MOLECULAR EPIDEMIOLOGY OF DENGUE VIRUS INFECTION AND FULL LENGTH CHARACTERIZATION OF PREVALENT GENOTYPES IN PAKISTAN

AMJAD ALI

DEPARTMENT OF GENETICS
HAZARA UNIVERSITY MANSEHRA
2015
MOLECULAR EPIDEMIOLOGY OF DENGUE VIRUS INFECTION AND FULL LENGTH CHARACTERIZATION OF PREVALENT GENOTYPES IN PAKISTAN

SUBMITTED BY

AMJAD ALI
PhD Scholar

Research Supervisor

PROF. DR. HABIB AHMAD
Tamgha-e-Imtiaz
Dean Faculty of Science
Hazara University Mansehra

Co Supervisor

DR. IJAZ ALI
Assistant Professor
Institute of Biotechnology and Genetic Engineering
The University of Agriculture Peshawar

DEPARTMENT OF GENETICS
HAZARA UNIVERSITY MANSEHRA
2015
HAZARA UNIVERSITY, MANSEHRA

APPROVAL SHEET OF THE MANUSCRIPT

PHD THESIS SUBMITTED BY

Name: Amjad Ali
Fathers Name: Muhammad Ibrahim
Date of Birth: 21-07-1982
Postal Address: Department of Genetics, Hazara University Garden Campus, Mansehra
Permanent Address: Village Binkat P/O Fateh Pur Tehsil Khwaza Khela District Swat
Telephone: 0346-9457223 Residence: 0946-841223
Email: amjadaliswat@gmail.com
PhD Thesis title: Molecular epidemiology of Dengue virus infection and full length characterization of prevalent genotypes in Pakistan

Language in which the thesis has been written: English

APPROVED BY

1. Prof. Dr. Habib Ahmad TI (Supervisor)
2. Dr. Ijaz Ali (Co-supervisor)

RECOMMENDED BY

1. Prof. Dr. Habib Ahmad TI
   Dean Faculty of Science
2. Prof. Dr. Fida Muhammad Abbasi
   Department of Genetics
3. Dr. Hakim Khan
   Associate Professor
   Department of Genetics
4. Dr. Inamullah
   Assistant Professor
   Department of Genetics
In the Name of Allah, the Most Merciful, the Most Gracious.
DEDICATION

I DEDICATE THIS PIECE OF WORK TO MY GRANDPARENTS, PARENTS AND MY FAMILY ESPECIALLY TO MY LATE GRANDMOTHER. MAY HER SOUL REST IN PEACE.

AMJAD ALI
# CONTENTS

LIST OF ABBREVIATIONS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

ABSTRACT

Chapter 1

INTRODUCTION

1.1 THE DISEASE

1.1.1 Etymology

1.1.2 Symptoms

1.1.3 Pathogenesis

1.1.4 Tropism

1.1.5 The humoral immune response

1.1.6 The cellular immune response

1.1.7 Cytokines in dengue pathogenesis

1.1.8 Treatment and therapeutic approaches

1.1.9 Classification of the disease

1.1.10 Brief History of disease and virus

1.1.11 Vectors (mosquitoes) and Transmission Cycles

1.1.12 Geographical Distribution

1.1.13 Factors affecting the spread of the disease

1.1.14 Economic importance of dengue
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2 DENGUE, THE VIRUS</td>
<td>27</td>
</tr>
<tr>
<td>1.2.1 Taxonomic classification</td>
<td>27</td>
</tr>
<tr>
<td>1.2.2 Morphology of the virion</td>
<td>29</td>
</tr>
<tr>
<td>1.2.3 Dengue genomic organization</td>
<td>30</td>
</tr>
<tr>
<td>1.2.4 The structural proteins C, prM, and E</td>
<td>32</td>
</tr>
<tr>
<td>1.2.5 The non-structural proteins</td>
<td>35</td>
</tr>
<tr>
<td>1.2.6 Genetic diversity of DENV</td>
<td>36</td>
</tr>
<tr>
<td>1.2.7 Pakistan in dengue perspective</td>
<td>43</td>
</tr>
<tr>
<td>Chapter 2</td>
<td></td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>52</td>
</tr>
<tr>
<td>2.1 SAMPLING</td>
<td>52</td>
</tr>
<tr>
<td>2.1.1 2011 outbreak</td>
<td>52</td>
</tr>
<tr>
<td>2.1.2 2013 outbreak</td>
<td>53</td>
</tr>
<tr>
<td>2.2 PRIMER DESIGNING</td>
<td>53</td>
</tr>
<tr>
<td>2.2.1 Diagnostic primers</td>
<td>53</td>
</tr>
<tr>
<td>2.2.2 Primers for individual genes and complete genome amplification</td>
<td>54</td>
</tr>
<tr>
<td>2.3 RNA ISOLATION AND DETERMINATION OF THE ACTIVE INFECTION</td>
<td>55</td>
</tr>
<tr>
<td>2.4 DENGUE VIRUS SEROTYPE 3 ANALYSIS</td>
<td>56</td>
</tr>
<tr>
<td>2.5 RNA EXTRACTION AND cDNA SYNTHESIS FOR FULL LENGTH AMPLIFICATION</td>
<td>56</td>
</tr>
<tr>
<td>2.6 COMPLETE GENOME AMPLIFICATION</td>
<td>57</td>
</tr>
<tr>
<td>2.7 SEQUENCING</td>
<td>58</td>
</tr>
<tr>
<td>2.8 DETERMINATION OF THE OPEN READING FRAME (ORF) AND AMINO ACIDS PROFILE</td>
<td>59</td>
</tr>
<tr>
<td>2.9 DETERMINATION OF THE CONSERVED DOMAIN</td>
<td>59</td>
</tr>
</tbody>
</table>
2.10 PHYLLOGENETIC ANALYSIS

2.11 NUCLEOTIDES AND AMINO ACIDS SUBSTITUTIONS

2.12 NUCLEOTIDE AND AMINO ACIDS IDENTITIES DETERMINATION

2.13 SELECTION PRESSURE ANALYSIS ON PAKISTANI ISOLATES

2.14 RECOMBINATION ANALYSIS

Chapter 3

RESULTS

3.1 SEROTYPE SPECIFIC PCR

3.1.1 2011 outbreak

3.1.2 DENV 2013 outbreak

3.2 GENOME ANALYSIS OF DENGUE VIRUS

3.2.1 Genome analysis of the DENV-2

3.2.1.1 Open reading frame and encoded amino acid profile

3.2.1.2 Conserved domains architecture

3.2.1.3 Phylogenetic analysis

3.2.1.4 Nucleotide and amino acid identities

3.2.1.5 Recombination analysis

3.2.1.6 Selection pressure analysis

3.2.1.7 Phylogenetic analysis of DENV-2 based on NS3 gene
3.2.2 GENOME ANALYSIS OF DENGUE VIRUS TYPE 3

3.2.2.1 Phylogenetic analysis of DENV-3

3.2.2.2 Nucleotides and amino acids differences in the Core protein gene of DENV-3

3.2.2.3 Selection pressure analysis on DENV-3

Chapter 4
DISCUSSION

4.1 DISEASE ASSESSMENT

4.2 EPIDEMIOLOGY

4.3 CHARACTERIZATION OF THE VIRUS & PHYLOGENY

CONCLUSION AND RECOMMENDATIONS

FUTURE EXPERIMENTS

REFERENCES

ANNEX-I

ANNEX-II

ANNEX-III

ANNEX-IV

ANNEX-V

ANNEX-VI

ANNEX-VII

ANNEX-VIII
**ABBREVIATION**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADE</td>
<td>Antibody-Dependent Enhancement</td>
</tr>
<tr>
<td>aLRT</td>
<td>Approximate Likelihood-Ratio Test</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>C protein</td>
<td>Core protein</td>
</tr>
<tr>
<td>CCL4</td>
<td>Chemokine Ligand 4</td>
</tr>
<tr>
<td>CDD</td>
<td>Conserved Domain Database</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CDs</td>
<td>Coding Sequence</td>
</tr>
<tr>
<td>CFA</td>
<td>Cell Fusion Agent</td>
</tr>
<tr>
<td>CHIKV</td>
<td>Chikungunya Virus</td>
</tr>
<tr>
<td>DALYs</td>
<td>Disability-Adjusted Life Years</td>
</tr>
<tr>
<td>DCs</td>
<td>Dendritic Cells</td>
</tr>
<tr>
<td>DC-SIGN</td>
<td>Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin</td>
</tr>
<tr>
<td>DENV</td>
<td>Dengue virus</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethylpyrocarbonate</td>
</tr>
<tr>
<td>DF</td>
<td>Dengue Fever</td>
</tr>
<tr>
<td>DHF</td>
<td>Dengue Hemorrhagic Fever</td>
</tr>
<tr>
<td>DSS</td>
<td>Dengue Shock Syndrome</td>
</tr>
<tr>
<td>E protein</td>
<td>Envelop protein</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-Linked Immunosorbent Assay</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic Reticulum</td>
</tr>
<tr>
<td>FcγR</td>
<td>Fc-gamma Receptors</td>
</tr>
<tr>
<td>FEL</td>
<td>Fixed Effects Likelihood</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>GARD</td>
<td>Genetic Algorithm for Recombination Detection</td>
</tr>
<tr>
<td>GTR</td>
<td>General Time Reversible</td>
</tr>
<tr>
<td>hCF</td>
<td>Human Cytotoxic Factor</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C Virus</td>
</tr>
<tr>
<td>HI</td>
<td>Hemagglutination Inhibition</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>IBGE</td>
<td>Institute of Biotechnology &amp; Genetic Engineering</td>
</tr>
<tr>
<td>ICTV</td>
<td>The International Committee on Taxonomy of Viruses</td>
</tr>
<tr>
<td>IFNα/β</td>
<td>Interferon alpha/beta</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IgM</td>
<td>Immunoglobulin M</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IRFs</td>
<td>Interferon Regulatory Factors</td>
</tr>
<tr>
<td>JEV</td>
<td>Japanese Encephalitis Virus</td>
</tr>
<tr>
<td>KD</td>
<td>KiloDalton</td>
</tr>
<tr>
<td>KP</td>
<td>Khyber Pakhtunkhwa</td>
</tr>
<tr>
<td>KUN</td>
<td>Kunjin</td>
</tr>
<tr>
<td>M</td>
<td>Membrane</td>
</tr>
<tr>
<td>MDA5</td>
<td>Melanoma Differentiation-Associated gene5</td>
</tr>
<tr>
<td>MEGA</td>
<td>Molecular Evolutionary Genetics Analysis</td>
</tr>
<tr>
<td>MIP1β</td>
<td>Macrophage Inflammatory Protein 1 Beta</td>
</tr>
<tr>
<td>MTase</td>
<td>Methyltransferase</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NF-kB</td>
<td>Nucleic Factor-kappa B</td>
</tr>
<tr>
<td>NK cells</td>
<td>Natural Killer cells</td>
</tr>
<tr>
<td>NO</td>
<td>Nitric oxide</td>
</tr>
<tr>
<td>NS proteins</td>
<td>Nonstructural proteins</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>PrM</td>
<td>Precursor of Membrane</td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>REL</td>
<td>Random Effects Likelihood</td>
</tr>
<tr>
<td>RIG-1</td>
<td>Retinoic acid Inducible Gene 1</td>
</tr>
<tr>
<td>RT</td>
<td>Reverse Transcriptase</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcriptase PCR</td>
</tr>
<tr>
<td>SDs</td>
<td>Coding sequence</td>
</tr>
<tr>
<td>SH-like</td>
<td>Shimodaira–Hasegawa-like</td>
</tr>
<tr>
<td>SLAC</td>
<td>Single Likelihood Ancestor Counting</td>
</tr>
<tr>
<td>SLE</td>
<td>St. Louis encephalitis virus</td>
</tr>
<tr>
<td>STH</td>
<td>Saidu Teaching Hospital</td>
</tr>
<tr>
<td>TBE complex</td>
<td>Tick-borne Encephalitis virus complex</td>
</tr>
<tr>
<td>TBEV</td>
<td>Tick-Borne Encephalitis Virus</td>
</tr>
<tr>
<td>TCAG</td>
<td>The Centre for Applied Genomics</td>
</tr>
<tr>
<td>TGN</td>
<td>Trans-Golgi Network</td>
</tr>
<tr>
<td>TLRs</td>
<td>Toll-like receptors</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumor Necrosis Factor alpha</td>
</tr>
<tr>
<td>TS</td>
<td>Type Specific</td>
</tr>
<tr>
<td>UNEP</td>
<td>United Nations Environment Program</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WN</td>
<td>West Nile</td>
</tr>
<tr>
<td>WNV</td>
<td>West Nile Virus</td>
</tr>
<tr>
<td>YFV</td>
<td>Yellow Fever Virus</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO</th>
<th>DESCRIPTION</th>
<th>PAGE NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Grades of DHF according to WHO</td>
<td>18</td>
</tr>
<tr>
<td>1.2</td>
<td>Lengths (in amino acids) of DENV structural and non-structural proteins</td>
<td>31</td>
</tr>
<tr>
<td>1.3</td>
<td>NS proteins and their role</td>
<td>36</td>
</tr>
<tr>
<td>1.4</td>
<td>Genotypes of DENV-1</td>
<td>39</td>
</tr>
<tr>
<td>1.5</td>
<td>Genotypes of DENV-2 according to Twiddy <em>et al.</em>, 2002</td>
<td>41</td>
</tr>
<tr>
<td>1.6</td>
<td>Genotypes of DENV-3 as described by Lanciotti <em>et al.</em>, 1994</td>
<td>42</td>
</tr>
<tr>
<td>1.7</td>
<td>Genotypes of DENV-4</td>
<td>43</td>
</tr>
<tr>
<td>3.1</td>
<td>Distribution of Dengue virus (DENV) serotypes during the 2011 and 2013 outbreaks in Pakistan</td>
<td>68</td>
</tr>
<tr>
<td>3.2</td>
<td>Conserved domains and functional sites of DENV-2 characterized in this study</td>
<td>104</td>
</tr>
<tr>
<td>3.3</td>
<td>Percent nucleotide sequence identity among full length DENV2, cosmopolitan genotypes</td>
<td>108</td>
</tr>
<tr>
<td>3.4</td>
<td>Percent amino acids sequence identity among full length DENV2, cosmopolitan genotypes</td>
<td>109</td>
</tr>
<tr>
<td>3.5</td>
<td>Results of the selection pressure analysis acting on the ORF of the DENV-2 cosmopolitan genotypes</td>
<td>111</td>
</tr>
<tr>
<td>3.6</td>
<td>Nucleotides substitution of Core protein among Pakistani isolates of DENV-3</td>
<td>118</td>
</tr>
<tr>
<td>3.7</td>
<td>Conserved amino acids in the Core protein of DENV-3 Pakistani Isolates</td>
<td>118</td>
</tr>
<tr>
<td>3.8</td>
<td>Analysis of the selection pressure on codons of the E gene of DENV-3</td>
<td>120</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO</th>
<th>DESCRIPTION</th>
<th>PAGE NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Types of the disease on the basis of clinical outcome (WHO, 1997)</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Methods of diagnosis and duration of the dengue infection</td>
<td>7</td>
</tr>
<tr>
<td>1.3</td>
<td>ADE mechanism for the replication of dengue (Whitehead et al., 2007)</td>
<td>9</td>
</tr>
<tr>
<td>1.4</td>
<td>T-cells response is changed by the variant epitopes</td>
<td>12</td>
</tr>
<tr>
<td>1.5</td>
<td>Pathogenesis of DHF and DSS</td>
<td>15</td>
</tr>
<tr>
<td>1.6</td>
<td>DENV transmission cycle</td>
<td>21</td>
</tr>
<tr>
<td>1.7</td>
<td>Areas affected by dengue throughout the world in 2005</td>
<td>23</td>
</tr>
<tr>
<td>1.8</td>
<td>Regions where the risk of dengue outbreaks persists</td>
<td>24</td>
</tr>
<tr>
<td>1.9</td>
<td>Number of dengue cases and countries affected from 1955 to 2007</td>
<td>25</td>
</tr>
<tr>
<td>1.10</td>
<td>ML tree of the E gene (123 flaviviruses) rooted on Aedes albopictus cell fusion agent (CFA) virus</td>
<td>29</td>
</tr>
<tr>
<td>1.11</td>
<td>Structures of the virion and E protein</td>
<td>30</td>
</tr>
<tr>
<td>1.12</td>
<td>(A). Domains of dengue E protein. Red is domain I, domain II is yellow, domain III is blue. (B). This is the conformation of an E dimer in the mature virus particle and in solution above the fusion pH. (C). Packing of E on the surface of the virus</td>
<td>34</td>
</tr>
<tr>
<td>1.13</td>
<td>(top) Genetic organizations, (bottom) configuration of the membrane and cleavage sites. Arrows indicate viral and cellular proteases</td>
<td>35</td>
</tr>
<tr>
<td>1.14</td>
<td>Factors responsible for genetic diversity</td>
<td>36</td>
</tr>
<tr>
<td>1.15</td>
<td>15 E gene-based minimum evolution tree using 554 flaviviruses (Unrooted)</td>
<td>38</td>
</tr>
<tr>
<td>2.1</td>
<td>Ampification and sequencing methodology for the complete genome characterization of DENV-2</td>
<td>58</td>
</tr>
<tr>
<td>3.1</td>
<td>Gel electrophoresis of the RT PCR of the second round amplification with type specific primers for genotyping</td>
<td>66</td>
</tr>
<tr>
<td>3.2</td>
<td>Proportion of anti-DENV IgG and/or IgM positive and DENV RNA-positive patients’ serum samples</td>
<td>70</td>
</tr>
<tr>
<td>3.3</td>
<td>Distribution of the DENV genotypes in patients of 2013 Outbreak</td>
<td>70</td>
</tr>
<tr>
<td>3.4</td>
<td>Phylogenetic tree generated by maximum Likelihood analysis of nucleotide sequences from the entire genomes of 67 DENV-2 viruses</td>
<td>76</td>
</tr>
<tr>
<td>3.5</td>
<td>GARD analysis plot of the cosmopolitan DENV-2 genotypes</td>
<td>80</td>
</tr>
<tr>
<td>3.6</td>
<td>Substitution summary of DENV-2 cosmopolitan genotypes by residue</td>
<td>82</td>
</tr>
<tr>
<td>3.7</td>
<td>Phylogenetic analyses of the DENV-2 Swat isolates, based on NS3 genes</td>
<td>84</td>
</tr>
<tr>
<td>3.8</td>
<td>Phylogenetic tree generated by maximum Likelihood analysis of nucleotide sequences from the complete E gene sequences of 57 dengue 3 viruses</td>
<td>86</td>
</tr>
<tr>
<td>3.9</td>
<td>Phylogenetic tree generated by maximum Likelihood analysis of nucleotide sequences from the complete NS3 gene sequences of 57 dengue 3 viruses</td>
<td>87</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

This thesis, the invaluable source of joy and satisfaction for me, is the end of the journey in obtaining my PhD. The God Almighty endowed me with His blessings and that’s why I reached this stage and achieved the most important goal of my life.

The Higher Education Commission (HEC), Government of Pakistan awarded me scholarship under Indigenous-5000 PhD Fellowships Program, without the availability of such grant; it would be too hard for me to complete my PhD in such a better way. The inspiration, I received from HEC in terms of fellowship, support to lab and support to University played a pivotal role in the successful accomplishment of the PhD program. Another outstanding facility provided by the HEC was abroad scholarship under IRSIP Program which supported me to visit the Department of Cell and Systems Biology, University of Toronto, Canada for refining my research findings. This was the most pleasant experience of my life and the memories will always be surrounding me. The Higher Education Commission, Government of Pakistan is acknowledged for all these.

At this moment, first of all, I would like to thank my PhD thesis supervisor, Prof. Dr. Habib Ahmad, Dean Faculty of Sciences and Chairman Department of Genetics, Hazara University, Mansehra, for the guidance, support and providing me the opportunity to work under his supervision. No doubt, he remained a source of enlightenment and motivation not only for me but for the community as well. He has innumerable contributions to the University in general and to the Department of Genetics in particular. Completing my PhD as well as MPhil studies under his supervision is really an honor and privilege for me. I learned a lot from him during the entire period my studies. It is due to his endeavors that students like me complete the projects and fulfill their dreams. To be honest, starting right from my enrollment in PhD to the present stage, all made possible due to the keen interest been shown by Prof. Dr. Habib Ahmad in developing my career. I still remember the day
when he picked me up as his PhD student and provided me the opportunity
to work under his guidance. In fact, I have no words to thank Prof. Dr. Habib
Ahmad, the way he supported me and paved the ways out of impossibilities
in reaching this stage.

I would like to extend my gratitude to my co-supervisor, Dr. Ijaz Ali,
Assistant Professor, Institute of Biotechnology and Genetic Engineering, The
University of Agriculture, Peshawar for designing my research layout and
support. Dr. Ijaz Ali provided me the space to work in his laboratory and
guided me throughout during my research work. He put all possible efforts
and endeavors in the successful completion of my PhD research work. He
extended full cooperation and support in terms of research methodology and
lab facilities. I was given a free hand in using every available facility in his lab.
In fact, it was a good time I spent in the laboratory of Dr. Ijaz Ali.

Adding to the context, I would like to extend my heartiest thanks to Prof. Dr.
AbouHaider, Professor of Virology, Department of Cell and Systems Biology,
University of Toronto, Canada for providing me the opportunity to be a part
of his team. He accepted me as a visiting PhD student and provided me all the
facilities in his lab. I worked under his supervision and learned a lot, starting
from basic molecular techniques to the advanced techniques like sequencing.
This was the wonderful and unforgettable experience of my life to be among
the world class scientists having advanced laboratory, equipped with modern
and sophisticated scientific instruments. I found Professor AbouHaider as the
most cooperative and nice man. He guided me throughout during my stay
there in Canada and equipped me with knowledge. It would not be an
exaggeration to say that I came to know what actually science means in a true
sense during my stay with Prof. AbouHaider. Besides helping me out in my
research, he gave me the permission to attend his lectures on virology in the
University for the whole period I spent there. It was an extraordinary
situation for me to be among the Canadian students and attending lectures of
the world renowned scientist. I absorbed all the facilities provided by the
University of Toronto for their students and all this made possible due to the cooperation extended by Prof. AbouHaider. In short, I have no words to thank Prof. AbouHaider for his invaluable support.

I am really thankful to Mr. Tauqeer Ahmad, PhD student, Department of Cell and Systems Biology, University of Toronto, for his support, guidance and company. I learned many basic techniques used in the laboratory from him.

I am highly obliged to all my family members for their unconditional support, encouragement and financial assistance. They extended every type of support whether moral or financial that led me to this stage. They kept all their problems aside and always prioritized my tasks. Without their keen interest and support, it was impossible for me to complete my PhD.

I am thankful to Mr. Rafi-ur-Rehman, Mr. Zahir Mian, Mr. Zeeshan (IBGE students), Mr. Hasan Riaz (PhD student, Arid Agriculture University, Rawalpindi) and Mr. Muhammad Ali (PhD student, NUST, Islamabad), for their cooperation and support.

This thesis has been kept on track and been seen through to completion with the support and encouragement of numerous people including my teachers, my well-wishers, friends, colleagues, my family and various institutions. I would like to thank all those people who made their contributions in many ways which led me towards the successful completion of my research and thesis.

AMJAD ALI
ABSTRACT

Dengue fever is one of the important and most common mosquito borne diseases in the tropical and subtropical areas of the world. Since 1994, the dengue fever is appearing as a fatal health hazard in Pakistan. Two severe outbreaks of dengue were recorded during the years-2011 in Punjab Province and 2013 in Swat District, wherein hundreds of people died and 1000s more remained with the potential risk of the fatal fever. This thesis encompasses molecular biology of the major dengue outbreaks of Pakistan in general and the first hand information on the dengue outbreak of the year 2013 in Swat District. In order to characterize the dengue virus (DENV), blood samples from 1795 dengue patients were obtained and analyzed through RT-PCR, for its epidemiology, evolutionary history and prevalent genotypes/serotypes of the outbreaks. Phylogenetic analyses were carried out through MEGA and PhyML software while selection pressure was determined through SLAC, REL and FEL methods implemented in Datamonkey software. Whereas Recombination Event Detection was analysed through GARD method and nucleotide and amino acid identities were determined through Clustal Omega. Results of the study were based upon the characterization of 4 complete sequences of DENV-2, 7 NS3 genome sequences of DENV-2 and 10 each of the C, E and NS3 gene sequences, of DENV-3. The results revealed that the active dengue infection in Swat District was 30.5% with the prevalence of DENV-2 and DENV-3 as 36.6 and 62.8 percent, respectively which shows that the major serotypes circulating in Swat were DENV-2 and DENV-3. Furthermore
DENV-1 and DENV-4 were not detected in any of the sample collected from Swat. In Punjab Province during the 2011 outbreak the active infection was 58.5%, of which the prevalence of DENV-2 and DENV-3 was 41.8 and 38.0 percent, respectively, which figured-out that these two serotypes were mainly involved in the dengue outbreak of the year-2011. A limited incidence of DENV-4 and DENV-1 in 2011 outbreak in Punjab at the rate of 9.5 and 4.5 percent respectively was also recorded. Phylogenetic analysis based on both complete genomes and the NS3 gene sequences of DENV-2 grouped the Pakistani isolates as cosmopolitan genotypes. Whereas no recombination was detected in the entire genome of DENV-2 of Pakistani isolates. The phylogenetic analyses, percent nucleotides and amino acids identities of the complete genomes of DENV-2 indicated that the 2013 dengue outbreak in Swat was a continuation of the 2011 dengue infections in Punjab. Furthermore the Pakistani isolates were found to be genetically more close to the Indian isolates based upon the nucleotide (97.1%) and amino acid (98.5%) identities. The selection pressure analyses revealed no positively selected sites in any of the genes of the DENV-2 among Pakistani isolates. However codon 65 of C, 155 of PrM, 203 of E and 254 of NS1 genes were under negative selection pressure. The highest substitution rate observed among amino acids residues of the ORF of all DENV-2 Pakistani isolates, was between arginine and lysine residues, followed by valine and isoleucine. Phylogenetic analysis of DENV-3 on the basis of the complete E gene and NS3 gene sequences grouped the Pakistani isolates into the genotype III; recognized that the year’s 2013
outbreak in Swat was the continuation of the previous outbreaks in Pakistan and the Pakistani DENV-3 isolates are more closely related to Indian and Chinese isolates. Analyses of the C gene of DENV-3 of the year-2011 were mostly conserved for lysine, arginine and asparagine. The viruses of the 2013 outbreak of Swat District showed the amino acids change for the nonpolar P (proline) and polar K (lysine) instead of R and N in the C gene. All the viruses isolated from the year-2011 dengue outbreak of Punjab showed a conserved pattern of nucleotides/amino acids as compared to the DENV-3 isolates of the dengue outbreak in Swat. This pattern may be due to one of the reasons that viruses established a severe infection in the relatively colder area of Swat and its adaptation to the new environment. Selection pressure analyses of NS3 genes of DENV-3 revealed no negatively or positively selected sites using all the three analytical tools. However codon 94 and 477 of the E genes turned out to be under negative selection pressure. From the selection pressure analysis it is assumed that the emergence and spread of dengue in Pakistan is more due to permissive ecological conditions than the drastic genetic changes in the viral genome. Hence it is recommended that concerted efforts in terms of public awareness, vector control and household protection measures against dengue should be adopted to check its spread to other areas of Pakistan.
Chapter 1
INTRODUCTION

1.1 THE DISEASE

The disease dengue is the common and most important mosquitoes transmitted viral disease in the world. Around 2.5 billion inhabitants in tropics and subtropics of the world are living with the risk of getting the infection, makes up to 2/5th of the world’s population (Dash et al., 2013, WHO, 2009; Gubler & Clark, 1995). An estimated global rate of dengue infection is 50-100 million every year with 24,000 mortalities and 500,000 cases requiring hospitalizations (Halstead, 1988; WHO, 1997). The recent study conducted on dengue infection, suggests that annually 400 million people are infecting with the disease around the globe, out of which 100 million are clinically apparent (Bhatt et al., 2013). Furthermore, the fact that the number of infections be doubled in those people who are living in the tropical and subtropical areas (Anonymous, 2009), makes the disease a major threat around the globe.

