

Copyright  
by  
Andrew Stephen Beck  
2015

The Dissertation Committee for Andrew Stephen Beck  
certifies that this is the approved version of the following dissertation:

**Population Structure of Yellow Fever Virus:  
Influence of Viral Diversity on Vaccine Attenuation**

Committee:

---

Alan D.T. Barrett, Ph.D., Supervisor

---

Richard B. Pyles, Ph.D.

---

Tian Wang, Ph.D.

---

Alexander Freiberg, Ph.D.

---

Michael Holbrook, Ph.D.

**Population Structure of Yellow Fever Virus:  
Influence of Viral Diversity on Vaccine Attenuation**

by

**Andrew Stephen Beck, B.S.**

**DISSERTATION**

Presented to the Faculty of the Graduate School of Biomedical Sciences of

The University of Texas Medical Branch at Galveston

in Partial Fulfillment

of the Requirements

for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS MEDICAL BRANCH AT GALVESTON

December 2015

To my wife, Courtney. Your support of my studies has been utterly sacrificial, amounting to a debt that no man could repay. Your faith in me has been, without comparison, the most healing and constructive human influence I have experienced.



## Acknowledgments

Gratitude is due to my parents, Leon and Denise Beck. Dad, I'll not forget that my first view through a microscope happened under your supervision. Mom, you were unyielding in your demands that myself and the brothers should value education. Many thanks to you both for your sacrifices that were so frequently measured in combinations of time, money, and aggravation.

To my mentor, Alan Barrett. I am very, very grateful for your insistence that my weaknesses should not be indulged. Behind your back, you are known as a strong advocate for your student's interests. I believe this, too. So let it be known in print.

To the faculty and staff at UTMB, especially the department of Pathology. Studying at this institution has been one of my life's greatest privileges. Many thanks for taking the risk on me. I hope to justify the investment.

To my dissertation committee members, Rick Pyles, Tina Wang, Alex Freiberg, and Mike Holbrook: Many thanks for your time and service. I hope the content is at least interesting enough to for you to read without falling asleep.

To my in-laws, no order implied. Deb and John, you've been endlessly supportive of Courtney and myself, especially during the holiday trips. You've also shown interest in my work, which has been really nice. Dave and Cindy, you have also showed graciousness to Court and I in the last few; I'm glad to hear about your own present adventures. I'm especially thankful for the four of you to have visited us as many times as you have.

Sam and Kris, you've allowed me to move Court far away, which has had it's own difficulties. I am incredibly grateful for your kindness to us at great distance. Whenever we return to Maryland, you somehow treat us as if we've never left. Also, your kids are adorable.

To my two best friends, Alex Myers and Rob Cramer, no order implied. No man should be without friends of such quality. I look forward to our shared milestones, and many raised glasses of scotch.

To my lab members, past and present. This includes Li Li, Melissa White-man, Vanessa Sarathy, Trevor Pitcher, Amy Schuh, Allison McMullen, Brian Mann, Natalie Collins, Daniele Swetnam, Courtney Parker, Jackie Blanc, and others.

To my brothers Tim, Seth, James, John, and David: I am thrilled for us to be adults, and to finally relate as peers. You guys have been a constant and positive influence, proving that among my close relationships, I have much for which to be thankful.

To Darren and Heather Parker, you are first among many unequals.

To Walt and Debbie Mills, who shine brightly in the constellation of my surrogate family members.

To my community at Galveston Bible Church, you have been a adoptive home and family during these years. To have functioned as a member in this lovely little group has been one of my life's other great privileges. Particular thanks are owed to Clay and Karla Thomas, Bob and Jeanette Duke, Rick and Lonni Mallahan, Tim and Susie Bjorklund, Nick and Bekka Bergren, Danny and Marian Davis, John and Virginia Weber, Stephen and Jessica Villanueva, Ben and Sarah Thomas, Brad and Kim Linsley, Jason and Kella Dohring, Richard "Dixie" D'Albergo, Matt Irwin, James Beebe, Andy Krauss, and David Pauling.

To my unborn daughter, Eliana. Although you will likely not read this work, and you may even think of it as that boring stuff your father does, remember that it is virtuous to consider the health of others. Also, we will be fully vaccinating you.

**Population Structure of Yellow Fever Virus:  
Influence of Viral Diversity on Vaccine Attenuation**

Publication No. \_\_\_\_\_

Andrew Stephen Beck, Ph.D.

The University of Texas at Medical Branch at Galveston, 2015

Supervisor: Alan D.T. Barrett, Ph.D.

Yellow fever virus (YFV) is the prototype of all arthropod-borne viruses, the study of which encompasses some of the landmark paradigm shifts in modern vaccinology. Recently, studies on the lifecycle of RNA viruses have indicated that population structure of these viruses is in many cases complex, which in itself may confer fitness within hosts or host ecologies. Thus, these dissertation studies were performed under a set of hypotheses that the YFV population structure is in itself complex, and that features of these populations define a signature that differentiates wild-type from serially-passaged and attenuated vaccine strains. Using deep sequencing techniques, a comparison was made between the prototype West African YFV strain Asibi, and a commercial aliquot of the YFV 17D-204 vaccine. *In silico* methods revealed the vaccine to be of lower diversity than the parental strain, leading to subsequent hypotheses that the level of complexity in the vaccine populations would be related to phenotypic stabilities. A deep sequencing study was performed comparing a set of rare 17D vaccine seeds, which represent all main lineage (substrain) relationships in modern production of the vaccine, and millions of doses

administered. The population structures of the vaccine seeds were found to be stable, showing both low diversity and repeatability of subpopulation content. A deep sequencing study was performed on a set of institutional collection strains representing the discontinued YFV French neurotropic vaccine (FNV), which was developed concurrently to 17D. FNV strains with high neurovirulence contained homogeneity in their population structures, showing that the neurotropic adverse events caused by FNV were the result of mouse brain adaptation to attenuate. A deep sequencing study of an adapted strain of Asibi which causes viscerotropy in hamsters was shown to be of low complexity, suggesting that YFV disease of the liver is rare due to low content of viscerotropic alleles. Finally as confirmation, the Japanese encephalitis vaccine strain SA14-14-2 was compared to its wild-type parental strain SA14 by deep sequencing, revealing that loss of diversity for the vaccine was localized to regions of the genome previously thought to influence pathogenicity. The combined results from both JEV and YFV suggest that population studies are useful in the control of empiric vaccine stability and monitoring of safety.

## Table of Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>viii</b>
<b>Acronyms</b>	<b>xvii</b>
<b>List of Tables</b>	<b>xxiii</b>
<b>List of Figures</b>	<b>xxv</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Historical Significance of the Virus . . . . .	1
1.2 YFV is the Prototype Flavivirus . . . . .	2
1.3 Replication Cycle and Molecular Biology of YFV . . . . .	2
1.3.1 Structural Genes . . . . .	3
1.3.2 Nonstructural Genes . . . . .	5
1.3.3 Untranslated Regions . . . . .	8
1.4 Ecology and Epidemiology . . . . .	8
1.5 Pathogenesis and Tissue Tropism . . . . .	9
1.6 Early Development of YFV Vaccines . . . . .	10
1.6.1 Development of the French Strain . . . . .	11
1.6.2 Development of the 17D Strain . . . . .	12
1.7 Efficacy and Use to the Public Health . . . . .	13
1.8 Lineage of 17D and the Current Vaccine Standard Genotype . . . . .	14
1.9 Nonclinical Comparison of 17D Substrains . . . . .	17
1.10 Clinical Comparison of 17D Substrains . . . . .	18
1.11 Use of 17D as a Recombinant Backbone of Other Vaccines . . . . .	20
1.12 New Report of A Serious Adverse Reaction to 17D . . . . .	22
1.13 New Vaccine Candidate Technologies and Monitoring of Safety . . . . .	23

1.14	New Insights on Durability of Protection for 17D . . . . .	26
1.15	New Insights on Immunogenicity for 17D . . . . .	27
1.16	New Paradigms and Legacy Technologies . . . . .	28
1.16.1	Quasispecies Theory and Fitness of RNA Viruses . . . . .	30
1.16.2	Viral Diversity and Pathogenesis . . . . .	32
1.16.3	Viral Diversity and Ecology of Arboviruses . . . . .	32
1.16.4	Deep Sequencing and Technical Approaches . . . . .	34
1.16.5	Connecting Viral Diversity to Flavivirus Vaccinology . . . . .	34
1.17	Specific Aims . . . . .	35
1.17.1	Specific Aim 1 . . . . .	37
1.17.2	Specific Aim 2A . . . . .	37
1.17.3	Specific Aim 2B . . . . .	37
1.17.4	Specific Aim 3 . . . . .	38
1.17.5	Specific Aim 4 . . . . .	38
<b>Chapter 2.</b>	<b>Materials and Methods</b>	<b>44</b>
2.1	Virus Strains and Passage Histories, By Chapter . . . . .	44
2.1.1	Chapter 3: Comparison of Asibi and 17D-204 . . . . .	44
2.1.2	Chapter 4: Comparison of 17D Substrain Seeds and Vaccines . . . . .	44
2.1.3	Chapter 5: Comparison of French Neurotropic Vaccine Strains . . . . .	45
2.1.4	Chapter 6: Analysis of Subpopulation Selection in a Hamster-Adapted Strain . . . . .	45
2.1.5	Chapter 7: Comparison of Wild-Type with Derived Vaccine Strains of Japanese Encephalitis Virus . . . . .	46
2.2	RT-PCR . . . . .	46
2.3	Pipeline . . . . .	46
2.3.1	Illumina Sequencing, <i>de novo</i> Assembly, and Preprocessing . . . . .	46
2.3.2	Reference Sequences and Genome Layouts . . . . .	47
2.3.3	Preprocessing and Alignment of Illumina Reads for Downstream Analysis . . . . .	48
2.3.4	Generation and Analysis of Consensus Sequences . . . . .	49
2.4	Diversity and Divergence of Sequence Alignments . . . . .	49
2.4.1	Shannon's Entropy . . . . .	49

2.4.2	Simpson's $1 - D$ . . . . .	50
2.4.3	Root-Mean Squared Distance . . . . .	51
2.4.4	Error Rate . . . . .	51
2.5	Statistical Methods for Comparison of Sitewise Nucleotide Diversity . . . . .	52
2.6	Resolution and Classification of Single-Nucleotide Variants . . . . .	53
2.6.1	Resolution and Classification of Wild-Type Revertants . . . . .	53
2.7	Selection Analyses . . . . .	54
2.8	Phylogenetic Methods . . . . .	57
2.9	Data Processing Environment and Graphical Methods . . . . .	57
<b>Chapter 3. Comparison of the Live-Attenuated Yellow Fever Vaccine 17D-204 Strain to Its Virulent Parental Strain Asibi By Deep Sequencing</b>		<b>59</b>
3.1	Abstract . . . . .	59
3.2	Introduction . . . . .	60
3.3	Results . . . . .	62
3.3.1	Assembly and Inspection of Viral Genomes . . . . .	62
3.3.2	Consensus Sequences . . . . .	62
3.3.3	Heuristic Identification of Variants . . . . .	62
3.3.4	Comparison of Quasispecies Diversity . . . . .	63
3.3.5	Intersection of Population Structure For Asibi and 17D-204 . . . . .	63
3.3.6	Recovery of Small Variants By A Paired-Test Model . . . . .	64
3.4	Discussion . . . . .	65
3.4.1	The Wild Type Parental Strain Asibi Is A Diverse Quasispecies . . . . .	65
3.4.2	Variability in Transition to The 17D Population is Localized . . . . .	66
3.4.3	Wild-Type Sequence Content is Not Contained in the Vaccine Population . . . . .	66
3.4.4	Variable YFV Sequence Content is Shared by Other Models of Attenuation . . . . .	67
3.4.5	Attenuation of Asibi Was Likely By Discrete Selection . . . . .	68



<b>Chapter 4. Traversal of Yellow Fever Vaccine Strain 17D Seed Lineages by Deep Sequencing Reveals Limited and Stable Patterns of Subpopulation Diversity</b>	<b>85</b>
4.1 Abstract . . . . .	85
4.2 Introduction . . . . .	86
4.3 Results . . . . .	89
4.3.1 17D-204 Substrain Coverages . . . . .	89
4.3.2 17D-213 Substrain Coverages . . . . .	90
4.3.3 17DD Substrain Coverages . . . . .	90
4.3.4 Asibi Strain Coverage . . . . .	90
4.3.5 Consensus Sequences . . . . .	90
4.3.6 Diversity of Strains . . . . .	93
4.3.7 Divergence of Strains and Stability of Vaccine Genotypes . .	94
4.3.8 Identification of Variant Structure and Substitution Ratios <i>In Situ</i> . . . . .	94
4.3.9 Traversal of 17D Lineage Using Sequentially Passaged Strains: Seed-lot Relationships . . . . .	95
4.3.10 Comparison of 17D Lineage Strains At The Same Passage Level	97
4.4 Discussion . . . . .	97
4.4.1 17D Seeds and Vaccines Represent A Phenotypically Stable Framework To Investigate Population-Level Effects . . . . .	97
4.4.2 Previously Unobserved Sequence Content of the Vaccines Is Revealed by Deep Sequencing . . . . .	99
4.4.3 Differences in Vaccine Diversity Are Associated With Passage Distance . . . . .	100
4.4.4 Purifying Selection Contributes to Maintenance of the Vaccine Genotype . . . . .	101
4.4.5 Naturally Arising Diversity In Flaviviral Vaccines Influences Manufacture . . . . .	102
4.5 Conclusions . . . . .	103
<b>Chapter 5. Deep Sequencing of French Neurotropic Yellow Fever Vaccine Strains Reveals A Conserved Genotype and Dissimilarity of Population Structure</b>	<b>129</b>
5.1 Abstract . . . . .	129
5.2 Introduction . . . . .	130

5.3	Results . . . . .	134
5.3.1	FNV Strain Coverages . . . . .	134
5.3.2	Consensus Sequences . . . . .	134
5.3.3	Diversity of Strains and Stability of Vaccine Strain Genotypes	136
5.3.4	Single Nucleotide Variant Structure . . . . .	137
5.3.5	Reversion of Amino Acid Residues from Vaccine to Wild-Type	137
5.4	Discussion . . . . .	138
5.4.1	Complex Populations in FNV Are Consistent With Observations Using Other YFV Strains . . . . .	138
5.4.2	Divergence of FNV Strains in Negative Association With Neurovirulent Phenotype . . . . .	138
5.4.3	Inference of a Compact Genotype Underlying Neurotropism of FNV . . . . .	140
5.4.4	Amino Acid Substitutions are Similar Between FNV and 17D, Although Attenuated By Different Passage Tissues . . . . .	141
5.4.5	Reversion to Wild-Type is Associated with Loss of Neurotropism for FNV . . . . .	142
5.4.6	Parental Strains FVV and Asibi Are Less Divergent than Previously Described and Likely Originated from the Same Epidemic Event . . . . .	142
5.5	Conclusions . . . . .	143
<b>Chapter 6. Intrahepatic Passage of YFV Asibi Strain Selects a Strain With Reduced Population Complexity</b>		<b>166</b>
6.1	Abstract . . . . .	166
6.2	Introduction . . . . .	166
6.3	Results . . . . .	169
6.3.1	Consensus Sequences . . . . .	169
6.3.2	Diversity of Strains . . . . .	169
6.3.3	Single Nucleotide Variant and Wild-Type Revertant Content of Strains . . . . .	170
6.3.4	Ratio of Nonsynonymous to Synonymous Mutations . . . . .	171
6.4	Discussion . . . . .	172
6.4.1	Perturbation of Viral Diversity Is Frequently Associated With Change of Phenotype . . . . .	172

6.4.2	Intersection of Consensus and Subpopulation Sequence Content With Other Models of YFV Phenotype Adaptation . . .	172
6.4.3	Transition from Asibi to Hamster/P7 Indicates Unique Local Selection Patterns Across Parent and Selected Genomes . . .	174
6.4.4	A Rationale for the “Iceberg Effect” By Observed Relaxation of Evolutionary Tradeoffs . . . . .	174
6.5	Conclusions . . . . .	175
<b>Chapter 7. Comparison of the Live-Attenuated Japanese Encephalitis Vaccine SA14-14-2 to Its Virulent Parental Strain SA14 By Deep Sequencing</b>		<b>186</b>
7.1	Abstract . . . . .	186
7.2	Introduction . . . . .	187
7.3	Results . . . . .	190
7.3.1	Read Coverages . . . . .	190
7.3.2	Consensus Sequences . . . . .	190
7.3.3	Diversity and Genetic Distance of Vaccines, Relative to the Wild-Type Strain . . . . .	191
7.3.4	Variant and Wild-Type Revertant Structure of SA14-14-2 Strains	193
7.3.5	Selection Analysis and $dN/dS$ Ratios . . . . .	193
7.4	Discussion . . . . .	194
7.4.1	Resolution of Consensus for SA14 and SA14-14-2 Strains Using Limited Sequence Information . . . . .	194
7.4.2	Previous Investigations of Neurovirulence Determinants . . .	195
7.4.3	Features of Stability and Low Diversity for SA14-14-2 Vaccine Strains . . . . .	196
7.4.4	Patterns of Selective Pressure Are Typical For Arboviruses .	197
7.5	Conclusions . . . . .	199
<b>Chapter 8. Discussion</b>		<b>217</b>
8.1	Rationale for Studies Performed in the Dissertation . . . . .	217
8.2	Main Findings and Future Directions . . . . .	219
8.2.1	Chapter 3 . . . . .	219
8.2.2	Chapter 4 . . . . .	220
8.2.3	Chapter 5 . . . . .	222

8.2.4	Chapter 6 . . . . .	223
8.2.5	Chapter 7 . . . . .	224
8.3	Critical Methods Analysis for Strategies Used in the Dissertation . .	225
8.3.1	RT-PCR and Isolation of Vaccine Nucleic Acids . . . . .	225
8.3.2	Diversity Indices . . . . .	226
8.3.3	<i>In Silico</i> Methods for Resolution of Population Structure . . .	227
8.3.4	Selection Analyses . . . . .	228
8.4	Statistical Analyses . . . . .	229
8.5	Experimental Issues in Flavivirus Vaccinology That May Benefit From A Diversity Framework . . . . .	230
8.5.1	Reconciling Population-Level Effects WIth Typical Assessments of YFV Vaccine Attenuation . . . . .	231
	<b>Appendix</b>	<b>232</b>
	<b>Appendix 1. Extended Variant Tables</b>	<b>233</b>
	<b>Bibliography</b>	<b>238</b>
	<b>Vita</b>	<b>264</b>
	<b>Curriculum Vitae</b>	<b>265</b>

## Acronyms

- ACIP** Advisory Committee on Immunization Practices. 26
- ALV** Avian Leukosis Virus. 15, 29, 100, 101
- ASP-2** *Trypanosoma cruzi* Amastigote Surface Protein. 20
- AST** Average Survival Time. 132, 133, 139
- ATCC** American Type Culture Collection. 15, 97
- BAM** Binary Alignment/Map File. 47, 48, 51–55
- BLAST** Basic Local Alignment Search Tool. 44, 47
- C** Flavivirus Capsid Protein. 2–4, 63, 134, 186, 190
- CDC** Centers for Disease Control and Prevention. 132
- cDNA** Complementary DNA. 19, 46
- CEF** Chick Embryonic Fibroblast. 28, 102, 197
- CERF** Central Emergency Response Fund. 14
- CI** Confidence Interval. 9, 13, 24
- d.f.** Degrees of Freedom. 93, 190, 191
- DC** Dendritic Cell. 9, 27
- DENV** Dengue Virus. 2, 4, 5, 21, 34, 66

**DNA** Deoxyribobonucleic Acid. 25, 60, 187

**DPI** Days Post-infection. 144

**dVV** d4R Defective Vaccinia Virus. 25

**E** Flavivirus Envelope Protein. 2, 4, 5, 63, 64, 66–68, 91, 92, 99, 134–136, 138–141, 167, 168, 170–172, 186, 189–191, 193, 194, 196

**ECHO** European Community Humanitarian Office. 14

**ER** Endoplasmic Reticulum. 4, 7

**Fiocruz** Fundação Oswaldo Cruz. 15, 18, 19

**FMDV** Foot–and–Mouth Disease Virus. 30, 31

**FNV** French Neurotropic Virus. 11–13, 16, 36, 37, 45, 53, 67, 68, 83, 84, 129–145, 148, 150, 156–159, 161–164, 172, 221, 222, 234

**FVV** French Viscerotropic Virus. 10, 11, 45, 47, 53, 67, 68, 91, 129–145, 148, 156–162

**GAVI** Global Alliance for Vaccines and Immunizations. 13, 14

**GMT** Geometric Mean Titer. 24

**HBV** Hepatitis B Virus. 31

**HCV** Hepatitis C Virus. 31

**HIV** Human Immunodeficiency Virus. 20, 31

**i.c.** Intracerebral. 10, 12, 17, 25, 131–133, 139, 140

**i.d.** Intradermal. 14

**i.h.** Intrahepatic. 86, 166, 167, 172

**i.n.** Intranasal. 132, 133, 139, 140

**i.p.** Intraperitoneal. 45, 144, 166

**IFN** Interferon. 5, 7, 23, 28, 66, 166, 196

**IHR** International Health Regulations. 26

**JE** Japanese Encephalitis Disease. 185–187, 196

**JEV** Japanese Encephalitis Virus. 2, 36–38, 46–48, 185–189, 192, 194–196, 198, 209, 211, 213, 214, 223, 225–227

**kb** Kilobases. 46, 186

**kDa** Kilodaltons. 3–7

**KUNV** Kunjin Virus. 66

**LAMP-1** Human Lysosomal-Associated Membrane Protein 1. 25

**LaV** Lassa Virus. 20

**LGTV** Langat Virus. 7

**M** Flavivirus Membrane Protein. 4, 16, 134, 138, 140, 141

**MeV** Measles Virus. 88

**MVA** Modified Vaccinia Virus Ankara. 24

**MVEV** Murray Valley Encephalitis Virus. 186

**NDC** National Drug Company. 89, 92–95, 100, 101, 220

**NGS** Next-Generation Sequencing. 29, 33, 85, 88, 103, 143, 148, 175, 188, 225, 227, 228, 230

**nm** Nanometers. 186

**NS** Nonstructural. 2, 5–7, 16, 17, 28, 62–64, 66, 90–92, 99, 100, 134–136, 138–141, 167, 168, 170, 172, 186, 189–194, 196, 197, 221, 227

