2015

Patterns of Selection Amongst Dengue Virus Serotypes and Efficacy of Computational Epitope Prediction Applications

April Christina Hall South
University of South Carolina - Columbia

PATTERNS OF SELECTION AMONGST DENGUE VIRUS SEROTYPES AND EFFICACY OF COMPUTATIONAL EPITOPE PREDICTION APPLICATIONS

by

April Christina Hall South

Bachelor of Science
Clemson University, 2004

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biological Sciences

College of Arts and Sciences

University of South Carolina

2015

Accepted by:

Robert Friedman, Major Professor

Austin L. Hughes, Committee Member

Joseph M. Quattro, Committee Member

Deanna S. Smith, Committee Member

Jijun Tang, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

DEDICATION

To my parents, Rita and Vaughn Hall, for their continued generosity and support;

To my brother, Johnathon Hall, for his encouragement;

To my son, Robert Truedell South, for his astonishing ability to bring joy to almost every moment of my life;

To my husband, Sean South, for giving me all the time in the world and for being the greatest defender of my abilities;

To some of my best friends, especially Melissa P. Davis, Kimberly R. Shorter, and Grandon T. Wilson, for sharing their struggles and for always believing in me.

ACKNOWLEDGEMENTS

There are many people who are in some way responsible for the completion of this dissertation and I am so incredibly thankful for every last one of them. To my advisor, Dr. Robert Friedman, for his continued support and advice. To my committee members: Dr. Austin L. Hughes for his assistance and daily guidance; to Dr. Joseph M. Quattro for his enthusiasm and encouragement as well as for accepting the position of Committee Chair; to Dr. Deanna S. Smith for her optimism and zeal; and to Dr. Jijun Tang for his time and consideration. This goal would have been unobtainable without all of your support.

I would also like to thank the following collaborators, colleagues, professors and friends including Terry Acton, Dr. Priyasma Bhoumik, Dr. Amy Bodycott, Jason Bodycott, Sue Carstensen, Dr. Margaret Cirtain, Sandra Clifton, Milan Dhameja, Dr. Trey Franklin, Dr. Megan Gamble, Brittany Garvin, Rebecca Givens, Margaret Goulding, Shadow Gulledge, Zachary Haberman, Dr. Jerry Hilbish, Dr. Stephanie Irausquin, Dr. Elina Levina, Bryan Mathis, Chase Nelson, Dr. Amanda Polson, Saravanan Rajabojan, Dr. Thomas Reeves, Robert Robertson, Brittany Snee, Dr. Richard Vogt, Dr. David Wethey, and Dr. Sarah Woodin as well as all of the biology department staff members and all of my previous teaching assistants for the Biology 102 labs. I owe all of you my deepest gratitude for your continued encouragement.

ABSTRACT

Dengue virus (DENV) is a mosquito-borne virus of global concern that leads to infections with symptoms ranging from high fevers and headaches to death. Current estimates place almost half the world's population living in areas that are at-risk for infection by dengue and that area continues to expand over time. There are four distinct serotypes of dengue (DENV1, DENV2, DENV3, and DENV4). Disease severity from dengue infection is significantly worse if there is a secondary infection by a heterotypic serotype. The problems of increased severity with secondary infection as well as host immune response evasion by the virus itself have made vaccine development especially difficult.

Analysis of complete polyprotein sequences of all four DENV serotypes revealed that all serotypes are currently subject to ongoing purifying selection and have experienced past purifying selection. However, several noteworthy exceptions within some genes of varying serotypes suggest conflicting evolutionary pressures. Most notably, the envelope (E) gene of DENV4 showed nucleotide polymorphism patterns suggestive of positive selection or relaxed purifying selection in its epitope regions. The E protein interacts directly with the host cell and is necessary for viral entry suggesting that its polymorphism pattern is indicative of evasion from immune system recognition. Both the NS1 gene of DENV2 and NS3 gene of DENV3, which are critical in viral replication, exhibited significantly higher medians of nonsynonymous gene diversity in

their epitope regions implying escape mutations in these proteins as well. Significant conflicting ratios of π_N (mean nonsynonymous nucleotide diversity) to π_S (mean synonymous nucleotide diversity) between pairs of serotypes indicate that each individual serotype is under conflicting evolutionary pressures and help to explain the severity of secondary heterotypic infections.

A comparison of three epitope prediction applications using datasets of both DENV genomes coupled with known cytotoxic T lymphocytes (CTL) DENV epitopes revealed significant differences in RANK and sensitivity measures of the programs. As these programs are used to predict which epitopes should be further studied in the creation of subunit vaccines, it was apparent that applications containing a combination of several prediction methods in concert are much more efficient than programs involving only one approach.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

FOREWORD

The use of bioinformatics in immunology has become one of the most rapidly evolving fields in biology today. Recent advances in technology coupled with appreciable interest in preservation of human life have created a tremendous increase in the number of viral genomes available for analysis and the tools with which to study them. Using a combination of statistical and computational techniques, it is possible to quantify changes between individual viral genomes, infer phylogenetic and ancestral relationships, predict host immune response, and complete a plethora of other analyses (Volz et al., 2013). As viral infections are typically more difficult to treat in comparison to bacterial infections and have higher rates of mutation, these analyses may be able to assist in the treatment and prevention of viral illness (Volz et al., 2013).

The major histocompatibility complex (MHC) found in all vertebrates is the primary adaptive immune response to viral infection (Andersen et al., 2006; Dimmock et al., 2007; Hughes and Yeager, 1998). Nucleated somatic cells express MHC class I molecules which consist of proteins that either identify the cell as belonging solely to the host (self) or as being bound to a foreign peptide (nonself) (Andersen et al., 2006; Dimmock et al., 2007; Hughes and Yeager, 1998). This presentation of a foreign peptide by the MHC on an antigen presenting cell (APC) is called an epitope (Dimmock et al., 2007). Cytotoxic T lymphocytes (CTL) have a T cell receptor (TCR) that recognizes these

epitopes and elicit several responses (Andersen et al., 2006; Dimmock et al., 2007). These responses include proliferation of T cells, differentiation into memory or effector T cells, and the eradication of the infected cell (Andersen et al., 2006; Dimmock et al., 2007).

Considering viruses, RNA viruses are particularly poised to evade immune system recognition in order to replicate and proliferate as much as possible. RNA viruses have higher mutation rates, estimated rates of mutation being on the order of 10^{-3} to 10^{-5} per 10Kb genome, than those of DNA viruses (Domingo and Holland, 1997; Drake et al., 1998; Hughes and Hughes, 2007). This is most likely due to the fact that RNA polymerase, RNA replicase, and RNA transcriptase are not nearly as efficient as their DNA counterparts resulting in many errors when replication occurs (Domingo and Holland, 1997).

Utilizing sequence data of RNA viruses, patterns in nucleotide polymorphism, particularly at the population level, can give insight to which mechanisms of selection may be acting on various proteins and known epitope regions within the virus' genome (Hughes and Hughes, 2007; Lambkin et al., 1994; Moore et al., 2002; Rodriguez-Roche and Gould, 2013; Suzuki, 2006). One of the most prevalent patterns is the mean number of synonymous substitution per synonymous site (d_S) exceeding the mean number of nonsynonymous substitutions per nonsynonymous site (d_N) as most nonsynonymous changes can disrupt function of the resulting protein and are typically deleterious therefore this pattern is evidence of purifying selection (Graur and Li, 2000; Kimura, 1977). There are several studies showing evidence of purifying selection in viral

populations (Holmes, 2003; Hughes and Hughes, 2007). While many nonsynonymous mutations are deleterious, this does not apply to all nonsynonymous mutations and it is possible for some that are only slightly deleterious to persist in populations (Ohta, 1973). Epitope regions of viruses would benefit immensely from retaining slightly deleterious variants leading researchers to search for patterns of positive selection (Bedi et al., 2013; Murray et al., 2013; Ross et al., 2002).

A particular RNA virus of increasing international concern is the mosquito-borne Dengue virus (DENV). DENV is both positive sense and single-stranded RNA with a genome that is a little over 10Kb in length (not including any untranslated regions) that encodes for one large polyprotein (Fields et al., 2013). This polyprotein is eventually cleaved into eleven products. Three are structural proteins: AnchC (capsid), prM (precursor/membrane), and E (envelope); while eight are nonstructural and play several different roles: NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5 (Fields et al., 2013). Of these proteins, E, NS3 and NS5 have been shown to be particularly important in viral entry into the cell and successful replication (Iglesias et al., 2014). The E protein is necessary for viral entry into host cells while NS3 is particularly integral in RNA replication (Iglesias et al., 2014). NS5, while also essential to RNA replication, has the ability to block interferon (Iglesias et al., 2014).

With a primary DENV infection the epitope portions of the virus itself are recognized and consequently bound by TCR eventually leading to clearance of the infection after a short illness that is termed dengue fever typically with some symptoms resembling those to influenza (Dimmock et al., 2007, p. 197). Dengue fever conveys

symptoms that are similar to those experienced with a case of influenza. However, there is a much more drastic response to secondary DENV infections if the subsequent infection is that of a heterotypic serotype (World Health Organization, 2012a). The illness that results from these secondary infections is termed dengue hemorrhagic fever or severe dengue and will be accompanied by several serious symptoms (World Health Organization, 2012a). If these symptoms are left untreated, the illness may progress to dengue shock syndrome which ultimately results in death in up to 20% of cases (World Health Organization, 2012a).

There are three prominent hypotheses regarding the mechanism behind the severity of secondary heterologous infection that patterns in the genome of nucleotide and gene diversity may help support or refute. Suppose CTLs generated during primary infection resurface when a secondary serotype is presented. However, these CTLs have a difficult time discerning cells that are infected with the secondary serotype as its sequence differs ever so slightly from the initial infection. These CTLs may not bind as well or even at all to peptides presented by class I MHC in the newly infected cells (Vaughan et al., 2010). This theory is termed original antigenic sin and it is thought to lead to higher viral loads (Halstead et al., 1983; Vaughan et al., 2010). Another very similar theory is known as altered peptide ligand (APL) antagonism.

APL antagonism occurs when there are slight changes in the secondary infection's RNA sequence that alter the epitope presented to TCR by an infected cell (Jameson and Bevan, 1995). CTLs will still bind to the antigen presenting cell but will not elicit an immune response. Instead, the interaction may cause the T-cell to become

entirely functionally inactive or allow for only partial activation which also allows for a higher viral load (Andersen et al., 2006; Kalergis and Nathenson, 2000; Sloan-lancaster and Allen, 1996).

Antibody-dependent enhancement is somewhat similar to the two previous concepts except that in this phenomenon replace CTLs that do not recognize heterotypic serotypes well with antibodies that are produced during a primary infection (Kliks et al., 1989; Vaughan et al., 2010; Whitehead et al., 2007). The antibodies from the primary infection bind to the new DENV particles and remain bound although are unable to neutralize the virus (Kliks et al., 1989; Vaughan et al., 2010; Whitehead et al., 2007). This complex then attaches to fragment, crystallizable regions of monocytes, allowing the DENV particles to enter the white blood cell and replicate also creating an increased viral load (Kliks et al., 1989; Vaughan et al., 2010; Whitehead et al., 2007).

As the prevention and control of DENV is of increasing concern, creation of a vaccination that would generate immunity for all four serotypes without causing adverse reactions is of extreme interest. The discovery of vaccinations has changed widely since Jenner's initial vaccine to smallpox (Hilleman, 2000). Employing genomes and computer algorithms along with known immunological data has brought about a more efficient process in vaccine creation by allowing for identification of possible epitopes that can be used to assist the production of effective subunit vaccines (De Gregorio and Rappuoli, 2012; Delany et al., 2014; Gilbert, 2011; He et al., 2010; Hilleman, 2000; Yang and Yu, 2009). Epitope prediction applications use various

methods in order to predict which will have the most significant response and range widely in their efficacy (De Gregorio and Rappuoli, 2012; Yang and Yu, 2009).

Two separate studies were performed to compare gene and nucleotide diversity within and between DENV serotypes and to evaluate the performance of three epitope prediction applications. The first study was an analysis of a large dataset of 1615 whole DENV genomes. This study examined the patterns of both nucleotide polymorphism and gene diversity in all eleven proteins of all four serotypes. Patterns within CTL epitope regions were compared to those in non-epitope regions in order to detect immune evasion between serotypes as secondary heterotypic infections tend to be more severe and elicit little to no immune response. The second study employed portions of the larger DENV dataset and known epitopes from the initial study to compare scores given to predicted epitopes from three epitope programs in both RANK measures and sensitivity percentages to determine which was most efficient.

CHAPTER 2

DIFFERENCES IN GENE DIVERSITY AND NUCLEOTIDE POLYMORPHISM IN EPITOPE REGIONS OF HETEROLOGOUS DENGUE VIRUS SEROTYPES[1]

[1] South, A.C.H., Friedman, R., and A.L. Hughes. To be submitted to *Infection, Genetics, and Evolution*.

2.1. Introduction

The Dengue virus (DENV) is a positive-sense, single-stranded RNA virus classified in the *Flavivirus* genus. This genus contains several detrimental viruses including Yellow Fever virus for which the genus is named and the West Nile virus among others (Kuno et al., 1998). The majority of the members of the *Flavivirus* genus are transmitted via arthropod vectors where DENV is of immense concern as it is currently the most rapidly-spreading mosquito-borne virus in the world (World Health Organization, 2012b).

DENV infection results in the onset of dengue fever (DF) which conveys flu-like symptoms and is typically remedied over a period of several days without much medical intervention, however, DENV infection may also produce a more critical disease known as dengue hemorrhagic fever (DHF) that is characterized by petechial hemorrhaging, high fever, and thrombocytopenia (Rodriguez-Roche and Gould, 2013; World Health Organization, 2012b). The severity of DHF is distinguished by plasma leakage. If not successfully treated when signs of circulatory failure appear DHF can progress to dengue shock syndrome (DSS) which can cause respiratory distress and organ impairment due to excessive plasma leakage and fluid accumulation and can culminate in death (World Health Organization, 2012a).

The DENV genome codes for one single polyprotein that is eventually cleaved into 11 major products, three of which are primarily structural: AnchC (capsid), prM (precursor/membrane), and E (envelope); while the other eight are nonstructural: NS1, NS2A, NS2B, NS3, NS4A, 2K (sometimes denoted as a fragment of NS4A), NS4B, and NS5 (Fields et al., 2013). There are four distinct serotypes of DENV denoted DENV1, DENV2,

DENV3, and DENV4 that differ by approximately 30-35%. Secondary infection with a heterologous serotype is more likely to cause DHF/DSS (Rodriguez-Roche and Gould, 2013; World Health Organization, 2012b). Due to the difficulty of a secondary infection causing increased likelihood for disease, a vaccine for DENV is not currently available although there are recently completed trials of a potentially successful tetravalent vaccine (Durbin et al., 2013). There are several theories as to how heterotypic infection results in increased disease severity rather than inferring a stronger host immune response (Goncalvez et al., 2007; Halstead, 1979; Halstead et al., 1983; Kliks et al., 1989; Whitehead et al., 2007).

Typically, adaptive immune responses to viruses elicit class I major histocompatibility complex (MHC) molecules in the infected cells of vertebrates to bind foreign peptides and present them extracellularly (Hughes and Yeager, 1998). These presented peptides are known as epitopes (Dimmock et al., 2007). Epitopes are recognized by T cell receptors (TCR) of cytotoxic T lymphocytes (CTL) which initiate the destruction of infected cells (Hughes and Yeager, 1998). In the case of DENV, this immune response works well with the initial infection yet poses significant health issues with a secondary heterotypic infection. As immune response to viral infection is increasingly important to vaccine development, the epitope glycoproteins presented by class I MHC molecules that are recognized by CTL are therefore of the utmost importance in predicting epitopes and designing vaccines particularly *in silico* (Patronov and Doytchinova, 2013).

Patterns in nucleotide polymorphism of RNA viruses can be utilized to postulate how evolution may be acting on these epitope regions such as escape from immune system recognition or shed light on differences between serotypes, either of which may lead to more knowledge regarding the difficulties associated with secondary heterologous infection (Hughes and Hughes, 2007; Lambkin et al., 1994; Moore et al., 2002; Rodriguez-Roche and Gould, 2013; Suzuki, 2006). In protein-coding nucleotide sequence data the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous (amino acid-altering) substitutions per nonsynonymous site (d_N) are valuable indicators for revealing the mechanism of selection (Graur and Li, 2000; Hartl, 2000; Hughes, 1999; Kimura, 1994; Nei, 1975). The prevalent pattern displayed in the majority of these coding regions is that of mean d_S exceeding mean d_N This pattern is indicative of past purifying selection and is due to the fact that coding for a different amino acid with a nonsynonymous mutation can lead to disruptions in the creation of the resulting protein and are typically deleterious in nature (Graur and Li, 2000; Kimura, 1977). Not all nonsynonymous mutations are strongly deleterious, therefore slightly deleterious variants may persist within populations (Ohta, 1973). Patterns of positive selection may be found in epitope regions as slightly deleterious variants could be selected for in attempts to evade immune system recognition.

As the severity of a secondary heterotypic DENV infection is of increasing global concern, the objective of the present study was to analyze patterns of both gene diversity and nucleotide polymorphism in known CTL epitope regions as well as to

further examine the similarities and differences within and amongst complete genomes of all four serotypes of DENV.

2.2. Methods

2.2.1 Sequences and epitopes analyzed

Complete unaligned genome sequences of 1615 DENV isolates (640 of DENV1, 598 of DENV2, 341 of DENV3 and 36 of DENV4) were obtained from the Dengue Virus Database available at http://www.denguedb.org/index.asp (Viral Bioinformatics Resource Center, 2009). Artemis was utilized to arrange sequences into the correct reading frame (Rutherford et al., 2000). Sequences were aligned at the amino acid level using ClustalW (Thompson et al., 1994) within the MEGA6 program's Alignment Explorer (Tamura et al., 2013). Aligned protein sequences were superimposed back onto their nucleotide frames. Sequences with premature stop codons or indeterminate nucleotides were excluded from analysis (for a list of all utilized GenBank accession numbers see Table A.1).

Known epitopes were obtained from the Immune Epitope Database and Analysis Resource available at http://www.iedb.org/ (Vita et al., 2010). Epitopes used were limited to those presented by MHC Class I and found only in humans. B cell epitopes and those listed as epitope containing regions were also excluded as CTL epitopes have been noted as especially important in immunity in several studies of dengue pathogenesis (Piazza et al., 2014; Vaughan et al., 2010; Weiskopf et al., 2013; Yauch et al., 2009). 525 of the remaining epitopes were 100% conserved within at least one of

the sequences within our DENV dataset (Table A.2). It is of note that some epitopes were located either overlapping or entirely within other epitopes.

2.2.2 Ancestral reconstruction

A neighbor-joining (NJ) tree (Saitou and Nei, 1987) of all 1615 DENV isolates was constructed using the PAUP* program, version 4.0 (Swofford, 2003), rooted with the DENV4 serotype. The bootstrap method (Felsenstein, 1985) was used with 1000 bootstrap replicates in order to verify statistical significance of internal branches (tree not shown). The NJ tree was imported into MEGA6 (Tamura et al., 2013) where the last common ancestor for each of the four serotypes was reconstructed using maximum parsimony (Eck and Dayhoff, 1966; Fitch, 1969). Each of the reconstructed ancestors was then inspected for presence of the 525 aforementioned epitopes that were 100% conserved within our dataset.

2.2.3 Nucleotide and gene diversity

Epitope and non-epitope domains were differentiated in MEGA6 (Tamura et al., 2013) for each serotype separately as not all CTL epitopes were found in all four serotypes. These domains were then utilized in estimating d_S and d_N values by the Pamilo, Bianchi, and Li method (Li, 1993; Pamilo and Bianchi, 1993) within the MEGA6 software (Tamura et al., 2013). These values were used to then determine both the synonymous nucleotide diversity (π_S), the mean of all d_S values, and the nonsynonymous nucleotide diversity (π_N), the mean of all d_N values. Statistical methods that did not rely on model dependence were used in order to avoid the assumptions that model-based methods have regarding various mechanisms of evolution (Hughes et al., 2006). Ratios

of π_N to π_S were computed for each serotype separately. These ratios were then compared by examining the differences between pairs of serotypes utilizing the known CTL epitope as the unit.

Gene diversity was estimated using PolyAna (Knapp et al., 2011) for polymorphic synonymous and nonsynonymous nucleotide sites for epitope and non-epitope domains of each serotype separately using the following formula:

$$1 - \sum_{i=1}^{n} x_i^2$$

where *n* is the number of alleles and x_i is the population frequency of the *i*th allele (Nei, 1987, p. 177). Any site at which nucleotide polymorphism could be considered either synonymous or nonsynonymous was excluded from analysis. Non-parametric methods were utilized in the analysis of gene diversity as it was not normally distributed (Hollander et al., 2013). The Minitab statistical software package version 13.1 was used for all statistical analyses ("Minitab 13.1 Statistical Software," 2000).

2.3. Results

2.3.1 Nucleotide diversity

In comparisons of nucleotide diversity, π_S was greater than π_N in CTL epitope regions for all four serotypes, significantly greater in 40 of 43 comparisons (Z-tests, two-tailed, $P<0.05$ in 3 comparisons, $P<0.01$ in 5 comparisons, $P<0.001$ in 32 comparisons), except in the NS2B protein of DENV1 as there were no known epitopes found (Tables 2.1-2.4). In non-CTL epitope regions of all four serotypes π_S was also greater than π_N,

significantly greater in 42 of 44 comparisons (Z-tests, two-tailed, P<0.05 in 5 comparisons, P<0.01 in 1 comparisons, P<0.001 in 36 comparisons, Tables 2.1-2.4).

When comparing the nucleotide diversity of CTL epitope regions and non-CTL epitope regions, only the E protein in DENV4 had a significantly higher π_N value (Z-test, two-tailed, P<0.05) in the CTL epitope regions (Fig. 2.1a). The NS1 protein of DENV1 , NS2B proteins of DENV3 and DENV4, and the NS3 protein in DENV4 all had significantly higher π_N values in the non-CTL epitope regions (Fig. 2.1b-e). For synonymous diversity, the NS2A and 2K proteins had significantly higher π_S values in CTL epitope regions of DENV1 while the NS2A protein in DENV4 showed significantly higher π_S values in the non-CTL epitope regions (Fig. 2.2a-c).

