

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

Project Report of 2022: SVP-2247

**“Comparative Genomics and Proteomics of Dengue virus
and Zika virus”**



IQAC

Sri Venkateswara College

University of Delhi

Dhaura Kuan




New Delhi -110021

SRIVIPRA PROJECT 2022

Title: Comparative Genomics and Proteomics of Dengue virus and Zika virus

<p>Name of Mentor: Dr. Mansi Verma Name of Department: Zoology Designation: Assistant Professor</p>	<p>Photo</p>  A portrait of Dr. Mansi Verma, a woman with long dark hair, wearing a blue and gold patterned sari, smiling. The background shows a tree trunk.
<p>Name of Mentor: Dr Rajendra Phartyal Name of Department: Zoology Designation: Assistant Professor</p>	<p>Photo</p>  A portrait of Dr. Rajendra Phartyal, a man with short dark hair, wearing a light blue shirt, smiling. The background is a plain wall.

List of students under the SRIVIPRA Project

S.No	Name of the student	Course	Photo
1	Madhav Gupta	B.Sc.(H) Zoology	
2	Srianshu Kumar Panda	B.Sc.(H) Zoology	
3	Deepanshu	B.Sc.(H) Zoology	
4	Vidhi Wadhawan	B.Sc.(H) Biological Sciences	
5	Jeevan	B.Sc. Life Science	



(Dr. Mansi Verma)
Department of Zoology



(Dr. Rajendra Phartyal)
Department of Zoology

Signature of Mentor

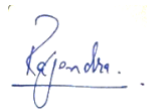
Certificate

This is to certify that the aforementioned students from Sri Venkateswara College have participated in the summer project SVP-2247 titled “**Comparative Genomics and Proteomics of Dengue virus and Zika virus**”. The participants have carried out the research project work under my guidance and supervision from 21st June 2022 to 25th September 2022. The work carried out is original and carried out in an online mode.



(Dr. Mansi Verma)

Department of Zoology



(Dr. Rajendra Phartyal)

Department of Zoology

Signature of Mentor

Acknowledgements

We would like to express our sincere gratitude to Dr. Mansi Verma of the Department of Zoology, Sri Venkateswara College, who provided guidance, in-depth insights and expertise during this internship period which helped us to learn a lot many things.

We would also like to acknowledge the support of our IQAC Coordinator, Prof. Swarn Singh, for his constant encouragement and support.

We are thankful to Dr. Sharda Pasricha and Dr. S Krishnakumar, the Convenors of Sri Vipra- 2022 summer internship program for successfully organizing this activity and providing support throughout.

This Summer Internship was supported by our respected Principal, Professor. C. Sheela Reddy. We thank her for giving us this opportunity.

We would also like to extend our gratitude to Miss Pranjal Vats for her immense support and guidance throughout the project.

CONTENTS

S.No	Topic	Page No.
1	Abstract	7
2	Introduction	8
	2.1 Flaviviridae	8
	2.2 Flaviviral Genome Organisation	9
	2.3 DENV: General account	9
	2.4 ZIKV: General account	10
	2.5 DENV and ZIKV: Comparative account	11
	2.6 mRNA	11
	2.7 Differentially Expressed mRNA	12
3	Objective	12
4	Methodology	12
	4.1 Searching a Microarray Dataset	12
	4.2 Data analysis in R	12
	4.3 Selection of genes on the basis of significance level	12
	4.4 Finding out intersecting subset of DENV and ZIKV	13
	4.5 Generation of Genetic interaction network	13
	4.6 Identification of hub genes and pathway analysis	13
	4.7 Gene Ontology of hub genes	14
5	Results	14
	5.1. Screening for Differentially Expressed Genes (DEGs)	14
	5.2. Identification of Hub genes from PPI network	16
	5.3. Enrichment and Pathway Analysis	17
6	Discussion	18
7	Conclusion	21
8	References	22

1. Abstract

The present study is an in-silico approach for the comparative genomics and translational dynamics of dengue virus (DENV) and Zika virus (ZIKV). Till now comparative genomics of foot-and-mouth disease virus, Ebola virus, Epstein-Barr virus, and many other viruses have been conducted to study the mechanism of action with respect to humans. Comparative studies of different serotypes of dengue- DENV 1, DENV 2, and DENV 3 have also been performed individually. DENV and ZIKV belong to the same sister taxa. Despite the differences in clinical features of dengue and Zika, they both share about 90% of their genome sequence. For the purpose of the study, GEO Data is retrieved from the NCBI Gene Expression Omnibus Database. By combining the results of functional analysis, MCODE, CytoHubba, and ShinyGO – We have found a few important genes that show most of the significant pathways for the virulence of these viruses. Using Cytoscape analysis, DENV and ZIKV were described to actively initiate autophagy in several cellular models, possibly through the expression of NS4A, NS4B, and/or NS1. Another noteworthy aspect of the research was the identification of similar signalling pathways, which are an essential step in their invasion. The application of this research would be to understand the mechanism of infection, gene ontology, and phylogeny, and to determine common target genes for the treatment of diseases caused by these flaviviruses.

Keywords: Dengue, Zika, Flavivirus, *in silico*, differentially expressed genes, Protein-protein interaction, bioinformatics.

2. Introduction

A virus conveyed by mosquitoes called dengue has recently expanded quickly in all WHO areas. Dengue virus is an arbovirus (arthropod-borne virus), which are responsible for the escalating hazards to public health. These unusual viruses cause considerable harm to civilization and become a more severe problem as there is no specific treatment to stop them. Therefore, the creation of a potent remedy for infections caused by them is urgently needed.