1.1.1 Etymology

According to Christie, 1881, the word ‘dengue’ is cosidered to be taken from Swahili term ‘ki denga pepo’, means an abrupt and sudden attack of an evil spirit. Due to intense joint pain experienced by the dengue patients, the breakbone fever is also frequently used for this disease.
1.1.2 Symptoms

Dengue caused by one of the four known serotype of DENV (DENV1-4), is an acute febrile disease. Mostly infections caused by dengue virus go asymptomatic whilst the rest develops a wide spectrum of disease from mild undifferentiated called classical dengue fever (DF) to the severe form of the disease that is dengue haemorrhagic fever (DHF) as well as dengue shock syndrome (DSS) as indicated in figure 1.1.

Fatigue, headache, fever, rash, retro-ocular pain, myalgia, vomiting, arthralgia, nausea and leukopenia are the common symptoms associated with classic dengue fever. In extreme cases, severe hemorrhage, brain, abnormal and liver malfunctions occur (Rigau-Perez et al., 1998).

Figure 1.1 types of the disease on the basis of clinical outcome (WHO, 1997).
Cardinal signs of the DHF are thrombocytopenia (low platelet count), plasma leakage and hemorrhagic complications. According to Halstead, 2007, increased vascular permeability associated with DHF is considered to be the result in the increase in cytokines whenever dengue-infected cells attacked by T-lymphocytes.

DSS, a severe and fatal type of the disease is characterized by intense and prolong pain in the abdomin, restlessness or lethargy, insistent vomiting, an abrupt change from hyperthermia to hypothermia, sweating and exhaustion along with all the symptom of DF and DHF. Finally due to very low blood pressure, shock is evident in DSS (Rigau-Perez et al., 1998).

1.1.3 Pathogenesis

DENV infection having a wide clinical spectrum is a systemic and dynamic disease. The major phatological findings in DHF and DSS are the hemorrhages in subcut tissues, skin, heart and gastrointestinal tract (Bhamarapravati, 1989). Hemorrhages, congestion as well as dilatation of blood vessels in the infection are common while hemorrhagic complications in rest of the body organs, with accumulation of fluids in body cavities, might be consequential (Hotta, 1969, Bhamarapravati, 1967).

The basic mechanism responsible for vascular leakage and hemorrhage is still not completely understood. Increased levels of vasoactive and pro-inflammatory cytokines in blood of pateints suffering from DHF at times of plasma leakage imply that vascular permeability is due to the excessive cytokine production (cytokine storm).
Data, available on DENV infection indicate that the fate of the disease relies upon the nature of the immune response, whether it is favourable or unfavourable. The favourable immune response controls the replication of the virus while the unfavourable response enhances inflammatory and vascular permeability. The main hindrances in the understanding of dengue pathogenesis are the unavailability of credible markers (immunological) for pathological as well as protective immune responses to virus and suitable animal model for dengue disease, leaving behind the options to look at the clinical and epidemiological studies for the understanding of immune response against DENV infection.

1.1.4 Tropism

The target cells for DENV replication are yet not clearly defined. The initial round of viral replication, after inoculation by an infected mosquito (incubation period from 3 and 14 days) is considered to come off in the subdermal Langerhans DCs (Ho et al., 2001, Libraty et al., 2001, Marovich et al., 2001, Wu et al., 2000). Then these cells after activation migrate to the draining lymph nodes (Johnston et al., 2000). The activated DCs call forth a robust IFNα/β and tumor necrosis factor alpha (TNFα) response along with an intense pro-inflammatory response to bound the dissemination of infection (Libraty et al., 2001). Replication of the virus still continues in undefined cells of lymph node. The macrophage-monocyte lineage is generally considered to be the candidate cell types. Human biopsies and autopsies analysis indicate that mononuclear phagocytic cells are probably the basic targets of DENV
infection, after the initial spread from the local skin site. DENV has also been recovered from infiltrating mononuclear cells in infected tissues (Boonpucknavig et al., 1979; Boonpucknavig et al., 1976; Sahaphong et al., 1980) and occasionally from peripheral blood leukocyte fractions (Scott et al., 1980).

Virus replication continues in the lymph nodes and leads to viremia if the virus gets entry into blood circulation through thoracic duct and efferent lymphatic system. When viremia is at its peak, the patient subsequently enters the painful febrile period. After the start of fever, viremia ends in 5-7 days, coexisting with defervescence. The common consensus is that viremia infects circulating monocytes in the blood, which further facilitates the spread of infection to secondary visceral organs and causes infection in the macrophages within liver, bone marrow and spleen. This is the crucial time because the sever form of the disease that is DHF and DSS more often develops at this time. The patients should be kept under strrrict observation and care as this is the most crucial time of the infection. Lack of proper attention to the patients, may results in shock and death within 24 hours if proper health measures were not taken into account. Dectection of IgG and IgM levels in blood are the primary diagnostic procedures used in the laboratory after defervescence (Jessie et al., 2004; Diamond et al., 2003; Solomon & Vaughn, 2002; Xiao et al., 2001a; Xiao et al., 2001b; Durbin et al., 2008). Schematic representation of the progression of the dengue disease is given in figure 1.2.
Dissemination of infection to solid organs is limited and inconsistent (Marchette et al., 1973). DENV antigen has been recovered from lymphocytes, hepatocytes, cerebral neurons, endothelium and astrocytes (Bhoopat et al., 1996; Jessie et al., 2004; Couvelard et al., 1999; Huerre et al., 2001; Miagostovich et al., 1997; Hall et al., 1991; Ramos et al., 1998).

There is controversy regarding the function of endothelial cells as the primary target for dengue virus. Systemic endothelial dysfunction associated with vascular leakage occurs in the severe form of the disease that is DHF and DSS but inspite of this generally the destructive lesion are not evident in fatal cases (Gubler, 1998a). According to Andrews et al., 1978 and Avirutnan et al., 1998, for DENV infection to persist, permissiveness of the endothelial cells of human as well as cell lines are evident but on the other hand apparently it is not neccessary that endothelial infection be needed for severe pathologic changes in human tissues (Balsitis et al., 2009).

Infection of various organs and cell types with dengue virus indicate that hosts receptors are broadly distributed. For DENV, the host receptors are believed to be chondroitin sulphate, Mannose binding protein, heparan sulphate and DC-SIGN (Wang et al., 2011; Miller et al., 2008; Avirutnan et al., 2007). Viremia and tissue dissemination is considered to be controlled initially by antibodies like IgM, complement and may be NK cells up to some extent after DENV infection. Cytotoxic T-lymphocytes recognize the infected cells and then the cellular immune system attack on it.
1.1.5  The humoral immune response

A person can contract dengue disease four times throughout his or her entire life, once for each serotype. Infection with any serotype of DENV whether primary (first) or secondary (subsequent), can results into either DF or DHF (Rosen, 1977).

For controlling DENV infection and dissemination, the humoral immune response is considered to be vital. Furthermore, disease with one serotype contributes to the ever-lasting immunity to the same infecting serotype and is called homotypic immunity while on the other hand if the same person is infected with another serotype, the immunity then, is short-lived, harmful and escalates developing the risk of severe dengue disease that is heterotypic immunity (Sabin, 1950). This phenomenon is called antibody-dependent enhancement (ADE). The common consensus is that in ADE the already
available antibodies to the primary infection which are sub-neutralizing and heterotypic infection give rise to complexes to cells having Fc gamma receptor (FcγR) that is monocytes along with B-lymphocytes. As a result instead of controlling the virus, an increase in the taking in of the virus as well as in replication occurs as indicated in figure 1.3 (Halstead, 1988).

In a recent study in 2010, a group of human monoclonal antibodies against DENV was generated. It was noticed that the immune response to DENV comprised of a large number of antibodies against prM protein which were highly cross-reactive among the dengue serotypes. Further more, it was concluded that these antibodies also play a major role in ADE phenomenon and having very low neutralization ability. This research further illustrates that the partial cleavage of prM decreases the availability of antigen for the antibodies against prM to neutralize the virus and hence the susceptibility of the viruses to ADE increases (Dejnirattisai et al., 2010).

Specific antibodies against the E protein of the virus, cross reactive in nature and protective when a certain threshold is reached, are considered to be the cause of the transient nature of heterotypic immunity (Whitehorn & Simmons, 2011). The main targets of the immune system in terms of antibodies production in response to infection are prM, E structural proteins and NS1 protein of DENV. The non-structural proteins like NS3 and NS5, also been reported to provoke a weak antibody response (Valdes et al., 2000; Churdboonchart et al., 1991). The viral E protein is targeted by neutralizing antibodies that prevent viral attachment, internalization and replication.
within cells. Within each of the three E domains, there are multiple epitopes (Roehrig et al., 1998; Sukupolvi-Petty et al., 2010) however these epitopes of the E protein are not easily accessible for the antibodies to attach with, due to its dimeric conformation on the surface of the virion and tight packaging in the mature state (Lok et al., 2008; Kaufmann et al., 2006; Cherrier et al., 2009).

Figure 1.3 ADE mechanism for the replication of dengue (Whitehead et al., 2007).

Domain III of the E protein which is extensively variable in amino acid composition among dengue virus serotypes, contains sites for binding to host receptors. Consequently, the greatest degree of serotype specificity is observed by antibodies specific for this domain (Lai et al., 2008). Furthermore, mutations are common in domain III of the E protein to escape neutralizing antibody (Lin et al., 1994; Lok et al., 2001). Similarly, sequence variation for C and NS2B proteins is responsible for the loss of an effective neutralizing antibody response (Wang et al., 2002a; Wang et al., 2002b).
Antibodies to DENV sometimes bind to complement proteins which further enhance the activation of these proteins. DENV infection could also be checked by anti-prM or E protein antibody-mediated complement fixation (Mehlhop et al., 2007). One of the features of severe dengue is the complement activation and sometimes related to the leakage. This explains complement activation has a major role in the pathogenesis of DSS (Bokisch et al., 1973; Malasit, 1987). Vascular leakage could also be contributed by the increase in complement activation at endothelial cell surfaces (Avirutnan et al., 2006). The modulator of the complement pathway is considered to be NS1 protein of the virus. Degradation of C4 to C4b indicates that NS1 plays the possible role in the protection of DENV from complement-dependent neutralization in solution (Avirutnan et al., 2010).

1.1.6 The cellular immune response

In dengue pathogenesis cellular immune responses are also vital along with the humoral immunity, can cause infection in CD4+ T cells as well as CD8+ T cells (Mentor & Kurane, 1997). Like specific antibodies to DENV, the cellular immunity may also act as protective in nature and sometimes harmful as well. T-cells response is very disparate in nature, ranging from proliferation and target cell lysis to the production of a diverse group of cytokines. The response of the CD4+ T-cells in production of TNFα, TNFβ, IFNγ, interleukin (IL)-2 and CC-chemokine ligand 4 (CCL4), may contribute to pathogenesis (Gagnon et al., 1999). While on the otherhand, the least detected cytokine group is the T-helper type 2 like IL-4 (Mangada et al., 2004; Bashyam
More CD8+ T-cells are present in uncomplicated DENV infections as a result the production of a less number of IFNγ and TNFα (Duangchinda et al., 2010). According to An et al., 2004, DENV specific CD8+ T-cell clones provides immunity to mice partially against lethal DENV challenge. In dengue disease the role of T-regulatory cells is unclear however according to Luhn et al., 2007, their function is evident and spread in acute DENV infection.

Memory T-cells whether they are serotype specific or serotype cross-reactive are produced, following primary infection. The memory T-cells, wheter protective or cross-reactive activate in response to secondary infection and the memory T-cells which are non-protective will enhance the infection further (Kurane et al., 1990). The conserved as well as altered peptide ligand epitopes are recognized by activated memory T-cells. Differences in the sequence of the antigen rely upon the epitope of DENV however it does not change the characteristic response of the effector T-cells. As a result the immunological respose changes that is considered playing a major role in the development of plasma leakage (Rothman & Ennis, 1999). A complete range of T-cell responses by full agonist peptide involve multiple cytokines production (IFNγ, TNF and CCL4) as well as infected cells while on the other hand there will be memory T-cells cross-reactivity and uneven functional response in a sense that some cytokines will be produced in much higher quantities than some other group of cytokines as well as an inefficient cell lysis to a partial agonist peptide (varying at one residue) as indicated in figure 1.4. Hence due
to the diversity found in the sequences among the serotypes of DENV, the memory T-cells and B cells which reactivated in secondary infection may not have maximum avidity for the epitopes of the subsequent infecting virus. Memory pertaining to the primary DENV infection changes the response of the immune system to the secondary infection which modifies the clinical outcome.

Figure 1.4 T-cells response is changed by the variant epitopes. The top part of the figure shows full agonist response while partial agonist response is shown in the bottom part (Alan et al., 2011).

There is a correlation between the T-cells response and disease severity. This mechanism of low affinity for the current infecting serotype but a high affinity for a past infection with a different serotype is called the original antigenic sin which is due to the net effect of the altered balance between a protective and pathological outcome (Zivna et al., 2002; Mongkolsapaya et al., 2003; Mangada & Rothman, 2005).
In secondary DENV infections, the pattern of antibody/T-cell responses is also affected by the sequence of the virus as well as gap between infections caused by DENV (Valdes et al., 1999; Halstead et al., 1983; Alvarez et al., 2006b; Guzman et al., 1999).

As the ADE scenario is concerned, after the primary infection long ago, the response of the memory T-cells which is cross reactive and proliferative in response to secondary infection may shift potentially the balance from being a protective immune response to the non protective and improper immune response. Interesting fact is that mostly the known epitopes pertaining to CD4+ and CD8+ T-cells contained in non-structural protein 3 (NS3) that accounts for only about 20% of the ORF (Duangchinda et al., 2010).

1.1.7 Cytokines in dengue pathogenesis

To raise an appropriate antiviral response, the host cells soon after the entry of the virus into the body, cognizes it. DENV sensing is mediated by two main families of receptors for antigen recognition, the extracellular (endosomal) toll-like receptors (TLRs) (Akira & Takeda, 2004; Bowie & Haga, 2005) as well as the cytoplasmic receptor family of DExD/H box RNA helicases (e.g. retinoic acid inducible gene 1 (RIG-1) and melanoma differentiation-associated gene-5 (MDA5) (Meylan & Tschopp, 2006). On binding to TLR, the nucleic factor-kappa B (NF-kB) along with interferon regulatory factors (IRFs) (transcriptional factors) are activated. Production of interferon alpha & beta and pro-inflammatory cytokines which do maturation of DCs by stimulating them as well as evoke an antiviral response
is activated by these signaling cascades (Severa & Fitzgerald, 2007; Libraty et al., 2001).

Cells of the DC/macrophage/monocyte lineage are considered primarily to be infected by DENV through receptor-mediated endocytosis or increased taking in through antibody-virus complexes, bound to Fc-gamma receptors (FcγR) (Anderson, 2003). The precise mechanism of DHF and DSS are still not well comprehended but the general agreement is that the TNFα and nitric oxide (NO) are produced by infected cells and activated endothelial cells that increase vascular wall permeability (Borish & Steinke, 2003; Carr et al., 2003; Espina et al., 2003; Charnsilpa et al., 2005; Neves-Souza et al., 2005). The co-occurrence of the severe form of the disease with suppression of fever and the control of the virus indicates that the outcome of the disease may be the results of the immune response to the virus instead of virus-induced cytopathology. In dengue virus infections, the high levels of various cytokines been studied, is consistent with the above hypothesis (Basu & Chaturvedi, 2008).

The raised levels of cytokines and chemokines in serum of patients include IL-2 (Kurane et al., 1991; Hober et al., 1993; Green et al., 1999), IL-6 (Iyngkaran et al., 1995; Juffrie et al., 2001), IL-8 (Talavera et al., 2004), IL-10 (Libraty et al., 2002a; Green et al., 1999), IL-13, IL-18 (Mustafa et al., 2001), IFNγ (Azeredo et al., 2001, Chakravarti & Kumaria, 2006), TNFα (Iyngkaran et al., 1995, Hober et al., 1996; Bethell et al., 1998; Azeredo et al., 2001) and monocyte chemotactic
protein-1 (MCP-1) (Sierra et al., 2010). In secondary infections and severe disease with DHF/DSS, cytokines are associated (figure 1.5).

Figure 1.5 Pathogenesis of DHF and DSS. Through receptor-mediated endocytosis, DENV initially infects a cell of the DC/macrophage/monocyte or increased uptake through antibody-virus complexes bound to FcγRs. The heterotypic antibody is not protective, but cross-reacting antibodies attach to the virus and enhance macrophage infection through Fc receptor-mediated endocytosis (as in the bottom part). Formation of phlogistic immune complexes occurs through the secondary antibody response to the new infecting DENV plus the enhanced virus output. The resulting activated complement components may cause damage directly, and also through interactions with the coagulation system. The cellular immune system is also important. Cross-reacting epitopes on the new DENV species recognized by the Memory T-cells (Tm) are activated, proliferate, and secrete IFNγ, TNF, and other cytokines. FcRs on the macrophages are upregulated by IFNγ, further increasing the ADE of infection of the dengue target cell. The T-cells also interact with macrophages, with cross-reactive CD8+ T cells able to lyse infected macrophages. The release of this extensive cytokine and other mediators result the abrupt onset of shock due to increased vascular permeability, called DSS (Peters et al., 2002).
Leakage of the plasma caused by the improper function of the vascular endothelial cells is not yet fully understood. If there is ATh1-type response then it leads to the recovery from the acute stage but on the other hand Th2-type response further exacerbates the infection and a poor clinical outcome. Those patients having DF have predominantly a Th1-type response. IL-10 and IFNγ primarily mediate cross regulation of Th1 and Th2 respectively (Mosmann & Sad, 1996). Moreover, CD4+ T-cells are recruited by activated macrophages, producing human cytotoxic factor (hCF), that as a result brings about a cytokine pour out leading to a Th1-type or Th2-type response (Chaturvedi et al., 2000). In severe dengue cases, Levels of hCF could be elevated and hCF auto-antibodies protect against severe disease (Chaturvedi et al., 2001). As the infection becomes severe a Th2-type response is initiated which in turn induces the release of IL-4, IL-5, IL-6, IL-10 and IL-13 into the blood. Those disease conditions which bring about a humoral immune response primarily cause an increase cytokines expression of Th2-type (Mosmann & Sad, 1996; Chaturvedi et al., 2000).

Pathological aspects associated with DHF include increased capillary permeability without morphological damage to the capillary endothelium, increased hematocrit and thrombocytopenia, an alter state of leukocytes both in number and function as well. As a result the thrombocytopenia leads to leakage of the plasma and also it deregulates the coagulation. This deregulated coagulation is supposed to be mediated by cytokines, e.g. TNFα (Chen et al., 2007). IL-6 and IL-8 with high levels in the disease are connected
to the deregulated coagulation as well as fibrinolysis (Lei et al., 2001; Martina et al., 2009). So the common belief is that the permeability of the vascular wall is indirectly the result of the effects produced by the virus. Endothelial cells permeability augmentation as well as increase in the expression of adhesion molecules on the endothelial cells is promoted by secreted TNFα from activated and infected cells (Kallmann et al., 2000; Madan et al., 2002; Dagia & Goetz, 2003; Javaid et al., 2003). Elevated level of IL-10 is associated with reduced levels of platelets and its function (Anderson et al., 1997; Libraty et al., 2002a). Liver may also be infected with DENV resulting in decrease of clotting factors hence the increase in bleeding.

**1.1.8 Treatment and therapeutic approaches**

In clinical use for dengue fever, neither targeted antiviral therapy or vaccines avaiblabe yet. The routine medical assistance provided to the patients, is supportive in nature that means to monitor the patients and to avoid the possible dehydration and shock as well, fluids are administered. To reduce pain, hyperthermia as well as for the control of bleeding related complications, medication is administered. 41.3% fatality rate associated with DHF has been reported in the late 1960s because health care providers were not familiar with the disease (Sumarmo, 1987). Today, without proper treatment, fatality rate due to DHF can outreach 20% but on time and with proper medical management, the rate can be reduced to 1% according to the 1997 report of WHO.
1.1.9 Classification of the disease

Though the word dengue is generally used to represent the whole spectrum of the disease, in 1974 a formal scheme for the classification of the disease has been devised by WHO. According to the scheme the disease, dengue is characterized into DF, DHF and DSS (WHO, 1975). Based on the number of hemorrhagic manifestations, the DHF category is further classified into grades (I-IV) depending upon the severity as indicated in Table 1.1. The DHF grade III along with grade IV are designated as dengue shock syndrome (DSS) in which severe leakage of the plasma occurs.

Table 1.1 Grades of DHF according to WHO

<table>
<thead>
<tr>
<th>Classification</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I</td>
<td>Fever with other symptoms like headache, vomiting, muscle and joint pain, skin rash. Positive tourniquet test is the only evidence of hemorrhaging</td>
</tr>
<tr>
<td>Grade II</td>
<td>Grade I symptoms with spontaneous bleeding</td>
</tr>
<tr>
<td>Grade III</td>
<td>Failure of circulatory system, clammy skin, rapid and weak pulse, restlessness</td>
</tr>
<tr>
<td>Grade IV</td>
<td>No measurable blood pressure or pulse, severe shock</td>
</tr>
</tbody>
</table>

These guidelines developed by WHO in 1960s, grounded on the reports of the Children’s Hospital, Bangkok, Thailand on pediatric cases. Currently there are difficulties for clinicians following the above guidelines for the classification of the disease because dengue can be found around the globe with a parallel change in the demographic picture of the patients (Deen et al., 2006; Rigau-Perez, 2006).
1.1.10 Brief History of disease and virus

Ren Kimura and Susumu Hotta isolated DENV-1 for the first time in 1943 in Japan (Kimura & Hotta, 1943). From 1942 to 1944 during the Second World War era, some port cities of Japan like Nagasaki, Osaka and Kobe, an epidemic of DF had occurred that involved at least 200,000 cases. According to Hotta in 2000, the persons, coming back from the tropics, particularly from Southeast Asian as well as Pacific islands were the cause of transmitting the disease. Albert Bruce Sabin and Walter Schlesinger soon after the detection of DENV-1 in Japan, analyzed Hawaiian samples and detected DENV-1 in them. In another experiment, they isolated DENV-2 by analyzing the samples of Papua New Guinean patients (Sabin & Schlesinger, 1945). Sabin and Schlesinger illustrated that these viruses though antigenically closely related to each other but still different from each other and could be easily differentiated using hemagglutination inhibition (HI) test. On historical grounds, though there were various speculations related to the initial recognition of diseases similar to dengue (Halstead, 1980; Henchal & Putnak, 1990) but it was in 1953, in Manila, Philippines that for the first time DHF recognized (Quinlos et al., 1954).

In 1956, William Hammon isolated viruses from the patients belonged to Manila. These viruses were comparable to DENV-1 and DENV-2 and he distinguished them to be DENV-3 and DENV-4. In 1958 in Bangkok, Thailand, DHF epidemic started and multiple serotypes from the patients were isolated and detected (Hammon et al., 1960).
Southeast Asia was the main hideout of DHF and DSS outbreaks until the early 1980s and the disease was limited to that area (Halstead, 1980). Since then, transmission of dengue has increased many folds and currently in tropical areas of the world, the epidemics of DHF and DSS are more common. Till now, the major means of hospitalization and mortalities in children in the Southeast Asian regions remains to be DHF/DSS. DHF confirmed cases have been reported in twenty four countries of Central America along with South America from 1981 to 1997 (Monath, 1994; Gubler & Clark, 1995; Gubler, 1998).

1.1.11 Vectors (mosquitoes) and Transmission Cycles

Through the bites of infected female mosquitoes, the disease is transmitted from one person to another. DENV, the causative agent of the disease, is thought to have been sustained in sylvatic or enzootic transmission cycles between nonhuman primates and mosquitoes species living in forests. When humans and mosquitos come into contact, the virus was transmitted to humans and hence the shift occurred from primates-mosquitoes to human-mosquitoes cycles near and around human populated areas (Figure 1.6).