Comparisons of differences of the ratios of π_N to π_S between pairs of serotypes yielded significant results between the majority of serotype pairs. Initial testing showed significant differences between four of the six comparisons (Sign test, P<0.001) with the exception of comparisons between DENV1 and DENV4 and between DENV2 and DENV3. Further examination showed significance in five of the six comparisons (Wilcoxon signed rank test, P<0.001) excluding only the pair containing both DENV2 and DENV3.

2.3.2 Gene diversity

Median synonymous gene diversity was higher in polymorphic sites of known CTL epitope and non-CTL epitope regions of all proteins of DENV1 and DENV2 (Fig. A.1-A.2) except for the CTL epitope regions of NS2B and 2K in DENV1. This was due to the fact that there were no known CTL epitopes found in NS2B and there were not any nonsynonymous polymorphic sites found in 2K in DENV1. The synonymous gene

diversity was statistically significant in the majority of the proteins. Similar results were noted for DENV3 (Fig. A.3) except in the case of the non-CTL epitope region of AnchC. The median nonsynonymous polymorphic gene diversity was higher in AnchC than its corresponding synonymous value however it was not of any significance.

DENV4 expressed quite different results with the median nonsynonymous polymorphic gene diversity exceeding the median synonymous polymorphic gene diversity in the CTL epitope regions of AnchC, prM, and E while also showing similar patterns in the non-CTL epitope regions of NS3 and 2K, none of which were of any significance. The median gene diversities were equal in the NS1 protein. There were not any nonsynonymous polymorphic sites found in the NS2B protein of DENV4. The remainder of the DENV4 proteins had greater median synonymous gene diversity but few were statistically significant (Fig. A.4)

Comparisons of median gene diversity between CTL epitope and non-CTL epitope regions were of interest in the AnchC, NS2A, and NS3 proteins of DENV1 (Fig. 2.3a-c). Median synonymous gene diversity was significantly higher in CTL epitope regions of AnchC and NS2A and in non-CTL epitope regions of NS3. The median nonsynonymous gene diversity in CTL epitope regions of NS1 was notably higher in DENV2 (Fig. 2.3d). DENV3 had higher median nonsynonymous gene diversity in both non-CTL epitope regions of AnchC and CTL epitope regions of NS3 (Fig. 2.3e-f). DENV4 had higher median synonymous gene diversity in non-CTL epitope regions of AnchC and NS2A (Fig. 2.3g-h) while the non-CTL epitope region of NS3 had higher nonsynonymous gene diversity (Fig. 2.3i).

2.3.3 CTL epitope presence in reconstructed ancestral sequences

The last common ancestor for each serotype was reconstructed using the maximum-parsimony method assuming the NJ tree. In the initial dataset, 19 known CTL epitopes were found in all four DENV serotypes while only 13 (68%) were found across all four of the reconstructed ancestral sequences. 163 known CTL epitopes were found in the DENV1 dataset and 129 (79%) in the reconstructed DENV1 ancestor. 267 known CTL epitopes were found in the DENV2 dataset and 191 (72%) in the reconstructed DENV1 ancestor. 196 known CTL epitopes were found in the DENV3 dataset and 151 (77%) in the reconstructed DENV3 ancestor. 147 known CTL epitopes were found in the DENV4 dataset and 129 (88%) in the reconstructed DENV4 ancestor.

2.4. Discussion

DENV has been studied significantly in an effort to stem the outbreaks that affect millions of individuals globally. The pursuit for understanding of the virus and how it is evolving over time may aide in disease prevention. An analysis of complete genomes of all four DENV serotypes showed several significant patterns in both nucleotide and gene diversity. Evidence of strong past purifying selection was shown by the extremely significant pattern of $\pi_S > \pi_N$ both within CTL epitope and non-CTL epitope regions for all four serotypes of DENV.

Similar indications of purifying selection were noted with the patterns in gene diversity. In all proteins of both DENV1 and DENV2, median synonymous gene diversity exceeded median nonsynonymous gene diversity in both CTL epitope and non-CTL epitope regions. The majority of the proteins in DENV3 follow the same pattern

regarding gene diversity reinforcing that purifying selection is definitely an effective force in the evolution of DENV just as it is in many other RNA viruses (Hughes and Hughes, 2007). DENV4 follows somewhat dissimilar patterns when it comes to gene diversity compared to the other three serotypes. Although none of the results for proteins or regions of DENV4 were significant, the differences in the patterns suggest that this serotype is under different evolutionary pressures.

Since secondary heterotypic infection tends to be one of the main concerns with DENV infection, it was pertinent to test whether serotypes that appeared related phylogenetically were indeed similar in their nucleotide diversities in CTL epitope regions. It was found that when comparing ratios of π_N to π_S for each pair of serotypes that there were significant differences (Wilcoxon signed rank test, $P<0.001$) between the median ratios of each pair across the entire genome excluding only the combination of DENV2 and DENV3. This result was not expected based on phylogenies of DENV as DENV1 and DENV3 cluster together and would therefore typically share similar ratios if selection was occurring comparably in both (Weaver and Vasilakis, 2009). This further signifies that each serotype is not subject to the same mechanisms of evolution in CTL epitope regions which aides in explaining the difficulties clearing a secondary heterologous infection.

As escape from immune recognition would be beneficial to the virus itself, one would expect to see either some relaxation of purifying selection or positive selection showing an increased number of nonsynonymous changes in CTL epitope regions versus non-CTL epitope regions as these regions are not readily identified by MHC. The only

protein that showed signs of either relaxation of purifying selection or positive selection when examining nucleotide diversity was the E protein in the DENV4 serotype. The E protein also had greater nonsynonymous gene diversity in CTL epitope regions as compared to non-CTL epitope regions of DENV4 but this result was not significant. It is possible that positive selection could be acting on the E protein as it is involved directly in virus attachment with the host cell and is more prevalent to host cell recognition (Hughes and Hughes, 2007; Idrees and Ashfaq, 2012; Iglesias et al., 2014; Piontkivska and Hughes, 2006). If this is indeed the case, it is surprising that only the E protein showed significance in DENV4 and not in other serotypes as well.

Another remarkable result was in the NS3 protein of DENV3. Median nonsynonymous gene diversity in CTL epitope regions of this protein significantly exceeded median nonsynonymous gene diversity in non-CTL epitope regions. This anomaly may also be due to positive selection or relaxation of purifying selection. Similar results have been noted for the NS3 protein in Hepatitis C, another RNA virus in the *Flaviviridae* family (Irausquin and Hughes, 2010). The NS3 protein is vital to the replication of the viral RNA and is the pivotal point for replication (Lescar and Lok, 2014).

Ancestral reconstructions of the last common ancestor for each serotype showed the presence of known CTL epitopes that were found in our dataset. While 100% of the CTL epitopes were not conserved in any ancestor, 72% was the minimum value for any of the serotypes. Since the ancestors do contain a large percentage of the known CTL epitopes, those that are found in both ancestors and contemporary

sequences may be optimal candidates for vaccine creation. However, ancestral reconstructions may lack escape mutations that the more divergent genomes may have now accumulated.

Analysis of the complete genomes of all four DENV serotypes in this study revealed that all serotypes are currently subject to ongoing purifying selection; however, there are several noteworthy exceptions within certain proteins of each serotype suggesting conflicting evolutionary pressures. As the mechanism for severity of secondary heterotypic infections is still under speculation, perhaps the significance conflicting ratios of π_N to π_S between pairs of serotypes when compared to the viral phylogeny can be of some explanation. Future studies could focus on particular proteins rather than the virus as a whole, particularly E and NS3, as E is instrumental in viral entry into the cell while NS3 is necessary for reproduction.

Table 2.1. Synonymous (π_S) and non-synonymsous (π_N) nucleotide diversity (± S.E.) in CTL epitope and non-CTL epitope regions of the 11 proteins of DENV1 (Z-tests, two-tailed, rejected the hypotheses that mean π_S = mean π_N for CTL epitope regions *P<0.05, **P<0.01, ***P<0.001; Z-tests, two-tailed, rejected the hypothesis that mean π_S = mean π_N for non-CTL epitope regions *P<0.05, **P<0.01, ***P<0.001).

DENV1		CTL Epitope Region		Non-CTL Epitope Region	
Protein	**n**	**π_S ± S.E.**	**π_N ± S.E.**	**π_S ± S.E.**	**π_N ± S.E.**
AnchC	640	0.132722 ± 0.045187	0.010144 ± 0.005066**	0.048414 ± 0.014175	0.012967 ± 0.004603*
preM	640	0.333525 ± 0.111373	0.002030 ± 0.001848**	0.155951 ± 0.021165	0.007147 ± 0.002189***
E	640	0.151728 ± 0.023791	0.005945 ± 0.002592***	0.148457 ± 0.010936	0.008248 ± 0.001546***
NS1	640	0.153716 ± 0.036706	0.001394 ± 0.000608***	0.165794 ± 0.015071	0.009230 ± 0.002289***
NS2a	640	0.275101 ± 0.049482	0.021878 ± 0.007099***	0.015913 ± 0.019196	0.013358 ± 0.003003
NS2b	640	N/A	N/A	0.163652 ± 0.024173	0.006573 ± 0.002478***
NS3	640	0.125992 ± 0.011680	0.002846 ± 0.001030***	0.152138 ± 0.012876	0.006220 ± 0.001721***
NS4a	640	0.140134 ± 0.038446	0.011301 ± 0.006231***	0.163706 ± 0.027722	0.007070 ± 0.003767***
2K	640	0.198848 ± 0.056279	0.000000 ± 0.000000***	0.065297 ± 0.037406	0.002092 ± 0.001557
NS4b	640	0.146876 ± 0.022916	0.002653 ± 0.001422***	0.133247 ± 0.022160	0.006453 ± 0.002647***
NS5	640	0.157413 ± 0.015610	0.004376 ± 0.001518***	0.131999 ± 0.008550	0.006949 ± 0.001235***

Table 2.2. Synonymous (π_S) and non-synonymsous (π_N) nucleotide diversity (± S.E.) in CTL epitope and non-CTL epitope regions of the 11 proteins of DENV2 (Z-tests, two-tailed, rejected the hypotheses that mean π_S = mean π_N for CTL epitope regions *P<0.05, **P<0.01, ***P<0.001; Z-tests, two-tailed, rejected the hypothesis that mean π_S = mean π_N for non-CTL epitope regions *P<0.05, **P<0.01, ***P<0.001).

DENV2		CTL Epitope Region		Non-CTL Epitope Region	
Protein	**n**	**π_S ± S.E.**	**π_N ± S.E.**	**π_S ± S.E.**	**π_N ± S.E.**
AnchC	598	0.156835 ± 0.038395	0.009034 ± 0.005002***	0.140824 ± 0.039421	0.024135 ± 0.009761**
preM	598	0.139154 ± 0.039400	0.013998 ± 0.006793**	0.228475 ± 0.034368	0.009684 ± 0.003204***
E	598	0.203800 ± 0.020994	0.009471 ± 0.002882***	0.210192 ± 0.017839	0.009921 ± 0.002218***
NS1	598	0.274804 ± 0.043861	0.016930 ± 0.004899***	0.204652 ± 0.018894	0.009653 ± 0.002289***
NS2a	598	0.254570 ± 0.053224	0.031210 ± 0.010899***	0.255409 ± 0.026457	0.014861 ± 0.003722***
NS2b	598	0.259382 ± 0.042397	0.012649 ± 0.005820***	0.178467 ± 0.038467	0.005037 ± 0.002661***
NS3	598	0.220837 ± 0.015193	0.006876 ± 0.001722***	0.220531 ± 0.022103	0.010276 ± 0.002799***
NS4a	598	0.185536 ± 0.029878	0.011192 ± 0.005276***	0.268489 ± 0.048939	0.009823 ± 0.005002***
2K	598	0.194842 ± 0.050233	0.000114 ± 0.000137***	0.093340 ± 0.037112	0.002399 ± 0.001539*
NS4b	598	0.270100 ± 0.030465	0.008696 ± 0.002939***	0.220691 ± 0.033157	0.009601 ± 0.003803***
NS5	598	0.205822 ± 0.015058	0.014183 ± 0.002563***	0.205426 ± 0.013982	0.010189 ± 0.001651***

Table 2.3. Synonymous (π_S) and non-synonymsous (π_N) nucleotide diversity (± S.E.) in CTL epitope and non-CTL epitope regions of the 11 proteins of DENV3 (Z-tests, two-tailed, rejected the hypotheses that mean π_S = mean π_N for CTL epitope regions *P<0.05, **P<0.01, ***P<0.001; Z-tests, two-tailed, rejected the hypothesis that mean π_S = mean π_N for non-CTL epitope regions *P<0.05, **P<0.01, ***P<0.001).

DENV3		CTL Epitope Region		Non-CTL Epitope Region	
Protein	**n**	**π_S ± S.E.**	**π_N ± S.E.**	**π_S ± S.E.**	**π_N ± S.E.**
AnchC	341	0.084428 ± 0.024962	0.012673 ± 0.006543**	0.054009 ± 0.017382	0.008089 ± 0.004138*
preM	341	0.115803 ± 0.045164	0.001566 ± 0.001232*	0.120089 ± 0.017533	0.005351 ± 0.002197***
E	341	0.173150 ± 0.040467	0.010361 ± 0.005097***	0.154298 ± 0.012253	0.007664 ± 0.002027***
NS1	341	0.142580 ± 0.026917	0.010503 ± 0.004659***	0.135931 ± 0.016604	0.005409 ± 0.001670***
NS2a	341	0.202068 ± 0.041102	0.012741 ± 0.006977***	0.162412 ± 0.023427	0.009603 ± 0.002371***
NS2b	341	0.153591 ± 0.065513	0.000502 ± 0.000472*	0.132889 ± 0.022444	0.002714 ± 0.000995***
NS3	341	0.134788 ± 0.013904	0.003846 ± 0.001227***	0.161881 ± 0.015291	0.003094 ± 0.001268***
NS4a	341	0.210068 ± 0.056317	0.015619 ± 0.010262***	0.140364 ± 0.026306	0.009322 ± 0.003781***
2K	341	0.108676 ± 0.065087	0.000000 ± 0.000000	0.216785 ± 0.078614	0.022914 ± 0.020093*
NS4b	341	0.138774 ± 0.022417	0.002753 ± 0.001854***	0.158675 ± 0.024709	0.002266 ± 0.001384***
NS5	341	0.137587 ± 0.013599	0.009476 ± 0.002329***	0.122250 ± 0.009058	0.004976 ± 0.001059***

Table 2.4. Synonymous (π_S) and non-synonymsous (π_N) nucleotide diversity (± S.E.) in CTL epitope and non-CTL epitope regions of the 11 proteins of DENV4 (Z-tests, two-tailed, rejected the hypotheses that mean π_S = mean π_N for CTL epitope regions *P<0.05, **P<0.01, ***P<0.001; Z-tests, two-tailed, rejected the hypothesis that mean π_S = mean π_N for non-CTL epitope regions *P<0.05, **P<0.01, ***P<0.001).

DENV4		CTL Epitope Region		Non-CTL Epitope Region	
Protein	**n**	**π_S ± S.E.**	**π_N ± S.E.**	**π_S ± S.E.**	**π_N ± S.E.**
AnchC	36	0.115883 ± 0.027116	0.020954 ± 0.012113**	0.073152 ± 0.015764	0.003404 ± 0.001473***
preM	36	0.169872 ± 0.040193	0.017278 ± 0.013210***	0.152543 ± 0.017262	0.005526 ± 0.001513***
E	36	0.162430 ± 0.027186	0.017320 ± 0.005083***	0.159463 ± 0.012028	0.006042 ± 0.001199***
NS1	36	0.200849 ± 0.049205	0.017177 ± 0.006232***	0.140621 ± 0.011346	0.011010 ± 0.002164***
NS2a	36	0.081079 ± 0.043610	0.045311 ± 0.027673	0.178837 ± 0.018938	0.019418 ± 0.004297***
NS2b	36	0.114396 ± 0.053882	0.000000 ± 0.000000*	0.153874 ± 0.021494	0.004968 ± 0.001848***
NS3	36	0.147172 ± 0.014668	0.002110 ± 0.000779***	0.162613 ± 0.012034	0.005374 ± 0.001312***
NS4a	36	0.203086 ± 0.038004	0.007198 ± 0.003811***	0.150597 ± 0.023333	0.004025 ± 0.002246***
2K	36	0.016857 ± 0.044667	0.000074 ± 0.005816	0.180962 ± 0.074187	0.016975 ± 0.012043*
NS4b	36	0.140868 ± 0.018186	0.004717 ± 0.001763***	0.185194 ± 0.023413	0.002268 ± 0.000771***
NS5	36	0.165307 ± 0.014475	0.005412 ± 0.001510***	0.143475 ± 0.008316	0.005784 ± 0.000950***

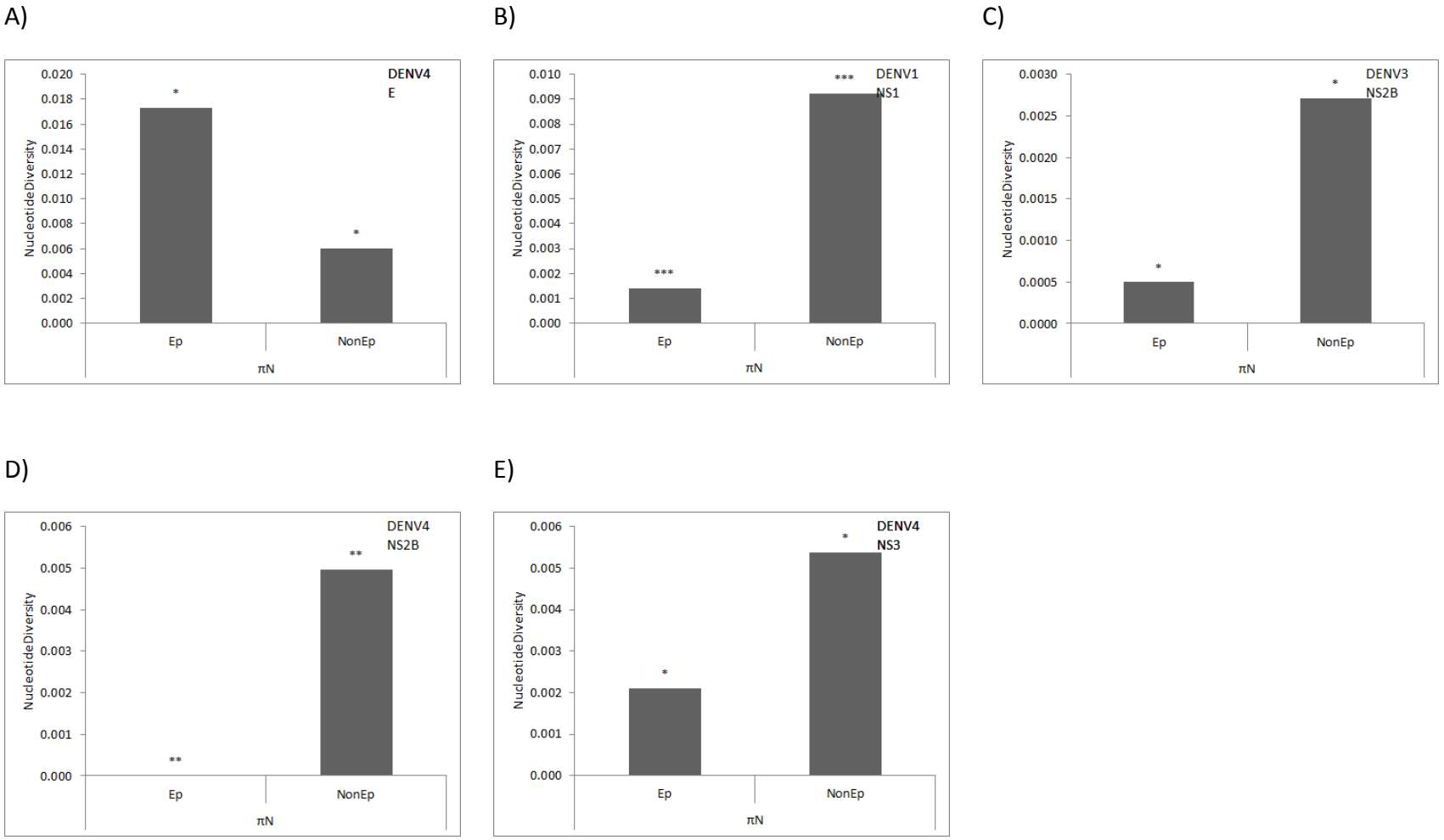


Figure 2.1. Average nucleotide diversity at nonsynonymous polymorphic sites (π_N) in regions of CTL epitopes and non-CTL epitopes in a) E of DENV4 b) NS1 of DENV1 c) NS2B of DENV3 d) NS2B of DENV4 e) NS3 of DENV4 (Z-test, two-tailed, * P<0.05, ** P<0.01, *** P<0.001).

A) B) C)

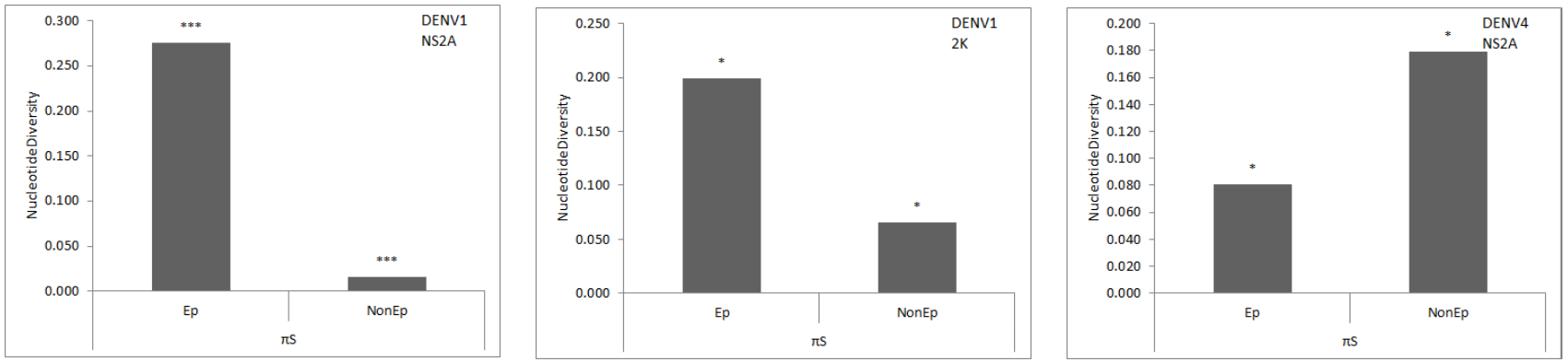


Figure 2.2. Average nucleotide diversity at synonymous polymorphic sites (π_S) in regions of CTL epitopes and non-CTL epitopes in a) NS2A of DENV1 b) 2K of DENV1 c) NS2A of DENV4 (Z-test, two-tailed, * P<0.05, ** P<0.01, *** P<0.001).