As the name suggests these viruses' spread infection by arthropods as their vector. Arthropods that transmit these viruses include bugs such as mosquitoes, ticks, fleas, and gnats but for our study we are more interested in mosquitoes. These viruses share some similarities, but there are also many differences. Arboviruses are classified based on these differences. The three main classifications of arboviruses are flavivirus, alphavirus, and bunyavirus. In this report we focus majorly on Dengue and Zika viruses which are examples of mosquito-borne flaviviruses, belonging to the family Flaviviridae.

2.1 Flaviviridae

Three antigenically diverse genera—Flavivirus, Pestivirus, and Hepacivirus—make up the family Flaviviridae. The viruses that cause Dengue haemorrhagic fever, Zika fever, yellow fever, West Nile fever/encephalitis, St. Louis encephalitis (SLE), and Japanese encephalitis, are human diseases of the genus Flavivirus (DHF). The genus Pestivirus has no relevance to humans, but the genus Hepacivirus contains the hepatitis C-causing agent. Many years ago, focus was paid to the creation of vaccines to manage diseases due to the public health burden of various flaviviral infections. They have significantly reduced the prevalence of that disease when they are available. The dengue virus has four different serotypes, which makes the creation of a vaccine challenging.

2.2 Flaviviral Genome Organisation

RNA viruses are known for their small-sized genomes that sustain high mutation rates and lack proofreading. Therefore, the occupancy of a large genome is not feasible. Another feature common to riboviruses is the presence of RNA-dependent-RNA polymerase (RdRp), required for the replication of (+) single-strand RNA to make (-) single-strand RNA. Flaviviruses are lipid-enveloped, icosahedral RNA viruses that have a single-stranded (+) RNA genome having a size of 11 kb approximately. Their GC content ranges from 43% (NKV flaviviruses) to 54% (TBFV). The

unsegmented RNA genome is capped but bears no polyA tail and comprises a single Open Reading Frame (ORF) flanked by 5' and 3' untranslated regions (UTRs). The ORF of around 10 kb is translated into a polyprotein of 3400 amino acids. The polyprotein is further processed by viral and host proteases to yield 10 viral proteins, comprising of three structural proteins viz. capsid (C), membrane (M), and envelope (E); and seven non-structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5).

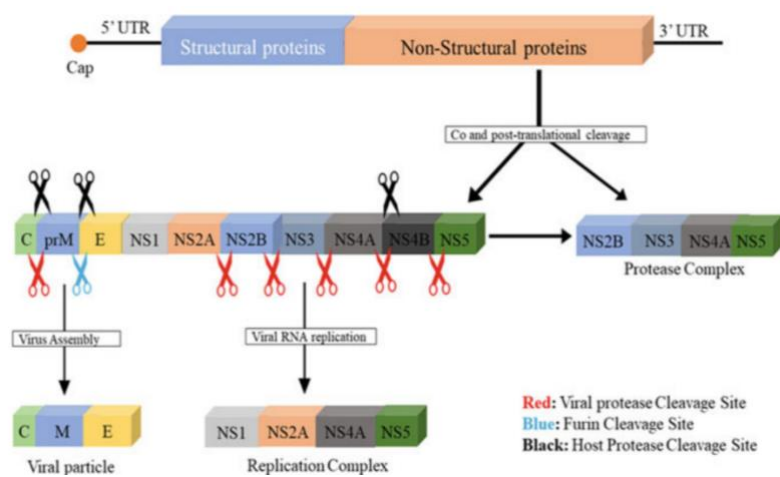


Fig. 1 Flaviviral genome is organised as 5'CAP(I)-5'UTR-C-prm-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'UTR. All coding region is translated as a polyprotein which is then acted upon by viral and host proteases to yield ten proteins. Structural proteins (C-M-E) are involved in viral particle formation whereas non-structural proteins form replication and protease complexes.

2.3 DENV: General account

The RNA virus of the Flaviviridae family, which causes the acute viral sickness known as dengue, is transmitted by *A. aegypti* and *A. albopictus*. Asymptomatic fever to severe consequences including hemorrhagic fever and shock are all possible presentation aspects. The most typical symptoms are a high temperature that comes on suddenly, myalgia, cutaneous rash, hemorrhagic bursts, and vascular shock. Oral signs of dengue infection are uncommon, although in some cases they may be the sole symptom that presents. For mortality to be reduced, early and precise diagnosis is essential.

Dengue is regarded by the Worldwide Health Organization (WHO) as a serious worldwide public health threat in tropics and subtropics countries. From 1960 and 2010, dengue cases surged by 30 times globally as a result of rapid urbanisation development, climate change, urban migration, ineffective mosquito prevention, frequent air transportation, and a dearth of

medical services. [3,4,5] There are 2.5 billion people who live in dengue-endemic areas [5], and 400 million infections happen year with a fatality rate that exceeds 5-20% in some places.

There are four different DENV serotypes: DENV-1, DENV-2, DENV-3, and DENV-4. However, there is a 60–80% commonality across the various serotypes. The adhesins of the serotypes show diversity in people. One of the reasons there is no effective treatment available is because once a serotype has recovered from infection, it guarantees lifetime immunity for that specific serotype but not for any other serotypes. There are two forms of dengue infection: mild dengue fever and severe hemorrhagic fever. Compared to other serotypes, secondary DENV-2 infection is more likely to cause severe illness. Since the first encounter of virus in Hawaii in 1944, there has been an ongoing effort to produce a dengue vaccine due to the disease's high danger around the world. Although Sanofi Pasteur's Dengvaxia, the first live attenuated tetravalent vaccination ever licenced, has been available since 2015, its effectiveness for all serotypes remains a concern.