DENV is known to be transmitted by various species belong to genus Aedes of the Culicidae family but still Aedes aegypti is the main vector that also transmits the yellow fever virus (YFV). Furthermore, chikungunya, a third arboviral disease, the causative agent of which is chikungunya virus (CHIKV) (an alpha virus of the Togaviridae family) is also transmitted by this species. Symptoms of the Chikungunya disease are similar to dengue that generally hinders in
the exact diagnosis of the two diseases. Thomas Lane Bancroft, the famous naturalist of Australia, based on epidemiological grounds suggested for the first time in 1906 that *Ae. aegypti* is the vector for dengue fever and later on John Burton Cleland in 1916 confirmed this (Cleland *et al.*, 1916).

*Ae. Aegypti*, loves to bite at daytime and preferably breed in domestic and peri-domestic water holders. Its desiccation-resistant eggs and adaptation to human habitats enables it to root in the urban areas.

![DENV transmission cycle](image)

Figure 1.6 DENV transmission cycle (Whitehead *et al.*, 2007).

*Aedes albopictus*, commonly known as the Asian tiger mosquito is the secondary vector for DENV. In Taiwan in 1917, Koizumi *et al.* first identified that in semi tropical localities, it acts as a dengue vector (Kuno, 2007). Countries where *Aedes aegypti* is absent, it is *Ae. albopictus* which acts as a primary vector and in rural areas as maintenance vector in situations in which both species are prevalent (Smith, 1956; Gratz, 2004).
According to Freier & Rosen, 1987 and Rosen et al., 1954 in Pacific islands *Ae. polynesiensis* act as the principle vector for the disease. On the other hand *Ae. scutellaris* acts as the jungle vector (Mackerras, 1946). *Ae. albopictus* like *Ae. aegypti* is equally responsible for carrying CHIKV and is designated as the major epidemic causing vector for chinkungunya in the recent history (Bessaud et al., 2006; Bonilauri et al., 2008).

Both *Ae. Aegypti* and *Ae. Albopictus* been known to be anthropophilic in nature, means they prioritize feeding on humans (Ponlawat & Harrington, 2005). The areas of the tropics and subtropics, designated as urban and semi urban, both the vector species could be found in abundantly. According to Moncayo et al., 2004, for DENV, a great vector competence between the two species has been found.

Currently, effective vector control (fogging for killing adult mosquitoes, larvicides to get rid of the larvae in the aquatic stage or destroying the source that is the breeding habitat) is the possible and only available procedure to tackle with the infection due to absence of vaccines and specific treatment.

1.1.12 Geographical distribution

Dengue is the most spread out disease in the world (Figure 1.7), caused by DENV and transmitted by Aedes mosquitoes. The distribution of the disease in turn linked to the prevalence of its vector and the habitate in which the vector found generally. The main regions where dengue is found in the world are the urban and semi urban areas of the tropical as well as subtropical habitates, famous for the abundant prevalence of the vectors which
consequently contribute to the disease prevalence. In general, in most of these urban areas, dengue is hyper-endemic and all the circulating serotypes of dengue could be found. In non-tropical regions where dengue is found, the returning international travelers from areas where dengue is endemic act as a source of contracting and further spreading the disease.

Figure 1.7 Areas affected by dengue throughout the world in 2005 (Halstead, 2007).

Under naturally changing temperature, the *Ae. Aegypti* are capable to maintain its life cycle due to the fact that the larvae are able to multiply and develop into adults at low temperature like 10°C. On the other hand *Ae. albopictus* are even more fit to survive because the larvae are viable and can result into adults even if the temperature is below 10°C (Tsuda & Takagi, 2001). That is the reason that both vector species could widely be seen in a winter isotherm (10°C) at latitudes 35°N to 35°S according to WHO report of 1997. At the start of the year 2009, dengue outbreak hit Buenos Aires, Argentina, lying very close to the isotherm and is the extreme South dengue ever reached.
Major parts of Australia and Africa and some parts of the US and Europe particularly the southern parts, are the areas where dengue epidemic may occur in future (Figure 1.8).

Figure 1.8 Regions where the risk of dengue outbreaks persists (WHO, 2006)

1.1.13 Factors affecting the spread of the disease

In mid 1950s, the causative agents of DF and DHF were recognized virologically and since then the frequency of dengue epidemics has intensified worldwide (Figure 1.9).

Factors contributing to the distribution of the disease globally include poverty, ineffective public health infrastructure, uncontrolled urbanization, expanding urban population, rapid and quick transportation, world wide business activities and hectic international visits (Gubler & Clark, 1995). The
single most important contributing factor is probably the rapid urbanization; the emerging population centers in majority of the cases have no piped water facility. As a result the people residing the area use water holders for the storage of water that turns into a suitable site for the mosquitoes to breed in. Inadequate sewage system is another problem in this regard.

Figure 1.9 Number of dengue cases and countries affected from 1955 to 2007 (http://www.who.int/disease/dengue/impact/en/index.html)

Another important factor which not only facilitates the dengue outbreak but also plays a major role in the development of sever disease is suggested to be virulence effecting genetic variations in the DENV strains (Rosen, 1977). A common example is in the Western Hemisphere where the Native American genotype of dengue virus serotype 2, has been replaced with another
genotype from the Southeast Asian lineage (Rico-Hesse et al., 1997). According to research, Southeast Asian genotypes of DENV are known to have well adaptation capability in order to be disseminated by Aedes mosquitoes leading to higher viremic manifestations in the dendritic cells of both humans and mosquitoes cell (Armstrong & Rico-Hesse, 2001; Cologna et al., 2005; Anderson & Rico-Hesse, 2006).

Some of the recognized, contributing factors which adds in mosquitoes competence towards the disease transmission are expansion of the vector range, relaxation of vector control efforts and increasing resistivity towards insecticides by the mosquitoes (Gubler & Clark, 1995; Kawada et al., 2009). The effect of environmental agents like favourable temperature, sufficient rainfall and suitable humid conditions are also identified well to contribute to the transmission and spread of mosquito-borne diseases (Watts et al., 1987). The local environmental factors are considered to be equally responsible towards the development of complex disease such as dengue along with the generalized climatic conditions such as global warning (Kuno, 1995; Reiter, 2008; Johansson et al., 2009).

1.1.14 Economic importance of dengue

Dengue brings about disturbances economically to the affected persons in terms of hospital related costs as well as discontinuation from earning apart from physical pain. The collective ecumber of the infectious diseases such malaria, AIDS and TB is almost equal to the losses claimed by dengue in Southeast Asia in terms of DALYs (disability-adjusted life years) (Gubler &
Meltzer, 1999). In order to provide medical assistance to patients, to eradicate the vector and to launch public awareness programs, a number of funds have to be installed. Loss of revenue through reduced tourism is another indirect cost (WHO, 1997).

1.2 DENGUE, THE VIRUS

Dengue disease is caused by DENV, actually four flaviviruses, which are antigenically deferentiable but closely related to each other. According to Wang et al., 2000, the hypothesis about the evolution of these viruses is that from 100-1,500 years back, these viruses evolved from the sylvatic lineage (ancestral) independently. These four viruses are know as serotypes and are denoted as DENV-1, DENV-2, DENV-3 and DENV-4.

1.2.1 Taxonomic classification

The Flaviviridae family (formerly known to be group B arboviruses) is comprised of three genera which are Flavivirus, Pestivirus and Hepacivirus. The dengue virus belongs to the genus Flavivirus that contains 55 recognized species of viruses (Büchen-Osmond, 1996). As far as the term Flavi is concerned, it is believed that the word has been taken from flavus which is a Latin word means ‘yellow’ and that’s why that the yellow fever virus (YFV) is a member of this genus. Jaundice is observed in yellow fever patients and that is the reason that these viruses termed flaviviruses (Burke & Monath, 2001; Gubler, 2002).

Flaviviruses mostly are the major human pathogens such as dengue viruses. Other important viruses included in the group are yellow fever virus, West...
Nile virus, tick-borne encephalitis virus and Japanese encephalitis virus. Predominantly mosquitoes as well as ticks are the major transmitting sources of the flaviviruses but still there are some viruses for which the vector is still not recognized (Cook et al., 2006; Cook et al., 2009).

Based on neutralization tests, according to the initial research done, the flaviviruses were grouped into eight antigenic complexes based on serological investigation and dengue was one of the groups classified. But many viruses like YFV were found to have no association with the formation of any of the above mentioned complexes (Calisher et al., 1989). The phylogenetic inference from molecular data as the sequence data became available confirmed the classification pattern based on antigenic complexes. Furthermore, this analysis clearly grouped the Flaviviruses into vector born and non vector borne virusis. The vector-borne viruses in turn divided into mosquito-borne and tick-borne viruses (Kuno et al., 1998). Moving on further, the group of viruses related to mosquito-born was further analyzed and it was found that this group contains dengue viruses, Japanese encephalitis virus and yellow fever virus as well, as indicated in figure 1.10 (Zanotto et al., 1996).

Based on antigenic properties, four groups, known as serotypes of dengue viruses were identified. This classification was reaffirmed later on from the molecular analysis of dengue that delivered a good comprehension of dengue serotypes regarding its phylogeny. According to the phylogenetic analysis, first of all it was DENV-4 which emerged from the common ancestor and then
DENV-2 split out. Later on DENV-1 and DENV-3 separated from the ancestral lineage (Zanotto et al., 1996).

Figure 1.10 ML tree of the E gene (123 flaviviruses) rooted on Aedes albopictus cell fusion agent (CFA) virus (Zanotto et al., 1996).

1.2.2 Morphology of the virion

Likewise other flaviviruses, dengue virion is 40-50 nm in diameter and has icosahedral symmetry. Virion is having a nucleocapsid of around 30 nm in diameter surrounded by a lipid envelope. As the name indicates the nucleocapsid is comprised of capsid and viral RNA. A lipid bilayer, an envelope protein (51 to 59 kD, serves for the attachment of the virus and fusion
to host cells as well as penetration) and an internal matrix protein (approximately 8.5 kD) which is non-glycosylated constitutes the lipid envelope.

In most of the flaviviruses the envelope protein in general is glycosylated as well as exposed on the surface of the virion. In mature form, dengue virions have comparatively smooth surfaces comprised of 180 copies of the envelop protein makes the icosahedral symmetry under electron microscopic examination as shown in figure 1.11 (Kuhn et al., 2002).

![Figure 1.11 Structures of the virion and E protein (Perera & Kuhn, 2008).](image)

**1.2.3. Dengue genomic organization**

Genomic organization of all flaviviruses, particularly of the dengue virus is comparatively simple than other arboviral families like *Togaviridae* (formerly called group A arboviruses), *Bunyaviridae* or *Rhabdoviridae* (Casals & Brown, 1954; Henchal et al., 1982).
Genome of the dengue virus is comprised of a single-stranded, positive sense RNA molecule about 10.7 kb in size. The genome has a single open reading frame which encodes a polypeptide chain of 3390 amino acids that later on catalytically split down and gives rise to ten (structural and non-structural) viral proteins as indicated in table 1.2. There has been no evidence of any alternative or overlapping reading frames. DENV has no hyper-variable region in its genome unlike that reported in the HCV genome (Pang et al., 2001a; Pang et al., 2001b; Alvarez et al., 2006a).

Table 1.2 Lengths (in amino acids) of DENV structural and non-structural proteins. C: Capsid, E: envelope, PrM: Precursor of membrane, M: Membrane, NS: Nonstructural, CDS: Coding sequence

<table>
<thead>
<tr>
<th>Proteins</th>
<th>DENV-1</th>
<th>DENV-2</th>
<th>DENV-3</th>
<th>DENV-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>114</td>
<td>114</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>prM/M</td>
<td>166</td>
<td>166</td>
<td>166</td>
<td>166</td>
</tr>
<tr>
<td>E</td>
<td>495</td>
<td>495</td>
<td>493</td>
<td>495</td>
</tr>
<tr>
<td>NS1</td>
<td>352</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>NS2A</td>
<td>218</td>
<td>218</td>
<td>218</td>
<td>218</td>
</tr>
<tr>
<td>NS2B</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>NS3</td>
<td>619</td>
<td>618</td>
<td>619</td>
<td>618</td>
</tr>
<tr>
<td>NS4A</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>NS4B</td>
<td>249</td>
<td>248</td>
<td>248</td>
<td>245</td>
</tr>
<tr>
<td>NS5</td>
<td>899</td>
<td>900</td>
<td>900</td>
<td>900</td>
</tr>
<tr>
<td>Length of CDS</td>
<td>3392</td>
<td>3391</td>
<td>3390</td>
<td>3387</td>
</tr>
</tbody>
</table>

ORF of the DENV is flanked by untranslated regions (UTRs) both at 5’ (approximately 95-135 nucleotides) and 3’ ends (114-650 nucleotides) of the
genome (Rice et al., 1985; Chambers et al., 1990). There is a type I cap (m7GpppAmp) at the 5’ terminus of the genome while at the 3’ terminus polyadenylation has not been observed (Chambers et al., 1990). Viral and host proteases cleave the translated polyprotein into 10 viral proteins, comprised of structural and non-structural protein. Capsid (C), precursor of membrane (prM/M) and envelop (E) are the three structural proteins towards the 5’ end of the genome. On the other hand, there are seven non-structural proteins pointing towards the 3’ end of the genome and are designated as NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 as shown in figure 1.12. For efficient translation and replication both the 5’ and 3’ UTRs are required (Khromykh et al., 2003; Edgil et al., 2006; Romero et al., 2006).

1.2.4 The structural proteins C, prM and E

The three structural proteins that is C, prM/M and E constitute the DENV virion, the capsid protein form the nucleocapsid by surrounding the viral RNA genome while the E proteins along with prM are enrooted in the lipid bilayer which constitutes the viral envelope.

DENV replication and translation are dependent on the intracellular membrane structures. The C, prM and E proteins on their C-terminal ends bear hydrophobic amino acids that act as signal sequences for the remaining protein to get entry into endoplasmic reticulum (Salonen et al., 2005). NS2B-NS3 viral protease and ER signal peptidase collectively process the structural and NS1 proteins to individual membrane-bound proteins (Lobigs, 1993;
Yamshchikov & Compans, 1993; Pethel et al., 1992). Expression of the non-structural proteins occurs only in those cells which are infected with the virus. In replication of the viral genome, there is no involvement of the structural proteins (Pang et al., 2001a; Pang et al., 2001b; Alvarez et al., 2006a).

Furin cleaves prM to its mature form that is M protein amid of viral release and for the production of mature infectious virions this process is a prerequisite (Stadler et al., 1997). The common agreement is that the E proteins are protected by prM during the release of the virus from premature attachment and pH induced reorganization. Thus, the maturation process is vital for the infectivity to be ensured (Guirakhoo et al., 1991; Zhang et al., 2003).

The general agreement about viral E glycoprotein is that it plays vital role in host cell receptor binding, entry of the virus and also is the prime target for humoral immunity. There are three domains of the E protein which are domains I, II and III. The fusion peptide are hold by domains I as well as II on their distal ends while domain III serves for the attachment of the virus to the receptor (Figure 1.12). The M glycoproteins also possess a small number of some of these epitopes (Figure 1.13) (Kaufman et al., 1989). The homodimeric conformation of the E proteins makes the fusion peptides inaccessible in the mature state. The trimerization, induced by Low pH, unveils the hydrophobic fusion peptides in a fashion congruent to the class II fusion protein mediated membrane fusion (Modis et al., 2004). Mutations in the ligand pocket-forming
residues at the coordinate of domain I and II, changes the mandatory pH value, thus influencing the virulence (Rey et al., 1995).

Figure 1.12 (A). Domains of dengue E protein. Red is domain I, domain II is yellow, domain III is blue. (B). This is the conformation of an E dimer in the mature virus particle and in solution above the fusion pH. (C). Packing of E on the surface of the virus (Modis et al. 2004).

The two important and potential asparagine (N)-linked glycosylation sites in the E proteins are at positions Asn-67 and Asn-153. Asn-67 is unique for DENVs whereas Asn-153 is conserved in most flaviviruses (Heinz & Allison, 2003). Based on DENV serotype, in various strains even and in those infected cells where the virus replicates, the glycosylation pattern differs. The antigenic properties of DENV are affected by the degree and position of N-linked glycans (Johnson et al., 1994; Lee et al., 1997; Navarro-Sanchez et al., 2003).
Figure 1.13 (top) Genetic organizations, (bottom) configuration of the membrane and cleavage sites. Arrows indicate viral and cellular proteases (Perera & Kuhn, 2008).

1.2.5 The non-Structural proteins

The seven non-structural (NS) proteins of the DENV genome are NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 at its 3’ end. Some of the non structural proteins play multiple functions while NS1, NS2A, NS4A and NS4B are not well characterized regarding their role (Perera & Kuhn, 2008). A generalized sketch regarding the functional aspects of the non-structural proteins are summarized in table 1.3.
Table 1.3 NS proteins and their role

<table>
<thead>
<tr>
<th>NS proteins</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS1</td>
<td>Important function in RNA replication complex of the virus as well as soluble complement-fixing antigen</td>
</tr>
<tr>
<td>NS2A</td>
<td>Component of replication complex</td>
</tr>
<tr>
<td>NS2B</td>
<td>Act as co-factor for NS3 protease</td>
</tr>
<tr>
<td>NS3</td>
<td>Serine protease &amp; RNA helicase as well beside RTPase and NTPase activities</td>
</tr>
<tr>
<td>NS4A</td>
<td>May inducts changes in the membrane; vital for the viral replication</td>
</tr>
<tr>
<td>NS4B</td>
<td>Possibly controls IFN α and β-induced signal transduction</td>
</tr>
<tr>
<td>NS5</td>
<td>Methyltransferase (MTase) as well as RNA-dependent RNA polymerase</td>
</tr>
</tbody>
</table>

1.2.6 Genetic Diversity of DENV

Genetic diversity to a greater extent has been observed in dengue viruses on the basis of available molecular data. The contributing factors for the genetic diversity are numerous (figure 1.14) and that’s why the epidemiological complications related to the diversity are also diverge in nature (Holmes and Burch, 2000).

![Figure 1.14](image)

Figure 1.14 Factors responsible for genetic diversity (Holmes & Burch, 2000).
All the four serotypes of DENV are regarded as one virus in general for most purposes and hence the resulting diseases are regarded as one disease with a similar outcome. Among the four serotypes however, the genetic distances are even much higher compared to the diversity among other known virus species, belong to the genus (Figure 1.15) (Kuno et al., 1998; Holmes & Burch, 2000).

Based on sequence diversity, each serotype of DENV can further be classified into many genetic groups called genotypes/subtype. Initially Rico-Hesse in 1990 on the basis of sequence diversity explained that a DENV genotype is a group of viruses which show not more than 6% diversity in sequences using E/NS1 junction (240 nucleotides) of DENV-1 and DENV-2. Starting from this initial work, different regions of the dengue virus genomes were analyzed through sequencing. The sequencing strategy varied greatly depending on the research group they were interested in, from complete genome sequencing of a single gene to the whole genome sequencing covering the entire genome of the virus.

Currently genotype specification is based on phylogenetic analysis instead of applying the the percent divergence values.
Figure 1.15 E gene-based minimum evolution tree using 554 flaviviruses (Unrooted).

For the four dengue serotypes, a detailed and excellent explanation for the genotypic classification has now been outlined using molecular techniques (Rico-Hesse, 2003; Vasilakis & Weaver, 2008).

Based on the complete E gene sequence, DENV-1 can be divided into five genotypes (Goncalvez et al., 2002). Initially DENV-1 was also classified using 240 nucleotides of the E and NS1 gene junction by Rico-Hesse in 1990 into five
groups but having little differences as compared to the new scheme introduced as in table 1.4.

Table 1.4 Genotypes of DENV-1

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Initially defined distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Found in Japan, in 1940s in Hawaii (prototype strains) as well as China, Southeast Asia and Taiwan</td>
</tr>
<tr>
<td>II</td>
<td>Found in Thailand (1950 - 1960s)</td>
</tr>
<tr>
<td>III</td>
<td>Found in Malaysia as Sylvatic strain</td>
</tr>
<tr>
<td>IV</td>
<td>Distributed in Australia, Nauru, Indonesia and the Philippines as well</td>
</tr>
<tr>
<td>V</td>
<td>Distributed in Southeast Asia, Africa and the Americas</td>
</tr>
</tbody>
</table>

DENV-1 genotypes in general show a widespread distribution pattern and can be found in many regions of the world except from genotypes III, a sylvatic strain and II, restricted to Thai in the era corresponds to 1950s and 1960s.

In the Pacific between 2000 and 2004, the major causative agents of epidemics recently, are determined to be genotypes I and IV (A-Nueoonpipat et al., 2004). During the dengue outbreaks in 2003 in the Americas, genotype V has been detected and isolated many times (Aviles et al., 2003). But still it is not concluded that out of the three genotypes of DENV-1 which one was resopnsible for causing the sever form of the disease (Rico-Hesse, 2003).

The most studied serotype among the dengue viruses is the DENV-2. On the basis of the E gene complete sequencing after the initial research of Rico-
Hesse in 1990 and that of Lewis et al., 1993, the existence of six genotypes of DENV-2 has been proposed by Twiddy et al., in 2002 as indicated in table 1.5. Closely related sylvatic DENV-2 strains from Malaysia and many countries in West Africa have been detected and characterized. Both these two regions are located far away from each other which hypothesized that in the Asian-Oceanic region, the DENV sylvatic ancestor arose prior to diverge into today’s four DENV serotypes (Wang et al., 2000).

The Asian II genotype emerged for the first time in 1981 in Cuba, resulted in the first DHF epidemic in the Americas (Guzman et al., 1995). Similarly, the already native and prevailing genotype V, now known as American genotype has been relieved by the genotype III, now known as American/Asin genotype in the Western Hemisphere with the highest epidemiological impact (Rico-Hesse et al., 1997; Rico-Hesse, 2003).
Table 1.5 Genotypes of DENV-2 according to Twiddy et al., 2002

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Initially recognized prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>Initially called subtype V. Prevalent in Latin America and old strains restricted to India (1957). From 1950 to 1970s, prevalent in Caribbean and the Pacific islands</td>
</tr>
<tr>
<td>American/Asian</td>
<td>Initially recognized as subtype III. Prevalent in Vietnam, China and Thailand. Since 1980s prevalent in Latin America</td>
</tr>
<tr>
<td>Asian I</td>
<td>Distributed in Thailand, Malaysia &amp; Myanmar</td>
</tr>
<tr>
<td>Asian II</td>
<td>Initially recognized as subtypes I and II. Prevalent in China, Philippines, Sri Lanka, Taiwan as well as Vietnam. New Guinean C prototype strain is also included in the group</td>
</tr>
<tr>
<td>Cosmopolitan</td>
<td>Initially recognized as genotype IV. This is widely distributed in many regions like the Pacific islands, Australia, Indian Ocean islands, Southeast Asia, the Indian subcontinent, Middle East, East Africa and West Africa as well</td>
</tr>
<tr>
<td>Sylvatic</td>
<td>Pertaining to Malaysia and West Africa from non-human primates</td>
</tr>
</tbody>
</table>

Lanciotti et al., in 1994, proposed the current genotype classification for DENV-3 based on prM/E sequences which recognized four genotypes (Table 1.6). This study is similar to the work conducted by Chungue et al., in 1993 which recognized four groups for DENV-3 based on the sequence analysis of the E gene towards the 5’ end of the genome covering 195 nucleotides.

In 1994, genotype III of DENV-3 introduced to the Americas through Nicaragua and is currently prevalent in Central America as well as Southern America to a larger extent (Balmaseda et al., 1999; Usuku et al., 2001; Messer et al., 2003). This genotype is regarded as the most common and the most virulent out of the four genotypes of DENV-3.
Interestingly, according to Lanciotti et al., 1994, DHF condition has not been reported to have any link with the genotype IV. However, based on the detection of antibodies against DENV-3 in non-human primates (canopy-dwelling), their existence is anticipated and so for the sylvatic strain of DENV-3 has not been identified (Rudnick, 1984).

Table 1.6 Genotypes of DENV-3 as described by Lanciotti et al., 1994

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Original identified prevalence (prior to 1993)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Distributed among Thailand, Vietnam, Indonesia, Malaysia, Burma, the Philippines &amp; the South Pacific islands (French Polynesia, Fiji and New Caledonia). H87 prototype strain is also included</td>
</tr>
<tr>
<td>II</td>
<td>Prevalent in Vietnam, Thailand and Bangladesh</td>
</tr>
<tr>
<td>III</td>
<td>Circulates in South Pacific islands, Singapore, Indonesia, Sri Lanka, India, Africa &amp; Samoa</td>
</tr>
<tr>
<td>IV</td>
<td>Common in Puerto Rico &amp; French Polynesia (Tahiti)</td>
</tr>
</tbody>
</table>

DENV-4 was classified into genotypes, I and II by Lanciotti et al., in 1997 on the basis of the complete E gene sequences. Subsequently, two other genotypes recognized, one in Malaysia in non-human primates and the other one in Bangkok, Thailand, the genotype III (Klungthong et al., 2004) as shown in table 1.7.

The most widespread of the four genotypes is the genotype II of DENV-4 that in 1981 the Westren hemisphere witnessed by contracting it possibly through Pacific islands (Lanciotti et al., 1997; Fosteret al., 2003). Though, out of the four serotypes, DENV-4 is not much identified but still during secondary
infection, hemorrhagic manifestations are correlated with it according to the study conducted by Vaughn in 2000.

Table 1.7 Genotypes of DENV-4

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Initial recognized prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Common in Malaysia, Philippines, Thailand, &amp; Sri Lanka H241 prototype strain is also included</td>
</tr>
<tr>
<td>II</td>
<td>Prevalence in Tahiti, the Caribbean islands (Puerto Rico and Dominica), Indonesia, Malaysia &amp; the Americas</td>
</tr>
<tr>
<td>III</td>
<td>Prevalent in Thailand particularly in Bangkok</td>
</tr>
<tr>
<td>Sylvatic</td>
<td>Detected in Malaysia in non-human primates</td>
</tr>
</tbody>
</table>

1.2.7 Pakistan, in dengue perspective

Pakistan, a country important for its geographical and climatic diversity, is located in the north-west of South Asia and is endemic to all four serotypes of DENV. During the post monsoon period between October and December, the infection reaches to its peak level (Jahan, 2011; Khan et al., 2010). Pakistan is stretched between latitudes 23.45° and 36.75° north and longitudes 61° and 75.5° east. The provinces of Punjab, Sindh, Khyber Pakhtunkhwa, Balouchistan, Gilgit-Baltistan as well as Azad Jammu & Kashmir constitute the territories of the country.

