 in GTOV, 59.406±2.51 and 40.593±2.51in SABV, 58.26±3.04 and 41.73±3.04 in CHPV respectively.

The average (%) of the nucleotides A (30.42±0.96 in JUNV, 30.89±1.23 in MACV, 30.78±1.44 in GTOV, 32.02±0.50 in SABV and 32.61±1.11 in CHPV) and U (26.23±2.72 in JUNV, 25.84±3.51 in MACV, 27.20±2.61 in GTOV, 27.38±2.72 in SABV and 25.64±4.04 in CHPV) were found to be substantially higher than G (22.44±1.62 in JUNV, 21.54±2.21 in MACV and 22.45±0.35 in GTOV, 21.18±0.85 in SABV and 20.88±1.77 in CHPV) and C (20.89±2.86 in JUNV, 21.71±5.48 in MACV , 19.55±2.39 in GTOV, 19.40±2.64 in SABV and 20.84±3.87 in CHPV). Also, occurrence of high frequency AU among preferred codons (19 out of 29 in JUNV; 16 out of 29 in MACV and 20 out of 28 in GTOV, 21 out of 29 in SABV and 20 out of 28 in CHPV) revealed AU richness.

It was also evident from results shown in Table 6.6 that (24 out of 29 in JUNV; 19 out of 29 in MACV and 24 out of 28 in GTOV, 26 out of 29 in SABV and 25 out of 28 in CHPV) preferred codons ended with A or U at third position. As third position of codon is highly redundant therefore, it is important to analyze whether the third position is purine (A or G) or pyrimidine (U or C). For example: AUU, AUC and AUA all with variation in third position code for isoleucine amino acid but AUG codes for methionine. So, study of positions gives information about occurrence of mutations arising in these three positions (Saier, 2019).

Further, codon analysis was also performed for natural hosts of JUNV, MACV and GTOV only as natural hosts for SABV and CHPV are still unknown. Results shown in Table 6.1, 6.2 and 6.3 conclude that out of all possible codon sets (excluding start and stop codons) 42.37 % (25 out of 59) in *C. musculinus* (host of JUNV), 44.06% (26 out of 59) by *C. Callosus* (host of MACV) *and* 44.06% (26 out of 59) were favoured by *Z. brevicauda* (host of GTOV) respectively. Also, occurrence of high frequency AU among preferred codons (14 out of 25 in *C. musculinus),* 14 out of 26 in *C. callosus* and 17 out of 26 in *Z. brevicauda* revealed AU richness in host genomes. It was also evident that (16 out of 25 in *C. musculinus,* 18 out of 26 codons

in *C. callosus* and 20 out of 26 codons in *Z. brevicauda*) preferred codons ended with A or U at third position.

Extensive analysis of genomic composition in the present study revealed a desire for AU rich codons over the GC in JUNV, MACV, GTOV, SABV, CHPV and hosts. Similar analysis was performed in other viruses also like in New castle virus showing preference for GC rich codons and A or T nucleotides are preferred at third position in codons (Kumar & Kumar, 2017). Further, Kumar and colleagues performed codons analysis in Equine influenza virus predicting preference for AT rich codons over GC rich (Kumar *et al.,* 2016). Furthermore, Roy and colleagues studied the codon preference in members of genus *Bifidobacterium* and all the members have shown preference for GC rich codons (Roy *et al.,* 2015) .

6.3 Factors contributing to codon usage analysis

6.3.1 GC3 versus ENc plot

GC3 versus ENc plot aid to elucidate the aspects affecting the codon usage patterns in viral genomes. This can be predicted from contingency of query gene values of viral genome in curve. If values prevail above or fall on the curve, mutational biasness is the only factor affecting the codon usage (Wright, 1990). However, values lying beneath the curve signify the occurrence of other aspects that includes natural selection. In-depth study of the GC3 versus ENc plot (Figure 6.1 (a, b, c, d and e)) for JUNV, MACV, GTOV, SABV and CHPV viral nucleotide sequences revealed the assemblage of genes below the ENc curve which indicates the integrated impact of mutational constraint and natural selection in influencing the codon usage of viral genomes. Genes showing ENc value less than 40 show strong codon biasness that can be explained as an outcome of pure compositional constraint (Comeron & Aguadé, 1998).

Table 6.1: Relative synonymous codon usage analysis of JUNV with hosts *H. sapiens* and *C. musculinus*

Aa	Codon	JUNV	*C. musculinus*	*H. sapiens*	Aa	Codon	JUNV	*C. musculinus*	*H. sapiens*
Phe	**TTT***	1.13	0.56	0.92	Ala	**GCT**	1.36	0.39	1.08
	TTC	0.87	1.44	1.08		GCC	0.71	2.25	1.60
Leu	**TTA**	1.09	1.49	0.48		**GCA***	1.71	1.29	0.92
	TTG*	1.69	0.39	0.78		GCG^	0.22	0.06	0.44
	CTT	0.97	0.58	0.78	Tyr	**TAT***	1.21	0.99	0.88
	CTC	0.60	0.94	1.20		TAC	0.79	1.01	1.12
	CTA	0.80	1.89	0.42	His	**CAT***	1.05	0.70	0.84
	CTG	0.85	0.70	2.40		CAC	0.95	1.30	1.16
Ile	**ATT***	1.26	1.23	1.08	Gln	**CAA***	1.09	1.70	0.54
	ATC	0.88	0.96	1.41		CAG	0.91	0.30	1.46
	ATA	0.86	0.81	0.51	Asn	**AAT***	1.11	0.72	0.94
Val	**GTT***	1.56	1.24	0.72		AAC	0.89	1.28	1.06
	GTC	0.73	0.57	0.96	Lys	**AAA***	1.07	1.73	0.86
	GTA^	0.53	1.94	0.48		AAG	0.93	0.27	1.14
	GTG	1.17	0.25	1.84	Asp	**GAT***	1.07	0.87	0.92
Ser	**TCT**	1.44	1.33	1.14		GAC	0.93	1.13	1.08
	TCC	1.01	1.34	1.32	Glu	**GAA***	1.13	1.64	0.84
	TCA	1.33	2.63	0.90		GAG	0.87	0.36	1.16
	TCG^	0.13	0.00	0.30	Cys	**TGT***	1.19	0.11	0.92
	AGT	1.06	0.09	0.90		TGC	0.81	1.89	1.08
	AGC*	3.17	0.61	1.44	Arg	CGT^	0.22	0.39	0.48
Pro	**CCT**	1.34	1.17	1.16		CGC^	0.06	1.75	1.08
	CCC	0.62	0.30	1.28		CGA^	0.32	2.82	0.66
	CCA*	1.79	2.22	1.12		CGG^	0.14	0.12	1.20
	CCG^	0.25	0.31	0.44		**AGA***	3.17	0.74	1.26
Thr	**ACT**	1.11	0.44	1.00		**AGG**	2.09	0.18	1.26
	ACC	0.89	0.81	1.44	Gly	**GGT**	1.19	0.62	0.64
	ACA*	1.93	2.47	1.12		GGC	0.88	0.94	1.36
	ACG^	0.06	0.29	0.44		**GGA***	1.04	1.81	1.00
						GGG	0.89	0.63	1.00

Aa stands for Amino acids; Codons (JUNV) having RSCU (Relative synonymous codon usage) > 1.00 have been marked in bold; codons (JUNV) rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; highly preferred codons (JUNV) for each amino acid has been marked with *; under-represented codons (JUNV) having RSCU value less than 0.60 has been marked with^; codons (JUNV) showing richness in G(Guanine) or C(Cytosine) nucleotides have been highlighted in green.

Table 6.2: Relative synonymous codon usage analysis of MACV with hosts *H. sapiens* and *C. callosus*

Aa	Codon	MACV	*C. callosus*	*H. sapiens*	Aa	Codon	MACV	*C. callosus*	*H. sapiens*
Phe	TTT	0.73	1.13	0.92	Ala	**GCT***	2.05	0.51	0.92
	TTC*	1.27	0.87	1.08		GCC	0.90	1.86	1.08
Leu	TTA	0.72	1.48	0.48		GCA	0.97	1.58	0.48
	TTG	0.91	0.43	0.78		GCG^	0.08	0.05	0.78
	CTT*	1.38	0.50	0.78	Tyr	**TAT***	1.58	1.02	0.78
	CTC	1.37	1.28	1.20		TAC^	0.42	0.98	1.20
	CTA	0.91	1.85	0.42	His	**CAT***	1.38	0.69	0.42
	CTG	0.70	0.46	2.40		CAC	0.62	1.31	2.40
Ile	**ATT**	1.07	1.17	1.08	Gln	CAA	0.80	1.10	1.08
	ATC*	1.23	1.11	1.41		**CAG***	1.20	0.90	1.41
	ATA	0.71	0.72	0.51	Asn	**AAT***	1.22	1.25	0.51
Val	**GTT**	1.12	0.79	0.72		AAC	0.78	0.75	0.72
	GTC	1.18	0.57	0.96	Lys	**AAA***	1.02	1.69	0.96
	GTA^	0.24	1.90	0.48		AAG	0.98	0.31	0.48
	GTG*	1.45	0.75	1.84	Asp	GAT	0.91	1.25	1.84
Ser	**TCT**	1.00	1.05	1.14		**GAC***	1.09	0.75	1.14
	TCC	0.71	1.19	1.32	Glu	**GAA***	1.02	0.86	1.32
	TCA*	1.84	2.85	0.90		GAG	0.98	1.14	0.90
	TCG^	0.21	0.04	0.30	Cys	**TGT***	1.61	0.70	0.30
	AGT	1.16	0.36	0.90		TGC	0.39	1.30	0.90
	AGC	1.08	0.51	1.44	Arg	CGT	0.28	0.81	1.44
Pro	**CCT**	1.33	0.63	1.16		CGC	0.00	1.26	1.16
	CCC	0.98	0.53	1.28		CGA	0.26	2.36	1.28
	CCA*	1.51	2.57	1.12		CGG	0.03	0.23	1.12
	CCG^	0.18	0.27	0.44		**AGA**	2.66	0.81	0.44
Thr	ACT	0.94	0.62	1.00		**AGG***	2.76	0.53	1.00
	ACC	1.35	1.37	1.44	Gly	GGT	0.86	0.47	1.44
	ACA*	1.70	1.83	1.12		**GGC**	1.01	0.72	1.12
	ACG^	0.01	0.18	0.44		**GGA***	1.53	2.22	0.44
						GGG	0.60	0.59	

Aa stands for Amino acids; codons (MACV) having RSCU (Relative synonymous codon usage) > 1.00 have been marked in bold; codons (MACV) rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; highly preferred codons (MACV) for each amino acid has been marked with *; under-represented codons (MACV) having RSCU value less than 0.60 has been marked with^; codons (MACV) showing richness in G (Guanine) or C (Cytosine) nucleotides have been highlighted in green.

Table 6.3: Relative synonymous codon usage analysis of GTOV with host *H. sapiens* and natural reservoir *Z. brevicauda*

Aa	Codon	GTOV	*Z. brevicauda*	*H. sapiens*	Aa	Codon	GTOV	*Z. brevicauda*	*H. sapiens*
Phe	**TTT***	1.17	0.98	0.92	Ala	**GCT**	1.49	0.94	1.08
	TTC	0.83	1.02	1.08		GCC	0.58	1.54	1.60
Leu	**TTA**	1.22	1.74	0.48		**GCA***	1.70	1.49	0.92
	TTG*	1.41	0.15	0.78		GCG	0.23	0.03	0.44
	CTT	1.12	1.29	0.78	Tyr	TAT	0.94	1.12	0.88
	CTC^	0.59	0.76	1.20		**TAC***	1.06	0.88	1.12
	CTA	0.82	1.89	0.42	His	**CAT***	1.17	0.99	0.84
	CTG	0.84	0.17	2.40		CAC	0.83	1.01	1.16
Ile	**ATT***	1.33	1.31	1.08	Gln	**CAA***	1.23	1.78	0.54
	ATC	0.77	0.53	1.41		CAG	0.77	0.22	1.46
	ATA	0.90	1.16	0.51	Asn	**AAT***	1.16	1.19	0.94
Val	**GTT***	1.53	1.31	0.72		AAC	0.84	0.81	1.06
	GTC	0.85	0.93	0.96	Lys	**AAA***	1.10	1.69	0.86
	GTA^	0.56	1.56	0.48		AAG	0.90	0.31	1.14
	GTG	1.06	0.20	1.84	Asp	**GAT***	1.25	0.95	0.92
Ser	**TCT**	1.46	1.99	1.14		GAC	0.75	1.05	1.08
	TCC	0.62	1.52	1.32	Glu	**GAA***	1.12	1.61	0.84
	TCA*	1.81	1.39	0.90		GAG	0.88	0.39	1.16
	TCG^	0.12	0.02	0.30	Cys	**TGT***	1.21	1.13	0.92
	AGT	1.35	0.49	0.90		TGC	0.79	0.88	1.08
	AGC	0.65	0.60	1.44	Arg	CGT	0.14	0.24	0.48
Pro	**CCT**	1.46	0.64	1.16		CGC	0.12	0.48	1.08
	CCC	0.75	1.03	1.28		CGA	0.17	4.14	0.66
	CCA*	1.47	2.17	1.12		CGG	0.08	0.24	1.20
	CCG^	0.32	0.16	0.44		**AGA***	2.83	0.71	1.26
Thr	**ACT**	1.00	0.89	1.00		**AGG**	2.65	0.19	1.26
	ACC^	0.59	0.93	1.44	Gly	**GGT***	1.39	0.64	0.64
	ACA*	2.23	2.04	1.12		GGC	0.77	0.83	1.36
	ACG^	0.18	0.14	0.44		**GGA**	1.05	2.13	1.00
						GGG	0.79	0.40	1.00

Aa stands for Amino acids; codons (GTOV) having RSCU (Relative synonymous codon usage) > 1.00 have been marked in bold; codons (GTOV) rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; highly preferred codons (GTOV) for each amino acid has been marked with *; under-represented codons (GTOV) having RSCU value less than 0.60 has been marked with^; codons (GTOV) showing richness in G (Guanine) or C (Cytosine) nucleotides have been highlighted in green.

Table 6.4: Relative synonymous codon usage analysis of SABV and *H. sapiens*

Aa	Codon	SABV	*H. sapiens*	Aa	Codon	SABV	*H. sapiens*
Phe	**TTT***	1.27	0.92	Ala	GCT	1.15	1.08
	TTC	0.73	1.08		GCC	0.62	1.60
Leu	*TTA*	1.28	0.48		**GCA***	2.13	0.92
	TTG*	1.65	0.78		GCG^	0.10	0.44
	CTT	1.12	0.78	Tyr	***TAT****	1.19	0.88
	CTC^	0.53	1.20		TAC	0.81	1.12
	CTA^	0.59	0.42	His	***CAT****	1.13	0.84
	CTG	0.82	2.40		CAC	0.87	1.16
Ile	***ATT****	1.27	1.08	Gln	***CAA****	1.20	0.54
	ATC	0.71	1.41		CAG	0.80	1.46
	ATA	1.02	0.51	Asn	***AAT****	1.31	0.94
Val	***GTT****	1.35	0.72		AAC	0.69	1.06
	GTC^	0.56	0.96	Lys	***AAA)****	1.09	0.86
	GTA	0.79	0.48		AAG	0.91	1.14
	GTG	1.30	1.84	Asp	***GAT****	1.29	0.92
Ser	*TCT*	1.58	1.14		GAC	0.71	1.08
	TCC	0.69	1.32	Glu	***GAA****	1.21	0.84
	TCA*	1.81	0.90		GAG	0.79	1.16
	TCG^	0.17	0.30	Cys	***TGT****	1.54	0.92
	AGT	1.13	0.90		TGC^	0.46	1.08
	AGC	0.63	1.44	Arg	CGT^	0.26	0.48
Pro	CCT	1.42	1.16		CGC^	0.13	1.08
	CCC	0.71	1.28		CGA^	0.25	0.66
	CCA*	1.51	1.12		CGG^	0.12	1.20
	CCG^	0.36	0.44		***AGA****	3.23	1.26
Thr	*ACT*	1.23	1.00		AGG	2.02	1.26
	ACC	0.87	1.44	Gly	**GGT***	1.48	0.64
	ACA*	1.79	1.12		GGC	0.60	1.36
	ACG^	0.10	0.44		GGA	1.06	1.00
					GGG	0.86	1.00

Aa stands for Amino acids; codons having RSCU (Relative synonymous codon usage) > 1.00 have been marked in bold; codons rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; highly preferred codons for each amino acid has been marked with *; under-represented codons having RSCU value less than 0.60 has been marked with^; codons showing richness in G (Guanine) or C (Cytosine) nucleotides have been highlighted in green.

Table 6.5: Relative synonymous codon usage analysis of CHPV and *H. sapiens*

Aa	Codon	CHPV	*H. sapiens*	Aa	Codon	CHPV	*H. sapiens*
Phe	***TTT****	1.24	0.92	Ala	***GCT****	1.55	1.08
	TTC	0.76	1.08		GCC	0.90	1.60
Leu	*TTA*	1.31	0.48		GCA	1.46	0.92
	TTG*	1.38	0.78		GCG	0.09	0.44
	CTT	1.04	0.78	Tyr	***TAT****	1.25	0.88
	CTC	0.67	1.20		TAC	0.75	1.12
	CTA	0.72	0.42	His	***CAT****	1.14	0.84
	CTG	0.89	2.40		CAC	0.86	1.16
Ile	***ATT****	1.27	1.08	Gln	***CAA****	1.12	0.54
	ATC	0.83	1.41		CAG	0.88	1.46
	ATA	0.90	0.51	Asn	***AAT****	1.14	0.94
Val	***GTT****	1.41	0.72		AAC	0.86	1.06
	GTC	0.80	0.96	Lys	***AAA****	1.18	0.86
	GTA	0.53	0.48		AAG	0.82	1.14
	GTG	1.26	1.84	Asp	***GAT****	1.20	0.92
Ser	*TCT*	1.34	1.14		GAC	0.80	1.08
	TCC	0.70	1.32	Glu	***GAA****	1.16	0.84
	TCA*	1.73	0.90		GAG	0.84	1.16
	TCG^	0.25	0.30	Cys	***TGT****	1.27	0.92
	AGT	1.29	0.90		TGC	0.73	1.08
	AGC	0.70	1.44	Arg	CGT^	0.19	0.48
Pro	CCT	1.20	1.16		CGC^	0.23	1.08
	CCC	0.80	1.28		CGA^	0.19	0.66
	CCA*	1.71	1.12		CGG^	0.11	1.20
	CCG	0.29	0.44		***AGA****	3.28	1.26
Thr	*ACT*	1.21	1.00		AGG	2.00	1.26
	ACC^	0.47	1.44	Gly	**GGT***	1.51	0.64
	ACA*	2.11	1.12		GGC^	0.52	1.36
	ACG^	0.21	0.44		GGA	1.14	1.00
					GGG	0.84	1.00

Aa stands for Amino acids; codons having RSCU (Relative synonymous codon usage) value >1.00 have been marked in bold; codons rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; highly preferred codons for each amino acid has been marked with *; under-represented codons having RSCU value less than 0.60 has been marked with^; codons showing richness in G (Guanine) or C (Cytosine) nucleotides have been highlighted in green.

Table 6.6: Showing average values of occurrence of nucleotides in viral genomes

	JUNV	MACV	GTOV	SABV	CHPV
%A(average)	30.42	30.89	30.78	32.02	32.61
%U (average)	26.23	25.84	27.204	27.38	25.64
%C (average)	20.89	21.71	19.55	19.407	20.84
%G (average)	22.44	21.54	22.45	21.18	20.88
%AU (average)	56.65	56.74	57.98	59.406	58.26
%GC (average)	43.34	43.25	42.011	40.59	41.73
%AU3(Average)	53.9923	54.40238	57.29544	58.82621	58.10367
%GC3(average)	46.0077	45.59762	42.70456	41.17379	41.89633

The average ENc values were found to be 50.78346 ± 1.92 for JUNV, 49.14698 ± 2.38 for MACV and 49.6744 ± 2.27 for GTOV, 50.144 ± 2.07 for SABV and 46.2375 ± 6.038 for CHPV; thus robust codon usage analysis of viruses predicts low codon usage biasness among viral sequences. Similar cases of RNA viruses showing low codon usage biasness have been reported earlier also (Duret 2000 ; Chen *et al.*, 2013). Low codon usage biasness in viral genome reduces the competition of the virus with its host for usage of host machinery for synthesis and also increases the efficiency of replication of the virus in host cells. Thus, a virus with low codon usage biasness might be able to adapt itself to host more easily and be able to survive in a broad range of hosts (Bahir *et al.*, 2009; Roy *et al.*, 2015).

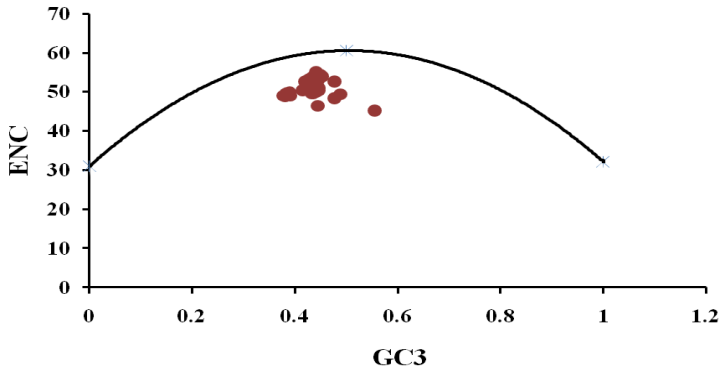


Figure 6.1a: GC3 versus ENc plot for JUNV

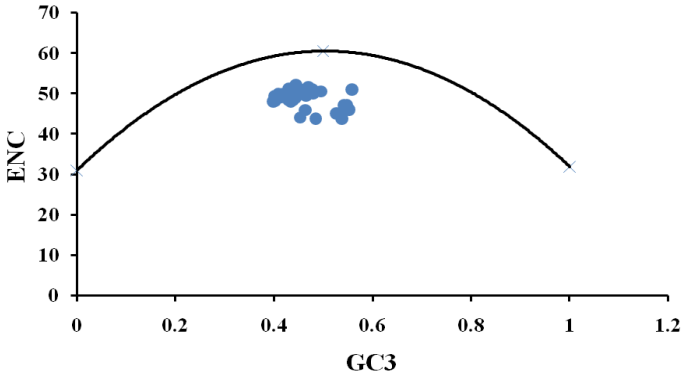


Figure 6.1b: GC3 versus ENc plot for MACV

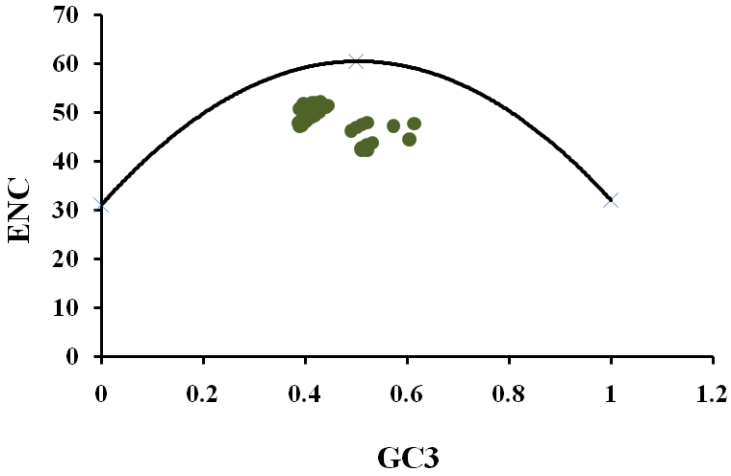


Figure 6.1c: GC3 versus ENc plot for GTOV

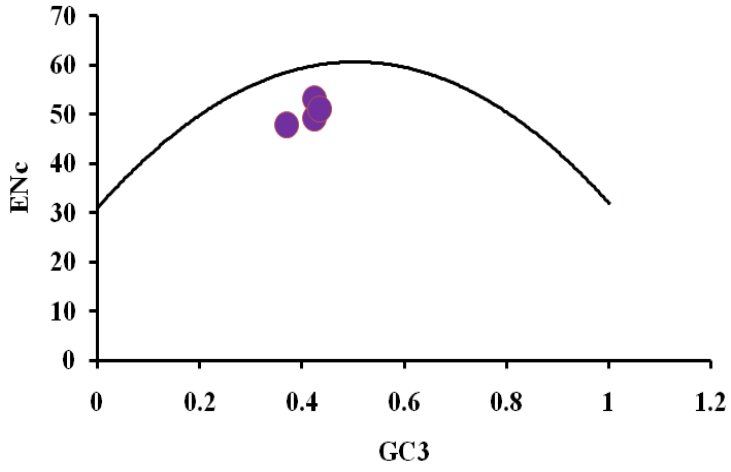


Figure 6.1d: GC3 versus ENc plot for SABV

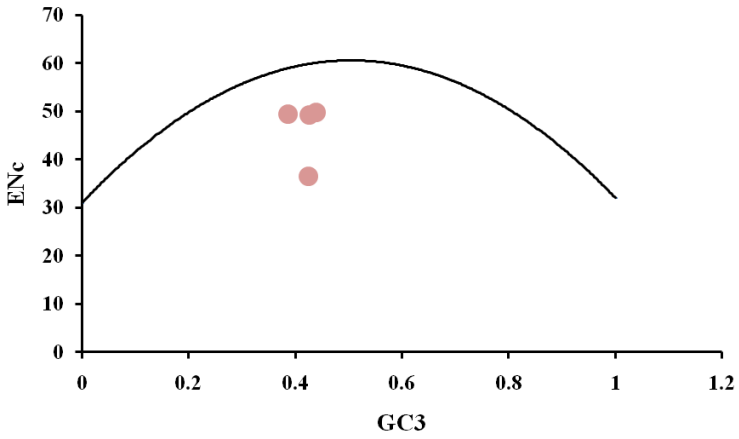


Figure 6.1e: GC3 versus ENc plot for CHPV

Analyzed genes have been marked as colored circles.The bell shaped solid line indicates the continuous Enc plot curve.

6.3.2 Neutrality plot

Neutrality plot revealed the effect of mutational constraints and natural selection on genes of viral genome (Khandia *et al*., 2019). The high value of slope of regression line (close to or above 1) reflects the effect of mutational constraint only; else value of slope (close to or below zero) provides information about natural selection as the dominating constraint affecting codon biasness (Nie *et al*., 2014). Comprehensive study revealed that the slope of the regression line in the neutrality plot (Figure 6.2 (a, b, c, d and e)) was around 0.699 signifying 69.9% influence of mutational constraint in JUNV, 0.683 signifying 68% influence of mutational biasness in MACV and 0.59 signifying 59 % in GTOV, 0.692 slope of regression line signifying 69.2% in SABV and 0.821 signifying 82.1% influence of mutational constraint in CHPV viral coding sequences.

Thus, it was evident from the analysis that compositional biasness dominated over natural selection for the codon usage patterns of viral genomes. In 2019, Khandia and colleagues performed similar analysis on codon patterns of Nipah virus where neutrality plot was constructed showing 0.405 values for slope of regression line. This analysis suggested 40% effect of mutational constraint and 60 % effect of natural selection on codon biasness of Nipah virus (Khandia *et al*., 2019).

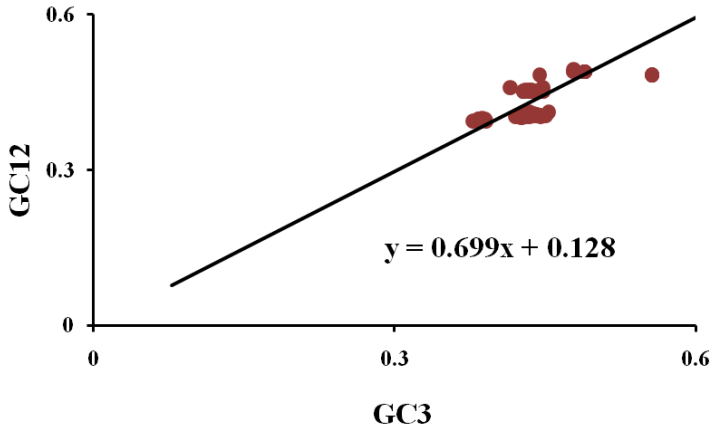


Figure 6.2a: Neutrality plot of JUNV

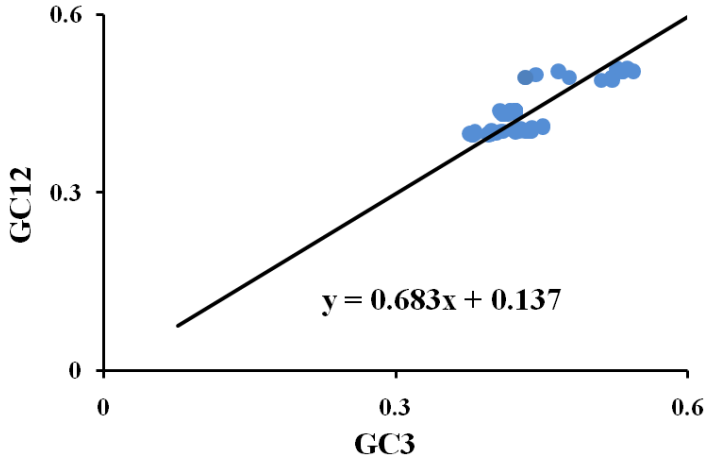


Figure 6.2b: Neutrality plot of MACV

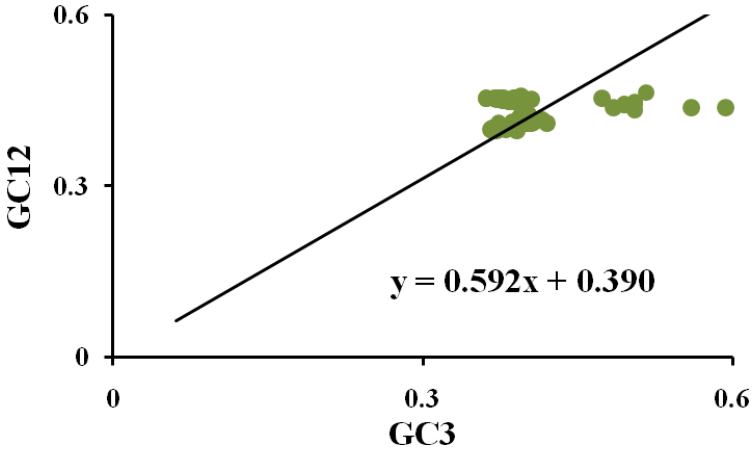


Figure 6.2c: Neutrality plot of GTOV

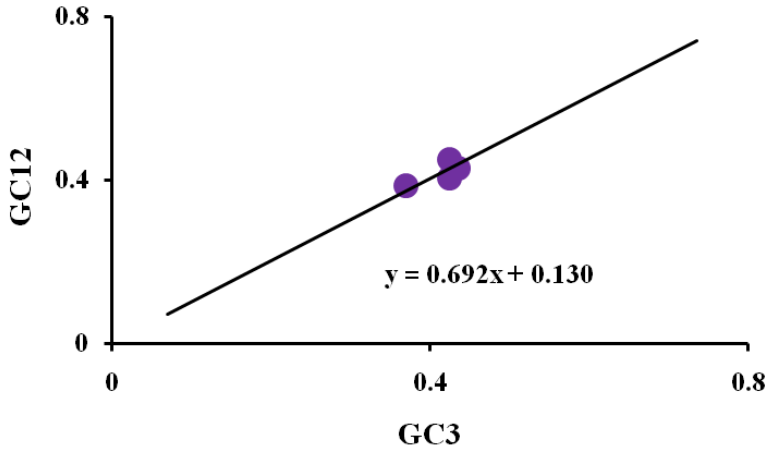


Figure 6.2d: Neutrality plot of SABV

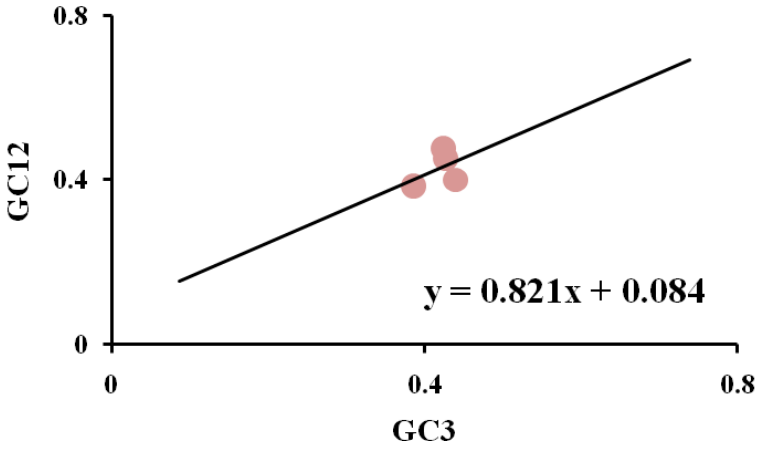


Figure 6.2e: Neutrality plot of CHPV

Analyzed genes have been marked as coloured circles. The slope of the plot defines the degree of compositional bias viable on the genomes of interest

6.3.3 Correspondence analysis revealed multiple determinants of codon usage in NW Arenaviruses

Correspondence analysis was performed on relative synonymous codon usage data (RSCU) of JUNV, MACV, GTOV, SABV and CHPV to determine the determinants of codon usage variation. High significant correlation of GC with axis1 and axis2 (the two major principle axes of separation of genes) of RSCU data was observed in JUNV, GTOV. In MACV GC was found to be highly correlated with axis1 only, showing the influence of compositional constraint. Immense level of significant correlation of GC with axis2 (one of the major axis of separation of genes) was observed in SABV and CHPV (Malhotra & Kumar, 2021).