2.4 ZIKV: General account

Zika is the next terrifying infection after dengue. The recent widespread outbreak of the Zika virus (ZIKV) has compelled health researchers to express extraordinary alarm about it. This virus, which has a single stranded positive sense RNA genome, is likewise a member of the Flaviviridae family and the genus Flavivirus. The Zika virus was isolated from a monkey (Rhesus) in the jungle of Zika, Uganda, in April 1947, giving rise to its name. The first large epidemic of Zika happened on Micronesia's Yap Island in 2007. Later outbreaks occurred in French Polynesia and the Pacific islands in 2013 and 2014, respectively. Similarly, Central and South America were affected by the Zika virus in 2015 when cases were identified in Brazil, while the United States saw its first epidemic in the state of Florida in 2016. [10]

2.5 DENV and ZIKV: Comparative account

In terms of how their genomes are organised and their vectors, DENV and Zika are too similar to one another. The open reading frame (ORF) of Zika, which is 10,794 kb in length, encodes 10 proteins that are identical to those found in the dengue virus genome. Their vectors for transmission, *Aedes aegypti* and *Aedes albopictus*, are also the same. Non-human primates and insects both maintain the ZIKAV endemicity lifecycle. However, there have

been a few isolated instances when certain cow, buffalo, camel, sheep, and bat species have demonstrated the existence of antibodies.

According to some studies, the virus can also spread between people by blood transfusion, breastfeeding, or maternofetal contact. About 18% of viral infections are symptomatic, with the rest showing only moderate symptoms as headache, retro-orbital discomfort, Edema, vomiting, rashes, conjunctivitis, high body temperature, and arthralgia.

The majority of the symptoms are similar to dengue, but major complications from ZIKAV infections include Guillain-Barre syndrome (GBS), an autoimmune disease that causes flaccid paralysis, and microcephaly in newborns. The former is a neurodevelopmental disorder, and the latter is a neuroimmune disease. ZIKAV, like all other flaviviruses, is treated symptomatically and lacks a particular vaccine. However, Inovio Pharmaceuticals and GeneOne Life Sciences have created a DNA vaccine, GLS-5700, by utilising various strains of infected virus. The trials are anticipated to be completed by 2018.

2.6 mRNA

One of the several forms of RNA present in cells is messenger RNA (mRNA), which is a single-stranded RNA molecule corresponding to one of a gene's DNA strands. This particular RNA is produced in the nucleus, like the majority of RNAs, and is then transported to the cytosol where the translational machinery (the equipment that actually produces proteins) attaches to the mRNA molecules and extracts the information on the mRNA to produce a particular protein. Every process requires regulatory points, from messenger RNA (mRNA) translation as well as mRNA and polypeptide catabolism in the cytoplasm to pre-mRNA biosynthesis in the nucleus.

2.7 Differentially Expressed mRNA

It is common practise to examine gene expression differences within cells or tissues of various types and under various circumstances by analysing messenger RNA and proteins. A gene is said to be differentially expressed if there is a statistically significant difference or change in read numbers or expression index over two experimental conditions. If any of the five time periods significantly differed from control (at FDR 0.05, termed to as a DEGs mRNA profile), then the mRNA is said to be differentially expressed within the condition. Differential gene expression is thought to be essential for the planned formation of new cells and tissues due to its exact start and stop periods.

3. Objectives of the study

The current research would be useful:

- To understand the mechanism of infection of DENV and ZIKV & their gene ontology.
- To identify common target genes involved in infection caused by DENV and ZIKV.
- To the finding of significant biomarkers and therapeutic targets for the assessment and treatment of Dengue and Zika Viruses.

4. Methodology

4.1. Searching a Microarray Dataset

A microarray is a laboratory tool used to detect the expression of thousands of genes at the same time. Microarray datasets are made up of complex and high-dimensional samples and genes, and the number of samples is often significantly lower than the number of genes. Gene selection is a difficult process for microarray expression data analysis because of the data imbalance. Finding a decent microarray dataset is thus critical for obtaining accurate study findings.

Gene Expression Omnibus (GEO) is a database supported by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) that accepts raw and processed data with written descriptions of experimental design, sample attributes, and methodology for studies of high-throughput gene expression and genomics. In this study, we identified mRNA profile of both dengue virus (DENV) and zika virus (ZIKV) for the differentially expressed genes (DEGs). The GEO Dataset GSE207347 was used for microarray analysis for the given viruses. This Dataset contains mRNA sequencing of the human neuronal progenitor cells (hNPCs) cells either uninfected or infected with ZIKV or DENV. Further for the retrieval of data from this dataset, R programming language was used.

4.2. Data analysis in R

The R version 4.2.1 and R studio 2022.07.1+554 was used for this purpose. Metadata was retrieved using GEOquery package and the sample information of 12 samples was printed in the tabular format. 13150 genes for DENV and 13009 genes for ZIKV were observed. The data was then normalised and the summary of expression values were obtained.

4.3. Selection of genes on the basis of significance level

Further the data was exported and analysed in Microsoft excel. For the tabulated data obtained from R, significant genes were selected (Significance level $P_{adj} < 0.05$) and for further analysis. 193 significant genes for DENV and 780 significant genes for ZIKV were identified after removing duplicated values. Venn of the data was then created for filtering out common genes.

4.4. Finding out intersecting subset of DENV and ZIKV

Venn diagram is a bioinformatics tool which can be used to calculate the intersections of a list of elements. By using this tool, we are able to generate a textual output indicating which elements are in each intersection or are unique to a certain list. In our study, this allows us to compare and contrast the differentially expressed genes (DEGs) in the two sequences of the viruses and thus, to filter out the common genes among them. We have been able to identify 138 common genes out of the selected significant genes of the two viral sequences being analysed. These common genes were then exported to STRING to identify the various protein-protein interactions and create a network between them.