As the history reveals, Pakistan since its creation, is facing with various vector-born diseases, notably the malaria, dengue, Crimean-Congo, Japanese encephalitis, leishmaniasis, scrub typhus, West Nile and others as well. The reasons behind the presence of such a wide range of diseases in Pakistan are;
the country is located in subtropics and the local habitat is suitable for the vectors to live in as well breed (Sugamata et al., 1987). In Pakistan, for more than half a century vector borne diseases such as malaria, have been recognized while dengue has recently established its roots in the country. No outbreaks of dengue disease reported until 1994, possibly without proper surveillance and diagnostic measures. However, considerable proportions of the apparently healthy population in Rawalpindi and Peshawar (Burney, 1966; Burney & Munir, 1966), parts of Punjab (Hayes et al., 1982; Hayes & Burney, 1981) and Karachi (Sugamata et al., 1986) turned positive for the neutralization and haemagglutination inhibition antibodies for DENV in sero-epidemiological studies directed in the 1960s and 1980s. In another study, directed towards West Nile, Japanese encephalitis and dengue diseases in order to determine their occurrence in the period from 1983 to 1985 in the city of Karachi, revealed that 65% people were infected with these diseases collectively, using haemagglutination inhibition (HI) assay, affecting people aged 6-65 years. Individuals aged between 6 and 20 years revealed an increase in the HI sero-antibody positivity rate for DENV-2 between July and October (Sugamata et al., 1987).

The first ever epidemic of dengue disease in Pakistan erupted in August 1994 and continued through November 1994 in Karachi, that affected thousands of people. No proper epidemiological data for this epidemic was recorded however thousands of people contracted the infection as examined by physicians. The mortalities associated with this outbreak are not well
estimated but at least two deaths occurred in Karachi (Chan et al., 1995). To investigate this epidemic, studies were directed on adults as well as children and DENV-1 and DENV-2 were indentified by implying ELISA for the detection of IgM antibodies (Chan et al., 1995; Akram et al., 1998; Ansari et al., 2001; Qureshi et al., 1997).

In 1995, DENV was circulating in area known as Hub, far away (60 km) from Karachi. Many workers, who were constructing a powerhouse, experienced pyrexia reported by the contractor (Anonymous, 2003). Upon examination, DENV serotypes 1 and 2 were confirmed by the detection of IgM and IgG antibodies in their blood using ELISA. Beside the local people, the expatriates engaged in the company were also infected that indicates that the disease was spreading locally. In general, the workers were making complains about the bites by mosquitoes during the daytime, though entomological studies were not undertaken. The disease, after destyoning the proposed breeding sites for mosquitoes around the power house, did not emerge in the area again (Paul et al., 1998; Humayoun et al., 2010).

During the year 2005, about 10 years later, DENV-3 was isolated in the country. According to the National Institute of Health Islamabad, in their laboratory in 2005, around 395 cases of dengue from Karachi were diagnosed. The turning point in the epidemiology of the disease was witnessed in 2006 when the virus expanded its roots to the Southern parts of Pakistan for the first time. With the introduction of the disease to the South of Pakistan, high epidemics were observed that caused huge morbiditie and mortalities. In
2006, the number of people suspected for the disease throughout Pakistan was around 5800 that ended with 3000 confirmed cases, out of which 52 people died (Tang et al., 2008). Soon after the monsoon with plenty of rain, from August to October a large number of people were infected with the disease. Again the city of Karachi turned out to be greatly affected. Around 4500 cases were registered, out of which 1500 were confirmed by antibodies against DENV in their blood and 50 individuals were died during this outbreak (Khan et al., 2007).

During the year 2006, dengue infection continued in the northern Pakistans as well. In Punjab, around 800 cases were registered mainly from Lahore and Rawalpindi which were greatly affected, almost 400 cases in each of the twin cities. From north-western province of KP, the number of people infected was 31 and claimed one death. 480 cases with one death were reported from Islamabad, the capital of Pakistan. This was the first outbreak in Pakistan that spread over a larger area and infected a large number of people as well.

In 2007, dengue outbreak erupted in the country again, covered both the South and North but the intensity of the infection was not as severe as it was in 2006. In this outbreak 3342 suspected cases were registered across the country, out of which 1208 patients were confirmed for dengue using ELISA and the death toll it resulted was 22.

From November to October, the frequency of infection was the highest though dengue cases were reported throughout the year. 2900 cases with 950 confirmed and 22 deaths were reported from Karachi, the most affected city
while 258 cases were detected in Lahore. From the interior of the Punjab, cases were also reported, seemingly the contracted cases from other areas because the rate was just one or two per each city.

In the north of Punjab in the city of Attock, the infection was reported but circulated only in Basal and Thatta Khalil villages, affected almost 80% of the population, complaining about symptoms similar to dengue in the months of August and September according to hospital records (Rasheed et al., 2013).

Another dengue epidemic erupted in 2008 too. This time, 3280 people were infected with the disease that led to 30 mortalities across Pakistan. This was the first time that the prevalence of the disease was higher in Northern than Southern Pakistan. In Punjab it infected 1450 people, out of which 20 died.

In this outbreak, ten patients had DENV-4, five DENV-2 and two DENV-3 in their blood out of 17 in Lahore, analyzed through real time PCR (Humayoun et al., 2010).

The country was devastated in 2010 by the worst flood in terms of loss of lives and properties as well. Inspite of this, the flood also contributed in the spreading of dengue disease by providing breeding sites for the mosquitoes. As as result, according to the reports of various health officials, the infections caused by dengue were the highest in 2010 by infecting 9000 cases in total across the country. In the province of Sindh, 5000 people were infected that resulted in 35 deaths, out of which 16 deaths were from Karachi.
In Punjab the number of people infected was 4000 and death toll recorded was 3.

The infection continued till November (1400 cases) and in the month of October, the number of the infected people was 1500. Karachi and Lahore once again found to be highly affected.

Dengue continued to affect the people of Pakistan in 2011 as well. This time the infection recorded was the highest in the history of Pakistan, even more than 2010. Provinces of Punjab and Sindh were mainly affected along with Khber Pakhtunkhwa. 22,562 individuals in total were infected with the disease that claimed 363 lives across Pakistan in 2011, according to the reports of health departments. According to independent surveys, the number of infected people was more than 35000 and number of deaths it caused was more than 420.

The outbreak erupted in the month of August and prevailed till November, with Punjab as the most hit Province by affecting more than 21,300 individuals out of which 337 people died. Though the entire Punjab was affected but Lahore was the most hit city with 17,493 people received the infection, out of which 290 people died. Rawalpindi, Faisalabad, Chakwal, Dera Ghazi Khan as well as Chichawatni were included in the list, hit by dengue virus and caused morbidities and mortalities.

Mainly Lahore was infected in 2011 however patients were also reported from KP as well as from Sindh. In Sindh a total of 952 cases diagnosed for dengue, leaved behind 18 deaths. Out of the 18 people died, 15 were from Karachi. The
major circulating serotypes detected in 2011 outbreak, were DENV-2 and DENV-3 along with less prevalence of DENV-4 as well (Koo et al., 2013).

In 2013, another epidemic of dengue infection hit Pakistan, notably district Swat and adjoining areas of Khyber Pakhtunkhwa for the first time. 8,343 cases were registered from the Swat, Shangla and Bisham districts with official and unofficial death tolls of 36 (District Health Office of Swat) (Anonymous, 2013) and 57 respectively reported up to the 18th October 2013. A detailed study directed to the outbreak in Swat, screened 414 clinically diagnosed dengue patients for dengue specific antibodies, out of which 366 were found to be serologically positive by the Standard Diagnostics (SD) Dengue Duo test. All dengue virus (DENV) positive samples were then analyzed using a DENV serotype-specific reverse transcriptase polymerase chain reaction to identify the DENV serotypes present and to figure out active infection. In this study, 28.69% (105/366) of the serologically positive patients had active infection and DENV type 3 (DENV-3) was the most abundant serotype followed by DENV-2; similar to the outbreak in Punjab during 2010-11 (Ali et al., 2013).

History and geographical origin of DENV lineage can often be determined through genotypic classification and characterization. By genetic characterization, tracking down the transmissibility of the virus as well nature of the infection becomes easy to understand. Studies relating to molecular epidemiology of dengue infection are based on genetic characterization, by the help of which it is possible to outline that whether an endemic strain is
involved in the epidemic or new strains have been introduced into a particular area. In case if new strains of viruses been introduced, then it is of particular importance to know that which genotype of DENV may possibly be involved in developing the severe form of the disease. Currently, Southeast Asian genotypes of DENV have been known to cause the severe disease (Guzman et al., 1995; Rico-Hesse et al., 1997). The correlation between the re-emerging endemic strains of dengue and the epidemics they cause needs further investigation and explanation. This can be answered by implying the science of comparative genomics along with classic epidemiology.

To address questions like whether the re-introduction of the virus is because of the environmental factors, immune responses of the population and vectorial elements or due to adaptation and evolution of the viral strains (which increase fitness as well as virulence), complete genetic characterization is required, without which it is impossible to address these questions.

So far in Pakistan, very only few, comprehensive molecular epidemiological studies have been conducted, related to the prevalence of DENV in the country. Studies encompassing the molecular epidemiology and evolutionary genetics are essential to determine the origin and spread of viruses, to enhance our understanding on the pathogenesis of the disease, the etiology of epidemics as well as the genetic basis of virulence. RNA viruses like DENV evolves rapidly (Steinhauer & Holland, 1987) and on rare occasions, some mutations lead to phenotypic changes in the viruses which alter their potential to cause epidemics associated with severe disease (Grenfell et al.,
Genotypic characterization is a useful tool in the determination of the evolutionary history of the DENV, identification of the circulating virus serotypes/strains in the endemic area and for the detection of new invading genotypes. Therefore, in the current study, a detailed molecular characterization of Pakistani DENV isolates was investigated from the 2011 and 2013 outbreaks in Pakistan in Punjab (Lahore) and KPK (Swat) respectfully with the following objective;

- Genotyping of the prevalent dengue virus in Pakistan
- Full length characterization of the representative dengue virus genotypes by sequencing
- Phylogenetic analysis of prevalent dengue viral genotypes
Chapter 2
MATERIALS AND METHODS

This thesis is a detailed, prospective study about dengue virus infection in Pakistan that covers the two major dengue outbreaks; the 2011 outbreak in Punjab and the 2013 in district Swat, KP.

2.1 SAMPLING

2.1.1 2011 outbreak

Already collected serum samples from the 2011 dengue outbreak in Punjab, stored in IBGE, Peshawar and Kohat University of Science and Technology were analyzed in this study. The samples were collected during the acute phase of the infection form hospitalized patients mainly from Lahore, Sargodha, Sheikhupura, Kasur and Rahim Yar Khan. 600 serum samples which had been collected from Enzyme Linked-immunosorbent assay positive patients were selected for further analysis. Out of the 600 individuals, 461 patients belonged to five districts in Punjab province as indicated above and 139 patients belonged to six districts of KP province (Peshawar, Mardan, Nowshera, Kohat, Charsadda and Swat), admitted at Lahore General Hospital. Patients that belonged to KP had been infected in Punjab province and were hospitalized in Lahore. Distribution of the samples and their detail is provided in table 3.1.
2.1.2 2013 outbreak

Samples from the recent dengue outbreak of 2013 in Swat, KP were collected in the acute phase of the disease mainly from Saidu Teaching hospital (STH). Informed written consent in the form of questionnaire was taken from each patient. Sample of the questionnaire is attached as (Annex-I). After the initial information, 5 ml of blood was collected in separate sterile syringes and labeled accordingly. The blood was carried to the Institute of Biotechnology and Genetic Engineering (IBGE), The University of Agriculture Peshawar for analysis. Blood was centrifuged for 3 minutes at 10000 rpm to separate serum. After separation, the sera were screened for dengue virus specific antibodies (IgG and IgM) and the DENV NS1 glycoprotein using SD Dengue Duo strips (Standard Diagnostics, Korea). The positive samples thus obtained were stored at -70° C in Eppendorf tubes and were labeled properly. Initially at the start of the outbreak in July, 2013, we collected 740 blood samples from patients who experienced dengue like symptoms and were immediately processed for molecular analysis (Table 3.1). Later on as the outbreak was progressing and new cases were admitting to the hospitals in Swat, we collected 414 blood samples and carried out the analysis.

2.2 PRIMER DESIGNING

2.2.1 Diagnostic primers

Two kinds of primer sets were used as diagnostic specific dengue primers, that is one already described by Lanciotti et al, 1992 and another set we designed specifically for the 3’ end confirmation of the virus genome by
aligning multiple sequences, retrieved from Genbank, using ClustalW software (Thompson et al., 1994). This set of primers for the 3’ end is unique in its kind that is conserved for all the four serotypes of dengue virus. Approximately 130 full length sequences were retrieved from the GenBank for the synthesis of the primer set which spans about 340 bp towards the 3 end of the virus genome (Annex-II). The most conserved regions were selected to design the primers. Primers were tested for specificity and BLAST was done using NCBI.

2.2.2 Primers for individual genes and complete genome amplification

Specific primers for the individual genes amplification like Capsid (C), Envelope (E) and Non-structural gene 3 (NS3) and for whole genome amplification were designed. Over 130 whole genome sequences were retrieved from the GenBank. Multiple sequence alignment was done using clustalW as discussed above. The conserved regions were selected for primer designing using primer3 program (Untergrasser et al., 2012). For non-conserved positions in the primers, the major base (the nucleotide most commonly found in the non-conserved position) was selected for primer synthesis. The other important criterion we employed for synthesizing the primers was that the non-conserved position on a primer kept far from the 3 end of the primer, by at least three bases. Furthermore, it was ensured that one primer may not contain more than three non-conserved nucleotides. Nine sets of overlapping primers were designed for whole genome amplification of DENV2 and BLAST search was done to ensure the specificity of the primers.
using NCBI. NS3 gene of DENV3 was amplified in two overlapping fragments while capsid and envelope genes were amplified as single fragments. List of the primers for DENV2 and DENV3 genome amplification is attached as Annex-III.

2.3 RNA ISOLATION AND DETERMINATION OF THE ACTIVE INFECTION

IgG and/or IgM positive samples were then analyzed for active DENV infection and genotyping using the diagnostic primers as indicated above. DENV-RNA was extracted from the DENV sero-positive patient’s samples using the Favorgen Viral Nucleic Acid Extraction Kit (Favorgen® Biotech Corp) according to the manufacturer’s instructions. Prior to RNA extraction, all the materials like tubes, tips etc were treated were DEPC treated water and the working bench was bleached to ensure ribonucleases free environment. The extracted DENV-RNA was then reverse transcribed into cDNA using the SuperScript III reverse transcriptase (Invitrogen) as indicated previously (Ong et al., 2008; Schreiber et al., 2009). Active infection was determined using the specific priming primers we designed attached as Annex-II while genotyping was done as indicated by Lanciotti et al., 1992 with slight modifications in the thermal and chemical methods used, using nested PCR. Briefly, cDNA was amplified using 2X Taq FroggaMix, Canada, (0.25U/ul of Taq DNA polymerase, 2X PCR Buffer, 0.4mM dNTPs, 3.2mM MgCl2, 0.02% bromophenol blue). The reaction was carried out in 25ul PCR mix, contained on 1.0ul of cDNA and 10ul of the double distilled water. Detail of the PCR conditions is attached as Annex-IV. For genotyping a second round of PCR
amplification was carried out using type specific primers (TS1, TS2, TS3 and TS4) as described by Lanciotti et al, 1992 with slight modifications.

2.4 DENGUE VIRUS SEROTYPE 3 ANALYSIS

Specific sense and antisense primers were designed for the amplification of C, E and NS3 genes as described above. The C (342 bp) protein gene was amplified using D3CF and D3CR as the sense and antisense primers respectively. Similarly, the E (1479 bp) protein gene was amplified as single PCR products using D3EF and D3ER as sense and antisense primers respectively. The NS3 (1857 bp) protein gene was amplified in two overlapping fragments, fragment 1 as 933 bp in length using the specific primers D3ns3F1 and D3ns3R1 while fragment 2 as 987 bp product using the primers D3ns3F2 and D3ns3R2. PCR amplification was done using 2X Taq FroggaMix, Canada, as explained above. Primers list is attached as Annex III and PCR conditions as Annex IV.

2.5 RNA EXTRACTION AND cDNA SYNTHESIS FOR FULL LENGTH AMPLIFICATION

On the basis of the confirmation of active dengue infection and genotyping, primers for the complete genome amplification of DENV2 were designed as indicated earlier. Dengue viral RNA was extracted using the Favorgen Viral Nucleic Acid Extraction Kit (Favorgen® Biotech Corp) according to the manufacturer’s instructions. Extreme precautionary measures were taken in handling of RNA as stated above. Extracted RNA was quantified using nanodrop (NanoDrop ND-1000 Spectrophotometer) and RNA integrity was checked using RNA gel electrophoresis. Good RNA samples were subjected
to cDNA synthesis and further studies. The extracted RNA was reversed transcribed to cDNA by the SuperScript III First-Strand Synthesis System (InvitrogenTM, Carlsbad, CA, USA) using specific primer ‘D23Rv5’ (Annex-III). Briefly, 5ul of the extracted RNA was mixed with 2uM of reverse primer, 0.5mM of the dNTPs mix and 3ul of DEPC-treated water in 0.2ml RNAse-free microtube and incubated at 65°C for 5 minutes. Then quickly transferred the tube to ice and kept for 3 minutes. Next, other components of the RT reaction mixture (20mM Tris-HCl pH 8.4, 50mM KCl, 5mM MgCl2, 0.01M DTT, 40U of RNaseOUT™ and 200 U of SuperScript® III RT enzyme) mixed in a separate RNAse-free 0.2 mL microtube and added to the RNA mixture on ice, made the total reaction volume to 20ul. Then the mixture was incubated at 50°C for 50 min for synthesizing cDNA and then 85°C for 5min to terminate the reactions on an Applied Biosystems™ 96 Well Thermal Cycler. The cDNA product was then chilled on ice for 5min and 1.0 ul of RNase H was added to each tube and incubated at 37°C for 20 min to degrade viral and carrier RNA. The resulting cDNA products were stored at −80°C.

2.6 COMPLETE GENOME AMPLIFICATION

The cDNA thus obtained was used as template for complete genome sequencing. The whole genome of the DENV-2 was amplified (in the University of Toronto, Canada) and from Macrogen (Korea) in nine overlapping fragments, using nine sets of overlapping primers as described in figure 2.1.
The PCR protocol and conditions for the complete genome amplification is attached as Annex-IV.

2.7 SEQUENCING

Quantification of PCR products was carried out using NanoDrop (ND-1000 Spectrophotometer). The integrity of the purified DNA was checked on Agarose gel. After this, the product were sent to ‘The Centre for Applied Genomics (TCAG), Canada or Macrogen (Korea) for automated capillary sequencing using serotype-specific sequencing primers for DENV-2 and DENV-3 as described earlier. 3730xl DNA Analyzer (Applied Biosystems) was used for sequencing with Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems).

Consensus sequences of the viruses were obtained using Bio Edit version 7.1.9. Initially, the consensus sequences of our study were compared to the already recognized sequences of the same serotype through alignment.
To ensure the accuracy, editing of the consensus sequences was carried out manually too.

2.8 DETERMINATION OF THE OPEN READING FRAME (ORF) AND AMINO ACIDS PROFILE
The single open reading frame which encodes the polyprotein was determined by the analysis tool of the ‘National Centre for Biotechnology Information’ using http://www.ncbi.nlm.nih.gov/projects/gorf.

2.9 DETERMINATION OF THE CONSERVED DOMAIN
The conserved domains footprints and functional sites in the sequences were determined using the conserved domain database (CDD) of NCBI (Marchler-Bauer et al., 2011).

2.10 PHYLOGENETIC ANALYSIS
The full length genomes of DENV-2 isolated, sequenced and characterized in this study, confirmed with blast search against global isolates and 67 DENV-2 isolates were retrieved from NCBI and multiple sequence alignment was done using Clustal Omega (Goujon et al., 2010, McWilliam et al., 2013). Isolate D2/Pk/A1/2011, D2/Pk/A2/2011, D2/Pk/Swat-01/2013 and D2/Pk/Swat-02/2013 are the complete genome sequences of DENV-2 reported in this study and deposited to GenBank with the accession numbers of KM217157, KM217156, KM217158 and KJ701507 respectively.

The phylogenetic analysis of the complete genome of DENV-2 isolates was conducted using PhyML software version 3.0 (Guindon et al., 2010) using the most complex General Time Reversible (GTR) model of nucleotide substitution. Approximate Likelihood-Ratio Test (aLRT SH-like) with 4
substitution rate categories, estimated proportion of invariable sites, estimated Gamma distribution parameter and an initial tree calculated by the BIONJ algorithm were employed. Tree and domain composition diagrams were drawn using Archaeopteryx (https://sites.google.com/site/cmzmasek/home/software/forester).

Seven DENV-2 (Swat isolates) were randomly selected for the analysis of NS3 gene. Sequences of the seven NS3 gene of DENV-2 Swat isolates (Swat 2X, 2T, 2U, 2V, 2W, 2Y and 2Z) reported in this study, were submitted to GenBank with their accession numbers [KJ611228, KJ716888, KJ716889, KJ716890, KJ716891, KJ716892, and KJ716893] respectively. BLAST search was done using the NCBI database for the isolated NS3 genes and sequences were retrieved covering the target gene. Multiple sequence alignment was done using Clustal Omega (Goujon et al., 2010; McWilliam et al., 2013). Phylogenetic analysis for the total 83 NS3 gene sequences was done using MEGA version 6.0 (Tamura et al., 2013). Maximum likelihood tree with a heuristic search and sub tree-Pruning and Regrafting (SPR) branch-swapping was generated using Tamura-Nei (Tn93) with gamma (G) distribution and invariant (I) sites as best fit DNA substitution model (AICc=11561.804). The reliability of the inferred phylogenetic tree was estimated by the bootstrap method, with 500 replications.

The Core protein, Envelope and Non-structural 3 protein genes of DENV-3 were isolated and analyzed. The full genome sequencing was not done due to insufficient titer of the virus in the blood of the patients.
Complete E gene (1479 bp) of the five dengue 3 viruses isolated from the 2011 and 2013 outbreaks each, were sequenced successfully. The sequences were confirmed using BLAST search at NCBI database (http://blast.ncbi.nlm.nih.gov) and 47 global DENV-3 isolates were retrieved. Sequences of the five E genes of the isolates of the 2011 outbreak namely D3/Pk/E-51/2011, D3/Pk/E-52/2011, D3/Pk/E-53/2011, D3/Pk/E-54/2011 and D3/Pk/E-55/2011 reported in this study, were deposited to GenBank with their accession numbers of KM226344, KM226345, KM226346, KM226347 and KM226348 respectively. Similarly accession numbers of the five complete NS3 genes of the 2013 DENV-3 isolates (D3/Pk/Swat-01B/2013, D3/Pk/Swat-02B/2013, D3/Pk/Swat-03B/2013, D3/Pk/Swat-04B and D3/Pk/Swat-05B/2013) are KM217135, KM217134, KM217133, KM217132 and KM217159 respectively. Multiple sequence alignment was done using Clustal Omega (Goujon et al., 2010; McWilliam et al., 2013). Phylogenetic analysis for the total 57 NS3 gene sequences was done using MEGA version 6.0 (Tamura et al., 2013). Maximum likelihood tree with a heuristic search and sub tree-Pruning and Regrafting (SPR) branch-swapping was generated using Tamura-Nei (Tn93) with gamma (G) distribution and invariant (I) sites as best fit DNA substitution model (AICc=9518.602). The reliability of the inferred phylogenetic tree was estimated by the bootstrap method, with 500 replications.

Complete NS3 gene (1857 bp) of the five dengue 3 viruses isolated from the 2011 and 2013 outbreaks each, were sequenced successfully. The sequences
were confirmed using BLAST search at NCBI database (http://blast.ncbi.nlm.nih.gov) and 65 global DENV-3 isolates were retrieved. Ten NS3 gene sequences of the dengue 3 viruses (five each from 2011 and 2013) reported in this study, were deposited to GenBank. Accession numbers of the five NS3 gene of the DENV-3, 2011 isolates (D3/Pk/F-01/2011, D3/Pk/F-02/2011, D3/Pk/F-03/2011, D3/Pk/F-04/2011 and D3/Pk/F-05/2011) are KM217150, KM217149, KM217148, KM217147 and KM217146 respectively. Similarly accession numbers of the NS3 genes of dengue 3 viruses of the 2013 outbreak (D3/Pk/Swat-01/2013, D3/Pk/Swat-02/2013, D3/Pk/Swat-04/2013 and D3/Pk/Swat-05/2013) are KM217155, KM217154, KM217153, KM217152 and KM217151 respectively. Multiple sequence alignment was done using Clustal Omega (Goujon et al., 2010; McWilliam et al., 2013). Phylogenetic analysis for the total 65 NS3 gene sequences was done using MEGA version 6.0 (Tamura et al., 2013) with DENV-2 isolate as outgroup for rooting. Maximum likelihood tree with a heuristic search and sub tree-Pruning and Regrafting (SPR) branch-swapping was generated using General Time Reversible model (GTR) with gamma (G) distribution and invariant (I) sites as best fit DNA substitution model (AICc=12655.935). The reliability of the inferred phylogenetic tree was estimated by the bootstrap method, with 500 replications.

2.11 NUCLEOTIDES AND AMINO ACIDS SUBSTITUTIONS

Nucleotides and amino acids substitutions and conserved sites were determined using MEGA version 6 (Tamura et al., 2013). The core protein
gene (342 bp) of DENV-3 was characterized in this study. A total of 10 C gene sequences (five from 2011 and five from 2013 outbreaks) were analyzed. Accession numbers of the five dengue 3 viruses isolated from the 2011 dengue outbreak in Punjab (D3/Pk/F-10/2011, D3/Pk/F-11/2011, D3/Pk/F-12/2011, D3/Pk/F-13/2011 and D3/Pk/F-14/2011) analyzed for the C gene are KM217140, KM217139, KM217138, KM217137 and KM217136 respectively.