Similarly, significant correlation of GC3 with axis1 and axis2 (the two major principle axes of separation of genes) of RSCU data was observed in GTOV. In JUNV and MACV GC3 was found to be highly correlated with axis1 only, showing the influence of compositional constraint. Immense level of significant correlation of GC3 with axis2 (one of the major axis of separation of genes) was observed in SABV and CHPV (Malhotra & Kumar, 2021) (Table 6.7). Similar analysis was also performed in 2013 in *T. pisiformis* predicting high level of correlation of GC and GC3 with genes on axis1 (Chen *et al*., 2013).

Assessment of CAI in viral genes and genomes efficiently portray the role of natural selection operating on them. RSCU data on axis1 commence to show significant correlation with CAI, RCDI of SABV and CHPV genomes; RCDI of GTOV, MACV (Table 6.7); whereas axis 1 and 2 of RSCU data were found to correlate significantly with CAI of the JUNV, MACV and GTOV genomes (Table 6.7), RCDI of JUNV and RCDI of MACV, GTOV, SABV and CHPV correlate significantly with RSCU data on axis1. Thus, depicting an indubitable influence of natural selection.

Similar analysis was performed in *E. festucae* to determine correlation of codon biasness with various parameters showing negative correlation of CAI with data on axis 1 and 2. Finally, inferring the effect of ENc as one of the major factor affecting codon biasness (Li *et al.,* 2016).

Also, other elements such as GRAVY (grand average of hydropathicity) and aromaticity show significant level of correlation with RSCU data on axis2 of SABV and CHPV. It was evident from Table 6.7 that factors like hydropathicity index (GRAVY [positive GRAVY (hydrophobic), negative GRAVY (hydrophilic)]) correlates with axis 1 and axis 2 of RSCU data of JUNV, MACV and GTOV viral genomes and aromaticity of encoded viral gene products correlated significantly with RSCU data on axis 2 of all viral genomes. Length of the viral coding sequences correlated significantly with axis 1 of RSCU of JUNV, MACV and GTOV (Table 6.7). Similar results were observed in codon analysis of *T. pisiformis,* showing significant effect of factors aromaticity and hydrophobicity on codon usage patterns (Chen *et al*., 2013).

Thus, codon usage patterns of the JUNV, MACV, GTOV, SABV and CHPV emerged to be a complex interplay of various crucial determinants with compositional

bias playing the most dominant role. Codon usage patterns of the JUNV, MACV, GTOV, SABV and CHPV appeared to be affected by various factors like compositional bias, natural selection, length, hydropathicity and aromaticity of the viral coding sequences.

In spite of a convoluted interplay of various determinants, compositional constraint was found to play the most dominant role in shaping codon usage of JUNV, MACV, GTOV, SABV and CHPV. In 2019, Roy and colleagues also studied CoA multivariate statistical analysis on codon data of genus *Puccinia* and predicted the dominating influence of mutational pressure on codon biasness (Roy and van Staden 2019).

6.4 Relative Dinucleotide Abundance in NW Arenaviral

Analysis of relative dinucleotide abundance vigorously in NW Arenaviral revealed that UpG and CpA dinucleotides were over-represented (Figure 6.3 (a, b, c, d, e)). Same dinucleotides were also observed to be highly preferred in *H. sapiens* (Figure 6.3 (a, b, c, d, e)), a characteristic pattern of vertebrate genomes *whereas C. musculinus, C. callosus and Z. brevicauda* (natural reservoirs) show GpG dinucleotides as over-represented.

Dinucleotides showing P_{xy} (observed/expected frequencies of dinucleotides xy) > 1.25 were considered to be over-represented, whereas, (xy) dinucleotides having $P_{xy} < 0.78$ were inferred as under-represented.

Thorough analysis of RSCU revealed that UpG containing codons like UUG, UGU and GUG in JUNV, GTOV, SABV and CHPV; GUG, UGU in MACV and CpA containing codons like UCA, ACA, CCA, GCA, CAU and CAA in JUNV, GTOV, SABV and CHPV; UCA, CCA, ACA, CAG and CAU in MACV were preferred (RSCU > 1.00) as shown in Table 6.1, 6.2, 6.3, 6.4 and 6.5. This result correlated well with the over-representation of the concerned dinucleotides. Dinucleotide abundance was also analyzed in members of *Puccinia* genus revealing TpA dinucleotides as underrepresented and TpG, TpC, GpA, CpA as highly preferred dinucleotides (Roy and Staden, 2019).

CpG dinucleotides were observed to be highly under-represented in viral genomes, a pattern consistent with its host *H. sapiens* (Figure6.3 (a, b, c, d, e)). Thorough study revealed that CpG containing codons like GCG, UCG, CGC, CGA, CGG, CCG ,CGU and ACG containing CpG were noted to be under-represented

(RSCU < 0.60) in viruses and this study was found to be consistent with host *H. sapiens, C. musculinus, C. callosus and Z. brevicauda.*

Dinucleotides have a great influence on codon usage pattern and such feature of under-representation of CpG dinucleotide has been observed in various genomes of RNA viruses. It has been suggested that unmethylated CpG containing coding sequences of the viral pathogens have been recognized as pathogen signature's by host intracellular pattern recognition receptor Toll like receptor 9 (TLR9) and stimulates innate immune responses in host(human) (Pakula 2019). Interestingly, a significant share of the over-represented and under-represented codon pairs in viruses matched with host *H. sapiens* and this decrease in CpG dinucleotides did not allow stimulation of immune response in humans.

Table 6.7: Correspondence analysis of codon and amino acid usage in NW Arenavirus

Organism		GC	GC3	RCDI	CAI	Length	GRAVY	Aromaticity
JUNV	**Axis1 (RSCU)**	0.662**	0.691**	.940**	0.899**	-0.405**	-0.235*	-0.062
	Axis 2 (RSCU)	-.509**	-0.111	-280**	0.202*	-0.066	0.842**	0.983**
MACV	**Axis1 (RSCU)**	-0.921**	-0.866**	-.967**	-0.929**	0.490**	0.699**	-0.062
	Axis 2 (RSCU)	-0.130	0.164	-.069	0.261**	-0.151	0.489**	0,983**
GTOV	**Axis1 (RSCU)**	0.816**	0.922**	.934**	0.881**	-0.457**	-0.632**	-0.188
	Axis 2 (RSCU)	**0.411****	-0.134	-.057	-0.396**	-0.127	-0.691**	-0.956**
SABV	**Axis1 (RSCU)**	-0.452	-0.539	-.855**	-.952**	0.572	0.414	0.607
	Axis 2 (RSCU)	**.633***	0.203	0.045	-0.259	-0.178	-.779**	-.789**
CHPV	**Axis1 (RSCU)**	-0.452	-0.539	-.855**	-.952**	0.572	0.414	0.607
	Axis 2 (RSCU)	**.633***	0.203	0.045	-0.259	-0.178	-.779**	-.789**

**: statistically significant at P < 0.01; *: statistically significant at P < 0.05; RSCU: Relative synonymous Codon usage; Length: Length of protein coding sequences; GRAVY: Grand Average Hydropathicity Score; Aromaticity: Aromaticity of encoded proteins; CAI codon adaptation index; RCDI relative codon deoptimization index.

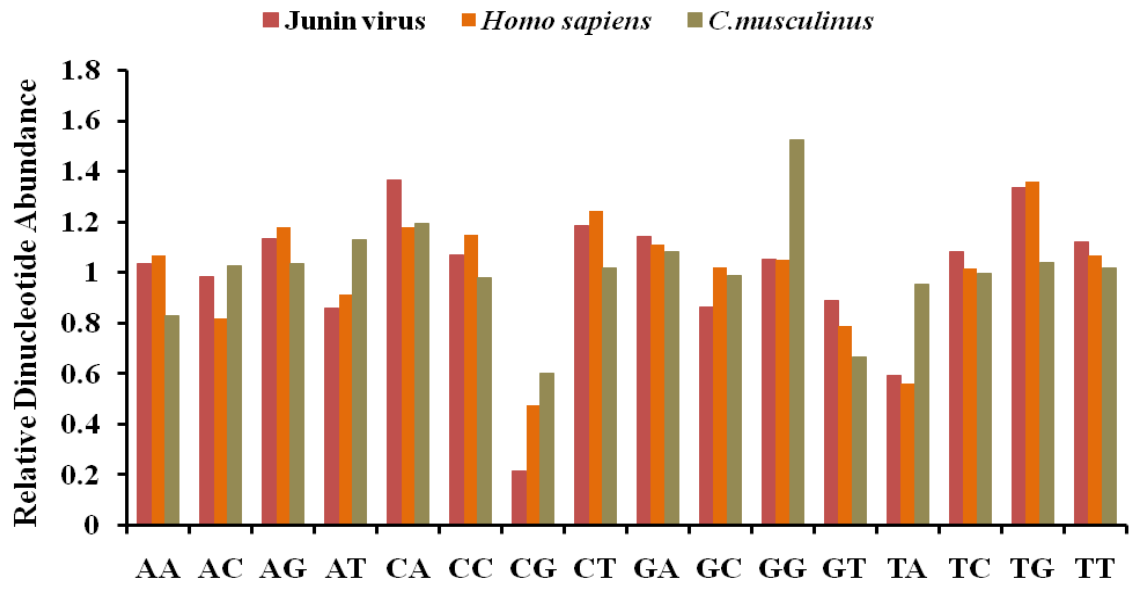


Figure 6.3a: Relative Dinucleotide analysis of JUNV

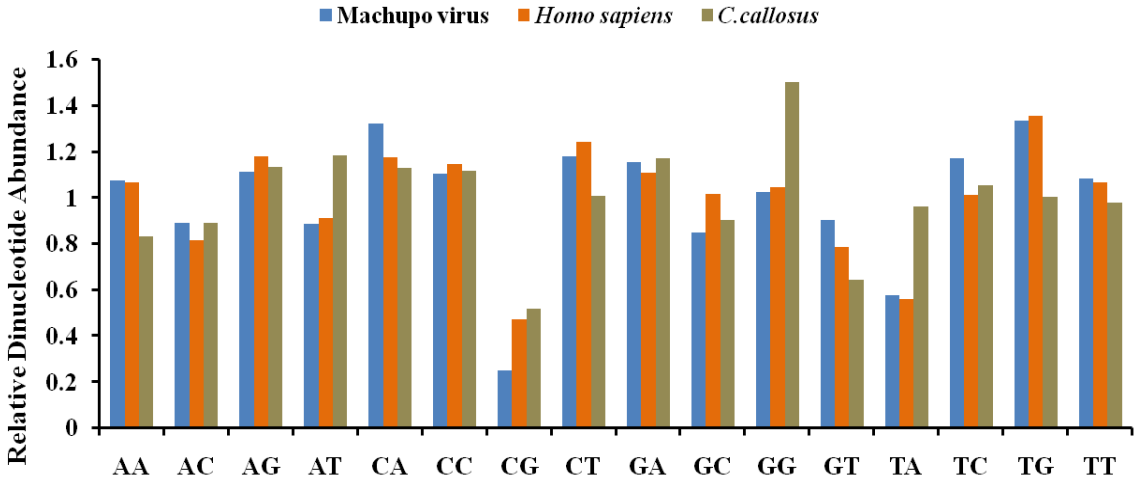


Figure 6.3b: Relative Dinucleotide analysis of MACV

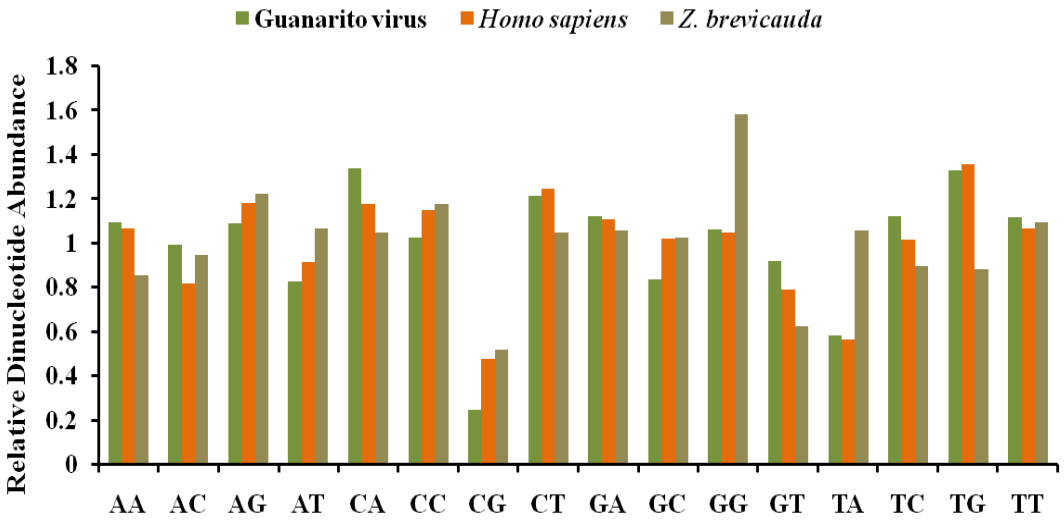


Figure 6.3c: Relative Dinucleotide analysis of GTOV

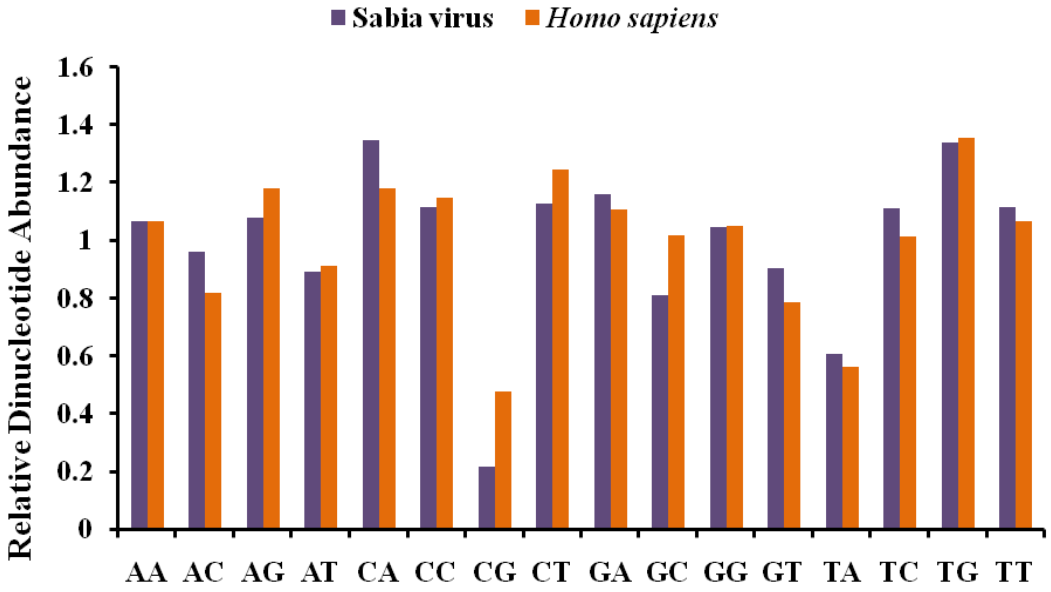


Figure 6.3d: Relative Dinucleotide analysis of SABV

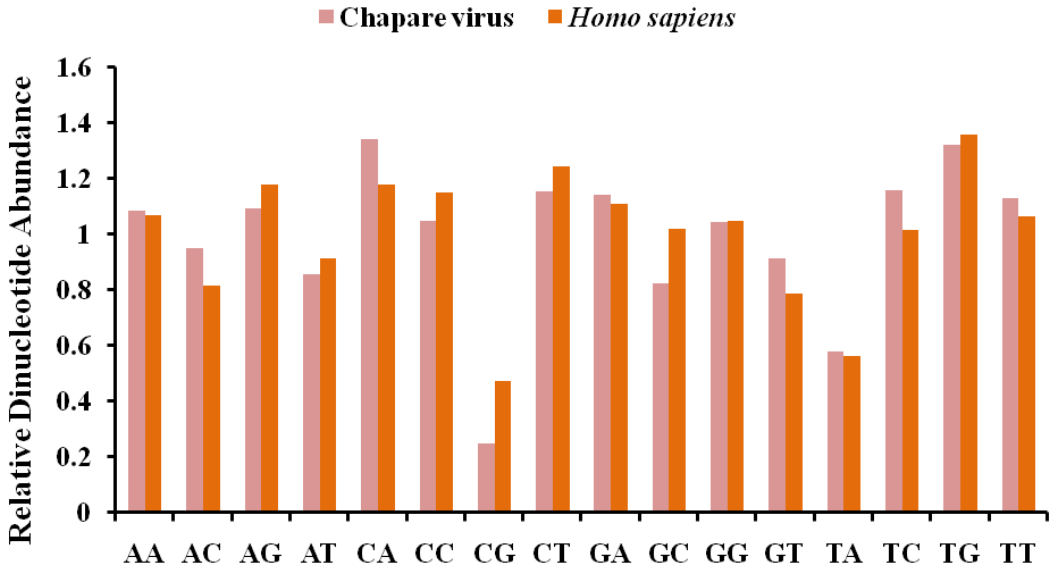


Figure 6.3e: Relative Dinucleotide analysis of CHPV

6.5 Estimation of relative synonymous codon pair usage (RSCPU) and codon pair score (CPS)

Extensive analysis of relative synonymous codon pair usage (RSCPU) of 3721 (61×61) codon pairs (excluding stop: stop and stop: sense codon pairs) revealed that 1310 codon pairs were over-represented in JUNV, whereas, 2411 were under-represented. Codon pair ACG-CAU coding for the amino acid pair Thr-His was noted

to be the most over-represented one with a codon pair score (CPS) of 3.27. On the contrary, codon pair AAU-UUU, encoding for the amino acid pair Asn-Phe, displayed the lowest CPS of -3.93 and was inferred to be the most under-represented in JUNV.

Interestingly, 738 out of 1310 over-represented codon pairs in JUNV (56.33%) matched with the over-represented codon pairs in *H. sapiens* and 45.4% matched with *C. musculinus* one of the natural reservoir of JUNV. Similar trend was also evident among under-represented codon pairs where 1394 out of 2411 under-represented viral codon pairs matched (57.8%) with that of the under-represented codon pairs of the human genome and 55% similarity with under-represented codon pairs of *C. musculinus.*

This analysis of codon pair scores of host shows that virus has more similarity with human genome so it can adapt the human machinery more efficiently causing infection among humans. Investigation of RSCPU analysis was also performed in *Puccinia* genus showing CGG-ACG coding for Arg-Thr as most preferred codon pair (Roy and Staden, 2019).

Similar analysis in MACV revealed that 1334 codon pairs were over-represented whereas, 1368 were under-represented. Codon pair CGC-UAC coding for the amino acid pair Arg-Tyr was noted to be the most over-represented one with a codon pair score (CPS) of 4.08. On the contrary, codon pair GGU-AAG, encoding for the amino acid pair Gly-Lys, displayed the lowest CPS of -3.7 and was inferred to be the most under-represented in MACV.

Interestingly, 926 out of 1334 over-represented codon pairs in MACV (69.4%) matched with the over-represented codon pairs in *H. sapiens* and 43.9% matched with natural reservoir *C. callosus*. Similar trend was also evident among under-represented codon pairs where 749 out of 1368 under-represented viral codon pairs matched (54.75%) with that of the under-represented codon pairs of the human genome and 53.3% matched with natural reservoir *C. callosus*. Recently, in 2021 similar investigation was performed in SARS-CoV2 predicting CGG-CGG coding for Arg-Arg codon pair as most preferred and AAU-AUA coding for Asn-Ile as under-represented codon pair with score of -4.56 (Roy *et al*., 2021).

Furthermore, analysis revealed that 1317 codon pairs were over-represented in GTOV, whereas, 1415 were under-represented. Codon pair ACG-CGU coding for the amino acid pair Thr-Arg was noted to be the most over-represented one with a

codon pair score (CPS) of 3.3. On the contrary, codon pair UUC-GAG, encoding for the amino acid pair Phe-Glu, displayed the lowest CPS of -3.83 and was inferred to be the most under-represented in GTOV.

Interestingly, 856 out of 1317 over-represented codon pairs in GTOV (64.9%) matched with the over-represented codon pairs in *H. sapiens* and 44.5% matched with natural reservoir *Z. bevicauda*. Similar trend was also evident among under-represented codon pairs where 775 out of 1415 under-represented viral codon pairs matched (54.7%) with that of the under-represented codon pairs of the human genome and 54.1% matched with natural reservoir *Z. bevicauda* showing high similarity with humans.

Similarly, in SABV 1237 out of total 3721 codon pairs (excluding stop: stop and stop: sense codon pairs) were found to be over-represented and 519 were under-represented. GCG-ACC codon pair coding for Alanine-Threonine was utmost over-represented with a CPS of 3.9. On the contrary, codon pair ACA-AAG coding for Threonine-Lysine have low codon pair score of -1.62 and was defined as the utmost under-represented in SABV. Interestingly, in SABV where 687 out of 1237 over-represented codon pairs (55.5%) matched with the over-represented codon pairs in *H. sapiens.* Similar trend was also evident among under-represented codon pairs where 250 out of 519 under-represented viral codon pairs matched (48.16%) with that of the under-represented codon pairs of the human genome.

Similarly, thorough study of RSCPU values of CHPV explained that 1249 out of 3721were found to be over-represented, 533 were under-represented. CGG-CCC codon pair coding for Arginine-Proline was utmost over-represented having codon pair score of 4.19. Although, codon pair UUC-GAG, encoding for Phenylalanine-Glutamate pair, displayed the lowest CPS of -3.83 and was examined as utmost under-represented in CHPV. Interestingly, in CHPV 708 out of 1249 (56.6%) matched with the over-represented codon pairs in *H. sapiens.* Similar trend was also evident among under-represented viral codon pairs as 254 out of 533 (47.65%) matched with that of the under-represented codon pairs of the human genome.

Dinucleotide pattern NNU-GNN (representing UpG dinucleotide at the codon pair junction) was established to be most frequent (10.8% in JUNV, 11.09% in MACV and 11.8 % in GTOV, 10.6% in SABV and 11.04% in CHPV) among the

over-represented codon pairs. Similar analysis was performed for Equine influenza virus predicting UpG and CpA as overrepresented patterns (Kumar *et al.*, 2016).

Also, methodical inspection of dinucleotide bias at the codon pair interface (cP3-cA1) depicted that the dinucleotide patterns NNU-GNN, NNC-ANN, NNC-UNN, representing the dinucleotides UpG, CpA, CpU respectively, were prevalent at the codon-codon junctions in JUNV, MACV, GTOV, SABV and CHPV (Figure 6.4 (a, b, c, d, e)). Interestingly, dinucleotide patterns NNU-GNN, NNC-ANN, NNC-UNN were also noted to be predominant among the over-represented codon pairs in *H. sapiens.* Codon Pair Score (CPS) data revealed more similarity of virus coding pairs with *H. sapiens*, revealing efficient adaptation of virus in humans. Similarly, analysis was performed in 2021 on SARS-CoV-2 predicting patterns NNU-GNN, NNC-ANN, NNC-UNN and NNA-CNN as preferred and UpG was predicted to be most occurred dinucleotide pattern with 11.24% (Roy *et al.*, 2021).

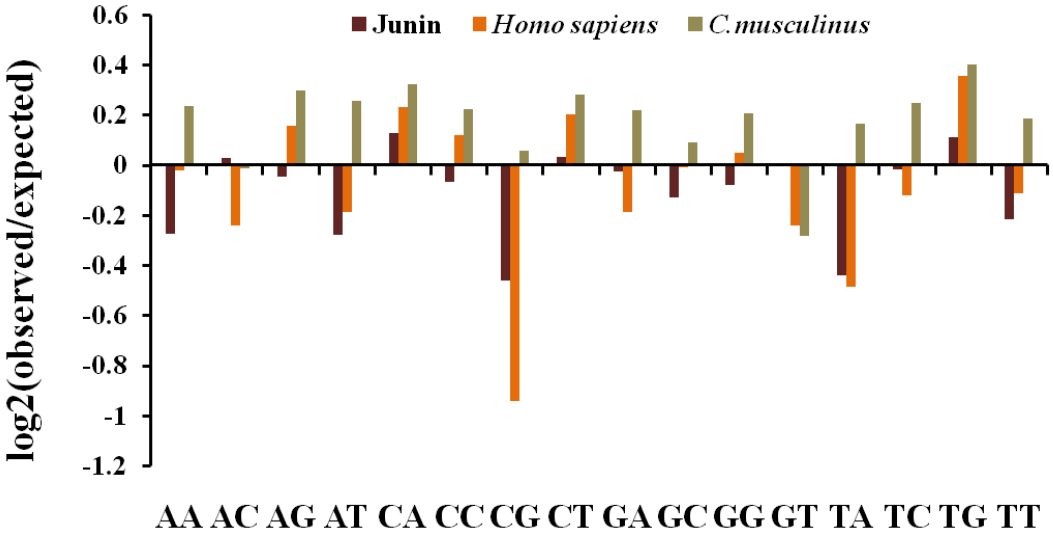


Figure 6.4a: CPS (Codon pairing Score) results of JUNV

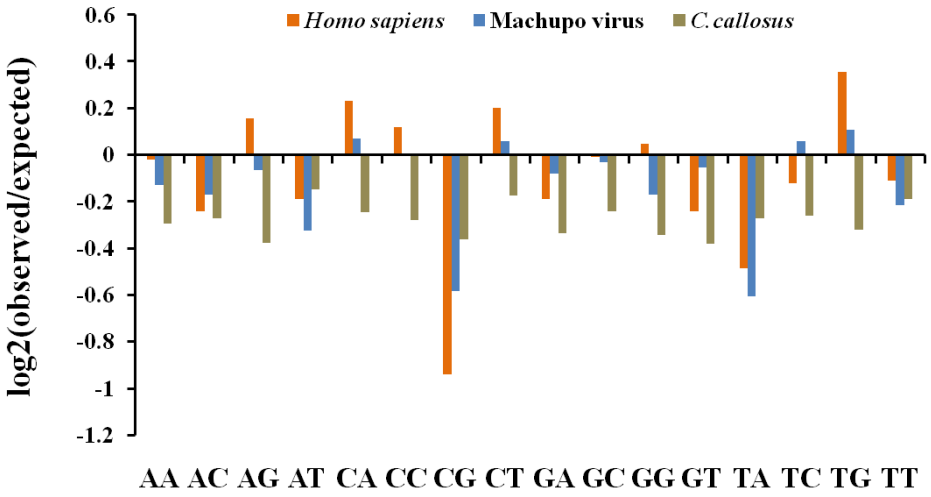


Figure 6.4b: CPS (Codon pairing Score) results of MACV

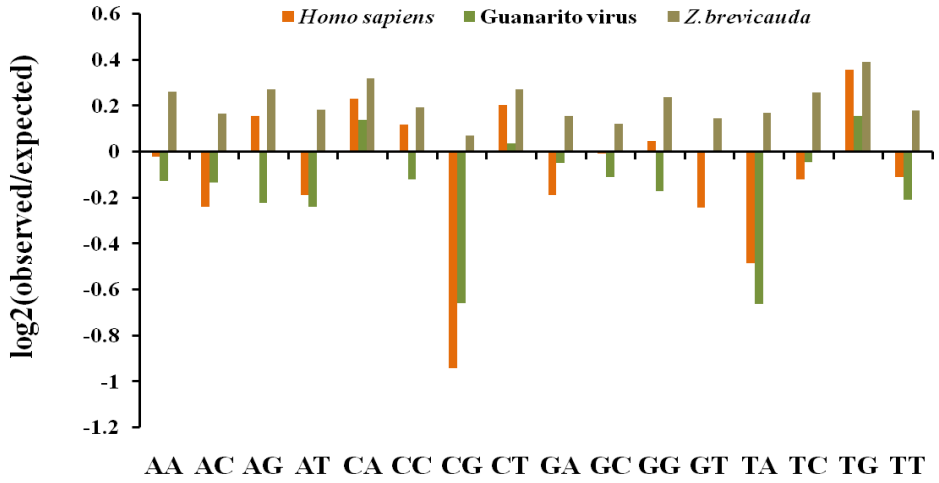


Figure 6.4c: CPS (Codon pairing Score) results of GTOV

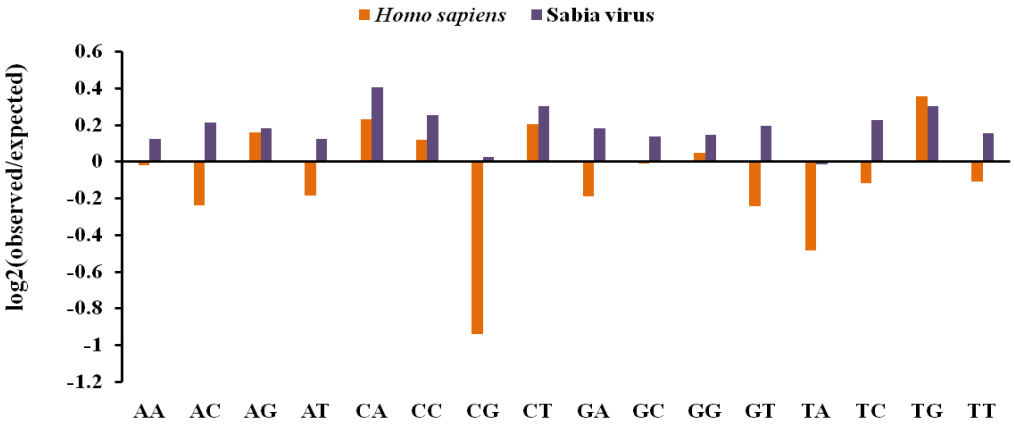


Figure 6.4d: CPS (Codon pairing Score) results of SABV

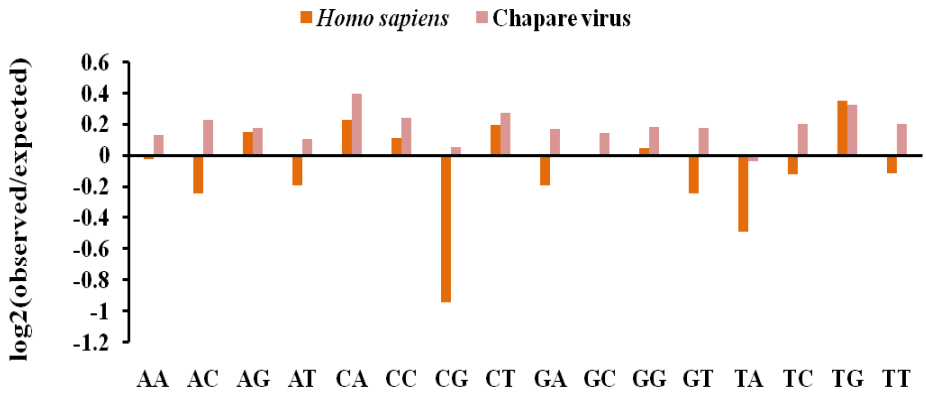


Figure 6.4e: CPS (Codon pairing Score) results of CHPV

X-axis showing dinucleotides and legends on right showing name of virus, host: human.