**NSE** Normalized Shannon’s Entropy. 50, 93, 121–123, 136, 158, 168, 169, 204, 205, 211, 212

**NTP** Nucleoside Triphosphate. 66

**PCR** Polymerase Chain Reaction. 48, 49, 101

**PFU** Plaque-Forming Units. 144, 187

**PHK** Primary Hamster Kidney. 187, 223

**PoV** Poliovirus. 31, 87

**prM** Flavivirus Premembrane Protein. 134, 138, 140

**prM/M** Flavivirus Premembrane/Membrane Protein. 2, 4, 63, 64, 186

**PRNT<sub>50</sub>** Plaque Reduction Neutralization Test–50 Percent. 17, 19, 24, 25, 27

**RABV** Rabies Virus. 33

**RbV** Rubella Virus. 88



**RKI** Robert Koch Institute. 15, 19, 86

**RMSD** Root–Mean Square Distance. 51, 93, 94, 123, 124, 136, 140, 168, 169, 181, 191, 213, 220

**RNA** Ribonucleic Acid. 2, 6, 7, 19, 29, 30, 34, 46, 59, 60, 65, 66, 87, 88, 99, 185–187, 195, 223

**RT–PCR** Reverse Transcription Polymerase Chain Reaction. 29, 46, 48, 224, 225

**RVFV** Rift Valley Fever Virus. 33

**s.d.** Standard Deviation. 61, 62, 89, 90, 133, 134, 136, 168, 169, 189–191

**SAE** Serious Adverse Event. 21, 22

**SAGE** WHO Strategic Advisory Group of Experts. 26

**SAM** Sequence Alignment/Map File. 48

**SE** Shannon’s Entropy. 49, 50, 63, 112, 225

**SIV** Simian Immunodeficiency Virus. 20

**SNP** Single-Nucleotide Polymorphism. 22, 67

**SNV** Single–Nucleotide Variant. 51–53, 95, 125, 128, 150, 160, 163, 164, 177, 182, 192, 214, 218, 220, 225, 226, 232

**TLR** Toll-like Receptor. 27

**UsV** Usutu Virus. 186

**UTMB** University of Texas Medical Branch. 45, 47, 219

**UTR** Untranslated Region. 2, 8, 16, 17, 62, 68, 91, 93, 105, 107, 109, 111, 134, 145, 168, 186, 200

**VEEV** Venezuelan Equine Encephalitis Virus. 33

**vRNA** Viral RNA. 3, 7, 46, 225, 226

**WHO** World Health Organization. 9, 13, 15, 17, 22, 25, 26, 28, 43, 86, 90–92, 94, 96, 100, 123, 131, 188

**WNV** West Nile Virus. 2, 7, 21, 32, 66, 186

**WRCEVA** World Reference Center for Emerging Viruses and Arboviruses. 44, 45, 139, 219

**WT** Wild-Type. 3, 8, 10, 12, 15–17, 21, 29–33, 35–38, 45, 53, 59–61, 65–68, 73, 88, 90–95, 97–99, 102, 103, 129–134, 137, 138, 140, 141, 143, 149, 150, 160, 165–169, 171–174, 185, 187–192, 194–197, 200, 216–225, 228, 229, 233

**YEL-AND** Yellow Fever Vaccine-Associated Neurotropic Disease. 21, 100

**YEL-AVD** Yellow Fever Vaccine-Associated Viscerotropic Disease. 21–23, 32

**YF** Yellow Fever Disease. 1, 2, 9–11, 13, 18, 21, 26, 130, 143, 166, 196

**YFV** Yellow Fever Virus. 1–5, 7–11, 13, 20, 22–26, 28, 29, 32–38, 44, 47, 53, 59, 60, 66–69, 81–89, 91, 97–99, 115, 117–120, 123, 129–133, 138–140, 143, 165, 167, 171–174, 188, 189, 195–198, 216–219, 221–224, 226, 227, 229, 230

## List of Tables

1.1	Amino acid substitutions contained in the WHO standard genotype for YFV vaccines. . . . .	39
1.2	Published clinical and nonclinical data comparing the 17D substrains. . . . .	40
2.1	Primers for RT-PCR amplifications. . . . .	58
3.1	Coding Variants observed in the Asibi strain alignment, using a 1.00 percent frequency cutoff heuristic. . . . .	70
3.2	Coding variants observed in the 17D-204 strain alignment, using a 1.00 percent frequency cutoff heuristic. . . . .	71
3.3	Diversity index comparison for individual genes between Asibi and 17D-204, Using the Mann-Whitney U Test. . . . .	72
4.1	Origin information for the seeds used in the study. . . . .	104
4.2	Amino acid and 3'UTR substitutions between Asibi and all 17D strains sequenced in the study. . . . .	105
4.3	17D-204 strain amino acid and 3'UTR substitutions using the newly sequenced strains, in reference to publicly available sequences. . . . .	107
4.4	17D-213 strain amino acid and 3'untranslated region (UTR) substitutions using the newly sequenced strains, in reference to publicly available sequences. . . . .	109
4.5	17DD strain amino acid and 3'UTR substitutions using the newly sequenced strains, in reference to publicly available sequences. . . . .	111
4.6	Shannon's entropy and RMSD estimations for the 17D strains considered in the study. . . . .	112
4.7	Wild-type revertant variants observed in the vaccine strains. . . . .	113
4.8	$dN/dS$ summary for all strains considered in the study, considering the YFV open reading frame and each gene segment individually. . . . .	115
5.1	Source laboratories and <i>in vivo</i> lethality measurements for the FNV strains considered in the study. . . . .	145
5.2	Amino acid and 3'UTR substitutions for FNV strains, relative to the parental strain FVV. . . . .	146

5.3	Diversity estimates by normalized Shannon's entropy, considering the consensus sequences re-computed by NGS methods. . . . .	149
5.4	Counts of WT revertants recovered under the heuristic method, using a one percent frequency cutoff. . . . .	150
5.5	Wild-type revertant SNVs observed in the FNV strain alignments. Each nucleotide position bearing wild-type (FVV, parental strain) identity is indicated with nucleotide or amino acid identity and percentage frequency. . . . .	151
6.1	Amino Acid Substitutions in the Hamster/P7 virus, relative to the unpassaged (P0) Asibi strain from which it was derived. . . . .	176
6.2	Summary of $dN/dS$ values for gene segments and complete open reading frame in both Asibi and Hamster/P7 strains. . . . .	177
6.3	Coding SNVs in Asibi and Hamster/P7 strains, in reference to the respective consensus of each strain. . . . .	178
6.4	20-codon sliding windows with high $dN/dS$ ratios in the strains considered. . . . .	179
7.1	Amino acid substitutions differentiating the SA14 strain sequenced in this study to that of three others previously reported. . . . .	200
7.2	Amino acid and untranslated region substitutions observed in SA14-14-2 vaccine strains, in reference to the newly-sequenced, wild-type SA14 strain. . . . .	201
7.3	Diversity summary for SA14, SA14-14-2/P1, and SA14-14-2/P2 using normalized Shannon's entropy, split by gene segment. . . . .	203
7.4	Summary of diversity by normalized Shannon's entropy for nucleotide positions bearing SA14-14-2 amino acid substitutions, by gene segment. . . . .	205
7.5	Summary of diversity by normalized Shannon's entropy for nucleotide positions with SA14-14-2 amino acid substitutions, by gene segment. . . . .	206
7.6	Amino acid variants concordantly observed in both SA14-14-2/P1 and SA14-14-2/P2, relative to the newly sequenced SA14 strain. . . . .	207
7.7	$dN/dS$ ratios for each gene segment and complete open reading frame for newly-sequenced SA14, SA14-14-2/P1, and SA14-14-2/P2 strains. . . . .	208
7.8	20-codon sliding windows with high $dN/dS$ ratios in the open reading frames of SA14, SA14-14-2/P1, and SA14-14-2/P2. . . . .	209
1.1	Complete (coding and silent) variants from the wild-type (WT) Asibi parental strain as an extension of the data shown in Table 3.1. . . . .	234
1.2	Complete (coding and silent) variants from the 17D vaccine strain as an extension of the data shown in Table 3.2. . . . .	236

## List of Figures

1.1	Canonical genome organization of YFV and other flaviviruses. . . . .	42
1.2	Current production for YFV 17D vaccines, including all manufacturers and substrains. . . . .	43
3.1	Diversity indices for each nucleotide position in Asibi and 17D-204, including both Shannon's entropy and Simpsons $1 - D$ . . . . .	73
3.2	Comparison of diversity indices for yellow fever virus Asibi and 17D-204 strains, at multiple scales. . . . .	74
3.3	Comparison of diversity indices for yellow fever virus Asibi and 17D-204 strains, at multiple scales. . . . .	75
3.4	Quantile-quantile comparisons of error rate along all nucleotide positions recovered for Asibi and 17D-204 strains. . . . .	76
3.5	Quantile-quantile comparisons of error rate along all nucleotide positions recovered for Asibi and 17D-204 strains. . . . .	77
3.6	Intersection of variant identity between Asibi and 17D-204 strain quaspecies is unidirectional, depicted as a barplot of marginal frequency. . . . .	78
3.7	Intersection of variant identity between Asibi and 17D-204 strain quaspecies is unidirectional, as box and scatterplot. . . . .	79
3.8	Scatterplots showing change in nucleotide frequency for paired samples comparison. . . . .	80
3.9	Recovery of local plasticity in the YFV genome using the paired= $\chi^2$ -test model, as Manhattan plot. . . . .	81
3.10	Recovery of local plasticity in the YFV genome using the paired= $\chi^2$ -test model, as a sliding-window summary. . . . .	82
3.11	Venn diagram showing overlap of significantly changed nucleotide frequencies with prototype YFV strains, showing intersection with parental strains only. . . . .	83
3.12	Venn diagram showing overlap of significantly changed nucleotide frequencies with prototype YFV strains, showing intersection with derived strains only. . . . .	84
4.1	Tree of seeds and vaccines used in the study, organized by substrain and passage level relative to the parental strain Asibi. . . . .	116

4.2	Alignment for newly sequenced and publicly available sequences showing vaccine divergence at E331. . . . .	117
4.3	Alignment for newly sequenced and publicly available sequences showing vaccine divergence at NS179. . . . .	118
4.4	Alignment for newly sequenced and publicly available sequences showing vaccine divergence at NS2B37. . . . .	119
4.5	Alignment for newly sequenced and publicly available sequences showing vaccine divergence at NS4B232. . . . .	120
4.6	Heat map of Kruskal-Wallis p-values for all pairwise comparisons of normalized Shannon's entropy, considering values along the whole genome extent of each strain. . . . .	121
4.7	Heat map of Kruskal-Wallis p-values for all pairwise comparisons of normalized Shannon's entropy, considering values only along the open reading frame of each strain. . . . .	122
4.8	Scatter plot of normalized Shannon's entropy versus RMSD relative to Asibi for all nucleotide sites recovered in each strain. . . . .	123
4.9	Neighbor-joining phylogeny calculated by pairwise mean RMSD for all seeds considered in the study. . . . .	124
4.10	Modeled variants for the strains, with coding/silent classifications and wild-type revertants indicated. . . . .	125
4.11	Boxplots of summarized $dN/dS$ values along each gene segment. . .	126
4.12	Heatplot of $dN/dS$ values for all strains considered. . . . .	127
4.13	Density plot for counts of variants, derived from modeled SNVs in the population. . . . .	128
5.1	Plot of read coverages for each virus strain. . . . .	156
5.2	Normalized Shannon's entropy along the full genomes of FVV and all FNV strains. . . . .	157
5.3	Pairwise quantile-quantile plots of sequencing error rate for FVV and FNV strains. . . . .	158
5.4	Scatter plot for all nucleotide positions of normalized Shannon's entropy (diversity) versus $RMSD$ (increasing genetic distance) relative to the parental strain FVV. . . . .	159
5.5	Diversity estimate by normalized Shannon's entropy for each of 26 nucleotide positions containing amino acid substitutions. . . . .	160
5.6	Scatter plots of nucleotide polymorphisms recovered in each virus strain, using a probabilistic model. . . . .	161

5.7	Venn diagram showing the intersection of counts for all variant sites recovered for FVV and FNV strains. . . . .	162
5.8	Venn diagram showing the intersection of count for only coding variant sites for FVV and FNV strains. . . . .	163
5.9	Density plots showing relative accumulations of all SNVs as recovered by modeled criteria. . . . .	164
5.10	Density plots showing relative accumulations of coding SNVs as recovered by modeled criteria. . . . .	165
6.1	Derivation of Hamster/P7 strain. . . . .	180
6.2	Line chart of diversity estimates for each nucleotide site in the genomes of Asibi and hamster/P7 strains. . . . .	181
6.3	Scatter plot considering entropy and genetic distance from Asibi for each nucleotide position recovered by deep sequencing in Asibi and hamster/P7 strains. . . . .	182
6.4	Variant populations identified by a model in the subpopulation structure of the strains Asibi, and hamster/P7. . . . .	183
6.5	$dN/dS$ estimation summarized along each gene for all viruses considered in the study. . . . .	184
6.6	$dN/dS$ estimation along the full genome for each strain considered in the study, using a sliding window. . . . .	185
7.1	Derivation of JEV Vaccine, showing the passage history of newly analyzed strains. . . . .	210
7.2	Read coverage along the length of genomes, for each virus analyzed in the study. . . . .	211
7.3	Sitewise diversity estimation, measured by normalized Shannon's entropy, for each strain considered in the study. . . . .	212
7.4	Summary of diversity, measured by normalized Shannon's entropy, for each nucleotide position containing an amino acid substitution between the newly sequenced wild-type strain SA14 and vaccine strains SA14-14-2/P1 or SA14-14-2/P2. . . . .	213
7.5	A plot of normalized Shannon's entropy versus RMSD relative to the wild-type parental strain SA14, for each nucleotide position recovered in the three JEV strains. . . . .	214
7.6	A plot of the frequency of each SNV recovered from the JEV strains studied by a probabilistic model in V-Phaser v. 2.0, in order of nucleotide position. . . . .	215
7.7	20-codon sliding window analysis of $dN/dS$ , along the open reading frames of JEV strains SA14. SA14-14-2/P1, and SA14-14-2/P2. . . .	216

# Chapter 1

## Introduction<sup>1</sup>

### 1.1 Historical Significance of the Virus

Yellow fever virus (YFV) is the causative agent of yellow fever (YF) disease in humans, which occupies a singular and fearsome place in the study and management of tropical illness. YF is noted in colonial-era literature as a febrile disease of the tropics, from which patients would suffer fever, and jaundice, eventually succumbing to hemorrhagic disease (Vainio and Cutts 1998). Disease occurrence was recorded in the late 1700s through early 1900s, in which the disease phenotype and eventually the complex mosquito-primate transmission cycle became described. Notably, losses to the disease stemming from American military deployments to Cuba during the Spanish-American War and construction of the Panama Canal prompted close study of the disease in the late nineteenth century (Bazin 2011). YF occurred along the North American east coast and Gulf of Mexico until 1905, when the last outbreak event of the United States mainland was recorded New Orleans, Louisiana. Transatlantic trade in both goods and persons are considered to have seeded, and eventually permitted the permanent establishment of the virus in South American circulation from West Africa (Bryant, Holmes, and A. D. T. Barrett 2007).

---

<sup>1</sup>This chapter contains material representing the author's accepted manuscript of an article published as the version of record in Expert Review of Vaccines ©14 Sep 2015, Taylor & Francis Group. Publisher's reference number for reprint permission: LA/IERV/P5790. Persistent URL to publisher's version: <http://www.tandfonline.com/doi/full/10.1586/14760584.2015.1083430>. (A. S. Beck and A. D. Barrett 2015)



## 1.2 YFV is the Prototype Flavivirus

YFV is the causative agent of YF. The virus is a member of the family *Flaviviridae*, genus *flavivirus*, and as such bears similarities of size and genome organization to the related Dengue virus (DENV), Japanese encephalitis virus (JEV), and West Nile virus (WNV), and others of public health significance. The particle is enveloped, measuring between 50 and 60nm by transmission electron microscopy (Burlaud-Gaillard et al. 2014). The YFV genome is organized as a single-stranded, positive-sense strand of RNA (RNA), comprising in read order, a 5' UTR (118 bases), three structural proteins capsid (C), premembrane/membrane (prM/M), envelope (E), 7 nonstructural (NS) proteins NS1, NS2A, NS3, NS4A, NS4B, NS5, and a 3'UTR (507 bases) [Figure 1.1]. Following clathrin-dependent endocytosis and subsequent low pH endosomal fusion, the genome is immediately translated upon entry to the host cytoplasm, producing a single 3411 amino acid polyprotein that is co- and post-translationally cleaved by a combination of host and viral proteases. Non-structural genome products contribute to the replication complex and serve various roles in genome replication, which occurs on endoplasmic reticular membrane structures; the assembled virus is transited, and finally exocytosed through the trans-golgi network.

## 1.3 Replication Cycle and Molecular Biology of YFV

Owing to the success of vaccination in the prevention of YF disease, the specific molecular biology of YFV is poorly understood and infrequently researched; in many cases YFV molecular biology is interpreted or reviewed by expectation of homology to other flaviviruses, usually in the context of vaccinology (Monath, M. Gershman, et al. 2013). For YFV, the pathogenic influences exerted by specific protein components of the virus may indicate specific mechanisms of attenuation for

YFV vaccines, the central interest of this work. As such, the following subsections describing functional properties of flavivirus proteins are described in reference to several members of the family *Flaviviridae*. The topology of the replication and assembly complex is itself obscure, with considerable research efforts spent on elucidation of both the orientation and interactions of virus components in the complex. Listed biophysical properties of the peptides are calculated from Genbank accession AY640589, the full-length WT prototype Asibi strain (McElroy et al. 2005).

### 1.3.1 Structural Genes

Three structural genes are arranged sequentially in read order along the canonical flavivirus genome [Figure 1.1], which in combination confer the physical structure of the mature particle.

#### 1.3.1.1 Capsid (C)

C arises from a peptide of 121 residues, with a molecular weight of 13.6kilodaltons (kDa), and pI of 13.05. Canonically, C binds and encapsidates viral RNA (vRNA) in the mature particle (Monath, M. Gershman, et al. 2013). The protein is observed as a dimer in solution, and the ablation of domains influencing the dimer arrangement were associated with loss of infectivity for cloned YFV (Patkar et al. 2007). Mature flavivirus C (11kDa) possesses an internal hydrophobic domain that mediates association of the C/vRNA complex to the viral membrane (Ma et al. 2004). Flavivirus C is also observed to aggregate on intracellular lipid structures and as such is thought to influence early stages of particle assembly (Samsa et al. 2009).

### 1.3.1.2 Premembrane (prM) and Membrane (M)

PrM/M arises from a peptide of 164 residues, with a molecular weight of 18.7kDa and pI of 10.15. PrM/M is a surface component of the nascent virus particle, of which maturation before assembly requires highly regulated cleavage of prM/M (endoplasmic reticulum (ER) lumen) and C (cytoplasmic) from both sides of an intervening, signal peptide which spans the ER membrane (Lobigs et al. 2010). Maturation of the nascent virus particle through the trans-golgi complex is mediated in part by furin cleavage of the N-terminal prM peptide from the surface of the virion, with membrane (M) remaining. If cleavage of prM/M is not carried through to completion, infectivity is reduced (Wengler 1989). An apoptosis-promoting domain was observed in DENV M, which in YFV 17D strain vaccines (later discussed) an amino acid substitution at residue 36 is observed (Catteau 2003).

### 1.3.1.3 Envelope (E)

YFV E arises from a peptide of 493 residues, with a molecular weight of 53.2kDa and pI of 8.11. Flavivirus E canonically forms the bulk of outward-facing surface structure for the virion, and as such the protein exerts pathogenically significant roles in physical interaction with the host cell and immunogenicity of the particle. Three domains have been described for the protein, of which domains I and II are discontinuously arranged proceeding from the amino terminus, domain III is continuously translated balance of 100 amino acids oriented C-terminally to domains I and II, with the terminal 8 amino acids consisting of a stem-loop membrane anchor (Monath, M. Gershman, et al. 2013).

Envelope domains II and III occupy, in their folded states, opposite ends of domain I, of which the entire complex is arranged in homodimeric structures along the surface of the virion. Domain II contains both dimerization and fusion properties,

e.g. it confers the capacity of the mature E complex to dimerize and to fuse with the host endosome, respectively. Following receptor-mediated endocytosis of the particle, endosomal fusion takes place via a pH-dependent mechanism in which the low pH of the endosome provokes a rearrangement of the dimer structure into trimers that present the domain II fusion loop and execute the fusion into the host cytoplasm (Monath, M. Gershman, et al. 2013). Domain III is thought to govern interactions with host receptors, however very few studies have demonstrated either a definitive receptor or the particular residues influencing the binding affinity of domain III (E. Lee and Lobigs 2008). Both domains II and III are immunogenic and stimulate neutralizing antibody against the particle, however domains I and II are observed to solicit the predominant bulk of neutralizing antibody response (Vratskikh et al. 2013).

### **1.3.2 Nonstructural Genes**

Seven structural genes are arranged sequentially in read order along the canonical flavivirus genome following the structural genes; they have diverse and sometimes poorly understood roles within the infected host cell.

#### **1.3.2.1 Nonstructural Protein 1 (NS1)**

NS1 arises from a peptide of 352 residues, with a molecular weight of 39.7kDa and pI of 7.98. The protein is found in both intracellular and secreted forms, of which the extracellular form is highly immunogenic. For YFV, the NS1 is found in membrane-bound (Schlesinger and Brandriss 1990) and secreted forms, which may be glycosylated (Post, Carvalho, and Galler 1991). Secreted forms of NS1 are observed in a hexameric form, which in DENV infection of interferon (IFN)- $\alpha$  receptor-knockout mice was found to contribute to vascular leakage, a manifestation

of severe disease (Beatty et al. 2015). Intracellular NS1 is required for replication competence of the virus and associates with NS4B (discussed later), providing evidence for association of the protein with the replication complex (Youn, T. Li, et al. 2012; Youn, Ambrose, et al. 2013).

#### **1.3.2.2 Nonstructural Protein 2A (NS2A)**

NS2A arises from a peptide of 224 residues, with a molecular weight of 24.1kDa and pI of 11.03. The protein is hydrophobic, and likely plays a role in the replication complex, in assembly, and in membrane induction (Leung et al. 2008). Destructive mutagenesis of the interaction between NS2A and NS3 ablates virion assembly (Vobmann et al. 2015).

#### **1.3.2.3 Nonstructural Protein 2B (NS2B)**

NS2B arises from a peptide of 130 residues, with a molecular weight of 13.8kDa and pI of 4.40. NS2B functions in complex with NS3 and is required for protease activity to occur (Monath, M. Gershman, et al. 2013).

#### **1.3.2.4 Nonstructural Protein 3**

NS3 arises from a peptide of 623 residues, with a molecular weight of 69.013.8kDa and pI of 8.69. The protein contains three functional domains. First a serine protease domain assists in cleavage of the single virus polypeptide (Monath, M. Gershman, et al. 2013). A second nucleotide triphosphatase domain catalyses the removal of phosphate groups from nascent nucleotide triphosphate during replication. Finally, an RNA helicase is employed to unwind RNAs during minus strand synthesis. The protein contains immunodominant CD8<sup>+</sup> T cell epitopes (Most et al. 2002).