2.12 NUCLEOTIDE AND AMINO ACIDS IDENTITIES DETERMINATION

Based on the evolutionary analysis of the DENV-2 in this study, complete genome nucleotides and amino acids identities of the 17 cosmopolitan genotypes were determined by Clustal Omega (Goujon et al., 2010; McWilliam et al., 2013). Prior to the analysis the cosmopolitan genotypes group of DENV-2 were aligned using Clustal Omega.

2.13 SELECTIONPRESSURE ANALYSIS ON PAKISTANI ISOLATES

To check whether any codon or ORF of the Pakistani isolates are under selection pressure, the Hyphy package was used, implemented in the online version of Datamonkey facility (http://www.datamonkey.org) (Pond et al., 2005; Delport et al., 2010; Pond & Frost, 2005a).

To check the overall selection pressure on the entire coding sequence and on individual codons, the methods in datamonkey software were used like the single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL) by identifying the ratio of the non-synonymous nucleotide substitutions (dN) the synonymous nucleotide
substitutions per site (dS) indicated as dN/dS for the whole ORF (Pond & Frost, 2005b).

2.14 RECOMBINATION ANALYSIS

To check whether any recombination event has taken place on any gene of the dengue viruses, Genetic Algorithm for Recombination Detection (GARD) method within the HyPhy package was employed using Datamonkey software (Pond et al., 2006).
Chapter 3
RESULTS

This thesis is a detailed and thorough analysis of the dengue virus infection in Pakistan. Dengue virus hit the country twice within a span of 3 years that is first in 2011 in Punjab province and second in 2013 in district Swat, KP. We conducted a comprehensive analysis into the molecular epidemiology of the aforementioned dengue outbreaks in Pakistan.

3.1 SEROTYPE SPECIFIC PCR

3.1.1 2011 outbreak

461 ELISA positive serum samples from patients belonged to Punjab and 139 from Punjab-based KP patients were included in the study. PCR assay was done of all the ELISA positive samples to investigate the active infection and genotyping. Serotype specific PCR assay detected all the four DENV serotypes from blood samples when analyzed on 2% Agrose gel electrophoresis (Figure 3.1). The presence or absence of specific PCR bands for all serotypes in every sample was recorded based on gel electrophoresis. All four DENV serotypes were present among the samples collected from the Punjab province during 2011 (Table 3.1).

In Punjab province, 74% of the samples were positive by RT-PCR and indicated to have active dengue infection that comprised 56% males and 44% female patients. Majority of the actively infected people were from Lahore (83%) followed by other districts like Sargodha (9%), Sheikhupura (5%), Kasur (4%) and Rahim Yar Khan (0.5%) (Table 3.1). The most dominant DENV
serotypes were DENV2 (42%) and DENV3 (41%) and followed by DENV4 (9%) and DENV1 (4%) in Punjab province.

![Gel electrophoresis of the RT-PCR of the second round amplification with type specific primers for genotyping.](image)

**Figure 3.1** Gel electrophoresis of the RT-PCR of the second round amplification with type specific primers for genotyping. M is the 100bp DNA ladder (GeneRuler). Lane 1 shows specific band of 482 bp and is DENV serotype 1. Similarly, lane 2 (119 bp) represents DENV serotype 2, lane 3 (290 bp) shows DENV serotype 3 and lane 4 (392 bp) shows DENV serotype 4.

Similarly, 43% patients of the total 139 tested that belonged to KP province were confirmed by RT-PCR to have active dengue infection. More than 80% of these individuals were males while only 20% were females that had active dengue infection (Table 3.1). Majority of the patients that were positive by RT-PCR belonged to Peshawar (43%), followed by Mardan (18%), Nowshera (13%), Swat (13%), Charsadda (8%) and Kohat (3%) (Table 3.1).

In KP province DENV-2 serotypes had the highest incidence (41.66%) compared to DENV-3 (35%), DENV-1 (10%) and DEVN-4 (5.0%). Both DENV-
2 and DENV-3 were the most abundant serotypes in Punjab and KP provinces during the 2011 outbreak and had significantly high incidence than DENV-1 and 4 serotypes. A small number of patients were infected with DENV-1 and DENV-4 in both Punjab and KP provinces (Table 3.1). Mixed infection of DENV-2 and 3 serotypes was found only in two cities in Punjab and three cities in KP province. Tracking down the patients with mixed infections, no one was recovered from the illness and died, which further strengthens the role of ADE in causing mortalities. Individuals between the ages of 15 to 45 years were mostly infected while individuals with less than 15 years or more than 45 years old were the least affected (Table 3.1).
Table 3.1 Distribution of Dengue virus (DENV) serotypes during the 2011 and 2013 outbreaks in Pakistan

<table>
<thead>
<tr>
<th>Province/Year</th>
<th>Name of District</th>
<th>Samples collected</th>
<th>Distribution of PCR + samples in DENV Serotypes</th>
<th>M/F</th>
<th>Age Group [Years]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ELISA-positive samples</td>
<td>PCR positive samples</td>
<td>DENV 1</td>
<td>DENV 2</td>
</tr>
<tr>
<td>Punjab/2011</td>
<td>Lahore</td>
<td>322</td>
<td>282</td>
<td>11</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>Sargodha</td>
<td>58</td>
<td>28</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>S.Pura²</td>
<td>41</td>
<td>16</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Kasur</td>
<td>29</td>
<td>13</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>RYK³</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>461</strong></td>
<td><strong>341</strong></td>
<td><strong>14</strong></td>
<td><strong>142</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>73.96%</strong></td>
<td></td>
<td><strong>4.11%</strong></td>
<td><strong>41.64%</strong></td>
</tr>
<tr>
<td>KP/2011</td>
<td>Peshawar</td>
<td>63</td>
<td>26</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Mardan</td>
<td>25</td>
<td>11</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Nowshehra</td>
<td>19</td>
<td>8</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Kohat</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Charsadda</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Swat</td>
<td>16</td>
<td>8</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>139</strong></td>
<td><strong>60</strong></td>
<td><strong>6</strong></td>
<td><strong>25</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>43%</strong></td>
<td></td>
<td><strong>10.0%</strong></td>
<td><strong>41.66%</strong></td>
</tr>
<tr>
<td>KP/2013</td>
<td>Swat</td>
<td>618</td>
<td>200</td>
<td>0</td>
<td><strong>80</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>32.36%</strong></td>
<td></td>
<td><strong>40.0%</strong></td>
<td><strong>60.0%</strong></td>
</tr>
</tbody>
</table>
3.1.2 DENV 2013 outbreak

Dengue virus hit the Swat district of KP for the first time, caused high mortalities and morbidities. The epidemic started in July, 2013 and continued till mid October, 2013. Initially, 740 serum samples were collected from Saidu Teaching Hospital in Saidu Sharif Swat.

Analysis of the 740 blood samples showed that only 618 patients were positive for antibodies. The remaining 122 patients were having dengue like symptoms but on investigation they were not positive for the antibodies against dengue virus. However, active infection was confirmed by RT-PCR in only 200 patients of the 618 positive by serology. Both DENV-2 and DENV-3 were the most prevalent serotypes in district Swat (Table 3.1). Like the 2011 outbreak of Punjab, the people in between 15 and 45 years of age were the most infected (44.50%) in this outbreak.

As the dengue outbreak progressed in district Swat, more patients were included in the study and 414 blood samples were processed from clinically diagnosed patients.

Of the 414 serum samples, 88.40% were found to be anti-DENV IgG and/or IgM positive (Figure 3.2). A total of 28.68% patients were found to have active DENV infection using the reverse transcriptase polymerase chain reaction (RT-PCR). Of these 105 RT-PCR positive samples, the majority (65.71%) contained DENV-3 followed by DENV-2 (33.33%). One sample contained both DENV-2 and DENV-3 and died suggesting ADE phenomenon to have
occurred. While DENV-1 or DENV-4 was not identified in any of these samples (Figure 3.3).

Figure 3.2 Proportion of anti-DENV IgG and/or IgM positive and DENV RNA-positive patient’s serum samples

Figure 3.3 Distribution of the DENV genotypes in patients of 2013 outbreak
3.2 GENOME ANALYSIS OF DENGUE VIRUS

On the basis of the genotypic picture of dengue virus as indicated in our study, genome analysis of the prevalent genotypes that is DENV2 and DENV3 was determined.

3.2.1 Genome analysis of DENV-2

Complete genome characterization of DENV-2 was conducted and four whole genomes were successfully characterized. Of the four complete genomes, two (D2/Pk/A1/2011 and D2/Pk/A2/2011) were from the 2011 outbreak of Punjab and the remaining two viruses (D2/Pk/Swat-01/2013 and D2/Pk/Swat-02/2013) were isolated from the patients, suffered from the 2013 outbreak of Swat, KP.

3.2.1.1 Open reading frame and encoded amino acid profile

Open reading frames and encoded amino acids profiles of all the four complete genomes, sequenced in this study, were determined. Coding sequence and amino acid profile of the strain D2/Pk/A1/2011 is attached as Annex-V. The coding sequence lies between 65-10240 nucleotide positions which encodes a polyprotein of 3391 amino acids. Length of the coding sequence is 10176 nucleotides.

Coding sequence of the DENV-2 isolate D2/Pk/A2/2011 is attached as Annex-VI. ORF begins at nucleotide position number 65 and ends at 10240. The coding region encodes a polyprotein of 3391 amino acids.
Open Reading Frame and encoded amino acids profile of the DENV-2, isolate D2/Pk/Swat-01/2013 is attached as Annex-VII. Length of the coding sequence is 10176 bp.

ORF and encoded amino acid of the DENV-2 isolate D2/Pk/Swat-02/2013 is attached as Annex-VIII. The encoded amino acid is a polyprotein of 3391 amino acids.

3.2.1.2 Conserved domains architecture

Annotation of the proteins of the complete characterized DENV2 in our study with the identification of conserved domain footprints and functional sites inferred from these footprints are summarized in table 3.2.

In NS3 (amino acid position from 1493-1643), our sequences corresponded to the peptidase family S7 (Flavivirus serine protease) domain, DEAD-like helicases superfamily associated with ATP-dependent helicase activity and sites predicted to be required for cell attachment and targeting signal for microbodies. In NS5 from 2546 to 2713 corresponded to the conserved FtsJ-like methyltransferase (FtsJ is a member of a universally conserved heat shock protein family). N-terminal domain of NS5 has been suggested to be methyltransferase and involved in capping of viral RNA (Egloff et al., 2002). Flavivirus glycoproteins turned out to be conserved for central and dimerization domain (281-576). Similarly, amino acids from 585-674 and 679-751 corresponded to be conserved for Immunoglobulin-like domain III (C-
terminal domain) of Flavivirus envelope glycoprotein E and homodimer interface (polypeptide binding) of E protein.
<table>
<thead>
<tr>
<th>Strain</th>
<th>ORF</th>
<th>length</th>
<th>1Flavi capsid</th>
<th>2Flavi Propep</th>
<th>3Flavi M</th>
<th>4Flavi glycoprot</th>
<th>5Flavi E_C</th>
<th>6Flavi E_stem</th>
<th>7Flavi NS1</th>
<th>8Flavi NS2A</th>
<th>9Flavi NS2B</th>
<th>10Peptidase S7</th>
<th>11DEXDc</th>
<th>12DEXDc</th>
<th>13HELICc</th>
<th>14Flavi NS4A</th>
<th>15Flavi NS4B</th>
<th>16FtsJ</th>
<th>17Flavi NS5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2/A1</td>
<td>60-10240</td>
<td>3391</td>
<td>5-114</td>
<td>119-204</td>
<td>206-280</td>
<td>281-576</td>
<td>585-674</td>
<td>679..&gt;751</td>
<td>777..1131</td>
<td>1138..1345</td>
<td>1349..1475</td>
<td>1493..1643</td>
<td>1657..1800</td>
<td>1664..1793</td>
<td>&lt;1830..1939</td>
<td>2097..2240</td>
<td>2244..2490</td>
<td>2546..2713</td>
<td>2742..3387</td>
</tr>
<tr>
<td>D2/A2</td>
<td>60-10240</td>
<td>3391</td>
<td>5-114</td>
<td>119-204</td>
<td>206-280</td>
<td>281-576</td>
<td>585-674</td>
<td>679..&gt;751</td>
<td>777..1131</td>
<td>1138..1345</td>
<td>1349..1475</td>
<td>1493..1643</td>
<td>1657..1800</td>
<td>1664..1793</td>
<td>&lt;1830..1939</td>
<td>2097..2240</td>
<td>2244..2490</td>
<td>2546..2713</td>
<td>2742..3387</td>
</tr>
<tr>
<td>D2/Swat-01</td>
<td>60-10240</td>
<td>3391</td>
<td>5-114</td>
<td>119-204</td>
<td>206-280</td>
<td>281-576</td>
<td>585-674</td>
<td>679..&gt;751</td>
<td>777..1131</td>
<td>1138..1345</td>
<td>1349..1475</td>
<td>1493..1643</td>
<td>1657..1800</td>
<td>1664..1793</td>
<td>&lt;1830..1939</td>
<td>2097..2240</td>
<td>2244..2490</td>
<td>2546..2713</td>
<td>2742..3387</td>
</tr>
<tr>
<td>D2/Swat-02</td>
<td>60-10240</td>
<td>3391</td>
<td>5-114</td>
<td>119-204</td>
<td>206-280</td>
<td>281-576</td>
<td>585-674</td>
<td>679..&gt;751</td>
<td>777..1131</td>
<td>1138..1345</td>
<td>1349..1475</td>
<td>1493..1643</td>
<td>1657..1800</td>
<td>1664..1793</td>
<td>&lt;1830..1939</td>
<td>2097..2240</td>
<td>2244..2490</td>
<td>2546..2713</td>
<td>2742..3387</td>
</tr>
</tbody>
</table>

Note:
1: Flavivirus capsid protein C
2: Flavivirus polyprotein propeptide
3: Flavivirus envelope glycoprotein M
4: Flavivirus glycoprotein, central and dimerization domain
5: Immunoglobulin-like domain III (C-terminal domain) of Flavivirus envelope glycoprotein E:
   a) (590,591,593,594,601,602,646): homodimer interface [polypeptide binding]
   b) (597,614,632,635,637): low pH domain interface [polypeptide binding]
   c) (598,614,630,632,650): low pH trimer interface [polypeptide binding]
6: Flavivirus envelope glycoprotein E, stem/anchor domain
7: Flavivirus non-structural Protein NS1
8: Flavivirus non-structural protein NS2A
9: Flavivirus non-structural protein NS2B
10: Peptidase S7, Flavivirus NS3 serine protease
11: DEAD-like helicases superfamily
12: DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region:
   a) (1671..1675):
   b) (1759..1762):
13: Helicase superfamily c-terminal domain; associated with DEXDc, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins:
   a) (1838..1841,1883..1885):
   b) (1891,1931,1935,1938):
14: Flavivirus NS4A
15: Flavivirus NS4B
16: FtsJ-like methyltransferase
17: Flavivirus RNA-directed RNA polymerase

74
3.2.1.3 Phylogenetic analysis

The full length genomes of DENV2 isolated, sequenced and characterized in this study were blasted against global isolates and phylogenetic analysis was conducted for the total of 67 DENV-2 full length sequences. Figure 3.4 refers to the results of the evolutionary analysis.

The tree is rooted on the DENV-3 Indian isolate with accession no. JX475906.1. Approximate Likelihood-Ratio Test (aLRT SH-like) were employed as branch support values. aLRT values equal or higher than 0.8 are shown on the branches.

The analysis clearly distributed the DENV-2 isolates into five separate clades. The phylogenetic analysis indicates that our dengue virus isolates (from this study) grouped into the cosmopolitan genotype of DENV-2. Dengue virus isolates of this study are marked with star as indicated in figure 3.4. This analysis shows that all the Pakistani isolates are more closely related to Indian isolates. The close genetic relationship of the Pakistani isolates (2011-2013) with those circulating in India indicates that Pakistani DENV-2 isolates have descended from an ancestral strain, previously prevalent in other areas of the India subcontinent. The phylogenetic analysis further illustrates the fact that the outbreak of 2011 and 2013 is a continuation of the previous dengue infections, traveled from Sindh province to Punjab and expanded to the remote areas of Khyber Pakhtunkhwa province.
Figure 3.4 Phylogenetic tree generated by maximum Likelihood analysis of nucleotide sequences from the entire genome of 67 DENV-2 viruses. Representative strain of DEN-3 was used to root the tree. The names of the DENV-2 isolates refer to the accession, year of isolation and the country of origin. Support values greater than 0.8 based on aLRT-SH like branch support are shown for key nodes. Isolates of the current study are indicated with star.
3.2.1.4 Nucleotide and amino acid identities

The phylogenetic analysis based on the complete genome of DENV-2, characterized in this study, grouped the isolates into cosmopolitan genotype. All the 17 DENV-2 of the cosmopolitan lineage were further analyzed for the nucleotides and amino acids identities. Percent nucleotide identities among DENV-2 full sequence cosmopolitan genotypes indicate that viruses isolated from the 2013 outbreak is continuation of previous outbreaks of 2011 (99.96%). Similarly this analysis indicates that the Pakistani isolates of DENV-2 are closely genetically identical to the Indian isolates (97.10%), followed by Sri Lankan strain of 1996 (96.11%) as presented in table 3.3. Results of the percent amino acids identities are given in table 3.4. Amino acids identities show a similar pattern of genetic relatedness to that of percent nucleotide identities. It shows that DENV-2 genotypes circulating in district Swat, KP and Punjab are the continuation of series of dengue outbreaks hit the country previously. Furthermore these circulating DEN-2 genotypes are more closely related to Indian (98.59%) and Sri Lankan (98.42) isolates.
Table 3.3 Percent nucleotide sequence identity among full length DENV-2, cosmopolitan genotypes

| Isolates                        | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1: JX475906.1/India/2009        |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 2: FJ882602.1/LK/1996           | 96.62 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 3: JQ955623.1/India/2009        | 95.88 | 96.61 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 4: KF479233.1/China/2013        | 95.64 | 96.20 | 96.48 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5: FJ898454.1/INDIA/2006        | 96.03 | 96.65 | 96.90 | 97.29 |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 6: KF041233.1/Pakistan/2011     | 95.77 | 96.45 | 96.64 | 96.94 | 97.43 |    |    |    |    |    |    |    |    |    |    |    |    |
| 7: KF041234.1/Pakistan/2011     | 95.67 | 96.34 | 96.48 | 96.85 | 97.21 | 98.77 |    |    |    |    |    |    |    |    |    |    |    |
| 8: KF041235.1/Pakistan/2009     | 95.77 | 96.34 | 96.64 | 97.01 | 97.39 | 98.97 | 98.95 |    |    |    |    |    |    |    |    |    |    |
| 9: KF041232.1/Pakistan2011      | 95.69 | 96.34 | 96.51 | 96.93 | 97.43 | 98.67 | 99.12 | 98.87 |    |    |    |    |    |    |    |    |    |
| 10: KF041237.1/Pakistan/2009    | 95.79 | 96.42 | 96.79 | 97.12 | 97.52 | 99.06 | 98.89 | 99.06 | 98.85 |    |    |    |    |    |    |    |    |
| 11: KM217156/Pk/A2/2011        | 95.46 | 96.12 | 96.41 | 96.68 | 97.07 | 98.39 | 98.58 | 98.57 | 98.62 | 98.52 |    |    |    |    |    |    |    |
| 12: KM217157/Pk/A1/2011        | 95.53 | 96.11 | 96.42 | 96.82 | 97.10 | 98.51 | 98.62 | 98.60 | 98.57 | 98.69 | 99.21 |    |    |    |    |    |    |
| 14: KM217158/Pk/Swat/2013       | 95.51 | 96.07 | 96.36 | 96.72 | 97.07 | 98.47 | 98.50 | 98.68 | 98.51 | 98.61 | 99.12 | 99.27 | 100.00 |    |    |    |
| 15: KF041236.1/Pakistan/2008    | 95.70 | 96.30 | 96.16 | 95.89 | 96.31 | 95.97 | 95.93 | 96.16 | 95.81 | 96.25 | 95.63 | 95.85 | 96.45 | 95.75 |    |    |
| 16: AF276619.1/China            | 96.37 | 96.97 | 96.75 | 96.55 | 96.97 | 96.75 | 96.74 | 96.89 | 96.69 | 96.97 | 96.56 | 96.52 | 97.23 | 96.43 | 96.89 |    |
| 17: AF359579.1/China            | 96.24 | 96.85 | 96.73 | 96.48 | 96.78 | 96.63 | 96.59 | 96.71 | 96.46 | 96.77 | 96.35 | 96.45 | 97.12 | 96.37 | 96.82 | 98.99 |
Table 3.4 Percent amino acids sequence identity among full length DENV-2, cosmopolitan genotypes

<table>
<thead>
<tr>
<th>Isolates</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: KF041236.1/Pakistan/2008</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2: JX475906.1/India/2009</td>
<td>97.32</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: FJ882602.1/LK/1996</td>
<td>97.26</td>
<td>98.51</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4: AF359579.1/China</td>
<td>97.17</td>
<td>98.12</td>
<td>97.90</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5: JQ955623.1/India/2009</td>
<td>97.51</td>
<td>98.46</td>
<td>98.23</td>
<td>98.15</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6: AF276619.1/China</td>
<td>97.75</td>
<td>98.57</td>
<td>98.40</td>
<td>98.77</td>
<td>98.32</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7: KF479233.1/China/2013</td>
<td>97.31</td>
<td>98.26</td>
<td>98.04</td>
<td>98.10</td>
<td>98.43</td>
<td>98.43</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9: KF041233.1/Pakistan/2011</td>
<td>97.51</td>
<td>98.23</td>
<td>98.18</td>
<td>98.01</td>
<td>98.10</td>
<td>98.13</td>
<td>98.49</td>
<td>6.99</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10: KF041234.1/Pakistan/2011</td>
<td>97.36</td>
<td>98.21</td>
<td>98.23</td>
<td>98.21</td>
<td>98.27</td>
<td>98.24</td>
<td>98.71</td>
<td>6.84</td>
<td>98.94</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11: KF041235.1/Pakistan/2009</td>
<td>97.45</td>
<td>98.26</td>
<td>98.26</td>
<td>98.07</td>
<td>98.38</td>
<td>98.41</td>
<td>98.57</td>
<td>6.83</td>
<td>98.91</td>
<td>98.99</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12: KF041232.1/Pakistan2011</td>
<td>97.58</td>
<td>98.43</td>
<td>98.24</td>
<td>98.10</td>
<td>98.27</td>
<td>98.24</td>
<td>98.52</td>
<td>6.83</td>
<td>98.83</td>
<td>99.10</td>
<td>99.02</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13: KF041237.1/Pakistan/2009</td>
<td>97.50</td>
<td>98.46</td>
<td>98.40</td>
<td>98.35</td>
<td>98.65</td>
<td>98.57</td>
<td>98.71</td>
<td>6.90</td>
<td>99.16</td>
<td>99.27</td>
<td>99.30</td>
<td>99.19</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14: KM217156/Pk/A2/2011</td>
<td>96.91</td>
<td>97.81</td>
<td>97.65</td>
<td>97.76</td>
<td>98.06</td>
<td>97.95</td>
<td>97.98</td>
<td>6.94</td>
<td>98.43</td>
<td>98.77</td>
<td>98.65</td>
<td>98.52</td>
<td>98.49</td>
<td>__</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15: KM217157/Pk/A1/2011</td>
<td>96.89</td>
<td>97.78</td>
<td>97.59</td>
<td>97.62</td>
<td>98.06</td>
<td>97.87</td>
<td>97.96</td>
<td>6.97</td>
<td>98.46</td>
<td>98.71</td>
<td>98.65</td>
<td>98.46</td>
<td>98.32</td>
<td>99.16</td>
<td>__</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16: KJ701507/Pk/swat/2013</td>
<td>97.49</td>
<td>98.36</td>
<td>98.42</td>
<td>98.36</td>
<td>98.59</td>
<td>98.65</td>
<td>98.81</td>
<td>6.92</td>
<td>99.21</td>
<td>99.24</td>
<td>99.29</td>
<td>99.35</td>
<td>99.92</td>
<td>99.92</td>
<td>__</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17: KM217158/Pk/Swat/2013</td>
<td>97.02</td>
<td>97.95</td>
<td>97.70</td>
<td>97.70</td>
<td>98.09</td>
<td>97.98</td>
<td>98.15</td>
<td>6.92</td>
<td>98.54</td>
<td>98.85</td>
<td>98.82</td>
<td>98.60</td>
<td>98.46</td>
<td>99.24</td>
<td>98.97</td>
<td>99.97</td>
<td>__</td>
</tr>
</tbody>
</table>
3.2.1.5 Recombination analysis

To explore any recombination breakpoints in the entire coding region of all the 17 cosmopolitan DENV-2, Genetic Algorithm for Recombination detection (GARD) method was employed using TrN93 as the best fit nucleotide substitution bias model. The analysis indicated no recombination event in the viruses indicated in figure 3.5.

![GARD analysis plot of the cosmopolitan DENV-2 genotypes. Evidence of 0 breakpoints was found using AICc; an improvement of 0 points was achieved over the model without recombination.](image)

3.2.1.6 Selection pressure analysis

All the 17 full length cosmopolitan sequences of DENV-2 were further analyzed for selection pressure on any codon or ORF using Hyphy package. To be more precise and accurate, care was taken to analyze the dataset with various methods. The three different methods of selection pressure analysis used in this study were SLAC, FELL and REL for calculating the ratio of
nonsynonymous substitution (dN) to the synonymous nucleotide substitutions per site (dS) which is dN/dS as explained previously (Zhang et al., 2006; Kosakovsky & Frost 2005). Only those results are presented here on which the three methods agreed upon using the best fit TrN93 substitution model by setting significance levels to p≤0.1 or Bayes factor ≥50. Our results indicate that codon 65 of the capsid gene, 155 of PrM, 203 of the E gene and 254 of NS1 are under negative selection pressure as shown in table 3.5. On the other hand there no codon was detected to be under positive selection pressure as indicated by the ratio of dN/dS, means that the Pakistani isolates are under purifying selection pressure.