4.5. Generation of Genetic interaction network

Genetic interaction networks are valuable for understanding the link between genotype and phenotype because they show the functional interactions between pairs of genes in an organism. The vast majority of genes do not encode for specific traits. Instead, phenotypes are frequently the product of the interplay of multiple genes. Interactions between genetic variations, as well as environmental factors, are anticipated to have a significant influence in defining the phenotype resulting from a particular genotype. By discovering such connections between pairs of genes, genetic interaction networks aid in the understanding of genetic interactions.

STRING is a biological database and online resource which shows protein-protein interactions that have been observed and predicted in molecular biology. Information from various sources, including as experimental data, computer prediction techniques, and open text collections, is present in the STRING database. For comparison, the most recent string database version (11.5) is used. A detailed insight of gene interactions between the sorted or common genes identified from Venn was observed in STRING. Data of medium confidence (0.400) is taken in tab-separated values file, which was further analysed in cytoscape.

4.6. Identification of hub genes and pathway analysis

Cytoscape is an open-source bioinformatics software used for visualizing complex networks and integrating these with any type of attribute data. Various extensions like cytoHubba, MCODE, BiNGO, including 243 others are further available for more convenient analysis in network biology. We can confine our study by restricting to hub genes which are the most functional or interacting genes in complete network. From 138, top ten hub genes are sorted in cytoHubba on the basis of degree.

4.7. Gene Ontology of hub genes

Further analyses include gene ontology for hub genes, which allows us to describe a gene/gene product in detail, considering three main aspects: its molecular function, the biological process in which it participates, and its cellular location. It can be done either by inbuilt app in cytoscape like BiNGO or other online tool including KEGG (Kyoto Encyclopaedia of Genes and Genomes) and ShinyGO.

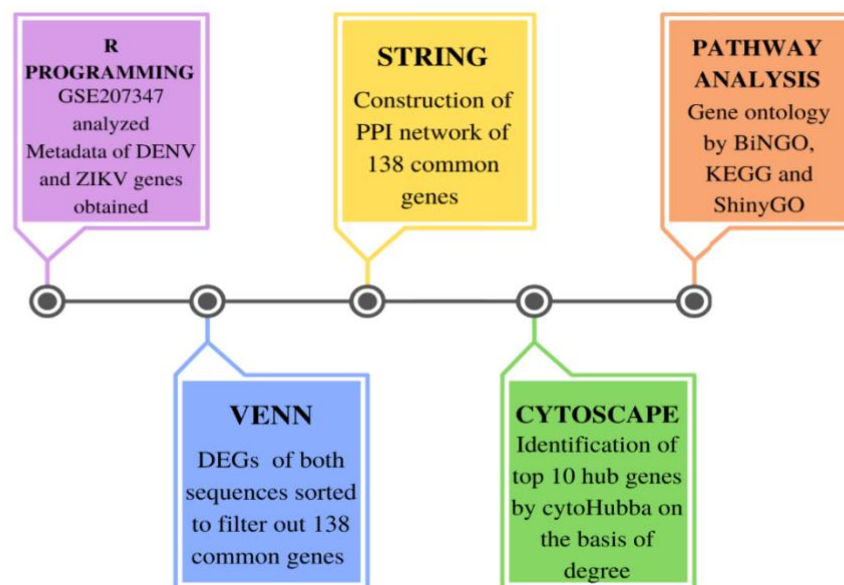


Fig. 2 Work flow

5. Results

5.1. Screening for Differentially Expressed Genes (DEGs)

Our survey of the GEO database led us to the dataset GSE207347 containing mRNA sequencing of both Dengue and Zika viruses. This recently published data involved research

on human neuronal progenitor cells, either infected or non-infected by DENV or ZIKV. This data was used to study cellular stress responses and infection pathways of the two viruses.

NCBI > GEO > **Accession Display** [?](#) Not logged in | [Login](#) [?](#)

GEO help: Mouse over screen elements for information.

Scope: Format: Amount: GEO accession:

Series GSE207347 [Query DataSets for GSE207347](#)

Status	Public on Jul 06, 2022
Title	Transcriptional and Translational Dynamics of Zika and Dengue Virus Infection
Organism	Homo sapiens
Experiment type	Expression profiling by high throughput sequencing
Summary	Zika virus (ZIKV) and dengue virus (DENV) are members of the Flaviviridae family of RNA viruses and cause severe disease in humans. ZIKV and DENV share over 90% of their genome sequences, however the clinical features of Zika and dengue infections are very different reflecting tropism and cellular effects. Here, we used simultaneous RNA sequencing and ribosome footprinting to define the transcriptional and translational dynamics of ZIKV and DENV infection in human neuronal progenitor cells (hNPCs). The gene expression data showed induction of aminoacyl tRNA synthetases (ARS) and the translation-activating PIM1 kinase indicating an increase in RNA translation capacity. The data also reveal activation of different cell stress responses, with ZIKV triggering a BACH1/2 redox program, and DENV activating the ATF/CHOP endoplasmic reticulum (ER) stress program. The RNA translation data highlight activation of polyamine metabolism through changes in key enzymes and their regulators. This pathway is needed for eIF5A hypusination and has been implicated viral translation and replication. Concerning the viral RNA genomes, ribosome occupancy readily identifies highly translated open reading frames and a novel upstream ORF (uORF) in the DENV genome. Together, our data highlight both the cellular stress response and also the activation of RNA translation and polyamine metabolism during DENV and ZIKV infection.
Overall design	We performed ribosome footprinting and RNA sequencing on human neuronal progenitor cells (hNPCs) cells either uninfected or infected with ZIKV or DENV in two biological replicates that were collected 72 hours post infection. Uninfected samples are indicated as A1 and B1, ZIKV infected samples are indicated as A2, B2 and DENV infected samples are indicated as A3 and B3.
Contributor(s)	Singh K , Martinez G , Lin J , Gregory J , Abdelaal R , Kang K , Brennand K , Grunweller A , Ouyang Z , Phatnani H , Kielian M , Wendel H
Citation(s)	Singh K, Martinez MG, Lin J, Gregory J et al. Transcriptional and Translational Dynamics of Zika and Dengue Virus Infection. <i>Viruses</i> 2022 Jun 28;14(7). PMID: 35891396

Fig. 3 GEO profile Interface [GSE207347]

Successful analysis of the dataset samples was done by using R programming. The top significant genes for both viruses were identified by considering the P-adj.<0.05 value. The differentially expressed genes were sorted by exporting the gene data from R to Excel, for further analysis by other bioinformatics tools.