#### **1.3.2.5 Nonstructural Protein 4A**

NS4A arises from a peptide of 149 residues, with a molecular weight of 16.0kDa and pI of 6.10. NS4A associates with NS4B, and thus is thought to have a role in the replication complex (J. Zou, X. Xie, Q.-Y. Wang, et al. 2015).

#### **1.3.2.6 Nonstructural Protein 4B**

NS4B arises from a peptide of 250 residues, with a molecular weight of 27.3kDa and pI of 9.87. NS4B is very hydrophobic and is localized to ER membranes during viral replication. The protein is thought to interact with the replication complex and provide stability or orientation as a homodimeric structure (J. Zou, X. Xie, L. T. Lee, et al. 2014). The C-terminus of flavivirus NS4B is reported to be an IFN antagonist (Muñoz-Jordán et al. 2005).

#### **1.3.2.7 Nonstructural Protein 5**

NS5 arises from a peptide of 906 residues, with a molecular weight of 103.4kDa and pI of 9.68. NS5 exerts two main functions, which are, respectively, (1) a methyltransferase domain, which participates in 5' capping of the nascent vRNAs, and (2), the RNA-dependent RNA polymerase, which synthesizes RNA from an RNA template (Monath, M. Gershman, et al. 2013). For YFV, NS5 contains determinants of neurovirulence in small mammalian models (H. Xie, Ryman, et al. 1998). Evidence has been put forth to suggest the involvement of flavivirus NS5 in antagonism of innate IFN responses in the mammalian host by the blockade of JAK-STAT signaling in Langkat virus (LGTV) infection (Best et al. 2005). Further, decreased surface expression of the IFN- $\alpha$  receptor was observed in WNV infection (Lubick et al. 2015). Specifically for YFV, a positive feedback mechanism has been described for NS5 in which the propagation of type 1 IFN signals causes ubiquitina-

tion of NS5, permitting its sequestration of STAT2, dampening the transduction of IFN responses (Laurent-Rolle et al. 2014).

### **1.3.3 Untranslated Regions**

#### **1.3.3.1 5' Untranslated Region(UTR)**

The 5'UTR is 119 nucleotides in length (YFV), and is highly conserved. The region contains secondary structures attributed to cyclization of the virus genome *in vivo*.

#### **1.3.3.2 3' Untranslated Region(UTR)**

The 3'UTR is 508 nucleotides in length, and also contains secondary structures, however there is variability in the sequence and secondary structures among flaviviruses and between YFV genotypes. Most interestingly, a set of tandem repeat deletions in the 3'UTR were associated geographically, meaning that some ecological factor exerts pressure to keep the deletions (Bryant, Vasconcelos, et al. 2005).

## **1.4 Ecology and Epidemiology**

WT YFV is principally vectored by *Aedes spp.* mosquitoes in Africa and *Haemagogus* and *Sabethes spp.* in South America and the non-human primate hosts differ by geographic region. The taxonomy of YFV implies considerable ecological and geographic restriction for the virus; sustained transmission has only been observed in tropical and subtropical belts of South America, Africa, and certain Caribbean islands (Jentes et al. 2011). Surveillance of WT YFV is rarely done in any comprehensive way, however the WT lineage is divided into 4 basal genotypes, East and West Africa, East and West South America (Bryant, Holmes, and A. D. T. Barrett 2007). Risk of YFV infection is reasonably thought to associate with prox-

imity to vector habitat, which requires both sufficient rain and low altitude (Jentes et al. 2011).

## 1.5 Pathogenesis and Tissue Tropism

YFV infection begins from the bite of an infected hematophagous mosquito, typically *Aedes spp.*. Then, the virus is thought to infect resident antigen-presenting cells in the skin, which transport the virus to a draining lymph node; flaviviruses are frequently reported to infect Langerhans and other dendritic cell (DC) types (Johnston, Halliday, and King 2000; T. Querec et al. 2006; Wu et al. 2000). From there, a primary viremia and febrile signs may develop between 3 and 6 days post-infection. After a resolution phase in which signs and symptoms decrease in severity, a minority of cases (12%) may advance to a toxic phase, featuring liver damage and hemorrhage, for which the mortality rate has been recently estimated to be 47 percent (Johansson, Vasconcelos, and Staples 2014). Length of YF disease course has been described for nonhuman primates, (Monath, Brinker, et al. 1981), however the most recent data on intrinsic incubation periods in humans (time from infection to first sign of fever) was inferred by a modeling study to be 4.6 days (95%confidence interval (CI)=4.1-5.1) (Johansson, Arana-Vizcarrondo, et al. 2010). Secondary viremia targeting the liver is referred to as *viscerotropic* disease. YFV is directly tropic to hepatocytes, which in histological sections reveals the presence of viral antigen, councilman bodies, and an ordered pattern of hepatocyte necroses and apoptosis in midzonal region of the infected acinus (Quaresma, Duarte, and Vasconcelos 2006; Quaresma, Barros, et al. 2006). The destruction of regular liver parenchymal architecture and consequent alteration to hemoglobin recycling produces characteristic jaundice; the yellowed skin in advanced cases is the color from which the disease derives its name (Latin *flavus*: “yellow”). The World Health Organization (WHO)-recommended case definition for



laboratory workup of suspected YF disease includes only the observation of jaundice within two weeks of acute fever (World Health Organization 2003). Tissue tropisms for YFV have been extensively reviewed (Monath and Vasconcelos 2015).

## 1.6 Early Development of YFV Vaccines

Live-attenuated vaccines to prevent YF were among the first to be successfully developed by empiric serial passage, and as such occupy a place of historical significance in the study and control of tropical diseases. The development pathway of early YFV vaccines was notable for several technical advances that rendered the virus amenable to laboratory manipulation. The discovery that Indian crown (then identified as *Macaca sinica*) and rhesus (*Macaca mulatta*) macaques were susceptible to YFV infection permitted the isolation of WT strains in the absence of a cold chain (Stokes, Bauer, and Hudson 1928). Max Theiler first demonstrated that the virus could be propagated in the brains of mice, also showing that the phenotype of the virus would change with serial passage (Theiler 1930). The susceptibility of mice to intracerebral (i.c.) challenge by YFV was then usefully employed in serological protection assays, in which sera were titrated against lethality of a YFV strain with neurotropic phenotype (Theiler 1931). These methods were used to successfully isolate two WT strains in 1927, these strains forming the basis of two concurrently developed, and then successfully deployed live-attenuated YFV vaccines. The first of these WT strains, called Asibi, was isolated from a Ghanaian of the same name, who suffered a mild case of YF and fully recovered (Stokes, Bauer, and Hudson 1928). The second was isolated from the Syrian Francois Mayali, which became known as the French viscerotropic virus (FVV), and is referred to in literature of the time as the “French” strain, whether as WT or in passaged, attenuated strain forms (Sellards and Hindle 1928).

### 1.6.1 Development of the French Strain

FVV was adapted mouse brain (formerly also called as “fixed”) by 128 passages by Theiler before transfer to the Institut Pasteur at Dakar, Senegal (Theiler 1930; Durieux 1956b). WT viscerotropic properties of the strain were attenuated, leading to the investigation of the strain as a human vaccine. Neurotropism in monkeys was enhanced, a property reasonably attributed to fixation of the virus to mouse brain (W. Lloyd and Penna 1933; Ni, Ryman, et al. 2000). At attenuated passage levels, the vaccine strain is referred to as the “neurotropic” strain, and later the “French neurotropic virus (FNV)”. FNV served as the platform by which to investigate YFV vaccine delivery methods, including the co-administration with the virus of human immune serum (Sawyer, Kitchen, and W. Lloyd 1932). The attendant impracticality of serovaccination, in conjunction with both the limited supply of serum and risk of pathogen contamination precluded widespread adoption of this technique. Through the 1940’s onward, FNV was used extensively in French west Africa, with 84 million doses produced between 1939 and 1954, which were administered by cutaneous scarification from an Arabic gum suspension of infected and desiccated mouse brain; FNV was frequently co-administered with smallpox vaccine (Durieux 1956b). Reference materials for FNV are exceedingly rare, however a number of FNV collection strains were partially sequenced in reference to the parental strain FVV and compared for neurotropic properties *in vivo* (E. Wang et al. 1995).

Although FNV was an excellent vaccine, there was recognition that it was responsible for unacceptable incidence of post-vaccinal encephalitis in children (approximately 3:1000)(Stones and Macnamara 1955). Initially, this was deemed acceptable due to the greater risk to children in endemic areas of contracting YF. As the vaccine gained success in the control of YF disease in Francophone Africa, administration of FNV to children under 10 years was prohibited in 1961 (Vainio and

Cutts 1998). In 1980, FNV was discontinued by the Institut Pasteur.

### 1.6.2 Development of the 17D Strain

The Asibi strain was passaged 53 times in monkeys before attempting to attenuate by passage in non-native tissues. From this juncture in handling of the WT strain, it was referred to as the “pantropic” or Asibi strain, meaning that the virus possessed both neurotropic and viscerotropic affinities. Under supervision of Max Theiler at the Rockefeller Foundation laboratories in New York, the Asibi strain was passaged in successive tissue preparations; a passage series called 17D was performed in minced mouse embryo (18x), minced whole chick embryo (58x) and finally minced whole chick embryo without brain and spinal cord (100x) (Theiler and H. H. Smith 1937). At stages along the passage series, the virus was assayed for neurotropism in mice by i.c. injection and pathogenicity in monkeys by both i.c. and extraneural administration. At the 114th passage (in reference to the parental strain Asibi), the virus was observed to have lost pathogenicity for both mice and monkeys. For monkeys challenged by i.c. or subcutaneous routes with the 114th passage of 17D, the strain produced a survivable encephalitis, with transient fever. 17D did not cause significant viremia in monkeys, and as such the virus was considered to have lost WT viscerotropism. The 176th passage of the virus was less lethal by average survival time in mice than a comparator strain at the 114th passage, when delivered i.c. and was selected for use as a vaccine. 17D infects *Aedes spp.* mosquitoes, however the dose required is greater than that required of for WT strains (Whitman 2015). The attenuated strain does not disseminate to the tissues of the mosquito, preventing transmission (B. R. Miller and Adkins 1988). The simultaneous loss of viscerotropism, neurotropism, and infectivity to mosquitoes ultimately rendered 17D to be safer than FNV, although immunogenicity of 17D was lower than that of the French strain (World Health Organization 1946; Stuart 1953). As with FNV, field

immunogenicity trials of 17D were first undertaken using co-administration of normal human serum with the virus (non-aqueous 17D). 17D preparations without added serum (aqueous-base) were standardized in 1945, and is the current form of the vaccine; administration is by the subcutaneous route (UNRAA Expert Commission on Quarantine 1945). There have been few changes in the seed-lot system since 1945 and over 600 million doses of 17D have been distributed (A. D. T. Barrett and Teuwen 2009).

## 1.7 Efficacy and Use to the Public Health

17D vaccines are considered an essential component of childhood vaccination schedules in YF endemic countries, although the early 21st century has brought increased attention on programmatic lapses in YFV vaccination as a matter of economic development. Though surveillance of YFV is incomplete across the endemic range, recent estimates of YF disease burden modeled from African serosurveys and case data indicate that 1.3 million cases (95%CI = 850,000-1.8 million) would have occurred in 2013, resulting in 78,000 deaths (95%CI = 19,000-180,000) (Garske et al. 2014). Vaccination coverage was estimated to have prevented 450,000 cases (95%CI = 340,000-560,000) and 28,000 deaths (95%CI = 7,200-62,000) in the same year, attesting *in silico* to a quantified, reduced burden of disease as a consequence of YF vaccination campaigns. Widespread mass vaccination campaigns in a number of African countries with FNV and 17D vaccines between the 1940s and 1960s had resulted in the almost-complete disappearance of yellow fever. However, immunization campaigns waned in the mid-1960s. Consequently, since the mid-2000s the nongovernmental Global Alliance for Vaccines and Immunizations (GAVI) has committed considerable resources to support of childhood YFV vaccination in endemic countries, including efforts supporting increased production of 17D vaccines from

WHO-prequalified manufacturers (Global Alliance for Vaccines and Immunization (GAVI) 2015). Since 2006, 12 countries have completed preventive yellow fever vaccination campaigns involving over 70 million doses of vaccine with financial support from GAVI Alliance, and its partners including the European Community Humanitarian Office (ECHO) and the Central Emergency Response Fund (CERF). These programs are ongoing, including a major program of vaccination in Nigeria.

Currently, over 35 of the 44 countries at risk for yellow fever in Africa and the Americas have routine infant immunization programs. Large scale vaccination campaigns require enormous amount of vaccine. To meet increasing demand for vaccine, intradermal (i.d.) delivery of 17D has been investigated as an alternative to subcutaneous administration under a hypothesis that less virus would be needed to induce protection; the strategy represents both a dose-saving approach to vaccination and a potential administration route for recipients with egg allergy (the vaccine is produced in embryonated chicken eggs). In healthy adult recipients, immunogenicity and safety measurements were equivalent in a comparison of i.d. delivery (1/5 typical dose) of 17D to the same lot of vaccine administered conventionally (Roukens, Vossen, Bredenbeek, et al. 2008). Additionally, a retrospective cohort (n=7) of individuals who failed the 17D skin allergy test (i.d. administration of 1/5 typical dose) were found to have seroconverted subsequent to the test (Roukens, Vossen, Dissel, et al. 2009). Results are promising, however more studies are needed to ascertain the true dose-response relationship of i.d. vaccine load with adverse events, immunogenicity, and viremia.

## **1.8 Lineage of 17D and the Current Vaccine Standard Genotype**

The lineage of 17D has been extensively described, consisting of seed and vaccine lots from the three “substrains” 17D-204, 17D-213, and 17DD (World Health

Organization 2010; Monath, M. Gershman, et al. 2013), where *seed lot* is a regulatory term used to define a particular batch of vaccine strain virus that is rigorously qualified for both phenotype and genotype before use in the production of bulk vaccine batches (World Health Organization 2010). The aforementioned technical standards define the use of *primary* and *secondary* seed lots, which define a single-passage chain to arrive at the vaccine, e.g. secondary seed is a single egg passage from the primary seed, and the vaccine is a single passage from the secondary seed, respectively. By means of the seed-lot system, the exact passage level of vaccine production batches is known.

Briefly, Theiler’s originating 17D strain at passage 176 was taken and the basal substrains 17D-204 and 17DD originate from subculture levels 204 and 195, respectively [Figure 1.2]. The 17D-213 substrain was produced by the Robert Koch Institute (RKI) in 1977 as a derivative of 17D-204, to be free from contamination by avian leukosis virus (ALV), a retrovirus producing sarcomas in chickens, but with no known pathogenic effects in humans (Bres and Koch 1987). All 17D seeds in current use are free of ALV, and standard production methods require use of eggs from ALV-free flocks (World Health Organization 2010; Monath, M. Gershman, et al. 2013). Since the 1940s there have been many vaccine producers from over 10 countries but currently there are only six producers: United States (17D-204, YF-Vax<sup>®</sup>, Sanofi-Pasteur), France (17D-204, Stamaril<sup>®</sup>, Sanofi-Pasteur), Senegal (17D-204, Institut Pasteur, Dakar), the People’s Republic of China (17D-204, Tiantan<sup>®</sup>, Wuhan Institute of Biological Products), the Russian Federation (17D-213, Chumakov Institute of Poliomyelitis and Viral Encephalitides), and Brazil (17DD, Bio-Manguinhos/Fundação Oswaldo Cruz (Fiocruz)). The French, Russian, Senegalese and Brazilian producers are prequalified by the WHO, serve international markets, and are used for mass vaccination campaigns. The U.S. and Chinese vaccines are

used in domestic markets only.

Initial genomic sequence comparison of 17D in reference to the WT parental Asibi strain was performed using a 17D-204 isolate obtained from the American type culture collection (ATCC) (Hahn et al. 1987). Similar information was later reported for the 17DD and 17D-213 substrains of the vaccine, which permitted elucidation of the conserved sequence features between all substrains (C. N. d. Santos et al. 1995). Current data indicates that the three vaccine substrains share 20 common amino acid substitutions, and four nucleotide substitutions in the 3'UTR [Table 1.1]. The contribution of the various amino acid substitutions and 3'UTR nucleotide changes to the attenuated phenotype is poorly understood (see below). Minimal consensus sequence divergence has been noted between substrains of 17D. Bayesian phylogenetic analysis for a set of 17D substrain consensus sequences was unable to resolve taxonomic groupings for 17D-204 and 17D-213 isolates, indicating very limited sequence divergence between these substrains for the clustering model employed (Stock et al. 2012). These data are consistent with the very effective seed-lot system developed in the 1940s and probably contribute to the excellent safety record of 17D vaccine.

Little is known about the molecular basis of attenuation for the 17D vaccine. The 17D and FNV vaccines share two common attenuated mutations in reference to their parental strains (M-L36F and NS4B-I95M), however the influence of these residues on attenuation is unknown. The residue M-36 (L in Asibi, F in 17D) is contained in a pro-apoptotic sequence described for dengue virus (Catteau 2003). Using a series of 17D/Asibi point mutants localized in the envelope protein, Lee and colleagues observed that the capacity of 17D to bind to heparin sulfate was reduced for viruses containing the WT residues E-380T and E-325P (the latter only found in 17D-204 and 17D-213 substrain vaccines); the substitutions also produced a low-neurovirulence phenotype in mice (E. Lee and Lobigs 2008). In summary,

determinants of attenuation have been explored to a limited extent. Findings thus far have indicated the influence of alterations in binding of virus to cells, but the effects are likely multigenic. The WT and vaccine strain may differ in capacity to antagonize innate immune responses in the host, although this has not been definitively shown for any 17D genotype (Muñoz-Jordán et al. 2005).

Genetic determinants of mosquito infectivity for the vaccine were investigated by Higgs and coworkers, in which Asibi-17D chimeric infectious clones were constructed to contain swapped residues in NS2A (four amino acids), NS4B (one amino acid), and the 3'UTR (four nucleotides) (McElroy et al. 2006). Dissemination of the chimeric viruses to salivary glands of orally infected *Aedes aegypti* was reduced for WT (Asibi) backbone constructs bearing 17D residues. In *Ae. aegypti* the 17D envelope residue E-380R was paradoxically observed to increase titers of the Asibi strain in mosquito salivary glands (Y.-J. S. Huang et al. 2014). Although not explaining attenuation in the vaccinee, reduced mosquito infectivity of 17D is hypothesized to contribute to the safety profile of the vaccine by preventing transmission of the attenuated virus post-administration.

## 1.9 Nonclinical Comparison of 17D Substrains

Nonclinical comparison of 17D substrain vaccines is infrequently reported. Vaccine seeds are evaluated using a standard WHO monkey neurovirulence assay, in which 10 rhesus or cynomolgous (*Macaca fascicularis*) macaques are administered a known quantity of seed virus i.c., and monitored for 30 days. Clinical and histological outcomes of neurotropism are compared to an equivalently sized group challenged with a reference strain of known acceptable properties; viscerotropism (serum viremia) and immunogenicity (plaque reduction neutralization test–50 percent (PRNT<sub>50</sub>)) are assessed by numeric limits. Minor and coworkers used this



method to compare three 17D seeds of differing substrain origin (Minor 2011). The WHO reference virus strain 168-73 (17D-213 substrain) was of equivalent immunogenicity to secondary seeds originating from Senegal and the United Kingdom (both 17D-204), however greater viremias and lower histological scores were observed for 168-73 when compared to the seeds.

### **1.10 Clinical Comparison of 17D Substrains**

Very little clinical information exists on the comparative performance of vaccines derived from the 17D substrains in human subjects, as modern efficacy trials would be considered unethical due to known historical properties of the vaccine and the severity of YF disease. Some non-inferiority trials have been reported for 17D-derived products, both within and between substrains. A comparison of two 17D-204 substrain products Stamaril® and Arilvax® (Chiron; originally produced by Wellcome labs in the U.K. and no longer manufactured) was performed in a healthy adult cohort (n=211) (Lang et al. 1999). Both vaccines were found to be immunogenic, observing similar rates of local and systemic adverse reactions; the test group receiving Stamaril® developed moderately but significantly higher neutralizing antibody titers at days 10 and 28 post-administration. Again for two products within the 17D-204 substrain, Arilvax® and YF-Vax® were compared in a cohort of Peruvian children of ages ranging from 9 to 10 years, observing similar rates of adverse events for both products; seroconversion rates differed significantly and were 94.9% and 90.6%, respectively (Belmusto-Worn et al. 2005).

For studies comparing 17D vaccines produced from different substrains, a series of trials were reported by Fiocruz comparing the immunogenicity and reactogenicity of two serially generated Brazilian 17DD substrain products with that of a 17D-213 substrain derivative vaccine (Camacho, M. d. S. Freire, et al. 2004; Cama-

cho, Aguiar, et al. 2005). In a healthy adult cohort (n=1087), neutralizing antibody titers were equivalent for the 17DD vaccines investigated, while moderately higher for the vaccine derived from the 17D-213 substrain, although immunogenicity results for all groups would be considered highly protective. In seronegative recipients, seroconversion rates for the three test groups were equivalent ( $\geq 98\%$ ), and no significant differences in rates of post-vaccinal adverse events were reported. Comparison of the 17D-204 substrain Stamaril<sup>®</sup> with 17D-213 substrain products RKI-YF<sup>®</sup> (RKI, Germany) and Berna-YF(Flavimun<sup>®</sup>) (neither vaccine produced currently) was reported in a cohort of healthy adults (n=304), from which was observed statistical dissimilarity of neutralizing titers between the groups receiving Berna-YF<sup>®</sup> and Stamaril<sup>®</sup> (Pfister et al. 2005). Statistically significant difference of neutralizing titers was not observed between groups receiving RKI-YF<sup>®</sup> and Berna-YF<sup>®</sup>, and all subjects seroconverted (PRNT<sub>50</sub> $\geq 1:10$ ); the result is of particular interest for the reason that the two vaccines are identically related on the 17D lineage, arising from passage 228 relative to the parental strain and from the same secondary seed. For a cohort of healthy Argentinian adults and children (n=2514, ages=1-70 years), the Fiocruz 17DD vaccine was analyzed for comparative safety to a vaccine derived from 17D-213 substrain; no statistical differences in either local or systemic adverse events were reported, immunogenicity data were not analyzed (Ripoll et al. 2014). In summary, there is little evidence to suggest divergence of phenotype across the 17D substrain lineage. Differences of immunogenicity observed in clinical trials exist at very high ranges of neutralizing antibody titer, and as such would not likely represent any meaningful change in protection. Trials comparing outcomes for 17D vaccines are summarized in Table 1.2.