Table 3.5 Results of the selection pressure analysis acting on the ORF of the DENV-2 cosmopolitan genotypes

<table>
<thead>
<tr>
<th>Codon</th>
<th>Protein and amino acid residues</th>
<th>DENV-2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SLAC</td>
<td>FEL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dN-dS</td>
<td>p-value</td>
</tr>
<tr>
<td>65</td>
<td>C&lt;sub&gt;65&lt;/sub&gt;</td>
<td>-85.4528</td>
<td>0.0083</td>
</tr>
<tr>
<td>155</td>
<td>PrM&lt;sub&gt;155&lt;/sub&gt;</td>
<td>-23.1677</td>
<td>0.0967</td>
</tr>
<tr>
<td>203</td>
<td>E&lt;sub&gt;203&lt;/sub&gt;</td>
<td>-33.8781</td>
<td>0.0284</td>
</tr>
<tr>
<td>254</td>
<td>NS1&lt;sub&gt;254&lt;/sub&gt;</td>
<td>-35.8083</td>
<td>0.0257</td>
</tr>
</tbody>
</table>

The highest substitution rate observed in this study among amino acid residues of the ORF of all the 17 cosmopolitan DENV-2 sequences was between arginine (Arg) and lysine (Lys) residues, followed by valine (Val) and isoleucine (Ile) as indicated in figure 3.6.
Figure 3.6 Substitution summary of DENV-2 cosmopolitan genotypes by residue. 'Mix' represents a codon with one or more nucleotide mixture. Color intensity is proportional to the number of substitutions between two residues relative to the highest number for all pairs (brighter = higher, lighter = lower).
3.2.1.7 Phylogenetic analysis of DENV-2 based on NS3 gene

The Non-structural gene 3 (1854 bp long) of DENV-2, isolated from the patients of 2013 outbreak of Swat, KP, characterized and its phylogenetic analysis was determined as indicated in figure 3.7. The tree is rooted on sylvatic dengue virus sequence (accession number EU056810.1). NS3 sequences characterized in this study are presented in red color (figure 3.7). NS3 gene based phylogenetic tree of DENV-2 grouped all the Pakistani isolates into the cosmopolitan genotype group with other types from geographically distinct regions such as East Asia, Australia, South Asia etc. Only the bootstrap values higher than 75 are shown on the main nodes of the tree. The analysis shows that the 2013 dengue outbreak in Swat KP was a continuation of the 2011 outbreak of Lahore, Punjab. This phylogenetic analysis also revealed the fact that Pakistani DENV-2 isolates are genetically more close to Indian and Sri Lankan strains.
Figure 3.7 Phylogenetic analyses of the DENV-2 Swat isolates, based on NS3 genes.
3.2.2 GENOME ANALYSIS OF DENGUE VIRUS TYPE 3

3.2.2.1 Phylogenetic analysis of DENV-3

Phylogenetic analysis of DENV-3 based on complete E gene and NS3 gene sequences are shown in figure 3.8 and 3.9 respectively. Five complete E and NS3 gene sequences, isolated and sequenced from the 2011 dengue outbreak in Punjab and five from the 2013 outbreak in Swat, KP are highlighted in the phylogenetic trees (figure 3.8 and 3.9). The phylogenetic analysis grouped all the DENV-3 types (analyzed in this study) into genotype III. The results show that the outbreaks of 2011 and 2013 are the continuation of the previous outbreaks in Pakistan (2006-2013). Furthermore, all the Pakistani dengue virus type 3 isolates are closer genetically to the Indian and Chinese isolates as indicated in figure 3.8 and 3.9.
Figure 3.8 Phylogenetic tree generated by maximum Likelihood analysis of nucleotide sequences from the complete E gene sequences of 57 dengue 3 viruses. Representative strain of DENV-2 was used to root the tree. DENV-3 isolates are represented by accession numbers, year of isolation and the country of origin. Branch support values are shown for key nodes. Isolates of the current study are highlighted.
Figure 3.9 Phylogenetic tree generated by maximum Likelihood analysis of nucleotide sequences from the complete NS3 gene sequences of 57 dengue 3 viruses. Representative strain of DENV-2 was used to root the tree. DENV-3 isolates are represented by the accessions, year of isolation and the country of origin. Branch support values are shown for key nodes. Isolates of the current study are highlighted.
3.2.2.2 Nucleotides and amino acids differences in the Core protein gene of DENV-3

The core protein gene (342 bp) of DENV-3 was characterized in this study. A total of 10 C gene sequences (five from 2011 and five from 2013 outbreaks) were analyzed. The results based on nucleotides position are displayed in table 3.6 while that of amino acids conservation is given in table 3.7. The analysis involved 16 dengue 3 isolates from Pakistan (2006-2013). The isolates reported in this study are highlighted. The nucleotide changes (A → G, G → C, T → A and G → A) observed were among sites at positions 134, 179, 291 and 327 respectively.

Table 3.6 nucleotides substitution of Core protein among Pakistani isolates of DENV-3

<table>
<thead>
<tr>
<th>Isolates</th>
<th>Nucleotides substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM217140/D3/Pk/F-10/2011</td>
<td>A  G  T  G</td>
</tr>
<tr>
<td>KM217136/D3/Pk/F-14/2011</td>
<td>.  .  .  .</td>
</tr>
<tr>
<td>KF041259.1/Pk/2006</td>
<td>.  C  A .</td>
</tr>
<tr>
<td>KF041257.1/Pk/2006</td>
<td>.  C  A .</td>
</tr>
<tr>
<td>KF041256.1/Pk/2006</td>
<td>G  C  A .</td>
</tr>
<tr>
<td>KF041254.1/Pk/2008</td>
<td>.  C  A  A</td>
</tr>
</tbody>
</table>

As for as the amino acids description of the C gene is concerned, it also portrays a similar pattern. The DENV-3 of the 2011 outbreak was the most conserved for the amino acids K (lysine) R (arginine) and N (asparagine). The
viruses of the 2013 outbreak of district Swat showed the amino acids change for the nonpolar P (proline) and polar K (lysine) instead of R and N respectively. The important observation here is that all the viruses isolated from the 2011 dengue outbreak of Punjab showed a conserved pattern of nucleotides/amino acids as compared to the DENV-3 isolates of the recent 2013 outbreak. This pattern may be one of the reasons that viruses established a severe infection in the relatively colder area of district Swat and adapted to the new environment.

Table 3.7 Conserved amino acids in the Core protein of DENV-3 Pakistani Isolates

<table>
<thead>
<tr>
<th>Isolates</th>
<th>Conserved amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM217140/D3/Pk/F-10/2011</td>
<td>K R N</td>
</tr>
<tr>
<td>KM217136/D3/Pk/F-14/2011</td>
<td>. . .</td>
</tr>
<tr>
<td>KM217145/D3/Pk/Swat-01A/2013</td>
<td>. P K</td>
</tr>
<tr>
<td>KM217144/D3/Pk/Swat-02A/2013</td>
<td>. P K</td>
</tr>
<tr>
<td>KM217143/D3/Pk/Swat-03A/2013</td>
<td>. P K</td>
</tr>
<tr>
<td>KM217142/D3/Pk/Swat-04A/2013</td>
<td>. P K</td>
</tr>
<tr>
<td>KM217141/D3/Pk/Swat-05A/2013</td>
<td>. P K</td>
</tr>
<tr>
<td>KF041259.1/Pk/2006</td>
<td>. P K</td>
</tr>
<tr>
<td>KF041258.1/Pk/2009</td>
<td>. P K</td>
</tr>
<tr>
<td>KF041257.1/Pk/2006</td>
<td>. P K</td>
</tr>
<tr>
<td>KF041255.1/Pk/55505/2007</td>
<td>. P K</td>
</tr>
<tr>
<td>KF041256.1/Pk/2006</td>
<td>R P K</td>
</tr>
<tr>
<td>KF041254.1/Pk/2008</td>
<td>. P K</td>
</tr>
</tbody>
</table>

3.2.2.3 Selection pressure analysis on DENV-3

The E and NS3 genes were subjected to check whether selection pressure is acting on any of the codon or evolving neutrally.
The three different methods of selection pressure analysis used in this study were SLAC, FEL and REL. Only those results are presented here on which the three methods are agreed upon using the best fit general time reversible (REV) substitution model by keeping significance levels to $p \leq 0.1$ or Bayes factor $\geq 50$.

Two codons (94 and 477) of the E gene were found under negative pressure whereas no codon was identified to be under any positive selection pressure (Table 3.8). For DENV-3 NS3 gene, no rates with $dN > dS$ were inferred for this dataset, suggesting that all sites are under purifying selection.

Table 3.8 Analysis of the selection pressure on codons of the E gene of DENV-3

<table>
<thead>
<tr>
<th>Codon</th>
<th>Protein and amino acid residues</th>
<th>SLAC</th>
<th>FEL</th>
<th>REL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$dN-dS$</td>
<td>$p$-value</td>
<td>$dN-dS$</td>
</tr>
<tr>
<td>94</td>
<td>E$_{94}$</td>
<td>-54.8518</td>
<td>0.0951</td>
<td>-1020.95</td>
</tr>
<tr>
<td>477</td>
<td>E$_{477}$</td>
<td>-58.0619</td>
<td>0.0957</td>
<td>-1046.07</td>
</tr>
</tbody>
</table>
Chapter 4
DISCUSSION

4.1 DISEASE ASSESSMENT

Dengue is one of the common and most important mosquitoes transmitted viral disease in the world. Around 2.5 billion inhabitants of the tropics and subtropics of the world are living with the risk of getting the infection, makes up to 2/5th of the world’s population (Dash et al., 2013, WHO, 2009; Gubler & Clark, 1995). An estimated global rate of dengue infection is 50-100 million every year with 24,000 mortalities and 500,000 cases requiring hospitalizations (Halstead, 1988; WHO, 1997). The recent study conducted on dengue infection, suggests that annually 400 million people are infecting with the disease around the globe, out of which 100 million are clinically apparent (Bhatt et al., 2013). Furthermore, the fact that the number of infection be doubled in those people who are living in the tropical and sub-tropical areas (Anonymous, 2009), makes the disease a major threat around the globe.

DENV is known to be transmitted by various species belong to genus Aedes of the Culicidae family but still Aedes aegypti is the principle vector that also transmits the yellow fever virus (YFV) (Cleland et al., 1916). Both Ae. Aegypti and Ae. Albopictus been known to be anthropohilic in nature (Ponlawat & Harrington, 2005). The areas of the tropics and subtropics, designated as urban and semi urban, both the vector species could be found in abundantly.

DENV is endemic in Pakistan. During the post monsoon period between October and December, the infection reaches to its peak level (Khan et al., 2010). The first confirmed dengue outbreak was observed in Karachi Pakistan.
in 1994. More recent (2013) outbreaks were also described. (Burney, 1966; Burney & Munir, 1966; Hayes et al., 1982; Hayes & Burney, 1981; Sugamata et al., 1986; Chan et al., 1995; Ali et al., 2013).

So far in Pakistan, only few, detailed molecular epidemiological studies have been conducted, related to the prevalence of DENV in the country. Studies encompassing the molecular epidemiology and evolutionary genetics are essential to determine the origin and spread of viruses, to enhance our understanding on the pathogenesis of disease, the etiology of epidemics as well as the genetic basis of virulence. Therefore, in the current study, a detailed molecular characterization of Pakistani DENV isolates was investigated from the 2011 and 2013 outbreaks in Pakistan in Punjab (Lahore) and KP (Swat) respectfully.

4.2 EPIDEMIOLOGY

Pakistan witnessed the worse of dengue infection in 2011 and in 2013. The geographical location of Pakistan, with hot and humid summers, favors the breeding of the Aedes species. Factors responsible for the increased dengue incidence includes the lack of effective mosquito control in the areas where dengue is endemic, unprecedented global population growth and the associated unplanned and uncontrolled urbanization, increased air travel, which provides the ideal mechanism for the transport of DENV infected humans and mosquitoes between different population centers of the world (Gubler & Trent, 1993; Gubler, 1996). Lahore is the second largest city in Pakistan, is a commercial hub of Pakistan and is linked to all parts of the
country through various communication and trading routes. Due to lack of any effective control strategy, DENV also spread from this city into other parts of the country. During the recent 2013 DENV outbreak in Swat a total of 8,343 cases with 57 deaths were reported (Anonymous, 2013).

Here we report for the first time in Pakistan that all the four DENV serotypes circulating in Punjab province unlike other studies like (Koo et al., 2013 & Fatima et al., 2013) probably involving less number of patients while in district Swat only DENV-2 and DENV-3 were confirmed to be circulating. In 2011 in Punjab province, 74% of the samples were positive by RT-PCR and indicated to have active dengue infection that comprised 56% males and 44% female patients. Majority of the actively infected people were from Lahore (83%) followed by other districts like Sargodha (9%), Sheikhupura (5%), Kasur (4%) and Rahim Yar Khan (0.5%) (Table 3.1). DENV-2 (42%) and DENV-3 (41%) turned out to be the most prevalent in Punjab province. Similarly, 43% KP-based people in Punjab tested and confirmed by RT-PCR to have active dengue infection. More than 80% of these individuals were males while only 20% were females that had active dengue infection. Majority of the patients that were positive by RT-PCR belonged to Peshawar (43%), followed by Mardan (18%), Nowshera (13%), Swat (13%), Charsadda (8%) and Kohat (3%). Both DENV-2 and DENV-3 were the most abundant serotypes in Punjab and KP provinces during the 2011 outbreak and had significantly high incidence than DENV-1 and 4 serotypes.
In this study we document for the first time the 2013 dengue outbreak in
district Swat. This outbreak was considered to be unique because the dengue
was reported to hit the far remote and colder area of district Swat as
compared to warm environment of Punjab. Initially at the onset of the
infection we collected 740 blood samples from patients, admitted to various
hospitals of district Swat. Analysis of the 740 blood samples showed that only
618 patients were positive for antibodies. The remaining patients (122), having
flue like symptoms and were supposed to have dengue infection but on
investigation, no antibodies were detected in their blood. However, active
infection was confirmed by RT-PCR in only 200 patients of the 618 positive by
serology. The 418 patients in which DENV was not detected by RT-PCR but
were positive for antibodies needs further investigation. One possibility is
that the blood samples from these patients might be taken when the
antibodies were fully developed against DENV, particularly when IgG titer is
high then detection of the virus in the blood is almost negligible as indicated
previously in another study (Laue et al., 1999). Both DENV-2 and DENV-3
were the most prevalent serotypes in district Swat. Like the 2011 outbreak of
Punjab, the people in between 15 and 45 years of age were the most infected
(44.50%) in this outbreak.

As the dengue outbreak progressed in district Swat, 414 serum samples were
collected from clinically diagnosed patients. Of the 414 serum samples,
88.40% were found to be anti-DENV IgG and/or IgM positive (Figure 3.2). A
total of 28.68% patients were found to have active DENV infection using the
reverse transcriptase Polymerase chain reaction (RT-PCR). The results obtained indicated that both DENV-2 and particularly DENV-3 were solely found in the outbreak. Surprisingly, DENV-1 and DENV-4 were not detected in any of the sample. This is an interesting development and the role of Aedes mosquitoes should be investigated in spreading the disease in Swat. Swat is relatively colder area and far away from Punjab, the establishment of the infection in Swat needs attention. We presume that A. albopictus may have a dominant role in the establishment of the disease in district Swat rather than A. aegypti because the literature survey indicates that A. albopictus is mostly common in highly urbanized areas while Ae. albopictus in rural, suburban and vegetated urban areas (Braks et al., 2003; Paupy et al., 2009). One sample contained both DENV-2 and DENV-3 and on contacting the person we came to know that has been died, indicating the role of ADE. The person might have infected earlier with one serotype and on this time infected again by another serotype of DENV that leads to ADE and shock. Lack of awareness, health facilities, investigation, proper record keeping and observation are the key points which could lead to disease severity and ADE. Good supportive medication such as parenteral fluid therapy even in case of DSS can reduce mortalities rate to less than 1% (Wills et al., 2005; Lam et al., 2013).

The analysis of the 2011 and 2013 outbreaks in Pakistan confirmed that DENV-2 and DENV-3 were the most prevalent serotypes circulating in the country. Detailed molecular analysis of the prevalent dengue serotypes in Pakistan was our main focuss to cope with.
Four complete genomes (two each from Punjab and Swat) of DENV-2 were sequenced, characterized and submitted to GenBank. NS3 gene from seven DENV-2 Swat isolates (2013 outbreak) were also successfully sequenced, characterized and submitted to GenBank.

4.3 CHARACTERIZATION OF THE VIRUS AND PHYLOGENY

The complete genome characterization of the DENV-2 (10.6 kb) reported in this study showed that the coding sequence lies between 65-10240 nucleotide positions which encode a polyprotein of 3391 amino acids. Length of the coding sequence/ORF (10176 nucleotides) was determined using the online available tool in NCBI database. The first 64 nucleotides correspond to the 5’ UTR and last 416 nucleotides to the 3’ UTR of the genome. On BLAST search using NCBI database, the complete sequences characterized in this study showed 99-100% homology to the sequences previously reported from Pakistan. The complete DENV-2 sequences isolated from the patients in Swat also showed 100% homology to the sequences we reported or been previously reported from Punjab.

Phylogenetic analysis of the 67 complete genomes of DENV-2 indicates that our dengue virus isolates (from this study) grouped into the cosmopolitan genotype of DENV-2. The phylogenetic analysis illustrates the fact that the outbreak of 2011 and 2013 is a continuation of the previous dengue infections, traveled from Sindh province to Punjab and expanded to the remote areas of Khyber Pakhtunkhwa province. This analysis further shows that all the Pakistani isolates are more closely related to Indian isolates. The close genetic
relationship of the Pakistani isolates (2011-2013) with those circulating in India indicates that Pakistani DENV-2 viruses emerged from an ancestral lineage, previously prevalent in other parts of the India subcontinent.

The percent nucleotide identities among DENV-2 full sequence cosmopolitan genotypes indicate that viruses isolated from the 2013 outbreak is continuation of previous outbreaks of 2011 (99.96%). Similarly this analysis indicate that the Pakistani isolates of DENV-2 are closely genetically identical to the Indian isolates (97.10%), followed by Sri Lankan strain of 1996 (96.11%) (Table 3.3). Amino acids identities show a similar pattern of genetic relatedness to that of percent nucleotide identities (Table 3.4). It shows that DENV-2 genotypes circulating in district Swat, KP and Punjab are the continuation of series of dengue outbreaks hit the country previously. Furthermore these circulating DEN-2 genotypes are more closely related to Indian (98.59%) and Sri Lankan (98.42) isolates.

Recently, NS3 gene has been used for the phylogenetic analysis of the DENV (Zhang et al., 2014). In an attempt to determine whether phylogenetic analysis based on the NS3 gene of DENV-2 portrays the exact classification in case of Pakistani DENV-2 isolates, 7 samples (Swat isolates) were characterized randomly for NS3 gene. The phylogenetic analysis based on NS3 gene grouped the Pakistani isolates into the cosmopolitan genotypes like the whole genome phylogeny explained above. The group contained the closely related previously characterized Pakistani isolates as well as Indian, Chinese and Sri Lankan isolates (Figure 3.7). This analysis indicates that NS3 gene based
phylogeny may be used as an alternative to the complete genome phylogeny of DENV-2 Pakistani isolates.

10 complete sequences each of C (342 bp), E (1479 bp) and NS3 (1857 bp) genes of DENV-3 both from Swat and Punjab isolates were also characterized in the study. The attempt to characterize and sequence the complete genome of DENV-3 was not successful. The possible causes for this may be the low titer of the DENV-3 in the blood of the patients, the samples be collected after the acute phase of the disease been passed, the presence of high titers of the IgM and IgG in the blood or some probable contamination of the blood sample during collection time. The phylogenetic analysis based on the complete E and NS3 genes revealed that the isolates of the current study fall into the genotype III of DENV-3. This analysis indicates that dengue infection migrated from Punjab to Swat and established its roots there. The E gene based phylogeny relates the Pakistani isolates more closely to Indian isolates while the NS3 based phylogeny links it to Chinese and then Indian isolates. This issue needs further elucidation and confirmation through whole genome sequencing of DENV-3. Over all the analysis illustrates the close genetic relationship of Pakistani DENV-3 isolates to that of Indian and Chinese isolates which strengthens the belief that dengue infection has been transported to Pakistan from the neighboring countries. Examining the whole scenario it is evident that dengue infection started in Karachi and then with the passage to time gradually introduced into Punjab province and finally now established in district Swat. The possible reasons behind this expansion
may be the hectic visits of the people to the big cities like Karachi and Lahore for jobs/settlements and trade.

Nucleotides and amino acids conservation and substitutions pattern of the C gene of DENV-3 Pakistani isolates indicates consistency in the 2011 isolates as compared to the 2013 isolates. The nucleotide changes (A to G, G to C, T to A and G to A) observed were among sites at positions 134, 179, 291 and 327 respectively. An interesting point here is that newly isolated viruses from the recent outbreak in Swat (2013) showed nucleotide changes of C instead of G and A instead of T. On the other hand all the viruses isolated from Punjab (2011) revealed a conserved nature for nucleotides A, G, T and G (Table 3.6).

Analysis of basic amino acids (R, K, and N) of DENV-3 of the 2011 outbreak was the most conserved. The viruses of the 2013 outbreak of district Swat showed the amino acids change for the nonpolar P (proline) and polar K (lysine) instead of R and N respectively (Table 3.7). The important observation here is that all the viruses isolated from the 2011 dengue outbreak of Punjab showed a conserved pattern of nucleotides/amino acids as compared to the DENV-3 isolates of the recent 2013 outbreak. This pattern may be one of the reasons that viruses established a severe infection in the relatively colder area of district Swat and adapted to the new environment. We suggest that these changes at the molecular level contributed for the dengue viruses to establish its roots in the far away, ruler and relatively colder areas of district Swat and adjacent territories.
Screening of the aligned sequences for recombination prior to phylogeny is essential because in the presence of recombination in sequences can mislead the evolutionary history of the samples (Pond et al., 2006; Shriner et al., 2003). Recombination may seriously change the accuracy and power of fundamentally essential tools of molecular evolutionary analyses such as molecular clock inference (Schierup & Hein, 2000), phylogenetic reconstruction (Posada & Crandall, 2002) and the identification of positively selected sites (Shriner et al., 2003). DENV-2 and DENV-3 isolates reported in the current study, showed no recombination event in any of the gene and in the entire ORF using the GARD method by employing TrN93 as the best fit nucleotide substitution bias model (Figure 3.5). The GARD analysis indicates no evidence of any breakpoints in the entire genomes of DENV-2 and DENV-3 Pakistani isolates, strengthens the fact that recombination has been rarely reported in positive-strand RNA viruses (Lai 1992). This is an important observation particularly towards the production of dengue vaccines because recombination is considered as one of the important hindrance in vaccine development like in case of HIV.

Isolates of DENV-2 and DENV-3 were screened for any evidence of selection pressure acting on any of the codon. The entire coding region of DENV-2 and the individual genes of DENV-3 were analyzed for the selection pressure by employing SLAC, FEL and REL methods keeping the significance level to \( p \leq 0.1 \) for SLAC and FEL and Bayes factor as \( p \geq 50 \) for REL. This is the first report on the selection pressure using a large dataset (\( n = 17 \)) on the entire ORF.
of the Pakistani cosmopolitan genotypes of DENV-2. The analysis indicated no codon under positive selection pressure keeping the set significance value in mind. However, codon 65 of the core gene \((C_{65})\), 155 of the PrM gene \((PrM_{155})\), 203 of the Envelope gene \((E_{203})\) and 254 of the NS1 gene \((NS1_{254})\) were identified to be under negative selection pressure among DENV-2 isolates. Apart from the negative selected sites, other sites revealed to be neutrally evolving (Table 3.5). The substitution rate observed in this study among amino acids residues of the ORF of all the 17 cosmopolitan DENV-2 sequences, was the highest between arginine (Arg) and lysine (Lys) residues, followed by valine (Val) and isoleucine (Ile) (Figure 3.6). Similarly no codon was found to under positive selection pressure in case of DENV-3. However two codons (94 and 477) of the E gene detected to be under negative selection pressure (Table 3.8). Collectively this suggests that these dengue viruses are under strong purifying selection pressure. In case of NS3 gene of DENV-3, we did not find any clue of positive or negative selection pressure acting on any of the codon which means that the gene is evolving neutrally. The low ratios of nonsynonymous to synonymous substitution indicate that dengue populations in nature are generally subject to strong selective constraints (Yang et al., 2000; Zanotto et al., 1996). Another aspect might be the week immune response to the virus as a result the virus evolving neutrally but this needs further elucidation in Pakistani perspective. From this entire scenario we assume that emergence and spreading of dengue in Pakistan is more due to permissive ecological conditions than natural selection.
CONCLUSION AND RECOMMENDATIONS

Dengue brings about disturbances economically to the affected persons in terms of hospital related costs as well as discontinuation from earning apart from physical pain. The collective ecumber of the infectious diseases such as malaria, AIDS and TB is almost equal to the losses claimed by dengue in Southeast Asia in terms of DALYs (disability-adjusted life years) (Gubler & Meltzer, 1999). In order to provide medical assistance to patients, to eradicate the vector and to launch public awareness programs, a number of funds have to be installed. Loss of revenue through reduced tourism is another indirect cost (WHO, 1997). The recent study conducted on dengue infection, suggests that annually 400 million people are infecting with the disease around the globe, out of which 100 million are clinically apparent (Bhatt et al., 2013). Dengue has become endemic to Pakistan. Favorable conditions like rainy season, humidity and temperature favors the dengue outbreak as we witnessed in the post monsoon period. In case of no epidemics, mosquitoes act as the potential reservoirs of dengue, although there are evidences that monkeys and some rodents may also act as reservoirs (Lavergne et al., 2009). The outbreaks started in Karachi (South) in 1994 and rocketed to the far-fetched corners of the country in the North as in 2014. All the four major serotypes of DENV were recorded during the 2011 outbreak in both genders with DENV-2 and DENV-3 the most dominant serotypes. Prevalence of all major types of DENV provides a reason for high morbidity and mortality. Consequently, this type of studies presented here becomes crucial to
determine the possibility of antibody-dependent enhancement (ADE), which usually leads to DSS. In ADE cross-reactive antibodies from previous infection not only fail to neutralize the current infecting serotype but also enhance the uptake of the virus into cells. As a result the viral burden is increased both in blood and in tissues that provokes an intense immune response leading to storm of vasoactive mediators (Halstead & O’Rourke, 1997; Libraty et al., 2002). Ultimately this phenomenon leads to DSS and Shock.