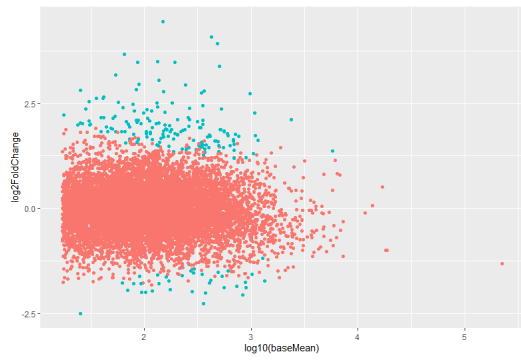


Fig. 4

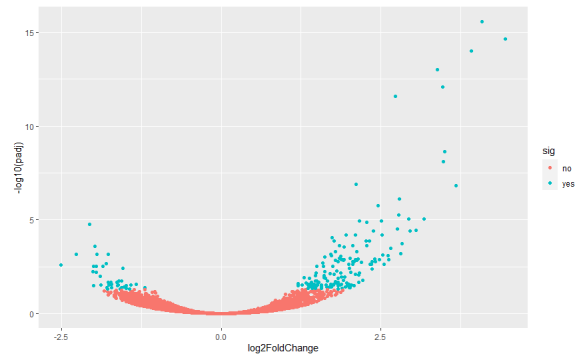


Fig. 5

Fig. 4 & Fig. 5 showing MA Plot & Volcano Plot respectively for DENV DESeq2 Data. Here, blue dot represents Differentially expressed genes (Sig. Level Padj < 0.05) while red dots represent insignificant genes (Padj > 0.05).

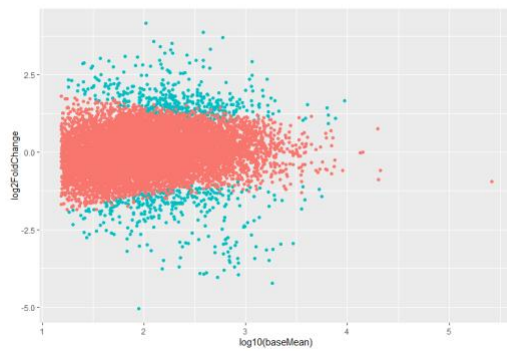


Fig. 6

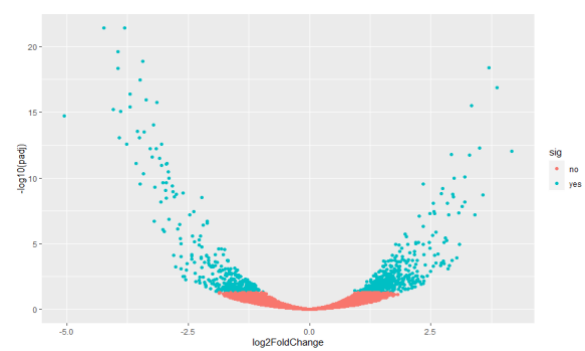


Fig. 7

Fig. 6 & Fig. 7 showing MA Plot & Volcano Plot respectively for ZIKV DESeq2 Data. Here, blue dot represents Differentially expressed genes (Sig. Level Padj < 0.05) while red dots represent insignificant genes (Padj > 0.05).

5.2. Identification of Hub genes from PPI network

Further, by considering significant level Padj. < 0.05, the DEGs for DENV and ZIKV were selected from the tabulated form of gene data obtained. Through a Venn analysis, we were able to screen out 138 common DEGs in our sample of 193 significant genes for DENV and 780 significant genes for ZIKV.

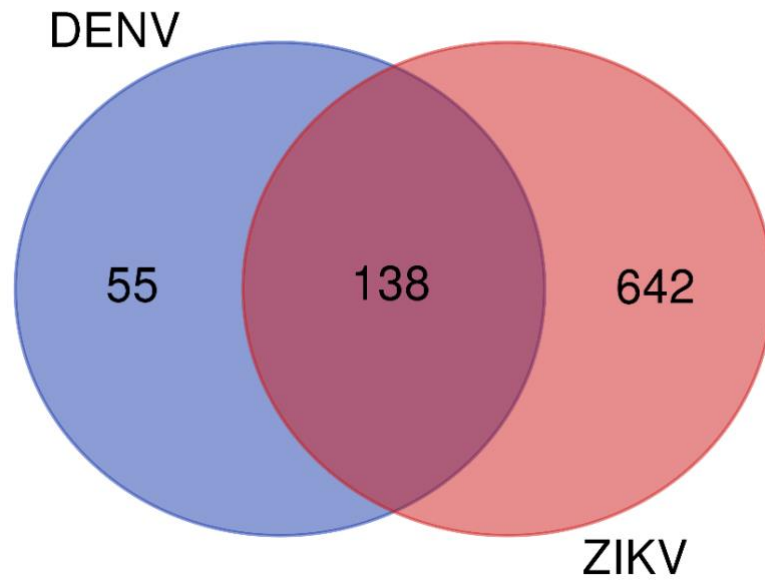


Fig. 8 Venn diagram showing overlapping genes of DENV & ZIKV.