6.6 Host-adaptation

6.6.1 Investigating the patterns of NW Arenaviral adaptation in human host

6.6.1.1 Antagonistic codon usage patterns of Arenavirus towards human host

Detailed analysis of RSCU data of overrepresented codons (for each amino acid) in NW Arenavirus and *H. sapiens* disclose a well defined trend of antagonism between the viral and hosts codon usage patterns (Table 6.1, 6.2, 6.3, 6.4 and 6.5). 27 out of 29 most preferred codons in JUNV were found to display antagonism with *H. sapiens*, whereas, codon AGA (coding for Arg), AGC (coding for Ser) two codons display coincidence. 24 out of 29 most preferred codons in *C. musculinus* show antagonism with JUNV and five codons ATT (Ile), CAA (Gln), CCA (Pro), AAA (Lys), GAA (Glu) show coincidence 24 out of 29 most preferred codons in MACV were found to exhibit antagonism with *H. sapiens*, whereas, codons GAC (coding for Asp), CAG (coding for Gln), GTG (coding for val), ATC (coding for Ile), TTC (coding for Phe) codons display coincidence.

MACV with natural reservoir *C. callosus* display 19 out of 26 codons as antagonism and codons TCA (Ser), CCA (Pro), ACA (Thr), TAT (Tyr), AAT (Asn), AAA (Lys), GGA (Gly) examined to show coincidence. 26 out of 28 most preferred codons in GTOV were found to display antagonism with *H. sapiens*, whereas, codon AGA (coding for Arg), TAC (coding for Tyr) display coincidence. In SARS-CoV-2 also seventeen out of eighteen preferred codons show antagonism with host *H. sapiens* and only AGA show coincidence (Roy *et al*., 2021).

GTOV also displayed antagonism with natural reservoir *Z. brevicauda* as 18 out of 26 show antagonism and codons ATT (Ile), CCA (Pro), ACA (Thr), CAA (Gln), AAT (Asn), AAA (Lys), GAA (Glu), TGT (Cys) show coincidence28 out of 29

highly preferred codons in SABV and CHPV were examined showing antagonism with codon patterns of *H. sapiens*, whereas AGA (coding for Arg) display coincidence in both species.

Codons showing coincident with host have efficient translation of that particular amino acid as compare to antagonistic codons which can be folded properly but efficiency of translation reduces. It has been suggested that coincident codon usage patterns between a virus and its host show increase in translational efficiency, despite the fact that antagonism appeared to increase the folding of viral proteins whereas the efficacy of translation might be reduced. Hu and colleagues similarly predicted codon patterns of hepatitis C virus as both coincident and antagonistic with host (Hu *et al.*, 2011).

6.6.2 CAI-RCDI-SiD

Codon Adaptation Index (CAI) analyzes the degree of adaptation of viral genes in host cell. CAI values range between 0-1 and higher values predict better adaptation of virus in host cellular environment (Roy *et al*., 2015).

The average CAI value of JUNV with respect to *H. sapiens* was 0.76 ± 0.02 and with *C. musculinus* was 0.51±0.012; MACV shows 0.76±0.002 and 0.611±0.02 CAI value with *H. sapiens* and *C. callosus*; GTOV shows 0.74±0.02 and 0.58±0.017 CAI value with *H. sapiens* and *Z. brevicauda,* SABV was 0.76±0.03 and CHPV was 0.75±0.02. These results predict high level of adaptation of Arenavirus in *H. sapiens* as average values of CAI was significantly higher ($P < 0.01$) with humans as compare to natural reservoir (mouse).

CAI value for GP, NP, Z and L proteins of JUNV was 0.76±0.004, 0.75±0.004, 0.806±0.006, 0.76±0.002. In MACV CAI at protein level was calculated as 0.76±0.005 for GP, 0.74±0.005 for NP, 0.74±0.002 for L and 0.80±0.01 for Z. CAI values for GTOV GP, NP, L and Zs were 0.74±0.005, 0.72±0.005, 0.73±0.003 and 0.79±0.012. CAI values for SABV GP, NP, L and Zs were 0.75±0.00, 0.73±0.0003, 0.73±0.0003 and 0.8±0.000. CAI values for CHPV were found to be 0.75±0.00 for GP, 0.73±0.0005 for NP, 0.73±0.0003 for L and 0.78±0.000 for Z. Results shows Z-protein has high level of CAI value as compare to other proteins in JUNV, MACV, GTOV, SABV and CHPV. Similarly, CAI was computed for SARS-CoV-2 as 0.701±0.004 with host *H. sapiens* and it was found to be lower than SARS-CoV and MERS-CoV showing lower adaptation of SARS-CoV2 in humans (Roy *et al*., 2021).

The Relative Codon Deoptimization Index (RCDI) values help to evaluate rate of translation of viral genes in host (Yi *et al.*, 2018). Lower the RCDI value more is the adaptation of virus inside host cellular environment. The average RCDI value of JUNV was 1.307±0.12 and 2.24±0.5 with *H. sapiens, C. musculinus;* MACV was 1.32±0.15 and 1.62±0.118 with *H. sapiens and C. callosus*; GTOV was 1.34±0.08 and 2.24±0.5 with *H. sapiens, Z. brevicauda*; the average RCDI value of SABV was 1.40±0.04 and CHPV was 1.41±0.23. The lower RCDI values in *H. sapiens* as compare to mouse shows more adaptation of viruses in *H. sapiens*. Similarly, computation of RCDI value in Nipah virus predicted high level of adaptation to African green monkey (Khandia *et al.,* 2019).

RCDI value for GP, NP, Z and L proteins of JUNV was 1.23±0.018, 1.25±0.013, 1.57±0.039, 1.28±0.007. In MACV RCDI at protein level was calculated as 1.23±0.02 for GP, 1.24 ±0.023 for NP, 1.28±0.012 for L and 1.61±0.05 for Z. CAI values for GTOV GP, NP, L and Zs were 1.22±0.005, 1.26±0.005, 1.30±0.003 and 1.60±0.002. RCDI values for SABV GP, NP, L and Zs were 1.27±0.0013, 1.31±0.0014, 1.34±0.0013 and 1.49±0.0012. RCDI values for CHPV were found to be 1.27±0.0014 for GP, 1.28±0.00135 for NP, 0.131±0.0013 for L and 1.79±0.0014 for Z. Results shows GP has low level of RCDI value as compare to other proteins in JUNV, MACV, GTOV, SABV and CHPV.

A Similarity index (SiD) analysis was performed to determine shaping of virus codons in host machinery. Value of SiD ranged between 0 and 1 (Khandia *et al.,* 2019). SiD values were found to be 0.065±0.5 (JUNV), 0.064±0.5 (MACV) and 0.07±0.5 (GTOV) 0.072 (SABV) and 0.073 *(*CHPV*)* with humans and 0.12±0.012 (JUNV), 0.10±0.01 (MACV) and 0.122±0.01 (GTOV) with *C. musculinus, C. callosus* and *Z. brevicauda.* These values predict natural reservoirs (*C. musculinus, C. callosus and Z. brevicauda)* had greater impact on the virus codon bias. Similarly, SiD value of Nipah virus was found to be highest with host dog (0.139) and lowest with African green monkey (0.054) (Khandia *et al.,* 2019).

SiD value for GP, NP, Z and L proteins of JUNV was 0.061±0.008, 0.060±0.003, 0.102±0.005 and 0.073±0.003. In MACV SiD at protein level was calculated as 0.062±0.002 for GP, 0.063±0.002 for NP, 0.098±0.001 for Z and 0.073±0.005 for L. In GTOV SiD at protein level was calculated as 0.076±0.02 for GP, 0.079±0.003 for NP, 0.104±0.002 for Z and 0.085±0.005 for L. SiD values for

SABV GP, NP, L and Zs were 0.082±0.003, 0.076±0.004, 0.083±0.003 and 0.133±0.002. SiD values for CHPV were found to be 0.080±0.004 for GP, 0.078±0.0035 for NP, 0.143±0.003 for Z and 0.081±0.004 for L. Results shows Z-protein has high level of SiD value as compare to other proteins in JUNV, MACV, GTOV, SABV and CHPV.

Various parameters such as CAI of viral genes, RCDI and SiD estimated the adaptation of viruses among various hosts. Data proposed the high level of adaptation of viruses with humans as compare to natural reservoir (*C. musculinus, C. callosus, Z. brevicauda*). Results shows GP and Z-protein has high level of adaptation as compare to other proteins in JUNV, MACV, GTOV, SABV and CHPV. The transmissibility of the virus is a complex interplay of several crucial factors like binding efficacy of viral with the host cellular receptors, host immune responses and apt control and proper, drastic prevention of rapid spread of infection (Emrani *et al*., 2020).

6.7 Most Preferred Codons in New World Arenavirus and Human Isoacceptor tRNAs

Identification of the highly preferred codons (for each amino acid) in JUNV, MACV, GTOV, SABV, CHPV and the most abundant isoacceptor tRNAs in human cells revealed that 9 out of the 18 most preferred codons in JUNV, namely, UCC, UGG, UGU, AAC, GCU, UCU, UUU, UUC, AAU (coding for the amino acids Gly, Pro, Thr, Val, Ser, Arg, Lys, Glu and Ile respectively). Eight out of eighteen in MACV, namely, AGC, UGG, UGU, CAC, AAG, GAA, GUC, CUG (coding for the amino acids Ala, Pro, Thr, Val, Leu, Phe, Asp and Gln, respectively). Eight out of eighteen in GTOV , namely, UGG, UGU, AAC, UCU, CAA, UUU and AAU (coding for the amino acids Pro, Thr, Val, Arg, Leu, Lys, Ile and Tyr, respectively). Nine codons out of eighteen highly favoured codons in SABV virus, namely, UCC, UGG, UGU, AAC, UCU, CAA, UUU, UUC, AAU (code for Gly, Pro, Thr, Val, Arg, Leu, Lys, Glu and Ile amino acids). Ten out of eighteen in CHPV virus, namely, AGC,UCC, UGG, UGU, CAC, UCU, AAG, UUU,UUC and AAU (code for Ala, Gly, Pro, Thr, Val, Arg, Leu, Lys, Glu and Ile amino acids) optimally matched with the respective most abundant isoacceptor tRNAs in human hosts (Table 6.8, 6.9, 6.10, 6.11, 6.12). Most of the preferred codons in JUNV, MACV, GTOV, SABV and

CHPV coding sequences use suboptimal isoacceptor tRNAs from human cells (Table 6.8, 6.9, 6.10, 6.11, 6.12).

A similar pattern has also been reported for Nipah virus to recognize the usage of suboptimal tRNA isotype (Khandia *et al*., 2019). It has been suggested that during the initial phase of an infection the usage of suboptimal isoacceptor host tRNAs might lead to slow but precise translation, which yields the synthesis of accurate and properly folded viral.The information gathered in present study with codon bias of JUNV, MACV, GTOV, SABV and CHPV may be used to design vaccine in human vaccination programs.

Table 6.8: Analysis of highly preferred codons in JUNV and iso-acceptor tRNAs in *H. sapiens*

Amino acids	**Most preferred codons in JUNV**	**tRNA isotypes in *H. sapiens***
Ala	GCA	AGC (22), GGC (0), CGC (4), UGC (8)
Gly	GGT	ACC (0), GCC (14), CCC (5), **UCC (9)**
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTT	**AAC (9),** GAC (0), CAC (11), UAC (5)
Ser	AGC	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), **GCU (8)**
Arg	AGA	ACG (7), GCG (0), CCG (4), UCG (6), CCU (5), **UCU (6**)
Leu	TTG	AAG (9), GAG (0), CAG (9), UAG (3**),**CAA (6**),** UAA (4)
Phe	TTT	AAA (0), GAA (10)
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), **UUU (12)**
Asp	GAT	AUC (0)**,** GUC (13)
Glu	GAA	CUC (8), **UUC (7)**
His	CAT	AUG (0), GUG (10)
Gln	CAA	CUG (13), UUG (6)
Ile	ATT	**AAU (14**), GAU (3), UAU (5)
Tyr	TAT	AUA (0), GUA (13)
Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *Homo sapiens* matching the most preferred codons of JUNV are marked in bold.

Table 6.9: Analysis of highly preferred codons in MACV and iso-acceptor tRNAs in *H. sapiens*

Amino acids	Most preferred codons in MACV	tRNA isotypes in *H. sapiens*
Ala	GCT	**AGC (22**), GGC (0), CGC (4), UGC (8)
Gly	GGA	ACC (0), GCC (14), CCC (5), UCC (9)
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTG	AAC (9)**,** GAC (0), **CAC (11**), UAC (5)
Ser	TCA	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), GCU (8)
Arg	AGG	ACG (7), GCG (0), CCG (4), UCG (6), CCU (5), UCU (6)
Leu	CTT	**AAG (9),** GAG (0), CAG (9), UAG (3**),**CAA (6**),** UAA (4)
Phe	TTC	AAA (0), **GAA (10)**
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), UUU (12)
Asp	GAC	AUC (0)**, GUC (13)**
Glu	GAA	CUC (8), UUC (7)
His	CAT	AUG (0), GUG (10)
Gln	CAG	**CUG (13**), UUG (6)
Ile	ATC	AAU (14), GAU (3), UAU (5)
Tyr	TAT	AUA (0), GUA (13)
Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *H. sapiens* matching the most preferred codons of MACV are marked in bold.

Table 6.10: Analysis of highly preferred codons in GTOV and iso-acceptor tRNAs in *H. sapiens*

Amino acids	Most preferred codons in GTOV	tRNA isotypes in *H. sapiens*
Ala	GCA	AGC (22), GGC (0), CGC (4), UGC (8)
Gly	GGT	ACC (0), GCC (14), CCC (5), UCC (9)
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTT	**AAC (9),** GAC (0), CAC (11), UAC (5)
Ser	TCA	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), GCU (8)
Arg	AGA	ACG (7), GCG (0), CCG (4),

		UCG (6), CCU (5), **UCU (6)**
Leu	TTG	AAG (9), GAG (0), CAG (9), UAG (3**),CAA (6),** UAA (4)
Phe	TTT	AAA (0), GAA (10)
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), **UUU (12)**
Asp	GAT	AUC (0)**,** GUC (13)
Glu	GAA	CUC (8), UUC (7)
His	CAT	AUG (0), GUG (10)
Gln	CAA	CUG (13), UUG (6)
Ile	ATT	**AAU (14**), GAU (3), UAU (5)
Tyr	TAC	AUA (0), **GUA (13)**
Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *Homo sapiens* matching the most preferred codons of GTOV are marked in bold.

Table 6.11: Analysis of highly preferred codons in SABV and iso-acceptor tRNAs in *H. sapiens*

Amino acids	**Most preferred codons in SABV**	**tRNA isotypes in *H. sapiens***
Ala	GCA	AGC (22), GGC (0), CGC (4), UGC (8)
Gly	GGT	ACC (0), GCC (14), CCC (5), **UCC (9)**
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTT	**AAC (9),** GAC (0), CAC (11), UAC (5)
Ser	TCA	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), GCU (8)
Arg	AGA	ACG (7), GCG (0), CCG (4), UCG (6), CCU (5), **UCU (6**)
Leu	TTG	AAG (9), GAG (0), CAG (9), UAG (3**),CAA (6),** UAA (4)
Phe	TTT	AAA (0), GAA (10)
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), **UUU (12)**
Asp	GAT	AUC (0)**,** GUC (13)
Glu	GAA	CUC (8), **UUC (7)**
His	CAT	AUG (0), GUG (10)
Gln	CAA	CUG (13), UUG (6)
Ile	ATT	**AAU (14**), GAU (3), UAU (5)
Tyr	TAT	AUA (0), GUA (13)
Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *H. sapiens* matching the most preferred codons of SABV virus are marked in bold.

Table 6.12: Analysis of highly preferred codons in CHPV and iso-acceptor tRNAs in *H. sapiens*

Amino acids	Most preferred codons in CHPV	tRNA isotypes in *H. Sapiens*
Ala	GCT	**AGC (22**), GGC (0), CGC (4), UGC (8)
Gly	GGT	ACC (0), GCC (14), CCC (5), **UCC (9)**
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTG	AAC (9)**,** GAC (0), **CAC (11**), UAC (5)
Ser	TCA	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), GCU (8)
Arg	AGA	ACG (7), GCG (0), CCG (4), UCG (6), CCU (5), **UCU (6)**
Leu	TTG	**AAG (9),** GAG (0), CAG (9), UAG (3**),**CAA (6**),** UAA (4)
Phe	TTT	AAA (0), GAA (10)
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), **UUU (12)**
Asp	GAT	AUC (0**),** GUC (13)
Glu	GAA	CUC (8), **UUC (7)**
His	CAT	AUG (0), GUG (10)
Gln	CAA	CUG (13), UUG (6)
Ile	ATT	**AAU (14),** GAU (3), UAU (5)
Tyr	TAT	AUA (0), GUA (13)
Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *H. sapiens* matching the most preferred codons of CHPV are marked in bold.

6.8 Phylogenetic analysis of Arenaviruses

Phylogenetic tree was generated by maximum likelihood (ML) method separately for all the proteins (GPC, NP, Z and L) of JUNV, MACV, GTOV, SABV, and CHPV of NW Arenaviruses with a motif to execute a comparative evolutionary analysis of their proteins. Phylogenetic analysis of GPC protein of JUNV has been shown in Figure 6.5a. It was evident that different strains of JUNV were clustered together in distinct clades showing *JNM-6682* as the out-group. Phylogenetic analysis of L has shown *P2290 Ledesma* as out-group as shown in Figure 6.5b; Z-protein phylogenetic analysis has predicted *P35032* strain as out-group as shown in Figure 6.5c. Further, tree of N-protein as shown in Figure 6.5d predicted *JNM-7354* strain as out-group.

Phylogenetic tree of the pathogenic member of NW Arenaviruses i.e. MACV has also been displayed in Figure 6.6a, b, c and d. It was perceptible that the strains of this virus were separated in different groups. The analysis has shown *Mallele* for NP, *9530537* for Z-protein, *Mallele* for L and *SPB201004275* for GPC protein as out-group strains. Similarly, phylogenetic analysis of Zika virus was performed by using Maximum Likelihood based method in MEGA software predicting the grouping of Zika virus isolates into three different genotypes as: Asian, East African and West African (Butt *et al.*,2016).

Phylogenetic trees for the pathogenic members of NW Arenavirus served as the platform for evolutionary investigations. Evolutionary rates of the protein components of the selected members of NW Arenavirus in the concerned phylogenetic trees of GTOV have analyzed as shown in Figure 6.7a, b, c, d and estimated *260456* for GPC, NP, L and *CVH-950801* for Z as out-group member. It was evident from comparative evolutionary study of all proteins (Table 6.13) that evolutionary rates of the GPC in pathogenic members of NW Arenavirus differed significantly from the other proteins.

Phylogenetic analysis was also performed on all the available genomes of H7A9 avian influenza virus by maximum likelihood method using RaxML software predicting host tropism i.e. occurrence of virus from avian to humans due to presence of convergent evolution at branches (Xiang *et al.*,2018).

As dataset available for SABV and CHPV was not enough to perform phylogenetic analysis of protein sequences separately. Overall phylogenetic analysis considering all sequences of SABV and CHPV was performed showing NP as diverged from other protein sequences.

6.9 Evolutionary analysis of protein sequences and selection of target protein

Phylogenetic tree analysis of the JUNV has provided the information of strain *P2045* for GPC, *MC2* for NP, *P2290 Ledesma* for L and *P35032* for Z-protein to be the out-group member among all the strains of JUNV for all the proteins separately as shown in Table 6.13. Further, evolutionary dynamics study of out-group strain with all other strains of JUNV was computed as shown in Table 6.13 for all proteins separately using PAML software showing dN/dS of GPC to be lowest and results were found to be significant by employing statistical analysis. These results predicted GPC as highly conserved protein in JUNV and can be used as target for drug

designing against JUNV. Similar analysis was performed in genus *Bifidobacterium* to predict the variation in evolutionary rate of PHX and PLX genes. Results investigated low evolutionary rate in PHX as compare to PLX genes (Roy *et al.*, 2015).

Similarly, phylogenetic analysis of the MACV has shown *MARU22688* for GPC, *Mallele* for NP and L, *9530537* for Z-protein as out-group strains and evolutionary results (Table 6.14) showing average value of dN/dS (Non-synonymous/synonymous sites) of out-group strain with all other strains of MACV. Computed values have been predicted to be lowest for GPC as shown in Table 6.14 and results were found to be significant. These results predicted GPC as highly conserved proteins in MACV and can be used as target for drug designing against MACV.

Further, phylogenetic analysis of GTOV has shown strain *VHF-1750* for GPC, NP, L and *VHF-1986* for Z as out-group among all the strains and dN/dS computed was found to be lowest for GPC and NP with high level of significance as shown in Table 6.15. These results predicted NP and GPC as highly conserved proteins in GTOV and can be used as target for drug designing against virus. Similar analysis was also performed in H7N9 avian influenza virus by using codeml program of PAML software. Results predicted positive selection in three genes (NP, NS and PA) out of eight (Xiang *et al.*, 2018).

As dataset available for SABV and CHPV was small, similar analysis when performed on protein sequences showed identical value of dN/dS for all the protein sequences and results were found to be statistically insignificant.

Present observation of varying evolutionary features of the proteins (GPC, NP, Z, L) among all the viruses might be indicative of the adaptive strategy of the pathogens for a better adaptation inside a diverse range of host community. Because of the conserved forces in GPC and also its importance in communication with host receptor protein to cause infection modulate GPC as target protein for *in silico* drug designing against pathogenic Arenavirus (Damte *et al.* 2013; Amineni *et al.*, 2010). Further, GP1 which is the subunit of GPC that mainly interact with the host cell receptor was considered as final target for *in silico* based drug designing.

Table 6.13: Comparative display of evolutionary rates for four proteins of JUNV of NW Arenavirus

JUNV	dN/dS	p-value
GPC *(P2045)*	0.058±0.01	0.00011
NP(*MC2*)	0.1275±0.02	0.00053
L (*P2290Ledesma*)	0.135±0.02	0.0074
Z-protein(*P35032*)	0.091±0.01	0.00016

Table 6.14: Comparative display of evolutionary rates for four proteins of MACV of NW Arenavirus

MACV	dN/dS	Significance (p-value)
GPC (*MARU22688*)	0.04985±0.01	0.00063
NP(*Mallele*)	0.052596±0.01	0.00062
L (*Mallele*)	0.082246667±0.01	0.0019
Z-protein(*9530537*)	0.12614±0.02	0.0016

Table 6.15: Comparative display of evolutionary rates for four proteins of GTOV of NW Arenavirus

GTOV	dN/dS	p-value
GPC (*VHF-1750*)	0.089±0.01	0.0001
NP(*VHF-1750*)	0.079±0.01	0.0003
L (*VHF-1750*)	0.11372667±0.02	0.00063
Z-protein(*VHF-1986*)	0.4605±0.03	0.063

Results showing average value of dN/dS (Non-synonymous/synonymous sites) for GP, NP, L and Z proteins

6.10 Molecular Docking of ligands with GP1 protein of JUNV

Molecular Docking of all ligands downloaded from ZINC15 database (Ren *et al.* 2020) in PDBQT format was performed separately with active site of target structure of GP1 protein of JUNV (5W1K) (Clark *et al*., 2018) in Autodock Vina software (Trott & Olson, 2010). Active site was selected by defining grid box dimensions as centre_X=-37.414, centre_Y=-0.048, centre_Z=-85.385; size_x=126, size_y=126, size_z=126 in AutoDock Vina software (Trott & Olson, 2010). Binding energy score of each ligand docked with target structure was calculated showing five best scored ligands from FDA approved drug library (Table 6.16) and five best scored ligands from investigational drug library (Table 6.17); as both libraries were

downloaded from ZINC15 database (Ren *et al.* 2020). Docking was also performed with ligands as shown in Table 5.9 and binding energy was computed with each separately.

Further, interactions of ligands with target were visualized in Pymol visualization tool (Seeliger *et al*., 2010) (Figure 6.10). Five best ligands after screening of FDA-approved drug library with their binding energy score with target protein GP1 of JUNV are shown. Hydrophobic interactions shown in italics, Hydrogen bonds are marked with * and π-stacking marked with ^.

6.11 Molecular Docking of ligands with GP1 protein of MACV

3D structures of all ligands separately were docked with 3D structure of GP1 target protein of MACV (5W1M) (Clark *et al*., 2018). Docking was performed at active site of target protein which was selected by defining grid dimensions as center_X=75.663, center_Y=222.274, center_Z=221.976; size_x=126, size_y=104, size_z=126. Binding energy of each ligand docked with GP1 protein of MACV was calculated separately in AutoDock Vina software (Trott & Olson, 2010). Out of 2115 FDA-approved ligand structures 5 ligands (Table 6.19) show good docking scores with GP1 target protein of MACV and among 3754 investigational ligands best 5 ligands were selected for further analysis (Table 6.20). Docking was also performed with ligands as shown in Table 5.9 and binding energy was computed with each separately. Interaction between final selected ligand and target was visualized in Pymol visualization tool (Figure 6.11). *In silico* structure based drug designing was performed against Nipah virus by using iGEMDOCK software predicting bioisosteres of favipiravir as potential drugs against target glycoprotein (Shah *et al.*, 2018).

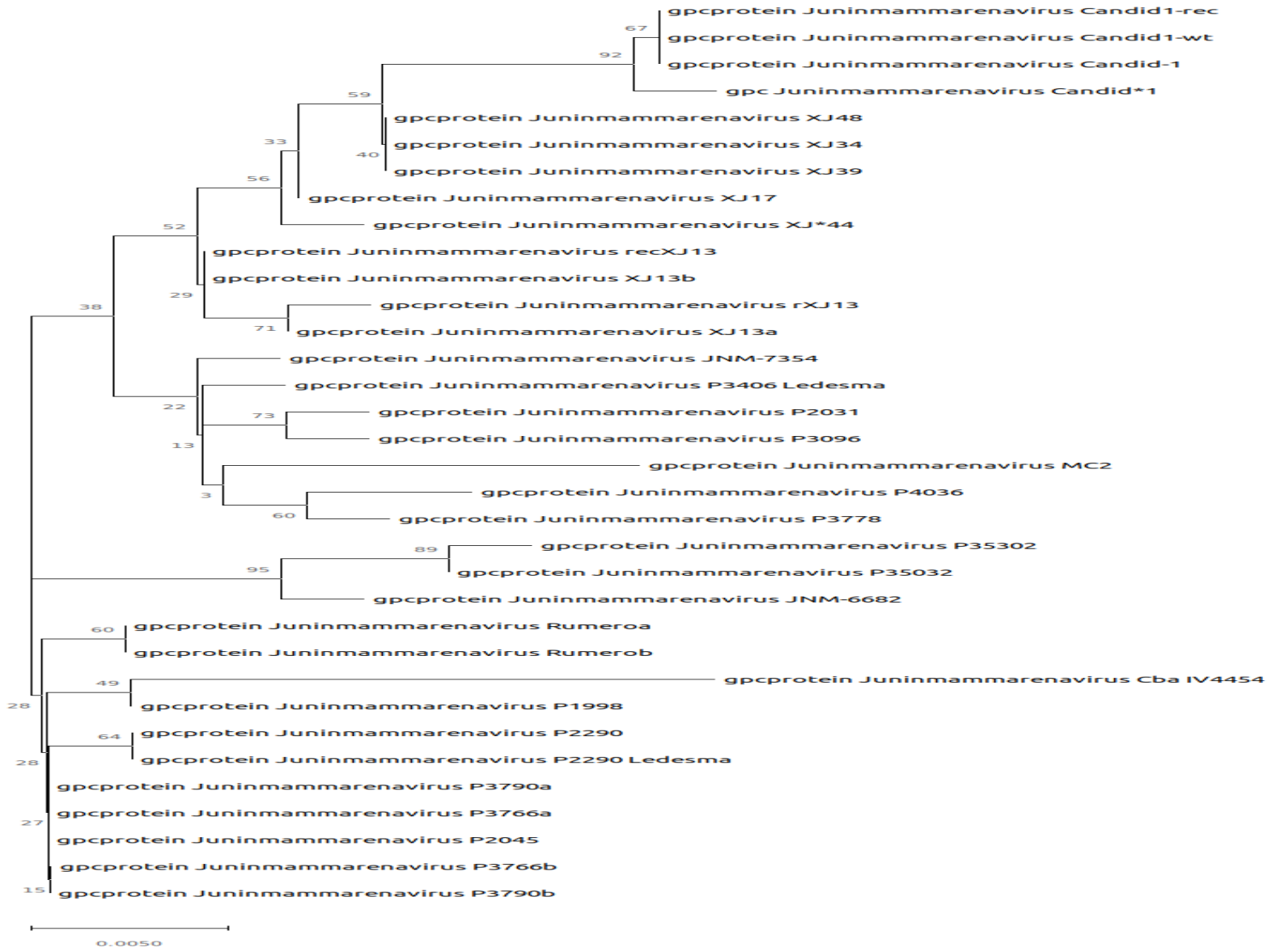


Figure 6.5a: Phylogenetic tree of GPC sequences of JUNV

(Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution).Bootstrap values are shown (1000 replicates).