DENV-2 and DENV-3 isolates found in Swat district share closer genetic similarity (Table 3.3, Figures 3.8 & 3.9) with isolates from Lahore which indicates that the virus spread to Swat district from Lahore.

Currently, effective vector control (fogging for killing adult mosquitoes, larvicides to get rid of the larvae in the aquatic stage or destroying the source that is the breeding habitat) is the possible and only available procedure to tackle with the infection due to absence of vaccines and specific treatment.
FUTURE EXPERIMENTS:

Effect of primary and secondary dengue infections on dengue virus
Correlation of serotype and disease outcome
Role of ADE on viral load
To find out which Aedes species is involved in transmitting of disease from Lahore (hot environment) to Swat (relative cold)
Complete genome of DENV-1, DENV-3 and DENV-4
Tetravalent vaccine development
REFERENCES


cultured rabbit and human endothelial cells. *Infect. Immun.* 20: 776-781


survey team secretarial, 2nd official report. Karachi: Karachi Encephalitis Survey Team Secretariat 22-26


Talavera, D., A.M. Castillo, M.C. Dominguez, A.E. Gutierrez and I. Meza. 2004. IL8 release, tight junction and cytoskeleton dynamic reorganization conducive to permeability increase are induced by dengue virus infection of microvascular endothelial monolayers. J. Gen. Virol. 85: 1801-1813


Annex-I

QUESTIONNAIRE

S. No ____________ Date ____________

Name ____________ Age ____________

Gender ____________ Profession ______

Contact No. ________

Address __________________________________________________________
                                                                                                        
History __________________________________________________________
                                                                                                        
Symptoms _________________________________________________________
                                                                                                        
Serology _________________________________________________________
                                                                                                        

Signature ____________
Annex-II

DIAGNOSTIC PRIMERS

(Lanciotti et al, 1992)

DI 5’-TCAATATGCTGAAACGCGGAGGAACCG-3’
D2 5’-TTGCACCAACAGTCAATGTCTTCAGGTT-3’
TS1 5’-CGTCTCAGTGATCCGGGG-3’
TS2 5’-CGCCACAGGGCCATGAACAG-3’
TS3 5’-TAACATCATCATGAGACAGAGC-3’
TS4 5’-CTCTGTGTTCTTAAACAAGAGA-3’

(We designed)

D23Rv5  5’-TGCGCTCTCTGTGCCTGGAATGAT-3’
DAF  5’-TGTGAGCCCCGTCCAAGGACGTGA-3’
Annex-III

PRIMERS FOR WHOLE GENOME AMPLIFICATION OF DENV-2

- **D2RG** 5’-CTATGGCTTAATCCGACCTGAC-3’
- **D2FG** 5’-CGGCACGTGAGGCTTTGAAGA-3’
- **D2RF** 5’-TCACACTTTCTTCAGATG-3’
- **D2FF** 5’-CTCTGCTGACTCAAGT-3’
- **D2RE** 5’-GCCCATGTAGTCTCATCAT-3’
- **D2FE** 5’-GCAGCACAAGAGAGGAG-3’
- **D2RD** 5’-CCCATGTATATGTACTGTC-3’
- **D2FD** 5’-GGAGTTTACAAAGAAGGA-3’
- **D2RC** 5’-CTCCTTCCTTATTCCCTTC-3’
- **D2FC** 5’-ACAGAATGTTGCTGGGCGATC-3’
- **D2RB** 5’-TAAGTCACGGCCATACCTAT-3’
- **D2FB** 5’-TGGATAGGAATGAATTCA-3’
- **D2RA** 5’-TACTAGTGACACAGACTGA-3’
- **D2FA** 5’-CCATAATGGCAGCAATCTTG-3’
- **D2RT** 5’-CTTCCTCCTGAAACCTTC-3’
- **D2FT** 5’-TTAGAGAGCAGATCTCTG-3’

PRIMERS FOR THE NS3 GENE AMPLIFICATION OF DENV-2

- **D2NS3F1** 5’-TGGTACCTGTGGAAGTGA-3’
- **D2NS3R1** 5’-ATGGGTCTCTGCTTCCCG-3’

PRIMERS FOR THE INDIVIDUAL GENE AMPLIFICATION OF DENV-3

**Capsid gene**

- **D3CF** 5’-TGAAACACCAACGGAAGAAGCAGGGA-3’
- **D3CR** 5’-AGCAAGTGCTGCGGCAATATC-3’

**E Protein gene**

- **D3EF** 5’-ATGAGATGTGGGAGGATGGG-3’
- **D3ER** 5’-CCCTGTACCACAGCTCCCAG-3’

**NS3 gene**

- **D3ns3F1** 5’-TCCGTTTATGATGAGGACGTACC-3’
- **D3ns3R1** 5’-GGCTGCTCTCCCATCAGCCAC-3’
- **D3ns3F2** 5’-AGACCCAGCCAGTATAGCGGC-3’
- **D3ns3R2** 5’-TTTCTGACCACGCGGCAATG-3’
Annex-IV

DETAIL OF THE PCR METHODOLOGY

<table>
<thead>
<tr>
<th>Serotypes/ Gene</th>
<th>Fragment size</th>
<th>Primers</th>
<th>Thermocycling Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Denaturing</td>
</tr>
<tr>
<td>DENV2&amp;3</td>
<td>340 bp</td>
<td>D23Rv5, DAF</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td>DENV 1-4</td>
<td>511 bp</td>
<td>D1,D2</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td>DENV1</td>
<td>482 bp</td>
<td>D1, TS1</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td>DENV2</td>
<td>119 bp</td>
<td>D1, TS2</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td>DENV3</td>
<td>290 bp</td>
<td>D1, TS3</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td>DENV4</td>
<td>392 bp</td>
<td>D1, TS4</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>------</td>
</tr>
<tr>
<td>DENV-2 NS3 gene amplification</td>
<td></td>
<td></td>
<td>30-40 Cycles</td>
</tr>
<tr>
<td>NS3 gene</td>
<td>1003 bp</td>
<td>D2NS3F1, D2NS3R1</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td></td>
<td>1014 bp</td>
<td>D2NS3F2, D2NS3R2</td>
<td>95°C 5 min</td>
</tr>
<tr>
<td>Complete genome amplification of DENV-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DENV2</td>
<td>1.5 kb</td>
<td>D2FG,D2RG</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>1.5 kb</td>
<td>D2FF,D2RF</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>1.5 bp</td>
<td>D2FE,D2RE</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>1.3 kb</td>
<td>D2FD, D2RD</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>1.4 kb</td>
<td>D2FC,D2RC</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>1.3 kb</td>
<td>D2FB,D2RB</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>1.5 kb</td>
<td>D2FA,D2RA</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>915 bp</td>
<td>D2FT,D2RT</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td>DENV-3 genes amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capsid gene</td>
<td>342 bp</td>
<td>D3CF, D3CR</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td>E gene</td>
<td>1479 bp</td>
<td>D3EF, D3ER</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td>NS3 gene</td>
<td>933 bp</td>
<td>D3ns3F1, D3ns3R1</td>
<td>94°C 4 min</td>
</tr>
<tr>
<td></td>
<td>987bp</td>
<td>D3ns3F2, D3ns3R2</td>
<td>94°C 4 min</td>
</tr>
</tbody>
</table>
Annex-V

CODING SEQUENCE AND DEDUCED AMINO ACID OF THE D2/PK/A1/2011 DENV2 ISOLATE. * INDICATES THE TERMINATION CODON

65 atgaataaccaacgaaaaaaaggggagaaatagccctttcaatat
M N N Q R K K A R N T P F N M
110 ctgaacccggagaaaaaacccggctgtcaactgtcaacacagtgcaca
L K R E R N V S T V Q L T
115 aagagattctctattgaatgtctgagggagagaggaggtttgtaaaa
KR F S L G M L Q G R G P L K
200 ctgctcaagtgccttgggtgcccattccttcgctttcaataacccca
LF M A L V A F L R F L T I P
245 ccaacacccggatctaaaaagatgggaaaccatcataaaaaagtca
PT A G I L K R W G T I K S
290 aaggcacataagttctctgagggtctcagaaagatgggaagg
K A I N V L R G F R K E I G R
335 atgcgtgaacatcttgaataggagacgcagaaaccgcggcgtgatac
M L N I L N R R R T A G V I
380 atagtcgtgatttccacacagcagttgcttccattttaaccacacgc
I M L I P T A M A F H L T T R
425 aacgtagacaacacaacagttgcttcgaaacacagagaaagggaa
NGE P H M I V S K Q E K G K
470 agttctcttggttttaaacagaggatgtgtgtaacatgtagcccttc
SL F K T E D G V N M C T L
515 atggcagatgaccccttgtgaatctgtgtaagagacaacatctttatat
M A M D L G E L C E D T I Y
560 aactgtctctttctagccgcaaatgaaaccagagatatagattgt
NC P L L R Q N E P E D I D C
605 ttggtataaccccgcaacgttccatggttaactttgtggacatgcact
W C N A T S T W V T Y G T C T
650 gcacacagagaaaccagaaaggggaaachtgctgcactagcttc
AT G E H R R E K R S V A L V
695 ccacatgctgggaattgaggtcggacagcaactgaaacgttggtag
PC H V M G L E T R T E T W M
740 tcatcagaagggccttgaaacacgcccagaggattgaacacttg
S S E G A W K H A Q R I E T W
785 attcttgagacatcaggtttttacattaatgtgcagcaacatcttgca
IL R H P G F T I M A A I L A
830 tacacatagagaaacagcataattttccaaagagttctctgtatattc
c Y I G T T Y F Q R V L I F I
875 ttactgtaacagctgtcctcttcaaatgacaatgtcttttgatttggga
L L T A V A P S M T M R C I G
920 atatcaatatagagacttctgtggaggggttctcaggaggaagcttg
IS N R D F V E G V S G G S W
965 gttgacatgtctctttagaacatgtaagctctgtgtaacacagtgca
V D I V L E H G S C V T T M A
1010 aaaaaaacaacacattggactttaagactataaaaaacagagccc
K N K P T L D F E L I K T E A
1055 aagcacaactctgactttaagaggagtctctggactagggcaaacgctg
K Q P A T L R K Y C I E A K L
1100 accacacaacaacacagctctctctgagccacaaacagagggaaacc
T N T T A S R C P T Q G E P
1145 agcttaaatagaaacacacagcataaaaaaggttttgctgaaacacttcc
S L N E E Q D K R F V C K H S
1190 atggtacagagagagtgggaaatgattggtgattttggaaag
atgacagccctcaaaacacagactggaactgctgctcctgtaaatagaggggtggagacaggtggtggtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
A  W  K  V  S  C  T  I  L  A  V  S  V
3980 ccacctgccttaaacatctttcgacgacagaagc gagattggatacca
PL L LT SSQQKAKDWIP
4025 ttggcattgacgatcaaaaggtctctcaaataacagctttttctt
LA LT K K G L N P T A I F L
4070 acacaacttttaagaactaaacaagaaagagagtggccactaat
TT L SR TNK K R SWPLN
4115 gaggctatcataagctgcggagtgtgacacattttgactagtgtct
E AIMAVGMVSILASS
4160 ctctttgagaatgacattcccatacgaggccacattagtggctgga
LLKNDIMPMTGPLVAG
4205 gggcttctcactgttgttgactgtcactggaagatcgcggcac
GLLTVCYVLTGRSAD
4250 ctggaactagagagagccagggcgcttggagatccagacgca
LELERAADVRWEDQA
4305 gaaatcaggaagattcctaatctgtaataaacaatattcgaa
EISSPSPILOTSISE
4340 gatggcagcagttcgataaaaaataagagagagaacaacactg
DGSMSIKNEEEEQTL
4385 accatactctatgagaacaggtgctgtgatcaggaacttttc
TILIRTGLLVISGLF
4430 cccgtatcaaatcacaatccagccgagcagcgctgtacgtgga
PVSIPIIITAALAWYLWE
4475 gttgagaagacaagggcttgaggtatgttgggtgtctcctccccc
VKKQRAGVLWDVPSP
4520 ccacctgtgggaagaagcctggaagatggctgtaaagatc
PVPVKAELDGAYRI
4565 aagccagaagggatttctttgatcttcccagatccggagctggagtt
KQKGLGYSQIGAV
4610 takaagagaagacaatcttcaccaataatgtcagctcaccagttgt
YKEGTFHMTMWHVTRG
4655 gctgatttctatgacacaaagaggaagatttgacaccatactagggcg
AVLMHKGKRIEPSWA
4700 gatgtcaggaagagacctaatctgatgagagagctggctgtaaagatc
DVRKDLISYGGGWKL
4745 gaggagataggaagagagagagatcgcggcttctgcacattta
EGEWKEGEEVQLAL
4790 gaggctcgaggaagataccagagcctgccaacaaaacaacccggtccttt
EPGKNPRAVQTGPGL
4835 ttaagactacacccggaaaccataggcctgtgctgtccctagacatttt
FKTNGTIGAVSLODF
4880 ttctctggaagtgcaggtctacacaaaaagggaaaspGTSGPSPIVDKKKGK
4925 gctgattttctctagttgatattggtgctgacaggaagagttgaca
VVGGLYNNGVVRTSGT
4970 tagttgatgtgcataagcaggagagtgaaccacatgtggcgcac
YVSAIAQTEKSIEDN
5015 ccagagattggaagacagatatttcctgaaaaaagagtagcaccatt
PEIDEFFKRRKRTLTI
5060 atggacctctccacccagaggcggaaagagacgaagagatatccttcca
MDLHPPAGKTKRYP
5105 gcagatgttagagagccataaaaaacgaggcttaagaacattaac
AIVREAIKRGRLTRLI
5150 ctggccccacactagatgctgagcagctggaagatggaggacttt
LAPTRYVVAAMEMEAL
5205 agagagctggccagaattacgtagctgatatttataatgtgtctagccacatttt
HTGREIVDLMCATFP
5240 accatcagggccagaaatcagctgatttaataatgtgtcagcacaatttt
TMRLSPVRVVPNYNL
I M D E A H F T D P A S I A
ARGYINSTREVMGEAA
GGAGATTCATGACAGCCACCTCCCGGGAAGACAGACAGCATTT
GIFMTATTPGSRDPF
CCCTCAAGCAATGCACCAATCATGATGTTACATTCAATTTCA
PQSNAPIMDEEREIP
GAGAACGTCTGGAACTCTGGACATGTGTGGTACATTCAATTTCA
ERSWNSGHEWVTDKF
GGGAAGACTGGTTTGGTTTTTCGCAATAAAAGCAGTTATGAT
GTKVTWVFVPSIKAGND
ATAGCAGCTGCTGGAGAAAAATGGAAAGAATGGTATACATC
IAACLRKNGGKKVQIOL
AGTGAAGACCTTGTTAGTTGCTGACACTCATTTGCAAATGGTCT
SRKTFDSEYIKTRTN
GATTTGGACCTTGTGCTGACACTCATTTGCAAATGGTCT
DWDFTVTTTDISEMGAPAACCTCAAGGCTGAAAGGTTTATAGACCCAGGGCGCTGATGAAA
NFKAERVIDPRRCMK
CCAGCTCATACATACGGACGGTGGAAGCGGTGATTCTGCGAGGA
PVILTGDGEERVIAG
CCATGCCAGTGACCCACCTCTGACTGACCCAGCAAGAGGAGA
PMVPVTHSSAAQRGR
ATAGGAAGAAATCCAAAAATGAAATAGCAGTACTATACATATG
IGRNPKNENDQYIYM
GGGAACCTCTGGAAATAATGAGAAGACTGCACACTGGAAGAAG
GEPLENDEDCAHWKE
gccagatgctttttagacacacacacagctggaagaatcatt
AKMLLDNINTPEGII
CCAGCATGCTTTGGACACGCAGCGTGAAAGGTTGGATGCTATTTGAT
PSMFEPERKEKVDADI
GGTTGAATACCGCTTAAGAGGAGAAGCGAGGAAACCTTTGCGAT
GEYRLRGEARKTFV
CTATGAGAAGAAGGAGCCTACGCTGCTGCGCTCACAGATG
LMRRGDLPVWLAYRV
GCACCAGGGCATTAATCTGACAGACAAGGTGGTGTGGTAT
AAEGINYADRRWCFD
GGTAATCCGAAACAACAACTTGGAAAGAATGTGGAGTGGAA
GGIKNQILEENVVE
ATCTGGACAAAGAAGGGGAAAGGAAAGAAGCTGAACACACAGATG
IWTKEGERKLLKPRW
CTGTAGCTCGGATCTACTGACCTCCACTGCGCTCAAGAATTCT
LDARIYSDPALKEF
AAAGATTTGTCAGCTGCAAGAAGAAATCCCTGACCTGGAATTATAT
KEFAAARKSSLTLLNLI
ACAGAATATGGTAGGCTCACATTTTATGACTCGAAGAAGCAAG
TEMGRLPFTMTQKAR
AACGCTCGGACAATCTGCGACTCAGGCGTGAAGACAGGC
NALDNLAVLHTAAEG
GGAAAGGCTGTAACACATCTGCTGTAAGACTGCCAGAAGCTGCTG
GRAYNHALSELPETL
GAGACATTGTCTTTTGGCTGACACTTTTGCCACAGTCCAGGGCGGA
ETLLLTTLLATVTTGG
ATCTTTTATTTTCGATGGCGCAAAGTATAGGGAATATGAC
IFLFMLMSGKGGIKM
CTGGAGATGTGCTGCTAATACCGCCAGGATTTCTCTATGTAAC
LGMCCITASILLWY
GCACAATAACAGCCACACTGCGATAAGCGCTTCAATAATACGTGAG
AQIQPHWIAASILE
TTTTTCTCATAGTCTTTGCCTCATTCCAGACAGCCAGAAGAAGAAGA
8075 ataaaggttcttaaccctatacatggccctcaagttcttaacccatagctgctcatatactacagtga
IKVLNPYMPSVIEKMI
8120 gaaacgtcgcaagaaaatagggagaaatgctgctggtaagagaatccc
ETLQRKYGGAALVRNP
8165 ctctcaagaattcaccacatgagatgtattggatatccatagctct
LSRNSTHEMBWVSNA
8210 tccgggaacatagtcgatccagagatgtattttcaagaggatgg
SGNIVSVVNMISSRM
8255 atcaatatgattaatgagacataagaaagcgccacatagaccca
INRFTMRHKKATYDP
8300 gatgtgactcctggagagcggacccgcaacatcggatcgaaggt
dvdglsgtrnigies
8345 gaaacaccacacctagacataatattgggtatccaatgct
ETPNLDIIGKRIEKI
8390 aaacaagcatagcatagcatcagcgcaactatgacccaaagaccca
KEHETSRHDQDHP
8435 tataaaacttgggcctctactgagtacatgagagggccacataatg
yktwashgsyetskq
8480 ggatcgcagctcatccacacggtgatgagttagtactgtcgaca
gsasstvngvvrltl
8525 aaaccttggagcgtctctcctatgtgcagacagatgcgaacata
tkpwdvimpmtqmat
8570 gacacgcaccccattggcagacggtgttttttcagagagaggtg
dttpgqqrqrichp
8615 gacacgagacccagacgagacgagacagagacgagacagactata
tgkwahtgkaugkka
8660 daaatcagggcagatigtgcttggagagaaactagggagagaaag
kitaewlwkkelgkk
8705 acacctagagatgtgcacagagagagatctgcaagagagagatggaga
tprmcctefarkvr
8750 agcaatgcagccctaggtgccatattcactgagagaagatgggg
snaaliqftdnekw
8795 aatttgccagctgaggtttctgtaagacagacagaattttgggaactaca
tgsravedserflew
8840 gttgacagagaaagagaaatctccatcttgaaagagaagtagtgagacaca
vdkernlhlegkcet
8885 tgtgtgtcaacacatgtggaaagagagagagaagagataggggaac
cvynmmgkrekkiige
8930 ttgcggcaagaaaaaggacgcagagcatgtgtactatgtgctgcttt
fgkgakgsraiwymwl
8975 ggagccagctgttttaggttgagaagccctagagattttcagagaga
garflfealgflnen
9020 gcactgttttctccagagagaacattttctctgagcggagtaagagga
dhdfsrenslsgveg
9065 gaagggctgcacaagctgggtttacatcttaagagaagttcgacagag
eglhklgyilorvsk
9110 aaagaagagagagcatactgacgcgcagacacagagcagatgggat
kegamyaddtagwbd
9155 acaaatcactacataaggacccataaataaattgaagagatgttaaca
tritledlkneeimvt
9200 aaccatatgtggagagaaagacacaaagaaactttgtctggacacatcc
nhmeghekklaeaif
9245 aatataacgtccaaaaaaaggtgtgtcggctgcaaaagacccaca
kltynknkvrvqrt
9290 ccagagggcagagataattatatctcagagagacacaagaga
prgtvmdiiiirsdqrr
9335 ggtgtgtgaattttggtattttttcccttttacttacaatgtgcaattttggtcagagacca
atsggagccccactaatcagacagatggagggaggagtcttt
mgalirqmemegvfps
9425 aaaggcatccagcatctgcagccacagacagaaaggttgcagtgcaaa

145

*INDICATES THE TERMINATION CODON.
ENEVKLSIMTDIGK
2675 atcatgcaagcaggaagaaaacagatcctgcgctccagccacactgag
IMQAGKRSRLRPQPTE
2720 ctgaaatatctggaaaaaagttgggcaaaagcaatatgtcttccc
LKYSWKTKWGKAKMMLS
2765 acagacgctctataccacagacctttttcattagtgcccccgaaca
TEPHNQTFLIDGPE
2810 gcacaatgtcccaacacaaacagagttggaactcactacaatttt
AQCPNTNRANWSLEV
2855 gaagactatgctttttgggtatatccacacaaacatagagtctgaag
EDYGFVFFTTTNIWLK
2900 ttgaaagaagggcaggtgtgcttttctactcactaanaactcagttca
LKERQDVFCDSKLMS
2945 gcggccataaaaaagacaacagagcccctcatgccgatatggttat
AAIKDNRADVADMGY
2990 tggataagaaagccacatcataaatcagactaggaagagaacctct
WIESALNDTWKIEKA
3035 ttcttttatcagatattaagaatttgccacttgccaaatccacacact
SFIEVKSCHWPKSHT
3080 ctcctggagtaacgaggtctgtaaaagttgagatgataatccaaag
LWSNGVLESEMIPK
3125 aatattttgtggaccagtatcgcaacaaacataacagcaggctat
NFAGPVSQHNYRPGY
3170 cacacacaaagcgaggctcctgtgcctaggcaagtttgagatg
HTQTAGPWHLGKLEM
3215 gacctttgttttttgaaagagagaggctgtggtgtgactaggaac
DFDFCEGGTTTVVVED
3260 tcggaatagagacccctttcataagaaacactactgtccttgga
CGNRGPSLRTTASG
3305 aaactcataacagatggtctgccgcatctgcacattttcagga
KLITEWCCRSCTLP
3350 ctcagatacagacggcaggtgatctgatcaggaagatac
LRYRGEDGCWYGMEI
3395 agaccattgaaaggaagagagagactttgtgtaacattcctttgtc
RPLKEKEENLVNSLV
3440 acacgccgccacagcatgagctatcactccccactagccatttgc
TAGHGQIDNFSLGV
3485 ggaatgtcagctgctctgaggaagatctcagagcccgattgaga
GMALFEEMLRTRVG
3530 acaaaaaacagttaaaatgctttttttcatagcatag
TKHAVLLVAVSFMTL
3575 atcacaagggacatgtttttttccagctggggaagatgttt
tagitaryGNMSFRDLGRVMV
3620 atgtgagggcgcctatcgagcgtgacatagatgtgctggctag
MVGAAMTDDIGMGT
3665 tatccttgcatctactagcagcccttaaagattgaggacaaacttttca
TAGHGLYLLAFAFKVRRPTFA
3710 gctgtgacgtctttttggaagaggtaagctctcctgaaagattagtagatg
AGLRRKLTSKELMM
3755 accacataagatcgattctccctccagacacccacccaaacagag
TTIGIVLLLSQSTTP
3800 actatactgtaagctgtagctgtgggtctttttgggcatagatt
TILELTDALALGMMI
3845 ctcagatagtggaaaaaatctgatacttggacttggatgctgatc
LKVIRNMEKYQLAVT
3890 atcatgtctactttatatgctttttccaaatcagatgatattcacaataa
IMAILCVPNAVILQN
3935 gcagatggaggctgcacatcacttgccacgctgtgctccgtatatcc
AWKVSCSTILAAAVSVS
3980 ccactgtctttacatcctgcaagaaagaggattgtgataccca
PLLTLSSQQKADWIP
ttggcattgacgatcaatccccacagctattttctct
t L A L T I K G L N P A I F L
acacccctttcaagaactaacaagaagaagagttgacgcactaaat
T T L S R T N K K R S W P L N
qaggctatcatgtcgaatcggatggtgacatttttagctagttct
t E A I M A V G M V S I L A S S
ctttgagaatgacattccccatagcagaccattagttgctgga
L L K N D I P M T G P L V A G
gggtcttcctacgtttgtaagctctagaaagatgacgccac
G L L T V C Y V L T G R S A D
tgtaagacagagagcgcgccgcgttagatgtaggaagatgcagca
L E L E A D V R W E D Q A
gaatatcaggaagatgttcataattcgtcaataacaatacgagaa
E I G S S P I L S I T I S E
gatcagcagcatgtcgatatacctgacgagccttttc
D G S M S I K N E E E E Q T L
accatactcattagaacaggtgtctgtgtgatcatacctgacgcttttc
T I L I R T G L L V I S G L F
cctgatcaataacatcagccagcaagcatagttgacgcga
P V S I P I T A A A W Y L W E
tgtaagaaacacggggtctgagatattggtgagttctctaccccc
V K K Q R A G V L W D V P S P
ccaccctgtggaaagcgcagctgagggcgcttacgaatc
P P V G K A E L D G A Y R I
aagcagaaaggggtcttctgtgatctcccaagctggagctgtgggt
K A K G I L G Y S Q I G A G V
tacaaagaaagacaattccacacatgtgactgctgaggctggt
t K E G T F H T M W H V T R G
gctgtcttaatgcacaagggagagaagattgaacacattcatggccg
A V L M H K G K R I E P S W A
gatgtaagggaaagacattatatcgtatggagaggctggaagcta
d V R K D L I S Y G G W K L
qaaggagaatggaggaagagaagttcagctgctgagtaa
E G E W K E G E V Q V L A L
agcctgggaagaaatccaagagccgctccaaacaaacacccggtcttt
E P G K N P R A V Q T K P G L
tttaagactaacaccggaacccataagccgctgtgtccctcagacttt
F K T N T G T I G A V S L D F
tctctgtgaaagctgagattctctcagtgcacaaaaagggaga
S P G T S G S P I V D K K G K
gtcgtggtgctctctgtagaatggtgctcataagggctggaaca
V V G L Y G N G V V T R S G T
atgtgtagtgccataagccagctgaaagacagtcgacat
Y V S A I A Q T E K S I E D N
ccagagattgaggacgtatcttttcgaaaaaaagattgaccatt
P E I E D D I F R K K R L T I
atgacactttccaccagccagcggaagagaagagatatacttcca
M D L H P G A G K T K R Y L P
gcaatagtttagagggccataaaaaacagggcttaagaacattaaatc
A I V R E A I K R G L R T L I
tctggccccactagagctgctgcgcagttgaaatgtgagagctttt
L A P T R V V A A E M E E A L
agagacttccataataagataacaaacccagcactacagctgag
R G L P I R Y Q T P A I R A E
cacacagcgccccagagatcgtggtatattaatgtgtctgacaatcttt
H T G R E I V D L M C H A T F
accatatagcagctttccacagacagcctagcagcttaattacatgt
t M R L L S P V R V P N Y N L
atcattaggtgaaagccccattttccacagacagcagataqca
I I M D E A H F T D P A S I A
gctcagaggtacatttcattcagagttagaaatgtgtaagcagct
9515  atcagtggagatgattgttgatgaaaccccttagatgacaggttt
9560  gcaagcgccttaacagctctaaatgacatgggaaaggttaggaaa
9605  gacatacaacaatgggaaccccaaaggagtgacaccggagaaca
9650  diqweptgdndtwtdcaatctccacacttcatgaggttaatcatgaa
9700  qpschhhf helimmk
9745  gacggtcgtggtgctctgtagtgtgcatgtagaaggccaaaggtgaaactg
9790  dgrvlnvvaecnqdel
9835  attgtagacgctgaatttccccaggagccggtggtctttgcgga
9880  igrarisqgagwsler
9925  gagagcgcctgtctggggaacctttcaagaggatggaacgactggaca
9970  diqwegpsrwrgsnhr
10015  gacggtcgtggtgctctgtagtgtgcatgtagaaggccaaaggtgaaactg
10060  dgrvlnvvaecnqdel
ORF AND ENCODED AMINO ACID PROFILE OF D2/PK/SWAT-1/2013.