A protein-protein interaction network was created between the 138 common DEGs by using the STRING database. This database allowed us to obtain a network of 138 nodes or genes and 326 edges at medium confidence (0.400). Further, this network was exported to Cytoscape software, which was then used to identify the top 10 hub genes according to degree by cytoHubba application. The identified hub genes were JUN, ASNS, DDIT3, ATF3, EIF2AK3, TRIB3, CEBPB, VEGFA, PPP1R15A, and H2AFZ.

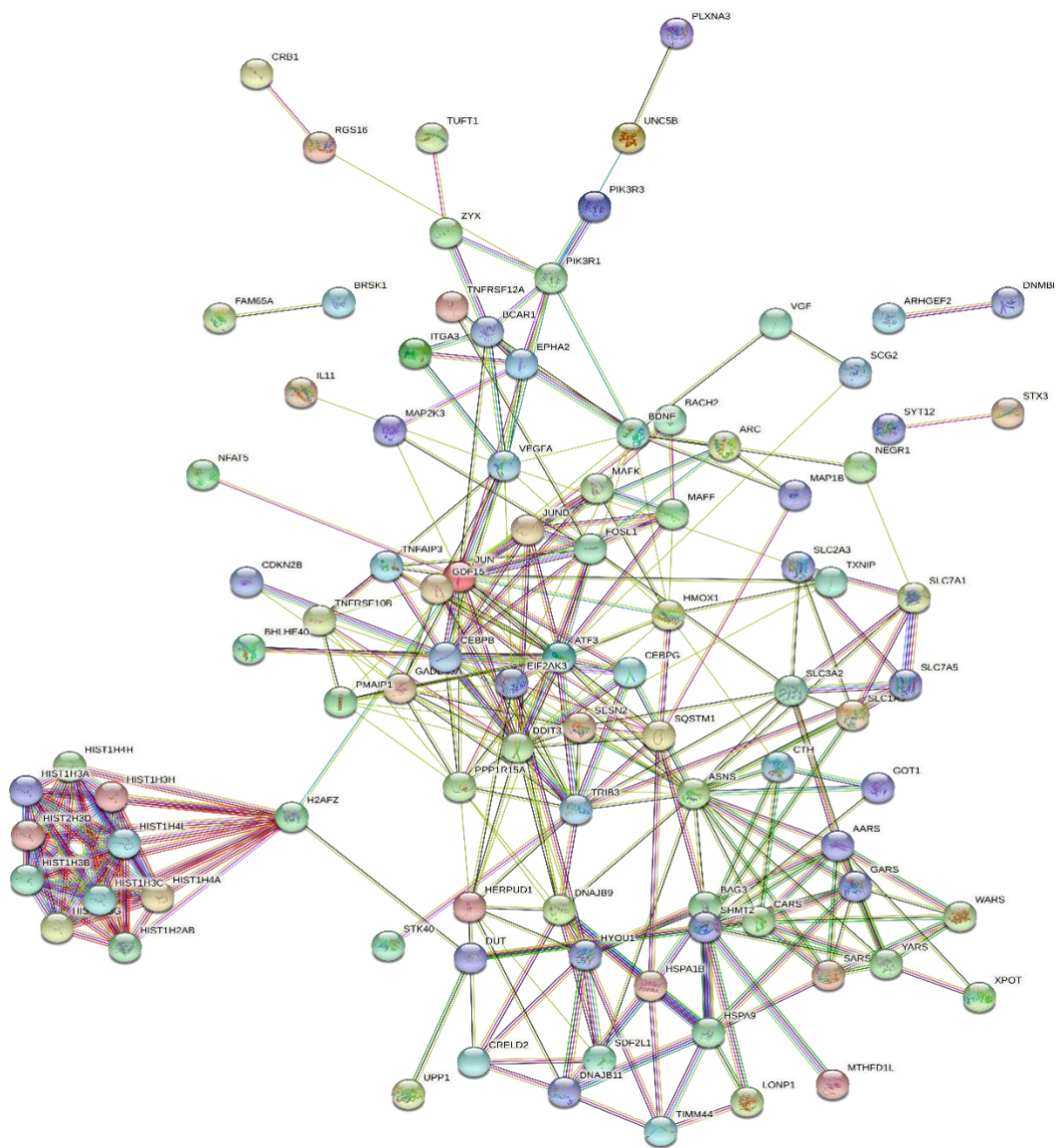


Fig. 9 STRING Network of 138 common genes showing interactions.

5.3. Enrichment and Pathway Analysis

Enrichment analysis and pathway analysis from ShinyGO, KEGG and BiNGO in Cytoscape was conducted based on biological processes associated with the hub genes. According to our results from pathway analysis, the genes JUN, VEGFA, CEBPB and EIF2AK3 appear to be the most significant. These are involved in processes such as mitophagy, receptor development and differentiation, and apoptosis. These genes may also play role in certain signalling pathways including IL-17, tyrosine kinase, MAPK, Toll-like receptor, NOD-like receptor, Wnt, Ras, CAMP and TNF signalling.