### 1.11 Use of 17D as a Recombinant Backbone of Other Vaccines

Owing to the considerable safety profile of 17D, the nonstructural genes of the vaccine have been used as a backbone in a number of vaccine candidates designed to deliver both flaviviral and non-related antigens. Specifically, infectious clones of 17D virus have been generated with the 17D genome complementary DNA (cDNA) in one or more plasmids; using reverse genetics the cDNA can be mutated to swap or insert foreign sequences. The cDNA is transcribed to RNA *in vitro* (which is equivalent to YFV genomes), transfected into cells, and is able to replicate in cells to make a recombinant virus carrying desired mutations (Rice, Grakoui, et al. 1989). Thus, the 17D genome is variably amenable to insertion of heterologous antigenic elements; the strategy is used in a number of flavivirus vaccine candidates at various preclinical and clinical development stages. Recombinant 17D virus bearing inserts of known *Plasmodium* epitopes showed moderate immunogenicity to the insert peptides in monkeys, while both immunogenicity to the 17D backbone and viremia were reduced in comparison to the parental 17D vaccine, which was hypothesized to result from partially abrogated endosomal fusion in the recombinant virus (Bonaldo, Garratt, et al. 2005). A recombinant 17DD construct expressing from the amastigote surface protein 2 (ASP-2) of *Trypanasoma cruzi* was partially protective in a susceptible mouse model, a significant result considering the dearth of preventatives and treatments for Chagas' disease (Nogueira et al. 2013). A recombinant 17D was investigated as a backbone to stimulate cellular immunity against human immunodeficiency virus (HIV) by delivering a known CD8<sup>+</sup> T cell epitopes (Bonaldo, Martins, et al. 2010). The epitope was derived from a known immunodominant section of the simian immunodeficiency virus (SIV) Gag polyprotein, ligated to be expressed by a nonstructural region of the 17D genome. With this construct, CD8<sup>+</sup> T cell responses to the 17D vector were stimulated in a nonhuman primate model, with

specific cellular immunogenicity to the SIV/gag polyprotein. Recombinant 17D virus with Lassa virus (LaV) glycoproteins were generated by insertion of the LaV GP1 and GP2 genes into the C-terminal region of 17D E protein gene; the construct was immunogenic in a guinea pig model of LaV infection, although less protective (80%) by comparison to a live-attenuated Mopeia virus/LaV reassortant (Bredenbeek et al. 2006).

The ChimeriVax<sup>®</sup> (Sanofi-Pasteur, Lyon, France) platform is perhaps the most advanced in development of this class of vaccines, and is derived by swap chimeras of heterologous flavivirus prM/M-E structural cassettes into the nonstructural backbone of the 17D genome (Bray and Lai 1991). Resulting recombinant vaccines are efficiently delivered in a live-attenuated context with a safety profile afforded by the 17D nonstructural genes (Guy et al. 2011). To date, one chimera, ChimeriVax-JE (Imojev<sup>®</sup>), has progressed to licensure and is registered in nine countries (Appaiahgari and Vрати 2010). The most advanced candidate in development using this technology is the tetravalent DENV preparation. Two phase III, multi-center trials of the ChimeriVax<sup>®</sup> DENV candidate were recently reported, testing efficacy to prevent hospitalized dengue disease in Asian and Latin American pediatric cohorts, for which combined-serotype efficacy was 56.5% and 60.8%, respectively (Capeding et al. 2014; Villar et al. 2015). Similarly, the ChimeriVax<sup>®</sup> WNV candidate WNV02 was recently assessed in a dose-ranging, phase II safety trial enrolling 208 healthy adults for three age cohorts, revealing 96 percent seroconversion at 28 days that was correlated with vaccine dose (Monath, Liu, et al. 2006; Biedenbender et al. 2011). While these studies with mosquito-borne flaviviruses suggest that the nonstructural protein genes of 17D encode the attenuated phenotype, utilization of the ChimeriVax<sup>®</sup> technology for a tick-borne encephalitis virus prM/M-E chimera did not produce the desired attenuated phenotype in preclinical

studies (Rumyantsev et al. 2013).

### **1.12 New Report of A Serious Adverse Reaction to 17D**

Serious adverse events (SAEs) following vaccination with 17D are rare, and generally are categorized as either neurotropic, referring to infection of the central nervous system (yellow fever vaccine-associated neurotropic disease (YEL-AND)) or viscerotropic (yellow fever vaccine-associated viscerotropic disease (YEL-AVD)), referring to multi organ disease with infection of the liver as is similarly observed in WT YF. Clinical data on SAE are consequently rare, but support a paradigm of SAE occurrence that is considerably influenced by host factors rather than reversion of the vaccine virus to a virulent phenotype.

A recent case of fatal YEL-AVD, the 65th reported case of this type, occurred in 2014, in a female recipient receiving the vaccine for travel to South America (DeSilva et al. 2015). The patient developed signs consistent with the WHO case definition of YEL-AVD, eventually succumbing to systemic disease (M. D. Gershman et al. 2012). YFV antigen was observed in recovered tissues by immunostain, with positive IgM titer of 1:640. On autopsy, a thymoma was found in association with serology indicating myasthenia gravis, which if previously identified would potentially have contraindicated receipt of the vaccine. The directly precedent reported cases of YEL-AVD occurred in 2009, which was not reported in open literature (Monath, M. Gershman, et al. 2013, p. 952). Preceding this in 2007, in Peru, five individuals were treated for systemic viscerotropic disease following 17D vaccination in a large campaign, with four fatalities (Whittembury et al. 2009). The consensus genomic sequence of the originating vaccine lot and three of the fatal cases were determined, and there was no sequence divergence from the secondary seed lot used in manufacture. Single-nucleotide polymorphism content from the 2007 cluster and

previous YEL-AVD case tissues were sequenced and compared at loci hypothesized to influence innate immune responses, but the results did not describe association of particular single-nucleotide polymorphisms (SNPs) with the cases. Because the case descriptions include evidence of altered immune status for the vaccinees in association with viscerotropic disease, host factors are likely contributors.

Post-vaccinal SAE have been extensively reviewed (A. D. T. Barrett and Teuwen 2009). Risk factors for YEL-AVD are estimated from the limited clinical data available and include age (>60 years), systemic lupus erythematosus, thymectomy/thymoma, or other autoimmune condition (Monath, Cetron, et al. 2005; Staples, M. Gershman, and Fischer 2010). A hypothesis that competent innate immune responses are required to prevent dissemination of 17D is supported by animal models of YFV disease, in which 17D vaccine virus is lethal in mice lacking  $\alpha/\beta/\gamma$  IFN receptors (AG129) (Meier et al. 2009). Similarly, widespread dissemination of genetically tagged 17D strains was observed in a mouse model lacking  $\alpha/\beta$  IFN receptors (Erickson and Pfeiffer 2013). In a study comparing naive 17D recipients in young and elderly age groups, post-administration viremias in the older subjects were found to be relatively greater (Roukens, Soonawala, et al. 2011). Mutations localized to CCR5/RANTES genes or their regulatory elements are associated with variable response to viral infection, and thus have been hypothesized to influence development of YEL-AVD, however the association of known alleles with the adverse syndrome is inconclusive (Pulendran et al. 2008).

### **1.13 New Vaccine Candidate Technologies and Monitoring of Safety**

As aforementioned, the occurrence of adverse events is rare, however the use of inactivated YFV vaccines has been proposed to ameliorate concerns of adverse reactions for recipients for whom administration of live 17D strains would be con-

traindicated, including the elderly, immunosuppressed, or those with egg allergy. Design of inactivated YFV vaccines have typically been frustrated by necessity to deliver multiple doses to achieve protection in typical neurovirulent mouse models of YFV infection.

Nonclinical data for two inactivated, whole-virion 17D candidates has been reported. Characterization of the  $\beta$ -propiolactone-inactivated 17D-204 candidate XRX-001 (Xcellerex/GE Healthcare, Marlborough, MA) was performed, showing that the vaccine was protective in Syrian golden hamsters against challenge by the hamster-adapted Jimenez strain of YFV (Monath, C. K. Lee, et al. 2010). Immunogenicity of XRX-001 was assessed in cynomolgous macaques for two and three-dose administration schedules, from which sera variably neutralized the heterologous 14-FA (Angola) strain. Results of a phase I safety and dose-ranging trial of XRX-001 in 60 recipients were reported in 2011, observing seroconversion (endpoint/baseline PRNT<sub>50</sub>>4) for all participants receiving a booster dose at 21 days (Monath, Fowler, et al. 2011). Although neutralizing antibody titers for the two-dose administration of XRX-001 did not exceed those typically found in clinical studies of live-attenuated 17D vaccines, the seropositivity criterion used (PRNT<sub>50</sub>>10) is thought to confer protection in humans (Camacho, M. d. S. Freire, et al. 2004). Two inactivated 17D candidates have been reported by the Oswaldo Cruz Foundation (Brazil). The first of these to be characterized was a pressure-inactivated, 17DD-substrain candidate; immunogenicity data in mice revealed that three doses were necessary to produce equivalent neutralization titers to those subjects receiving live 17DD vaccine (Gaspar et al. 2008). The second, a  $\beta$ -propiolactone-inactivated 17DD substrain derivative was tested for immunogenicity in outbred mice (subcutaneous route), using a live-attenuated 17DD vaccine as comparison. Greatest immunogenicity (geometric mean titer (GMT)=922, 95%CI = 666-1274, 43.75% seroconversion) was observed in a

group receiving three doses, adjuvanted with aluminum hydroxide (Pereira et al. 2015). No information is available on the durability of protection offered by any inactivated 17D vaccine, however under reported test conditions, immunogenicities for these exceed typical standards for seroconversion, and as such would be expected to be protective.

Recombinant YFV vaccine candidates have been investigated using diverse vectoring strategies. Very commonly, constructs are designed to deliver the 17D prM/M-E structural cassette, which contains the predominant surface-exposed neutralizing epitopes of YFV and all flaviviruses. Delivery of YFV antigens by live-attenuated, recombinant vaccinia constructs was shown to be feasible in the late 1990s (Pincus et al. 1992). A recent effort used the replication-deficient modified vaccinia virus Ankara (MVA) and defective vaccinia virus (dVV) to vector the 17D prM/M-E antigen protected mice from i.c. challenge with 17D when administered in a single dose (17D substrains, though highly attenuated in humans, are virulent for mice when directly introduced to the brain) (Schäfer et al. 2011). The study not only demonstrated protection and immunogenicity of the constructs, but also high apparent safety in mice, as the vaccines themselves were not neurotropic when administered at doses exceeding those required to induce protection. A plasmid-vectored DNA (DNA) vaccine was constructed by fusion-ligation of the 17DD substrain prM/M-E onto that of human lysosomal-associated membrane protein 1 (LAMP-1), a strategy designed to target the antigen structural subunit for efficient trafficking to MHC-II compartments and consequent presentation and stimulation of cellular immune responses (Maciel et al. 2015). A 45 day, three-dose course of the LAMP-1 plasmid construct was protective when assayed in mice against live 17DD virus delivered i.c..



### 1.14 New Insights on Durability of Protection for 17D

Production of YFV vaccines is regulated using a seed-lot system, developed in 1945 in response to evidence that phenotypic instability of the vaccine could be introduced by improper handling resulting in excessive passaging of the vaccine virus (UNRAA Expert Commission on Quarantine 1945). Among other contributing findings, an early Brazilian trial cohort observed absence of immunogenicity for a group receiving vaccine from a seed at extended passage levels (Soper and H. H. Smith 1938). Presently, the protection conferred by 17D production seed lots is characterized for immunogenicity by neutralization antibody titers during performance of the WHO monkey neurovirulence test (criterion:  $\text{PRNT}_{50} > 1:10$ ), and by comparison trials in humans for lots of vaccine originating from newly generated seeds (World Health Organization 2010). Presently, a neutralization titer of 1 in 10 or a fourfold increase of neutralizing antibody from baseline is considered evidence of seroconversion. International health regulations (IHR) require that, for certain travelers in areas of YFV endemicity, a certificate of prophylaxis (“yellow card”) be issued by the vaccinator, which is considered valid for ten years following the tenth day post-administration (World Health Organization 2006).

The WHO Strategic Advisory Group of Experts (SAGE) working group for YF vaccination evaluated open and closed-literature findings on the durability of immunogenicity conferred by 17D (World Health Organization 2013). Specifically, the working group recommended revision to the 2003 WHO position on 17D booster schedules, recommending that both the WHO position and IHR be revised to reflect an expectation that 17D confers lifetime protection without the necessity of 10-year booster intervals. The amendment to the IHR was approved by the World Health Assembly, to be included in a 2016 revision. In February 2015, the Advisory Committee on Immunization Practices (ACIP) working group on yellow fever vaccines presented

these findings to the ACIP, which were adopted by unanimous vote of the committee (Centers for Disease Control and Prevention/Advisory Committee on Immunization Practices 2015). The 2013 SAGE findings were assembled into an extensive review of immunogenicity for 17D, which in numerous cases of seropositivity have been observed in vaccinees at intervals greater than 10 years post-administration (Gotuzzo, Yactayo, and Córdova 2013). Briefly, of 8 open-literature studies considered by the working group, duration of 17D immunity was maintained for 10 years in all datasets, with an upper reported boundary of 38 years (Niedrig et al. 1999). Some countries have questioned the proposed removal of booster doses and it remains to be seen whether or not all will elect to adopt the WHO recommendation.

### 1.15 New Insights on Immunogenicity for 17D

Mechanisms underlying the long-lasting immunogenicity of 17D are poorly understood, but are attributed in part to the combined stimulation of innate, humoral and cellular immunities from the presentation of endogenous antigen expressed by the live, replicating virus (Mudd et al. 2010). In a passive serum transfer experiment using the hamster adapted challenge model, neutralizing PRNT<sub>50</sub> titers of <1:40 were sufficient to confer protection, supporting the canonical model for protective humoral seroconversion in humans (Julander, Trent, and Monath 2011). Systems biology approaches have recently been deployed to interrogate these mechanisms, in order to more closely elucidate determinants of immunogenicity for the vaccine in the host. Significantly, studies of this type have identified key cytokines and innate response elements that determine both the diversity and longevity of the immune response to 17D. A series of *in silico* studies reported associations of networked antiviral response elements that were predictive of CD8<sup>+</sup> T cell responses to the vaccine (T. D. Querec et al. 2009). In a comparative study of several vaccines, humoral re-

sponses to 17D were partially correlated to cellular stress responses, specifically that of transcription factor ATF2, and translation factor kinase EIF2AK4, both markers for stress response and amino-acid starvation, which contributes to programming of DCs to present antigens to stimulate CD8<sup>+</sup> T cell immunity (S. Li et al. 2014). The cellular response to 17D is highly diverse, involving activation of multiple toll-like receptor (TLR) response elements and lineages of DCs (Gaucher et al. 2008). This results in the induction of various pro-inflammatory cytokines, including IL-12p40, IL-6, and interferon- $\alpha$ , and consequently the stimulation of both T and B cell responses. In a cohort of healthy adults receiving 17D vaccine, the viral load at day 14 post-administration correlated with effector CD8<sup>+</sup> T cell responses; a saturation effect for CD8<sup>+</sup> T cell response is observed above a certain viremia sustained in the vaccinee, above which no changes were observed (Akondy et al. 2015). The study is especially significant to understanding the immunological responses of 17D recipients under conditions of dose-sparing. In a comparison of 17DD and 17D-213 substrains in a pediatric cohort, the cytokine responses of the vaccines were similar, however revealing an enhanced inflammatory component for the group receiving the 17D-213 substrain vaccine (Campi-Azevedo et al. 2012). IFN-gamma responses are observed to correlate with humoral, CD4<sup>+</sup> T cell, and CD8<sup>+</sup> T cell responses (Neves, Rudersdorf, et al. 2010; Neves, J. R. Santos, et al. 2013). It is hoped that, by interrogation of specific immune response elements, predictive models of response to live-attenuated vaccines could be constructed, and in such a manner offer personalized insight to the safety and eventual protection available to the recipient.

### **1.16 New Paradigms and Legacy Technologies**

As aforementioned, 17D is manufactured in embryonated chicken eggs, a legacy technology that remains unchanged since standardization of the vaccine in

the 1940s. The adaptation of the vaccine to other cell based systems has been investigated, with the goal of efficiently producing the vaccine *in vitro*. The adaptive capacity of YFV has been investigated since the early development of all YFV vaccines, with multiple studies recognizing that passage history could reverse the attenuation or immunogenicity of the vaccine (Findlay and Clarke 1935; Soper and H. H. Smith 1938). A 17DD substrain seed was adapted to chick embryonic fibroblasts (CEFs), to permit growth of the vaccine lot by microcarrier bead culture technology (M. S. Freire et al. 2005). The CEF-prepared vaccine lot was tested using the standard WHO monkey neurovirulence test; the authors noted elevated neurotropism scores in the monkeys tested with respect to a conventionally prepared seed, although these results were not statistically significant. Adaptation of the inactivated 17D-204 candidate XRX-001 to Vero cell culture was performed, revealing some instability in the vaccine genome localized to the envelope, NS2A, and NS4B genes, although differences in phenotype conferred by mutations have not been tested in an infectious clone system (Beasley et al. 2013). The substitutions were stably observed between three independently conducted passage series, associated with greater yields of the virus from Vero cell cultivation. In both of these cases, the vaccines were safe and immunogenic in nonclinical studies involving nonhuman primates, however the known potential for instability of live virus phenotype under *in vitro* passage adaptation warrants close attention.

To this end, massively parallel sequencing (next-generation sequencing (NGS)) approaches have been deployed to interrogate the fine population structure of vaccines, including the presence of contaminants (Neverov and Chumakov 2010). These methods partially derive from theoretical insight of the *quasispecies* paradigm, in which viral diversity is considered to influence fitness *in vivo* (Andino and Domingo 2015). Recently, a commercial lot of YF-Vax<sup>®</sup> 17D-204 vaccine (Sanofi-Pasteur,

Swiftwater, USA) was sequenced by massively parallel methods recovering evidence of very low ALV contamination, confirmed by reverse transcription polymerase chain reaction (RT-PCR) (Victoria et al. 2010). Low coverages for YFV ( $n < 100$ ) prevented complete assessment of population structure for the targeted viruses. Comparison of a 17D-204 commercial lot with a collection lot of the parental Asibi strain was performed at ultra deep coverage ( $> 5000$ ), revealing that the WT virus population structure differs from that of the vaccine by virtue of relative homogeneity suggesting that the lack of diversity in the 17D vaccine virus population may contribute to the attenuated phenotype (study is presented in chapter 3) (A. Beck et al. 2014). Resolution of population structure for viral vaccines by NGS or other methods is a potential avenue by which to measure the presence of virulent subpopulations which naturally occur in RNA viruses, and in doing so predict safety profiles of the vaccines before use.

#### 1.16.1 Quasispecies Theory and Fitness of RNA Viruses

The complexity of RNA virus populations, although speculated by observations of intermediate phenotype on passage, was first rigorously shown in a set of passage experiments using the RNA bacteriophage Q $\beta$  (Domingo, Sabo, et al. 1978). For a set of specific variant populations identified through T1 oligonucleotide fingerprints, assays were performed to competitively passage the variants in equivalent multiplicity to that of a diverse WT phage, observing that the WT phage was more fit in *E.coli* with respect to growth rate. Additionally in this study, clonally selected populations of the virus were homogenous when subjected to T1 oligonucleotide fingerprinting, indicative of a bottleneck effect. In a later review, it was proposed that the 1978 findings of Domingo et. al. had observed an effect in which rare variants were negatively selected, and by that means had achieved only low equilibrium frequency in the WT competition assay (J. Holland et al. 1982). Likewise, a population

of minority variants arising in foot-and-mouth disease virus (FMDV) were serially passaged and observed to be unstable (Sobrino et al. 1983); the authors interpreted the results as a confirmation of the findings of Domingo et. al. (1978), only in the context of a virus causing mammalian disease (Domingo, Sabo, et al. 1978).

The apparent cloud of variably fit genomes is formally referred to as a *mutant spectrum*, from which the particular dynamics of the population are measured. Within the spectrum are considered to exist *minority variants*, which may arise randomly by biophysical mechanisms; the generation, timescale, and persistence of these variants in the viral population is considered a possible signal of not only the selection pressures encountered by the virus population, but also specific determinants of the fitness of the population as a whole. For RNA viruses specifically, population complexity is considered to arise from a combination of large population size, short generation times, and a lack of error-proofing or correction in genome replication. These effects have been extensively reviewed (Andino and Domingo 2015; Domingo, Sheldon, and Perales 2012).

Several possible mechanisms are observed by which viral fitness is influenced by diversity of the population. First, the presence of random, discrete minority variants may confer increased fitness to the particles carrying the subpopulation allele. Second, the viral population may itself be a unit of selection; this is a formal quasispecies effect. An effect of this type was demonstrated by Vignuzzi et. al, in which a high-fidelity polymerase mutant of poliovirus (PoV) was shown to be of low fitness to produce neurotropic disease in a mouse model; chemical mutagenesis rescued the effect in the absence of the capacity of the virus to mount a typical, WT population structure (Vignuzzi et al. 2006). These results were interpreted as a *cooperativity* effect, in which the change of fitness was attributed to the viral population *en masse*.

### 1.16.2 Viral Diversity and Pathogenesis

In addition to early studies using FMDV, numerous studies attest to the pathogenic influence of viral minority variants; these effects may be diverse, differentially influencing tissue tropism or immune responses. In many cases, studies are designed to reconstruct changes in viral diversity under selective pressure of passage or treatment. It is likely that pharmaceutical treatment of some viral infections is effective in part due to a population-level effect. In human samples, the population complexity of hepatitis B virus (HBV) was reduced in response to entecavir treatment, while high viral diversity was associated with advanced liver fibrosis (Nishijima et al. 2012). Likewise, in a group of transplant patients, diversity of hepatitis C virus (HCV) was associated with an immunocompromised state (Pessoa et al. 1999). The study of both HCV and HIV-1 quasispecies effects admit the role of “founder” and “transmitter” genotypes, which are thought to be minority variant populations with sufficient tropism to surmount an initial bottleneck when productively infecting a new host (Stoddard et al. 2015). In a study in which mice lacking  $\alpha/\beta$ -interferon receptors were challenged with YFV 17D, the subjects contracted a disseminated infection, suggestive that some low level of diversity in the attenuated vaccine is insufficient to traverse the innate immune response. For YFV this is substantiated by infrequent occurrence of YEL-AVD, which in some cases is associated with administration of 17D to persons with history of immunocompromise (Whittembury et al. 2009; DeSilva et al. 2015).

### 1.16.3 Viral Diversity and Ecology of Arboviruses

Resolution of population structure effects for arthropod-borne viruses like YFV is complicated by the complexity of the WT transmission cycle, which for YFV may include, at minimum, productive infection of (1) a short-lived mosquito

vector and a (2) a mammalian host, likely nonhuman primate, which may or may not result in a dead-end infection. Although quasispecies effects have not been rigorously described for arboviruses, one prominent framework is the “trade-off” hypothesis, which indicates that a certain level of diversity propels infection across barriers to which the virus is poorly adapted; e.g. viral fitness is partially compromised in all hosts, if the virus must infect them all to propagate. This phenomenon is shown by cell culture passage studies, in which adaptation of the virus to one passage host compromises fitness, while alternating passage in multiple, representative culture systems does not (Coffey et al. 2008a).