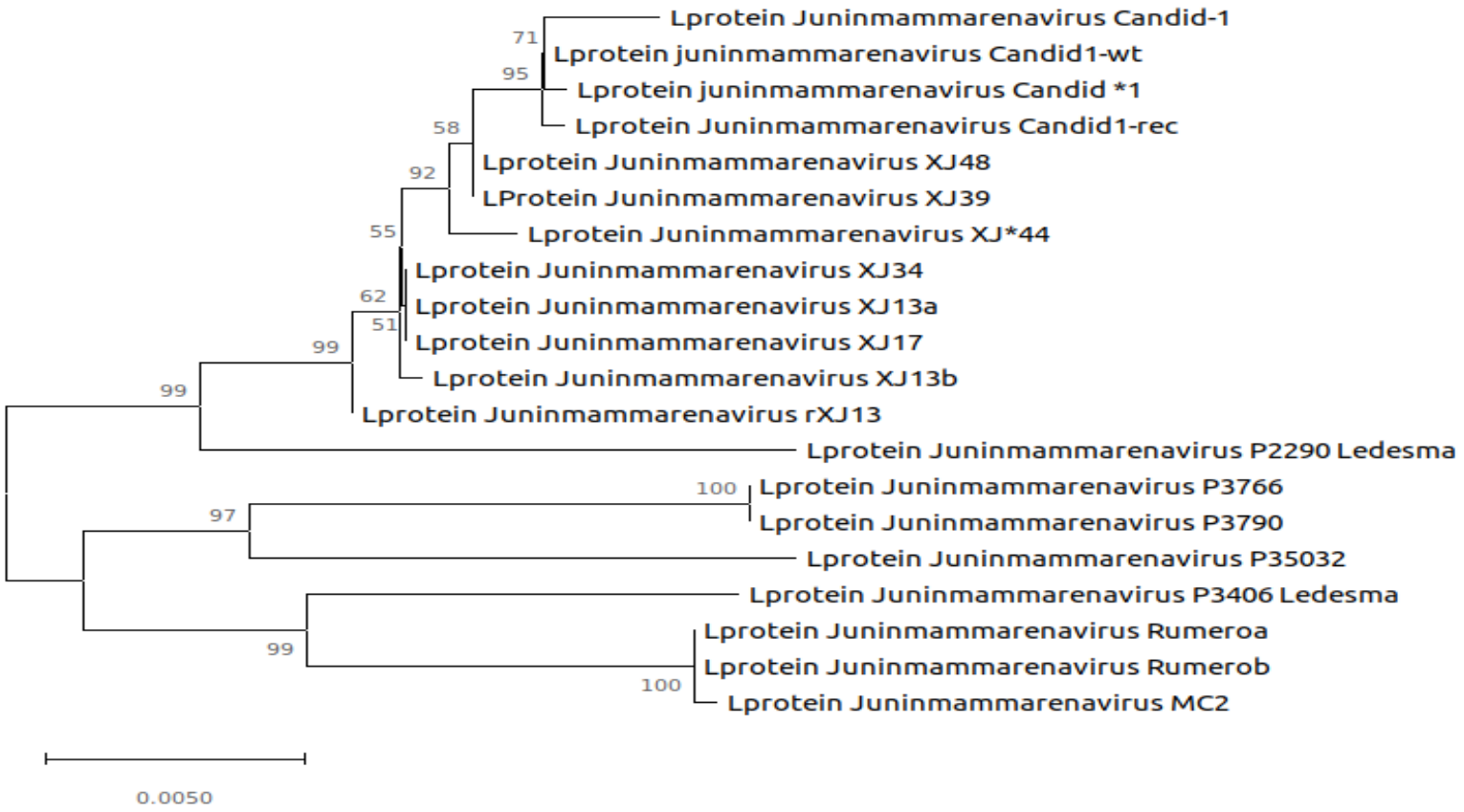


Figure 6.5b: Phylogenetic tree of L sequences of JUNV

(Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates). Legends: N=NP, gpc=glycoprotein, L=RNA polymerase, Z=Z

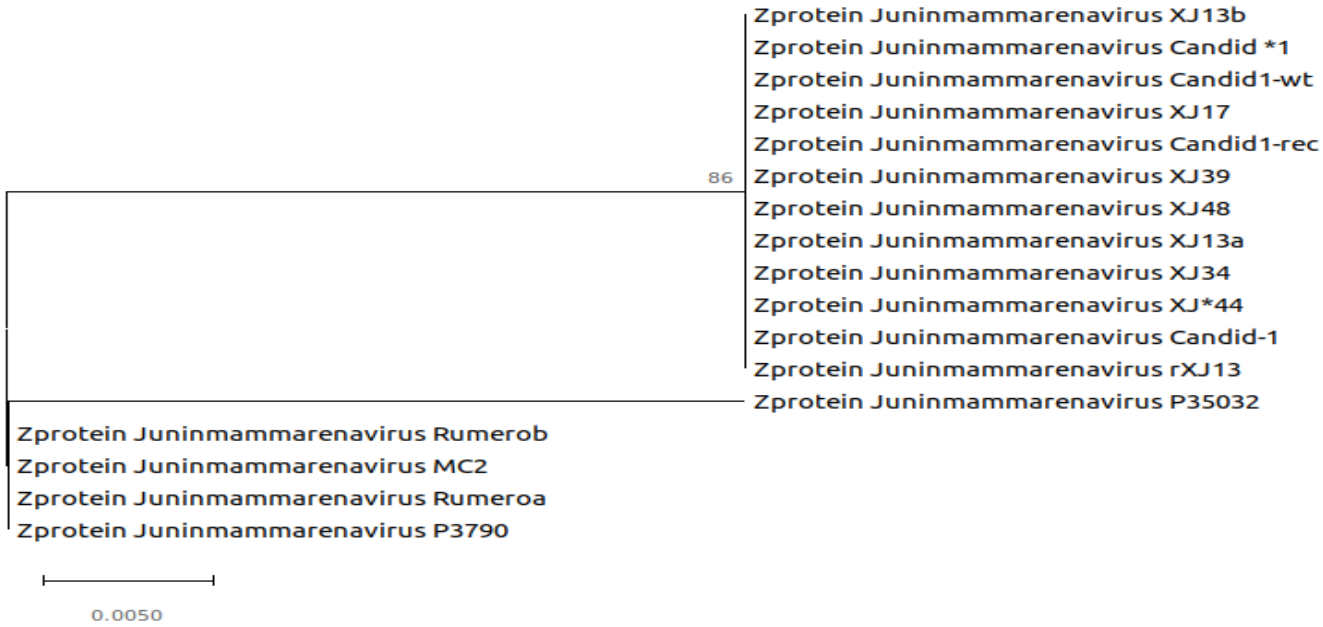


Figure 6.5c: Phylogenetic tree of Z sequences of JUNV

(Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

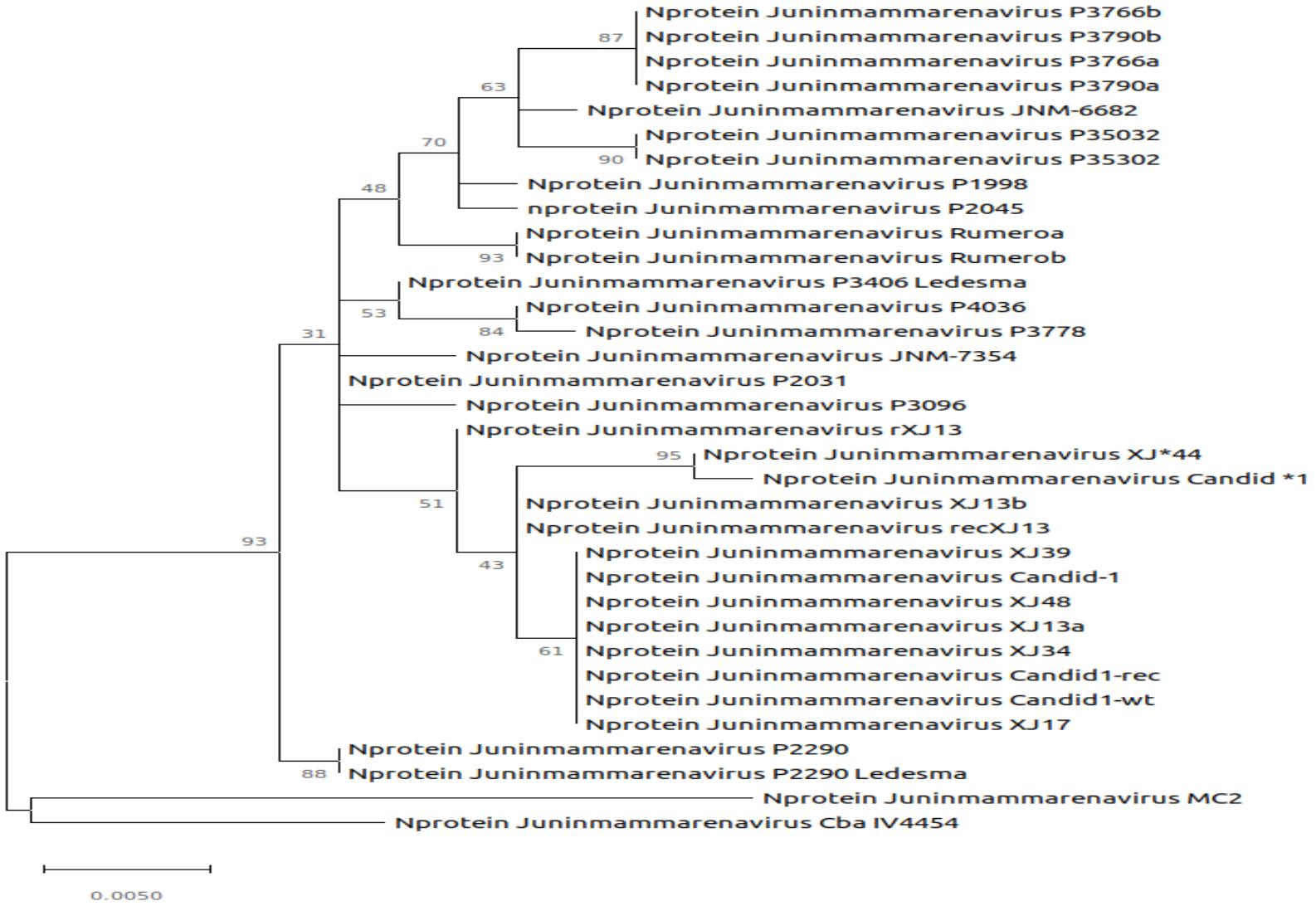


Figure6.5d: Phylogenetic tree of NP sequences of JUNV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

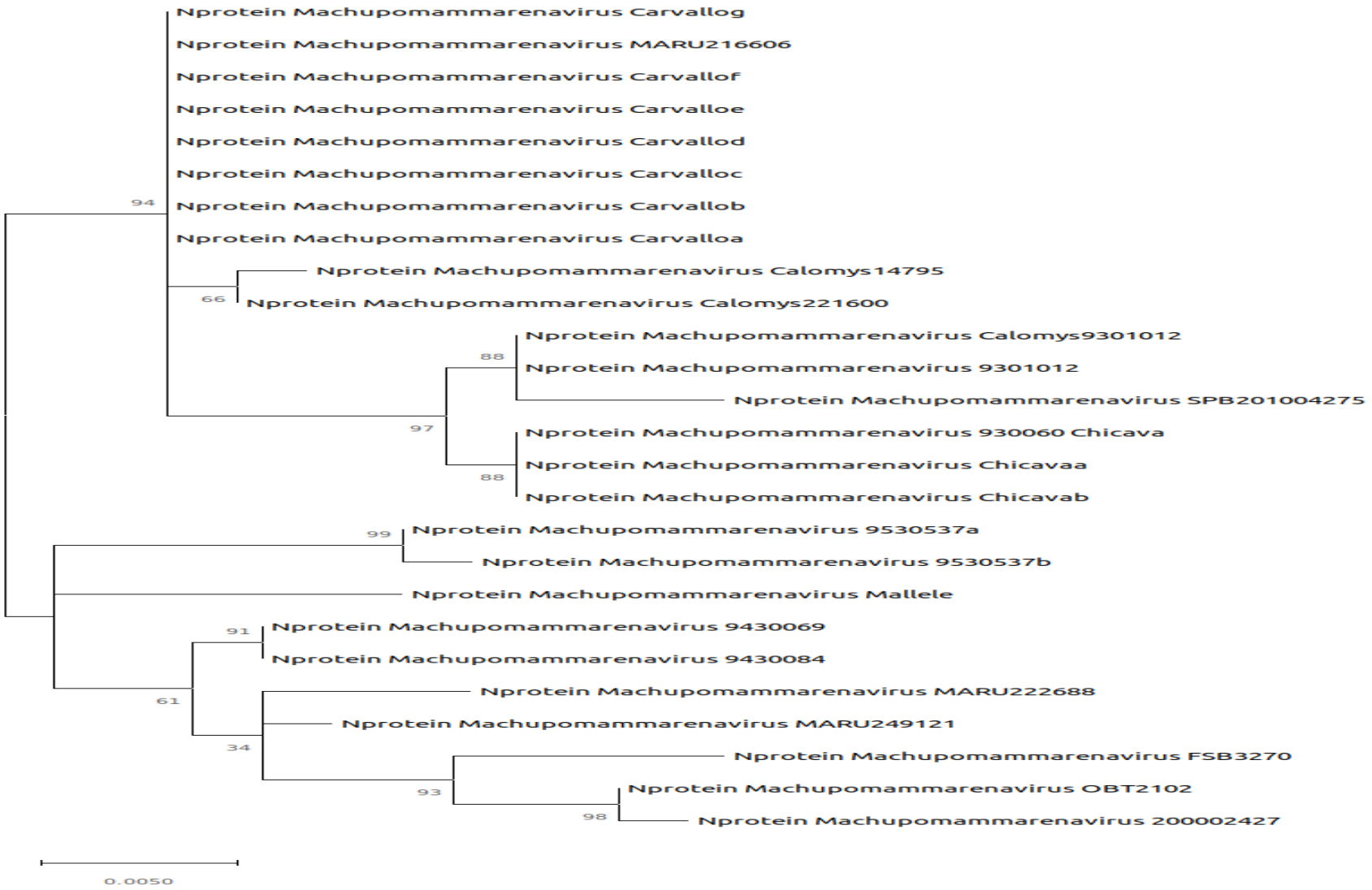


Figure 6.6a: Phylogenetic tree of NP sequences of MACV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

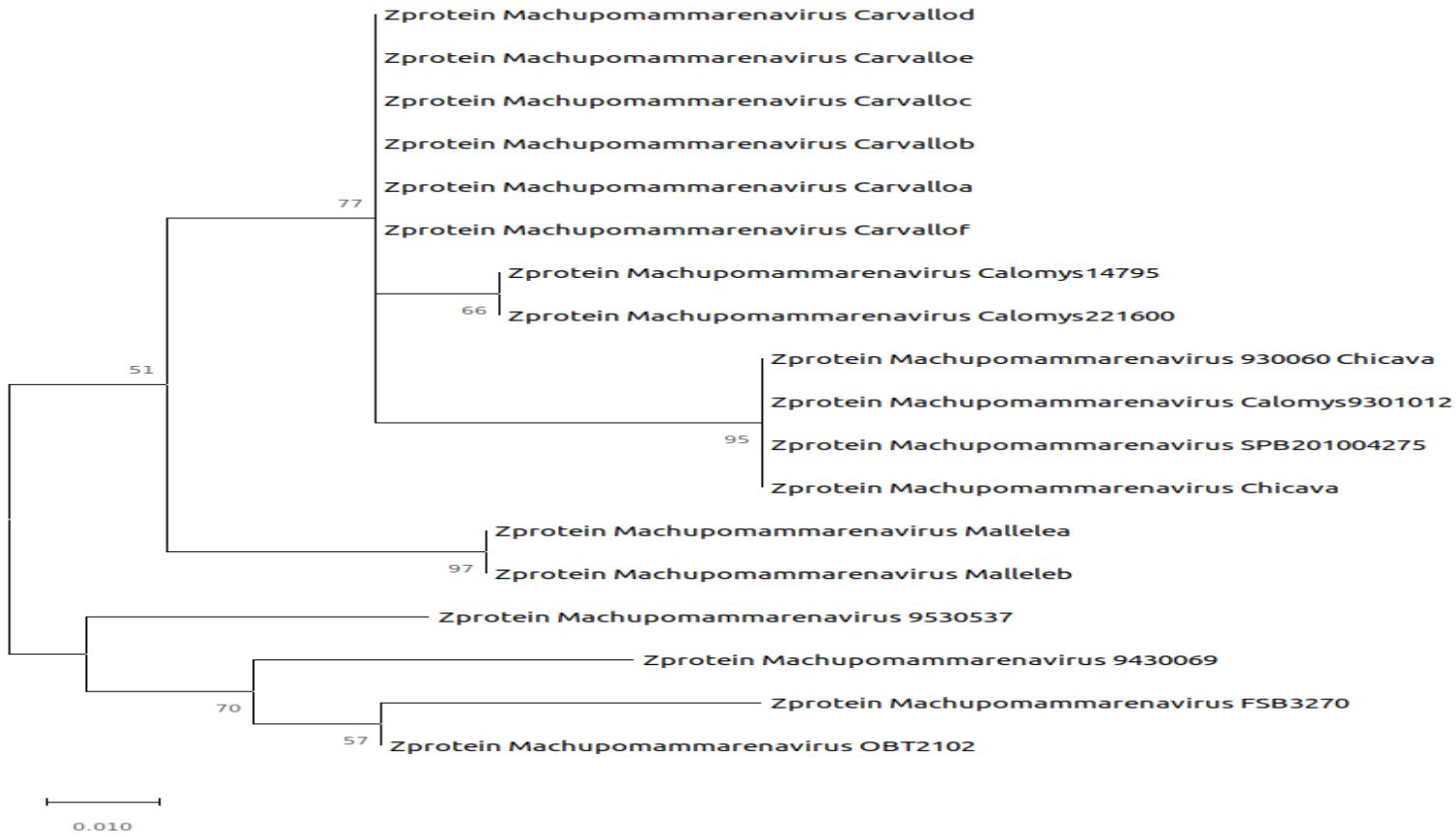


Figure 6.6b: Phylogenetic tree of Z sequences of MACV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

Figure 6.6c: Phylogenetic tree of L sequences of MACV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

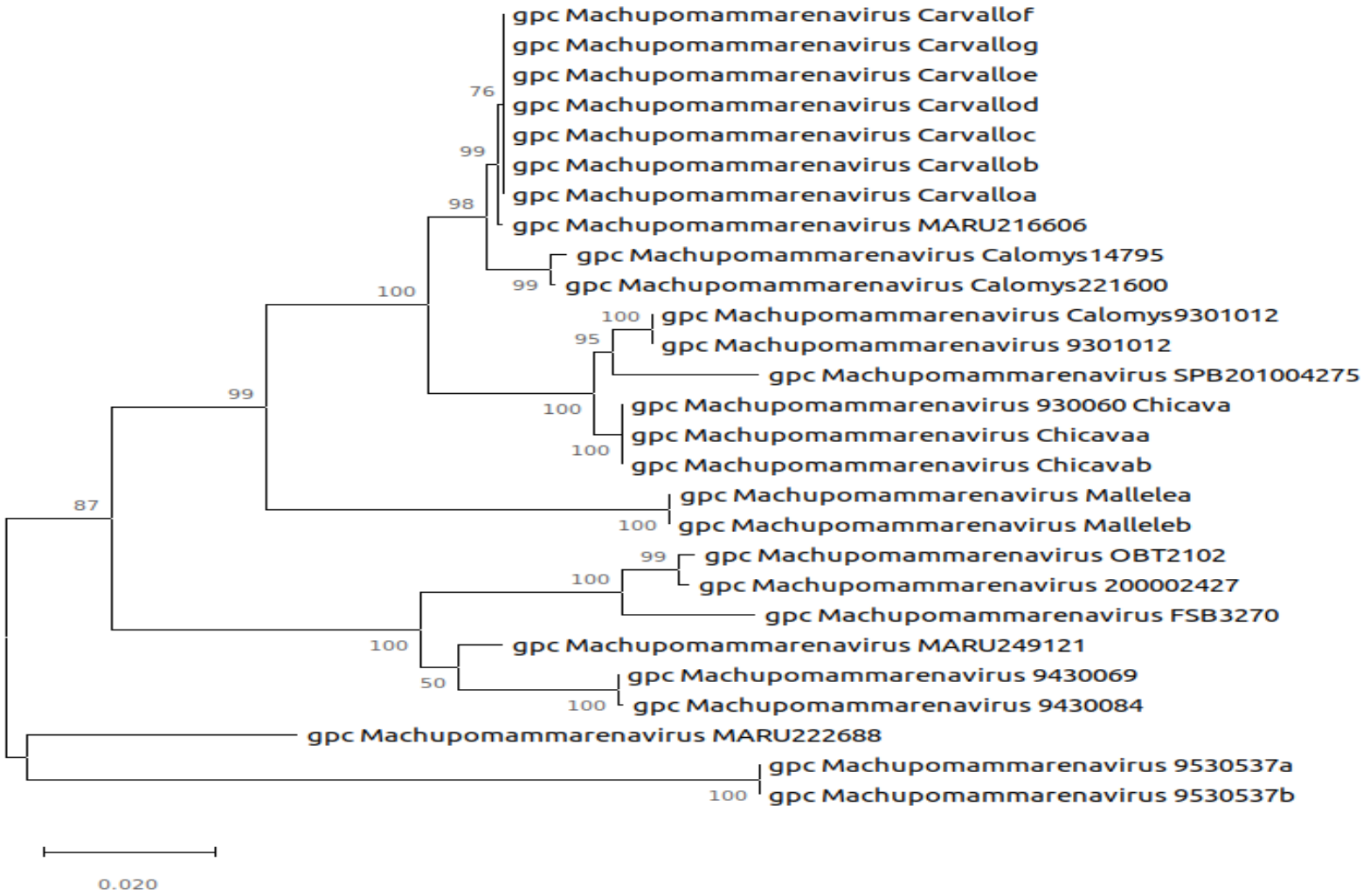


Figure 6.6d: Phylogenetic tree of GPC sequences of MACV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

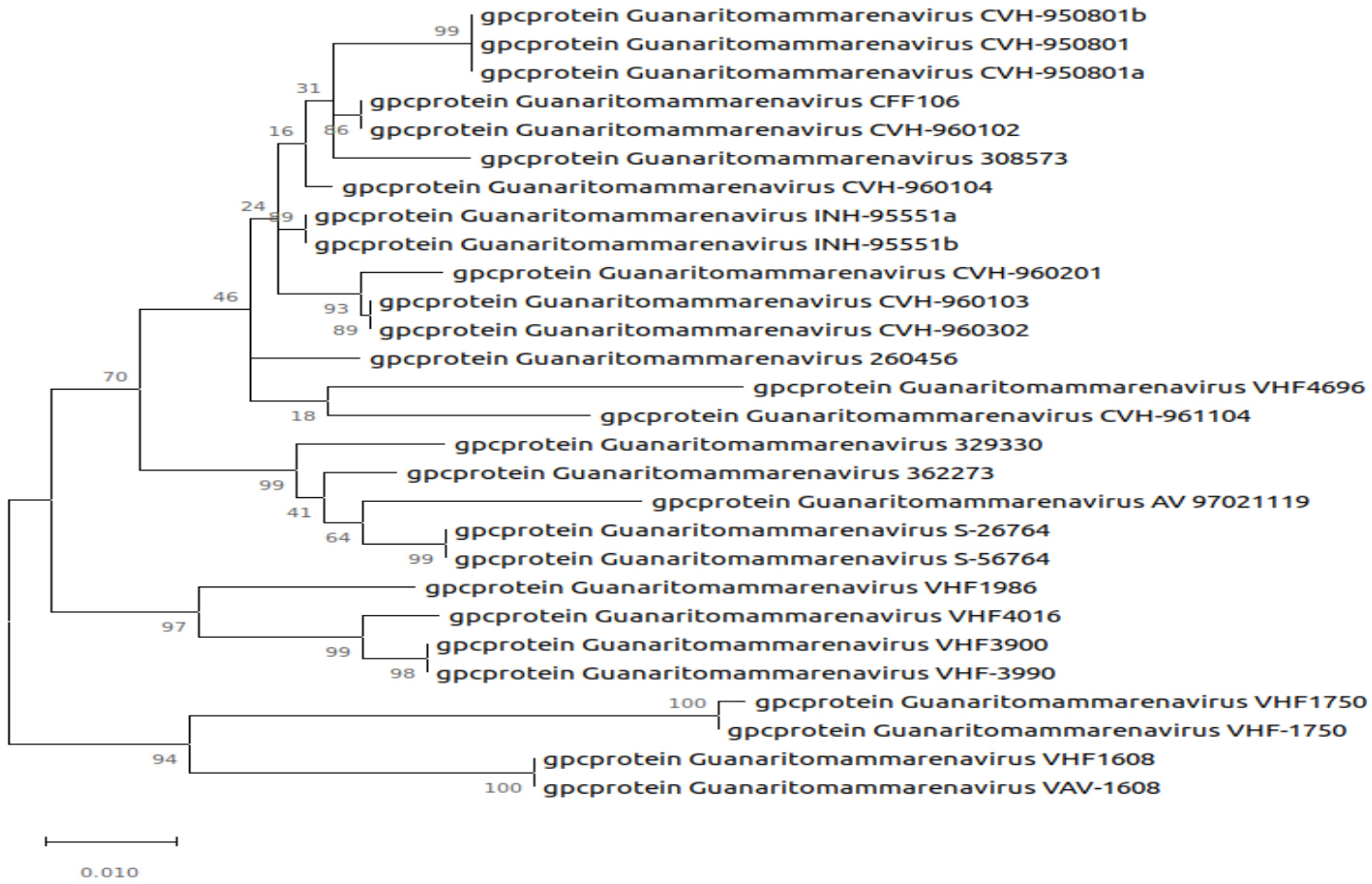


Figure 6.7a: Phylogenetic tree of GPC sequences of GTOV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

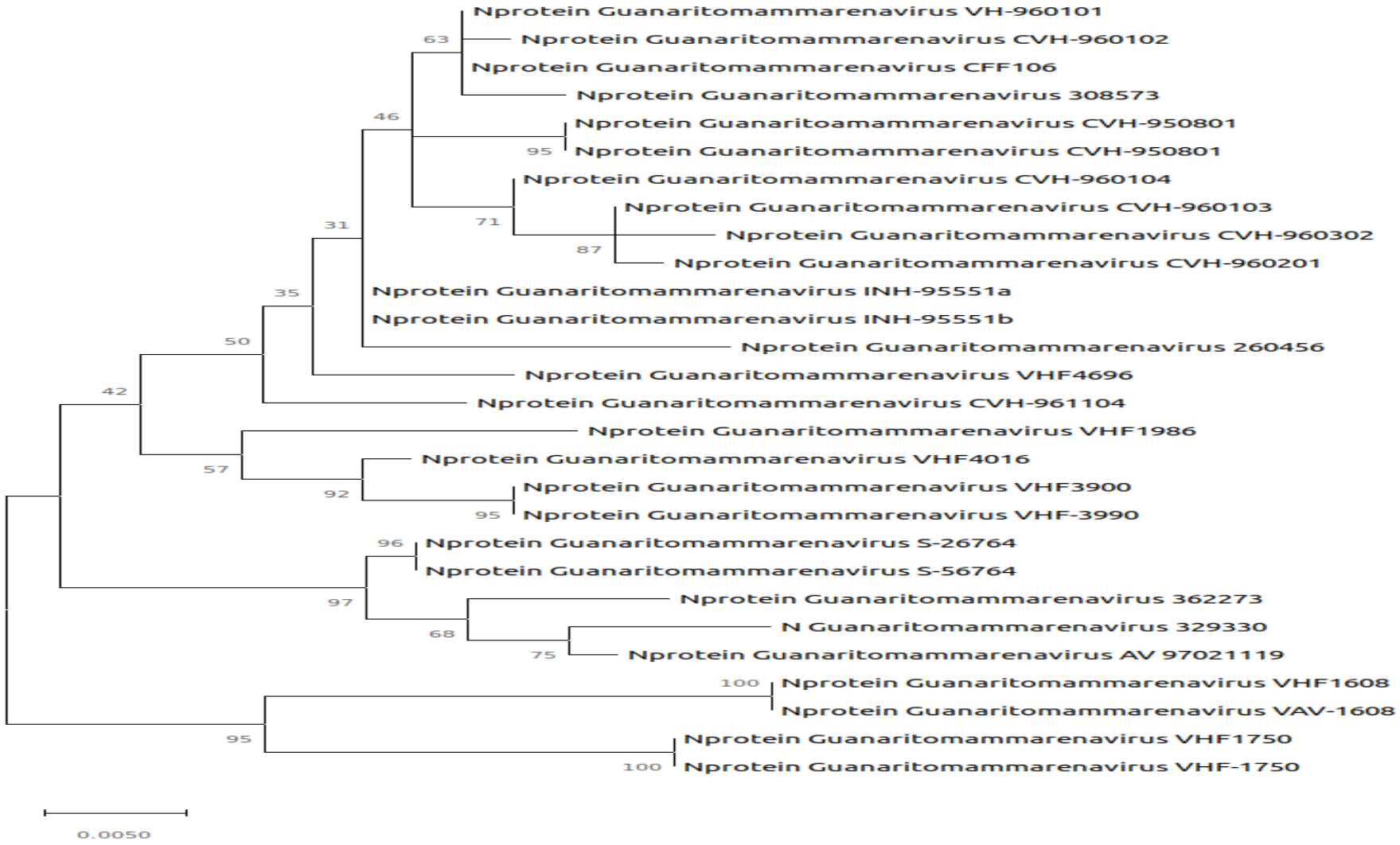


Figure6.7b: Phylogenetic tree of NP sequences of GTOV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