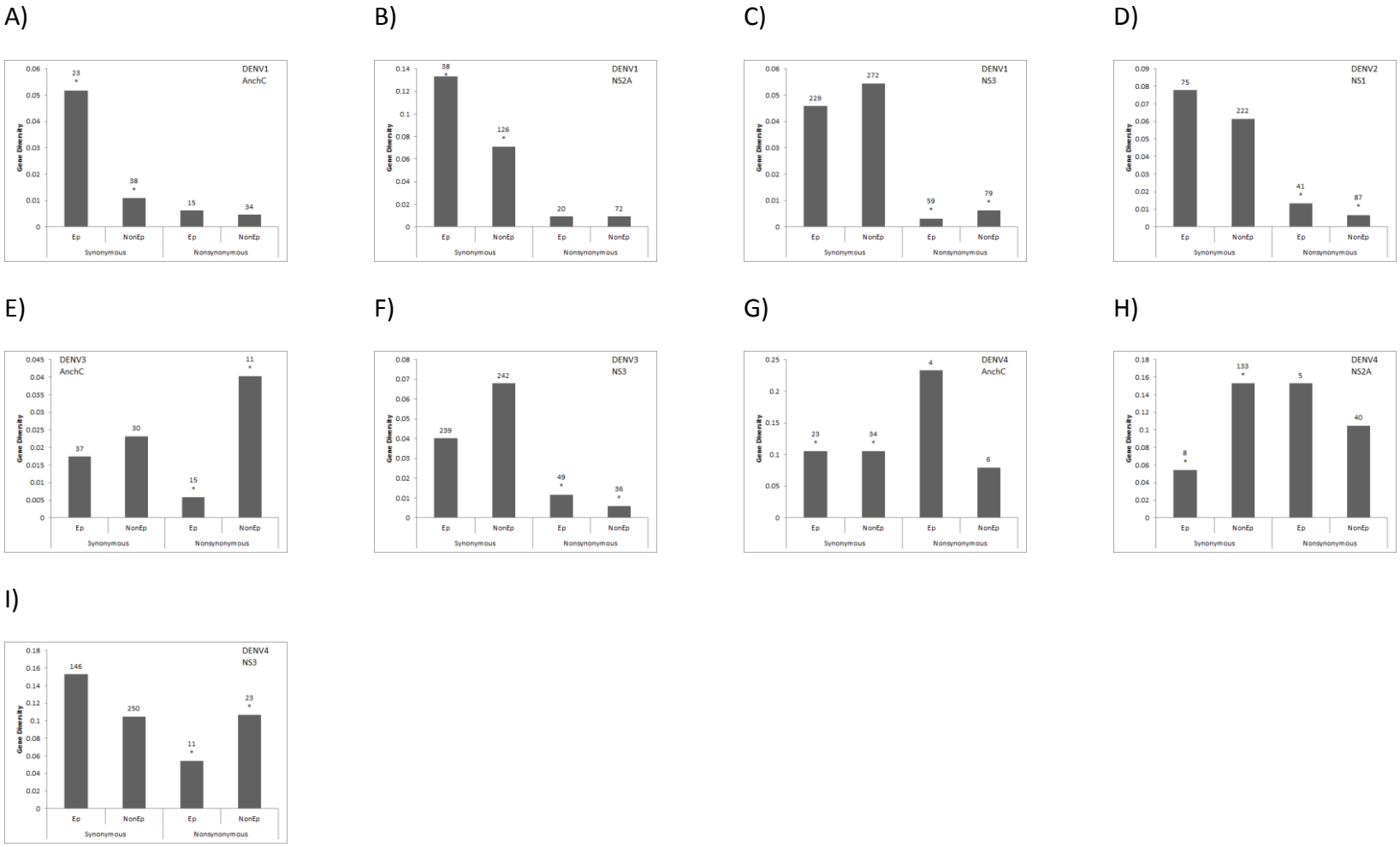


Figure 2.3. Median gene diversity at synonymous and nonsynonymous polymorphic sites in CTL epitope (Ep) and non-CTL epitope (NonEp) regions of a) Anch C of DENV1 b) NS2A of DENV1 c) NS3 of DENV1 d) NS1 of DENV2 e) AnchC of DENV3 f) NS3 of DENV3 g) AnchC of DENV4 h) NS2A of DENV4 i) NS3 of DENV4 (Kruskal-Wallis test * $P<0.05$, ** $P<0.01$, *** $P<0.001$).

CHAPTER 3

COMPARISON OF THREE EPITOPE PREDICTION SOFTWARE PROGRAMS UTILIZING A DENGUE VIRUS DATASET[2]

[2] South, A.C.H., Friedman, R., and A.L. Hughes. To be submitted to *Bioinformatics*.

3.1. Introduction

Vaccines are incredibly important in disease prevention saving millions of lives per year. As there are always new and emerging pathogenic bacteria and viruses as well as currently existing ones for which we have not yet successfully engineered a vaccine, new methods of vaccine development are invented. The advent of the genomic era led to one such modern approach that deals with the creation of subunit vaccines by mining genomes to predict immune response epitopes.

The Dengue virus (DENV) is a single-stranded RNA virus that has become quite an issue with the immune system. There are four serotypes of DENV (DENV1-4) and while infection with a single serotype elicits the normal adaptive immune response, secondary infection by a heterologous serotype any time post primary infection tends to produce a much more severe reaction and may even lead to death (Rodriguez-Roche and Gould, 2013; World Health Organization, 2012a). Due to this severe reaction coupled with the issue that the mechanisms behind secondary infection are currently speculated but unknown, vaccine development for DENV has become a high priority. In a previous study analysis of both a large dataset of DENV genomes and known epitopes was used in speculating evolutionary pressures additionally this data can now be utilized in testing several epitope prediction programs.

The major histocompatibility complex (MHC) of vertebrates (termed human leukocyte antigen (HLA) in humans) is a cluster of genes that are integral in identification of foreign pathogens (Hughes and Yeager, 1998). Molecules from class I MHC are present in all nucleated cells. When cells are infected, the MHC encodes for

various glycoproteins known as epitopes that are presented outside the cell (Andersen et al., 2006; Hughes and Yeager, 1998). Cytotoxic T cells (CTL) have specialized T cell receptors (TCR) that are able to identify these epitopes and trigger T cell proliferation, differentiation of these T cells into both effector and memory T cells as well as the eradication of the infected cell (Andersen et al., 2006; Hughes and Yeager, 1998). Generation of these differentiated T cells by using subunit vaccines of the specific protein epitopes known to generate immune response has already been used successfully in combating hepatitis B since 1986 and has the potential to aide in prevention of other diseases (Hilleman, 2000).

Prediction of these CTL epitopes has become an increasingly expanding field. These *in silico* methods are particularly important in speedily narrowing the search for possible immunologically responsive proteins before conducting extensive, costly wet lab experiments (Yang and Yu, 2009). Prediction methods vary between programs as views on which method is best differ between researchers. This study compared the efficacy of three epitope prediction software programs to identify known DENV epitopes. Each program used different prediction methods in determining overall prediction scores.

3.2. Methods

3.2.1 Known epitopes and sequences utilized in comparisons

Previously, 525 known human CTL epitopes presented by MHC class I were obtained from the Immune Epitope Database and Analysis Resource (IEDB) available at http://www.iedb.org/ (Kim et al., 2012; Vita et al., 2010). Also, 1615 complete genome

sequences of DENV were obtained from the Dengue Virus Database available at http://www.denguedb.org/index.asp (Viral Bioinformatics Resource Center, 2009), 640 belonging to DENV1, 598 to DENV2, 341 to DENV3 and 36 to DENV4. These sequences were placed into the correct reading frame and aligned at the amino acid level followed by superimposition back onto their nucleotide frames as noted in Chapter 2. Each serotype was inspected to find the complete genome or genomes within that serotype that contained the most epitope coverage (Table B.1). Two serotypes had several sequences with the same amount of coverage therefore all of these were used in analysis.

3.2.2 Epitope prediction programs

Several epitope prediction applications were selected based on various traits including that they were all accessible as freeware. NetCTL was among these programs as it has been shown to be more effective in the prediction of MHC class I CTL epitopes for HIV than several other alternative programs including EpiJen, MAPPP, MHC-pathway, and WAPP (Larsen et al., 2007). Initially NetCTL was administered on a LINUX platform but the results were the same when utilizing its web-based terminal which had a more user-friendly interface. Complete polyprotein sequences of those individual genomes containing the most epitope coverage were processed by the program. NetCTL ranks predictions utilizing several scores, one of which is the combination of all the more specific sub scores. The top 25% of results with the highest combination scores were recorded for comparison.

The second program chosen was SVMHC as it also used to predict MHC Class I CTL epitopes (Dönnes and Kohlbacher, 2006). The program was also ranked as having 91-100% accuracy in the Receiver Operating Characteristic curve value of Yang and Yu's review of several T cell epitope prediction methods (Yang and Yu, 2009). Two sets of results were obtained from SVMHC as it utilizes two separate databases, the SYFPEITHI database for MHC ligands (Rammensee et al., 1999) and the MHCPEP database of MHC class I and II binding peptides (Brusic et al., 1994), when determining prediction scores. The same complete polyprotein sequences of the genomes with the most known epitope coverage were processed using SVMHC's web-based interface. The MHCPEP and SYFPEITHI results were combined for analyses and the top 25% saved for comparison.

The IEDB has its own MHC Class I T cell epitope prediction software and was utilized as the final program as it is also the site where the known CTL epitopes being used for comparison were obtained. IEDB's tool incorporates several prediction methods including artificial neural networks (Lundegaard et al., 2008; Nielsen et al., 2003), stabilized matrix methods (Peters and Sette, 2005), and positional scanning of combined peptide libraries (Sidney et al., 2008). The same polyprotein sequences were employed again and only the top 25% of the results retained. It is of note that IEDB's program does not allow input sequences to possess gaps therefore gaps were removed from protein sequences prior to analysis.

3.2.3 Sensitivity measure of performance

Sensitivity of each program was determined by computing true positive predictions (the number of correctly predicted epitopes) and dividing this by actual positives obtained from the IEDB. This was performed amongst both the top 25% of results as well as the top 5% for each of the three programs. Not all available MHC supertypes were utilized in computing sensitivity. Seven supertypes were common among all three programs (Table 3.1) therefore comparison was restricted to results from the seven common supertypes in order to prevent bias.

3.2.4 RANK measure

Similarly to a measure utilized in NetCTL's HIV dataset comparison (Larsen et al., 2007), each true positive epitope identified by a pair of programs was observed individually. Whichever program of the pair gave the true positive epitope the higher score (standardized by number of scores generated per each program in either top 25% or top 5%), was counted as having a better predictive performance in each case. These counts were utilized as a RANK measure to determine overall predictive performance and were compared using a binomial distribution assuming that each software had a 50% chance of having the higher RANK measure.

3.3. Results

3.2.1 Sensitivity measures

Performance comparison of sensitivity of the three programs showed that IEDB had much higher percentages in the top 25% of results followed by NetCTL then SVMHC (Table 3.2a, Fig. 3.1a). When examining the top 5%, NetCTL slightly outperformed IEDB

in the datasets containing DENV1, DENV2 and DENV4 although the results were within 4% for all three serotypes (Table 3.2b Fig. 3.1b). IEDB and NetCTL had almost identical sensitivity (30.10% and 29.59% respectively) for DENV3 (Table 3.2, Fig. 3.1b). SVMHC again had drastically lower sensitivity (Table 3.2b, Fig. 3.1b).

3.2.2 RANK measures

Comparisons of RANK scoring for the top 25% of predictions were highly significant across all four serotypes as well as all four combined in favor of IEDB when comparing IEDB to SVMHC and in favor of NetCTL when comparing NetCTL to SVMHC (P<0.001 in all cases, Table B.2a, Fig. 3.2a-e). When comparing IEDB and NetCTL, results were significantly in favor of IEDB in DENV1, DENV3 (P<0.05 in both cases, Fig. 3.2a, Fig. 3.2c) and in all serotypes combined (P<0.01, Fig. 3.2e). RANK measure results were similar for the top 5% of epitope prediction results (Table B.2b). IEDB versus SVMHC had significance in favor of IEDB for DENV1 (P<0.05, Fig. 3.3a), DENV2, DENV4 (P<0.01 in both cases, Fig. 3.3b, Fig. 3.3d) and in all serotypes combined (P<0.001, Fig 3.3e). The comparisons of NetCTL and SVMHC only significantly favored NetCTL in DENV3 (P<0.01, Fig. 3.3c) and the combination of all serotypes (P<0.001, Fig. 3.3e). RANK measures of IEDB and NetCTL showed that IEDB was favored significantly in DENV3 (P<0.05, Fig. 3.3c) and in all serotypes combine (P<0.01, Fig. 3.3e).

3.4. Discussion

Epitope prediction software programs are becoming more sophisticated with advances in technology (Delany et al., 2014; Gilbert, 2011; He et al., 2010). The expansion of our understanding immunological response is aiding the efficacy of these

programs and their ability to help in the prevention of future illness. Due to the severity of a secondary heterotypic DENV infection causing extreme reactions in humans, vaccine creation for the virus has been extremely difficult which makes it a prime candidate for vaccine informatics in order to unearthing potential epitopes that may exist amongst all four DENV serotypes.

Of the three programs (IEDB, NetCTL, and SVMHC) compared utilizing complete genomes of all four serotypes of DENV as well as known epitopes obtained from the IEDB, there were clear differences in both sensitivity and RANK measures for each program. Percentages of sensitivity when comparing the top 25% of results from each program showed IEDB to have the highest performance. However, if only the top 5% of prediction results are compared, IEDB and NetCTL both have similar sensitivity percentages. Both IEDB and NetCTL are clearly capable of predicting true epitopes at least 20% of the time within the top 5% of their total results.

The IEDB program produces an exorbitant number of results in comparison to both NetCTL and SVMHC. In order to control for the discrepancy in the number of results as well as for identification of different true epitopes, the RANK measure was employed to determine if the higher sensitivity percentages were wholly due to having a larger number of results. IEDB outperformed both NetCTL and SVMHC in the top 25% and 5% of prediction results when comparing RANK measures although NetCTL also performed well. Differences between IEDB and NetCTL results were not always significant. The IEDB may have had superior performance as its CTL epitope database was the one utilized in the search for the known DENV epitopes used in this study.

SVMHC had the lowest performance across all sensitivity percentages and RANK measures at both the 5% and 25% mark. This program was initially chosen due to its high accuracy which also takes specificity of prediction and false positives into account. Specificity is defined by dividing the number of true negative epitopes (proteins that are correctly predicted as non-epitopes) by the actual number of non-epitopes in the dataset. This study solely examined the ability to predict true epitopes and was not concerned with either false positive results or predicted non-epitope proteins. SVMHC's results were only available at three peptide lengths (octamer, nonamer and decamer) while IEDB and NetCTL could both predict larger epitopes. Either of these facts may explain SVMHC's poor performance.

As subunit vaccines are typically more efficient to manufacture than other types of vaccines as well as having little to no adverse effects, the ability to precisely predict which epitopes will cause the most significant immune response is critical to the future of vaccine development. Therefore the expanding field of software programs that predict epitopes will be vital to the development of new subunit vaccines. Both IEDB and NetCTL are such programs that employ several immunological methods in concert for their predictions. Both outperformed SVMHC which only utilizes support vector machines. This study suggests that programs combining several methods, both immunologically and computationally based are the most successful in terms of both sensitivity percentage and RANK measure.

Table 3.1. Representative shared HLA supertype alleles for each of the three software programs compared. Each program had many other alleles available however these were not shared amongst all three and therefore not used in comparison.

Serotype	IEDB	NetCTL	SVMHC
A1	HLA-A*0101	HLA-A*0101	HLA-A1
A2	HLA-A*0201	HLA-A*0201	HLA-A2, HLA-A*0201
A3	HLA-A*0301	HLA-A*0301	HLA-A3, HLA-A*0301
A24	HLA-A*2402	HLA-A*2402	HLA-A*2402
B7	HLA-B*0702	HLA-B*0702	HLA-B7, HLA-B*0702
B8	HLA-B*0801	HLA-B*0801	HLA-B8
B27	HLA-B*2705	HLA-B*2705	HLA-B*2705

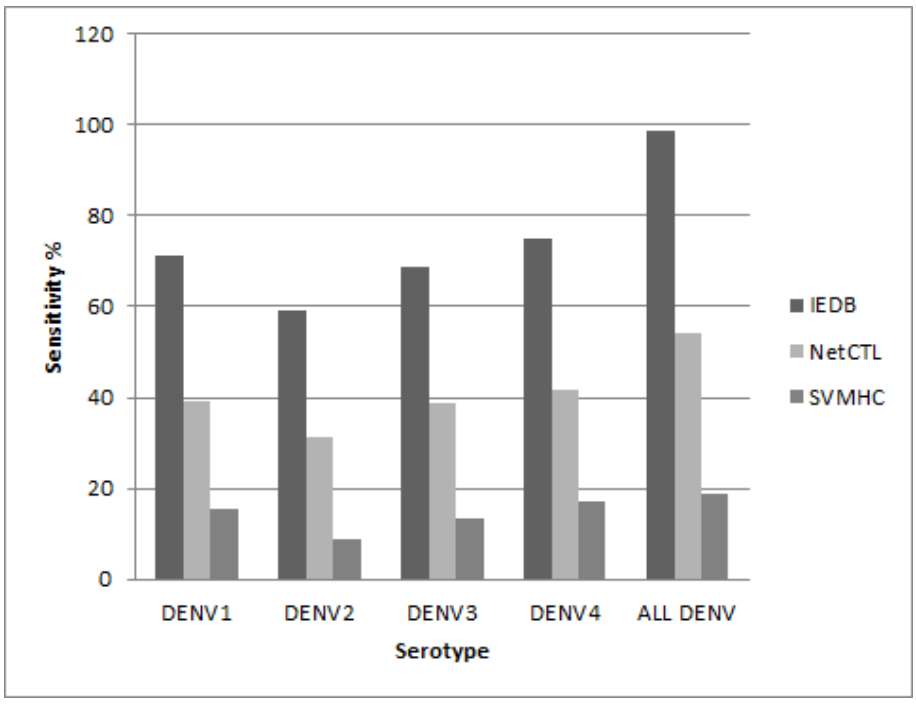
Table 3.2a. Sensitivity percentages at top 25% of prediction results for all four DENV serotypes individually as well as all serotypes combined.

Program	DENV1	DENV2	DENV3	DENV4	All DENV
IEDB	71.17	59.18	68.88	74.83	98.86
NetCTL	39.26	31.46	38.78	41.50	54.29
SVMHC	15.34	8.61	13.27	17.01	18.86

Table 3.2b. Sensitivity percentages at top 5% of prediction results for all four DENV serotypes individually as well as all serotypes combined.

Program	DENV1	DENV2	DENV3	DENV4	All DENV
IEDB	27.61	20.97	30.10	28.57	38.48
NetCTL	31.29	23.97	29.59	31.29	41.71
SVMHC	3.68	3.37	3.57	6.80	6.10

A)



B)

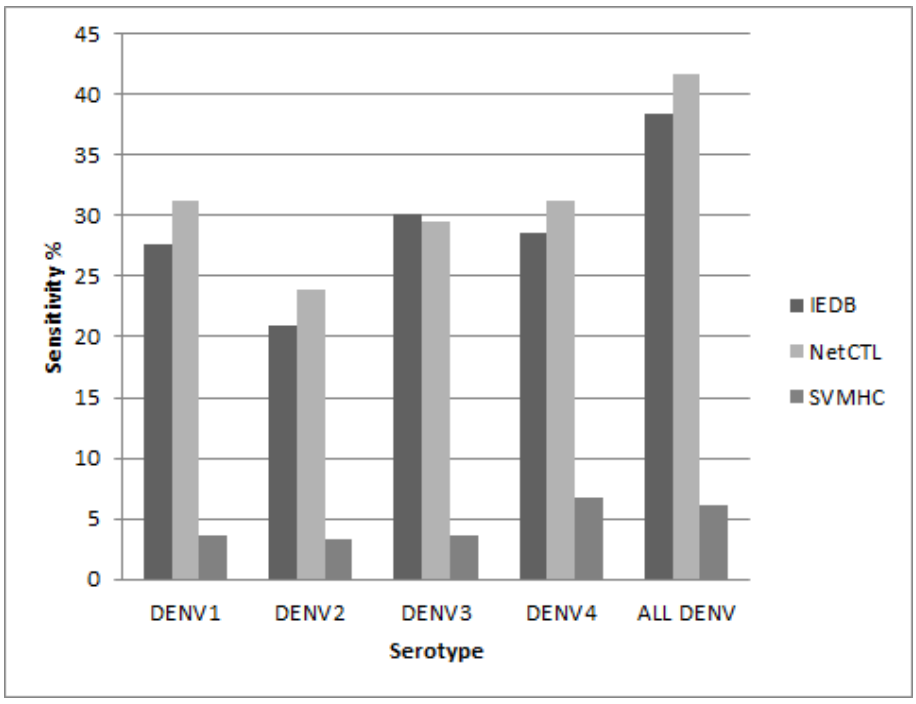
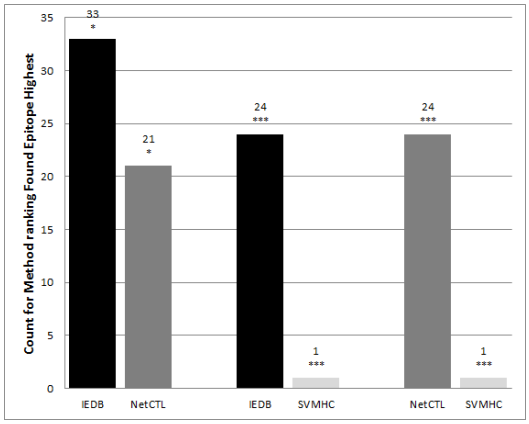
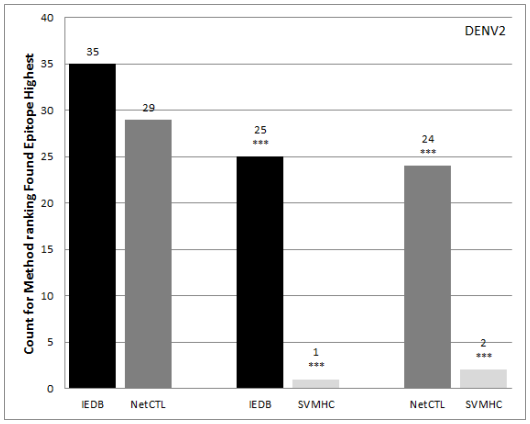


Figure 3.1. Sensitivity percentages for all four serotypes individually and combined in comparisons of IEDB, NetCTL, and SVMHC epitope prediction software at a) top 5% of predicted results and b) top 25% of predicted results.