Mechanistic origins of viral diversity through the full transmission cycle are poorly understood; some evidence has been reported to suggest that for mosquito-borne viruses, alternating patterns of diversification (in one host) with purifying selection and (in the mammalian host) is typical; this was shown for WNV in alternating passage of chicks and *Culex pipiens* mosquitoes (Deardorff et al. 2011). The phenomenon was observed as a loss of fitness to infect mice for strains of Rift Valley fever virus (RVFV) that were passaged exclusively in either Aag (*Aedes aegyptii*) or BHK21 cells; this implies that for RVFV, the multiple host transmission cycle maintains the stability of the genome (Moutailler et al. 2011). Local tissue effects are likely to govern the ecologies of specific transmission cycles; a study in which Venezuelan equine encephalitis virus (VEEV) was passaged in mosquitoes revealed that stepwise infection events in the body of the insect represent population bottlenecks for the virus, increasing homogeneity of the population (Forrester et al. 2012). For YFV, the dynamics of population complexity with respect to the virus ecology are not known, but it is likely that some level of high diversity exists in WT strains due the numerous host and vector types that are infected in nature (Ellis and A. D. Barrett 2008).



#### **1.16.4 Deep Sequencing and Technical Approaches**

The dynamics and content of subpopulations for YFV are unknown, but could influence the phenotype and safety of vaccine products; the resolution of vaccine population structure represents a novel class of methods to ensure the stability of attenuating genotypes for vaccine strains. Recently, the use of deep sequencing techniques has been deployed to assay viral diversity directly from nucleic acids isolated from preparations of a veterinary Rabies Virus (RABV) vaccine, with subsequent measurements of distance applied to assess divergence (Höper et al. 2015). Deep sequencing of vaccine has been extensively reviewed (Luciani, Bull, and A. R. Lloyd 2012; Kennedy and Poland 2011; Neverov and Chumakov 2010). For YFV, the ideal use case for NGS would be in the qualification of vaccine lots once filled. More essentially in cases of regeneration for primary or secondary seed lots, NGS would ideally be used to ensure negligible divergence of the vaccine population genotype from one previously accepted or qualified. Significantly, NGS methods would facilitate the development of a predictive model by which to assess the phenotype conferred by subpopulation alleles in the vaccine under consideration.

#### **1.16.5 Connecting Viral Diversity to Flavivirus Vaccinology**

For YFV specifically, some evidence has been put forth to suggest that the virus is genetically stable as a general phenomenon; e.g. YFV strains are observed to incorporate very few mutations per generation. This property is also inferred phylogenetically; a study comparing consensus sequence divergence over time for DENV and YFV revealed that, over the timescale of the phylogeny, YFV sequences mutated at a significantly lower substitution rate per unit time (Sall et al. 2010). It is likely, however that this result occurred due to the differing ecologies of YFV and DENV; YFV, though occasionally emergent, is consistently maintained in nature

and is not known to be influenced by expansion dynamics, whereas the four DENV serotypes are in geographic flux (Messina et al. 2014; Afreen et al. 2015). Therefore, the reported phylogenetic stability of YFV is more likely a consequence of the limited host range to which the virus had previously expanded and adapted.

Studies on the 17D vaccine have shown the strain to be of high stability, with a very limited amount of sequence diversity arising across passage. In a study of the 17D strain, in this case the Arilvax<sup>®</sup> 17D-204 substrain vaccine product, Sanger sequencing revealed a set of nucleotide positions in the vaccine with noncoding heterogeneity, as evidenced by divided chromatogram populations (Pugachev, Ocran, et al. 2002). In particular, the mutation at E-A240V, a site previously attributed to influence neurovirulence of the vaccine strain (Ryman et al. 1997). A later study by the same laboratory passaged the 17D vaccine under controlled conditions with subsequent conventional sequencing; the mutation rate of 17D was estimated by this study to be  $1.9 \times 10^{-7}$  -  $2.3 \times 10^{-7}$  substitutions per replicated RNA molecule (Pugachev, Guirakhoo, et al. 2003). The aforementioned study of 17D strains by Bayesian phylogenetic methods found poor evidence for diversification of the vaccine along the length of the vaccine lineage; there was notably poor resolution of known branch points in the lineage, indicating either an equilibrium of selective pressures along the genotype or very strong purifying selection of the vaccine genotype in egg culture (Stock et al. 2012). Owing to the uniquely stable features of 17D, the population structure of the virus is likely to contain key signals of attenuation.

### 1.17 Specific Aims

The **overall objective** of the dissertation research is to characterize typical population structures for YFV vaccine strains, and in doing so provide context for the influence of population-level effects on the attenuation mechanisms of 17D and

other live-attenuated viral vaccines. The **central hypothesis** of the dissertation research is that population structures of WT YFV strains are distinct from those of the vaccine strain 17D. This hypothesis was tested by the analysis *in silico* of the population structures for YFV 17D vaccine strains of known phenotype, with the ultimate goal of resolving elements of YFV quasispecies effects that may influence the phenotype of the vaccine. The overall hypothesis for Aim 1 is that by using deep sequencing to compare a commercial lot of 17D vaccine with the parental strain from which the vaccine was derived, evidence of the tissue selection process will be resolved, most evidently by changes in nucleotide diversity. The overall hypothesis for Aim 2A is that by comparing a set of standardized and phenotypically stable 17D vaccine seed strains, evidence will be provided for the stability of the vaccines on a subpopulation level; these results will provide a framework for evaluating small changes in population structure for the vaccines over long passage histories represented by the seeds. Also as a question of passage stability, the overall hypothesis for Aim 2B is that in a comparison of quasispecies structure between the prototype WT strain Asibi and a hamster-adapted (hamster/P7) strain of the same, specific viral subpopulations influencing viscerotropy in the hamster model will be observed. The overall hypothesis for Aim 3 is that by using a deep sequencing study to compare the population structures of collection strains of FNV, it will be found that previously described differences in neurotropism measures for the strains will be accompanied by destabilization of the strain populations. The overall hypothesis for Aim 4 is that when using deep sequencing to compare the population structures of the WT JEV SA14 strain to the vaccine derivative SA14-14-2, patterns of diversity change recovered in the original Asibi/17D comparison will similarly be observed.

#### 1.17.1 Specific Aim 1

**Aim:** Compare population structure of the YFV parental strain Asibi with that of the vaccine derivative 17D-204.

**Hypothesis:** Quasispecies structures of YFV strains Asibi and 17D-204 strains of YFV are distinct, and informative for studies of vaccine attenuation.

**Rationale:** There is previous evidence by classical techniques that YFV has some level of population diversity, but this is completely undescribed for YFV.

#### 1.17.2 Specific Aim 2A

**Aim:** Analysis by deep sequencing of YFV vaccine strains and seeds from a historical collection representing the entire lineage of the 17D vaccine substrains.

**Hypothesis:** Population structures of standardized 17D production seeds will show evidence of stability, with considerable similarity of nucleotide content between strains.

**Rationale:** It is not known as to whether the subpopulations of YFV vaccine seeds are stable in relation to each other.

#### 1.17.3 Specific Aim 2B

**Aim:** Analyze changes in population structure for a viscerotropic YFV strain produced by serial passage adaptation of the Asibi strain in hamster liver.

**Hypothesis:** Population structures of the WT Asibi strain and that of the Hamster/P7 strain are distinct, and will particularly identify changes in diversity during serial passage.

**Rationale:** Due to other evidence that passage, this dataset is of particular importance to understanding the population dynamics involved in the adaptation of

YFV to cause severe liver disease.

#### 1.17.4 Specific Aim 3

**Aim:** Translate findings of population structure for 17D to characterize the French neurotropic virus, a discontinued and unsafe YFV vaccine.

**Hypothesis:** Instability of population structure is associated with instability of neurovirulent phenotype for a set of FNV collection strains.

**Rationale:** Since FNV strains were previously characterized to be unequally neurotropic, *in silico* analyses will reveal instability of consensus identity, diversity, nucleotide selection, and single-nucleotide variation.

#### 1.17.5 Specific Aim 4

**Aim:** Extend analyses of YFV strains by resolving population structure of JEV strain SA-14-14-2, an empiric flaviviral vaccine.

**Hypothesis:** Patterns observed *in silico* for YFV vaccines will be recapitulated for a comparison of the WT JEV strain SA-14 and the live-attenuated vaccine strain SA-14-14-2.

**Rationale:** Previous work observed that the YFV vaccine 17D-204 was of homogenous population structure relative to the WT parental strain Asibi. As SA14-14-2 was attenuated by serial passage, this pattern will be similarly observed in the JEV dataset.

**Table 1.1:** Amino acid substitutions contained in the WHO standard genotype for YFV vaccines.

Position	Gene	Codon	Asibi	17D
854	M	36	L	F
1127	E	52	G	R
1482		170	A	V
1491		173	T	I
1572		200	K	T
1870		299	M	I
1887		305	S	F
2112		380	T	R
2193		407	A	V
3371	NS1	307	I	V
3860	NS2A	118	M	V
4007		167	T	A
4022		172	T	A
4056		183	S	F
4505	NS2B	109	I	L
6023	NS3	485	D	N
6876	NS4A	146	V	A
7171	NS4B	95	I	M
10142	NS5	836	E	K
10338		901	P	L
10367	3'UTR	-	U	C
10418		-	U	C
10800		-	G	A
10847		-	A	C

**Table 1.2:** Published clinical and nonclinical data comparing the 17D substrains.

<b>Trial Type and Reference</b>	<b>Cohort</b>	<b>Endpoints</b>	<b>Strain (Substrain, National Origin)</b>	<b>Results</b>
Nonclinical <sup>a</sup>	Cynomolgous Macaques	Viremia (Viscerotropism), Immunogenicity	1.168-73 (17D-213, WHO Reference Strain) 2.17D-204 Secondary Seed, UK 3.17D-204 Secondary Seed, Senegal	1.All groups of equivalent immunogenicity. 2.Greater viremias observed for group receiving 168-73. 3.Lower histological scores for group receiving 168-73.
Clinical <sup>b</sup>	Healthy Adults (n=211)	Immunogenicity, adverse reactions	1.Stamaril <sup>®</sup> (17D-204, France) 2.Arivax <sup>®</sup> (17D-204, UK)	1.Similar rates of adverse reactions and seroconversion for both vaccines. 2.Higher immunogenicity for group receiving Stamaril <sup>®</sup> .
Clinical <sup>c</sup>	Children, 9mos.-10yrs. (n=1,107)	Immunogenicity, Adverse reactions	1.YF-Vax <sup>®</sup> (17D-204, USA) 2.Arivax <sup>®</sup> (17D-204, UK)	1.Greater seroconversion rates for Arivax <sup>®</sup> . 2.Similar rates of adverse events. 3.Similar immunogenicity for both vaccines.
Clinical <sup>d</sup>	Healthy Adults (n=1087)	Immunogenicity, Adverse Reactions	1.17D-WHO (17D-213, Brazil) 2.17DD013Z (17DD, Brazil) 3.17DD10284 (17DD, Brazil)	1.Greater immunogenicity for 17D-213 substrain vaccine. 2.Similar rates of seroconversion for all vaccines considered. 3. Slightly higher rates of minor adverse reactions and detectable viremia for 17D-213 substrain.

Clinical <sup>e</sup>	Healthy Adults (n=304)	Immunogenicity, Adverse Reactions	1.Stamaril <sup>®</sup> (17D-204, France) 2.RKI-YF <sup>®</sup> (17D-213, Germany) 3.Berna-YF <sup>®</sup> (17D-213, Netherlands)	1.All groups immunogenic (GMT >1:10). 2.Higher ab titers for males receiving 17D-213 substrain products. 3.No difference in immunogenicities between 17D-213 products administered. 4.Higher rate of injection site erythema and pain for RKI-YF <sup>®</sup> , compared to Berna-YF <sup>®</sup> .
Clinical <sup>f</sup>	Healthy Adults and Children (n=2514)	Adverse Reactions	1.17DD/Fiocruz (17DD, Brazil) 2.Stamaril <sup>®</sup> (17D-204, France)	No significant differences in rates of local or systemic adverse events.

<sup>a</sup> (Minor 2011)

<sup>b</sup> (Lang et al. 1999)

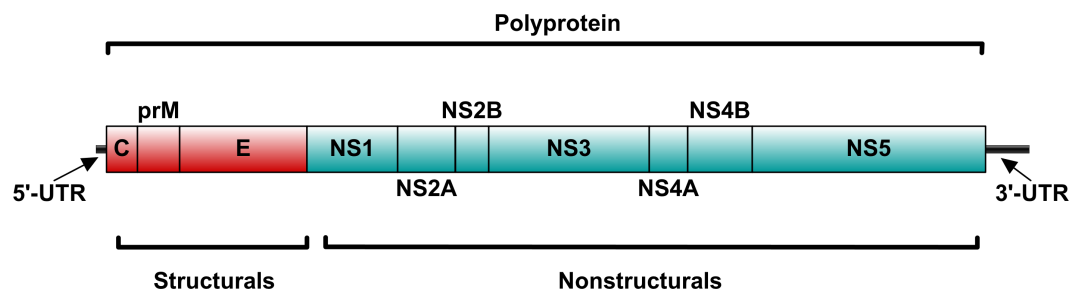
<sup>c</sup> (Belmusto-Worn et al. 2005)

<sup>d</sup> (Camacho, M. d. S. Freire, et al. 2004; Camacho, Aguiar, et al. 2005)

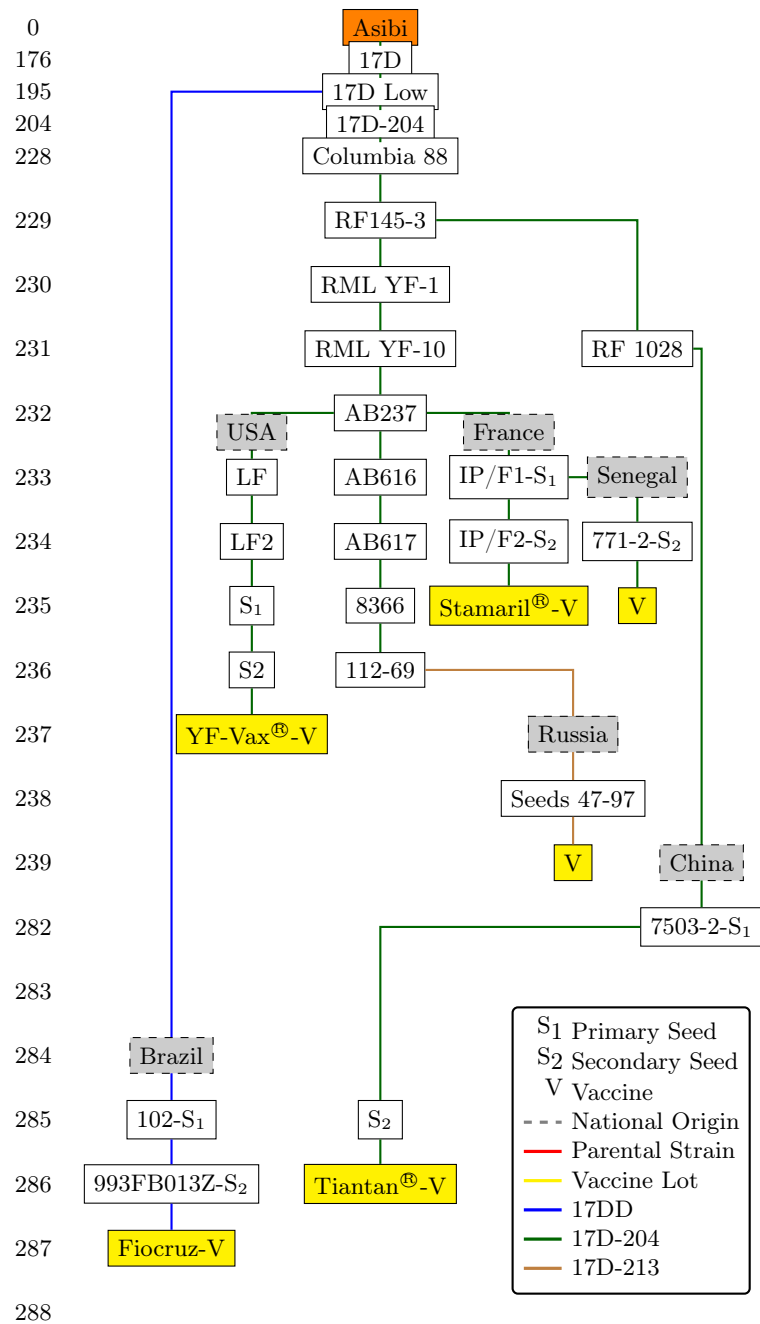
<sup>e</sup> (Pfister et al. 2005)

<sup>f</sup> (Ripoll et al. 2014)





**Figure 1.1:** Canonical genome organization of YFV and other flaviviruses.



**Figure 1.2:** Current production for YFV 17D vaccines. Includes all countries and producers, with substrain origins indicated. Intervening seeds are shown, but are not exhaustive. Adapted from WHO technical standards (World Health Organization 2010).

## Chapter 2

### Materials and Methods

#### 2.1 Virus Strains and Passage Histories, By Chapter

##### 2.1.1 Chapter 3: Comparison of Asibi and 17D-204

The Asibi strain of YFV was obtained as lyophilized cell culture supernatant from the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) (Galveston, TX). The passage history of Asibi strain virus was as follows from the original isolate: six times in rhesus macaque (*Macaca mulatta*) (Stokes, Bauer, and Hudson 1928), then 3 times in C6/36 (*Aedes albopictus*) cells. 17D-204 virus was reconstituted from a single, unexpired pharmaceutical dose of lyophilized vaccine (YF-Vax<sup>®</sup>, lot number UF795AA, Sanofi Pasteur, USA) using the provided injection diluent. Asibi virus was reconstituted in sterile, deionized water.

##### 2.1.2 Chapter 4: Comparison of 17D Substrain Seeds and Vaccines

A panel of primary and secondary 17D strain seed lots plus some vaccine production lots was assembled, representing examples of the vaccine lineages 17D-204, 17D-213, and 17DD [Table 4.1]. Collection strains ampules were obtained as unpassaged, lyophilized stocks from producers, which were reconstituted in molecular grade water for seed ampules or injection diluent for commercial vaccine preparations, without passage. Passage levels were obtained from open literature referencing the strains, with the exception of the 17D-213 substrain vaccine produced in Nigeria, which was of unknown substrain origin and resolved using basic local alignment search tool (BLAST). Strains are identified using a nomenclature sub-

strain/origin/lot name/year (year of manufacture is included for strains of same or similar origin).

### **2.1.3 Chapter 5: Comparison of French Neurotropic Vaccine Strains**

WT FVV and Asibi strains were obtained from WRCEVA as lyophilized cell culture supernatant. “FNV-IP” was obtained from the Institut Pasteur, (Paris, France). “FNV-Yale” was obtained from the Yale Arbovirus Collection, (New Haven, CT, USA). “FNV-FC” was obtained from the Centers for Disease Control and Prevention (Fort Collins, CO). “FNV-NT” was obtained from what is now called Public Health England, Porton Down (Salisbury, UK) (Fitzgeorge and Bradish 1980). Vaccine strain 17D-204 was obtained from a commercial ampule of YF-Vax<sup>®</sup> lot UH356AA (Sanofi-Pasteur, Swiftwater, PA). The passage history of all four FNV strains from FNV vaccine is unknown but their phenotypic properties have been described (E. Wang et al. 1995). All vaccine strains were passaged once only in Vero cells in this study to produce working stocks using Eagle’s minimal essential media, supplemented with 2% fetal bovine serum (5% CO<sub>2</sub>, 37°C).

### **2.1.4 Chapter 6: Analysis of Subpopulation Selection in a Hamster-Adapted Strain**

The parental Asibi strain was obtained from the WRCEVA (University of Texas Medical Branch (UTMB), Galveston, TX). Asibi was sequenced without passage by next-generation methods by Beck et. al.; reads were reprocessed for this study (A. Beck et al. 2014). Adapted Asibi strain Hamster/P7 was obtained from frozen archived stock as reported by McArthur (McArthur et al. 2003). Briefly, the virus was passaged from a large plaque variant of the parental strain Asibi (Shope strain as in Chapter 3, WRCEVA, UTMB) seven times intraperitoneal (i.p.) in Syrian golden hamsters [Figure 6.1].

### **2.1.5 Chapter 7: Comparison of Wild-Type with Derived Vaccine Strains of Japanese Encephalitis Virus**

The parental JEV strain SA14 was obtained from a stock passaged once in suckling mouse brain, and then once in Vero cells. Two strains of the SA14-14-2 vaccine were analyzed in the study, the first of which was, from the vaccine ampule, passaged once in C6/36 (*Aedes albopictus*) cells, and is called SA14-14-2/P1. The second analyzed vaccine strain was passaged once in Vero cells from SA14-14-2/P1, and is called SA14-14-2/P2.

## **2.2 RT-PCR**

Viral RNA was extracted using the QiAmp<sup>®</sup> viral RNA column isolation kit (Qiagen, Gaithersburg, MD). RT-PCR amplicons were generated using the listed primers and amplification strategy [Table 2.1], which were designed in reference to Genbank accession X03700.1, the earliest entry for 17D (Hahn et al. 1987). Overlapping RT-PCR amplicons were generated from extracted vRNA directly, without intervening passage of either virus. Six amplicons were designed to be at least 2 kilobases (kb) in size, covering the entire viral genome, with 250b overlap. The amplification program was as follows: reverse transcription for 30 minutes at 50°C, initial denaturation for 2 minutes at 94°C, 40 cycles of 10s denaturation at 94°C, anneal for 30s at a primer-dependent temperature, and extension for 2 minutes at 68°C. Final extension was at 68°C for 7 minutes.

## **2.3 Pipeline**

### **2.3.1 Illumina Sequencing, *de novo* Assembly, and Preprocessing**

Amplicons were diluted to molar equivalence, and indexed cDNA libraries generated. Libraries were sequenced on an Illumina Hi-Seq-1000 or 1500 instrument

(the instrument upgraded during the course of the dissertation studies), obtaining paired-end, 50b reads. Barcode sequences and three 3' bases were removed from all reads, retaining those read pairs with average quality scores (Phred scale) above 35. Reads were assembled *de novo* using ABySS v.1.3.4, using the paired-end function and k-mer size of 20-40 (Simpson et al. 2009). Contigs recovered from assembly were screened by BLAST search to detect relative abundance of expected viral templates in the sequencing reaction. Reads were received from the sequencing core lab (UTMB) as a preliminary assembly against a typical reference, compressed in binary alignment/map file (BAM) format.

### 2.3.2 Reference Sequences and Genome Layouts

Reference sequences were selected for alignment based on historical reliability, e.g. prototype strains were used for alignment references in all cases. For chapter 3, the alignment reference was Genbank accession AY640589, a standard reported sequence for Asibi (McElroy et al. 2005). For chapter 4, the alignment reference was Genbank accession X03700, the original sequence of 17D as reported by Hahn and Rice (Hahn et al. 1987). For chapter 5, the alignment reference was Genbank accession YFU21056, the wild-type parental FVV (E. Wang et al. 1995). For chapter 6, the alignment reference was AY640589. For chapter 7, the alignment reference was Genbank accession JEU14163, a sequence for the wild-type JEV prototype strain SA14 (Ni, Burns, et al. 1994).