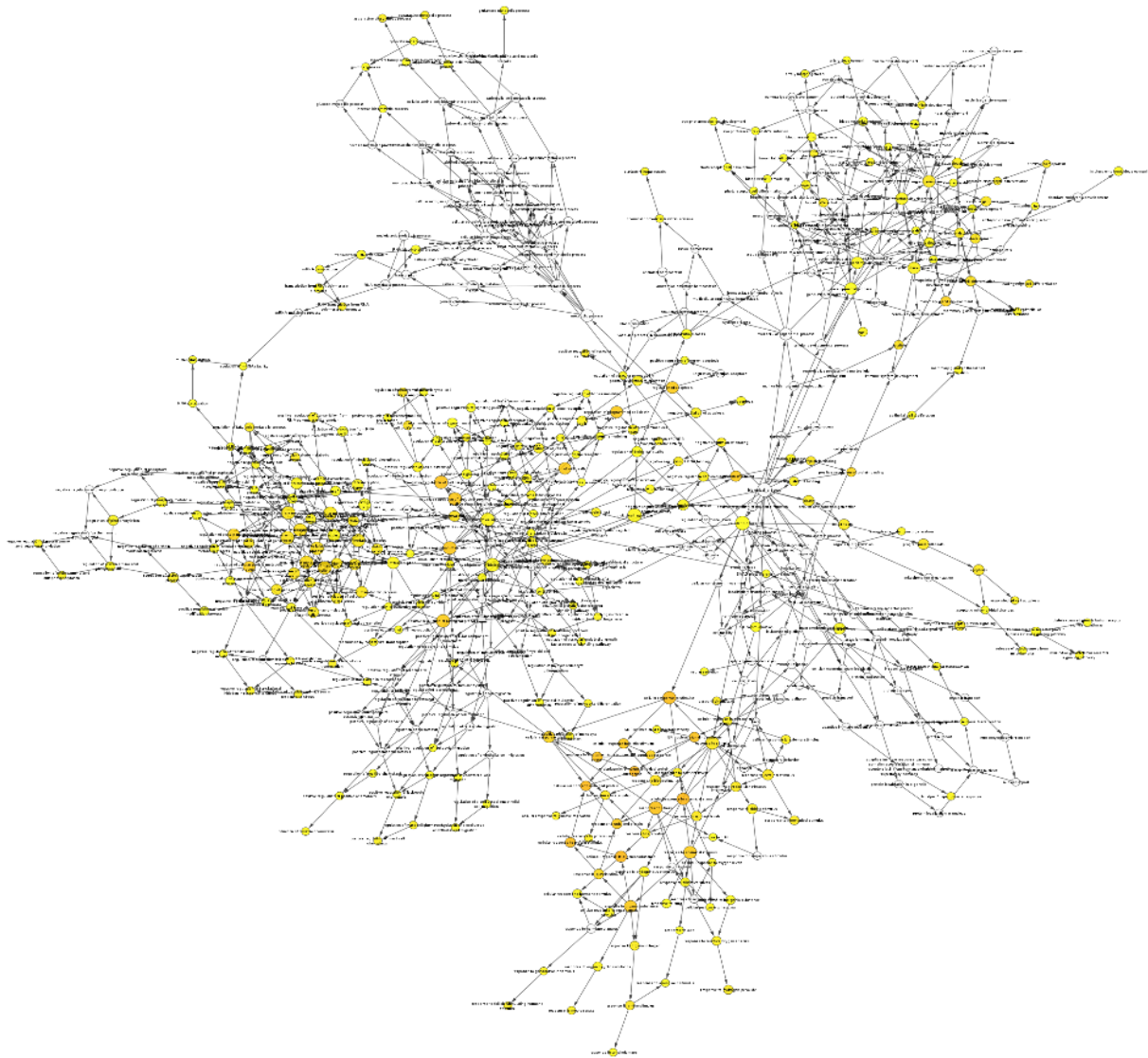


Fig. 10 BiNGO pathway enrichment map for top 10 Hub genes namely- JUN, ASNS, DDIT3, ATF3, EIF2AK3, TRIB3, CEBPB, VEGFA, PPP1R15A, and H2AFZ.

6. Discussion

Insect-borne viruses have been the cause of epidemics in various parts of the world due to recent population expansion and globalisation. With absence of an effective vaccine or treatment, recent research on the family Flaviviridae has been focussed on dengue and zika virus. Despite the similar genome organisation, the different manifestation of these viruses in humans and other primates and non-primates has been a particular matter of interest. The application of various bioinformatics tools and comparative genomics techniques to identify biomarkers can be further studied to develop effective treatments for the two diseases. [11]

Comparative genomics is simply the comparison of biological data acquired from whole-genome sequencing. Thus, comparative genomics began in 1995, with the publication of the first two entire organism genomes (for the bacteria *Haemophilus influenzae* RD and *Mycoplasma genitalium* G37). Soon after, bioinformatics tools were developed to compare genome sequences and the RNAs, proteins, and gene annotations that may be extracted from them. These tools are continually changing in order to keep up with the exponential expansion of sequenced genomes caused by breakthroughs in sequencing technology, as well as to become more thorough and user-friendly. With approximately 80,000 species having sequenced genomes, nearly 10,000 of which are complete, the application of comparative genomic techniques is growing.

Over a third of all genes in sequenced genomes have unknown functions, and over 2000 recognised enzymes do not have a corresponding gene. Correctly connecting gene and function is a critical problem for postgenomic biology, especially for the hundreds of genes that encode protein families that are extensively conserved. Comparative genomics predicts function by employing the 'guilt by association' concept, which involves discovering correlations between unknown and known genes and inferring functions for the unknowns. The most generally helpful indication to function is gene clustering—the closeness of genes in the genome. Functionally related genes in prokaryotes are frequently grouped in operons or divergently transcribed from the same promoter region, or they may simply be neighbours or near-neighbours. Plants can also have genes that encode biosynthetic pathways for specific compounds like alkaloids and terpenes. Gene fusions, in which independent parent gene products are fused together, also strongly suggest functional relationship, such as enzymes for related processes. A common protein that recognises a certain DNA motif or a common riboswitch frequently regulates genes from the same pathway or activity. Thus, shared regulatory sites can be used to identify groups of functionally related individual genes. Finally, there is phylogenetic co-occurrence, which states that genes that collaborate in a process will be either all present or all absent, or none at all in a particular organism. Integrating these many lines of comparative genomic evidence allows for substantial functional conclusions.

Although the number of potential biomarkers or genes decreases as we progress from variation identification to variant validation, sample size gets bigger. We need to investigate a finding on a larger sample set to confirm it. Larger sample sets demanded the use of multi-omics approaches for data analysis.