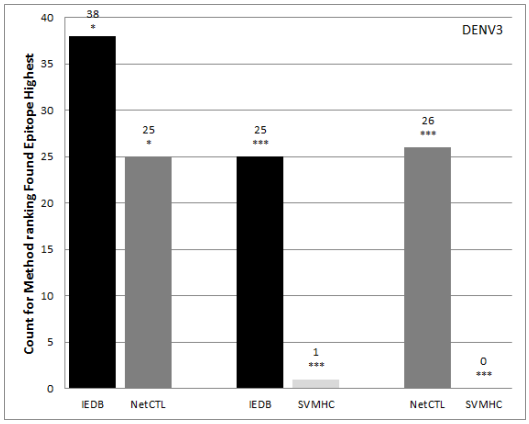
A)



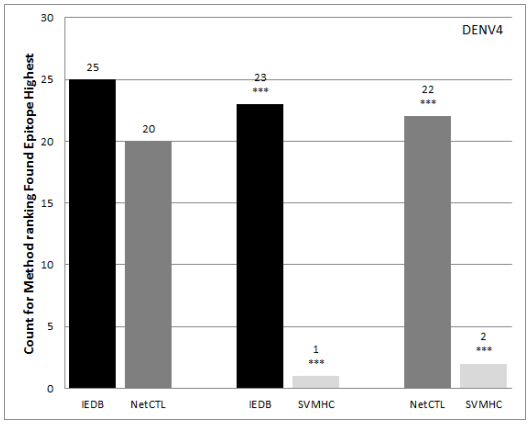
B)



C)



D)



E)

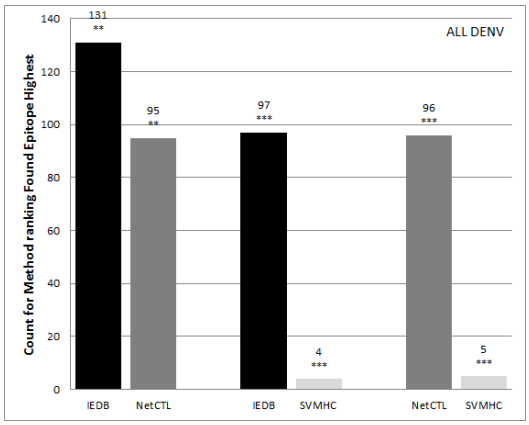
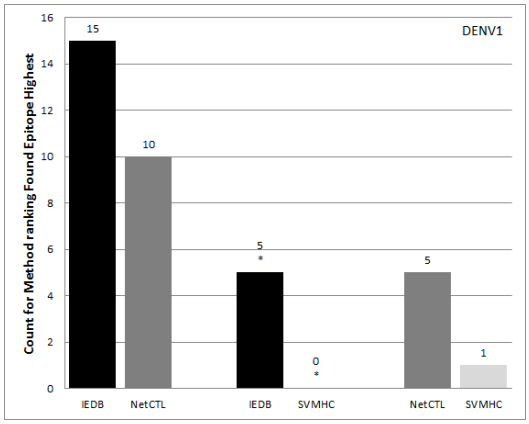
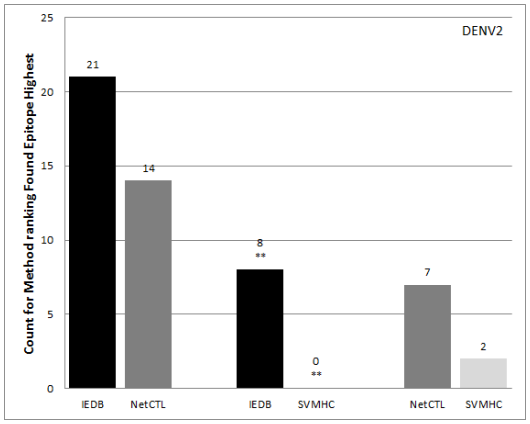


Figure 3.2. Rank measures of top 25% of predictions for comparisons of counts of epitope rank by pairs of epitope prediction programs for datasets of a) DENV1 b) DENV2 c) DENV3 d) DENV4 and e) all four serotypes combined (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).

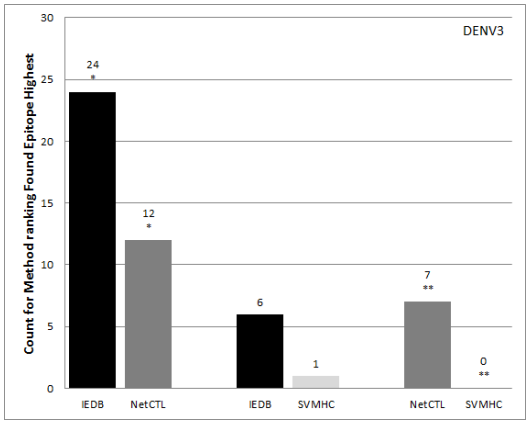
A)



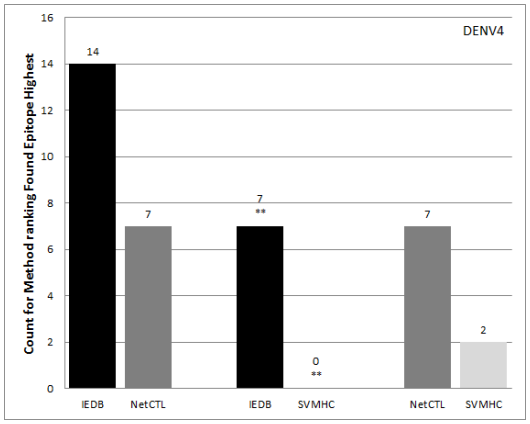
B)



C)



D)



E)

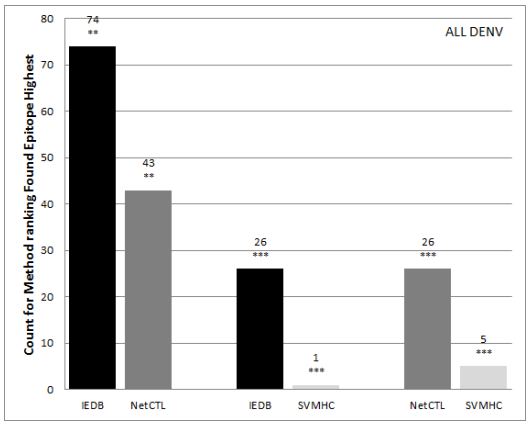


Figure 3.3. Rank measures of top 5% of predictions for comparisons of counts of epitope rank by pairs of epitope prediction programs for datasets of a) DENV1 b) DENV2 c) DENV3 d) DENV4 and e) all four serotypes combined (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).

CHAPTER 4

SUMMARY

The results of our first study support that there are significant differences in both nucleotide polymorphism and gene diversity between DENV serotypes facilitating understanding of severe immune response to a secondary heterologous infection. This data suggests conflicting evolutionary pressures on each serotype including an abundance of purifying selection as well as positive selection coupled with relaxation of purifying selection in attempts to escape immune system recognition.

Not only are there significant differences between the serotypes but also between CTL epitope and non-CTL epitope regions of the E protein of DENV4, NS1 protein of DENV2, and NS3 protein of DENV3. These results indicate that the conflicting evolutionary pressures exist within each serotype as well, particularly affecting proteins that are of interest in class I MHC identification but interestingly enough not across all four serotypes equally.

Results comparing the efficacy of epitope prediction applications demonstrated that programs incorporating several methods including algorithms, neural networks, matrices and peptide library searches were more capable of assigning higher scores to known DENV epitopes. The two multiple method programs, IEDB and NetCTL, significantly outperformed a single method program, SVMHC, in both RANK measures

and sensitivity percentages. As epitope prediction software programs are becoming more readily used in assistance with subunit vaccine creation, the efficacy of each program is critical to its continued value.

REFERENCES

Andersen, M.H., Schrama, D., Straten, P. thor, Becker, J.C., 2006. Cytotoxic T Cells. J. Invest. Dermatol. 126, 32–41.

Bedi, S.K., Prasad, A., Mathur, K., Bhatnagar, S., 2013. Positive selection and evolution of dengue type-3 virus in the Indian subcontinent. J. Vector Borne Dis. 50, 188–196.

Brusic, V., Rudy, G., Harrison, L.C., 1994. MHCPEP: a database of MHC-binding peptides. Nucleic Acids Res. 22, 3663–3665.

De Gregorio, E., Rappuoli, R., 2012. Vaccines for the future: learning from human immunology. Microb. Biotechnol. 5, 149–155.

Delany, I., Rappuoli, R., De Gregorio, E., 2014. Vaccines for the 21st century. EMBO Mol. Med. 6, 708–720.

Dimmock, N.J., Easton, A.J., Leppard, K.N., 2007. Introduction to Modern Virology, 6th ed. Blackwell Publishing, Malden, MA.

Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51, 151–178.

Dönnes, P., Kohlbacher, O., 2006. SVMHC: A server for prediction of MHC-binding peptides. Nucleic Acids Res. 34, 194–197.

Drake, J.W., Charlesworth, B., Charlesworth, D., Crow, J.F., 1998. Rates of spontaneous mutation. Genetics 148, 1667–1686.

Durbin, A.P., Kirkpatrick, B.D., Pierce, K.K., Elwood, D., Larsson, C.J., Lindow, J.C., Tibery, C., Sabundayo, B.P., Shaffer, D., Talaat, K.R., Hynes, N. a, Wanionek, K., Carmolli, M.P., Luke, C.J., Murphy, B.R., Subbarao, K., Whitehead, S.S., 2013. A single dose of any of four different live attenuated tetravalent dengue vaccines is safe and immunogenic in flavivirus-naive adults: a randomized, double-blind clinical trial. J. Infect. Dis. 207, 957–65.

Eck, R.V., Dayhoff, M.O., 1966. Atlas of Protein Sequence and Structure, Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Silver Spring, Maryland.

Felsenstein, J., 1985. Confidence Limits on Phylogenies : An Approach Using the Bootstrap Author (s): Joseph Felsenstein Stable URL : http://www.jstor.org/stable/2408678 . Evolution (N. Y). 39, 783–791.

Fields, B.N., Knipe, D.M., Howley, P.M., 2013. Fields Virology, 6th ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.

Fitch, W., 1969. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20, 406–416.

Gilbert, S.C., 2011. T-cell-inducing vaccines - what's the future. Immunology 135, 19–26.

Goncalvez, A.P., Engle, R.E., St Claire, M., Purcell, R.H., Lai, C.-J., 2007. Monoclonal antibody-mediated enhancement of dengue virus infection in vitro and in vivo and strategies for prevention. Proc. Natl. Acad. Sci. U. S. A. 104, 9422–9427.

Graur, D., Li, W.-H., 2000. Fundamentals of Molecular Evolution, 2nd ed. Sinauer Associates, Sunderland, Massachusetts.

Halstead, S.B., 1979. In Vivo Enhancement of Dengue Virus Infection in Rhesus Monkeys by Passively Transferred Antibody. J. Infect. Dis. 140, 527–533.

Halstead, S.B., Rojanasuphot, S., Sangkawibha, A., 1983. Original Antigenic Sin in Dengue. Am. J. Trop. Med. Hyg. 32, 154–156.

Hartl, D.L., 2000. A Primer of Population Genetics, 3rd ed. Sinauer Associates, Sunderland, Massachusetts.

He, Y., Rappuoli, R., De Groot, A.S., Chen, R.T., 2010. Emerging vaccine informatics. J. Biomed. Biotechnol. 2010, 1–26.

Hilleman, M.R., 2000. Vaccines in historic evolution and perspective: a narrative of vaccine discoveries. Vaccine 18, 1436–1447.

Hollander, M., Wolfe, D.A., Chicken, E., 2013. Nonparametric Statistical Methods, 3rd ed. John Wiley & Sons, New York, New York.

Holmes, E.C., 2003. Patterns of Intra- and Interhost Nonsynonymous Variation Reveal Strong Purifying Selection in Dengue Virus. Society 77, 11296–11298.

Hughes, A.L., 1999. Adaptive Evolution of Genes and Genomes. Oxford University Press, New York, New York.

Hughes, A.L., Friedman, R., Glenn, N.L., 2006. The Future of Data Analysis in Evolutionary Genomics. Curr. Genomics 7, 227–234.

Hughes, A.L., Hughes, M.A.K., 2007. More effective purifying selection on RNA viruses than in DNA viruses. Gene 404, 117–125.

Hughes, A.L., Yeager, M., 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. 32, 415–435.

Idrees, S., Ashfaq, U. a, 2012. A brief review on dengue molecular virology, diagnosis, treatment and prevalence in Pakistan. Genet. Vaccines Ther. 10, 1–10.

Iglesias, N.G., Byk, L.A., Gamarnik, A. V., 2014. Molecular virology of dengue virus, in: Gubler, D.J., Ooi, E.E. (Eds.), Dengue and Dengue Hemorrhagic Fever. CAB International, Boston, Massachusetts, p. 624.

Irausquin, S.J., Hughes, A.L., 2010. Conflicting selection pressures target the NS3 protein in hepatitis C virus genotypes 1a and 1b. Virus Res. 147, 202–7.

Jameson, S.C., Bevan, M.J., 1995. T cell receptor antagonists and partial agonists. Immunity 2, 1–11.

Kalergis, A.M., Nathenson, S.G., 2000. Altered peptide ligand-mediated TCR antagonism can be modulated by a change in a single amino acid residue within the CDR3 beta of an MHC class I-restricted TCR. J. Immunol. 165, 280–285.

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P.E., Nielsen, M., Peters, B., 2012. Immune epitope database analysis resource. Nucleic Acids Res. 40, 525–530.

Kimura, M., 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267, 275–276.

Kimura, M., 1994. Population genetics, molecular evolution, and the neutral theory. The University of Chicago Press, Chicago, Illinois.

Kliks, S.C., Nisalak, A., Brandt, W.E., Wahl, L., Burke, A.S., 1989. Antibody-dependent enhancement of dengue virus growth in human monocytes as a risk factor for dengue hemorrhagic fever. Am. J. Trop. Med. Hyg. 40, 444–451.

Knapp, E.W., Irausquin, S.J., Friedman, R., Hughes, A.L., 2011. PolyAna: analyzing synonymous and nonsynonymous polymorphic sites. Conserv. Genet. Resour. 3, 429–431.

Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N., Cropp, C.B., 1998. Phylogeny of the genus Flavivirus. J. Virol. 72, 73–83.

Lambkin, R., McLain, L., Jones, S.E., Aldridge, S.L., Dimmock, N.J., 1994. Neutralization escape mutants of type A influenza virus are readily selected by antisera from mice immunized with whole virus: A possible mechanism for antigenic drift. J. Gen. Virol. 75, 3493–3502.

Larsen, M. V, Lundegaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8, 1–12.

Lescar, J., Lok, S.-M., 2014. The structural biology of dengue virus, in: Gubler, D.J., Ooi, E.E. (Eds.), Dengue and Dengue Hemorrhagic Fever. CAB International, Boston, Massachusetts, p. 624.

Li, W.-H., 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36, 96–99.

Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., Nielsen, M., 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res. 36, 509–512.

Minitab 13.1 Statistical Software, 2000.

Moore, C.B., John, M., James, I.R., Christiansen, F.T., Witt, C.S., Mallal, S. a, 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science (80-.). 296, 1439–1443.

Murray, G.G.R., Kosakovsky Pond, S.L., Obbard, D.J., 2013. Suppressors of RNAi from plant viruses are subject to episodic positive selection. Proc. Biol. Sci. 280, 1–10.

Nei, M., 1975. Molecular Population Genetics and Evolution. American Elsevier Publishing Company, New York, New York.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York, New York.

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S.L., Lamberth, K., Buus, S., Brunak, S., Lund, O., 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 12, 1007–1017.

Ohta, T., 1973. Slightly deleterious mutant substitutions in evolution. Nature 246, 96–98.

Pamilo, P., Bianchi, N.O., 1993. Evolution of the Zfx and Zfy Genes: Rates and Interdependence between the Genes. Mol. Biol. Evol. 10, 271–281.

Patronov, A., Doytchinova, I., 2013. T-cell epitope vaccine design by immunoinformatics. Open Biol. 3, 1–13.

Peters, B., Sette, A., 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6, 132–140.

Piazza, P., Campbell, D., Marques, E., Hildebrand, W.H., Buchli, R., Mailliard, R., Rinaldo, C.R., 2014. Dengue virus-infected human dendritic cells reveal hierarchies of naturally expressed novel NS3 CD8 T cell epitopes. Clin. Exp. Immunol. 177, 696–702.

Piontkivska, H., Hughes, A.L., 2006. Patterns of sequence evolution at epitopes for host antibodies and cytotoxic T-lymphocytes in human immunodeficiency virus type 1. Virus Res. 116, 98–105.

Rammensee, H.-G., Bachmann, J., Emmerich, N.P.N., Bachor, O.A., Stevanović, S., 1999. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50, 213–219.

Rodriguez-Roche, R., Gould, E. a, 2013. Understanding the dengue viruses and progress towards their control. Biomed Res. Int. 2013, 1–20.

Ross, H. a, Ross, H. a, Rodrigo, A.G., Rodrigo, A.G., 2002. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J. Virol. 76, 11715–11720.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., 2000. Artemis: sequence visualization and annotation. Bioinforma. Appl. note 16, 944–945.

Saitou, N., Nei, M., 1987. The Neighbor-joining Method : A New Method for Reconstructing Phylogenetic Trees ’. Mol. Biol. Evol. 4, 406–425.

Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A., Peters, B., 2008. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res. 4, 2–15.

Sloan-lancaster, J., Allen, P.M., 1996. ALTERED PEPTIDE LIGAND – INDUCED PARTIAL T CELL ACTIVATION : Molecular Mechanisms and Role in T Cell Biology 1–27.

Suzuki, Y., 2006. Natural selection on the influenza virus genome. Mol. Biol. Evol. 23, 1902–1911.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–9.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–80.

Vaughan, K., Greenbaum, J., Blythe, M., Peters, B., Sette, A., 2010. Meta-analysis of All Immune Epitope Data in the Flavivirus Genus : Inventory of Current Immune Epitope Data Status in the Context of Virus Immunity and Immunopathology. Viral Immunol. 23, 259–284.

Viral Bioinformatics Resource Center, 2009. Dengue Virus Database [WWW Document]. URL http://www.denguedb.org/index.asp

Vita, R., Zarebski, L., Greenbaum, J. a, Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., Peters, B., 2010. The immune epitope database 2.0. Nucleic Acids Res. 38, D854–62.

Volz, E.M., Koelle, K., Bedford, T., 2013. Viral phylodynamics. PLoS Comput. Biol. 9, e1002947.

Weaver, S.C., Vasilakis, N., 2009. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect. Genet. Evol. 9, 523–40.

Weiskopf, D., Angelo, M. a, de Azeredo, E.L., Sidney, J., Greenbaum, J. a, Fernando, A.N., Broadwater, A., Kolla, R. V, De Silva, A.D., de Silva, A.M., Mattia, K. a, Doranz, B.J., Grey, H.M., Shresta, S., Peters, B., Sette, A., 2013. Comprehensive analysis of

dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc. Natl. Acad. Sci. U. S. A. 110, E2046–E2053.

Whitehead, S.S., Blaney, J.E., Durbin, A.P., Murphy, B.R., 2007. Prospects for a dengue virus vaccine. Nat. Rev. Microbiol. 5, 518–529.

World Health Organization, 2012a. Dengue and severe dengue. Geneva.

World Health Organization, 2012b. Global strategy for dengue prevention and control 2012-2020. Geneva.

Yang, X., Yu, X., 2009. An introduction to epitope prediction methods and software. Rev. Med. Virol. 19, 77–96.

Yauch, L.E., Zellweger, R.M., Kotturi, M.F., Qutubuddin, A., Sidney, J., Peters, B., Prestwood, T.R., Sette, A., Shresta, S., 2009. A Protective Role for Dengue Virus-Specific CD8+ T Cells. J. Immunol. 182, 4865–4873.

APPENDIX A – CHAPTER 2 SUPPLEMENTARY INFORMATION

Table A.1. GenBank accession numbers for complete genome sequences retrieved from the Dengue Virus Database.