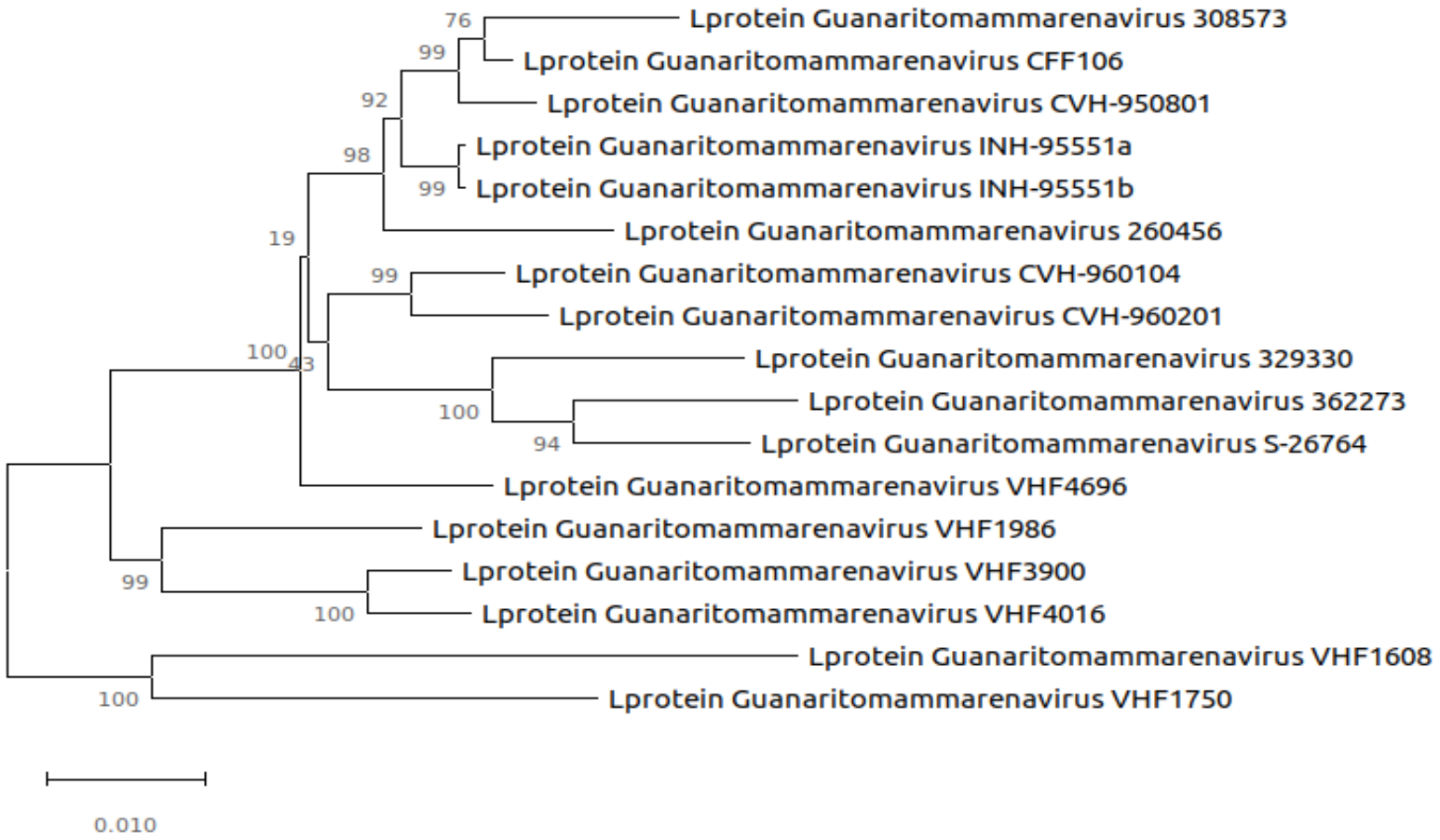


Figure 6.7c: Phylogenetic tree of L sequences of GTOV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

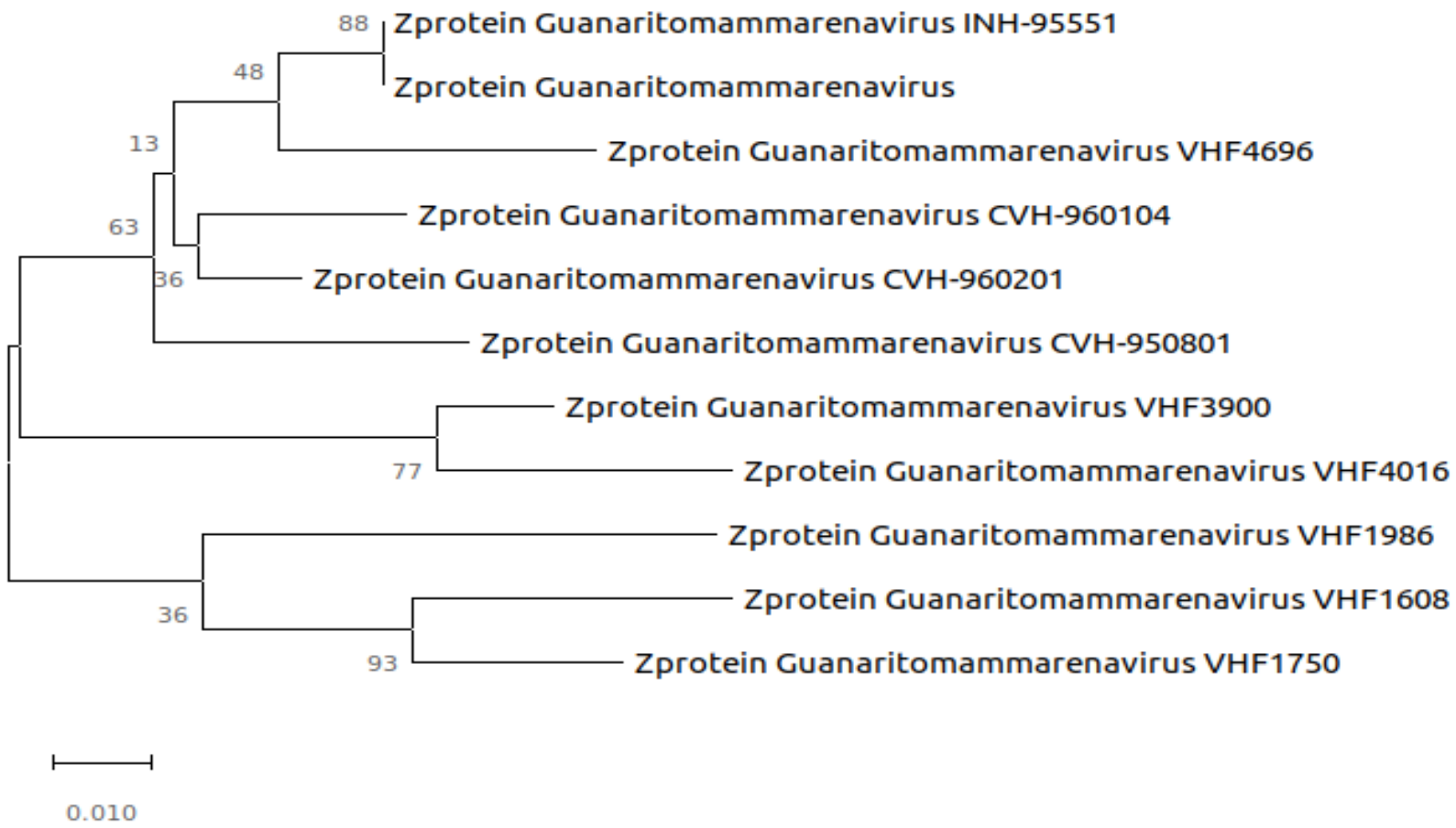


Figure 6.7d: Phylogenetic tree of Z sequences of GTOV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

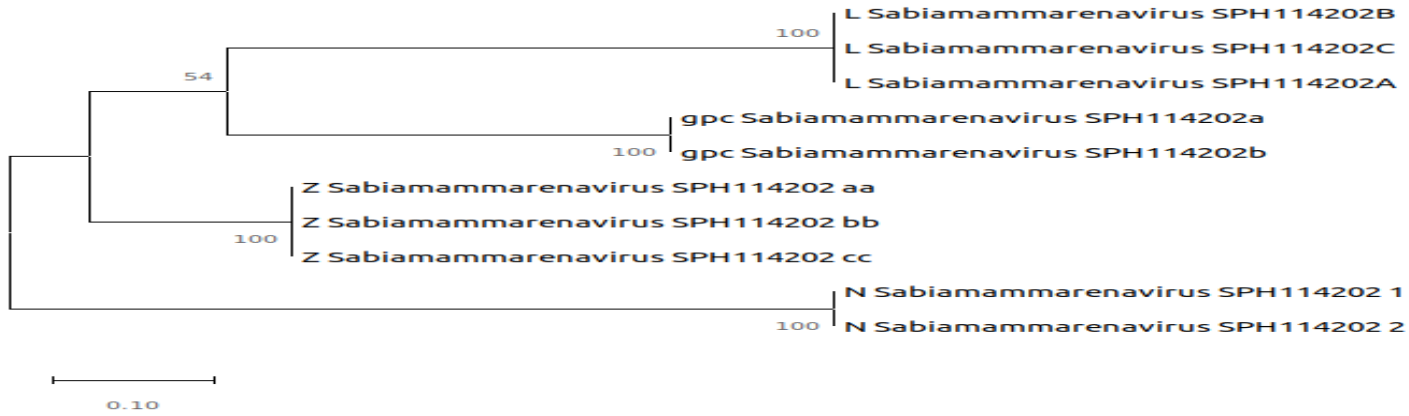


Figure 6.8: Phylogenetic tree of protein sequences of SABV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

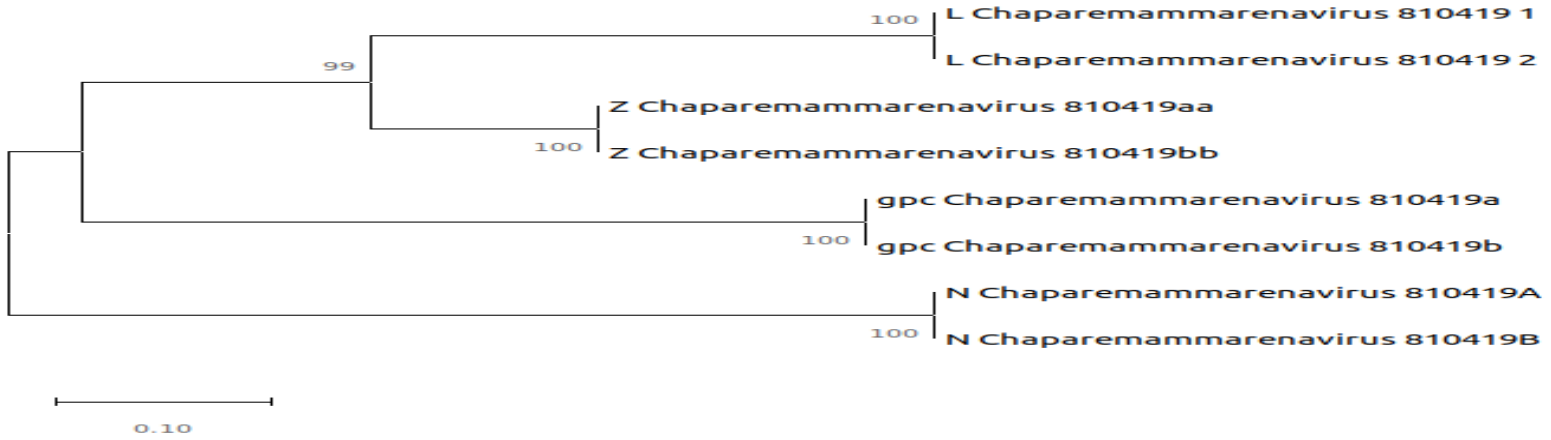


Figure 6.9: Phylogenetic tree of protein sequences of CHPV (Maximum Likelihood method and Jones-Taylor-Thornton (JTT model) model of substitution). Bootstrap values are shown (1000 replicates).

Table 6.16: Showing results of best scored ligands (FDA approved library) for JUNV

ZINC ID	Name	Binding energy (kcal/mol)	Interacting residues
ZINC169289767	Trypan Blue	-9.4±0.01	*Lys102,Thr170,Pro219,Trp222;*Asn105*,Lys137*,Ser138*,Gln141*,Arg167*,Thr168*,Thr220*,Leu228*
ZINC27990463	Lomitapide	-8.6±0.01	*Pro120,Leu163,Asn178,Thr182,Leu212*;Ser180*,Asn185*
ZINC000011679756	Eltrombopag	-8.4±0.01	*Pro160,Leu163,Asn178,Thr182,Leu212*;Pro161*,Leu163*,Ser180*,Asn185*
ZINC1612996	Irinotecan	-8.1±0.02	*Ala106,Gln141,Arg167,Thr170, Pro219*;Lys137*,Ser138*
ZINC3978005	Dihydroergotamine	-8.1±0.01	*Lys137,Phe173,Pro219;*Gln141*

5 Best ligands after screening of FDA-approved drug library with their binding energy score with target protein GP1 of JUNV are shown. Hydrophobic interactions shown in italics, Hydrogen bonds are marked with *.

Table 6.17: Showing results from best ligand results (investigational drug library) for JUNV

#	ZINCID	Name	Binding Energy score	Interactions
1	ZINC000003975327	Telomestatin	-9.1±0.01	Ser138*
2	ZINC000012358610	Phthalocyanine	-9.7±0.01	*Ala116,Pro120,Ile125,Pro160, Pro161,Leu163,Leu214*; Asn178*,Asn185*
3	ZINC000043203371	MK-3207	-8.6±0.01	*Phe173,Pro219;*Ser107*, *Lys137*,* Ser138*,Gln141*
4	ZINC000003922429	Adozelesin	-8.8±0.02	*Lys137*,Arg167,Phe173, Pro219;*Lys102*,Gln141*
5	ZINC000095561192	Unii-I6KF9AF7F7	-8.8±0.02	*Lys137*,Phe173,Pro219, Trp222,*Asn105*,Ser107*, Gln141*

5 Best ligands after screening of investigational drug library with their binding energy score with target protein GP1 of JUNV are shown. Hydrophobic interactions shown in italics, Hydrogen bonds are marked with *.

Table 6.18: Showing binding energy of selected ligands with GP1 protein of JUNV

#	Molecular Formula	Name	Binding energy kca/mol
1.	C18H23N3O3	SCHEMBL18497780	-6.5±0.03
2.	C17H14ClNO2	3-Chloro-5-methoxy-10-allylacridine-9(10H)-one	-4.6±0.03
3.	C13H17N3O4S3	1,1-Dioxo-2-propyl-3-[(prop-2-ynylamino)methyl]thieno[3,2-e]thiazine-6-sulfonamide	-5.3±0.02
4.	C26H36N2O9	Antimycin A3	-6.6±0.02
5.	C12H9F3N2O2	A771726	-5.8±0.02
6.	C3H5NO	Acrylamide	-3.2±0.04

7.	C10H11N5O4	Adenosine dialdehyde	-5.2±0.03
8.	C9H8N4O4	5-(Aziridin-1-yl)-2-nitro-4-nitrosobenzamide	-5±0.03
9.	C25H27N5O5	Avn-944	-5.6±0.03
10.	C2H4N4O2	1,2-Diazenedicarboxamide	-4.1±0.04
11.	C28H48O6	24-Epibrassinolide	-6.3±0.02
12.	C16H24O4	Brefeldin A	-6.2±0.02
13.	C8H10N4O2	Caffeine	-4.7±0.04
14.	C10H5F3N4O	Carbonyl cyanide p-trifluoromethoxyphenylhydrazone	-4.1±0.02
15.	C12H16N4O3	3-(4-Aminoimidazo[4,5-c]pyridin-1-yl)-5-(hydroxymethyl)cyclopentane-1,2-diol	-5.2±0.03
16.	C34H36N4O6	Chlorin e6	-6.3±0.02
17.	C17H19ClN2S	Chlorpromazine	-4.7±0.03
18.	C6H3Cl3	1,2,4-Trichlorobenzene	-3.8±0.04
19.	C7H8O4	3-(Dihydroxymethyl)benzene-1,2-diol	-5±0.02
20.	C12H11F3N2O2	Antiproliferative agent A771726	-5.8±0.03
21.	C2H5O6P	2-Phosphoglycolic Acid	-3.2±0.02
22.	C19H28O2	Dehydroepiandrosterone	-5.7±0.03
23.	C48H78N7O20P3S	(25R)-24-Oxo-DHCA-CoA	-5.2±0.03
24.	C11H9N5O2	DHPA	-6.3±0.02
25.	C19H30O2	Epiandrosterone	-5.3±0.02
26.	C12H24O2	Lauric acid	-3.5±0.02
27.	C9H5ClN4	m-Chlorophenyl carbonylcyanide hydrazone	-5.7±0.02
28.	C17H20O6	mycophenolic acid	-4.9±0.03
29.	C14H14N4O2S2	NSC 4492	-4.6±0.03
30.	C17H14Br2O	NSC 14560	-5.6±0.03
31.	C10H12N5Na2O8P	NSC 20265	-6.3±0.02
32.	C14H18N6O4S3	NSC 71033	-5.7±0.03
33.	C9H13N3O6	Pyrazofurin	-5.0±0.04
34.	C8H12N4O5	Ribavirin	-5±0.02
35.	C24H25N3O	ST-294	-6±0.03
36.	C5H4FN3O2	T-705	-4.6±0.03
37.	C10H20N2S4	Tetradine	-3.7±0.04
38.	C21H24F3N3S	TFP(trifluoperazine)	-6.2±0.02

Binding energy score of each ligand was calculated separately by docking with 3D structure of GP1 protein (5W1K) of JUNV.

Table 6.19: Showing results of best hit ligands (from FDA approved drug library) for MACV

S.No	ZINCID	Name	Binding energy score (kcal/mol)	Interactions
1	ZINC000052955754	Ergotamine	-10.7±0.02	*Leu91,Pro160*;Met158*, Cys237*;**Tyr127**
2	ZINC000003978005	Dihydroergotamine	-11±0.02	*Met93,Pro160, Pro161, Arg201,* Leu91*; **Tyr127**
3	ZINC000066166864	Alectinib	-10±0.01	*Leu91,Pro160, Arg201*; **Tyr127**
4	ZINC000003914596	Saquinavir	-10±0.01	*Leu88,Tyr127,Pro160,Pro161,Arg 201*;Pro89*,Leu91*,Met93*,Lys1 20*,Leu157*, Arg201Q*,Gly202*
5	ZINC000100013130	Midostaurin	-11.1±0.01	*Leu88,Met93,Tyr127,Pro161;* **Trp147**

5 Best ligands after screening of FDA-approved drug library with their binding energy score with target protein GP1 of MACV are shown. Hydrophobic interactions shown in italics, Hydrogen bonds are marked with *and and π-stacking marked with bold.

Table 6.20: Showing results from screening of investigational drug library for MACV

S.No	ZINCID	Name	Binding energy score(kcal/mol)	Interactions
1	ZINC000012358610	Phthalocyanine	-11.4±0.01	*Leu88,Tyr127,Pro160,Leu199, Arg201*;Lys120*
2	ZINC000095608296	Unii-G9Z22EU5FK	-10.6±0.02	*Lys120,Tyr127,Trp147,Pro160, Pro161,Leu163,Asp184,Ala185,Phe 200*;Leu91*,Ser125*,Asn178*; Leu199^
3	ZINC000043203371	MK-3207	-10.4±0.01	*Leu88,leu91,Tyr127,Phe200,Arg201* ;Ser125*,Tyr127*,Met158*;Leu91^
4	ZINC000100341584	Setrobuvir	-10.2±0.01	*Met93,Tyr127,Pro161,Phe200, Arg201*;Leu91*,Lys120*,Ser125*
5	ZINC000059749972	Radotinib	-10.2±0.01	*Met93,Met158,Pro160,Val187, Phe200*;Lys191*

5 Best ligands after screening of investigational drug library with their binding energy score with target protein GP1 of MACV are shown. Hydrophobic interactions shown in italics, hydrogen bonds are marked with *and halogen interaction with ^.

Table 6.21: Binding energy score of selected ligands with GP1 protein of MACV

#	Molecular Formula	Name	Binding energy in kcal/mol
1.	C18H23N3O3	SCHEMBL18497780	-6.5±0.01
2.	C17H14ClNO2	3-Chloro-5-methoxy-10-allylacridine-9(10H)-one	-6.8±0.02
3.	C13H17N3O4S3	1,1-Dioxo-2-propyl-3-[(prop-2-ynylamino)methyl]thieno[3,2-e]thiazine-6-sulfonamide	-6.5±0.02

4.	C26H36N2O9	Antimycin A3	-6.1±0.01
5.	C12H9F3N2O2	A771726	-6.4±0.02
6.	C3H5NO	Acrylamide	-3.3±0.03
7.	C10H11N5O4	Adenosine dialdehyde	-4.8±0.03
8.	C9H8N4O4	5-(Aziridin-1-yl)-2-nitro-4-nitrosobenzamide	-4.7±0.03
9.	C25H27N5O5	Avn-944	-6.7±0.01
10.	C2H4N4O2	1,2-Diazenedicarboxamide	-4.2±0.02
11.	**C28H48O6**	**24-Epibrassinolide**	**-7.6±0.01**
12.	C16H24O4	Brefeldin A	-5.4±0.02
13.	C8H10N4O2	Caffeine	-5.3±0.02
14.	C10H5F3N4O	Carbonyl cyanide p-trifluoromethoxyphenylhydrazone	-5.5±0.04
15.	C12H16N4O3	3-(4-Aminoimidazo[4,5-c]pyridin-1-yl)-5-(hydroxymethyl)cyclopentane-1,2-diol	-6.3±0.03
16.	C34H36N4O6	Chlorin e6	-6.3±0.01
17.	C17H19ClN2S	Chlorpromazine	-4.6±0.02
18.	C6H3Cl3	1,2,4-Trichlorobenzene	-4±0.03
19.	C7H8O4	3-(Dihydroxymethyl)benzene-1,2-diol	-5.1±0.03
20.	C12H11F3N2O2	Antiproliferative agent A771726	-5.8±0.02
21.	C2H5O6P	2-Phosphoglycolic Acid	-4.1±0.03
22.	C19H28O2	Dehydroepiandrosterone	-5.8±0.02
23.	C48H78N7O20P3S	(25R)-24-Oxo-DHCA-CoA	-5.9±0.02
24.	C11H9N5O2	DHPA	-6.5±0.01
25.	C19H30O2	Epiandrosterone	-6.9±0.02
26.	C12H24O2	Lauric acid	-4.2±0.04
27.	C9H5ClN4	m-Chlorophenyl carbonylcyanide hydrazone	-5.6±0.03
28.	C17H20O6	mycophenolic acid	-5.2±0.03
29.	C14H14N4O2S2	NSC 4492	-6.1±0.03
30.	C17H14Br2O	NSC 14560	-6.3±0.02
31.	C10H12N5Na2O8P	NSC 20265	-6.1±0.01
32.	C14H18N6O4S3	NSC 71033	-5.7±0.02
33.	C9H13N3O6	Pyrazofurin	-6.9±0.01
34.	C8H12N4O5	Ribavirin	-6.1±0.01
35.	**C24H25N3O**	**ST-294**	**-7.3±0.01**
36.	C5H4FN3O2	T-705	-4.2±0.02
37.	C10H20N2S4	Tetradine	-3.7±0.04
38.	C21H24F3N3S	TFP(trifluoperazine)	-5.2±0.03

Results showing binding energy scores of each ligand docked with GP1 protein structure (5W1M) of MACV downloaded from PDB database. 2 best hits have been highlighted.

6.12 Calculation of physiochemical properties and toxicity of selected ligands for JUNV and MACV

Physiochemical properties of ligands showing good docking scores with GP1 target protein of JUNV and MACV were evaluated. Physiochemical properties based on Lipinski's Rule of Five (Lipinski *et al*., 2001) which includes the following criteria that molecular weight must be less than 500, number of hydrogen-bond donors less than 5, number of hydrogen bond acceptors less than 10 and Log P value must be less than 5. These properties help in evaluation of drug-likeliness of ligand structures.

Analysis of 5 best hit ligands from FDA-approved drug library (Table 6.16) and 5 from investigational-drug library (Table 6.17) for JUNV showed that only compound ZINC000011679756 with docking score of -8.4kcal/mol (Table 6.16) follow the Lipinski's rule of five (Lipinski *et al*., 2001) (Table 6.22 and 6.24). However, ligands showing mild variations in physiochemical properties (Table 6.22 and 6.24) can also be considered. As modifications in physiochemical properties are also one of the techniques to increase the bioavailability of drug (Fasinu *et al*., 2011). Further, *in silico* evaluation of toxic parameters (mutagenecity, carcinogenicity and cytotoxicity) was also performed on best selected ligands from FDA-approved drug library and investigational drug library and compounds active for toxic parameters were not considered further (Table 6.23, 6.25).

Toxicity parameters and physiochemical properties for *in silico* drug designing against Nipah virus was also analyzed by using ProTox -II and SWISSADME server (Shah *et al.,* 2018).

One of the other parameter Pan-assay interference structures (PAINS), that include fluorescence of small molecules, redox reactivity and covalent modifications of target protein was also evaluated. Only one ligand compound ZINC000011679756 was predicted to possess PAINS value 1 (Table 6.23, 6.25) and was not considered further. Thorough analysis of interaction, physiochemical properties and toxicity predicts ligand with Zinc ID ZINC000043203371 (MK-3207) and docking score -8.6kcal/mol as safe and best candidate for further studies against GP1 protein of JUNV (Table 6.17, 6.24 and 6.25) (Malhotra *et al.*,2022). In 2014 Li and colleagues also studied MK-3207 as drug against target CGRP (calcitonin gene related peptide)

receptor to increase clinical efficiency and predict the efficiency of drug to cure migraine (Li *et al*., 2014).

Extensive analysis of physiochemical properties of best docked ligands for MACV predicted five best hit ligands from FDA-approved drug library (Table 6.19) and five from investigational-drug library (Table 6.20). Compound ZINC000066166864 with docking score of -10kcal/mol (Table 6.19) follow the Lipinski's rule of Five (Table 6.26). Other ligands showing mild variations in physiochemical properties (Table 6.26 and 6.28) can also be considered for further analysis. Durairaj and colleagues in 2017 analysed the physiochemical properties of various ligands used for docking against target Mtb-KasA enzyme for *Mycobacterium tuberculosis* and C-14 was predicted as best lead molecule (Durairaj, 2017).

Toxicity parameters (mutagenecity, carcinogenicity and cytotoxicity) were also analyzed and compounds showing toxicity were not considered further (Table 6.27 and 6.29). Further, no compound was predicted to possess structural alerts associated with false positive signals (Table 6.27 and 6.29).

Thorough analysis of physiochemical properties and toxicity predicts ZINC000003978005 (Dihydroergotamine) with docking score -11kcal/mol as safe and best candidate for further studies against GP1 protein of MACV (Malhotra *et al.,* 2022). Also, recently in 2020 *in silico* drug designing against coronavirus was performed and further, screening of FDA drug library was done inferring dihydroergotamine which is the anti-migraine drug as one of the potential inhibitor of three target enzymes of coronavirus (Gurung *et al.,* 2020).

Further, based on binding energy score and physiochemical properties selected ligands were studied for their interactions with the target protein. This analysed that MK-3207 selected as one of the prominent ligand show hydrophobic and hydrogen bond interactions with Phe173, Pro219, Ser107, Lys137, Ser138 and Gln141 residues of GP1 protein of JUNV. Similarly, Dihydroergotamine show hydrogen bond, hydrophobic interactions and Pi stacking interactions Met93, Pro160, Pro161, Leu91, Arg201 and Tyr 127 with residues of GP1 protein of MACV.

Table 6.22: Showing physiochemical properties of best selected ligands (FDA approved drugs) for JUNV

ZINC ID	Molecular weight	Log P	Number of hydrogen bond donor	Number of hydrogen bond acceptor
ZINC169289767	872.88	4.01	8	18
ZINC27990463	693.72	7.79	2	9
ZINC000011679756	**442.47**	**3.74**	**3**	**6**
ZINC1612996	*586.68*	*3.73*	*1*	*8*
ZINC3978005	*583.68*	*2.15*	*3*	*6*

Physiochemical properties of above 5 selected ligands are mentioned and ligand following Lipinski's rule of five is highlighted. LogP is logarithm of partition coefficient. Ligands showing minor variations in Lipinski's Rule of five (Molecular weight>500) has been italicised.

Table 6.23: Showing toxicity of best selected ligands (FDA approved drugs) for JUNV

ZINC ID	Mutagenecity	Cytotoxicity	Carcinogenecity	PAINS alert
ZINC169289767	**Yes**	**No**	**Yes**	**0**
ZINC27990463	No	No	No	0
ZINC000011679756	*No*	*No*	*No*	*1*
ZINC1612996	**No**	**No**	**Yes**	**0**
ZINC3978005	No	No	No	0

Ligands showing toxicity are highlighted. PAINS-Pan-assay-interference structure and ligand showing PAINS alert is italicised.

Table 6.24: Showing physiochemical properties of best ligands (investigational drug library) for JUNV

ZINCID	Molecular weight	Number of hydrogen acceptors	Number of hydrogen donors	LogP
ZINC000003975327	*582.5*	*15*	*0*	*2.21*
ZINC000012358610	518.57	2	6	5.88
ZINC000043203371	*557.59*	*7*	*3*	*3.32*
ZINC000003922429	*502.52*	*4*	*3*	*3.82*
ZINC000095561192	680.77	6	2	6.63

Physiochemical properties of above 5 selected ligands are mentioned. Ligand following Lipinski's rule of five with minor variation (Molecular Weight>500) has been italicised. LogP is logarithm of partition coefficient.

Table 6.25: Showing toxicity of 5 best ligands (investigational drug library) for JUNV

ZINCID	Mutagenecity	Carcinogenecity	Cytotoxicity	PAINS alert
ZINC000003975327	No	No	No	0
ZINC000012358610	**Yes**	**No**	**No**	**0**
ZINC000043203371	No	No	No	0
ZINC000003922429	**No**	**Yes**	**No**	**0**
ZINC000095561192	No	No	No	0

Ligands showing toxicity are highlighted. PAINS-Pan-assay-interference structure.

Table 6.26: Showing physiochemical properties of 5 best selected ligands (from FDA-approved drugs) for MACV

ZINCID	Molecular weight	Log P	Number of hydrogen donor	Number of hydrogen acceptor
ZINC000052955754	*581.66*	*2.26*	*3*	*6*
ZINC000003978005	*583.68*	*2.15*	*3*	*6*
ZINC000066166864	**482.62**	**4.33**	**1**	**4**
ZINC000003914596	670.84	2.87	5	7
ZINC000100013130	*570.64*	*4.11*	*1*	*4*

Physiochemical properties of above 5 selected ligands are mentioned. Ligand following Lipinski's rule of five has been highlighted. Ligands showing minor variations in Lipinski's ruke of five (Molecular weight>500) have been italicised. LogP is logarithm of partition coefficient.

Table 6.27: Showing toxicity of best selected ligands(from FDA approved drug library) for MACV

ZINCID	PAINS alert	Mutagenecity	Cytotoxicity	Carcinogenecity
ZINC000052955754	**0**	**No**	**No**	**Yes**
ZINC000003978005	0	No	No	No
ZINC000066166864	0	No	No	No
ZINC000003914596	0	No	No	No
ZINC000100013130	**0**	**No**	**Yes**	**No**

Ligands showing toxicity are highlighted. PAINS-Pan-assay-interference structure.

Table 6.28: Showing physiochemical properties of best ligands (from investigational drug library) for MACV

ZINCID	Molecular weight	Number of hydrogen acceptors	Number of hydrogen donors	LogP
ZINC000012358610	518.57	2	6	5.88
ZINC000095608296	771.87	9	0	5.15

ZINC000043203371	*557.59*	*7*	*3*	*3.32*
ZINC000100341584	*560.62*	*8*	*3*	*2.3*
ZINC000059749972	*530.05*	*9*	*2*	*4*

Ligands showing minor variations in Lipinski's rule of Five (Molecular weight>500) have been italicised. LogP is logarithm of partition coefficient.

Table 6.29: Showing toxicity of best selected ligands (from investigational drug library) for MACV

ZINCID	PAINS alert	Mutagenecity	Cytotoxicity	Carcinogenecity
ZINC000012358610	**0**	**Yes**	**No**	**No**
ZINC000095608296	0	No	No	No
ZINC000043203371	0	No	No	No
ZINC000100341584	0	No	No	No
ZINC000059749972	**0**	**No**	**No**	**Yes**

Ligands showing toxicity are highlighted. PAINS-Pan-assay-interference structure.