For calling of nucleotide and codon positions of YFV, the extent of gene segments was constructed based on inference of typical flavivirus cleavage sites in the YFV polyprotein (Rice, Lenches, et al. 1985). For calling of nucleotide and codon positions in JEV, the gene segment extents listed by Song, et. al were used (Song et al. 2012).

### 2.3.3 Preprocessing and Alignment of Illumina Reads for Downstream Analysis

The received alignments were disassembled into constituent, paired-end FASTQ files, using the Picard Tools SamToFastq.jar. For reads originating from RT-PCR amplicons, primer sequences were trimmed from the reads using Trimmomatic v.0.30. Alignment of reads was performed using Bowtie2, in either “-very-sensitive-local” or “-very-sensitive” modes; the latter use (nonlocal alignment) was only performed for Chapter 3 (Asibi/17D comparison) and Chapter 7 (JEV comparison) (Langmead and Salzberg 2012). Alignments were generated in sequence alignment/map file (SAM) format first, then compressed to BAM format and sorted by read nucleotide position using samtools/htslib “sort” command (H. Li et al. 2009). Polymerase chain reaction (PCR) duplicate removal was performed with Picard-tools MarkDuplicates.jar, using an optical pixel distance tolerance of 0 (Broad Institute, Cambridge, MA). Read coverage was then measured using the “coverage” command of bamtools v.2.4.0. Coverage measurements on processed alignments were used to generate a proportion of mean read coverage in reference to the BAM alignment file in the experiment (e.g. within a single chapter) with the lowest read coverage after PCR duplicate removal; the proportion was used to perform a random downsampling operation on alignments to provide a normalized equalization of coverage for all samples considered in a particular experiment. Downsampling was performed with Picard-Tools DownsampleSam.jar, using the aforementioned normalization ratio to determine random probability of read pair retention (Broad Institute, Cambridge, MA). Final coverage was verified. For Chapter 3 (Asibi/17D comparison), downsampling was performed directly on FASTQ files, before initial alignment, by a custom Python script.

### 2.3.4 Generation and Analysis of Consensus Sequences

After alignment, PCR duplicate removal, and downsampling for coverage normalization, consensus sequences were generated from coverage-normalized alignments by calls against a pileup file, which was generated using samtools v.0.1.9 and bcftools v. 0.1.9, using the biallelic consensus caller of bcftools, with base alignment quality filtering disabled for pileup generation. Final sequences were end-padded for end-artifact regions not covered by the sequencing procedure, aligned in reference to the parental sequence in MacVector 14.0.1. Substitutions were classified as silent or coding relative to the open reading frame and local codons of the parental strain used in the respective experiment.

## 2.4 Diversity and Divergence of Sequence Alignments

This section summarizes the use of diversity indices that were used in the studies to estimate both complexity of the viral populations, and their comparative genetic distance (Nishijima et al. 2012; K. Li et al. 2010). Nucleotide frequencies were calculated from a matrix of relative nucleotide counts at each nucleotide position, incorporating the four possible nucleotides with a gap symbol that may have been introduced by the aligner. The count matrix was generated using the “bam2R” parser in the R library deepSNV (Gerstung et al. 2012), and then the function “RF” of the same library was used to generate an R dataframe of relative nucleotide frequencies per nucleotide site. Diversity index and distance measurements are taken directly from these frequencies as generated in the R environment.

### 2.4.1 Shannon’s Entropy

Shannon’s entropy (SE)  $SE$  is calculated for a read alignment at every nucleotide site  $X$ , summarizing across the entire possible nucleotide mutation set  $p$ .



The probability  $P$  is the relative frequency of each possible nucleotide at the position  $i$  in the alignment taken as natural logarithm of the same multiplied by the same frequency, which are summed to give a single value for the position. The index can be summarized in multiple ways across a chosen group or extent of nucleotide positions, usually by descriptive statistics.

$$SE(X)_i = - \sum_{p=\{A,T,C,G,-\}} P[X_{i,p}] \ln(P[X_{i,p}])$$

Likewise, the normalized Shannon's entropy (NSE)  $NSE$  is generated by dividing 1.61, which is the maximum possible  $SE$  value for the five nucleotide set considered at each position. This reduces the range of the value to zero and 1.00 at minimum and maximum, respectively.

$$NSE(X)_i = - \sum_{p=\{A,T,C,G,-\}} P[X_{i,p}] \ln(P[X_{i,p}]) / 1.61$$

#### 2.4.2 Simpson's $1 - D$

Simpson's  $1 - D$  is a more stringent value, as the incorporation of zero values in the numerator removes the single-count of the summed nucleotide from the calculation by collapsing it to zero. Similar to  $NSE$ , the calculation ranges between zero and 1.000 for each nucleotide position considered in a particular alignment, however the natural log transform (as used in  $SE$  and  $NSE$ ) is not used as in  $SE$  or  $NSE$ , rendering the calculation less sensitive to changes in frequency at small values. In this case, the value  $n$  represents the count of each possible nucleotide at the considered position in the alignment, while  $N$  represents the total of those counts for the possible nucleotide set.

$$S_{1-D}(X)_i = 1 - \frac{\sum_{p=\{A,T,C,G,-\}} n(n-1)}{N(N-1)}$$

### 2.4.3 Root-Mean Squared Distance

Root-mean square distance (RMSD)  $RMSD$  is calculated using the set  $p$  of relative nucleotide frequencies at a chosen nucleotide position in the BAM alignment. The value accomodates positive and negative movement in frequency for all nucleotides at the position from a baseline alignment  $X$  to the frequencies of a resulting alignment  $Y$ . The value of  $RMSD$  increases with genetic distance between the two strain alignments considered. Several landmark values are seen in the calculation, for example 0.632 is the maximum travel of the value, indicating a complete swap of one nucleotide for another. Half of the maximum, 0.316, is indicative of a selection from an equally diverse state, or equivalent diversification from a completely fixed state.

$$RMSD(X, Y)_i = \sqrt{\frac{\sum_{p=\{A,T,C,G,-\}} (P[X_{i,p}] - P[Y_{i,p}])^2}{5}}$$

### 2.4.4 Error Rate

The sequencing error rate  $ER$  estimates the bulk of variant populations in the read alignment at a particular nucleotide position, and includes both true variants as well as false-positive, statistical noise single-nucleotide variants (SNVs) introduced by sequencing chemistry or read processing. The value is calculated from the set  $p$  of relative nucleotide frequencies at a chosen nucleotide position. By subtracting the maximum value in  $p$  (the consensus frequency, shown in the inset as  $Max(p)$ ),

the resulting value includes the sum of all frequencies not contained in the consensus nucleotide frequency at the considered nucleotide position. Since *ER* includes both the true variant populations and artifacts, it is useful as a rough estimation of alignment variant distribution before using more rigorous estimates.

$$ER(X)_i = 1 - Max(p)$$

## 2.5 Statistical Methods for Comparison of Sitewise Nucleotide Diversity

In all studies, diversity and distance indices are compared nonparametrically. For comparisons of entropy, the diversity index is summarized along a selected extent of the genome (whole-genome, open reading frame, individual genes, or selected sets of nucleotide positions) and analyzed using measures of central tendency. To compare the estimate of diversity between two strains, a Kruskal-Wallis/Mann-Whitney test was used with an alpha of 0.05. For comparisons of multiple sequences simultaneously, the Kruskal-Wallis H test was used to perform omnibus test, with an alpha of 0.05; pairwise comparison of strains was performed with the Wilcoxon Rank-Sum test with Dunn’s post-hoc correction of p-values.

To measure statistically significant changes in nucleotide frequencies between two read alignments, BAM files were parsed into an R session as nucleotide counts per nucleotide site for all possible values (A, C, G, U, -) and compared using the model of Gerstung and Beerenwinkel (deepSNV) (Gerstung et al. 2012). The model was used with an alpha of 0.05, with a beta-binomial distribution and Bonferroni correction of p-values. The output results from the model indicate nucleotides that have significantly changed in frequency in transition between the two alignments; this may indicate selection, or alternatively may show the presence of SNVs accurately at low

frequency, this depends on the appropriateness of chosen control sample alignment used in the model as a statistical noise floor. For YFV 17D and FNV strain vaccine samples, the WT parental strains Asibi and FVV are appropriate choices for control alignments considering the limited number of consensus mutations differentiating them from derived vaccines.

## 2.6 Resolution and Classification of Single-Nucleotide Variants

SNVs were analyzed from BAM alignments by two methods. A heuristic analysis was performed (retaining variants above one percent cutoff) by conversion of the BAM file to pileup format using the samtools/htslib “mpileup” command, with base alignment quality filtering disabled. VarScan v. 2.3.5 was then used to scan the pileup, retaining variants above one percent. For modeled detection of variants, V-Phaser v.2.0 was used on the BAM alignment directly, with default settings and an alpha level of 0.05. Outputs of the variant callers were simplified using GNU awk 4.1.0, condensing the raw output file(s) into a simplified, delimited table. Variants were classified in reference to the particular consensus sequence generated from the BAM alignment of each sample, using the Perl script SNPdat.pl (Doran and Creevey 2013). Classification tables were parsed into R sessions and merged with frequency data for graph preparation and depiction using ggplot.

### 2.6.1 Resolution and Classification of Wild-Type Revertants

Once generated, the WT or parental sequence was scanned in an R environment against the table of variants generated by either method. On each cycle of the script, if the variant was a match for the nucleotide identity of the parental sequence and position, it was considered a WT revertant. These revertants were classified as coding or silent by virtue of the previous variant detection and classification *in situ*,

which was performed by default on each alignment.

## 2.7 Selection Analyses

$dN/dS$  ratios were generated by a custom Python script, which incorporates the following series of calculations as described by Morelli (Morelli et al. 2013). The calculations use the consensus sequence of a chosen BAM alignment to establish a baseline for mutations detected in the read alignment, and subsequent generation of the final ratio.

A variable number of nonsynonymous or synonymous mutations may be available at a particular site, which depends on codon type and codon position of the possible mutation(s). These values  $n$  and  $s$  are the number of nonsynonymous and synonymous sites accessible in the reference population, which in this case is the consensus sequence of the BAM alignment used in the analysis. And so, the nonsynonymous sites are calculated over the three codon positions  $i$ .

$$n = \sum_{i=1}^3 f_i$$

The available synonymous sites  $s$  for an individual codon are likewise calculated from the reference by subtraction of  $n$ .

$$s = (3 - n)$$

Nonsynonymous sites  $N$  over a particular extent or number of codons in the reference sequence are added.

$$N = \sum_{i=1}^r n_i$$

Synonymous sites  $S$  are likewise calculated from the remainder of sites in the extent of the reference considered.

$$S = (3r - N)$$

For the sample read alignment  $d$  in BAM format, similar values  $N_d$  are calculated for nonsynonymous sites. This analysis requires the scanning of each read aligned to the region of the reference in the extent of codons from which the analysis is constructed.

$$N_d = \sum_{i=1}^r n_{d_i}$$

Likewise, synonymous sites  $S_d$  in the sample to be analyzed are scanned and counted along the extent of the reference used for analysis, resulting in a summation of silent nucleotide change in the sample region.

$$S_d = \sum_{i=1}^r s_i$$

A ratio  $P_N$  is expressed between the possible nonsynonymous substitutions in the sample and those detected in the reference.

$$P_N = \frac{N_d}{N}$$

Likewise, a ratio  $P_N$  is expressed between the possible synonymous substitutions in the reference and those detected in the sample.

$$P_S = \frac{S_d}{S}$$

The ratio  $d_N$  is expressed as a numeric correction of  $P_N$ .

$$d_N = -\frac{3}{4}\ln(1 - \frac{4P_N}{3})$$

The ratio  $D_S$  is expressed as a numeric correction of  $P_S$ .

$$d_S = -\frac{3}{4}\ln(1 - \frac{4P_S}{3})$$

The final ratio  $dN/dS$  is divided out from the previous ratios.

$$dN/dS = \frac{d_N}{d_S}$$

Calculations were performed with a custom Python script, which processed the reads in two stages. First, all reads in the alignment were scanned to generate a data structure containing the counts of nucleotides contained in the reads at each position in the reference sequence, including the reference identity; this structure was saved in .json format and subsequently used as a count index to generate the actual  $dN/dS$  values. Finally, the count index was processed to validate the correct extent of genome segments chosen, and also to summarize the ratio over the extent of the entire genome or discrete genes of the sample. Windowed analysis of  $dN/dS$

was performed by calculating the ratio along successive 20-codon windows; this was performed both to resolve selection effects on local domains of the viral genes and to detect smoothing effects caused by broad sampling of the ratio along long gene segments. Peak detection was performed on successive windows by the R package `quantmod`, with the function “FindPeaks” with default settings.

## 2.8 Phylogenetic Methods

*RMSD* was used to generate distance phylogenies by generating a matrix of pairwise *RMSD* between all strain alignments in an experiment. The matrix of pairwise mean distance was computed over the open reading frames of the strains considered, and used to generate a neighbor-joining distance phylogeny in the R package `ape`, exclusive of any consensus information.

## 2.9 Data Processing Environment and Graphical Methods

All analysis of genomics datasets was performed on an Intel MacBook Pro (late 2009), using OSX version 10.6 (Snow Leopard) or 10.7.5 (Lion). Package management and version handling for libraries and open-source sequencing applications was performed using Homebrew 0.9.5. Graphs were assembled using the R library `ggplot2`, with additional graphical settings provided by the R library `ggthemes`. Tables were constructed in a  $\text{\LaTeX}$  environment with the `booktabs` package.



**Table 2.1:** Primers for RT-PCR amplifications.

Primer	Sequence, 5'→3'	Oligo Length(b)	Amplicon Length(b)	Annealing Temperature, °C
17D-1-2005F	AGTAAATCCTGTGTGTGCTAATTGAGGTG	27	2005	58.0
17D-1-2005R	GATTGCCGCTGTAGATCATCAG	23		
17D-1749-3725F	CTGGGGCAATGAGGTTACAAA	22	1976	58.0
17D-1749-3725R	TTGAAAAGGCAGCAATCAACGC	22		
17D-3498-5506F	GGTTACAGCTGGAGAAATACATGC	25	2008	59.0 <sup>a</sup>
17D-3498-5506R	ATTGCCCTAGCTCTGTGCGCT	22		
17D-5249-7278F	GCCCCACGAGGTTGTTCTTTCT	24	2029	58.0
17D-5249-7278R	TGCTGGCTTTTGATTCAGGTA	22		
17D-6900-9100F	TGCTGGAGAAAACCAAGAGGA	22	2200	58.0
17D-6900-9100R	CTCAAACCTCAAGATACCGCGT	22		
17D-8749-10862F	TAGGAAGATCATGAAAAGTTGTCAACAGG	28	2113	58.0
17D-8749-10862R	AGTGGTTTGTGTGTTGTTCATCCAA	24		

<sup>a</sup> Temperature was used only for chapter with Asibi/17D comparison. All others use 58 degrees C.

## Chapter 3

### Comparison of the Live-Attenuated Yellow Fever Vaccine 17D-204 Strain to Its Virulent Parental Strain Asibi By Deep Sequencing<sup>1</sup>

#### 3.1 Abstract

The first comparison of a live RNA virus vaccine to its WT parental strain by deep sequencing is presented, using as a model the YFV live vaccine strain 17D-204 and its WT parental strain, Asibi. A commercial lot of 17D-204 vaccine and low passage collection isolate of the Asibi strain were obtained and subjected to sequencing analysis by massively parallel methods to compare population variability along multiple scales of the viral genomes. A modeled exploration of small frequency variants was performed to reconstruct plausible regions of mutational plasticity. Overt quasispecies diversity was observed to be a feature of the parental strain, whereas the live vaccine strain by contrast lacks diversity by multiple independent measurements. A lack of putative attenuating mutations in the Asibi population relative to that of 17D-204 was observed, demonstrating that the vaccine strain was derived by discrete mutation of Asibi and not by selection of genomes in the WT population. Analyses such as these of attenuated viruses improve our understanding of the molecular basis of vaccine attenuation and provide critical information on the stability of live

---

<sup>1</sup>This is a pre-copyedited, author-produced PDF accepted for publication in *The Journal of Infectious Diseases*. The version of record, (Andrew Beck et al., Comparison of the Live Attenuated Yellow Fever Vaccine 17D-204 Strain to Its Virulent Parental Strain Asibi by Deep Sequencing. *The Journal of Infectious Diseases*(2014) 209(3):334-344) is available online at <http://jid.oxfordjournals.org/content/209/3/334>. (A. Beck et al. 2014)

vaccines and the risk of reversion to virulence, and indicate that relative population structure is a plausible correlate of attenuation for live viral vaccines.

## 3.2 Introduction

All of the live attenuated vaccines in use today were derived empirically, and our understanding of molecular basis of attenuation is often rudimentary. Deep sequencing offers the opportunity to investigate population structures of live attenuated vaccine strains, contributions of these features to attenuation and stability, and the potential of reversion to virulence (Neverov and Chumakov 2010; Kennedy and Poland 2011; Luciani, Bull, and A. R. Lloyd 2012). The paradigm for RNA viruses is that they exist as highly diverse “quasispecies” populations, and that measurable features of viral diversity are plausible correlates to virulence and pathogenicity (Vignuzzi et al. 2006; J. J. Holland, De La Torre, and Steinhauer 1992). In this context, the YFV vaccine was investigated as a model. To date this technology has been applied to vaccines for DNA viruses, but not to vaccines for RNA viruses (Szpara et al. 2011). The live attenuated YFV vaccine strain 17D was derived by Max Theiler and coworkers in 1937, and is a critical tool for control of human YFV infection (Theiler and H. H. Smith 1937). Despite the production of over 550 million doses of vaccine in the last 70 years, our understanding of how the vaccine is attenuated or how the protective immune response is elicited are very limited. Incomplete deep sequencing of the 17D-204 and other live RNA virus vaccines has been performed at low coverage, without achieving full assembly of the viral open reading frame (Victoria et al. 2010). The presented results offer the first comprehensive coverage depth ( $>1000$  mean reads/nucleotide position aligned), permitting direct comparison of low-frequency variants between the rare, archived parental strain and commercial vaccine derivative.

The vaccine was derived from the WT strain Asibi, which was isolated in 1927 from the blood of a male Ghanaian patient of the same name. This parental virus was empirically attenuated by serial passage 18 times in mouse embryo tissue, 58 times in minced whole chick embryo tissue, and finally 128 times in minced whole chick embryo without nervous tissue (W. Lloyd, Theiler, and Ricci 1936; Theiler and H. H. Smith 1937). This “17D” strain was selected from a panel of other candidates for its desirable attenuated properties, which include loss of both viscerotropism and neurotropism in mammalian models (Monath, M. Gershman, et al. 2013). The 17D-204 vaccine substrain was derived from the original 17D strain at passage 204, and is now presently manufactured in the United States, France, Senegal, Russia, and China (World Health Organization 2010, p. 278). Standardization of 17D-derived vaccine is achieved by a seed-lot system, ensuring that all production lots are derived from qualified seeds by one passage in embryonated chicken eggs (World Health Organization 2010, p. 251). By consensus, comparison of Asibi and 17D-204 genomes identified 67 nucleotide differences, encoding for 31 amino acid substitutions (Hahn et al. 1987; Dupuy et al. 1989). Reasonably, it is hypothesized that the historical dataset of mutations differentiating Asibi and 17D-204 represent determinants of mammalian pathogenicity (E. Lee and Lobigs 2008). However to date, specific amino acids contributing to the attenuation of 17D have not been elucidated. The presented studies utilize a bioinformatics approach to compare nucleotide diversity of 17D-204 and Asibi, under a hypothesis that viral subpopulations distinguish the WT and live-attenuated strains.

### 3.3 Results

#### 3.3.1 Assembly and Inspection of Viral Genomes

Nucleotide counts of five randomly downsampled Asibi strain alignments were averaged and floor-rounded along all sites, and the median coverage of this alignment was 5531 (mean=5562.2, standard deviation (s.d.)=2054.1). Median coverage depth for the 17D-204 alignment was 5144 (mean=5682.1, s.d.=2434.6).

#### 3.3.2 Consensus Sequences

Consensus sequences generated by plurality from *de novo* assemblies were compared to reference Genbank sequences of Asibi and 17D-204 vaccines, inspecting the viral polyprotein and UTRs. Following comparison of our *de novo* assembled Asibi strain consensus to that of the published Asibi sequence (Genbank accession AY640589), recovery of nucleotide positions 11 through 10843 was observed, with the presence of one silent consensus mutation at nucleotide position A2704G. An independently produced *de novo* consensus sequence of 17D-204 was generated from the same commercial lot of vaccine stock as was used in this study and was found to be identical to the consensus sequence obtained here (Genbank accession JX503529.1). Recovery of nucleotide positions 14 through 10850 for the 17D-204 consensus was observed. Relative to a 17D-204 sequence published by Xie (H. Xie, Cass, and A. D. Barrett 1998) (Genbank accession AF052437.1), two nonsynonymous mutations were observed at positions U7496C and C7497U, both independently coding for the same amino acid substitution NS4B S67P.

#### 3.3.3 Heuristic Identification of Variants

For the Asibi strain alignment, 55 variant sites meeting the inclusion criteria were observed, distributed over the entire genome [Table 3.1 and Appendix Table 1].

Of these, 9 variants encode amino acid substitutions. For the 17D-204 alignment, 32 variant sites meeting the inclusion criteria were observed; of these, 5 encode amino acid substitutions [Table 3.2 and Appendix Table 1].

### 3.3.4 Comparison of Quasispecies Diversity

Estimates of diversity were compared at multiple scales [Figure 3.1]. For SE, median values for Asibi and 17D-204 strains were 0.0179 and 0.0170, respectively, and strains were significantly different ( $U=61791383$ ,  $p=1.60 \times 10^{-9}$ ) [Figure 3.2]. For Simpson's  $1 - D$ , median values for Asibi and 17D-204 strains were 0.0047 and 0.0044, respectively, and strains were significantly different ( $U=61095895$ ,  $p=7.81 \times 10^{-7}$ ). Using only the vaccine mutation sites described by Hahn (Hahn et al. 1987), median entropy values for Asibi and 17D-204 strains were 0.0194 and 0.0214, respectively, and strains were not significantly different ( $U=1728$ ,  $p=.127$ ). Median Simpson's  $1 - D$  for the mutation sites was 0.005 and 0.006 respectively, and strains were not significantly different ( $U=2611$ ,  $p=0.62$ ) [Figure 3.3]. Significant strain differences were observed for local comparison of 3' untranslated regions, and for viral genes C, prM/M, E, NS1, NS2A, NS2B, NS3, NS4B, and NS5 [Table 3.3]. Quantile-quantile comparisons demonstrate inequality of distribution for error rates, indicating a greater relative presence of high outlying variant frequencies for the Asibi strain. This pattern was recovered both along full genomes and individual genes [Figure 3.4,3.5].