	Accession	Serotype		Accession	Serotype
1	AB178040	1	809	EU687214	2
2	EU081229	1	810	EU687215	2
3	EU081230	1	811	EU687216	2
4	EU081231	1	812	EU726770	2
5	EU081232	1	813	EU687217	2
6	EU081233	1	814	EU781135	2
7	EU081234	1	815	EU687220	2
8	EU081235	1	816	EU726775	2
9	EU081236	1	817	EU726767	2
10	EU081237	1	818	EU687222	2
11	EU081238	1	819	EU687223	2
12	EU081239	1	820	EU687224	2
13	EU081240	1	821	EU687225	2
14	EU081241	1	822	EU687227	2
15	EU081242	1	823	EU687228	2
16	EU081243	1	824	EU687229	2
17	EU081244	1	825	EU687230	2
18	EU081245	1	826	EU687231	2
19	EU081246	1	827	EU687232	2
20	EU081247	1	828	EU687199	2
21	EU081248	1	829	EU687235	2
22	EU081249	1	830	EU687236	2
23	EU081250	1	831	EU687237	2
24	EU081251	1	832	EU687238	2
25	EU081252	1	833	EU687240	2
26	EU081253	1	834	EU687241	2
27	EU081254	1	835	EU687242	2
28	EU081255	1	836	EU687243	2
29	EU081256	1	837	EU687244	2
30	EU081257	1	838	EU687245	2
31	EU081258	1	839	EU687246	2
32	EU081259	1	840	EU660413	2

33	EU081260	1	841	EU660414	2
34	EU081261	1	842	EU687248	2
35	EU081262	1	843	EU660415	2
36	EU081263	1	844	EU660416	2
37	EU081264	1	845	EU726776	2
38	EU081265	1	846	EU660417	2
39	EU081266	1	847	EU677137	2
40	EU081267	1	848	EU677138	2
41	EU081268	1	849	EU677148	2
42	EU081269	1	850	EU677149	2
43	EU081270	1	851	EU687250	2
44	EU081271	1	852	FJ024473	2
45	EU081272	1	853	FJ024474	2
46	EU081273	1	854	FJ024475	2
47	EU081274	1	855	FJ182012	2
48	EU081275	1	856	FJ390389	2
49	EU081276	1	857	EU854293	2
50	EU081277	1	858	EU854294	2
51	EU081278	1	859	FJ024477	2
52	EU081279	1	860	FJ410288	2
53	EU081226	1	861	FJ024452	2
54	EU081227	1	862	FJ205877	2
55	EU081228	1	863	FJ024454	2
56	EU081280	1	864	FJ390384	2
57	EU081281	1	865	FJ205878	2
58	AF180817	1	866	FJ205879	2
59	AF180818	1	867	FJ390385	2
60	AF514883	1	868	FJ024458	2
61	AF514878	1	869	FJ205880	2
62	AY206457	1	870	FJ390387	2
63	AF514885	1	871	FJ024461	2
64	AF514889	1	872	FJ373299	2
65	AF514876	1	873	FJ205885	2
66	EF025110	1	874	FJ410193	2
67	AB189120	1	875	FJ410195	2
68	AB189121	1	876	FJ432726	2
69	AB074761	1	877	FJ562098	2
70	AF298807	1	878	FJ410200	2
71	AY277665	1	879	FJ410202	2
72	AY277666	1	880	FJ410208	2
73	AY277664	1	881	FJ461305	2
74	EU482615	1	882	FJ410215	2

75	EU482616	1	883	FJ461309	2
76	EU482617	1	884	FJ461311	2
77	EU482618	1	885	FJ410217	2
78	EU482619	1	886	FJ410219	2
79	EU482609	1	887	FJ461314	2
80	EU482610	1	888	FJ410221	2
81	EU482611	1	889	FJ547064	2
82	EU482567	1	890	FJ410223	2
83	EU660396	1	891	FJ410224	2
84	EU660390	1	892	FJ461321	2
85	EU660391	1	893	FJ410228	2
86	EU660392	1	894	FJ410233	2
87	EU660393	1	895	FJ410237	2
88	EU660394	1	896	FJ410241	2
89	EU660401	1	897	FJ410259	2
90	FJ373305	1	898	FJ547067	2
91	EU660402	1	899	FJ639697	2
92	EU660403	1	900	FJ639698	2
93	EU660397	1	901	FJ639699	2
94	EU687247	1	902	FJ639700	2
95	EU660395	1	903	FJ639701	2
96	EU660412	1	904	FJ639702	2
97	EU660418	1	905	FJ639703	2
98	EU677150	1	906	FJ639704	2
99	EU677151	1	907	FJ639705	2
100	EU677152	1	908	FJ639706	2
101	EU677153	1	909	FJ639707	2
102	EU677154	1	910	FJ639708	2
103	EU677155	1	911	FJ639709	2
104	EU677156	1	912	FJ639710	2
105	EU677157	1	913	FJ639711	2
106	EU726777	1	914	FJ639717	2
107	EU677158	1	915	FJ639718	2
108	EU677159	1	916	FJ639828	2
109	EU677160	1	917	FJ639829	2
110	EU677161	1	918	FJ639830	2
111	EU677162	1	919	FJ639831	2
112	EU677163	1	920	FJ639832	2
113	EU677164	1	921	FJ639732	2
114	EU677165	1	922	FJ639733	2
115	EU677139	1	923	FJ639734	2
116	EU677140	1	924	FJ639783	2

117	EU677166	1	925	FJ639788	2
118	EU677167	1	926	FJ639809	2
119	EU677168	1	927	FJ639822	2
120	EU677169	1	928	FJ639834	2
121	EU677170	1	929	FJ639835	2
122	EU726778	1	930	FJ639836	2
123	EU677171	1	931	FJ639837	2
124	EU677172	1	932	EU482748	2
125	EU677173	1	933	EU482749	2
126	EU677174	1	934	EU482750	2
127	EU726779	1	935	EU482751	2
128	EU660419	1	936	EU482752	2
129	EU677175	1	937	EU482753	2
130	EU677176	1	938	EU482770	2
131	EU726780	1	939	FJ226066	2
132	EU726781	1	940	EU482680	2
133	EU677177	1	941	EU482754	2
134	EU726782	1	942	EU482755	2
135	EU677178	1	943	FJ478455	2
136	EU687251	1	944	EU482756	2
137	FJ024425	1	945	EU482757	2
138	FJ024426	1	946	EU482758	2
139	FJ024427	1	947	EU482759	2
140	FJ024428	1	948	EU482681	2
141	FJ024429	1	949	EU482761	2
142	FJ024430	1	950	EU482762	2
143	FJ024431	1	951	EU482763	2
144	FJ024432	1	952	EU482771	2
145	FJ024433	1	953	EU482682	2
146	FJ024434	1	954	EU482444	2
147	FJ024435	1	955	EU482772	2
148	FJ024436	1	956	EU482773	2
149	FJ024437	1	957	EU482683	2
150	FJ024472	1	958	EU482684	2
151	FJ024438	1	959	EU482685	2
152	FJ024439	1	960	EU482686	2
153	FJ024440	1	961	EU482687	2
154	FJ024441	1	962	EU482688	2
155	FJ024442	1	963	EU482689	2
156	FJ024443	1	964	EU482690	2
157	FJ024444	1	965	EU482691	2
158	FJ373296	1	966	EU482692	2

159	FJ182018	1	967	EU482693	2
160	FJ410287	1	968	EU482694	2
161	FJ182019	1	969	EU529706	2
162	FJ432719	1	970	EU482719	2
163	FJ182020	1	971	EU482720	2
164	FJ024445	1	972	EU482721	2
165	FJ182021	1	973	EU482722	2
166	FJ182022	1	974	EU482723	2
167	FJ024446	1	975	EU482724	2
168	FJ182023	1	976	EU482725	2
169	FJ024447	1	977	EU482726	2
170	FJ182024	1	978	EU482727	2
171	FJ182025	1	979	EU482728	2
172	FJ373297	1	980	EU482729	2
173	FJ024448	1	981	EU482730	2
174	FJ182026	1	982	EU482731	2
175	FJ373298	1	983	EU596483	2
176	FJ182027	1	984	EU482597	2
177	FJ390381	1	985	EU482598	2
178	FJ182028	1	986	EU482603	2
179	FJ024449	1	987	EU482599	2
180	FJ390382	1	988	EU482632	2
181	FJ182003	1	989	EU482600	2
182	FJ182029	1	990	EU482633	2
183	FJ390383	1	991	EU482634	2
184	FJ024450	1	992	EU482635	2
185	FJ182030	1	993	EU482636	2
186	FJ182031	1	994	EU482637	2
187	FJ182032	1	995	EU596484	2
188	FJ024451	1	996	EU482638	2
189	FJ182033	1	997	EU482601	2
190	FJ182034	1	998	EU482631	2
191	FJ182035	1	999	EU482602	2
192	FJ182036	1	1000	EU482639	2
193	FJ410289	1	1001	EU482732	2
194	FJ024453	1	1002	EU482733	2
195	FJ205876	1	1003	EU482734	2
196	FJ024455	1	1004	EU482735	2
197	FJ024456	1	1005	EU482736	2
198	FJ024457	1	1006	EU482737	2
199	FJ390386	1	1007	EU482738	2
200	FJ024459	1	1008	EU482739	2

201	FJ205881	1	1009	EU482740	2
202	FJ024460	1	1010	EU482741	2
203	FJ205882	1	1011	EU482742	2
204	FJ205883	1	1012	EU482743	2
205	FJ205884	1	1013	EU482744	2
206	FJ024462	1	1014	EU482745	2
207	FJ024463	1	1015	EU482746	2
208	FJ390388	1	1016	EU482747	2
209	FJ024464	1	1017	EU482764	2
210	FJ390374	1	1018	EU482765	2
211	FJ390378	1	1019	EU482766	2
212	FJ205872	1	1020	EU482767	2
213	FJ205873	1	1021	EU482768	2
214	FJ390379	1	1022	EU482640	2
215	FJ390380	1	1023	EU482641	2
216	FJ205874	1	1024	EU482642	2
217	FJ205875	1	1025	EU482643	2
218	FJ410191	1	1026	EU482644	2
219	FJ432723	1	1027	EU482645	2
220	FJ410192	1	1028	EU482646	2
221	FJ410194	1	1029	EU482647	2
222	FJ432725	1	1030	EU482648	2
223	FJ432727	1	1031	EU482649	2
224	FJ432729	1	1032	EU482650	2
225	FJ432730	1	1033	EU482651	2
226	FJ432732	1	1034	EU482652	2
227	FJ432733	1	1035	EU482653	2
228	FJ410196	1	1036	EU482654	2
229	FJ410197	1	1037	EU482655	2
230	FJ432734	1	1038	EU482656	2
231	FJ432735	1	1039	EU482657	2
232	FJ432736	1	1040	EU482658	2
233	FJ410198	1	1041	EU482659	2
234	FJ432737	1	1042	EU482660	2
235	FJ410199	1	1043	EU482661	2
236	FJ432738	1	1044	EU482662	2
237	FJ432739	1	1045	EU482663	2
238	FJ432740	1	1046	EU482664	2
239	FJ410201	1	1047	EU482665	2
240	FJ410203	1	1048	EU482666	2
241	FJ410204	1	1049	EU482667	2
242	FJ432742	1	1050	EU482668	2

243	FJ410205	1	1051	EU482669	2
244	FJ410206	1	1052	EU482670	2
245	FJ410207	1	1053	EU482671	2
246	FJ432744	1	1054	EU482672	2
247	FJ547060	1	1055	EU482673	2
248	FJ432745	1	1056	EU482674	2
249	FJ410209	1	1057	EU482675	2
250	FJ410210	1	1058	EU482676	2
251	FJ432746	1	1059	EU482677	2
252	FJ410211	1	1060	EU482678	2
253	FJ432747	1	1061	EU482679	2
254	FJ432748	1	1062	EU482695	2
255	FJ432749	1	1063	EU482769	2
256	FJ461306	1	1064	EU482696	2
257	FJ410212	1	1065	EU482774	2
258	FJ461307	1	1066	EU482775	2
259	FJ410213	1	1067	EU482776	2
260	FJ461308	1	1068	EU482777	2
261	FJ410214	1	1069	EU482778	2
262	FJ461310	1	1070	EU482779	2
263	FJ410216	1	1071	EU482780	2
264	FJ562101	1	1072	EU482781	2
265	FJ461312	1	1073	EU482782	2
266	FJ461313	1	1074	EU482783	2
267	FJ547063	1	1075	EU482784	2
268	FJ410218	1	1076	EU482785	2
269	FJ461315	1	1077	EU482786	2
270	FJ461316	1	1078	EU482787	2
271	FJ461317	1	1079	EU482788	2
272	FJ410220	1	1080	EU482697	2
273	FJ410222	1	1081	EU482698	2
274	FJ461318	1	1082	EU482699	2
275	FJ461319	1	1083	EU482700	2
276	FJ410225	1	1084	EU482701	2
277	FJ410226	1	1085	EU482702	2
278	FJ410227	1	1086	EU482703	2
279	FJ547065	1	1087	EU482704	2
280	FJ410230	1	1088	EU482705	2
281	FJ410231	1	1089	EU482590	2
282	FJ461323	1	1090	EU482593	2
283	FJ461324	1	1091	EU482594	2
284	FJ461325	1	1092	EU482463	2

285	FJ410232	1	1093	EU482464	2
286	FJ461327	1	1094	EU482465	2
287	FJ410234	1	1095	EU482466	2
288	FJ410235	1	1096	EU482467	2
289	FJ410236	1	1097	EU482468	2
290	FJ410238	1	1098	EU482469	2
291	FJ410239	1	1099	EU482470	2
292	FJ410240	1	1100	EU482471	2
293	FJ410242	1	1101	EU482472	2
294	FJ410243	1	1102	EU482473	2
295	FJ410244	1	1103	EU482474	2
296	FJ410245	1	1104	EU482475	2
297	FJ461330	1	1105	EU482541	2
298	FJ410246	1	1106	EU482542	2
299	FJ410247	1	1107	EU482543	2
300	FJ410248	1	1108	AF489932	2
301	FJ410249	1	1109	AF100463	2
302	FJ410250	1	1110	AF100464	2
303	FJ410251	1	1111	AF100461	2
304	FJ410252	1	1112	AF100462	2
305	FJ410253	1	1113	AF119661	2
306	FJ410254	1	1114	FM210236	2
307	FJ461331	1	1115	FM210214	2
308	FJ410255	1	1116	AY702036	2
309	FJ410256	1	1117	AY702034	2
310	FJ410257	1	1118	AY702038	2
311	FJ461332	1	1119	AY702039	2
312	FJ410258	1	1120	AY702035	2
313	FJ461333	1	1121	AY702037	2
314	FJ410260	1	1122	EF105382	2
315	FJ410261	1	1123	EF105381	2
316	FJ461335	1	1124	EF105380	2
317	FJ410262	1	1125	EF105383	2
318	FJ410263	1	1126	EF105386	2
319	FJ410264	1	1127	EF105385	2
320	FJ410265	1	1128	EF457904	2
321	FJ461336	1	1129	FM210215	2
322	FJ410266	1	1130	FM210216	2
323	FJ410267	1	1131	FM210217	2
324	FJ410268	1	1132	FM210218	2
325	FJ410269	1	1133	FM210237	2
326	FJ410270	1	1134	FM210219	2

327	FJ410271	1	1135	FM210220	2
328	FJ410272	1	1136	FM210221	2
329	FJ410273	1	1137	FM210208	2
330	FJ410274	1	1138	FM210209	2
331	FJ410275	1	1139	FM210210	2
332	FJ410276	1	1140	FM210222	2
333	FJ410277	1	1141	FM210202	2
334	FJ410278	1	1142	FM210223	2
335	FJ410279	1	1143	FM210203	2
336	FJ410280	1	1144	FM210211	2
337	FJ410281	1	1145	AB122020	2
338	FJ410282	1	1146	AB122021	2
339	FJ410283	1	1147	AB122022	2
340	FJ410284	1	1148	EU179858	2
341	FJ410285	1	1149	EU179859	2
342	FJ410286	1	1150	EU179857	2
343	FJ461339	1	1151	AF276619	2
344	FJ461341	1	1152	AF359579	2
345	FJ639669	1	1153	AY702040	2
346	FJ639670	1	1154	EU056811	2
347	FJ639671	1	1155	AF100467	2
348	FJ639672	1	1156	AF100468	2
349	FJ639673	1	1157	M20558	2
350	FJ639674	1	1158	AF100459	2
351	FJ639675	1	1159	AF100460	2
352	FJ639676	1	1160	AF100466	2
353	FJ639677	1	1161	AF208496	2
354	FJ639678	1	1162	FM210231	2
355	FJ639679	1	1163	FM210239	2
356	FJ639680	1	1164	FM210232	2
357	FJ639681	1	1165	FM210233	2
358	FJ639682	1	1166	FM210240	2
359	FJ639683	1	1167	FM210241	2
360	FJ639684	1	1168	FM210242	2
361	FJ639685	1	1169	FM210243	2
362	FJ639686	1	1170	FM210244	2
363	FJ639687	1	1171	FM210234	2
364	FJ639688	1	1172	FM210213	2
365	FJ639689	1	1173	FM210235	2
366	FJ639690	1	1174	FM210245	2
367	FJ639691	1	1175	FM210205	2
368	FJ639692	1	1176	FM210206	2

369	FJ639693	1	1177	FM210246	2
370	FJ639694	1	1178	FM210207	2
371	FJ639695	1	1179	FM210238	2
372	FJ639696	1	1180	FM210224	2
373	FJ410173	1	1181	FM210225	2
374	FJ410174	1	1182	FM210226	2
375	FJ410175	1	1183	FM210227	2
376	FJ562105	1	1184	FM210212	2
377	FJ562106	1	1185	FM210204	2
378	FJ410179	1	1186	FM210228	2
379	FJ410180	1	1187	FM210229	2
380	FJ410181	1	1188	FM210230	2
381	FJ547086	1	1189	M29095	2
382	FJ410182	1	1190	AF038403	2
383	FJ410183	1	1191	AF038402	2
384	FJ410184	1	1192	EF105379	2
385	FJ410185	1	1193	DL138662	2
386	FJ547087	1	1194	CS477302	2
387	FJ410186	1	1195	CS479165	2
388	FJ410187	1	1196	CS479167	2
389	FJ478457	1	1197	FB730117	2
390	FJ410188	1	1198	GM059692	2
391	FJ410189	1	1199	U87412	2
392	FJ478458	1	1200	AJ968413	2
393	FJ410190	1	1201	EF105378	2
394	FJ639735	1	1202	M19197	2
395	FJ639740	1	1203	AY858036	2
396	FJ639741	1	1204	DQ181799	2
397	FJ639743	1	1205	DQ181802	2
398	FJ639794	1	1206	DQ181806	2
399	FJ639796	1	1207	DQ181798	2
400	FJ639797	1	1208	DQ181797	2
401	FJ639802	1	1209	DQ181805	2
402	FJ639806	1	1210	DQ181800	2
403	FJ639808	1	1211	DQ181801	2
404	FJ639811	1	1212	DQ181803	2
405	FJ639812	1	1213	DQ181804	2
406	FJ639813	1	1214	AF022434	2
407	FJ639814	1	1215	AF022437	2
408	FJ639815	1	1216	AF022438	2
409	FJ639818	1	1217	AF022439	2
410	FJ639819	1	1218	AF022440	2

411	FJ639820	1	1219	AF022435	2
412	FJ639821	1	1220	AF169678	2
413	FJ639823	1	1221	AF169679	2
414	FJ639824	1	1222	AF169680	2
415	FJ547088	1	1223	AF022436	2
416	FJ547068	1	1224	AF169682	2
417	FJ562104	1	1225	AF169681	2
418	FJ432721	1	1226	AF169683	2
419	FJ024483	1	1227	AF169684	2
420	FJ024478	1	1228	AF169685	2
421	EU596504	1	1229	AF169686	2
422	FJ024481	1	1230	AF169687	2
423	EU596502	1	1231	AF169688	2
424	FJ410290	1	1232	AF022441	2
425	FJ432720	1	1233	AY744147	2
426	FJ024423	1	1234	AY037116	2
427	FJ024485	1	1235	AY776328	2
428	FJ547089	1	1236	AF100465	2
429	FJ024484	1	1237	EU359009	2
430	EU596503	1	1238	EF051521	2
431	FJ024479	1	1239	EU081193	3
432	FJ182002	1	1240	EU081194	3
433	FJ024482	1	1241	EU081195	3
434	EU596501	1	1242	EU081196	3
435	FJ024480	1	1243	EU081197	3
436	EU482789	1	1244	EU081198	3
437	EU482790	1	1245	EU081199	3
438	EU482791	1	1246	EU081200	3
439	EU482792	1	1247	EU081201	3
440	EU482706	1	1248	EU081202	3
441	EU482707	1	1249	EU081203	3
442	EU482708	1	1250	EU081204	3
443	EU482709	1	1251	EU081205	3
444	EU482710	1	1252	EU081206	3
445	EU482711	1	1253	EU081207	3
446	EU482712	1	1254	EU081208	3
447	EU482713	1	1255	EU081209	3
448	EU482714	1	1256	EU081210	3
449	EU482715	1	1257	EU081211	3
450	EU482716	1	1258	EU081212	3
451	EU482717	1	1259	EU081213	3
452	EU482718	1	1260	EU081214	3

453	EU249490	1	1261	EU081215	3
454	EU249491	1	1262	EU081216	3
455	EU249492	1	1263	EU081217	3
456	EU249493	1	1264	EU081218	3
457	EU249494	1	1265	EU081219	3
458	EU249495	1	1266	EU081220	3
459	EU482793	1	1267	EU081221	3
460	EU482794	1	1268	EU081222	3
461	EU482795	1	1269	EU081223	3
462	EU482796	1	1270	EU081224	3
463	EU482797	1	1271	EU081225	3
464	EU482798	1	1272	EU081182	3
465	EU482799	1	1273	EU081183	3
466	EU482800	1	1274	EU081184	3
467	EU482801	1	1275	EU081185	3
468	EU482802	1	1276	EU081186	3
469	EU482803	1	1277	EU081187	3
470	EU482804	1	1278	EU081188	3
471	EU482805	1	1279	EU081189	3
472	EU482806	1	1280	EU081190	3
473	EU482807	1	1281	EU081191	3
474	EU482808	1	1282	EU081192	3
475	EU482809	1	1283	AY099337	3
476	EU482810	1	1284	NC_001475	3
477	EU482811	1	1285	AF317645	3
478	EU482812	1	1286	AY858038	3
479	EU482813	1	1287	DQ675519	3
480	EU482814	1	1288	AY858039	3
481	EU482815	1	1289	AB189125	3
482	EU482816	1	1290	AB189126	3
483	EU482817	1	1291	AB189127	3
484	EU482818	1	1292	AB189128	3
485	EU482819	1	1293	DQ675520	3
486	EU482820	1	1294	DQ675521	3
487	EU482821	1	1295	DQ675522	3
488	EU482822	1	1296	DQ675523	3
489	EU482823	1	1297	DQ675524	3
490	EU482824	1	1298	DQ675525	3
491	EU482825	1	1299	DQ675526	3
492	EU482826	1	1300	DQ675527	3
493	EU482827	1	1301	DQ675528	3
494	EU482828	1	1302	DQ675530	3