Table 6.30: Physiochemical properties and drug-like features of ligands selected for MACV based on Binding energy score with GP1 protein

#	Formula	MW	#H-bond acceptors	#H-bond donors	Log P	PAINS #alerts	Cytotoxicity	Mutagenecity
1	24-Epibrassinolide	480.68	6	4	3.62	0	No	No
2	ST-294	371.47	2	1	4.73	1	No	No

MW: molecular weight, Log P- Logarithm of partial coefficient; PAINS- Pan-Assay Interference Structures.

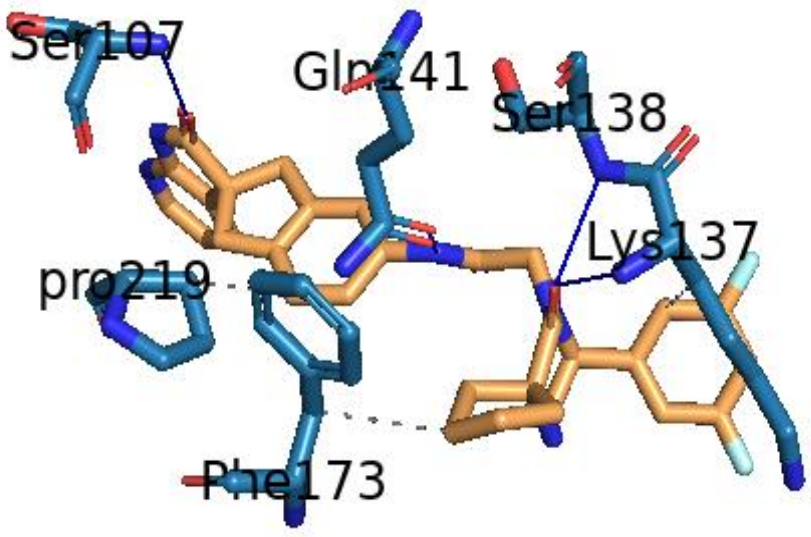


Figure 6.10: Mode of interaction of ligand MK-3207 with target protein GP1 of JUNV

Hydrophobic interactions and Hydrogen bonds are shown in dashed and blue lines.

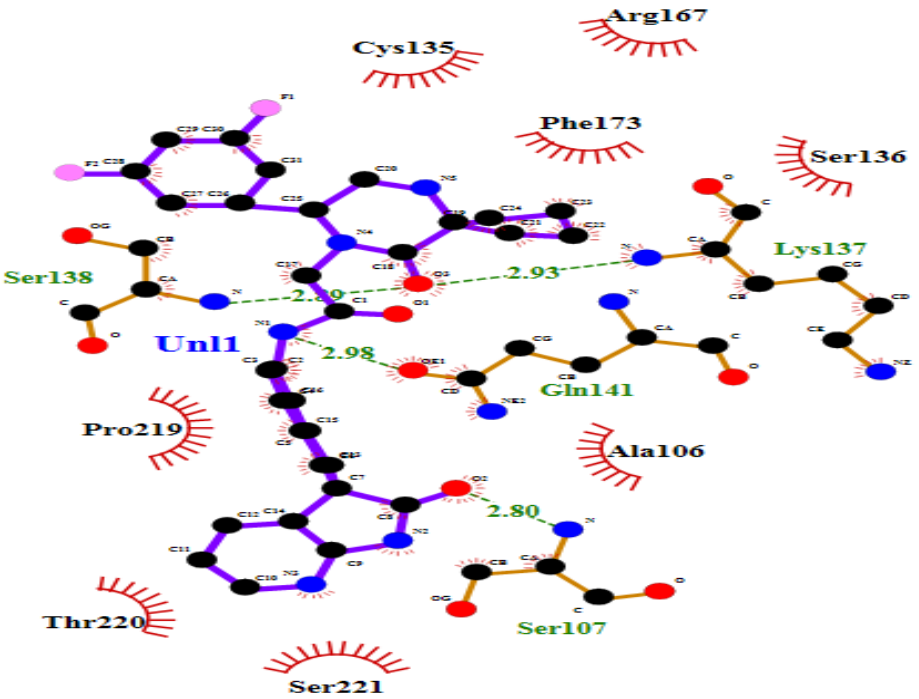


Figure 6.11: Molecular interaction between GP1 protein of JUNV and MK-3207

The hydrophobic interactions are indicated by red arcs with radiating spikes and green dashed lines correspond to hydrogen bonds.

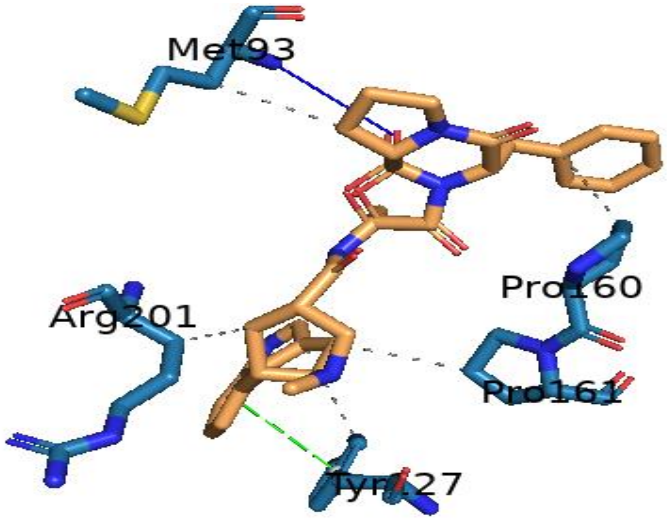


Figure 6.12: Mode of interaction of ligand Dihydroergotamine with target protein GP1 of MACV.

Hydrophobic interactions, hydrogen bonds and π-stacking shown as grey dashed, blue and green dashed lines.

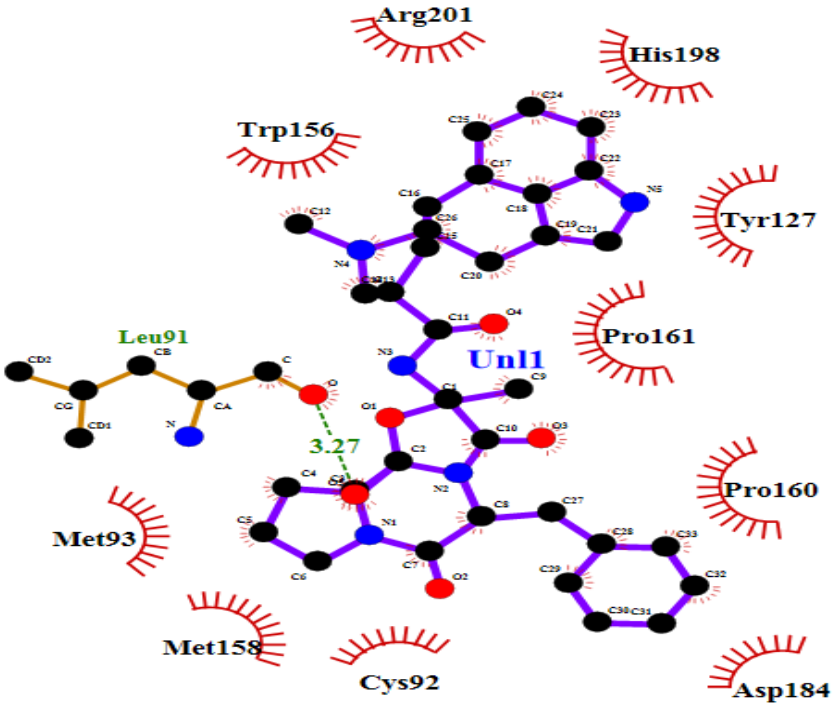


Figure 6.13: Molecular interaction between GP1 protein of MACV and Dihydroergotamine

The hydrophobic interactions are indicated by red arcs with radiating spikes and green dashed lines correspond to hydrogen bonds.

6.13 Molecular Dynamics

The compound MK-3207(ZINC000043203371) was selected for targeting the GP1 protein of JUNV on the basis of computation of binding energy and physiochemical properties (Table 6.17). Further, to validate the results of molecular docking the complex MK-3207 with GP1 protein of JUNV was selected for MD simulations along a timescale of 120 ns and based on this analysis the conformational stability of complex was assessed. The complex displayed fluctuations in RMSD values of Cα atoms until 102 ns and after that attained stability (Figure 6.14a). In 2021 similar dynamics analysis was performed on taraxerol complex which was predicted as one of the ligand against SARS-CoV-2 to validate the docking results by using GROMACS software (Kar *et al.*, 2021).

Similarly, compound dihydroergotamine (ZINC000003978005) selected on the basis of binding energy against target GP1 protein of MACV (Table 6.19) and further results of docking were validated by computation of dynamic simulations. The

complex displayed fluctuations in RMSD values of Ca atoms until 106 ns and attained stability thereafter (Figure 6.14b).

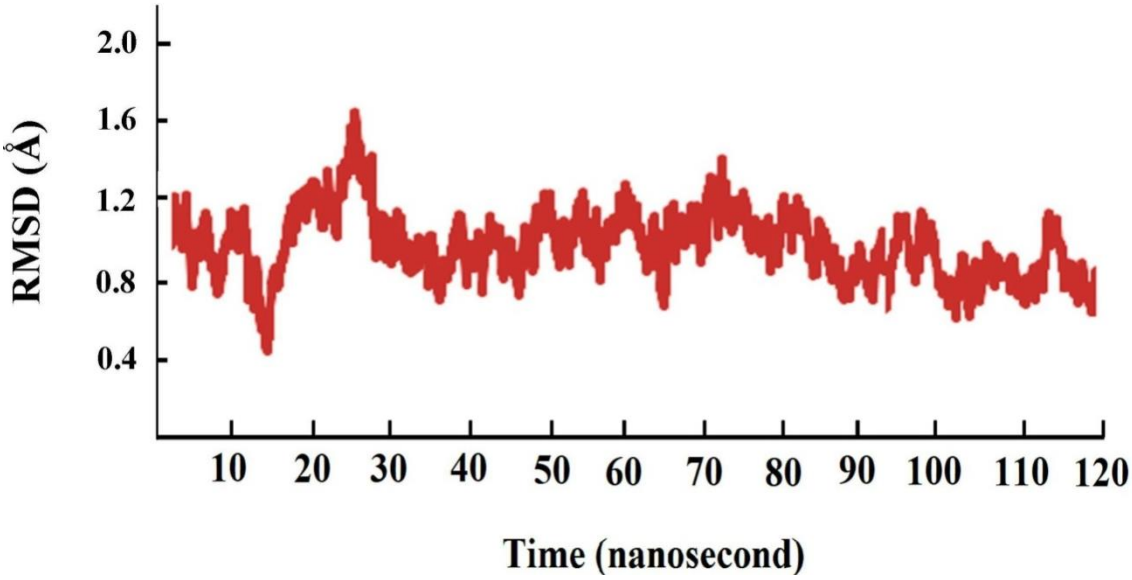


Figure 6.14a: RMSD analysis of the complex MK-3207 and JUNV GP1 protein (red)

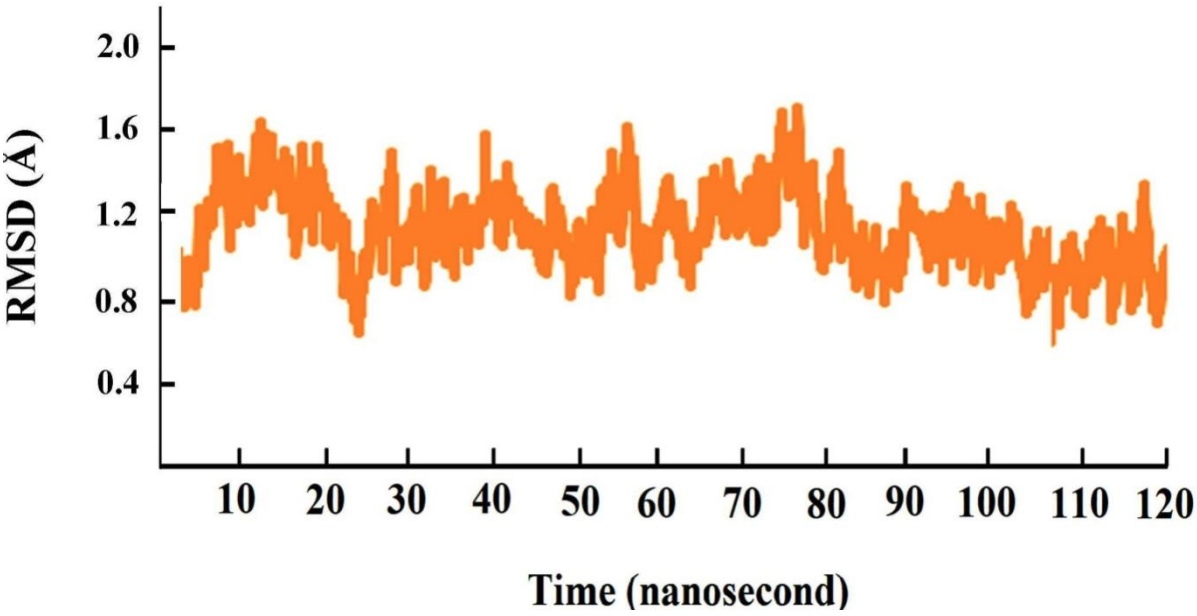


Figure 6.14b: RMSD analysis of the complex dihydroergotamine and MACV GP1 protein (orange)

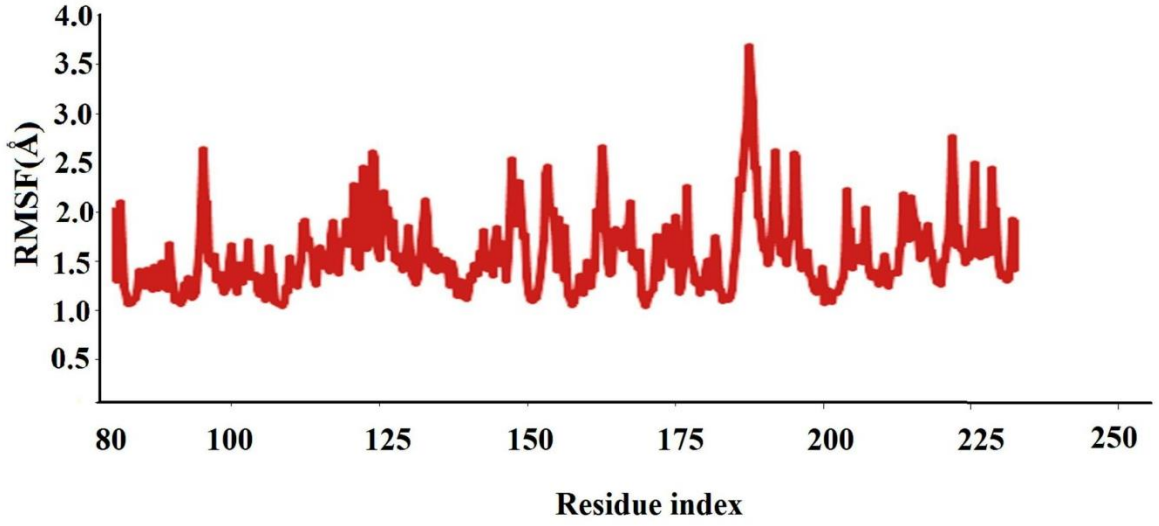


Figure 6.15a: RMSF analysis of the complex MK-3207 and JUNV GP1 protein (red)

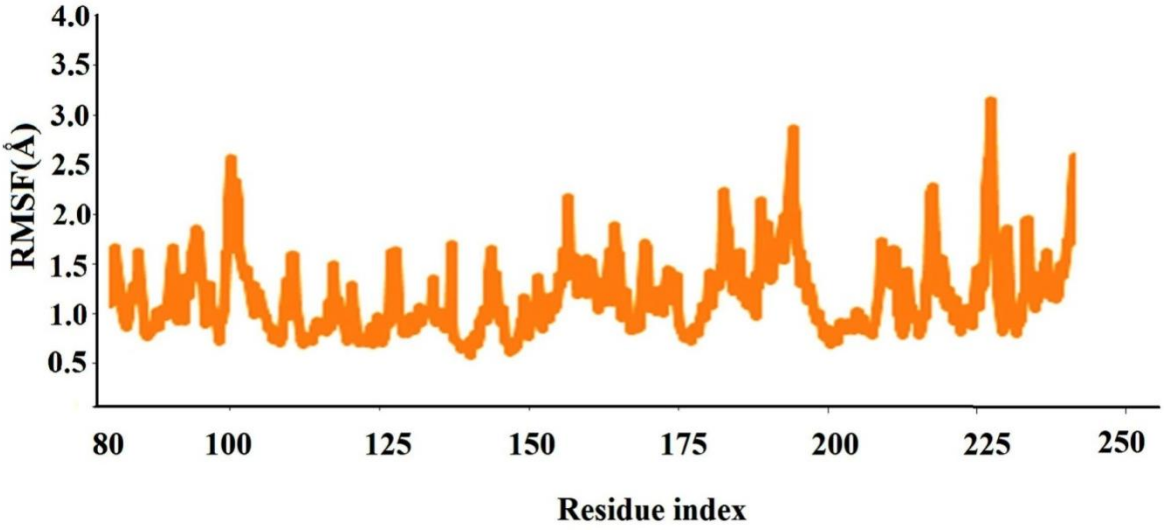


Figure 6.15b: RMSF analysis of the complex dihydroergotamine and MACV GP1 protein (orange)

Thorough RMSF analysis of the complexes revealed that the average RMSF values were found to be 2.0 Å for the complex MK-3207 and GP1 protein of JUNV (Figure 6.14b) and 1.7 Å for the complex dihydroergotamine and GP1 protein of MACV (Figure 6.15b) respectively. Residues which show high level of fluctuations during molecular dynamic study were analyzed from peaks shown in the plot (Figure

6.14b and 6.15b). It was analysed from the results that residues Gln141, Phe173 and Pro219 in the active binding pockets of the complex MK-3207 and JUNV GP1 protein as shown in Table 6.17 and residues Tyr127, Pro161 and Arg201 of the complex dihydroergotamine and MACV GP1 protein as shown in Table 6.19 were relatively stable (Kar *et al*., 2021).

In present study, *in silico* drug designing was performed against target GP1 (subunit of Glycoprotein) of JUNV and MACV; further screening of ligands on the basis of binding energy, physiochemical properties and toxicity was performed. Finally, MK-3207 for JUNV and dihydroergotamine for MACV were predicted as best fit ligand molecules. Validation was performed by dynamics study.

Chapter-7

Summary and Conclusion

Human beings are surrounded with plethora of viral species that depend on host machinery for replication and are highly infectious; so there is urgent need to study the viral species at genomic level. With the emergence of highly pathogenic NW Arenavirus: JUNV, MACV, GTOV, SABV and CHPV; a modest effort was made in this present work to realize the molecular underpinnings of viral establishment in humans and rodents.

The present study demonstrated the analysis of codon usage patterns of the NW Arenaviral pathogens: JUNV, MACV, GTOV, SABV and CHPV relative to host codon usage architecture and this would be instrumental in addressing the problem like how the viruses mimic the genomic signatures of their hosts and modulate, adapt themselves for better fitness and sustenance in host micro-environment.

Study of genomic codon usage data revealed preference for AT rich codons in NW Arenaviruses and also A or T nucleotides were preferred at third position of codons. Similar results of preference for AT rich codons were detected in hosts of JUNV (*C. musculinus*), MACV (*C. callosus*) and GTOV (*Z. brevicauda*).

Further, results from GC3 Vs ENc and neutrality plot have scrutinized that NW Arenavirus have weak codon usage biasness with Enc values 50.78 for JUNV, 49.14 for MACV, 49.67 for GTOV, 50.144 for SABV and 46.23 for CHPV. This result provides information about adaptation of virus to the diverse host or the varied environment. Further CoA analysis was performed to analyze various factors affecting codon biasness. Results predicted that biasness in codon usage patterns were deciphered by mutational pressure which dominates as compare to other factors like hydropathicity, aromaticity, and genomic length which also have their effect.

Comparative genomics and proteomics based analysis resulted in fruitful inferences regarding viral policies of adaptation in human host. Also, various parameters such as CAI, RCDI and SiD inferred high level of similarity between codon patterns of virus and humans. CAI value of all viruses with humans were shown as 0.76 for JUNV, 0.76 for MACV , 0.74 for GTOV, 0.76 for SABV and 0.75 for CHPV.

Also dinucleotide patterns such as UpG and CpA were found to be over-represented; CpG was observed to be under-represented. This result was found to be consistent with hosts also. NNU-GNN codon pair was found to be most preferred

among all with 10.8% in JUNV, 11.09% in MACV and 11.8 % in GTOV, 10.6% in SABV and 11.04% in CHPV). NNC-ANN and NNC-UNN were predicted as other most preferred codon pairs. Similarly, dinucleotide patterns NNU-GNN, NNC-ANN, NNC-UNN were also noted to be predominant among the over-represented codon pairs in *H. sapiens* revealing preference of virus for human host machinery.

Most of the preferred codons in JUNV, MACV, GTOV, SABV and CHPV coding sequences use suboptimal isoacceptor tRNAs from human cells as inferred from study of highly preferred codons (for each amino acid) in viruses and the most abundant isoacceptor tRNAs in human cells. Nine in JUNV, eight in MACV, eight in GTOV, nine in SABV and ten in CHPV codons matched with most abundant isoacepptor tRNAs from human cells.

The present information which is the first report regarding codon analysis of NW Arenaviral may help to understand the evolution of NW Arenaviral and also have potential for development of the virus vaccines. Scrutiny of potential therapeutic targets in pathogenic NW Arenavirus revealed GPC as target protein based on presence of conserved forces in GPC and also its importance in interaction with host receptor protein to cause infection.

Further, GP1 (subunit of GPC) was selected as target for computational dug designing against JUNV and MACV. Subsequent screening and molecular docking investigations revealed that drug namely ZINC000043203371 (MK-3207) with energy score of -8.6kcal/mol against GP1 protein of JUNV and ZINC000003978005 (Dihydroergotamine) with energy score of -11kcal/mol against GP1 protein of MACV as potential inhibitors against both viruses and can be considered for further *in vitro* and *in vivo* experimental studies to draw a definite conclusion.

Understanding the global health emergency and the immediate need for drugs and vaccines for the treatment of arenaviral infection, the present study is undertaken to identify promising inhibitors for glycoprotein of JUNV and MACV through molecular docking approach. Our study suggests that drugs such as MK-3207 (ZINC000043203371) and Dihydroergotamine (ZINC000003978005) are the most potent lead molecules which can be taken for further studies in wet lab experimentations.

Facts and facets of intricate viral adaptation that came out of this present endeavour might prove utility in unravelling the tangles of viral behaviour in human environment. However, with the availability of genomic data and rapid advancement in field of Bioinformatics this is just the beginning of the exciting journey pertaining to research associated with human viral flora.

Bibliography

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Acosta, E. G., Bruttomesso, A. C., Bisceglia, J. A., Wachsman, M. B., Galagovsky, L. R., & Castilla, V. (2008). Dehydroepiandrosterone, epiandrosterone and synthetic derivatives inhibit Junin virus replication in vitro. *Virus Research*, *135*(2), 203–212. https://doi.org/10.1016/j.virusres.2008.03.014

Aiewsakun, P., & Simmonds, P. (2018). The genomic underpinnings of eukaryotic virus taxonomy: Creating a sequence-based framework for family-level virus classification. *Microbiome*, *6*(1), 1–24. https://doi.org/10.1186/s40168-018-0422-7

Akalin, P. K. (2006). Introduction to bioinformatics. *Molecular Nutrition and Food Research*, *50*(7), 610–619. https://doi.org/10.1002/mnfr.200500273

Allsop, A. E. (n.d.). *Bacterial genome sequencing and drug discovery*. 637–642.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ambrosio, A. M., Enria, D. A., & Maiztegui, J. I. (1986). Junin virus isolation from lympho-mononuclear cells of patients with argentine hemorrhagic fever. *Intervirology*, *25*(2), 97–102. https://doi.org/10.1159/000149662

Ambrosio, A., Saavedra, M. C., Mariani, M. A., Gamboa, G. S., & Maiza, A. S. (2011). Argentine hemorrhagic fever vaccines. *Human Vaccines*, *7*(6), 694–700. https://doi.org/10.4161/hv.7.6.15198

Amineni, U., Pradhan, D., & Marisetty, H. (2010). *In silico identification of common putative drug targets in Leptospira interrogans*. 165–173. https://doi.org/10.1007/s12154-010-0039-1

Andrei, G., & De Clercq, E. (1990). Inhibitory effect of selected antiviral compounds on arenavirus replication in vitro. *Antiviral Research*, *14*(4–5), 287–299. https://doi.org/10.1016/0166-3542(90)90009-V

Andrei, G., & De Clercq, E. (1993). Molecular approaches for the treatment of

hemorrhagic fever virus infections. *Antiviral Research*, *22*(1), 45–75. https://doi.org/10.1016/0166-3542(93)90085-W

Baell, J. B., & Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, *53*(7), 2719–2740. https://doi.org/10.1021/jm901137j

Bahir, I., Fromer, M., Prat, Y., & Linial, M. (2009). Viral adaptation to host : a proteome-based analysis of codon usage and amino acid preferences. *Molecular Systems Biology*, *5*(311), 1–14. https://doi.org/10.1038/msb.2009.71

Bakke, J. E., Huwe, J. K., Mulford, D. J., & Bergman, A. (1992). Metabolism of 1,2,4-trichlorobenzene in rats: Examination of thiol formation. *Xenobiotica*, *22*(2), 199–210. https://doi.org/10.3109/00498259209046618

Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research*, *46*(W1), W257–W263. https://doi.org/10.1093/nar/gky318

Barradas, J. S., Errea, M. I., D'Accorso, N. B., Sepúlveda, C. S., & Damonte, E. B. (2011). Imidazo[2,1-b]thiazole carbohydrate derivatives: Synthesis and antiviral activity against Junin virus, agent of Argentine hemorrhagic fever. *European Journal of Medicinal Chemistry*, *46*(1), 259–264. https://doi.org/10.1016/j.ejmech.2010.11.012

Bartolotta, S., García, C. C., Candurra, N. A., & Damonte, E. B. (2001). Effect of fatty acids on arenavirus replication: Inhibition of virus production by lauric acid. *Archives of Virology*, *146*(4), 777–790. https://doi.org/10.1007/s007050170146

Barton, L. L., Mets, M. B., & Beauchamp, C. L. (2002). Lymphocytic choriomeningitis virus: Emerging fetal teratogen. *American Journal of Obstetrics and Gynecology*, *187*(6), 1715–1716. https://doi.org/10.1067/mob.2002.126297

Belalov, I. S., & Lukashev, A. N. (2013). Causes and Implications of Codon Usage Bias in RNA Viruses. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056642

Bell, T. M., Shaia, C. I., Bunton, T. E., Robinson, C. G., Wilkinson, E. R., Hensley, L.

E., & Cashman, K. A. (2015). Pathology of Experimental Machupo Virus Infection, Chicava Strain, in Cynomolgus Macaques (Macaca fascicularis) by Intramuscular and Aerosol Exposure. *Veterinary Pathology*, *52*(1), 26–37. https://doi.org/10.1177/0300985814540544

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, *46*(D1), D41–D47. https://doi.org/10.1093/nar/gkx1094

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank. A computer-based archival file for macromolecular structures. *European journal of biochemistry*, *80*(2), 319–324. https://doi.org/10.1111/j.1432-1033.1977.tb11885.x

Bigot, T., Temmam, S., Pérot, P., & Eloit, M. (2020). *RVDB-prot , a reference viraL database and its HMM profiles [version 2 ; peer review : 2 approved]*. 1–13.

Binkowski, T. A., Naghibzadeh, S., & Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Research*, *31*(13), 3352–3355. https://doi.org/10.1093/nar/gkg512

Biome, L. De, & Bernard, C. (1999). Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(April), 4482–4487.

Birney, E., Clamp, M., & Hubbard, T. (2002). Databases and tools for browsing genomes. *Annual Review of Genomics and Human Genetics*, *3*, 293–310. https://doi.org/10.1146/annurev.genom.3.030502.101529

Bodewes, R., Raj, V. S., Kik, M. J. L., Schapendonk, C. M., Haagmans, B. L., Smits, S. L., & Osterhaus, A. D. M. E. (2013). Updated Phylogenetic Analysis of Arenaviruses Detected in Boid Snakes. *Journal of Virology*, *88*(2), 1399–1400. https://doi.org/10.1128/jvi.02753-13

Bolken, T. C., Laquerre, S., Zhang, Y., Bailey, T. R., Pevear, D. C., Kickner, S. S., Sperzel, L. E., Jones, K. F., Warren, T. K., Lund, S. A., Kirkwood-watts, D. L., King, D. S., Shurtleff, A. C., Guttieri, M. C., Deng, Y., Bleam, M., & Hruby, D. E. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre*

with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January. https://doi.org/10.1016/j.antiviral.2005.10.008

Brinkmann, A., Ergünay, K., Radonić, A., Kocak Tufan, Z., Domingo, C., & Nitsche, A. (2017). Development and preliminary evaluation of a multiplexed amplification and next generation sequencing method for viral hemorrhagic fever diagnostics. *PLoS Neglected Tropical Diseases*, *11*(11). https://doi.org/10.1371/journal.pntd.0006075

Brisse, M. E., & Ly, H. (2019a). Hemorrhagic fever-causing arenaviruses: Lethal pathogens and potent immune suppressors. *Frontiers in Immunology*, *10*(MAR), 10–17. https://doi.org/10.3389/fimmu.2019.00372

Brisse, M. E., & Ly, H. (2019b). Hemorrhagic fever-causing arenaviruses: Lethal pathogens and potent immune suppressors. *Frontiers in Immunology*, *10*(MAR). https://doi.org/10.3389/fimmu.2019.00372

Brisse, M., Huang, Q., Rahman, M., Di, D., Liang, Y., & Ly, H. (2021). RIG-I and MDA5 Protect Mice From Pichinde Virus Infection by Controlling Viral Replication and Regulating Immune Responses to the Infection. *Frontiers in Immunology*, *12*(December), 1–12. https://doi.org/10.3389/fimmu.2021.801811

Buchmeier, M., Adam, E., & Rawls, W. E. (1974). Evidence of Infection by Pichinde. *Infection and Immunity*, *9*(5), 821–823.