*INDICATES TERMINATION CODON

65 atgatataaccaacgaaaaaagcgagaaataacgcctttcaaatatg
M N N Q R K K A R N T P F N M
110 ctgaaaccccgagaaaccccgagtctcaactgtcacaacagctgacca
L K R E R N R V S T V Q L T
155 aagagatctctcacttggaatatgctgacaggacagagccctgtgaaaa
K R F S L G M L Q G R G P L K
200 ctgttcatggtccttggtgccatactcttctttcttaacaatccccca
L F M A L V A F L R F L T I P
245 ccaacgcctgggtactaazaagaagatggaacacatctaaaaagtcgat
P T A G I L K R W G T I K S
290 aagggccatcaatgtcttgaggttcaggaacagcgacaaccgacaggctgtac
K A I N V L R G F R K E I G R
335 atgtcgtcaatcattggaataggacagggcataatgtcaccctc
c M L N I L N R R R R T A G V I
380 atcatgctgtatttccacacacggtgcgttccatttaaccacgc
I M L I P T A M A F H L T T R
425 aacggagaacaccacatagcgtcagcaaacaggaagagggaa
N G E P H M I V S K Q E K G K
470 agttctctgttttaaaacagaggtgatgtgtgaacatagtgacaccc
c S L L F K T E D G V N M C T L
515 atggcccatgtgcccttggtgaactgtgaagacacaatcatttat
c M A M D L G E L C E D T I T Y
560 aacgctctcttttcagcagcagatagcagcataatgattgt
c N C P L R Q N E P E D I D C
605 tgtgtgaacgcaccgtcctaggttaacatttggccatgcact
WC N A T S T W V T Y G T C T
650 gccacagagaacacagagaagggaaaaagatcagtcgcaacctgtc
AT G E H R R E K R S V A L V
695 ccacatgtgggaatggagtttgacagacagaaacagtgtgagatg
AT G M G L E T R T E T W M
740 tcatcgaaaggggctctgtgaaacccgcagaggtttcagcattcgg
S S E G A W K H A Q R I E T W
785 acttggacacatccaggttttacacatagtgccacagcatacttgac
c I L R H P G F T I M A A I L A
830 tacacatcaggaacagcataatctccccaaaggtctgtatattcctc
y T I G T T Y F Q R V L I F I
875 ttactgacagctgtctctcttcaatgcacaatgctttatattgga
L L T A V A P S M T M R C I G
920 atatcataatagcagattcgtggaagggattttcaggaggaagctgg
I S N R D F V E G V S G G S W
965 gttgacatagtctctgagaacatcggaagctgtgtgacacaatgtgc
V D I V L E H G S C V T T T M A
1010 aaaaacaaaaacccatgtgacttttgaactgtaaacacagaaagcc
K N K P T L D F E L I K T E A
1055 aagccacccctgactctcaggaaggtgtactgtatagagggaaacgctg
K Q P A T L R K Y C I E A K L
1100 accaacaacaaacacatcttccagtcgccccacacacaggggaaacc
T N T T T A S R C P T Q G E P
1145 agcttacatagtctctgagaacatcggaagctgtgtgacacaatgtgca
S L N E Q D K R F V K H S
1190 atgtctacagaggtgggaaaatgtgatggtggtatatttgggaag
M V D R G W G N G C G L F G K
1235 ggaggcatttggtgaccttgctatgtccacatcgaagaaaaatag

155
PLLLTTSSQQKADWP
4025 ttggcattgacgatcaaaggtctcaatccaacagcttatttctcta
4070 ttggcattgacgatcaaaggtctcaatccaacagcttatttctctta
4115 attttccta
4160 acaaccctttcaagaactaacaagaaaaggtgctacaatggtcttaac
4205 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4250 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4300 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4345 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4390 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4435 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4480 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4525 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4570 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4615 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4660 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4705 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4750 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4800 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4845 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4890 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4935 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
4980 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5025 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5070 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5115 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5160 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5205 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5250 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5295 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5340 gaggctatcgcagttgctgctcttgctcttgctgttaagagacgctggac
5375  gctagagggatacacatttcactccgagactagagaaatgggtgaagcgactcr  
  A R G Y I S T R V E M G E A A
5420  gggattttatgacacccactccccggaagcagagaccatatt
  C G F M T A T P P G S R D P F
5465  cctcaaacagatgcaccaatcatgtagaagagagggaaatccct
  P Q S N A P I M D E E R E I P
5510  gaacggtcgtgaacactctgacatagtggttacagatttccaac
  E R S W N S G H E W V T D F K
5555  gggagagctgttttgtgttccccgacataaaacaggttaatgat
  G K T V W F V P S I K A G N D
5600  atagagcctgcctgagaaaaaatggaagaaatgtatcaaacct
  I A A C L R K N G K K V I Q L
5645  agtagaaaagacatttttgtgatcctgaataacattagccaggcacaat
  S R K T F D S E Y I K T R T N
5690  gattgggaccttttggtgtacccgtacacttggaaataggggtct
  D W F V V T T D I S E M G A
5735  aacactacggctgaaggtttagatagccacacacttagaacaan
  N F K A E R V I D P R R C M K
5780  ccagctcataactacgccaggttagagagcgggtgtatcctgccagga
  P V I L T D G E R V I L A G
5825  ccacatgcaggctacacatctagtgcagcacaaagaagggaga
  P M P V T H S S A A Q R R G R
5870  ataggaagaaatccaanaatgaccagcatatattagcttag
  I G R N P K N E N D Q Y I Y M
5915  gggagacacccctggaagaaatgtagagactgcaacactggaagaa
  G E P L E N D E D C A H W K E
5960  gccaagatgctttttgataacacataacacacgcctgaaaggaatcatt
  A K M L L D N I N T P E G I I
6005  cccagcatgtttcagaccagagctgaaaaagttgatgtacctttagat
  P S M F E P E R E K V D A I D
6050  ggtgataaacgctttaagagggagaaagctgaaacccagattttgtgag
  G E Y R L R G E A R K T F V D
6095  ctaatgagagagagagacccctactggctctgcccatacagatgtg
  L M R R G D L P V W L A Y R V
6140  gcagccgagggcattaaactctgacagacaagaggttggttttgtat
  A A E G I N Y A D R R W C F D
6185  ggaatcagacaacacaaatatgttagagagctgaaagggaaatgtaggtgaa
  G I K N N Q I L E E N V E V E
6230  atctggaaaaaagaggggaagagagaaacgtaacccagaatgggg
  I W T K E G E R K K L K P R W
6275  cttcgatgtcctggatctactctgtacccacctggccctgaaatatcc
  L D A R I Y S D P L A L K E F
6320  aaagagtttgcagctgagaaaaaacctcctgtacccatgatttatact
  K E F A A G R K S L T L N L I
6365  acagaaatgaggtagctcccaacttttattgactcagaaagcaag
  T E M G R L P T F M T Q K A R
6410  aacgcaactggcaacatgacatgctcttcacagggctgagagcacc
  N A L D N L A V L H T A E A G
6455  ggagggcggtcataaacagctcctctaggtacacgtcagagactctg
  G R A Y N H A L S E L P E T L
6500  gagaacattgtctttttgggtagactccctttttggcagacacttggcagga
  E T L L L L L T L L A T V T G G
6545  accttttttatattttagatgtgacagaaaggtatatgggaaatgaacc
  I F L F L M S G K G I G K M T
6590  ctggagatgtctgtccataacacggccagttgtctctatgtgtactgat
  L G M C C I I T A S I L L W Y
6635  gcacaaatacacgccaacacagtgagatacgctctctataatacctagag
  A Q I Q P H W IA A S I I L E
6680  ttttttttcatatgctttgctcaattccagacagcagaaagcaagaa
  F F L I V L L I P E P E K Q R
6725  acaccccaagataaaaccatctagtcgagttgtgcttagccatcctc
GAAACGCTGCAAAGAAATACGGAGACTTTTGTTGAGGAATCCG
TCTCAAGAAGATTTCCACACGATGATGATTGGTGATCAATGCT
TCCGGAAACATAGTTGCACTACAGTGAACATGATTTCAGAGATTG
SNIIVSSVNMISRM
ATCAATAGATTACAGAGACATAAGAGGCACACATGCAGACACCA
INRFTRMRHKKATYDP
GATGTTGACCTCGGAAAGCGGAAACCAGCTGGATCAGAAAGTTG
DVDSGTSRNGIES
GAAACAAACCTAGACATATTGGGAAGAAATAGAAATA
ETPNLDIIIGKRIEKI
AAACAAGGCGATGAGGACATCCGCGACATGAGCAAGAAGACACCA
KCQETHTSRHYDQDHP
TATAAAACTGGGCTTACTGTCGACTGAACA
YKTYAWAGSYESCT
GGATCAGCATCATCCACGTTGAATGGAGTATTAGACTGCTGACA
GSASSVTNVGVVRLLT
AACCTTGGGACGTACATCCCTATGTTGACACAGATGGCAAACATA
KPDVDMVTQMAMTM
GACACGACCCCATTTGGGAAACAGCTGTTTTCCAAGAGAAAGTG
DDTPFGQRVFKEKV
GACACGAGAACCAGAAGCGAAGGCGCAGAAGACTATAATG
DTRTQEPKEGTKKLM
AAAATACGCGGCAATGCTGCTGGAAGAAGACTGGGAAAGAAAG
KITATEWLWKELGKKK
ACACCTAGAGTTGCACTAGAAGAATTCGCAAGAAGGTGAGA
TPRMCTREEFARKVR
AGCAATCGACCCCTAGTGCCATATTCACTGTGAAGAACCAGTG
SNAAALGAIFTDENKW
AAGTGGCGACGTGGAGCTGTTGGAAGACATAGATTTGGGAACTA
KASAREAVEDSRFWEL
GTTGACAAGGAAAGGAAATCTCCATTTGGGAAGAAGTTGAGACA
VDKERNLHEGKCET
TGTTGTATAACATGATGGGAAGAGAGAAGAAGATAGGGAA
CVYNMGMKREKKGIE
TTCCGGCAAGAAGAGGCGGCAAGGCTATGTTACATGTTGGCT
FGKAKGSRAIWYMWL
GGACACGCGTTCTTAGTTGGAACCTAGTTTTGCAATGAA
GARFLEFEALGFLENE
GATCACTGTGTTTTCCAGAGAAGACCTTTCTTGACCGGATGTAAGGA
DHWFSSRENSLSGVEG
GAAAGGCTCACAAGCTGTTACATCTTCAAGAAGATGGCAG
ELHLKHGLYILREVSK
AAAGAAGGAGGCAATGACGCCAGACGACACGAGATGGGAT
KEGMAMYADDTAGWD
AACCATATGGAAGGGAACACAAGAAACCTTGTGGGACCTTCC
NHMEGEHKKLAAEIF
AAATACGTACCCAAAGACAGGGTGCTGCTGGAACACCA
KLTYNKVKVRVRQRT
CCAAAGGCGACATGAGATACATATTATACGGAAGAGAACCAGAA
PRGTVMIDISRDRDQ
GGATGTTACATGAGCTGATCATTACGGCTGGCCATCTTACAT
GSVQVVTGLEGFTTN
ATGGAGCAGAAAATACAGACATGAGGAAGGAGGAGGAGTCTTT
MGALIRQMEGEVF
AAAGCACTACCGCATCTGGACACGCAGAAGAGGATTCGACTGCAA
KGIQHLTATEEAVVQ
GACTGTTAGCAGAAGATGGGCGGTGAGGTTATCAAGATGGC
D W L A R V G R E L S R M A
9515 atcagtggagagatgtgtgttgtaaaacccttagatgacaggttt
I S G D D C V V K P L D D R F
9560 gcaagcgctttaacagctcttaaagctgagaacaggaagtttagaaa
A S A L T A L N D M G K V R K
9605 gacatacaacaatgggaaccttcaagagatgggaacgactggaca
D I O Q W E P S R G W N D W T
9650 caagtgcctttttctgttacacaccatctcatgtaaatcatgaaa
Q V P F C S H H F H E L I M K
9695 gacggtcgtgtgtgctgtgctgtcagtctgatgaaacccaaagatgacgtg
D G R V L V V P C R N Q D E L
9740 attggtagagcttagttccacggagcagcgtgcttttgcgg
I G R A R I S Q G A G W S L R
9785 gacgacggcctgtctgggaagctttacgccccaaatgtggaagctgg
E T A C L G K S Y A Q M W S L
9830 atgtacctcccacagctgtgcctggctggacaaagttctatt
M Y F H R R D L R L A A N A I
9875 tgctcggcagtttcttacacattggtttccgacaagtcgaacacc
C S A V P S H W V P T S R T T
9920 tggtctatcatacgcggacatggtctcaacagaaagacatctg
W S I H A G H E W M T T E D M
9965 ctgatctgcttgaatgcgggttctttgattgatggagaacccccatgtg
L T V W N R V W I Q E N P W M
10010 gaagacaaacactcagtggaactatggaggaattcctcatcttg
E D K T P V E S W E E I P Y L
10055 ggaaaaagaggaagccaccttggtgctggcattgtagggtaaca
G K R E D Q W C G S L I G T
10100 agcaggggccacctgggaagaaacctcataacaagtaaaaac
S R A W A R I Q T A I N Q
10145 gttagatcctttaaataggagtggagatatcaacagagactacatgcca
V R S L I G S E E Y T D Y M P
10190 tccatgaaaaagattctggagagagaagaggcagggctttg
S M K R F G R E E E E A G V L
10235 tggtag 10240
W *
Annex-VIII

CODING SEQUENCE AND AMINO ACID PROFILE OF DENV-2, ISOLATE D2/PK/SWAT-02/2013

65 atgataaaccacacgaaaaagcgagaaatagcgcctttcaatatg
MNNQKRKKARMNTPFNM
110 ctgaaaccgcagagaaaccgcgtgtcaactgtaagccacagtgcaca
LKRERNRSTVQLT
155 aagagattctctacatgaaagctgcaggagcaggacgctgtgaa
KRFSLGMLQGRGPLK
200 ctgttcatgtgctttggtgcaattctctttcttaacactccc
LMALVIAMFLRLTIP
245 ccaaccacccggtactataaaagaatggggaacacatcaaaaaag\atca
PTAGILKRWGTIKKS
290 aagggcatacatgtcttgagaggtctcaggaaagagattggaag\gg
KAINVLRGFRIEGR
335 atgtgctgaaccatcctctgaaagagctgcagcagacagctgtgac\tc
MLNILNRRRTAGVI
380 atcatgtttaccacactcgggaaggtctgcttttaacactccc
IMLIPTAMAFHLTTTR
425 aacggaagaacccacatgactgtcagcgaacacaggggaag\gaaa
NGEPMIVSKQEGK
470 agttccctgttttaaaacagaggtatgtgtaagactatgctgcac\ctc
SLFLKTEDGVNMCTL
515 atggccagctgactctgttctgaagctgtgaagacactattcatctat\tat
MAMDLGELCEDITTY
560 aacgtcctctcttcagcagaaatagcagagataatagttgt
NCPLLRQNEPEDIDC
605 tgggttaaccgcagtcaccaacctagggaatctctctctctctcaac\ctgcact
WCNATSTWVTYGTCT
650 gccacagagagaacagaagggaaagaatcagtagcagtcactgtc\tc
ATGEHRREKRSVALV
695 ccacatgtttgggaatggaattgagcacagactgaaacgtag\gtg
PHTVGMGLETRTETWIM
740 tcaatccagagggcctggttaaaacagccgagaggataatagctgg\gg
SSEGAWKHAQRTEIW
785 atctgtgagacatcaccgagttttaccatataatggacagaaactt\ggca
ILRHPGFTIMAAILA
830 tacacatgagaaacagacatattctctactcagttcagaa\ggatcact
YITGTTFQFRVLIFI
875 ttactgtacagcgtgtcctcttcatgtacaatgcttttgattg\gga
LTLTAVPSMTMRICG
920atatcaatataagacactctgtgaaagggtttcagggaggaagtgg\tg
ISNRFDEVGVSAGSW
965 tgtacatagtctcttagaagctgtgaagacgtgtgacacaactg\atgca
VDIVLEHGCSTTTMA
1010 aaaaacaacccacatttggacttggataaaacacagaagcc\nKNKPTLDFELIKTEA
1055 aagcaacattgctgtcaacagatgtgagactggtgaagag\ggctg
KQPATLKRKCYICEAKL
1100 accaacaacaccacatcctgtgaccacacagggagaaaccc\ntnttarsrcptqgep
1145 agctataatagaaagacagggaagagtttggtctgcacacattcc\slneneqdkrfvchks
1190 atgtagcagagaggtagggaaatgtgtagttattttggaag\g
MVDRGWNGCGLFGK
1235 ggaggcattgtgaacctgtgcctagttcacaatgcagaaaa\catg

163
G G I V T C A M F T C K K N M
1280 gaagggaaatggtgcaaccgaaacacctggaatcacacatggatcgaagggcttag
E G K I V Q P E N L Y T I V
1325 gtaacacctacattcaggggaagagatacggaaacacctggaatcaccattgtgcaaccggaagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatgcgcagttggcgagatgtgggtgtgttgctgcaagagggctatggtagatg
5375 gctagaggatacaactttcaactggagtagaaagatggtagaagcgcagct  
ARG Y IST R V E M GEE A 
5420 gggattttccatgcagccacactccccgggaagcagagactccttttt 
G F M T A T P P G S R D P F 
5465 cctcaagcaatgcaccaact atgatgaagagagggaaatccct 
P Q S N A P I M D E E R E I P 
5510 gaacggtcatggactatcggaaagctcagatgtggttcacatttcc 
E S W N S G H E W V T D F K 
5555 gggagactgttttggatgttgttcccgagcatataaagacgtaaatgat 
G K T V W F V P S I K A G N D 
5600 atacagcgcttcctgagaaataatgaaagacgataaactc 
I A A C L R K N G K K V I Q L 
5645 agtaggaagacctttatgtatatcagataattatagcaaagccagg 
S R K T F D S E Y I K T R T N 
5690 gattgaggacttgttgtgcagacctagcactatcatcagaatggtgct  
D W F V V T T D I S E M G A 
5735 aacatcagtgaagaggttattagacccccaggctgcactag  
N F K A E R V I D P R R C M K 
5780 ccagctcataactacggagctgatagagggctcgccagga 
P V I L T D G E R V I L A G 
5825 ccagctccagtgaccaactctagtgtcagcaacaaagagggaga  
P M P V T H S S A A Q R R G R 
5870 ataggaagaatccaaaaatatgagaggactaatacagcactaatc  
I G R N P K N E N D Q Y I Y M 
5915 gggaaccttcctggaattaatgatgaagactgcgacacttgagaaagaa 
G E P L E N D E D C A H W K E 
5960 ggcaagagatggcttttttagacaacacacagcctgaggaatcatt 
A K M L L D N I N T P E G I I 
6005 ccagctacatggaggacagctggattttgagttgtctggcagga 
P S M F E P E R E K V D A I D 
6050 ggtgaatatccctttaagagggagagaggaaaaattttgtgtgat 
G E Y R L R G E A R K T F V D 
6095 ctataggaagagggagacctccccagcagatcagagggaaaaattttgat 
L M R R G D L P V W L A Y R V 
6140 gcagccgaaagcttaactatactagcagagaagagttgtttttagat 
A A E G I N Y A D R R W C F D 
6185 ggatcaagagcaaccaatcttagagaagaaatgtggaatgagaa 
G I K N N Q I L E E N V E V 
6230 atctgagaaaaagggaaaggaaagaaactggacacacagag 
A W T K E G G R K K L K Y P R W 
6275 ctgctagtcggtattactctcgtacccacctgcagcgtctgaagaatcc 
L D A R I Y S D P L A L K E F 
6320 aaagatggttcgagggagaaaaatccctctgacacctgtatttattac 
K E F A A G R K S L T L N L I 
6365 acagagaaagtaggtagcttccacacatatatataagacgagaggg 
T E M G R L P T F M T Q K A R 
6410 aacgcacgtggacacattcagcttcctgcatagctggactgaagcagcg  
A A C G C A L N L A V L H T A E A G 
6455 ggagggcgtcacaactccatctgcttgagacacactctcct 
G R A Y N H A S E L P E T L 
6500 gagacattgcttttttggacacaccttttggccacagcttcggcggaa  
E T L L L L T L L A T V T G G 
6545 atcttttatatttttttagaagtgcaaaagtataagggaaatagacc  
I F L F L M S G K G I G K M T 
6590 cttggaatagtgtgtgtatcataacccggcagagttttctctatgttac 
L G M C C I I T A S I L L W Y 
6635 gcacaaataacccacagctgtgaagcagctttcatatatactgcagag  
A Q I Q P H W I A A S I I L E 
6680 ttttttcattagcttttctccatcttcacgcccagccagaaaatgaaaga  
F F L I V L L I P E E P E K Q R 
6725 acaccccaagataaaccataatttagcgtgcttcagccatcctcct
D W L A R V G R E R L S R M A

9515 atcagtggagatgattgttggtaaacccccttagatgacagtttt
I S G D D C V V K P L D D R F

9560 gcaagcgctttacagcttcaatagtgggaaaggttgtagaaaa
A S A L T A L N D M G K V R K

9605 gacatacaacaatgggaaccttcaagaggatgggaacgtggaca
D I Q Q W E P S R G W N D W T

9650 caagtgcctttctgttccacaccatctccatgtagtaatcatgaaa
Q V P F C S H H F H E L I M K

9695 gacggtcgtgtgtcgtgtagtctccatgtagaaaccaagatgaaaatg
D G R V L V V P C R N Q D E L

9740 attgttagagctcgaatttcccagggagccgtgtgtcttttgccg
I G R A R I S Q G A G W S L R

9785 gagacgggctgtcttggaagttcctacgcacaaatcgtagagacctg
E T A C L G K S Y A Q M W S L

9830 atgtacttccccacagctgacctcaggtgcacagcaatgtcatt
M Y F H R R D L R L A A N A I

9875 tgctcggccagttccatcactattggttcgacaagtcgaacacc
C S A V P S H W V P T S R T T

9920 ttggttccacacgcggacatgaattgtagaacaacagagacgtag
W S I H A G H E W M T T E D M

9965 ctgacagcgttgggaacaggtgggtgattcagaaacaccagatgtag
L T V W N R V W I Q E N P W M

10010 gaagacaaacatcccaagtgaatcatggggaatcccaacctcttg
E D K T P V E S W E E I P Y L

10055 gggaaaaagagaagaccatgttgctggctcaattggagtttaaca
G K R E D Q W C G S L I G L T

10100 aqcagggccacctggcaggaacatccaaacacgaataaaacca
S R A T W A R N I Q T A I N Q

10145 gttagatccttaataagcagtgggaatcagagatcactagcag
V R S L I G S E E Y T D Y M P

10190 tccatgaaaaagattcggaaagagaagggcagcggtctctg
S M K R F G R E E E E A G V L

10235 tggtag 10240

W *