495	EU482591	1	1303	DQ675531	3
496	EU482592	1	1304	DQ675533	3
497	EU482476	1	1305	AY858037	3
498	EU482477	1	1306	AY496871	3
499	EU482478	1	1307	AY496873	3
500	EU482479	1	1308	AY496874	3
501	EU482480	1	1309	AY496877	3
502	EU482481	1	1310	EU687234	3
503	EU482482	1	1311	EU482452	3
504	EU482483	1	1312	EU482453	3
505	EU482484	1	1313	EU482454	3
506	EU482485	1	1314	EU482455	3
507	EU482486	1	1315	EU482456	3
508	EU482487	1	1316	EU482457	3
509	EU482488	1	1317	EU482458	3
510	EU482489	1	1318	EU482459	3
511	EU482490	1	1319	EU482460	3
512	EU482491	1	1320	EU482461	3
513	EU482492	1	1321	EU482462	3
514	EU482493	1	1322	EU482555	3
515	EU482494	1	1323	EU529692	3
516	EU482495	1	1324	EU482558	3
517	EU482496	1	1325	EU482559	3
518	EU482497	1	1326	EU482563	3
519	EU482498	1	1327	EU529696	3
520	EU482499	1	1328	EU529697	3
521	EU482500	1	1329	EU482564	3
522	EU482501	1	1330	EU529698	3
523	EU482502	1	1331	EU529699	3
524	EU482503	1	1332	EU482566	3
525	EU482504	1	1333	EU529702	3
526	EU482505	1	1334	EU529703	3
527	EU482506	1	1335	EU529704	3
528	EU482507	1	1336	EU529705	3
529	EU482508	1	1337	EU529683	3
530	EU482509	1	1338	EU529684	3
531	EU482510	1	1339	FJ182015	3
532	EU482511	1	1340	EU569688	3
533	EU482512	1	1341	EU529685	3
534	EU482513	1	1342	EU529686	3
535	EU482514	1	1343	EU529687	3
536	EU482515	1	1344	EU660407	3

537	EU482516	1	1345	EU660408	3
538	EU482517	1	1346	EU660409	3
539	EU482518	1	1347	EU660410	3
540	EU482519	1	1348	EU660411	3
541	EU482520	1	1349	EU596492	3
542	EU482521	1	1350	EU596493	3
543	EU482522	1	1351	EU596494	3
544	EU482523	1	1352	EU687218	3
545	EU482524	1	1353	EU726771	3
546	EU482525	1	1354	EU726772	3
547	EU482526	1	1355	FJ182013	3
548	EU482527	1	1356	EU781136	3
549	EU482528	1	1357	EU781137	3
550	EU482529	1	1358	EU726773	3
551	EU482530	1	1359	EU726774	3
552	EU482531	1	1360	EU687219	3
553	EU482532	1	1361	EU687221	3
554	EU482533	1	1362	EU726768	3
555	EU482534	1	1363	EU687226	3
556	EU482535	1	1364	EU687196	3
557	EU482536	1	1365	EU854298	3
558	EU482537	1	1366	FJ373306	3
559	EU482538	1	1367	EU687197	3
560	EU482539	1	1368	EU687198	3
561	EU482540	1	1369	EU726769	3
562	AF513110	1	1370	EU687239	3
563	AF226685	1	1371	FJ373303	3
564	AF311956	1	1372	FJ373304	3
565	AF311958	1	1373	EU854291	3
566	AF311957	1	1374	EU854292	3
567	AF309641	1	1375	FJ182004	3
568	EU863650	1	1376	FJ024465	3
569	DQ285562	1	1377	FJ024466	3
570	FJ176779	1	1378	FJ024467	3
571	FJ176780	1	1379	FJ024468	3
572	FJ384655	1	1380	FJ024469	3
573	EU280167	1	1381	FJ024470	3
574	AF226687	1	1382	FJ024471	3
575	EF122232	1	1383	FJ182005	3
576	AF226686	1	1384	FJ373302	3
577	EF122231	1	1385	FJ182006	3
578	DQ193572	1	1386	FJ182007	3

579	DQ672556	1	1387	FJ182008	3
580	DQ672557	1	1388	FJ182009	3
581	DQ672558	1	1389	FJ182010	3
582	DQ672559	1	1390	FJ182011	3
583	EF032590	1	1391	FJ182037	3
584	AF350498	1	1392	FJ182038	3
585	AF298808	1	1393	FJ182039	3
586	DQ672560	1	1394	FJ182040	3
587	DQ672562	1	1395	FJ182041	3
588	DQ672561	1	1396	FJ390371	3
589	DQ672564	1	1397	FJ390372	3
590	DQ672563	1	1398	FJ205870	3
591	AB204803	1	1399	FJ390373	3
592	AB074760	1	1400	FJ205871	3
593	AY708047	1	1401	FJ390375	3
594	AY713474	1	1402	FJ390376	3
595	AY713475	1	1403	FJ390377	3
596	AY722802	1	1404	FJ432722	3
597	AY713476	1	1405	FJ432728	3
598	AY726555	1	1406	FJ432731	3
599	AY726554	1	1407	FJ562097	3
600	AY722803	1	1408	FJ432741	3
601	AY726549	1	1409	FJ562099	3
602	AY726550	1	1410	FJ432743	3
603	AY713473	1	1411	FJ562100	3
604	AY722801	1	1412	FJ547061	3
605	AY726551	1	1413	FJ547062	3
606	AY726552	1	1414	FJ461322	3
607	AY726553	1	1415	FJ562102	3
608	EF457905	1	1416	FJ410229	3
609	CS477263	1	1417	FJ461326	3
610	CS477264	1	1418	FJ461329	3
611	CS477265	1	1419	FJ547066	3
612	CS479203	1	1420	FJ461334	3
613	CS479204	1	1421	FJ461337	3
614	FB667398	1	1422	FJ461338	3
615	FB730116	1	1423	FJ562103	3
616	GM059691	1	1424	FJ639712	3
617	DQ285558	1	1425	FJ639713	3
618	DQ285559	1	1426	FJ639714	3
619	AY145123	1	1427	FJ639715	3
620	AY145122	1	1428	FJ639716	3

621	AB195673	1	1429	FJ639719	3
622	DQ285561	1	1430	FJ639720	3
623	AY762084	1	1431	FJ639721	3
624	AY732483	1	1432	FJ639722	3
625	AY732482	1	1433	FJ639723	3
626	AY732481	1	1434	FJ639724	3
627	AY732480	1	1435	FJ639725	3
628	AY732479	1	1436	FJ639726	3
629	AY732478	1	1437	FJ639727	3
630	AY732477	1	1438	FJ639728	3
631	AY732476	1	1439	FJ639729	3
632	AY732475	1	1440	FJ639730	3
633	AY732474	1	1441	FJ639731	3
634	EU848545	1	1442	FJ547069	3
635	AY145121	1	1443	FJ547070	3
636	NC_001477	1	1444	FJ562107	3
637	U88535	1	1445	FJ547071	3
638	U88537	1	1446	FJ410176	3
639	EU359008	1	1447	FJ410177	3
640	AY835999	1	1448	FJ547072	3
641	EU081177	2	1449	FJ547073	3
642	EU081178	2	1450	FJ547074	3
643	EU081179	2	1451	FJ547075	3
644	EU081180	2	1452	FJ547076	3
645	DQ645543	2	1453	FJ547077	3
646	DQ645544	2	1454	FJ547078	3
647	DQ645545	2	1455	FJ547079	3
648	DQ645546	2	1456	FJ547080	3
649	AF100469	2	1457	FJ547081	3
650	EU056810	2	1458	FJ547082	3
651	DQ645547	2	1459	FJ547083	3
652	DQ645548	2	1460	FJ410178	3
653	M84727	2	1461	FJ547084	3
654	NC_001474	2	1462	FJ478456	3
655	M84728	2	1463	FJ547085	3
656	DQ645549	2	1464	FJ639746	3
657	DQ645550	2	1465	FJ639747	3
658	DQ645551	2	1466	FJ639749	3
659	DQ645552	2	1467	FJ639750	3
660	DQ645553	2	1468	FJ639751	3
661	DQ645554	2	1469	FJ639752	3
662	DQ645555	2	1470	FJ639753	3

663	DQ645556	2	1471	FJ639754	3
664	AF204178	2	1472	FJ639755	3
665	AF204177	2	1473	FJ639756	3
666	DQ645540	2	1474	FJ639757	3
667	DQ645541	2	1475	FJ639758	3
668	DQ645542	2	1476	FJ639759	3
669	AB189122	2	1477	FJ639760	3
670	AB189123	2	1478	FJ639761	3
671	AB189124	2	1479	FJ639762	3
672	AY858035	2	1480	FJ639763	3
673	AJ487271	2	1481	FJ639765	3
674	EU687249	2	1482	FJ639766	3
675	EU482760	2	1483	FJ639767	3
676	EU482445	2	1484	FJ639768	3
677	EU482446	2	1485	FJ639769	3
678	EU482447	2	1486	FJ639770	3
679	EU482448	2	1487	FJ639771	3
680	EU482449	2	1488	FJ639772	3
681	EU482450	2	1489	FJ639774	3
682	EU569721	2	1490	FJ639775	3
683	EU482451	2	1491	FJ639776	3
684	EU482544	2	1492	FJ639777	3
685	EU482545	2	1493	FJ639778	3
686	EU482546	2	1494	FJ639779	3
687	EU482547	2	1495	FJ639780	3
688	EU482548	2	1496	FJ639781	3
689	EU482549	2	1497	FJ639782	3
690	EU482550	2	1498	FJ639784	3
691	EU482551	2	1499	FJ639785	3
692	EU482552	2	1500	FJ639786	3
693	EU482553	2	1501	FJ639787	3
694	EU482554	2	1502	FJ639789	3
695	EU482556	2	1503	FJ639790	3
696	EU529693	2	1504	FJ639791	3
697	EU482557	2	1505	FJ639792	3
698	EU482560	2	1506	FJ639793	3
699	EU482561	2	1507	FJ639795	3
700	EU529694	2	1508	FJ639798	3
701	EU529695	2	1509	FJ639799	3
702	EU482562	2	1510	FJ639800	3
703	EU660398	2	1511	FJ639801	3
704	EU482565	2	1512	FJ639803	3

705	EU529700	2	1513	FJ639804	3
706	EU529701	2	1514	FJ639805	3
707	EU482604	2	1515	FJ639807	3
708	EU482605	2	1516	FJ639810	3
709	EU482606	2	1517	FJ639816	3
710	EU482607	2	1518	FJ639817	3
711	EU482608	2	1519	FJ639825	3
712	EU569704	2	1520	FJ639826	3
713	EU482568	2	1521	FJ639827	3
714	EU482569	2	1522	EU482595	3
715	EU482570	2	1523	EU482596	3
716	EU482571	2	1524	EU529688	3
717	EU482572	2	1525	EU482612	3
718	EU482573	2	1526	EU660420	3
719	EU482574	2	1527	EU482613	3
720	EU482575	2	1528	EU529689	3
721	EU482576	2	1529	EU529690	3
722	EU482577	2	1530	EU529691	3
723	EU482578	2	1531	EU569689	3
724	EU482579	2	1532	EU482614	3
725	EU482580	2	1533	EU569690	3
726	EU482581	2	1534	EU569691	3
727	EU482582	2	1535	EF629369	3
728	EU482583	2	1536	EF629366	3
729	EU482584	2	1537	EF629367	3
730	EU482585	2	1538	EF629368	3
731	EU482586	2	1539	EF629370	3
732	EU482587	2	1540	AY679147	3
733	EU482588	2	1541	AY876494	3
734	EU482621	2	1542	AY923865	3
735	EU569692	2	1543	EF643017	3
736	EU569693	2	1544	AY858040	3
737	FJ373300	2	1545	M93130	3
738	EU596495	2	1546	AB214879	3
739	EU660404	2	1547	AB214880	3
740	EU482622	2	1548	AB214881	3
741	EU569694	2	1549	AB214882	3
742	EU596496	2	1550	DQ401690	3
743	EU660405	2	1551	AY858043	3
744	EU596497	2	1552	DQ401689	3
745	EU569695	2	1553	DQ401691	3
746	FJ373301	2	1554	DQ401692	3

747	EU569696	2	1555	DQ401693	3
748	EU660406	2	1556	DQ401694	3
749	EU569697	2	1557	DQ401695	3
750	EU482623	2	1558	AY744677	3
751	EU596498	2	1559	AY744678	3
752	EU569698	2	1560	AY744679	3
753	FJ478459	2	1561	AY744680	3
754	EU596499	2	1562	AY744681	3
755	EU569699	2	1563	AY744682	3
756	EU569700	2	1564	AY744683	3
757	EU569701	2	1565	AY744684	3
758	EU621672	2	1566	AY744685	3
759	EU569702	2	1567	AY496879	3
760	EU482624	2	1568	AY858046	3
761	EU596500	2	1569	AY766104	3
762	EU482625	2	1570	AY648961	3
763	FJ410291	2	1571	EU081181	3
764	EU482626	2	1572	AY858047	3
765	EU482627	2	1573	AY676353	3
766	EU482628	2	1574	AY676352	3
767	FJ639833	2	1575	AY676351	3
768	EU482629	2	1576	AY676350	3
769	FJ390390	2	1577	AY676349	3
770	FJ390391	2	1578	AY676348	3
771	FJ182014	2	1579	AY776329	3
772	FJ547090	2	1580	AF375822	4
773	EU482630	2	1581	AF326826	4
774	EU569705	2	1582	AF326573	4
775	EU482589	2	1583	AF289029	4
776	EU569706	2	1584	FJ024424	4
777	EU569707	2	1585	EU854295	4
778	EU660399	2	1586	EU854296	4
779	EU569708	2	1587	EU854297	4
780	EU569709	2	1588	EU854299	4
781	EU596485	2	1589	FJ182016	4
782	EU569710	2	1590	EU854300	4
783	EU596486	2	1591	FJ182017	4
784	EU687212	2	1592	EU854301	4
785	EU660400	2	1593	FJ024476	4
786	EU569711	2	1594	FJ639736	4
787	EU569712	2	1595	FJ639737	4
788	EU596487	2	1596	FJ639738	4

789	EU569713	2	1597	FJ639739	4
790	EU569714	2	1598	FJ639742	4
791	EU569715	2	1599	FJ639744	4
792	EU569716	2	1600	FJ639745	4
793	EU569717	2	1601	FJ639748	4
794	EU569718	2	1602	FJ639764	4
795	EU569719	2	1603	FJ639773	4
796	EU569720	2	1604	FJ226067	4
797	EU596488	2	1605	AY947539	4
798	EU596489	2	1606	EF457906	4
799	EU596490	2	1607	NC_002640	4
800	EU596491	2	1608	AF326827	4
801	EU677141	2	1609	AY618989	4
802	EU677142	2	1610	AY618991	4
803	EU677143	2	1611	AY618990	4
804	EU677144	2	1612	AY618988	4
805	EU677145	2	1613	AY618992	4
806	EU677146	2	1614	AY618993	4
807	EU677147	2	1615	M14931	4
808	EU687213	2			

Table A.2. List of CTL epitopes retrieved from the Immune Epitope Database.

	Epitopes	Protein	Start	End	Serotype
1	RVSTVQQLTK	AnchC	22	31	2
2	GPMKLVMAF	AnchC	42	50	1, 3
3	GPMKLVMAFI	AnchC	42	51	1, 3
4	KLVMAFIAFLRFL	AnchC	45	57	1, 3
5	LFMALVAFLRFLTIP	AnchC	46	60	2
6	VLAFITFLR	AnchC	47	55	4
7	FMALVAFLRF	AnchC	47	56	2
8	MAFIAFLRF	AnchC	48	56	1, 3
9	IAFLRFLAI	AnchC	51	59	1, 3
10	TIKKSKAINVLRGFR	AnchC	71	85	2
11	KSGAIKVLK	AnchC	74	82	3
12	LRGFKKEISNML	AnchC	81	92	1, 3
13	LRGFRKEIGRML	AnchC	81	92	2
14	LKGFKKEISNML	AnchC	81	92	3
15	LIGFRKEIGRML	AnchC	81	92	4
16	FRKEIGRML	AnchC	84	92	2, 4
17	ISSMLNIMNR	AnchC	88	97	1
18	ITLLCLIPTV	AnchC	103	112	4
19	VTLLCLIPTV	AnchC	103	112	4
20	TLLCLIPTV	AnchC	104	112	4
21	CLMMMLPATL	AnchC	104	113	3
22	LMMMLPATL	AnchC	105	113	3
23	LMMILPAAL	AnchC	105	113	3
24	LMMMLPATLA	AnchC	105	114	3
25	MMMLPATLA	AnchC	106	114	3
26	MMMLPATLAF	AnchC/prM	106	115	3
27	MLIPTAMAF	AnchC/prM	107	115	2
28	MMLPATLAF	AnchC/prM	107	115	3
29	CLIPTAMAF	AnchC/prM	107	115	4
30	IPTVMAFHL	AnchC/prM	109	117	2, 4
31	TLMAMDLGEL	prM	149	158	2
32	VTYECPLLV	prM	163	171	4
33	MSSEGAWKHA	prM	225	234	2, 4
34	QRIETWILRHPGFTM	prM	235	249	2
35	WILRHPGFTMMAAIL	prM	240	254	2
36	HPGFTILALF	prM	244	253	3
37	FTIMAAILAY	prM	247	256	2
38	FTILALFLAH	prM	247	256	3

39	TLMAAILAY	prM	248	256	2
40	TIMAAILAY	prM	248	256	2
41	LAYTIGTTHF	prM	254	263	2
42	LLMLVTPSM	prM	271	279	1, 3
43	MLVTPSMTM	prM/E	273	281	1, 3
44	KPTLDFELI	E	318	326	2, 4
45	LDFELIKTEAKQPAT	E	321	335	2
46	LKTEVTNPAV	E	326	335	1
47	CPTQGEATL	E	354	362	1, 3
48	CPTQGEPSL	E	354	362	2
49	CPTQGEAVL	E	354	362	3
50	LPEEQDQNY	E	362	370	3
51	WGNGCGLFGKGGIVT	E	381	395	2
52	EPIEGKVVQY	E	403	412	3
53	YENLKYSVI	E	412	420	1
54	MENKAWLVHRQWFLD	E	481	495	2
55	WLVHRQWFLDLPLPW	E	486	500	2
56	FFDLPLPWT	E	493	501	3
57	FLDLPLPWL	E	493	501	2
58	FLDLPLPWT	E	493	501	1, 3, 4
59	LPLPWTSGA	E	496	504	1, 3
60	LPLPWLPGAD	E	496	505	2
61	LPGADTQGSN	E	501	510	2
62	TLVTFKNPHAKKQDV	E	516	530	2
63	TFKVPHAKR	E	519	527	4
64	KNPHAKKQDVVVLGS	E	521	535	2
65	KKQDVVVLGSQEGAM	E	527	540	2
66	QEGAMHTAL	E	536	544	1, 2, 3
67	QEGAMHSAL	E	536	544	4
68	MSYTMCSGK	E	577	585	4
69	MSYSMCTGKF	E	577	586	2
70	SPCKIPFEIM	E	611	620	2
71	IPFEIMDLEK	E	615	624	2
72	RDVNKEKVVGRVISSTPLAE	E	620	639	4
73	IVIGVGDSAL	E	658	667	4
74	EPGQLKLNWF	E	663	672	2
75	SSIGKMFEATARG	E	676	688	1
76	RMAILGDTAWDFGSL	E	691	705	2
77	MAILGDTAW	E	692	700	1, 2, 3
78	ILGDTAWDFG	E	694	703	1, 2, 3
79	LIHQVFGTAY	E	715	724	1

80	LVHQIFGTAY	E	715	724	1
81	FGAIYGAAF	E	720	729	2
82	YGVLFSGVSW	E	724	733	1
83	SWMVRILIGF	E	732	741	4
84	KILIGVIITWIGMNS	E	736	750	2
85	IGIGILLTW	E	737	745	1, 3
86	IGIGVLLTW	E	737	745	1, 3
87	MAMTCIAVG	E	755	763	4
88	SLVLVGVVTL	E	758	767	2
89	VQADMGCVV	E/NS1	773	781	3, 4
90	SPKRLATAIA	NS1	813	822	3
91	KQIANELNY	NS1	844	852	3
92	KQISNELNY	NS1	844	852	1
93	IMTGDIKGIMQA	NS1	863	874	2
94	LKYSWKTWGKAKMLS	NS1	886	900	2
95	TPEARNSTF	NS1	900	908	4
96	TPEAKNSTF	NS1	900	908	4
97	GVFTTNIWLKLKEKQ	NS1	936	950	2
98	VYTQLCDHR	NS1	949	957	3
99	VYTQLCDHRL	NS1	949	958	3
100	NRAVHADMGYWIESA	NS1	966	980	2
101	KAVHADMGYW	NS1	967	976	1, 3
102	RAVHADMGYW	NS1	967	976	1, 2, 3, 4
103	MGYWIESAL	NS1	973	981	2
104	LNDTWKIEKASFIEV	NS1	981	995	2
105	RASFIEVKTC	NS1	989	998	1
106	IPKAYAGPF	NS1	1018	1026	4
107	YGGPISQHNY	NS1	1022	1031	1
108	FAGPVSQHNY	NS1	1022	1031	2
109	RPGYHTQTA	NS1	1032	1040	2, 3
110	GPWHLGKLEL	NS1	1041	1050	1, 3
111	GPWHLGKLEM	NS1	1041	1050	2, 4
112	GPSLRTTTV	NS1	1070	1078	1, 3
113	RYMGEDGCWY	NS1	1097	1106	3
114	RPISEKEENM	NS1	1111	1120	3
115	FTMGVLCLAI	NS2A	1135	1144	3
116	RVGTKHAILLVAVSF	NS2A	1153	1167	2
117	HMIAGVLFTF	NS2A	1158	1167	3
118	MSFRDLGRVM	NS2A	1175	1184	2
119	RDLGRVMVMVGATMT	NS2A	1178	1192	2
120	TYLALIATF	NS2A	1200	1208	1, 3