Butt, A. M., Nasrullah, I., Qamar, R., & Tong, Y. (2016). Evolution of codon usage in Zika virus genomes is host and vector specific. In *Emerging Microbes and Infections* (Vol. 5, Issue 10, pp. 1–14). Nature Publishing Group. https://doi.org/10.1038/emi.2016.106

Canakoglu, A., Pinoli, P., Bernasconi, A., Melidis, D. P., Ceri, S., Alfonsi, T., Elettronica, D., Bioingegneria, I., Milano, P., & Ponzio, V. (2021). *ViruSurf : an integrated database to investigate viral sequences*. *49*(October 2020), 817–824. https://doi.org/10.1093/nar/gkaa846

Candurra, N. A., & Damonte, E. B. (1997). Effect of inhibitors of the intracellular exocytic pathway on glycoprotein processing and maturation of Junin virus. *Archives of Virology*, *142*(11), 2179–2193.

https://doi.org/10.1007/s007050050234

Cannarrozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G., & Barral, Y. (2010). A role for codon order in translation dynamics. *Cell*, *141*(2), 355–367. https://doi.org/10.1016/j.cell.2010.02.036

Carballal, G., Calello, M. A., Laguens, R. P., & Weissenbacher, M. C. (1987). Tacaribe virus: A new alternative for Argentine hemorrhagic fever vaccine. *Journal of Medical Virology*, *23*(3), 257–263. https://doi.org/10.1002/jmv.1890230308

Casals, J. (1975). *Arenaviruses ’. 140*, 115–140.

Chakraborty, S., Uddin, A., & Choudhury, M. N. (2017). Factors affecting the codon usage bias of SRY gene across mammals. Gene, 630, 13–20. https://doi.org/10.1016/J.GENE.2017.08.003.

Charrel, N., & Lamballerie, X. De. (2008). *Phylogeny of the genus Arenavirus*. https://doi.org/10.1016/j.mib.2008.06.001

Chem, R. F. J. B., & Chem, R. F. J. B. (1993). *Codon usage: mutational bias, translational selection, or both? 1992*, 693–699.

Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., Wu, X., Huang, X., Gu, X., Wang, S., Peng, X., & Yang, G. (2013). Biochemical and Biophysical Research Communications Analysis of codon usage patterns in Taenia pisiformis through annotated transcriptome data. *Biochemical and Biophysical Research Communications*, *430*(4), 1344–1348. https://doi.org/10.1016/j.bbrc.2012.12.078

Clark, L. E., Mahmutovic, S., Raymond, D. D., Dilanyan, T., Koma, T., Manning, J. T., Shankar, S., Levis, S. C., Briggiler, A. M., Enria, D. A., Wucherpfennig, K. W., Paessler, S., & Abraham, J. (2018). Vaccine-elicited receptor-binding site antibodies neutralize two New World hemorrhagic fever arenaviruses. *Nature Communications*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-04271-z

Coller, S. (2007). Hhs Public Health Emergency Countermeasure- Implementation Plan for Chemical, Bilogical and Radiological Ad Nuclear Threat. *Federal Register*, 1–22.

Comeron, J. M., & Aguadé, M. (1998). An evaluation of measures of synonymous codon usage bias. *Journal of Molecular Evolution*, *47*(3), 268–274.

https://doi.org/10.1007/PL00006384

Cordo, S. M., & Candurra, N. A. (2003). Intermediate filament integrity is required for Junin virus replication. *Virus Research*, *97*(1), 47–55. https://doi.org/10.1016/S0168-1702(03)00221-1

Cox, B. D., Stanton, R. A., & Schinazi, R. F. (2015). Predicting Zika virus structural biology: Challenges and opportunities for intervention. *Antiviral Chemistry and Chemotherapy*, *24*(3–4), 118–126. https://doi.org/10.1177/2040206616653873

Currie, G. M. (2018a). Pharmacology, part 1: Introduction to pharmacology and pharmacodynamics. *Journal of Nuclear Medicine Technology*, *46*(2), 81–86. https://doi.org/10.2967/jnmt.117.199588

Currie, G. M. (2018b). Pharmacology, part 2: Introduction to pharmacokinetics. *Journal of Nuclear Medicine Technology*, *46*(3), 221–230. https://doi.org/10.2967/jnmt.117.199638

D'Addato, M., Curti, T., & Pasquinelli, G. (1986). Ultrastructural investigation on Dacron-Gelseal graft implanted in sheep carotid. *Life Support Systems*, *4*(SUPPL. 2), 62–64.

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(October 2016), 1–13. https://doi.org/10.1038/srep42717

Damte, D., Suh, J. W., Lee, S. J., Yohannes, S. B., Hossain, M. A., & Park, S. C. (2013). Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of Mycoplasma hyopneumoniae. *Genomics*, *102*(1), 47–56. https://doi.org/10.1016/j.ygeno.2013.04.011

Das, S., Paul, S., Chatterjee, S., & Dutta, C. (2005). Codon and amino acid usage in two major human pathogens of genus Bartonella - Optimization between replicational-transcriptional selection, translational control and cost minimization. *DNA Research*, *12*(2), 91–102. https://doi.org/10.1093/dnares/12.2.91

De Manzione, N., Salas, R. A., Paredes, H., Godoy, O., Rojas, L., Araoz, F., Fulhorst, C. F., Ksiazek, T. G., Mills, J. N., Ellis, B. A., Peters, C. J., & Tesh, R. B.

(1998). Venezuelan hemorrhagic fever: Clinical and epidemiological studies of 165 cases. *Clinical Infectious Diseases*, *26*(2), 308–313. https://doi.org/10.1086/516299

de Mello Malta, F., Amgarten, D., de Seixas Santos Nastri, A. C., Ho, Y. L., Casadio, L. V. B., Basqueira, M., Selegatto, G., Cervato, M. C., Duarte-Neto, A. N., Higashino, H. R., Medeiros, F. A. F., Gendler, J. L. P. L., Levin, A. S., & Pinho, J. R. R. (2020). Sabiá virus-like mammarenavirus in patient with fatal hemorrhagic fever, Brazil, 2020. *Emerging Infectious Diseases*, *26*(6), 1332–1334. https://doi.org/10.3201/EID2606.200099

Delgado, S., Erickson, B. R., Agudo, R., Blair, P. J., Vallejo, E., Albariño, C. G., Vargas, J., Comer, J. A., Rollin, P. E., Ksiazek, T. G., Olson, J. G., & Nichol, S. T. (2008). Chapare virus, a newly discovered arenavirus isolated from a fatal hemorrhagic fever case in Bolivia. *PLoS Pathogens*, *4*(4), 1–6. https://doi.org/10.1371/journal.ppat.1000047

Drosten, C., Kümmerer, B. M., Schmitz, H., & Günther, S. (2003). Molecular diagnostics of viral hemorrhagic fevers. *Antiviral Research*, *57*(1–2), 61–87. https://doi.org/10.1016/S0166-3542(02)00201-2

Dunham, E. C., Leske, A., Shifflett, K., Watt, A., Feldmann, H., Hoenen, T., & Groseth, A. (2018). Lifecycle modelling systems support inosine monophosphate dehydrogenase (IMPDH) as a pro-viral factor and antiviral target for New World arenaviruses. *Antiviral Research*, *157*, 140–150. https://doi.org/10.1016/j.antiviral.2018.07.009

Durairaj, D. R. (2017). In Silico Drug Design of Thiolactomycin Derivatives Against Mtb-KasA Enzyme to Inhibit Multidrug Resistance of Mycobacterium tuberculosis. *Interdisciplinary Sciences: Computational Life Sciences*. https://doi.org/10.1007/s12539-017-0257-0

Duret, L. (2000). tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends in Genetics*, *16*(7), 287–289. https://doi.org/10.1016/S0168-9525(00)02041-2

Emonet, S., Lemasson, J., & Gonzalez, J. (2006). *Phylogeny and evolution of old world arenaviruses*. *350*, 251–257. https://doi.org/10.1016/j.virol.2006.01.026

Emrani, J., Ahmed, M., Jeffers-francis, L., Teleha, J. C., Mowa, N., Newman, R. H.,

& Thomas, M. D. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January*.

Enria, D. A., Briggiler, A. M., Levis, S., Vallejos, D., Maiztegui, J. I., & Canonico, P. G. (1987). Tolerance and antiviral effect of ribavirin in patients with argentine hemorrhagic fever. *Antiviral Research*, *7*(6), 353–359. https://doi.org/10.1016/0166-3542(87)90017-9

Enria, D. A., Briggiler, A. M., & Sánchez, Z. (2008). Treatment of Argentine hemorrhagic fever. *Antiviral Research*, *78*(1), 132–139. https://doi.org/10.1016/j.antiviral.2007.10.010

Enria, D. A., & Pinheiro, F. (2000). Rodent-Borne emerging viral zoonosis: Hemorrhagic fevers and hantavirus infections in South America. *Infectious Disease Clinics of North America*, *14*(1), 167–184. https://doi.org/10.1016/S0891-5520(05)70223-3

Escalera-Antezana, J. P., Rodriguez-Villena, O. J., Arancibia-Alba, A. W., Alvarado-Arnez, L. E., Bonilla-Aldana, D. K., & Rodríguez-Morales, A. J. (2020). Clinical features of fatal cases of Chapare virus hemorrhagic fever originating from rural La Paz, Bolivia, 2019: A cluster analysis. *Travel Medicine and Infectious Disease*, *36*(February), 101589. https://doi.org/10.1016/j.tmaid.2020.101589

Farthing, M. J. G. (n.d.). *Senegal*, *12*. 639.

Fasinu, P., Pillay, V., Ndesendo, V. M. K., Toit, L. C., & Choonara, Y. E. (2011). *Diverse approaches for the enhancement of oral drug bioavailability*. *209*(January), 185–209. https://doi.org/10.1002/bdd

Fedeli, C., Moreno, H., & Kunz, S. (2018). Novel Insights into Cell Entry of Emerging Human Pathogenic Arenaviruses. *Journal of Molecular Biology*, *430*(13), 1839–1852. https://doi.org/10.1016/j.jmb.2018.04.026

Fenstermacher, D. (2005). Introduction to bioinformatics. *Journal of the American Society for Information Science and Technology*, *56*(5), 440–446. https://doi.org/10.1002/asi.20133

Fever, L. (2021). Hemorrhagic Fevers Caused by Arenaviruses. *Red Book (2021)*, 362–365. https://doi.org/10.1542/9781610025225-part03-ch053

Flint, S. J., Racaniello, V. R., Rall, G. F., Hatziioannou, T., & Skalka, A. M. (2020). *Principles of virology, Volume 2: pathogenesis and control*. John Wiley & Sons.

Forli, S., Huey, R., Pique, M. E., Sanner, M., Goodsell, D. S., & Arthur, J. (2016). *00006565-201002000-00017*. *11*(5), 905–919. https://doi.org/10.1038/nprot.2016.051.Computational

Frey, F. (2017). SPSS (Software) . *The International Encyclopedia of Communication Research Methods*, *November 2016*, 1–2. https://doi.org/10.1002/9781118901731.iecrm0237

Fredrick, K., & Ibba, M. (2010). How the sequence of a gene can tune its translation. *Cell*, *141*(2), 227–229. https://doi.org/10.1016/j.cell.2010.03.033

Fukushi, S., Tani, H., Yoshikawa, T., Saijo, M., & Morikawa, S. (2012). Serological assays based on recombinant viraLs for the diagnosis of arenavirus hemorrhagic fevers. *Viruses*, *4*(10), 2097–2114. https://doi.org/10.3390/v4102097

Fulhorst, C. F., Ksiazek, T. G., Peters, C. J., & Tesh, R. B. (2002). Experimental Infection of the Cane Mouse Zygodontomys brevicauda (Family Muridae) with Guanarito Virus (Arenaviridae), the Etiologic Agent of Venezuelan Hemorrhagic Fever . *The Journal of Infectious Diseases*, *180*(4), 966–969. https://doi.org/10.1086/315029

García, C. C., Candurra, N. A., & Damonte, E. B. (2000). Antiviral and virucidal activities against arenaviruses of zinc-finger active compounds. *Antiviral Chemistry and Chemotherapy*, *11*(3), 231–237. https://doi.org/10.1177/095632020001100306

García, C. C., Candurra, N. A., & Damonte, E. B. (2003). Differential inhibitory action of two azoic compounds against arenaviruses. *International Journal of Antimicrobial Agents*, *21*(4), 319–324. https://doi.org/10.1016/S0924-8579(02)00390-4

Godini, R., & Fallahi, H. (2019). A brief overview of the concepts, methods and computational tools used in phylogenetic tree construction and gene prediction. *Meta Gene*, *21*(April), 100586. https://doi.org/10.1016/j.mgene.2019.100586

Golden, J. W., Beitzel, B., Ladner, J. T., Mucker, E. M., Kwilas, S. A., Palacios, G., & Hooper, J. W. (2017). An attenuated Machupo virus with a disrupted L-

segment intergenic region protects Guinea pigs against lethal Guanarito virus infection. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-04889-x

Gong, Z. J., De Meyer, S., Clarysse, C., Verslype, C., Neyts, J., De Clercq, E., & Yap, S. H. (1999). Mycophenolic acid, an immunosuppressive agent, inhibits HBV replication in vitro. *Journal of Viral Hepatitis*, *6*(3), 229–236. https://doi.org/10.1046/j.1365-2893.1999.00163.x

Gonzalez, J. P., Emonet, S., Lamballerie, X. De, & Charrel, R. (2007). Arenaviruses. *Current Topics in Microbiology and Immunology*, *315*, 253–288. https://doi.org/10.1007/978-3-540-70962-6_11

Gonzalez, J. P. J., Bowen, M. D., Nichol, S. T., & Rico-Hesse, R. (1996). Genetic characterization and phylogeny of Sabia virus, an emergent pathogen in Brazil. *Virology*, *221*(2), 318–324. https://doi.org/10.1006/viro.1996.0381

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, *5*(4), 536–544. https://doi.org/10.1038/s41564-020-0695-z

Gowen, B. B., & Bray, M. (2011). Progress in the experimental therapy of severe arenaviral infections. *Future Microbiology*, *6*(12), 1429–1441. https://doi.org/10.2217/fmb.11.132

Grant, A., Seregin, A., Huang, C., Kolokoltsova, O., Brasier, A., Peters, C., & Paessler, S. (2012). Junín virus pathogenesis and virus replication. *Viruses*, *4*(10), 2317–2339. https://doi.org/10.3390/v4102317

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., Mercier, R., Evolution, G., Biometrie, L. De, & I, U. L. (1981). Nucleic Acids Research Codon catalog usage is a genome strategy modulated for gene expressivity Nucleic Acids Research. *Nucleic Acids Research*, *9*(1).

Grobbelaar, A. A., Jardine, J., Burt, F. J., Shepherd, A. J., Shepherd, S. P., Leman, P. A., Kemp, A., Braack, L., Weyer, J., Paweska, J. T., & Swanepoel, R. (2021). Mammarenaviruses of Rodents, South Africa and Zimbabwe. *Emerging*

infectious diseases, *27*(12), 3092–3102. https://doi.org/10.3201/eid2712.211088

Gu, W., Zhou, T., Ma, J., Sun, X., & Lu, Z. (2004). The relationship between synonymous codon usage and protein structure in Escherichia coli and Homo sapiens. *BioSystems*, *73*(2), 89–97. https://doi.org/10.1016/j.biosystems.2003.10.001

Günther, S., & Lenz, O. (2004). Lassa virus. In *Critical Reviews in Clinical Laboratory Sciences* (Vol. 41, Issue 4). https://doi.org/10.1080/10408360490497456

Guo, F. B., & Yu, X. J. (2007). Separate base usages of genes located on the leading and lagging strands in Chlamydia muridarum revealed by the Z curve method. *BMC Genomics*, *8*, 1–8. https://doi.org/10.1186/1471-2164-8-366

Guo, F. B., & Yuan, J. B. (2009). Codon usages of genes on chromosome, and surprisingly, genes in plasmid are primarily affected by strand-specific mutational biases in lawsonia intracellularis. *DNA Research*, *16*(2), 91–104. https://doi.org/10.1093/dnares/dsp001

Guo, H., Pan, X., Mao, R., Zhang, X., Wang, L., Lu, X., Chang, J., Guo, J. T., Passic, S., Krebs, F. C., Wigdahl, B., Warren, T. K., Retterer, C. J., Bavari, S., Xu, X., Cuconati, A., & Block, T. M. (2011). Alkylated porphyrins have broad antiviral activity against hepadnaviruses, flaviviruses, filoviruses, and arenaviruses. *Antimicrobial Agents and Chemotherapy*, *55*(2), 478–486. https://doi.org/10.1128/AAC.00989-10

Gurung, A. B., Ali, M. A., Lee, J., Abul Farah, M., & Al-Anazi, K. M. (2020). In silico screening of FDA approved drugs reveals ergotamine and dihydroergotamine as potential coronavirus main protease enzyme inhibitors. *Saudi Journal of Biological Sciences*, *27*(10), 2674–2682. https://doi.org/10.1016/j.sjbs.2020.06.005

Hallam, S. J., Koma, T., Maruyama, J., & Paessler, S. (2018). Review of mammarenavirus biology and replication. *Frontiers in Microbiology*, *9*(AUG), 1–8. https://doi.org/10.3389/fmicb.2018.01751

Hass, M., Gölnitz, U., Müller, S., Becker-Ziaja, B., & Günther, S. (2004). Replicon System for Lassa Virus. *Journal of Virology*, *78*(24), 13793–13803. https://doi.org/10.1128/jvi.78.24.13793-13803.2004

Hesper, B., & Hogeweg, P. (2021). *Bio-informatics: a working concept. A translation of "Bio-informatica: een werkconcept" by B. Hesper and P. Hogeweg*. *November*, 28–29. http://arxiv.org/abs/2111.11832

Hu, J. song, Wang, Q. qin, Zhang, J., Chen, H. tai, Xu, Z. wen, Zhu, L., Ding, Y. zhong, Ma, L. na, Xu, K., Gu, Y. xing, & Liu, Y. sheng. (2011). The characteristic of codon usage pattern and its evolution of hepatitis C virus. *Infection, Genetics and Evolution*, *11*(8), 2098–2102. https://doi.org/10.1016/j.meegid.2011.08.025

Huang, C., Kolokoltsova, O. A., Yun, N. E., Seregin, A. V., Ronca, S., Koma, T., & Paessler, S. (2015). Highly Pathogenic New World and Old World Human Arenaviruses Induce Distinct Interferon Responses in Human Cells. *Journal of virology*, *89*(14), 7079–7088. https://doi.org/10.1128/JVI.00526-15

Hugh, S., & Jackson, D. (1994). *Pharmacodynamics in the elderly*. *87*(23), 5–7.

Ikemura, T. (1985). *Codon Usage and tRNA Content in Unicellular and Multicellular Organisms '*. *2*, 13–34.

Jenkins, G. M., & Holmes, E. C. (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Research*, *92*(1), 1–7. https://doi.org/10.1016/S0168-1702(02)00309-X

John, T. J., & Roberts, J. M. (n.d.). *age-specific poliovirus antibody prevalence*. 873–874.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993. https://doi.org/10.1038/nature06536

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Garcia Diez, F., Harte, N., Kulikova, T., Lin, Q., Lombard, V., … Apweiler, R. (2005). The EMBL nucleotide sequence database. *Nucleic Acids Research*, *33*(DATABASE ISS.), 29–33. https://doi.org/10.1093/nar/gki098

Kar, P., Kumar, V., Vellingiri, B., Sen, A., Jaishee, N., Anandraj, A., Malhotra, H., Bhattacharyya, S., Mukhopadhyay, S., Kinoshita, M., Govindasamy, V., Roy, A., Naidoo, D., & Subramaniam, M. D. (2020). Anisotine and amarogentin as

promising inhibitory candidates against SARS-CoV-2 proteins: a computational investigation. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–11. https://doi.org/10.1080/07391102.2020.1860133

Kar, P., Sharma, N. R., Singh, B., Sen, A., & Roy, A. (2021). Natural compounds from Clerodendrum spp. as possible therapeutic candidates against SARS-CoV-2: An in silico investigation. *Journal of Biomolecular Structure and Dynamics*, *39*(13), 4774–4785. https://doi.org/10.1080/07391102.2020.1780947

Karesh, W. B., Dobson, A., Lloyd-smith, J. O., Lubroth, J., Dixon, M. A., Bennett, M., Aldrich, S., Harrington, T., Formenty, P., Loh, E. H., Machalaba, C. C., Thomas, M. J., & Heymann, D. L. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- company ' s public news and information website . Elsevier hereby grants permission to make all its COVID-19-r. January*.

Karlin, S., & Mrázek, J. (1996). What drives codon choices in human genes?. *Journal of molecular biology*, *262*(4), 459–472. https://doi.org/10.1006/jmbi.1996.0528

Kerber, R., Reindl, S., Romanowski, V., Gómez, R. M., Ogbaini-Emovon, E., Günther, S., & ter Meulen, J. (2015). Research efforts to control highly pathogenic arenaviruses: A summary of the progress and gaps. *Journal of Clinical Virology*, *64*, 120–127. https://doi.org/10.1016/j.jcv.2014.12.004

Khan, M. T., Ali, A., Wang, Q., Irfan, M., Khan, A., Zeb, M. T., Zhang, Y. J., Chinnasamy, S., & Wei, D. Q. (2020). Marine natural compounds as potents inhibitors against the main protease of SARS-CoV-2—a molecular dynamic study. *Journal of Biomolecular Structure and Dynamics*, *39*(10), 1–11. https://doi.org/10.1080/07391102.2020.1769733

Khandia, R., Singhal, S., Kumar, U., Ansari, A., Tiwari, R., Dhama, K., Das, J., Munjal, A. O., & Singh, R. K. (2019). Analysis of nipah virus codon usage and adaptation to hosts. *Frontiers in Microbiology*, *10*(MAY). https://doi.org/10.3389/fmicb.2019.00886

Khandia, R., Singhal, S., Kumar, U., Ansari, A., Tiwari, R., Dhama, K., Das, J., Munjal, A., & Singh, R. K. (2019). Analysis of Nipah Virus Codon Usage and Adaptation to Hosts. *Frontiers in Microbiology*, *10*(May), 1–18.

https://doi.org/10.3389/fmicb.2019.00886

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2016). PubChem substance and compound databases. *Nucleic Acids Research*, *44*(D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951

Knox, R. J., Friedlos, F., Biggs, P. J., Flitter, W. D., Gaskell, M., Goddard, P., Davies, L., & Jarman, M. (1993). Identification, synthesis and properties of 5-(aziridin-1-YL)-2-nitro-4-nitrosobenzamide, A novel DNA crosslinking agent derived from CB1954. *Biochemical Pharmacology*, *46*(5), 797–803. https://doi.org/10.1016/0006-2952(93)90487-H

Koma, T., Patterson, M., Huang, C., Seregin, A. V., Maharaj, P. D., Miller, M., Smith, J. N., Walker, A. G., Hallam, S., & Paessler, S. (2016). Machupo Virus Expressing GPC of the Candid#1 Vaccine Strain of Junin Virus Is Highly Attenuated and Immunogenic. *Journal of Virology*, *90*(3), 1290–1297. https://doi.org/10.1128/jvi.02615-15

Koonin, E. V., Krupovic, M., & Yutin, N. (2015). Evolution of double-stranded DNA viruses of eukaryotes: From bacteriophages to transposons to giant viruses. *Annals of the New York Academy of Sciences*, *1341*(1), 10–24. https://doi.org/10.1111/nyas.12728

Koonin, E. V, Dolja, V. V, & Krupovic, M. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January*.

Kumar, C. S., & Kumar, S. (2017). Synonymous codon usage of genes in polymerase complex of Newcastle disease virus. *Journal of Basic Microbiology*, *57*(6), 481–503. https://doi.org/10.1002/jobm.201600740

Kumar, N., Bera, B. C., Greenbaum, B. D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B. N., & Virmani, N. (2016). Revelation of Influencing Factors in Overall Codon Usage Bias of Equine Influenza Viruses. *PloS One*, *11*(4), e0154376. https://doi.org/10.1371/journal.pone.0154376

Kumar, V., Dhanjal, J. K., Kaul, S. C., Wadhwa, R., & Sundar, D. (2020). Withanone

and caffeic acid phenethyl ester are predicted to interact with main protease (Mpro) of SARS-CoV-2 and inhibit its activity. *Journal of Biomolecular Structure and Dynamics*, *39*(11), 1–13. https://doi.org/10.1080/07391102.2020.1772108

Kunz, S. (2009). The role of the vascular endothelium in arenavirus haemorrhagic fevers. *Thrombosis and Haemostasis*, *102*(6), 1024–1029. https://doi.org/10.1160/TH09-06-0357.

Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 51, 2778–2786.
https://doi. org/10.1021/ci200227u.

Lee, J. Y., Han, G. G., Kim, E. B., & Choi, Y. J. (2017). Comparative genomics of Lactobacillus salivarius strains focusing on their host adaptation. *Microbiological Research*, *205*(April), 48–58. https://doi.org/10.1016/j.micres.2017.08.008

Lee, K. J., Novella, I. S., Teng, M. N., Oldstone, M. B. A., & de la Torre, J. C. (2000). NP and Ls of Lymphocytic Choriomeningitis Virus (LCMV) Are Sufficient for Efficient Transcription and Replication of LCMV Genomic RNA Analogs. *Journal of Virology*, *74*(8), 3470–3477. https://doi.org/10.1128/jvi.74.8.3470-3477.2000

Li, X., Song, H., Kuang, Y., Chen, S., Tian, P., Li, C., & Nan, Z. (2016). Genome-Wide Analysis of Codon Usage Bias in Epichloë festucae. *International journal of molecular sciences*, *17*(7), 1138. https://doi.org/10.3390/ijms17071138

Li, C., Vermeersch, S., Denney, W. S., Kennedy, W. P., Palcza, J., Gipson, A., Han, T. H., Blanchard, R., Lepeleire, I. De, Depré, M., Murphy, M. G., Dyck, K. Van, & Hoon, J. N. De. (2014). *Characterizing the PK / PD relationship for inhibition of capsaicin-induced dermal vasodilatation by MK-3207 , an oral calcitonin gene related peptide receptor antagonist*. 831–837. https://doi.org/10.1111/bcp.12547

Li, G., Zhang, W., Wang, R., Xing, G., Wang, S., Ji, X., Wang, N., Su, S., & Zhou, J. (2019). Genetic Analysis and Evolutionary Changes of the Torque teno sus Virus. *International Journal of Molecular Sciences*, *20*(12), 2881. https://doi.org/10.3390/ijms20122881

Li, K., Lin, X. D., Wang, W., Shi, M., Guo, W. P., Zhang, X. H., Xing, J. G., He, J. R., Wang, K., Li, M. H., Cao, J. H., Jiang, M. L., Holmes, E. C., & Zhang, Y. Z. (2015). Isolation and characterization of a novel arenavirus harbored by Rodents and Shrews in Zhejiang province, China. *Virology*, *476*, 37–42. https://doi.org/10.1016/j.virol.2014.11.026

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. *46*, 3–26.

Lobry, J. R., & Gautier, C. (1994). Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 escherichia coli chromosome-encoded genes. *Nucleic Acids Research*, *22*(15), 3174–3180. https://doi.org/10.1093/nar/22.15.3174

Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J., & Jiggins, F. M. (2014). The Evolution and Genetics of Virus Host Shifts. *PLoS Pathogens*, *10*(11). https://doi.org/10.1371/journal.ppat.1004395

López, N., Jácamo, R., & Franze-Fernández, M. T. (2001). Transcription and RNA Replication of Tacaribe Virus Genome and Antigenome Analogs Require N and Ls: Z Is an Inhibitor of These Processes. *Journal of Virology*, *75*(24), 12241–12251. https://doi.org/10.1128/jvi.75.24.12241-12251.2001

Lozano, M. E., Enria, D., Maiztegui, J. I., Grau, O., & Romanowski, V. (1995). Rapid diagnosis of argentine hemorrhagic fever by reverse transcriptase PCR-based assay. *Journal of Clinical Microbiology*, *33*(5), 1327–1332. https://doi.org/10.1128/jcm.33.5.1327-1332.1995

Lu, J., Han, Z., Liu, Y., Liu, W., Lee, M. S., Olson, M. A., Ruthel, G., Freedman, B. D., & Harty, R. N. (2014). A Host-Oriented Inhibitor of Junin Argentine Hemorrhagic Fever Virus Egress. *Journal of Virology*, *88*(9), 4736–4743. https://doi.org/10.1128/jvi.03757-13

Ly, H., & Liang, Y. (2015). Immune evasion mechanisms of arenaviruses. *Oncotarget*, *6*(38), 40386–40387. https://doi.org/10.18632/oncotarget.6367

Mackenzie, J. S., & Jeggo, M. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on*

Elsevier Connect , the company ' s public news and information . January.

Ma, J., Zhang, S., & Zhang, X. (2021). Structure of Machupo virus polymerase in complex with matrix protein Z. *Nature communications*, *12*(1), 6163. https://doi.org/10.1038/s41467-021-26432-3

Maes, P., Alkhovsky, S. V., Bào, Y., Beer, M., Birkhead, M., Briese, T., Buchmeier, M. J., Calisher, C. H., Charrel, R. N., Choi, I. R., Clegg, C. S., de la Torre, J. C., Delwart, E., DeRisi, J. L., Di Bello, P. L., Di Serio, F., Digiaro, M., Dolja, V. V., Drosten, C., … Kuhn, J. H. (2018). Taxonomy of the family Arenaviridae and the order Bunyavirales: update 2018. *Archives of Virology*, *163*(8), 2295–2310. https://doi.org/10.1007/s00705-018-3843-5

Mali, S. N., & Chaudhari, H. K. (2019). Molecular modelling studies on adamantane-based Ebola virus GP-1 inhibitors using docking, pharmacophore and 3D-QSAR. *SAR and QSAR in Environmental Research*, *30*(3), 161–180. https://doi.org/10.1080/1062936X.2019.1573377

Malhotra, H., & Kumar, A. (2021). Codon usage signatures in Sabia and Chapare for host adaptation. *Bioinformation*, *17*(10), 891–898. https://doi.org/10.6026/97320630017891

Malhotra, H., Kumar, A. & Afaq, Y. (2022). Molecular docking analysis of FDA approved drugs with the glycoprotein from Junin and Machupo viruses. *Bioinformation*, *18*(2), 119–126. https://doi.org/10.6026/97320630018119

Mammarenavirus, W., Virus, T., Ye, C., & De, C. (2020). *crossm Development of Reverse Genetics for the Prototype New*. *94*(L), 1–16.

Manuscript, A. (2005). *Function Prediction*. *5*(4), 45–67. https://doi.org/10.1201/9781420035001.ch3

Manuscript, A. (2008). 基因的改变NIH Public Access. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Manzione, N. De, Salas, R. A., Paredes, H., Godoy, O., Rojas, L., Araoz, F., Fulhorst, C. F., Ksiazek, T. G., Mills, J. N., Ellis, B. A., Peters, C. J., & Tesh, R. B. (1989). *Venezuelan Hemorrhagic Fever : Clinical and Epidemiological Studies of 165 Cases*. 308–313.

Martínez-Sobrido, L., & de la Torre, J. C. (2016). Reporter-expressing, replicating-competent recombinant arenaviruses. *Viruses*, *8*(7), 1–20.

https://doi.org/10.3390/v8070197

Martı, A. J., Wells, J. M., & Campa, A. G. De. (2004). *Relationship between codon biased genes , microarray expression values and physiological characteristics of Streptococcus pneumoniae Printed in Great Britain*. 2313–2325. https://doi.org/10.1099/mic.0.27097-0

Maskin, L. (1996). *Ql*. *31*, 149–158.