### 3.3.5 Intersection of Population Structure For Asibi and 17D-204

Eight of the previously observed mutant positions differentiating Asibi and 17D-204 strains share considerable variant identity between parent and vaccine, but only in the Asibi alignment. Variability at these shared sites is greater for the Asibi

strain alignment, while broadly these sites are fixed to a homogenous identity in the 17D-204 alignment [Figure 3.6]. Of these eight highly overlapping variants identified by heuristic criteria, the mutation A5153G, present in both viruses, codes for the only amino acid substitution, Ile195Val in the NS3 protein (1.39% Val, 98.6% Ile in 17D-204), with approximate reverse frequency in the Asibi strain population (98.0% Val, 1.89% Ile). Variant populations recovered from the Asibi strain intersect with 17D-204 to a greater extent by percentage than vice versa [Figure 3.7]. Considering percent variant frequency at 74 (67 fixed and 7 clonal) sites differentiating Asibi and 17D-204 (Hahn et al. 1987), Asibi identity intersects with the consensus 17D-204 identity at 2.89 percent of total coverage in the parental read set (n=11196), while conversely 17D-204 identity intersects with the Asibi consensus at 0.42 percent of total coverage in the vaccine read set (n=1595) [Figure 3.7].

### 3.3.6 Recovery of Small Variants By A Paired-Test Model

A total of 157 mutations were recovered by this criterion, after excluding insertions, deletions, and stop codons. 109 of the mutations were noncoding, of these 53 represent four-fold degenerate sites, while 48 represent amino acid substitutions [Figure 3.8]. Of nonsynonymous substitutions recovered, 2 were observed in prM/M, 17 in E, 4 in NS1, 6 in NS2A, 3 in NS2B, 5 in NS3, 1 in NS4A, 5 in NS4B, and 4 in NS5. Significantly represented frequency changes ( $p < 0.5$ ) were clustered in specific regions of the viral genome [Figure 3.9,3.10]. Mutations recovered from the paired test model were compared both to the consensus (50 percent plurality) sequences of Asibi and 17D-204 generated *de novo*, and also to lists of differentiating mutations observed previously. The paired-test model recovered all of the 63 of 64 mutations that differentiate the *de novo* assembled sequences by consensus, and 15 (11 plus 4 clonal sites) less than the 17D-204 mutation set described by Hahn, or 79.8 percent recovery (Hahn et al. 1987).

### 3.4 Discussion

#### 3.4.1 The Wild Type Parental Strain Asibi Is A Diverse Quasispecies

The WT Asibi virus was found to consist of diverse quasispecies, as would be expected of a RNA virus. By contrast, the 17D-204 vaccine strain population was homogeneous; very limited evidence was found for the existence of WT nucleotide identity within subpopulations of the vaccine strain. Production of 17D-204 occurs with some variability due to proprietary methods of manufacture, therefore it is important to ascertain the population identity of 17D-204 at the point of use, and later any effects of these mutations on the viral phenotype in experimental systems. Because analysis of viral quasispecies structure is expected to inform predictive models of pathogenicity and attenuation (Luciani, Bull, and A. R. Lloyd 2012), reconstruction of vaccine populations is a matter of public health importance. The standardized and attenuated properties of 17D-204 provide an excellent framework to construct such a comparative model.

Absolute quasispecies diversity was estimated by diversity indices; Asibi and 17D-204 differed by nonparametric comparison, a pattern that was recapitulated at the level of multiple discrete genes. Quantile-quantile comparisons of error rates demonstrated a greater presence in the Asibi strain of variable sites at the high ranges of the total error distribution [Figure 3.4,3.5]. As expected, these methods demonstrate a restricted pattern of variability for both genomes, but especially for the 17D-204 vaccine. This admits the possibility that a small number of large entropy values are present in an otherwise predictably distributed set, an expectation frequently utilized in construction of phylogenetic models of sequence divergence that accommodate among-site differences in mutation rate (Z. Yang 1993). This property of the sequencing dataset is consistent with a hypothesis that site variants with greatest mutation rates are localized to a small proportion of tolerated sites.



### 3.4.2 Variability in Transition to The 17D Population is Localized

In a paired-samples test, significant changes in variant frequencies were localized to several regions of the genome hypothesized to influence virulence and pathogenicity [Figure 3.9,3.10]. Clustering of variants by sliding window average reconstructed clusters of mutations that differentiate Asibi and 17D-204 by consensus, of which the highest peaks (indicating higher density of significant mutations) are located at sites in the viral genes E, NS2A, NS3, NS4B, and NS5. This pattern is rationally expected. Briefly of these, the flavivirus E gene encodes the major external structural element of the virus, of which domain III (nucleotides 1851-2168, containing the largest reconstructed peak) contains immunodominant epitopes and influences adhesion to the extracellular matrix (E. Lee and Lobigs 2008; Sil et al. 1992). The NS2A gene has been implicated as a factor in the assembly of flaviviruses; mutational analysis of Kunjin virus (KUNV) showed influence of the protein over particle egress and intracellular membrane ultrastructure development (Leung et al. 2008). NS3 encodes serine protease, helicase and nucleoside triphosphate (NTP) triphosphatase activities (Bera, Kuhn, and J. L. Smith 2007; Luo et al. 2008). The NS4B gene of flaviviruses (peak at 7171) has been implicated as an IFN antagonist *in vitro* following infection by DENV, WNV, and YFV (Muñoz-Jordán et al. 2005). NS5 encodes methyltransferase and RNA-dependent RNA polymerase functions (Zhou et al. 2007; G. Zou et al. 2011).

### 3.4.3 Wild-Type Sequence Content is Not Contained in the Vaccine Population

It is especially significant that variants in the vaccine overlap the virulent Asibi genotype at low or undetectable frequencies. The WT strain contains a subpopulation with representation of vaccine strain sequence identity. The relationship is unidirectional, meaning that the vaccine does not contain low populations of WT

variants at the coverage depths achieved. It is not possible to assess linkage of variants beyond read pair distance. However, even independent recovery of parental identity variants was not observed in the vaccine alignment, discounting even a low probability that multiple virulence-determinant sites would be linked to a single genome. The apparent fixation of multiple sites in the vaccine strain indicates an influence of selection pressures exerted at discrete, subgenomic scales. This fixation of the population may in part explain the excellent safety record of the vaccine, although mutation of the virus in vaccinees cannot be excluded. However, these data are supportive of the current opinion that serious adverse events giving rise to a WT viscerotropic disease phenotype are a consequence of host effects rather than of specific virulent properties of the vaccine (Barban et al. 2007). This also supports an interpretation that the genomic stability of YF-Vax<sup>®</sup> at the point of use is stably maintained by the seed-lot system.

#### **3.4.4 Variable YFV Sequence Content is Shared by Other Models of Attenuation**

Intersection of the recovered low-frequency SNP populations with other models of YFV attenuation was observed [Figure 3.11,3.12]. For the Asibi alignment, 12 sites share variant identity with two other models of YFV attenuation. The first of these models, the FVV, was isolated concurrently with Asibi, and then passaged 128 times in mouse brain to yield the FNV vaccine strain, which was distributed in francophone Africa until high rates of post-vaccinal encephalitis in recipients below 14 years of age prompted discontinuation of FNV in 1982. Additionally, passage of the Asibi strain in HeLa cells was performed to produce a virus with similar phenotypic characteristics to 17D, namely reduced virulence in both mice and cynomolgous macaques, and loss of vector competence in *Aedes aegypti* (A. D. Barrett et al. 1990; Dunster, H. Wang, et al. 1999). Sequences of these attenuation models were

compared to the reconstructed variants, recovering some similarities [Tables 3.1,3.2, Appendix 1 1]. The Asibi E gene variant nucleotide A1054C (E protein residue E27H, 9.74% H) was previously observed to arise in the attenuated HeLa p6 strain of YFV (Dunster, H. Wang, et al. 1999), plus two additional silent variants G1000A and A3274G. The Asibi E protein variant C2193U (E protein residue A407V, 13.11% V) was extinguished by consensus during attenuating passage of FVV(V) to produce FNV(A) (E. Wang et al. 1995). Four silent variants intersect with the FNV strain. For the 17D-204 alignment, one silent variant C10367U (1.58%U) was observed in the 3'UTR, intersecting the consensus identities of both FNV(U) and Asibi(U).

### 3.4.5 Attenuation of Asibi Was Likely By Discrete Selection

Of particular concern is whether the empirically selected vaccine genotype is the result of either discrete selection at specific sites or macroscopic selection of a population of parental virus. Selection pressures governing the virulence of Asibi are largely unknown. Such pressures are likely to be diverse in nature, involving tissue tropisms, transmission dynamics, host immune responses, and others. It is possible that some variants represent tolerable plasticity in the vaccine genotype, comprising a multivariate sequence space that is flexibly occupied by the virus (Vignuzzi et al. 2006). Little is known about the adaptive context of quasispecies distributions in arthropod-borne viruses, however selection for phenotypes that productively infect both arthropod and vertebrate hosts almost certainly propels this feature. Arthropod-borne viruses encounter population stresses during WT transmission cycles, which are relieved by single-tissue, *in vitro* passage at the cost of multihost fitness (Ebel et al. 2011; Coffey et al. 2008b). The attenuating passage of 17D outside of the normative mosquito-primate host cycle plausibly represents such an example.

The virulence and pathogenicity-determining features of specific viral quasispecies have largely not been addressed; however some evidence has been offered to suggest that both replication-dependent and replication-independent mechanisms exist and both may be parsed *in vivo* (Lancaster and Pfeiffer 2011). Since original derivation of the 17D vaccine was performed outside of naturally infected systems, it is reasonable to expect not only divergent adaptation to these tissues, but also drift from genotypes under selection in the natural mosquito-primate transmission cycle (Coffey and Vignuzzi 2011; Coffey et al. 2008b). Significantly, we have recovered common sequence identity occupied by multiple contexts of YFV attenuation. The distribution and nature of these effects along the YFV genome are significant targets for further investigation, with great potential to inform the study of flaviviral pathogenesis and rational vaccine design.

**Table 3.1:** Coding Variants observed in the Asibi strain alignment, using a 1.00 percent frequency cutoff heuristic.

Position	Base	Variant Base	Codon	Local Codon	Gene	Reference Amino Acid	Variant Amino Acid	Coverage Depth	Variant Reads	% Variant
620	U	C	168	47	prM	Y	H	8174	129	1.59
1054	A	C <sup>a</sup>	312	27	E	Q	H	6743	656	9.74
2193	C	U <sup>b</sup>	692	407	E	A	V	2640	346	13.11
2349	C	U	744	459	E	A	V	2985	121	4.07
2481	A	G	788	10	NS1	K	R	3356	420	12.53
3171	A	C	1018	240	NS1	E	A	2440	56	2.32
4416	G	A	1433	79	NS2B	S	N	6744	195	1.86
5153	A	G <sup>c</sup>	1679	195	NS3	I	V	6042	103	1.7
9716	G	A	3200	694	NS5	D	N	3918	118	3.02

<sup>a</sup> Intersection of sequence with the HeLa/P6 strain.

<sup>b</sup> Intersection of sequence with the French neurotropic vaccine.

<sup>c</sup> Intersection of sequence with the 17D-204 strain.

**Table 3.2:** Coding variants observed in the 17D-204 strain alignment, using a 1.00 percent frequency cutoff heuristic.

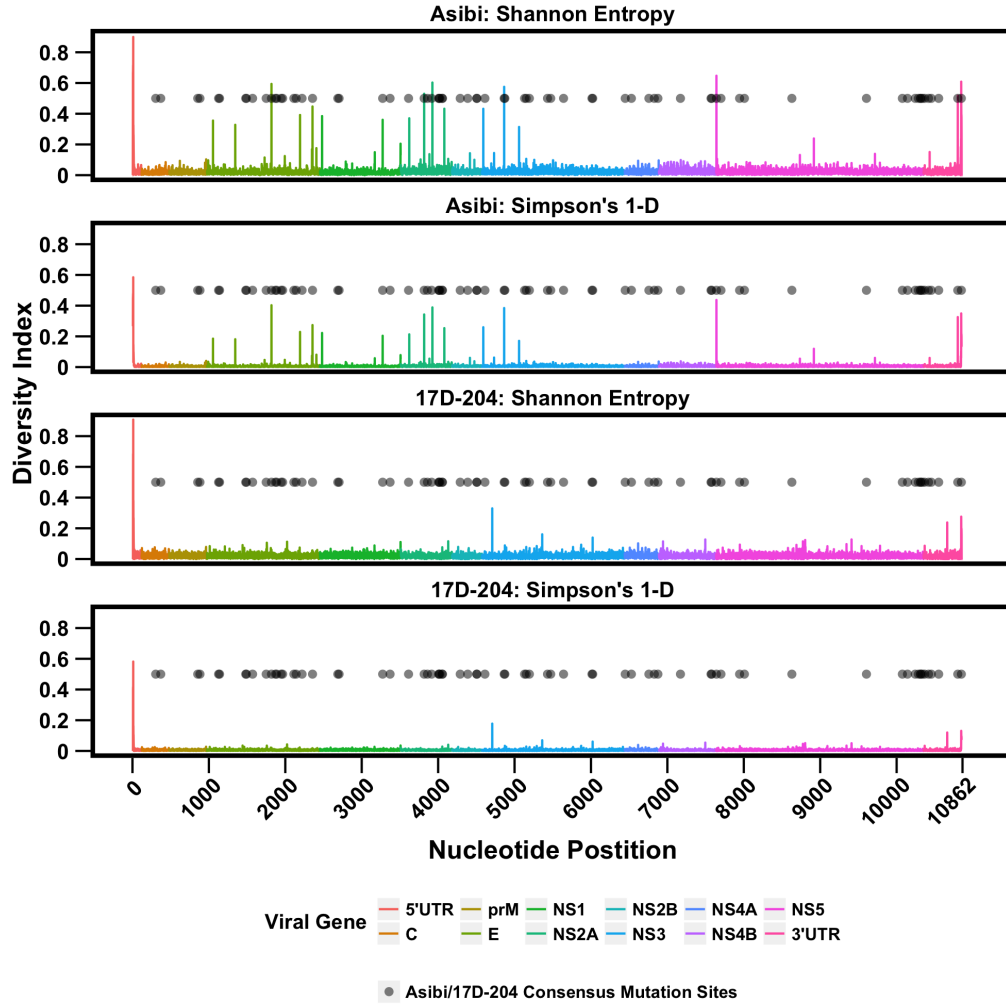
Position	Base	Variant Base	Codon	Local Codon	Gene	Reference Amino Acid	Variant Amino Acid	Coverage Depth	Variant Reads	% Variant
263	U	C	49	49	C	F	L	4591	46	1
3027	A	G	970	192	NS1	K	R	5476	60	1.1
5153	G	A <sup>a</sup>	1679	195	NS3	V	I	5057	72	1.42
5765	U	C	1883	399	NS3	Y	H	3037	31	1.01
9093	A	G	2992	486	NS5	E	G	2969	40	1.35

<sup>a</sup> Intersection of sequence with the Asibi strain.

**Table 3.3:** Diversity index comparison for individual genes between Asibi and 17D-204, Using the Mann-Whitney U Test.

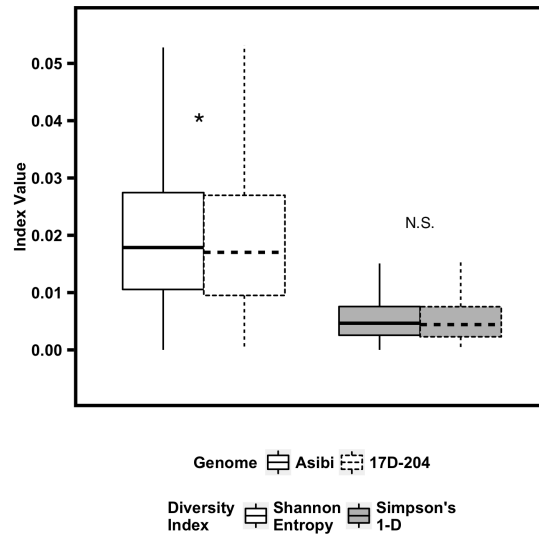
Gene	Length(b)	Shannon's Entropy				Simpson's 1-D			
		Median		Median		Median		Median	
		Asibi	17D-204	Mann-Whitney U	p	Asibi	17D-204	Mann-Whitney U	p
5'UTR	119	0.020	0.019	7058	9.7E-01	0.006	0.005	6615	1.0E+00
C	363	0.014	0.016	60022	3.8E-02*	0.003	0.004	59513	2.4E-02*
prM	492	0.014	0.016	111041	2.5E-02*	0.004	0.004	109925	1.3E-02*
E	1479	0.016	0.017	1039852	2.0E-02*	0.004	0.004	1031152	7.1E-03*
NS1	1056	0.014	0.019	436915	2.2E-16*	0.004	0.005	440043	2.2E-16*
NS2A	672	0.022	0.017	276584	9.4E-13*	0.006	0.004	273610	1.8E-11*
NS2B	390	0.018	0.014	89943	1.0E-05*	0.005	0.003	88781	5.2E-05*
NS3	1869	0.019	0.017	1927976	3.8E-08*	0.005	0.004	1899208	3.7E-06*
NS4A	447	0.020	0.020	99266	8.7E-01	0.005	0.005	98106	6.4E-01
NS4B	750	0.022	0.017	335441	1.0E-10*	0.006	0.004	333309	5.4E-10*
NS5	2715	0.020	0.017	4038540	9.9E-10*	0.005	0.005	4011186	1.7E-08*
3'UTR	510	0.018	0.015	150138	3.3E-05*	0.005	0.004	142853.5	3.3E-05*

\* Statistically significant at p<0.05.

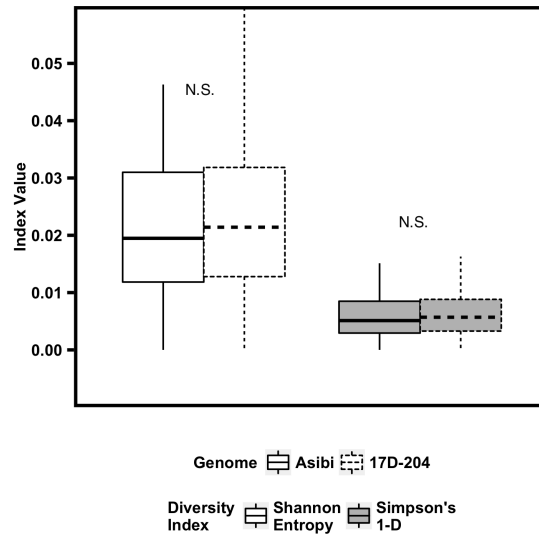


**Figure 3.1:** Diversity indices for each nucleotide position in Asibi and 17D-204, including both Shannon's entropy and Simpsons  $1 - D$ . For both diversity indices used, high diversity values are more frequently observed in the WT strain, and high values are observed at a greater number of nucleotide positions.

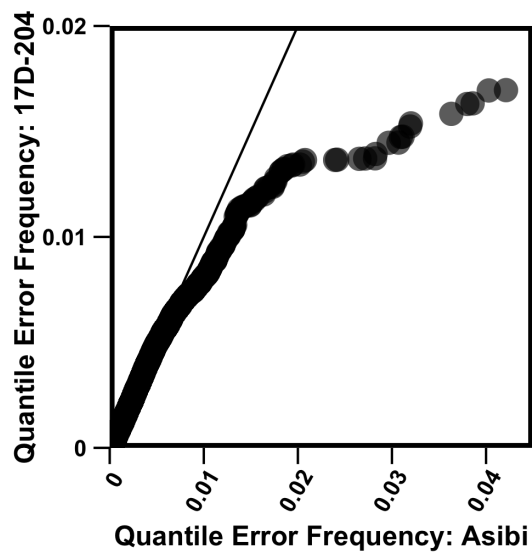




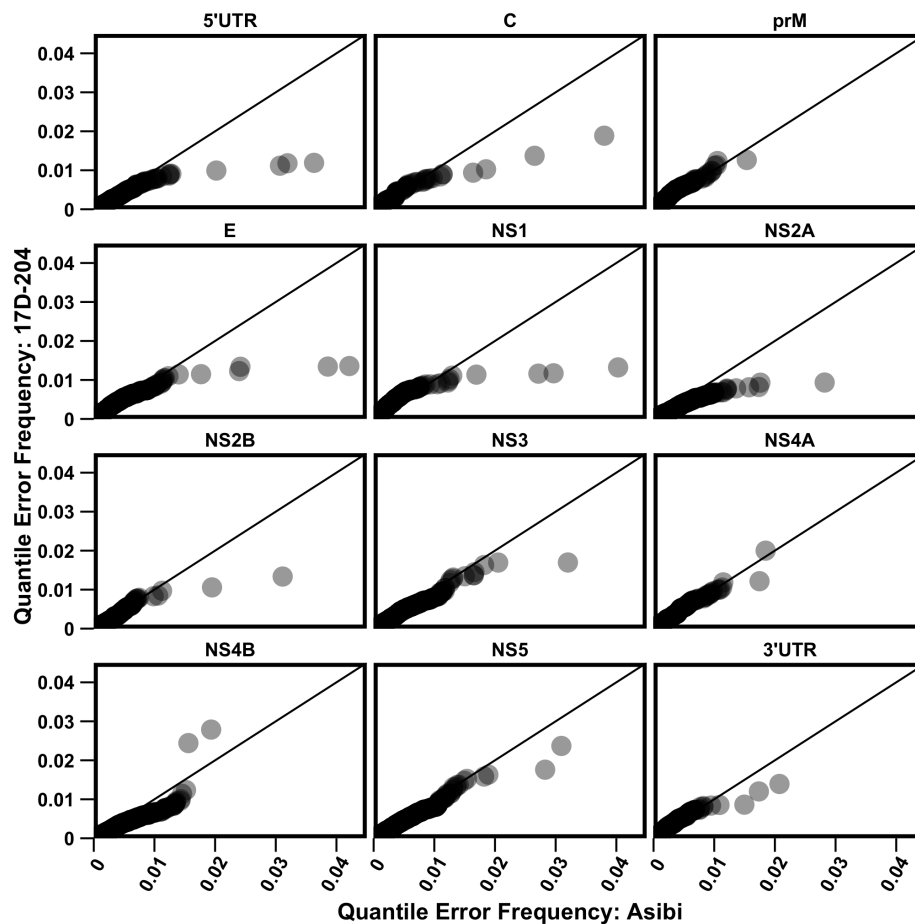
**Figure 3.2:** Comparison of diversity indices for yellow fever virus Asibi and 17D-204 strains, at multiple scales. All statistical comparisons were performed using the Mann–Whitney U test; 2-tailed P values of  $<.05$  were considered statistically significant. Downsampled Asibi read sets were averaged and floor-rounded by counts. Box plots show diversity indices computed for every nucleotide position along the entire genomes for Asibi and 17D-204 strains, and compared. All box plots depict the median, first quartile, third quartile, with whiskers showing  $\pm 1.5$  times the interquartile distance. \* $p < .05$ . Abbreviation: NS, not significant.