121	IQPFLALGF	NS2A	1210	1218	3
122	GFFLRKLTSR	NS2A	1217	1226	3
123	LMMTTIGVVL	NS2A	1228	1237	2
124	LMMATIGIAL	NS2A	1228	1237	2
125	MMATIGIAL	NS2A	1229	1237	2
126	MMLKLLTDF	NS2A	1259	1267	1
127	HQLWATLLSL	NS2A	1270	1279	1
128	YQLAVTITAI	NS2A	1270	1279	2
129	YQLWTALISL	NS2A	1270	1279	3
130	TTFSLHYAW	NS2A	1284	1292	1
131	MPLVMAWRTI	NS2A	1286	1295	4
132	AMALSIVSLF	NS2A	1296	1305	1
133	MALSIVSLF	NS2A	1297	1305	1
134	GASKRSWPLN	NS2A/NS2B	1342	1351	4
135	MAVGMVSIL	NS2B	1355	1363	2
136	IPMTGPLVAG	NS2B	1372	1381	2
137	GPLVAGGLL	NS2B	1376	1384	2, 3
138	ELERAADVK	NS2B	1398	1406	2
139	SPILSITISE	NS2B	1417	1426	2
140	ILIRTGLLVI	NS2B	1443	1452	2
141	ISGLFPVSIPITAAA	NS2B	1452	1466	2
142	FPVSIPITAA	NS2B	1456	1465	2
143	IPITAAAWY	NS2B	1460	1468	2
144	ITAAAWYLW	NS2B	1462	1470	2
145	YRILQRGLLGRSQ	NS3	1499	1511	1
146	RIKQKGIL	NS3	1500	1507	2
147	SQIGAGVYK	NS3	1510	1518	2
148	QKEGVFHTMW	NS3	1517	1526	3
149	KEGVFHTMW	NS3	1518	1526	3
150	MEGVFHTMW	NS3	1518	1526	4
151	MEGVFHTMWH	NS3	1518	1527	4
152	TFHTMWHVTRGAVLM	NS3	1521	1535	2
153	VTRGAVLMHK	NS3	1528	1537	2
154	RIEPSWADVK	NS3	1540	1549	2
155	SVKKDLISY	NS3	1547	1555	1, 3
156	DVKKDLISY	NS3	1547	1555	2
157	GGWRLSAQW	NS3	1557	1565	3
158	EGEWKEGEEVQVLAL	NS3	1562	1576	2
159	AVQTKPGLFK	NS3	1584	1593	2
160	MPGTFQTTTG	NS3	1588	1597	3
161	GEVGAIALDF	NS3	1597	1606	1

162	KPGTSGSPI	NS3	1607	1615	1, 3, 4
163	KPGTSGSPIV	NS3	1607	1616	1
164	KPGTSGSPII	NS3	1607	1616	3, 4
165	SPGTSGSPIIDKKGK	NS3	1607	1621	2
166	GTSGSPIADK	NS3	1609	1618	2
167	GTSGSPIIDK	NS3	1609	1618	2
168	GTSGSPIINR	NS3	1609	1618	3, 4
169	GTSGSPIVDK	NS3	1609	1618	2
170	GTSGSPIVDR	NS3	1609	1618	2
171	GTSGSPIVNR	NS3	1609	1618	1
172	GTSGSPIADKK	NS3	1609	1619	2
173	GTSGSPIIDKK	NS3	1609	1619	2
174	GTSGSPIINRE	NS3	1609	1619	3
175	GTSGSPIINRK	NS3	1609	1619	4
176	TSGSPIIDK	NS3	1610	1618	2
177	SPIINREGKV	NS3	1613	1622	3
178	DKKGKVVGL	NS3	1617	1625	2
179	NREGKIVGLYGNGVV	NS3	1617	1631	1, 3
180	NREGKVVGLYGNGVV	NS3	1617	1631	1, 3
181	DKKGKVVGLYGNGVV	NS3	1617	1631	2
182	KVVGLYGNGVVTRSG	NS3	1621	1635	2
183	VIGLYGNGV	NS3	1622	1630	4
184	YVSAIAQTEK	NS3	1637	1646	2
185	EPDYEVDEDI	NS3	1649	1659	4
186	NPEIEDDIF	NS3	1652	1660	2
187	RNLTIMDLHPGSGKT	NS3	1663	1677	1, 3
188	RKLTIMDLHPGSGKT	NS3	1663	1677	1, 3
189	RKLTIMDLHPGAGKT	NS3	1663	1677	2
190	KRLTIMDLHPGAGKT	NS3	1663	1677	2, 4
191	HPGAGKTKRY	NS3	1671	1680	2
192	TKRYLPAIVREAIKR	NS3	1677	1691	1, 2
193	RKYLPAIVRE	NS3	1678	1687	1, 3
194	KRYLPAIVREAIKRG	NS3	1678	1692	2
195	LPAIVREAI	NS3	1681	1689	1, 2, 3
196	AIVREAIKR	NS3	1683	1691	1, 2, 3
197	AIKRGLRTL	NS3	1688	1696	2
198	LAPTRVVAAEME	NS3	1698	1709	2, 3, 4
199	LAPTRVVAAEMEEAL	NS3	1698	1712	2, 3, 4
200	APTRVVAAEM	NS3	1699	1708	1, 2, 3, 4
201	APTRVVASEM	NS3	1699	1708	1
202	PTRVVASEMAEALKG	NS3	1700	1714	1

203	PTRVVAAEMEEALRG	NS3	1700	1714	2, 4
204	PTRVVAAEMEEAMKG	NS3	1700	1714	3
205	TRVVAAEMEEA	NS3	1701	1711	2, 3, 4
206	SEMAEALKGM	NS3	1706	1715	1
207	MEEALRGLPIRYQTP	NS3	1708	1722	2, 4
208	EALRGLPIR	NS3	1710	1718	2, 3, 4
209	ALRGLPIRY	NS3	1711	1719	2, 3, 4
210	AMKGLPIRY	NS3	1711	1719	3
211	LPIRYQTPA	NS3	1715	1723	2, 4
212	LPIRYQTPAI	NS3	1715	1724	2
213	PIRYQ	NS3	1716	1720	1, 2, 3, 4
214	IRYQTTATK	NS3	1717	1725	3
215	EHTGREIVDLMCHAT	NS3	1727	1741	1, 2, 3, 4
216	HTGREIVDLMCHATF	NS3	1728	1742	1, 2, 3, 4
217	EIVDLMCHAT	NS3	1732	1741	1, 2, 3, 4
218	EIVDLMCHATFTMRL	NS3	1732	1746	1, 2, 3
219	VDLMCHATFT	NS3	1734	1743	1, 2, 3, 4
220	LMCHATFTM	NS3	1736	1744	1, 2, 3
221	LLSPVRVPNY	NS3	1746	1755	1, 2, 3
222	SPVRVPNYNL	NS3	1748	1757	2, 3
223	VPNYNMIIM	NS3	1752	1760	1
224	PNYNLIIMDEAHFTD	NS3	1753	1767	2, 3
225	DPASIAARGY	NS3	1767	1776	1, 2, 3
226	PASIAARGYI	NS3	1768	1777	1, 2, 3
227	STRVEMGEAAGIFMT	NS3	1778	1792	2
228	GMGEAAAIF	NS3	1782	1790	1, 3
229	TPPGSRDPF	NS3	1794	1802	2
230	FPQSNAPIM	NS3	1802	1810	2
231	FPQSNAPIMD	NS3	1802	1811	2
232	APIMDEEREI	NS3	1807	1816	2
233	EERDIPERSW	NS3	1812	1821	1, 3
234	EREIPERSWNSGHEW	NS3	1813	1827	2
235	REIPERSWNT	NS3	1814	1823	2, 4
236	RSWNSGHEW	NS3	1819	1827	1, 2
237	GNEWITDFVGKTVWF	NS3	1824	1838	3
238	WITDFVGKTVW	NS3	1827	1837	3
239	VTDFKGKTVWFVPSI	NS3	1828	1842	2
240	TVWFVPSIK	NS3	1835	1843	1, 2, 3, 4
241	KVIQLSRKTF	NS3	1858	1867	2, 3, 4
242	YPKTKLTDW	NS3	1871	1879	4
243	YPKTKLTDWD	NS3	1871	1880	4

244	RVIDPRRCL	NS3	1898	1906	1, 3, 4
245	RVIDPRRCM	NS3	1898	1906	2
246	RVIDPRRCMK	NS3	1898	1907	2
247	RVIDPRRCLK	NS3	1898	1907	1, 3, 4
248	DPRRCLKPV	NS3	1901	1909	1, 3, 4
249	MPVTHSSAA	NS3	1924	1932	2
250	MPVTVASAAQ	NS3	1924	1933	1, 3
251	MPVTHSSAAQ	NS3	1924	1933	2
252	NPAQEDDQY	NS3	1941	1949	4
253	QYIFTGQPL	NS3	1948	1956	3
254	DNINTPEGIIPSMFE	NS3	1973	1987	2
255	YTPEGIIPTL	NS3	1976	1985	4
256	TPEGIIPAL	NS3	1977	1985	1, 3
257	TPEGIIPSL	NS3	1977	1985	2
258	TPEGIIPSM	NS3	1977	1985	2
259	TPEGIIPTL	NS3	1977	1985	4
260	TPEGIIPALF	NS3	1977	1986	1, 3
261	TPEGIIPSMF	NS3	1977	1986	2
262	TPEGIIPTLF	NS3	1977	1986	4
263	GEFRLRGEQR	NS3	1998	2007	4
264	RLRGEARKTF	NS3	2001	2010	1, 2
265	RGEARKTFVDLMRRG	NS3	2003	2017	2
266	GESRKTFVE	NS3	2004	2012	3
267	GEARKTFVDL	NS3	2004	2013	1, 2
268	GEARKTFVEL	NS3	2004	2013	1, 2
269	GESRKTFVEL	NS3	2004	2013	3
270	GEQRKTFVEL	NS3	2004	2013	4
271	KTFVDLMRR	NS3	2008	2016	2
272	ELMRRGDLPV	NS3	2012	2021	1, 2, 3, 4
273	LMRRGDLPVWLAYRV	NS3	2013	2027	2
274	LPVWLSYKV	NS3	2019	2027	1, 4
275	LPVWLAYRV	NS3	2019	2027	2
276	LPVWLAYKV	NS3	2019	2027	2, 3
277	LPVWLAYRVA	NS3	2019	2028	2
278	LPVWLAYKVA	NS3	2019	2028	2, 3
279	LPVWLAHKVA	NS3	2019	2028	3
280	YKVASAGISY	NS3	2025	2034	4
281	AAEGINYADRRWCFD	NS3	2028	2042	2
282	INYADRRWCF	NS3	2032	2041	2
283	FQYSDRRWCF	NS3	2032	2041	1
284	KYTDRKWCF	NS3	2033	2041	3

285	NYADRKWCF	NS3	2033	2041	2
286	NYADRRWCF	NS3	2033	2041	2
287	QYSDRRWCF	NS3	2033	2041	1
288	SYKDREWCF	NS3	2033	2041	4
289	REWCFTGERN	NS3	2037	2046	4
290	EENMDVEIW	NS3	2051	2059	1, 3
291	KEGERKKLRPRWLDA	NS3	2061	2075	1, 3
292	KEGERKKLKPRWLDA	NS3	2061	2075	2
293	KEGEKKKLRPRWLDA	NS3	2061	2075	3
294	RPRWLDART	NS3	2069	2077	1, 3
295	KPRWLDARI	NS3	2069	2077	2
296	LDARIYSDPLALKEF	NS3	2073	2087	2
297	MALKDFKEF	NS3	2082	2090	4
298	EFKEFAAGR	NS3	2086	2094	1, 2
299	EFKEFAAGRR	NS3	2086	2095	1
300	FASGRKSITL	NS3/NS4A	2090	2099	4
301	LPTFMTQKAR	NS4A	2108	2117	2
302	LPTYLSSRAK	NS4A	2108	2117	4
303	MTQKARNAL	NS4A	2112	2120	2
304	TAEAGGRAY	NS4A	2128	2136	2
305	ALSELPETL	NS4A	2139	2147	2
306	LPESLETLM	NS4A	2143	2151	4
307	LPETLETLLL	NS4A	2143	2152	2
308	LETLMLVAL	NS4A	2147	2155	4
309	LETLMLVALL	NS4A	2147	2156	4
310	TLMLLALIAV	NS4A	2149	2158	1
311	LMLLALIAVL	NS4A	2150	2159	1
312	LLLLTLLATV	NS4A	2150	2159	2
313	LLLGLMILL	NS4A	2151	2159	3
314	LLLTLLATV	NS4A	2151	2159	2
315	MLLALIAVL	NS4A	2151	2159	1
316	MLVALLGAM	NS4A	2151	2159	4
317	GAMLFLISGK	NS4A	2162	2171	3
318	LSMGLITIAV	NS4A	2176	2185	4
319	MASSALLWMA	NS4A	2184	2193	1
320	ASSVLLWMAS	NS4A	2185	2194	1
321	VASGLLWVAE	NS4A	2185	2194	4
322	MLWMAEIPL	NS4A	2189	2197	3
323	WVAEIQPQW	NS4A	2191	2199	4
324	YAQIQPHWI	NS4A	2192	2200	2
325	MAEIPLQWI	NS4A	2192	2200	3

326	QPHWIAASI	NS4A	2196	2204	2
327	QPHWIAASII	NS4A	2196	2205	2
328	SIILEFFLMV	NS4A	2203	2212	1, 4
329	IILEFFLMV	NS4A	2204	2212	1, 2, 4
330	IILEFFLMVL	NS4A	2204	2213	1, 2, 4
331	FLMVLLIPEP	NS4A	2209	2218	1, 2, 4
332	TPQDNQLAY	2K	2223	2231	1, 3
333	TPQDNQLTY	2K	2223	2231	2
334	TPQDNQLAYV	2K	2223	2232	1, 3
335	NQLIYVILTI	2K	2227	2236	4
336	MANEMGFLEK	2K/NS4B	2244	2253	2
337	LEKTKKDL	NS4B	2251	2258	2
338	TTKRDLGMSK	NS4B	2253	2262	3
339	KTDFGFYQV	NS4B	2255	2263	4
340	EPGVVSPTSY	NS4B	2263	2273	3
341	QPESNILDI	NS4B	2268	2276	2
342	TETTILDVDL	NS4B	2269	2278	4
343	RPASAWTLY	NS4B	2279	2287	1, 2, 4
344	HPASAWTLY	NS4B	2279	2287	1, 3
345	RPASAWTLYA	NS4B	2279	2288	1, 2, 4
346	PASAWTLYAV	NS4B	2280	2289	1, 2, 3, 4
347	TLYAVATTI	NS4B	2285	2293	1, 3, 4
348	VATTFVTPM	NS4B	2289	2297	2
349	ITPMLRHTI	NS4B	2294	2302	1, 3
350	TPMLRHTIEN	NS4B	2295	2304	1, 3, 4
351	IANQAAILM	NS4B	2314	2322	1
352	IANQATVLM	NS4B	2314	2322	2
353	VPLLAIGCY	NS4B	2337	2345	2
354	VPLLAMGCY	NS4B	2337	2345	4
355	NPLTLTAAV	NS4B	2349	2357	1, 3
356	NPITLTAAL	NS4B	2349	2357	2
357	TLTAAVLLLV	NS4B	2352	2361	3
358	LTAAVLLLI	NS4B	2353	2361	3
359	TAAVLLLITH	NS4B	2354	2363	3
360	AAVLLLVTHY	NS4B	2355	2364	3
361	LVMLLVHYA	NS4B	2357	2365	4
362	VLMLVAHYA	NS4B	2357	2365	1
363	VLLLVTHYA	NS4B	2357	2365	3
364	VLLLVTHYAI	NS4B	2357	2366	3
365	FLLVAHYAI	NS4B	2358	2366	2
366	AIIGPGLQAK	NS4B	2365	2374	1, 2, 3, 4

367	KATREAQKRA	NS4B	2374	2383	2
368	NPTVDGITV	NS4B	2390	2398	2, 4
369	VTIDLDPVVY	NS4B	2397	2406	1
370	TVIDLDPIPYDPKFE	NS4B	2397	2411	2
371	DPIPYDPKF	NS4B	2402	2410	2
372	EPISYDPKF	NS4B	2402	2410	4
373	MLLILCVTQV	NS4B	2418	2427	2
374	LILCVTQVLM	NS4B	2420	2429	2
375	ILLMRTTWA	NS4B	2427	2435	1
376	LLLMRTSWA	NS4B	2427	2435	3
377	LLLMRTTWA	NS4B	2427	2435	4
378	WALCESITL	NS4B	2434	2442	1
379	CEALTLATGPISTLW	NS4B	2437	2451	2
380	LATGPVLTLW	NS4B	2442	2451	4
381	ATGPLTTLW	NS4B	2443	2451	1
382	ATGPISTLW	NS4B	2443	2451	2
383	ATGPITTLW	NS4B	2443	2451	3
384	ATGPILTLW	NS4B	2443	2451	4
385	ATGPVLTLW	NS4B	2443	2451	4
386	LWEGSPGKF	NS4B	2450	2458	1, 3
387	SPGKFWNTTI	NS4B	2454	2463	1, 3
388	IAVSMANIF	NS4B	2463	2471	1, 2, 3
389	IAVSTANIF	NS4B	2463	2471	4
390	MANIFRGSY	NS4B	2467	2475	1, 2, 3
391	YLAGAGLAF	NS4B	2475	2483	1, 3, 4
392	SRLNALGKSEFQI	NS5	2509	2521	2
393	KKLNQLSRKEFDL	NS5	2509	2521	3
394	ETTKHAVSR	NS5	2543	2551	1
395	ETTHHAVSR	NS5	2543	2551	3
396	GSSKIRWIVE	NS5	2552	2561	4
397	GPGHEEPIPM	NS5	2601	2610	1, 2, 4
398	IPMATYGWNL	NS5	2608	2617	1, 4
399	IPMSTYGWNL	NS5	2608	2617	2
400	MSTYGWNIVK	NS5	2610	2619	3
401	MATYGWNLVK	NS5	2610	2619	1, 4
402	MSTYGWNLVRLQSGV	NS5	2610	2624	2
403	STYGWNIVK	NS5	2611	2619	3
404	ATYGWNLVK	NS5	2611	2619	1, 4
405	STYGWNLVR	NS5	2611	2619	2
406	QSGVDVFFTP	NS5	2621	2630	2
407	DVFFTPPEK	NS5	2625	2633	1, 2

408	ESSSNPTIEE	NS5	2643	2652	1, 4
409	SPSPTVEESR	NS5	2645	2654	3
410	TLRVLNLVENWLNNN	NS5	2655	2669	2
411	LRVLNLVENW	NS5	2656	2665	2
412	RVLKMVEPW	NS5	2657	2665	1, 3, 4
413	KVLNPYMPSV	NS5	2675	2684	1, 2
414	VLNPYMPSV	NS5	2676	2684	1, 2
415	MPSVIEKMET	NS5	2681	2690	2
416	LSRNSTHEM	NS5	2704	2712	1, 2, 3, 4
417	SRNSTHEMY	NS5	2705	2713	1, 2, 3, 4
418	VSSVNMVSRL	NS5	2723	2732	3
419	MVSRLLLNR	NS5	2728	2736	3
420	FTMRHKKATY	NS5	2737	2746	2
421	SWHYDQDHPYKTWAY	NS5	2785	2799	2
422	WHYDQDHPY	NS5	2786	2794	2
423	QENPYRTWAY	NS5	2790	2799	4
424	KTWAYHGSY	NS5	2795	2803	1, 2, 3
425	KTWAYHGSYETKQTG	NS5	2795	2809	2
426	WAYHGSYEV	NS5	2797	2805	1, 3
427	WAYHGSYET	NS5	2797	2805	2
428	ASSMVNGVVR	NS5	2811	2820	1, 2
429	ASSMVNGVVK	NS5	2811	2820	1, 2, 4
430	RLLTKPWDVVPMVTQ	NS5	2820	2834	2
431	KPWDIIPMV	NS5	2824	2832	2
432	KPWDVIPMV	NS5	2824	2832	1, 2, 4
433	KPWDVLPMV	NS5	2824	2832	2
434	KPWDVLPTV	NS5	2824	2832	2
435	KPWDVVPMV	NS5	2824	2832	2, 3, 4
436	KPWDVVPTV	NS5	2824	2832	3
437	VPMVTQMAM	NS5	2829	2837	2, 3
438	QIAMTDTTPF	NS5	2834	2843	1
439	MAMTDTTPFGQQRVF	NS5	2835	2849	1, 2, 3
440	DTTPFGQQR	NS5	2839	2847	1, 2, 3, 4
441	TPFGQQRVF	NS5	2841	2849	1, 2, 3, 4
442	TPRSMPGTRR	NS5	2857	2866	3
443	EPKEGTKKLM	NS5	2859	2868	2
444	MPGTRRVMGI	NS5	2861	2870	3
445	LMKITAEWLW	NS5	2867	2876	2
446	VMGITAEWLW	NS5	2867	2876	3
447	MKITAEWLW	NS5	2868	2876	2
448	MEITAEWLW	NS5	2868	2876	3