MICROARRAY Analysis can be carried out to aid in the knowledge of gene expression. The construction of a protein-protein network comprising differentially expressed genes, pathway enrichment mapping, and GO ontology are the next steps. Public libraries (GEO2R-avail the AFFYMETRIC microarray data), Cytoscape, and STRING are among of the tools used. The fact that the results are limited to the probes we have on the chip is a possible challenge for this technique. Furthermore, the data obtained has a low resolution.

The potential biomarker is suggested by integrating OMICS with Gene Ontology, pathway enrichment, and network analysis. Antibodies against the targeted protein can be used in further testing, such as Immunohistochemistry, to investigate tissue-specific protein expression.

This project aimed to identify the common infection pathways between the two viruses and potential biomarkers of these diseases. The GO and pathway analysis conducted on hub genes with respect to the related biological processes indicate their roles in response to cellular stress, unfolded proteins, chemical stimulus and ER signalling. These also show involvement in regulation of apoptosis, biosynthetic processes, cell cycle, gene expression, and in differentiation and development of cells, tissues and organs. Autophagy refers to the breakdown of cellular components for maintenance of homeostasis. DENV and ZIKV infect cells by causing autophagy to allow replication and organization of viral genome in host. Reticulophagy or ER degradation has been reported to be inhibited by NS2B3 proteases coded by the two viruses. NS3 viral proteins can also induce ER expansion to promote viral replication [12]. A few of the hub genes studied here have shown involvement in autophagy and endoplasmic reticulum stress response. Furthermore, immune responses such as Wnt, Ras, Rap1, MAPK and Type-1 interferon signalling pathways have been associated with DENV infection, and are shown to be involving the hub genes identified in our study. [13-15]

7. Conclusion

In this study, a number of differentially expressed genes have been identified to be associated with infection mechanisms of dengue and zika viruses, by utilising various bioinformatics tools. The interactions of these host genes with structural proteins of viral genome, and their ability to suppress immune responses serves to be a point of interest for the study of flaviviruses. Further research on these genes and infection pathways can be fruitful in development of treatment options based on the common target genes.

8. References

- [1] de Crécy-Lagard V, El Yacoubi B, de la Garza RD, Noiriél A, Hanson AD. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC genomics*. 2007 Dec;8(1):1-5.
- [2] Carrillo C, Tulman ER, Delhon G, Lu Z, Carreno A, Vagnozzi A, Kutish GF, Rock DL (2005) Comparative genomics of foot-and-mouth disease virus. *J Virol* 79:6487–6504
- [3] Jun S-R, Leuze MR, Nookaew I, et al (2015) Ebolavirus comparative genomics. Greber U, ed., *FEMS Microbiol Rev*, 39,764–778.
- [4] ChiaraM, ManzariC, Lionetti Cetal(2016) Geographic population structure in Epstein-Barr virus revealed by comparative genomics. *GenomeBiol.Evol*8:3284–3291
- [5] Tang Y, Rodpradit P, Chinnawirotpisan P et al (2010) Comparative analysis of full-length genomic sequences of 10 dengue serotype 1 viruses associated with different genotypes, epidemics, and disease severity isolated in Thailand over 22 years. *Am J Trop Med Hyg* 83: 1156–1165
- [6] Osman O, Fong MY, Devi S (2008) Complete genome sequence analysis of dengue virus type 2 isolated in Brunei. *Virus Res* 135: 48–52
- [7] Patterson JW. Weedon's skin pathology E-book. Elsevier Health Sciences; 2014 Dec 7.
- [8] Verma M, Phartyal R, Bhatt A. Introduction to Flaviviruses and Their Global Prevalence. In *Human Viruses: Diseases, Treatments and Vaccines 2021* (pp. 411-439). Springer, Cham.
- [9] Patterson, Jessica, et al. 'Dengue, Zika and Chikungunya: Emerging Arboviruses in the New World'. *Western Journal of Emergency Medicine*, vol. 17, no. 6, Nov. 2016, pp. 671–79.
- [10] Sukhralia, Shivani, et al. 'From Dengue to Zika: The Wide Spread of Mosquito-Borne Arboviruses'. *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 38, no. 1, Jan. 2019, pp. 3–14.

- [11] Fonseca, Vagner, et al. 'A Computational Method for the Identification of Dengue, Zika and Chikungunya Virus Species and Genotypes'. *PLoS Neglected Tropical Diseases*, vol. 13, no. 5, May 2019, p. e0007231.
- [12] Lennemann, Nicholas J., and Carolyn B. Coyne. 'Dengue and Zika Viruses Subvert Reticulophagy by NS2B3-Mediated Cleavage of FAM134B'. *Autophagy*, vol. 13, no. 2, Feb. 2017, pp. 322–32.
- [13] Xie, Li-Min, et al. 'Identification of Potential Biomarkers in Dengue via Integrated Bioinformatic Analysis'. *PLoS Neglected Tropical Diseases*, vol. 15, no. 8, Aug. 2021, p. e0009633.
- [14] Pong, Lian Yih, et al. 'MicroRNA Profiling of Mouse Liver in Response to DENV-1 Infection by Deep Sequencing'. *PeerJ*, vol. 7, Apr. 2019, p. e6697.
- [15] Byers, Nathaniel M., et al. 'Metabolomic Insights into Human Arboviral Infections: Dengue, Chikungunya, and Zika Viruses'. *Viruses*, vol. 11, no. 3, Mar. 2019, p. 225.
- [16] Hasan S, Jamdar SF, Alalowi M, Al Beaiji SM. Dengue virus: A global human threat: Review of literature. *Journal of International Society of Preventive & Community Dentistry*. 2016 Jan;6(1):1.