Mashima, J., Kodama, Y., Fujisawa, T., Katayama, T., Okuda, Y., Kaminuma, E., Ogasawara, O., Okubo, K., Nakamura, Y., & Takagi, T. (2017). DNA Data Bank of Japan. *Nucleic acids research*, *45*(D1), D25–D31. https://doi.org/10.1093/nar/gkw1001

McInerney J. O. (1998). Replicational and transcriptional selection on codon usage in Borrelia burgdorferi. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10698–10703. https://doi.org/10.1073/pnas.95.18.10698

McKee, K. T., Huggins, J. W., Trahan, C. J., & Mahlandt, B. G. (1988). Ribavirin prophylaxis and therapy for experimental Argentine hemorrhagic fever. *Antimicrobial Agents and Chemotherapy*, *32*(9), 1304–1309. https://doi.org/10.1128/AAC.32.9.1304

McKee, K. T., Mahlandt, B. G., Maiztegui, J. I., Eddy, G. A., & Peters, C. J. (1985). Experimental Argentine hemorrhagic fever in rhesus macaques: Viral strain-dependent clinical response. *Journal of Infectious Diseases*, *152*(1), 218–221. https://doi.org/10.1093/infdis/152.1.218

McKee, K. T., Mahlandt, B. G., Maiztegui, J. I., Green, D. E., & Peters, C. J. (1987). Virus-specific factors in experimental Argentine hemorrhagic fever in rhesus macaques. *Journal of Medical Virology*, *22*(2), 99–111. https://doi.org/10.1002/jmv.1890220202

McLay, L., Liang, Y., & Ly, H. (2014). Comparative analysis of disease pathogenesis and molecular mechanisms of New World and Old World arenavirus infections. *Journal of General Virology*, *95*(PART 1), 1–15. https://doi.org/10.1099/vir.0.057000-0

Methods, C. D. D. (2017). *Antibiotics*. *1520*, 85–106. https://doi.org/10.1007/978-1-4939-6634-9

Meyer, A. G., Sawyer, S. L., Ellington, A. D., & Wilke, C. O. (2014). Analyzing machupo virus-receptor binding by molecular dynamics simulations. *PeerJ*, *2*, e266. https://doi.org/10.7717/peerj.266

Meyer, B. J., De la Torre, J. C., & Southern, P. J. (2002). Arenaviruses: Genomic RNAs, transcription, and replication. *Current Topics in Microbiology and Immunology*, *262*, 139–157. https://doi.org/10.1007/978-3-642-56029-3_6

Moraz, M. L., & Kunz, S. (2011). Pathogenesis of arenavirus hemorrhagic fevers. *Expert Review of Anti-Infective Therapy*, *9*(1), 49–59. https://doi.org/10.1586/eri.10.142

Mori, T. (2009). *Reemergence of Bolivian Hemorrhagic Fever ,*. *15*(9), 2005–2007. https://doi.org/10.3201/eid1509.081336

Moriyama, E. N., & Powell, J. R. (1997). Codon usage bias and tRNA abundance in Drosophila. *Journal of Molecular Evolution*, *45*(5), 514–523. https://doi.org/10.1007/PL00006256

Murphy, F. A., Webb, P. A., Johnson, K. M., Whitfield, S. G., & Chappell, W. A. (1970). Arenoviruses in Vero Cells: Ultrastructural Studies. *Journal of Virology*, *6*(4), 507–518. https://doi.org/10.1128/jvi.6.4.507-518.1970

Nakamura, Y., Gojobori, T., & Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Research*, *28*(1), 292. https://doi.org/10.1093/nar/28.1.292

Nakauchi, M., Fukushi, S., Saijo, M., Mizutani, T., Ure, A. E., Romanowski, V., Kurane, I., & Morikawa, S. (2009). Characterization of monoclonal antibodies to junin virus nucleocapsid protein and application to the diagnosis of hemorrhagic fever caused by South American arenaviruses. *Clinical and Vaccine Immunology*, *16*(8), 1132–1138. https://doi.org/10.1128/CVI.00163-09

Nasrullah, I., Butt, A. M., Tahir, S., Idrees, M., & Tong, Y. (2015). Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evolutionary Biology*, *15*(1), 1–15. https://doi.org/10.1186/s12862-015-0456-4

Naya, H., Romero, H., Carels, N., Zavala, A., & Musto, H. (2001). Translational selection shapes codon usage in the GC-rich genome of Chlamydomonas reinhardtii. *FEBS Letters*, *501*(2–3), 127–130. https://doi.org/10.1016/S0014-

5793(01)02644-8

Nie, X., Deng, P., Feng, K., Liu, P., Du, X., You, F. M., & Weining, S. (2014). Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Molecular Biology Reporter*, *32*(4), 828–840. https://doi.org/10.1007/s11105-013-0691-z

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An Open chemical toolbox. *Journal of Cheminformatics*, *3*(10), 33. https://doi.org/10.1186/1758-2946-3-33

Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F., & Osawa, S. (1987). The ribosomaL gene cluster of Mycoplasma capricolum. *MGG Molecular & General Genetics*, *210*(2), 314–322. https://doi.org/10.1007/BF00325700

Organisms, U. (1986). *An Evolutionary Perspective on Synonymous Codon Usage in*. 28–38.

Ortiz-Riano, E., Ngo, N., Devito, S., Eggink, D., Munger, J., Shaw, M. L., de la Torre, J. C., & Martinez-Sobrido, L. (2014). Inhibition of Arenavirus by A3, a Pyrimidine Biosynthesis Inhibitor. *Journal of Virology*, *88*(2), 878–889. https://doi.org/10.1128/jvi.02275-13

Pakula, A. (2019). 乳鼠心肌提取 HHS Public Access. *Methods Molecular Biology*, *176*(5), 139–148. https://doi.org/10.1016/j.semcdb.2018.02.002.Modulation

Palareti, G., Legnani, C., Cosmi, B., Antonucci, E., Erba, N., Poli, D., Testa, S., & Tosetto, A. (2016). Comparison between different D-Dimer cutoff values to assess the individual risk of recurrent venous thromboembolism: Analysis of results obtained in the DULCIS study. *International Journal of Laboratory Hematology*, *38*(1), 42–49. https://doi.org/10.1111/ijlh.12426

Papageorgiou, N., Spiliopoulou, M., Van Nguyen, T. H., Vaitsopoulou, A., Laban, E. Y., Alvarez, K., Margiolaki, I., Canard, B., & Ferron, F. (2020). Brothers in arms: Structure, assembly and function of arenaviridae NP. *Viruses*, *12*(7). https://doi.org/10.3390/v12070772

Parvathy, S. T., Udayasuriyan, V., & Bhadana, V. (2022). Codon usage bias. *Molecular Biology Reports*, *49*(1), 539–565. https://doi.org/10.1007/s11033-021-06749-4

Paweska, J. T., Sewlall, N. H., Ksiazek, T. G., Blumberg, L. H., Hale, M. J., Lipkin,

W. I., Weyer, J., Nichol, S. T., Rollin, P. E., McMullan, L. K., Paddock, C. D., Briese, T., Mnyaluza, J., Dinh, T. H., Mukonka, V., Ching, P., Duse, A., Richards, G., De Jong, G., … Wamunyima, L. (2009). Nosocomial outbreak of novel arenavirus infection, Southern Africa. *Emerging Infectious Diseases*, *15*(10), 1598–1602. https://doi.org/10.3201/eid1510.090211

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(8), 2444–2448. https://doi.org/10.1073/pnas.85.8.2444

Peden JF (1999). Analysis of codon usage, University of Nottingham, UK. **PhD**.

Pellett, P. E., Mitra, S., & Holland, T. C. (2014). Basics of virology. In *Handbook of Clinical Neurology* (1st ed., Vol. 123). Elsevier B.V. https://doi.org/10.1016/B978-0-444-53488-0.00002-X

Peters, C. J. (2002). Human infection with arenaviruses in the Americas. *Current Topics in Microbiology and Immunology*, *262*, 65–74. https://doi.org/10.1007/978-3-642-56029-3_3

Peters, C. J., Salas, R. A., Duno, G., De Miller, E., Utrera, A., Ksiazek, T. G., Fulhorst, C. F., Bowen, M. D., Tesh, R. B., Vasquez, C., & De Manzione, N. M. (2017). Natural rodent host associations of Guanarito and pirital viruses (Family Arenaviridae) in central Venezuela. *The American Journal of Tropical Medicine and Hygiene*, *61*(2), 325–330. https://doi.org/10.4269/ajtmh.1999.61.325

Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., & Scheuermann, R. H. (2012). ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research*, *40*(D1), 593–598. https://doi.org/10.1093/nar/gkr859

Pires, D. E. V., Blundell, T. L., & Ascher, D. B. (2015). pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry*, *58*(9), 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

Plewe, M. B., Whitby, L. R., Naik, S., Brown, E. R., Nadezda, V., Gantla, V. R., York, J., Nunberg, J. H., Zhang, L., Kalveram, B., Freiberg, A. N., Boger, D. L., Henkel, G., Diego, S., States, U., States, U., Road, P., Jolla, L., States, U., …

States, U. (2020). *HHS Public Access*. *29*(22), 1–35. https://doi.org/10.1016/j.bmcl.2019.08.024.SAR

Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics*, *12*(1), 32–42. https://doi.org/10.1038/nrg2899

Polyak, S. J., Rawls, W. E., & Harnish, D. G. (1991). Characterization of Pichinde virus infection of cells of the monocytic lineage. *Journal of Virology*, *65*(7), 3575–3582. https://doi.org/10.1128/jvi.65.7.3575-3582.1991

Puigbò, P., Aragonès, L., & Garcia-Vallvé, S. (2010). RCDI/eRCDI: A web-server to estimate codon usage deoptimization. *BMC Research Notes*, *3*, 2–5. https://doi.org/10.1186/1756-0500-3-87

Puigbò, P., Bravo, I. G., & Garcia-Vallve, S. (2008). CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct*, *3*, 1–8. https://doi.org/10.1186/1745-6150-3-38

Pyle, J. D., & Whelan, S. P. J. (2019). RNA ligands activate the Machupo virus polymerase and guide promoter usage. *Proceedings of the National Academy of Sciences*, *116*(21), 10518–10524. https://doi.org/10.1073/pnas.1900790116

Raaben, M., Jae, L. T., Herbert, A. S., Kuehne, A. I., Stubbs, S. H., Chou, Y., Blomen, V. A., Kirchhausen, T., Dye, J. M., & Thijn, R. (2018). *HHS Public Access*. *22*(5), 688–696. https://doi.org/10.1016/j.chom.2017.10.002.NRP2

Radoshitzky, S. R., Abraham, J., Spiropoulou, C. F., Kuhn, J. H., Nguyen, D., Li, W., Nagel, J., Schmidt, P. J., Nunberg, J. H., Andrews, N. C., Farzan, M., & Choe, H. (2007). Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature*, *446*(7131), 92–96. https://doi.org/10.1038/nature05539

Radoshitzky, S. R., Bào, Y., Buchmeier, M. J., Charrel, R. N., Clawson, A. N., Clegg, C. S., DeRisi, J. L., Emonet, S., Gonzalez, J. P., Kuhn, J. H., Lukashevich, I. S., Peters, C. J., Romanowski, V., Salvato, M. S., Stenglein, M. D., & de la Torre, J. C. arlo. (2015). Past, present, and future of arenavirus taxonomy. *Archives of Virology*, *160*(7), 1851–1874. https://doi.org/10.1007/s00705-015-2418-y

Radoshitzky, S. R., Buchmeier, M. J., Charrel, R. N., Clegg, J. C. S., Gonzalez, J. P. J., Günther, S., Hepojoki, J., Kuhn, J. H., Lukashevich, I. S., Romanowski, V.,

Salvato, M. S., Sironi, M., Stenglein, M. D., & De La Torre, J. C. (2019). ICTV virus taxonomy profile: Arenaviridae. *Journal of General Virology*, *100*(8), 1200–1201. https://doi.org/10.1099/jgv.0.001280

Ramaiah, A., Dai, L., Contreras, D., Sinha, S., Sun, R., & Arumugaswami, V. (2017). Comparative analysis of protein evolution in the genome of pre-epidemic and epidemic Zika virus. *Infection, Genetics and Evolution*, *51*, 74–85. https://doi.org/10.1016/j.meegid.2017.03.012

Rathbun, J. Y., Droniou, M. E., Damoiseaux, R., Haworth, K. G., Henley, J. E., Exline, C. M., Choe, H., & Cannon, P. M. (2015). *Novel Arenavirus Entry Inhibitors Discovered by Using a Minigenome Rescue System for High-Throughput Drug Screening*. *89*(16), 8428–8443. https://doi.org/10.1128/JVI.00997-15

Reis, M., Wernisch, L., & Savva, R. (2003). *Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome*. *31*(23), 6976–6985. https://doi.org/10.1093/nar/gkg897

Ren, J., Huangfu, Y., Ge, J., Wu, B., Li, W., Wang, X., & Zhao, L. (2020). Computational study on natural compounds inhibitor of c-Myc. *Medicine*, *99*(50), e23342. https://doi.org/10.1097/MD.0000000000023342

Roccatagliata, A. J., Maier, M. S., Seldes, A. M., Pujol, C. A., & Damonte, E. B. (1996). Antiviral sulfated steroids from the ophiuroid Ophioplocus januarii. *Journal of Natural Products*, *59*(9), 887–889. https://doi.org/10.1021/np960171a

Rojek, J. M., Spiropoulou, C. F., & Kunz, S. (2006). Characterization of the cellular receptors for the South American hemorrhagic fever viruses Junin, Guanarito, and Machupo. *Virology*, *349*(2), 476–491. https://doi.org/10.1016/j.virol.2006.02.033

Romero, H., Zavala, A., Musto, H., & Bernardi, G. (2003). The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene*, *317*(1–2), 141–147. https://doi.org/10.1016/S0378-1119(03)00701-7

Romero, H., Zavala, A., & Musto, H. (2000). Codon usage in Chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic acids research*, *28*(10), 2084–2090.

https://doi.org/10.1093/nar/28.10.2084

Rosa, T., Vasconcelos, P. P. F. C., Pinheiro, F. P., Leduc, J. W., Rico-hesse, R., Gonzalez, J., Jahrling, P. B., & Tesh, P. R. B. (2012). *NIH Public Access*. *343*(8894), 391–392.

Roy, A., Banerjee, R., & Basak, S. (2017). HIV progression depends on codon and amino acid usage profile of envelope protein and associated host-genetic influence. *Frontiers in Microbiology*, *8*(JUN). https://doi.org/10.3389/fmicb.2017.01083

Roy, A., Guo, F., Singh, B., Gupta, S., Paul, K., Chen, X., Sharma, N. R., Jaishee, N., Irwin, D. M., & Shen, Y. (2021). Base Composition and Host Adaptation of the SARS-CoV-2: Insight From the Codon Usage Perspective. *Frontiers in Microbiology*, *12*(December 2019), 1–11. https://doi.org/10.3389/fmicb.2021.548275

Roy, A., Mukhopadhyay, S., Sarkar, I., & Sen, A. (2015). Comparative investigation of the various determinants that influence the codon and amino acid usage patterns in the genus Bifidobacterium. *World Journal of Microbiology and Biotechnology*, *31*(6), 959–981. https://doi.org/10.1007/s11274-015-1850-1

Roy, A., Sen, A., Chakrobarty, S., & Sarkar, I. (2018). Comprehensive profiling of functional attributes , virulence potential and evolutionary dynamics in mycobacterial secretomes. *World Journal of Microbiology and Biotechnology*, *34*(1), 1–19. https://doi.org/10.1007/s11274-017-2388-1

Roy, A., & van Staden, J. (2019). Insights into the riddles of codon usage patterns and codon context signatures in fungal genus Puccinia, a persistent threat to global agriculture. *Journal of Cellular Biochemistry*, *120*(12), 19555–19566. https://doi.org/10.1002/jcb.29263

Ruo, S. L., Mitchell, S. W., Kiley, M. P., Roumillat, L. F., Fisher-Hoch, S. P., & McCormick, J. B. (1991). Antigenic relatedness between arenaviruses defined at the epitope level by monoclonal antibodies. *Journal of General Virology*, *72*(3), 549–555. https://doi.org/10.1099/0022-1317-72-3-549

Salas, R., Pacheco, M. E., Ramos, B., Taibo, M. E., Jaimes, E., Vasquez, C., Querales, J., de Manzione, N., Godoy, O., Betancourt, A., Araoz, F., Bruzual, R., Garcia, J., Tesh, R. B., Rico-Hesse, R., & Shops, R. E. (1991). Venezuelan

haemorrhagic fever. *The Lancet*, *338*(8774), 1033–1036. https://doi.org/10.1016/0140-6736(91)91899-6

Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Research*, *43*(W1), W443–W447. https://doi.org/10.1093/nar/gkv315

Sanchez, A., Pifat, D. Y., Kenyon, R. H., Peters, C. J., McCormick, J. B., & Kiley, M. P. (1989). Junin virus monoclonal antibodies: Characterization and cross-reactivity with other arenaviruses. *Journal of General Virology*, *70*(5), 1125–1132. https://doi.org/10.1099/0022-1317-70-5-1125

Sarute, N., & Ross, S. R. (2020). *HHS Public Access*. *4*(1), 141–158. https://doi.org/10.1146/annurev-virology-101416-042001.New

Schrauwen, E. J. A., & Fouchier, R. A. M. (2014). Host adaptation and transmission of influenza A viruses in mammals. *Emerging Microbes and Infections*, *3*(000), 0. https://doi.org/10.1038/emi.2014.9

Schüttelkopf, A. W., & Van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, *60*(8), 1355–1363. https://doi.org/10.1107/S0907444904011679

Schyman, P., Liu, R., Desai, V., & Wallqvist, A. (2017). vNN web server for ADMET predictions. *Frontiers in Pharmacology*, *8*(DEC), 1–14. https://doi.org/10.3389/fphar.2017.00889

Sharp, P. M., & Li, W. H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of molecular evolution*, *24*(1-2), 28–38. https://doi.org/10.1007/BF02099948

Searls, D. B. (2000). *B t w g*.

Seeliger, D., Groot, B. L. De, & Pymol, V. (2010). *Ligand docking and binding site analysis with PyMOL and Autodock / Vina*. 417–422. https://doi.org/10.1007/s10822-010-9352-6

Sen, A., Sur, S., Bothra, A. K., Benson, D. R., Normand, P., & Tisa, L. S. (2008). The implication of life style on codon usage patterns and predicted highly expressed genes for three Frankia genomes. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *93*(4), 335–346.

https://doi.org/10.1007/s10482-007-9211-1

Sepúlveda, C. S., Fascio, M. L., Mazzucco, M. B., Docampo Palacios, M. L., Pellón, R. F., García, C. C., D'Accorso, N. B., & Damonte, E. B. (2008). Synthesis and evaluation of N-substituted acridones as antiviral agents against haemorrhagic fever viruses. *Antiviral Chemistry and Chemotherapy*, *19*(1), 41–47. https://doi.org/10.1177/095632020801900106

Sepúlveda, C. S., García, C. C., & Damonte, E. B. (2010). Inhibition of arenavirus infection by thiuram and aromatic disulfides. *Antiviral Research*, *87*(3), 329–337. https://doi.org/10.1016/j.antiviral.2010.06.005

Sepúlveda, C. S., García, C. C., & Damonte, E. B. (2018). Antiviral activity of A771726, the active metabolite of leflunomide, against Junín virus. *Journal of Medical Virology*, *90*(5), 819–827. https://doi.org/10.1002/jmv.25024

Sepúlveda, C. S., García, C. C., Levingston Macleod, J. M., López, N., & Damonte, E. B. (2013). Targeting of arenavirus RNA synthesis by a carboxamide-derivatized aromatic disulfide with virucidal activity. *PLoS ONE*, *8*(11), 1–11. https://doi.org/10.1371/journal.pone.0081251

Shackelton, L. A., Parrish, C. R., & Holmes, E. C. (2006). Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *Journal of Molecular Evolution*, *62*(5), 551–563. https://doi.org/10.1007/s00239-005-0221-1

Shah, A., Parmar, B., Ghodawala, M., & Seth, A. (2018). In silico drug discovery of novel small lead compounds targeting nipah virus attachment glycoprotein. *Journal of Integrated Health Sciences*, *6*(2), 60. https://doi.org/10.4103/jihs.jihs_21_18

Shao, J., Liang, Y., & Ly, H. (2015). Human hemorrhagic fever causing arenaviruses: Molecular mechanisms contributing to virus virulence and disease pathogenesis. *Pathogens*, *4*(2), 283–306. https://doi.org/10.3390/pathogens4020283

Sharpl, P. M., & Li, W. (1987). *potential applications Nucleic Acids Research*. *15*(3), 1281–1295.

Siegel, R. D. (2018). Siegel RD. Classification of Human Viruses. *Principles and Practice of Pediatric Infectious Diseases.*, *January*, 1044-1048e1. doi: 10.1016/B978-0-323-40181-4.00201-2. Epub 2017 Jul 18. PMCID:

PMC7151951.

Simulundu, E., Mweene, A. S., Changula, K., Monze, M., Chizema, E., Mwaba, P., Takada, A., Ippolito, G., Kasolo, F., Zumla, A., & Bates, M. (2016). Lujo viral hemorrhagic fever: considering diagnostic capacity and preparedness in the wake of recent Ebola and Zika virus outbreaks. *Reviews in Medical Virology*, *26*(6), 446–454. https://doi.org/10.1002/rmv.1903

Amineni, U., Pradhan, D., & Marisetty, H. (2010). *In silico identification of common putative drug targets in Leptospira interrogans*. 165–173. https://doi.org/10.1007/s12154-010-0039-1

Damte, D., Suh, J. W., Lee, S. J., Yohannes, S. B., Hossain, M. A., & Park, S. C. (2013). Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of Mycoplasma hyopneumoniae. *Genomics*, *102*(1), 47–56. https://doi.org/10.1016/j.ygeno.2013.04.011

Grobbelaar, A. A., Jardine, J., Burt, F. J., Shepherd, A. J., Shepherd, S. P., Leman, P. A., Kemp, A., Braack, L. E. O., Weyer, J., Paweska, J. T., & Swanepoel, R. (2021). Mammarenaviruses of rodents, South Africa and Zimbabwe. *Emerging Infectious Diseases*, *27*(12), 3092–3102. https://doi.org/10.3201/eid2712.211088

Gurung, A. B., Ali, M. A., Lee, J., Abul Farah, M., & Al-Anazi, K. M. (2020). In silico screening of FDA approved drugs reveals ergotamine and dihydroergotamine as potential coronavirus main protease enzyme inhibitors. *Saudi Journal of Biological Sciences*, *27*(10), 2674–2682. https://doi.org/10.1016/j.sjbs.2020.06.005

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Garcia Diez, F., Harte, N., Kulikova, T., Lin, Q., Lombard, V., … Apweiler, R. (2005). The EMBL nucleotide sequence database. *Nucleic Acids Research*, *33*(DATABASE ISS.), 29–33. https://doi.org/10.1093/nar/gki098

Kumar, C. S., & Kumar, S. (2017). Synonymous codon usage of genes in polymerase complex of Newcastle disease virus. *Journal of Basic Microbiology*, *57*(6), 481–503. https://doi.org/10.1002/jobm.201600740

Kumar, N., Bera, B. C., Greenbaum, B. D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B. N., & Virmani, N. (2016). Revelation of Influencing Factors in Overall Codon Usage Bias of Equine Influenza Viruses. *PloS One*, *11*(4), e0154376. https://doi.org/10.1371/journal.pone.0154376

Roy, A., Mukhopadhyay, S., Sarkar, I., & Sen, A. (2015). Comparative investigation of the various determinants that influence the codon and amino acid usage patterns in the genus Bifidobacterium. *World Journal of Microbiology and Biotechnology*, *31*(6), 959–981. https://doi.org/10.1007/s11274-015-1850-1

Roy, A., Sen, A., Chakrobarty, S., & Sarkar, I. (2018). Comprehensive profiling of functional attributes , virulence potential and evolutionary dynamics in mycobacterial secretomes. *World Journal of Microbiology and Biotechnology*, *34*(1), 1–19. https://doi.org/10.1007/s11274-017-2388-1

Saier, M. H. (2019). *crossm*. *April*, 1–12.

SS, M., JAK, M., M, W., CR, P., D, C., WB, K., C, Z.-T., WI, L., & Daszak P. (2012). Prediction and prevention of the next pandemic zoonosis. *The Lancet*, *380*(December), 1956–1965.

Stumm, G., Russ, A., Nehls, M., & Ag, I. P. (2002). *A Functional Approach to Identify Innovative Drug Targets in the Post-Genome Era*. *2*(4), 263–271.

Trott, O., & Olson, A. (2010). Autodock vina: improving the speed and accuracy of docking. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334.AutoDock

Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Ruppin, E., & Ziv-Ukelson, M. (2011). Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biology*, *12*(11), R110. https://doi.org/10.1186/gb-2011-12-11-r110

Umesh, Kundu, D., Selvaraj, C., Singh, S. K., & Dubey, V. K. (2021). Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target. *Journal of Biomolecular Structure and Dynamics*, *39*(9), 3428–3434. https://doi.org/10.1080/07391102.2020.1763202

Wachsman, M. B., López, E. M. F., Ramirez, J. A., Galagovsky, L. R., & Coto, C. E. (2000). Antiviral effect of brassinosteroids against herpes virus and arenaviruses. *Antiviral Chemistry and Chemotherapy*, *11*(1), 71–77.

https://doi.org/10.1177/095632020001100107

Walker, J. M. (n.d.). *IN Series Editor*.

Wang, M., Zhang, J., Zhou, J. H., Chen, H. T., Ma, L. N., Ding, Y. Z., Liu, W. Q., Gu, Y. X., Zhao, F., & Liu, Y. S. (2011). Analysis of codon usage in type 1 and the new genotypes of duck hepatitis virus. *BioSystems*, *106*(1), 45–50. https://doi.org/10.1016/j.biosystems.2011.06.005

Webb, P. A., Johnson, K. M., Hibbs, J. B., & Kuns, M. L. (1970). Parana, a new tacaribe complex virus from Paraguay. *Archiv Für Die Gesamte Virusforschung*, *32*(4), 379–388. https://doi.org/10.1007/BF01250066

Wilson, M.S. , Metink-Kane, M. M. (2012). 基因的改变NIH Public Access. *Bone*, *23*(1), 1–7. https://doi.org/10.3851/IMP2532.Activity

Winn, W. C., & Murphy, F. A. (1975). Tamiami virus infection in mice and cotton rats. *Bulletin of the World Health Organization*, *52*(4-5 6), 501–506.

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene*, *87*, 23–29. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T39-48T6HPK-5&_user=8077136&_coverDate=03/01/1990&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_acct=C000043575&_version

Wu, G., Culley, D. E., & Zhang, W. (2005). *Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism*. 2175–2187. https://doi.org/10.1099/mic.0.27833-0

Xiang, D., Shen, X., Pu, Z., Irwin, D. M., Liao, M., & Shen, Y. (2018). Convergent evolution of human-isolated H7N9 avian influenza a viruses. *Journal of Infectious Diseases*, *217*(11), 1699–1707. https://doi.org/10.1093/infdis/jiy082

Y Furuta, T komeno, T. N. (2017). Polymerase Activity (%) 100 μ mol / L Favipiravir Favipiravir-RMP Control. *Proc Jpn Acad Ser B Phys Biol Sci.*, *93*(7), 449–463. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5713175/pdf/pjab-93-449

Yamasaki. (2014a). 基因的改变NIH Public Access. *Bone*, *23*(1), 1–7.

https://doi.org/10.1016/j.coviro.2014.02.007.Epidemiology

Yamasaki. (2014b). 基因的改变NIH Public Access. *Bone*, *23*(1), 1–7. https://doi.org/10.1016/j.spen.2012.02.002.Lymphocytic

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yao, H., Chen, M., & Tang, Z. (2019). Analysis of Synonymous Codon Usage Bias in Flaviviridae Virus. *BioMed Research International*, *2019*, 1–12. https://doi.org/10.1155/2019/5857285

Yi, S., Li, Y., & Wang, W. (2018). Selection shapes the patterns of codon usage in three closely related species of genus Misgurnus. *Genomics*, *110*(2), 134–142. https://doi.org/10.1016/j.ygeno.2017.09.004

Zapata, J. C., & Salvato, M. S. (2013). Arenavirus variations due to host-specific adaptation. *Viruses*, *5*(1), 241–278. https://doi.org/10.3390/v5010241

Zeitlin, L., Geisbert, J. B., Deer, D. J., Fenton, K. A., Bohorov, O., Bohorova, N., Goodman, C., Kim, D., Hiatt, A., Pauly, M. H., Velasco, J., Whaley, K. J., Altmann, F., Gruber, C., Steinkellner, H., Honko, A. N., Kuehne, A. I., Aman, M. J., Sahandi, S., … Geisbert, T. W. (2016). Monoclonal antibody therapy for Junin virus infection. *Proceedings of the National Academy of Sciences*, *113*(16), 4458–4463. https://doi.org/10.1073/pnas.1600996113

Zhang, Y., Li, L., Liu, X., Dong, S., Wang, W., Huo, T., Guo, Y., & Rao, Z. (2019). *Crystal structure of Junin virus NP*. *2013*, 2175–2183. https://doi.org/10.1099/vir.0.055053-0

Zhang, Y. M., Shao, Z. Q., Yang, L. T., Sun, X. Q., Mao, Y. F., Chen, J. Q., & Wang, B. (2013). Non-random arrangement of synonymous codons in archaea coding sequences. *Genomics*, *101*(6),362–367. https://doi.org/10.1016/j.ygeno.2013.04.008

Zhou, M., & Li, Æ. X. (2009). *Analysis of synonymous codon usage patterns in different plant mitochondrial genomes*. 2039–2046. https://doi.org/10.1007/s11033-008-9414-1.

Appendix-A
Software used in the present study

Software	Execution	Description
CodonW (v1.4.4)	Windows	Program for codon and amino acid usage
AutoDock Vina	Windows/Linux	Suite of automated docking tools
MEGA	Windows/Linux	Tool for sequence alignment and phylogeny
PAML (v4.9)	Linux	Phylogenetic analysis using maximum likelihood
Pymol (v1.8.2.3)	Windows	Molecular visualization software written in python
SPSS	Windows	Software package used for statistical analysis
PRODRG server	Windows	Server for energy minimization of predicted structure
R-Package (v3.3.1)	Windows/Linux	Functional language for statistical analysis
CASTp	Windows	Tool for binding site prediction
SwissADME	Windows	Server for physiochemical properties prediction
ProTox-II	Windows	Server to predict toxicity parameter
GROMACS	Linux/Windows	Software for dynamics analysis
OpenBabel	Windows/Linux	File format converter
PLIP	Windows	To study protein ligand interactions

Appendix-B
Databases and web servers used in the present study

Software	Execution	Description
CAI Calculator2	www.userpages.umbc.edu/~wug1/codon/cai/cais.php	Calculation of codon adaptation Index
Clustal Omega	www.ebi.ac.uk/Tools/msa/clustalo	Sequence alignment
Codon Usage Database	www.kazusa.or.jp/codon	Database containing information regarding codon usage of several organisms
DrugBank	www.drugbank.ca	Database of drug information
NCBI	www.ncbi.nlm.nih.gov	National Center for Biotechnology Information
RCSB PDB	www.rcsb.org/pdb	RCSB Protein Data Bank
ZINC15 database	https://zinc15.docking.org/	Databse of commercially available compounds
PubChem database	www.pubchem.ncbi.nlm.nih.gov	Collection of chemicals