449	MTTTANWLW	NS5	2868	2876	4
450	KITAEWLWK	NS5	2869	2877	2
451	GKKKTprMCTREEFT	NS5	2880	2894	2
452	KPRICTREEF	NS5	2884	2893	1
453	NPRLCTREEF	NS5	2884	2893	4
454	TprMCTREEF	NS5	2884	2893	2
455	RPRLCTREEF	NS5	2884	2893	3
456	KPRLCTREEF	NS5	2884	2893	3, 4
457	RAAVEDEEF	NS5	2917	2925	3
458	EAVEDSRFWE	NS5	2918	2927	2
459	KAKGSRAIW	NS5	2961	2969	1, 2, 3
460	RAKGSRAIW	NS5	2961	2969	4
461	KGSRAIWYMW	NS5	2963	2972	1, 2, 3, 4
462	GSRAIWYMW	NS5	2964	2972	1, 2, 3, 4
463	RFLEFEALGF	NS5	2976	2985	1, 2, 3, 4
464	RYLEFEALGF	NS5	2976	2985	3
465	LEFEALGFM	NS5	2978	2986	1
466	LEFEALGFL	NS5	2978	2986	1
467	LEFEALGFMN	NS5	2978	2987	2, 3, 4
468	LEFEALGFLN	NS5	2978	2987	2, 3, 4
469	LGFLNEDHW	NS5	2983	2991	2, 3, 4
470	FSRENSLSGV	NS5	2992	3001	1, 2
471	GEGLHKLGY	NS5	3003	3011	1, 2, 3
472	YILRDVSKK	NS5	3011	3019	2
473	HALLATSIF	NS5	3055	3063	1
474	HKKLAEAIFKLTYQN	NS5	3055	3069	2
475	RQLANAIFK	NS5	3056	3064	3
476	KLAEAIFKL	NS5	3057	3065	2
477	YQNKVVKVLR	NS5	3067	3076	4
478	YQNKVVKVQR	NS5	3067	3076	3
479	RPTPKGAVM	NS5	3076	3084	4
480	RPTPKGTVM	NS5	3076	3084	2, 3
481	TPRGTVMDII	NS5	3078	3087	2
482	TPKGAVMDII	NS5	3078	3087	4
483	FTNMEAQLVR	NS5	3106	3115	1, 3
484	RQMEGEGIF	NS5	3115	3123	2
485	RQMEGEGVL	NS5	3115	3123	3
486	TITEEIAVQ	NS5	3130	3139	2
487	KVRKDIQQW	NS5	3181	3189	2
488	IPQWEPSKGW	NS5	3186	3195	1, 4
489	FMKDGRSLVV	NS5	3212	3221	4

490	KETACLGKSY	NS5	3244	3253	1, 2
491	RETACLGKSY	NS5	3244	3253	1, 2
492	RETACLGKAY	NS5	3244	3253	3, 4
493	ETACLGKSY	NS5	3245	3253	1, 2
494	ETACLGKAY	NS5	3245	3253	3, 4
495	ETACLGKSYA	NS5	3245	3254	1, 2
496	ETACLGKAYA	NS5	3245	3254	3, 4
497	KSYAQMWQL	NS5	3251	3259	1
498	KAYAQMWSL	NS5	3251	3259	3, 4
499	KSYAQMWSLMYFHRR	NS5	3251	3265	2
500	SYAQMWTLMY	NS5	3252	3261	2
501	YAQMWQLMY	NS5	3253	3261	1
502	YAQMWTLMY	NS5	3253	3261	2, 3
503	YAQMWSLMY	NS5	3253	3261	2, 3, 4
504	YAQMWQLMYF	NS5	3253	3262	1
505	YAQMWTLMYF	NS5	3253	3262	2, 3
506	YAQMWSLMYF	NS5	3253	3262	2, 3, 4
507	AQMWTLMYF	NS5	3254	3262	2, 3
508	ICSAVPVHW	NS5	3274	3282	3
509	PTSRTTWSIH	NS5	3284	3293	1, 2, 3, 4
510	WSIHAHHQW	NS5	3290	3298	1, 3, 4
511	NPWMEDKTH	NS5	3316	3324	1
512	NPWMEDKTPV	NS5	3316	3325	2, 3
513	NPNMIDKTPV	NS5	3316	3325	4
514	MIDKTPVHSW	NS5	3319	3328	4
515	TPVHSWEDI	NS5	3323	3331	4
516	VPYLGKREDQ	NS5	3331	3340	1, 2, 3
517	REDQWCGSL	NS5	3337	3345	1, 2, 3
518	REDLWCGSL	NS5	3337	3345	4
519	IGLTSRATW	NS5	3346	3354	2, 3, 4
520	NIQTAINQV	NS5	3357	3365	2
521	NEEYTDYMPSMKRFR	NS5	3371	3385	2
522	DYMPSMKRFR	NS5	3376	3385	2, 3
523	MPSMKRFRRE	NS5	3378	3387	2
524	MPVMKRYSAP	NS5	3378	3387	4
525	APFESEGVL	NS5	3386	3394	4

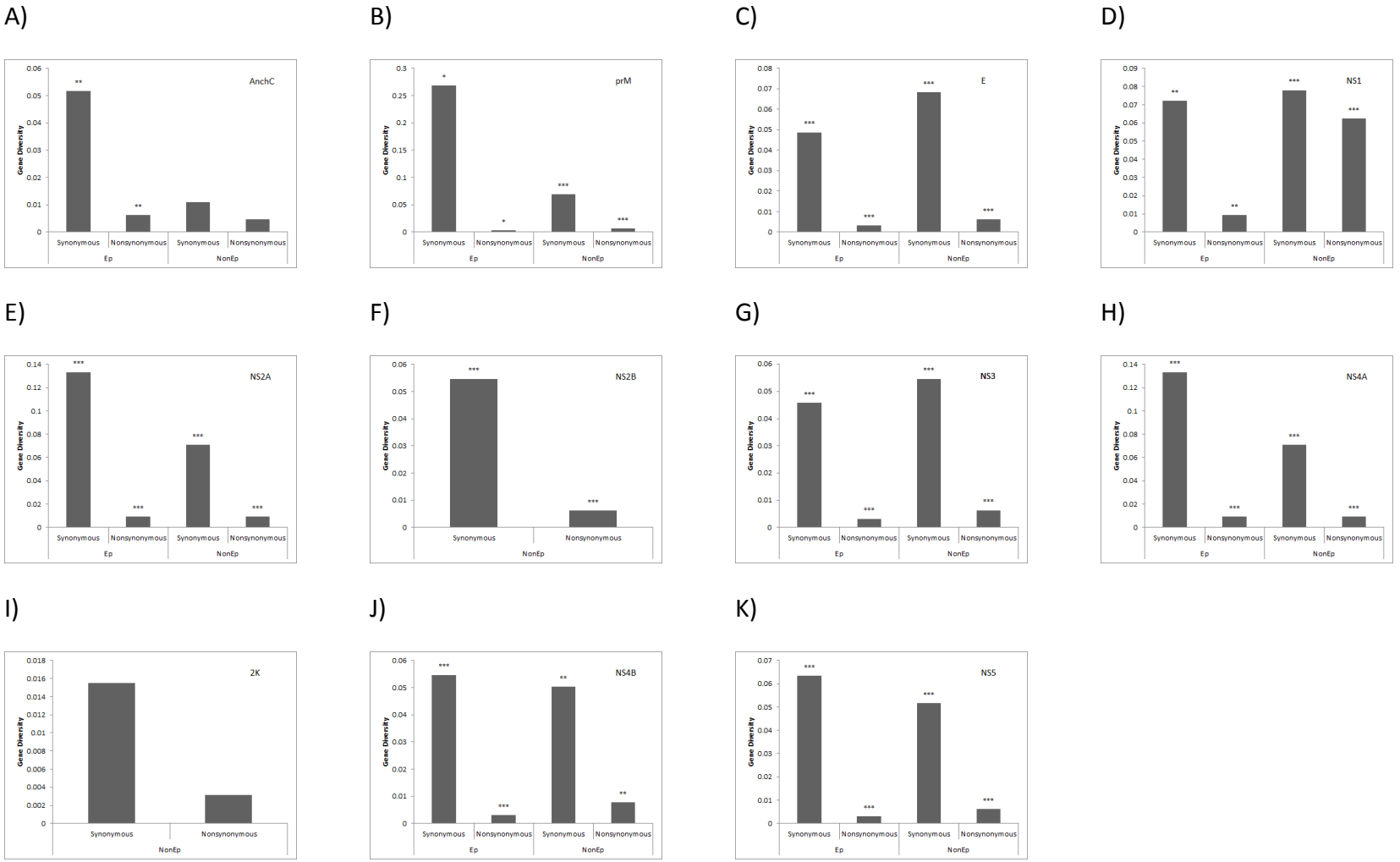


Figure A.1. Median gene diversity at synonymous and nonsynonymous polymorphic sites in CTL epitope (Ep) and non-CTL epitope (NonEp) regions of the 11 proteins of DENV1 (Kruskal-Wallis test * $P<0.05$, ** $P<0.01$, *** $P<0.001$).

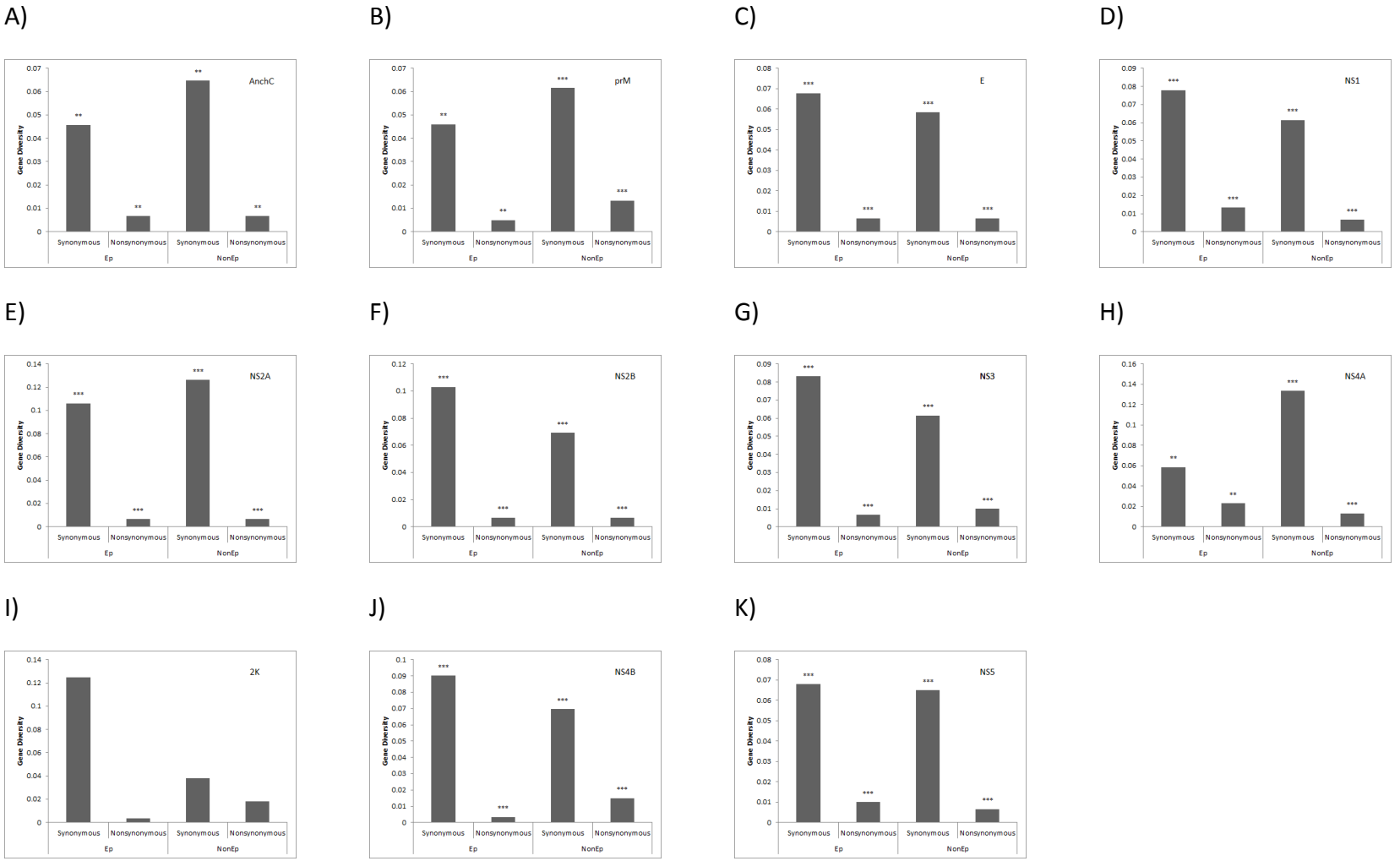


Figure A.2. Median gene diversity at synonymous and nonsynonymous polymorphic sites in CTL epitope (Ep) and non-CTL epitope (NonEp) regions of the 11 proteins of DENV2 (Kruskal-Wallis test * $P<0.05$, ** $P<0.01$, *** $P<0.001$).

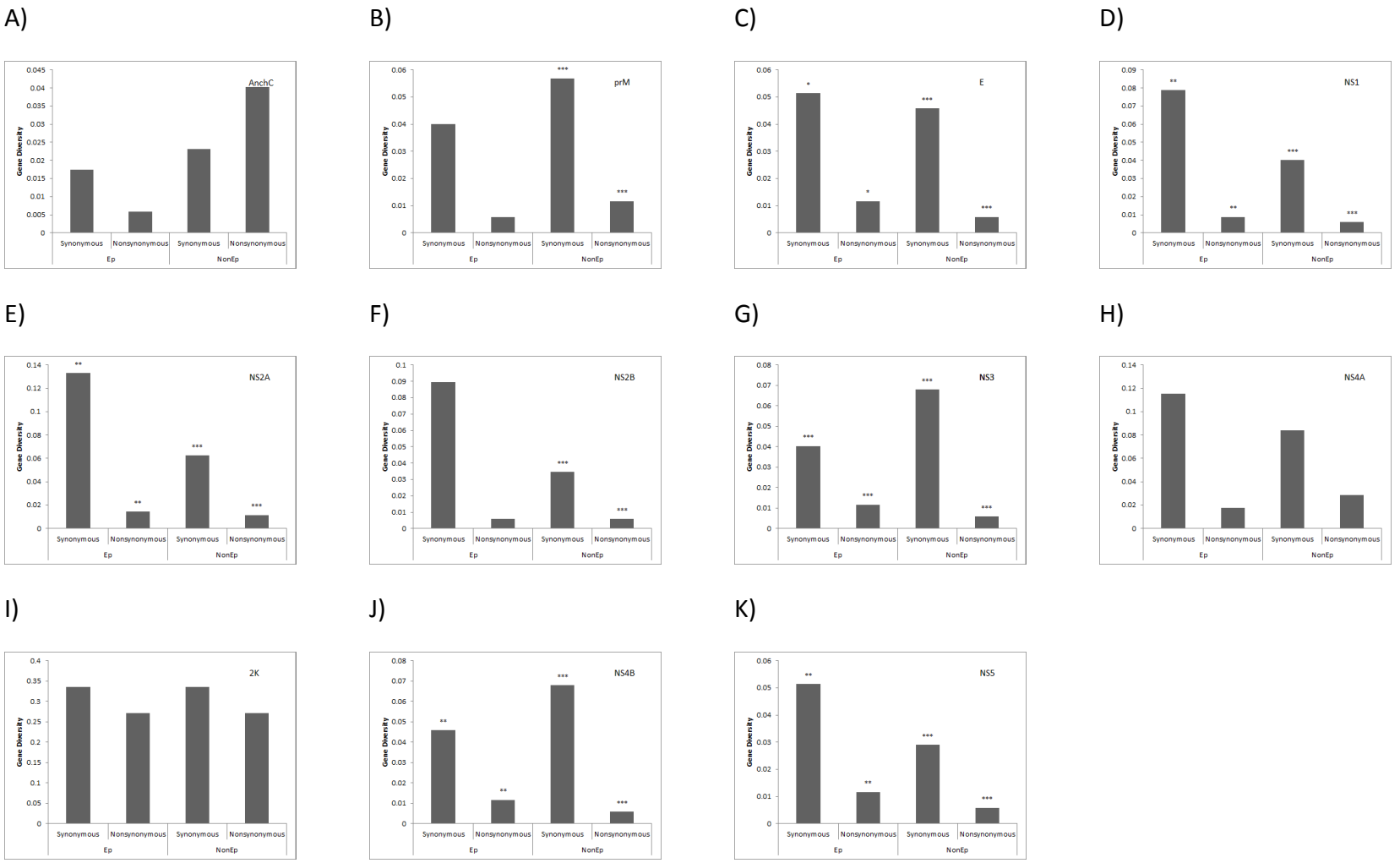


Figure A.3. Median gene diversity at synonymous and nonsynonymous polymorphic sites in CTL epitope (Ep) and non-CTL epitope (NonEp) regions of the 11 proteins of DENV3 (Kruskal-Wallis test * $P<0.05$, ** $P<0.01$, *** $P<0.001$).

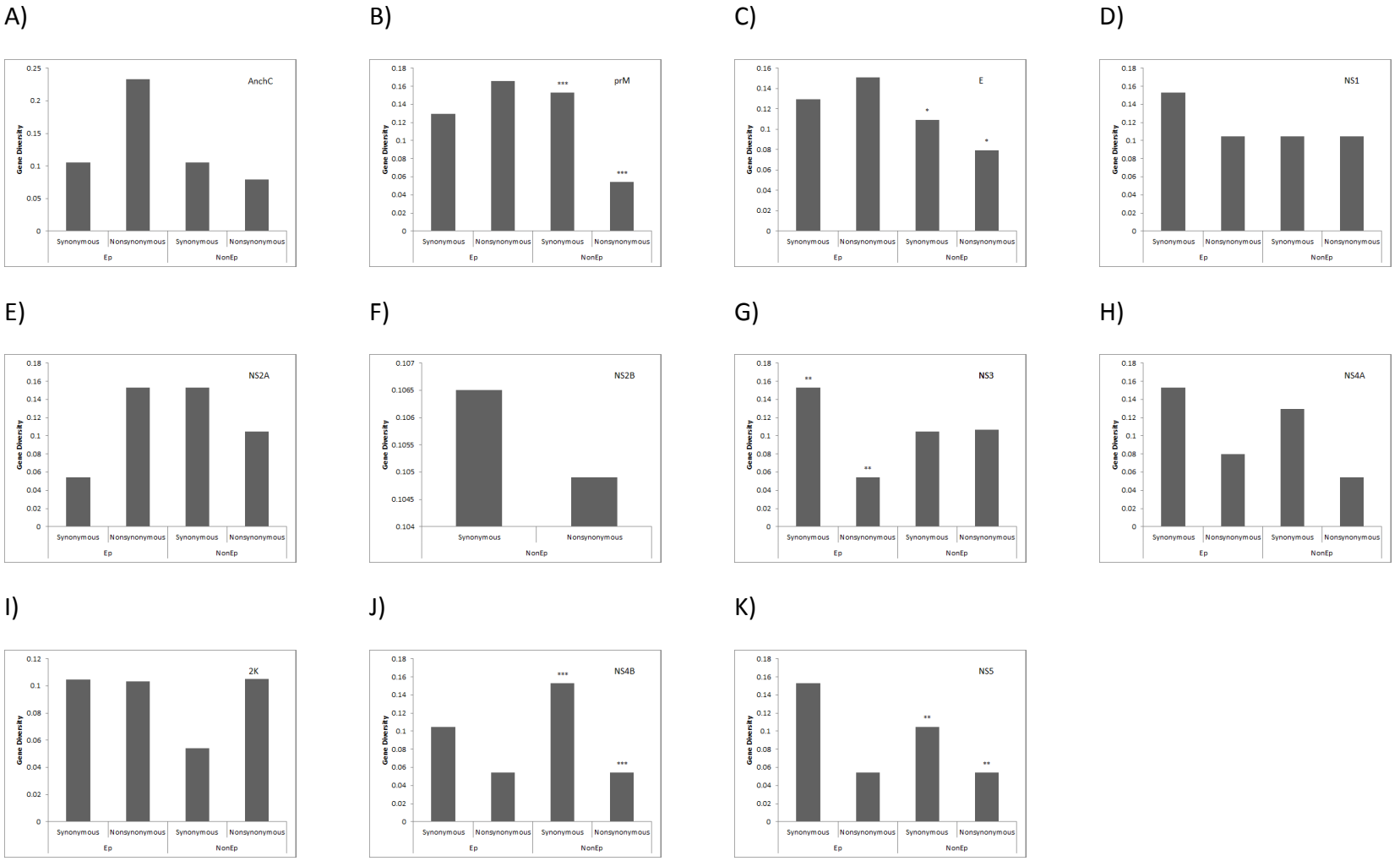


Figure A.4. Median gene diversity at synonymous and nonsynonymous polymorphic sites in CTL epitope (Ep) and non-CTL epitope (NonEp) regions of the 11 proteins of DENV4 (Kruskal-Wallis test * $P<0.05$, ** $P<0.01$, *** $P<0.001$).

APPENDIX B – CHAPTER 3 SUPPLEMENTARY INFORMATION

Table B.1. GenBank accession numbers for complete genome sequences retrieved from the Dengue Virus Database containing the highest percentage of epitope coverage. If more than one accession number is listed, these sequences contained equivalent percentages of epitope coverage across the genome.

Serotype			
DENV1	**DENV2**	**DENV3**	**DENV4**
EF457905	AF038402 AF038403	EU529691 EU687197 EU781136 FJ182006 FJ024469 FJ024468	AY947539

Table B.2a. Counts used to determine RANK measures at top 25% of prediction results between each pair of programs tested for all four DENV serotypes individually as well as all serotypes combined (* P<0.05, ** P<0.01, *** P<0.001).

Serotype	IEDB	NetCTL	IEDB	SVMHC	NetCTL	SVMHC
DENV1	33	21*	24	1***	24	1***
DENV2	35	29	25	1***	24	2***
DENV3	38	25*	25	1***	26	0***
DENV4	25	20	23	1***	22	2***
All DENV	131	95**	97	4***	96	5***

Table B.2b. Counts used to determine RANK measures at top 5% of prediction results between each pair of programs tested for all four DENV serotypes individually as well as all serotypes combined (* P<0.05, ** P<0.01, *** P<0.001).

Serotype	IEDB	NetCTL	IEDB	SVMHC	NetCTL	SVMHC
DENV1	15	10	5	0*	5	1
DENV2	21	14	8	0**	7	2
DENV3	24	12*	6	1	7	0**
DENV4	14	7	7	0**	7	2
All DENV	74	43**	26	1***	26	5***