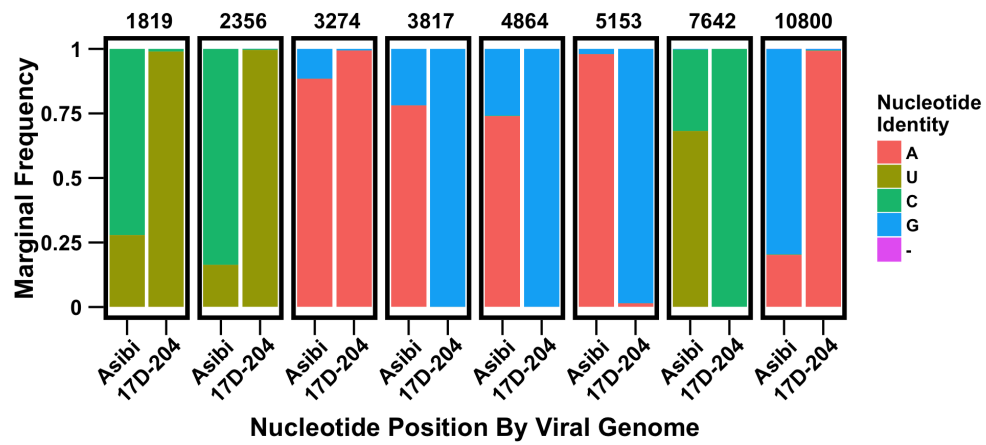
**Figure 3.3:** Comparison of diversity indices for yellow fever virus Asibi and 17D-204 strains, at multiple scales. All statistical comparisons were performed using the Mann–Whitney U test; 2-tailed P values of  $<.05$  were considered statistically significant. Downsampled Asibi read sets were averaged and floor-rounded by counts. Box plots show comparison of diversity indices for the vaccine mutations sites observed by Hahn et al. (Hahn et al. 1987). All box plots depict the median, first quartile, third quartile, with whiskers showing  $\pm 1.5$  times the interquartile distance. \* $p < .05$ . Abbreviation: NS, not significant.



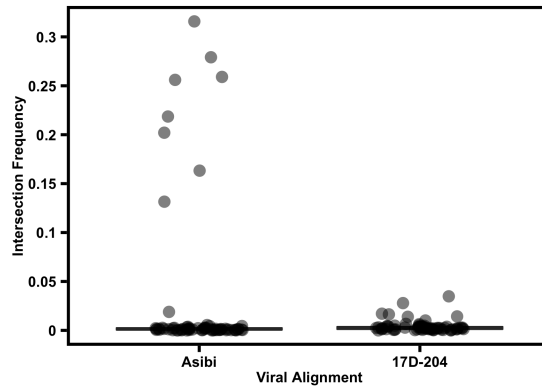
**Figure 3.4:** Quantile-quantile comparisons of error rate along all nucleotide positions recovered for Asibi and 17D-204 strains. Downsampled Asibi read sets were averaged and floor-rounded by counts. Error rates were calculated as frequency of base calls at each site not corresponding to the consensus base call of the alignment. Curve shapes show that a greater number of high outlying errors is present in the Asibi strain alignment, compared with that of 17D-204. This pattern is evident for both the entire genome lengths and at the scale of specific genes. Abbreviation: UTR, untranslated region.



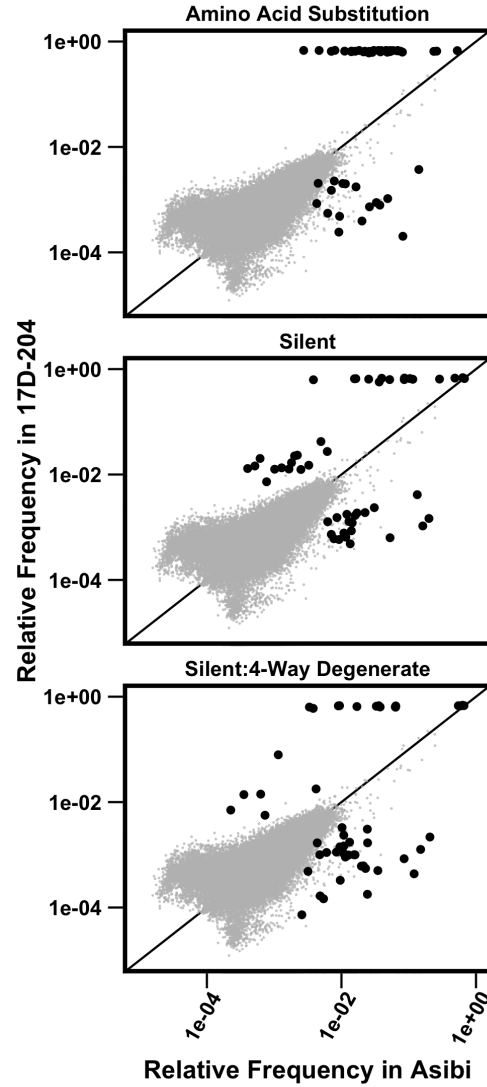
**Figure 3.5:** Quantile-quantile comparisons of error rate along all nucleotide positions recovered for Asibi and 17D-204 strains. Graphic shows each gene segment considered separately. Downsampled Asibi read sets were averaged and floor-rounded by counts. Error rates were calculated as frequency of base calls at each site not corresponding to the consensus base call of the alignment. Curve shapes show that a greater number of high outlying errors is present in the Asibi strain alignment, compared with that of 17D-204. This pattern is evident for both the entire genome lengths and at the scale of specific genes. Abbreviation: UTR, untranslated region.



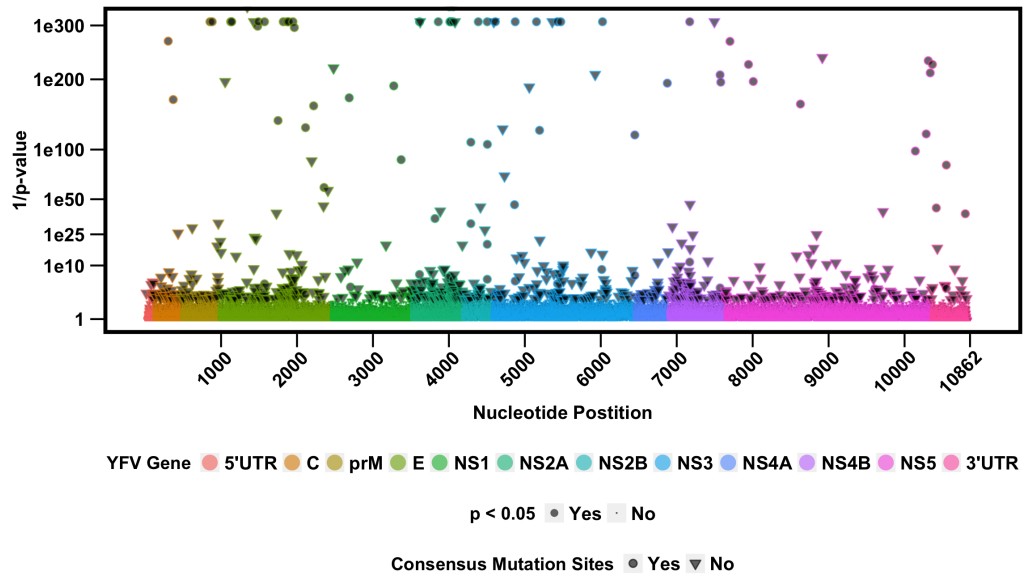
**Figure 3.6:** Intersection of variant identity between Asibi and 17D-204 strain quasispecies is unidirectional. Bar plot shows sites of overt quasispecies identity at sites that are fixed in the 17D-204 strain. By relative frequency, Asibi nucleotide identity intersects that of 17D-204 to a greater extent than the converse. Asibi frequencies are averaged for downsampled read sets.



**Figure 3.7:** Intersection of variant identity between Asibi and 17D-204 strain quasispecies is unidirectional. Graphic shows a box plot summary of intersection frequency for Asibi and 17D-204 strains, overlaid with scatterpoints, depicting the frequency of sites bearing the quasispecies identity of the opposing viral read set. By relative frequency, Asibi nucleotide identity intersects that of 17D-204 to a greater extent than the converse. Asibi frequencies are averaged for downsampled read sets.

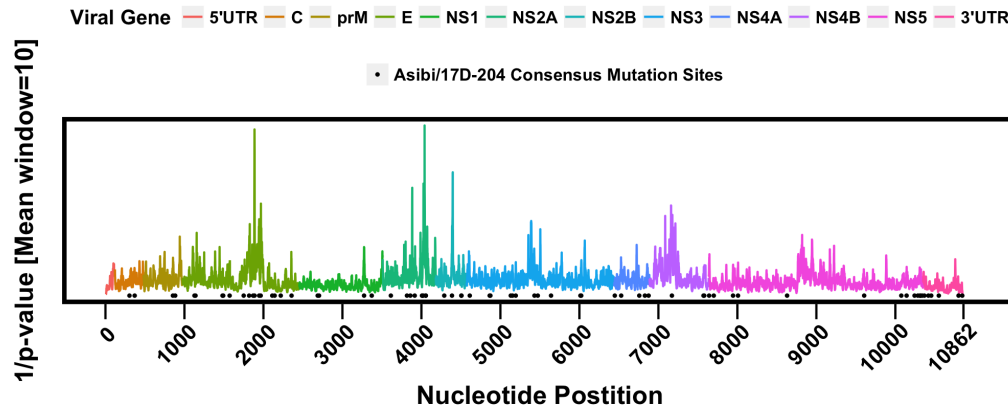


**Figure 3.8:** Scatterplots showing change in nucleotide frequency for paired samples comparison. Plot depicts the set of mutations recovered from a binomial model, using the Asibi strain alignment as a control, and 17D-204 as a test case. Filled polygons represent mutants (152) that were recovered concordantly from all downsampled Asibi strain read sets, at  $p < .05$ . Mutations are classified, relative to the Asibi (control) strain genome, as coding for amino acid substitutions, as silent, and as silent from four-way degenerate sites. The pattern of significant frequency changes shows fixation in the 17D-204 genome relative to the parental strain Asibi.

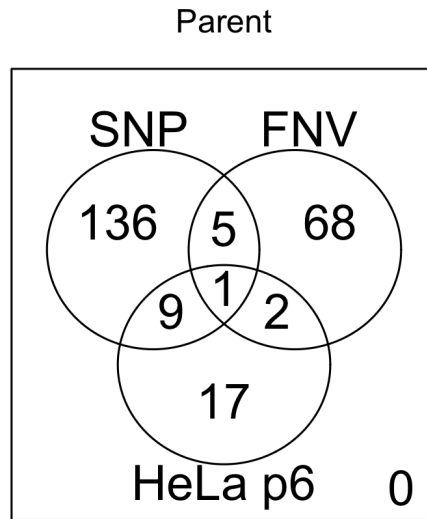


**Figure 3.9:** Recovery of local plasticity in the YFV genome using the paired= $t$ -test model. Graphic shows a Manhattan plot depicting all possible variant frequency changes in the modeled transition from YFV Asibi strain to 17D-204. Polygons represent significant frequency changes, expressed as  $\log(1/p)$ , following Bonferroni correction.

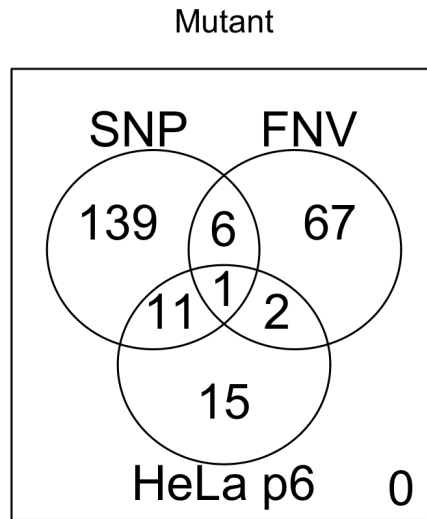




**Figure 3.10:** Recovery of local plasticity in the YFV genome using the paired= test model, as a sliding-window summary. Graphic shows a line plot of a sliding-window mean of the most significant variant recovered from each position in the paired test model. The graph permits visualization of mutationally diverse regions of the YFV genome under the selected model, in comparison to the expected mutation set (black points).



**Figure 3.11:** Venn diagram showing overlap of significantly changed nucleotide frequencies with prototype YFV strains, showing intersection with parental strains only. A paired-test model recovered patterns of local plasticity in the YFV genome. Graphic shows intersection between recovered variant populations with other models of YFV attenuation; this is separately performed for consensus sequences of *parent* (this figure) and derivative strains [Figure 3.12] for HeLa p6 virus and FNV (E. Wang et al. 1995; Dunster, H. Wang, et al. 1999).



**Figure 3.12:** Venn diagram showing overlap of significantly changed nucleotide frequencies with prototype YFV parental strains. A paired-test model recovered patterns of local plasticity in the YFV genome. Graphic shows intersection between recovered variant populations with other models of YFV attenuation; this is separately performed for consensus sequences of parent [Figure 3.11] and *derived* (this figure) strains for HeLa p6 virus and FNV (E. Wang et al. 1995; Dunster, H. Wang, et al. 1999).

## Chapter 4

### Traversal of Yellow Fever Vaccine Strain 17D Seed Lineages by Deep Sequencing Reveals Limited and Stable Patterns of Subpopulation Diversity

#### 4.1 Abstract

Attenuated phenotypic properties of the 17D vaccine lineage are well described and standardized using a seed-lot system developed in 1945; however, the discrete mechanisms giving rise to these features are poorly understood. Quasispecies theory is a useful paradigm by which to understand the influence of viral diversity upon vaccine attenuation by considering evidence that low diversity is a characteristic of attenuated viruses. To investigate these patterns for the YFV vaccine, an archived panel of 17D derivative seed and vaccine lots from different producers was assembled, representing the 17D substrains 17D-204, 17D-213, and 17DD. Massively parallel sequencing analysis was performed to assess the stability and homogeneity of subpopulation identity along divergent branches across the entire 17D lineage. Within each substrain, concordance of diversity and variant structure was observed; these patterns were recapitulated between substrains, but to a lesser degree. The results indicate that, despite production of multiple substrains of live-attenuated 17D vaccine, there is considerable homogeneity of the vaccine across multiple subculture passage histories, and that the seed-lot system is very effective at controlling the apparent diversity of the vaccine. Overall, the results suggest that NGS provides a technique to evaluate and resolve patterns of population structure in live attenuated vaccine preparations.

## 4.2 Introduction

All YFV vaccines in current use are derived from the 17D lineage, developed by the Rockefeller Foundation in 1937 (Theiler and H. H. Smith 1937; Theiler 1951). Following derivation of the 17D vaccine strain from the parental Asibi, the virus underwent a varied passage history with three substrains recognized, named 17D-204, 17DD and 17D-213 based on the passage level that they were derived (World Health Organization 2010). Briefly, the 17DD substrain was derived from 17D at passage 195, and is currently manufactured only in Brazil. The 17D-204 substrain refers to all vaccines derived from the original 17D strain at passage 204. The 17D-213 substrain refers to a derivative of 17D-204 developed by the RKI to be free of contamination by avian leucosis virus; 17D-213/77 is the prototype of this substrain, and is the primary reference seed of the WHO (Bres and Koch 1987). Over the last 75 years 17D vaccine has been produced in 14 different countries (Australia, Brazil, China, Colombia, France, Germany, India, Netherlands, Nigeria, Russia, Senegal, South Africa, United Kingdom, and United States) (Monath, M. Gershman, et al. 2013). Since the lineage relationships of all YFV vaccines by passage level is exactly known, the tree represented by the strains represents a dataset by which to interrogate the stability of the 17D genotype by traversal of known passage histories.

In early development of YFV vaccines, phenotypic stability was investigated by *in vivo* classical methods. 17D was shown to regain neurotropism to monkeys if serially passaged in mouse brain (Theiler 1951, p. 129). Intrahepatic (i.h.) passage of the French neurotropic vaccine strain, a competing live-attenuated product in Rhesus macaques revealed that viscerotropism could be efficiently selected and enhanced, leading to an early speculation as to the inherent diversity in tissue-fixed viral vaccines (Findlay and Clarke 1935). One early field trial in a Brazilian cohort observed a loss of immunogenicity for a 17D-derived vaccine batch produced from

seeds at extended subculture levels (Soper and H. H. Smith 1938).

These safety concerns prompted the development of a seed-lot system for the vaccine in 1945, a regulatory structure that remains in use today. Production lots of each 17D vaccine are derived from first a primary, then secondary, seed material qualified to possess acceptable phenotypic and genotypic characteristics (World Health Organization 2010). Because production lots of vaccine arise by single egg passage from secondary seed virus, subculture levels of production batches relative to the parental strain are exactly known, providing a reference by which to interrogate stability and consistency of the vaccine lineage over time with modern tools. A collection of vaccines and seeds from the 17D-204 and 17DD substrains were compared by T1 RNA oligonucleotide fingerprinting, revealing slight differences in electrophoretic patterns between the substrains (Monath, Kinney, et al. 1983).

The French Stamaril® vaccine (17D-204 substrain) was analyzed by sequence, multiplication kinetics, and plaque size distribution, revealing considerable stability for a set of production lots manufactured between 1990 and 2002 (Barban et al. 2007). Limited divergence of consensus sequence is observed across the entire 17D lineage; Bayesian phylogenetic analysis of consensus sequences for a set of 17D vaccines resolved clustering patterns by substrain history (Stock et al. 2012).

Quasispecies theory is now a prominent framework for understanding the lifecycle of RNA viruses such as YFV, which replicate under short generation times by means of error-prone RNA-dependent-RNA-polymerases (Domingo, Sheldon, and Perales 2012). Briefly, low-frequency sequence variants are observed to arise during the viral lifecycle, underlying a dominant genotype. The presence and relative abundance of these variants have been shown to be possible fitness determinants for the viral population, which is considered an aggregate unit of selection. This was shown most dramatically for PoV, in which chemical mutagenesis restored neuropathogenic-

ity of a strain bearing a high-fidelity polymerase and consequently impaired fitness (Vignuzzi et al. 2006). For YFV, a strain of the 17D vaccine was estimated to accumulate between  $1.9 \times 10^{-7}$  and  $2.3 \times 10^{-7}$  mutations per replicated genome, assessed by Sanger methods (Pugachev, Guirakhoo, et al. 2003). The potential significance of this feature to the safety profile of 17D has not been determined experimentally, however since viral subpopulations are hypothesized to facilitate adaptation to within-host infection barriers, low mutational rates or content for YFV vaccines are desirable to avoid the potential for WT reversion once administered. Empirically derived vaccine strains such as 17D therefore represent an ideal context by which to investigate the association of viral diversity with phenotype. The availability of standardized vaccine strains permits the reconstruction of typical population structures that may define an attenuated vaccine genotype.

Massively parallel sequencing (NGS) studies are frequently used to interrogate viral diversity, resolving subpopulation dynamics in response to various external factors, including tissue compartmentalization, immune system pressures, and pharmaceutical treatments (Nishijima et al. 2012). For vaccines, studies of this type have detected the presence of adventitious virus in commercial lots, after accomplishing low-coverage assembly ( $<100$  mean reads per nucleotide site considered in the alignment) of a number of live-attenuated RNA virus vaccine strains, including measles virus (MeV), rubella virus (RbV), and a 17D-derived vaccine lot (Victoria et al. 2010).

For YFV specifically, a deep sequencing study at high coverage ( $>5000$  mean reads per nucleotide site considered) compared to WT parental strain Asibi with a commercial lot of the 17D-204 substrain vaccine manufactured in the U.S. (YF-Vax<sup>®</sup>) and revealed that the vaccine is in fact largely homogenous, and devoid of hidden diversity relative to the WT strain (A. Beck et al. 2014). To investigate

whether or not the same homogeneity is found in 17D vaccines manufactured by other producers, we have performed a deep sequencing study of primary and/or secondary seed lots, and some vaccine production lots for all modern substrains in present use, 17D-204, 17DD, and 17D-213. The assembled panel of seeds and vaccines represents production of YFV vaccine from diverse subculture histories, permitting comparisons of both intra- and inter-lineage vaccine seed strains [Table 4.1 and Figure 4.1]. Findings represent the most complete record of YFV seed and vaccine relationships to date using these methods, and offers a framework by which the stability and consistency of the vaccine may be assessed over extended production histories *in silico*.

### 4.3 Results

Origins and nomenclature for the strains considered are listed in Table 4.1.

#### 4.3.1 17D-204 Substrain Coverages

For the secondary seed 17D-204/United Kingdom/YFS/11, median coverage was 7491 (mean = 7275.4, s.d.=1118.6). For the secondary seed 17D-204/South Africa/5114.15, median coverage was 7499 (mean = 7278.4, s.d.=1198.4). For the Connaught vaccine strain 17D-204/USA/1995 (YF-Vax<sup>®</sup> lot 4L51152), median coverage was 7400 (mean = 7286.2, s.d.=1400.7). For the Sanofi-Pasteur vaccine strain 17D-204/USA/2012 (YF-Vax<sup>®</sup> lot UH356AA), median coverage was 7534 (mean = 7279.2, s.d.=1154.9). For the Connaught vaccine strain 17D-204/USA/1995 (YF-Vax<sup>®</sup> lot 4L51152), median coverage was 7400 (mean = 7286.2, s.d.=1400.7). For the National Drug Company (NDC) secondary seed 17D-204/USA/NDC-6676, median coverage was 8704.5 (mean = 7279.2, s.d.=4146.0). For the NDC vaccine strain 17D-204/USA/NDC-6677, median coverage was 7350 (mean = 7277.3, s.d.=3582.6).



### 4.3.2 17D-213 Substrain Coverages

Consensus for the WHO primary seed lot 17D-213/WHO/213-77, median coverage was 7160 (mean = 7280.7, s.d.=1923.7). For the vaccine strain 17D-213/WHO/213-B2-81, median coverage was 7311 (mean = 7280.0, s.d.=1595.2). For the vaccine strain 17D-213/Russia/SW13, median coverage was 7699 (mean = 7279.7, s.d.=2400.9). For the vaccine strain 17D-213/Nigeria, median coverage was 7017 (mean=7278.3, s.d.=3966.0).

### 4.3.3 17DD Substrain Coverages

For the secondary seed 17DD/Senegal/SLII-75-1, median coverage was 7574 (mean=7281.0, s.d.=1230.7). For the vaccine strain 17DD/Colombia, median coverage was 7278 (mean = 7277.9, s.d.=1216.3).

### 4.3.4 Asibi Strain Coverage

For the WT parental strain Asibi, median coverage was 7311 (mean=7280.0, s.d.=1595.2).

### 4.3.5 Consensus Sequences

Consensus sequences for each vaccine virus were aligned to the WT parental strain Asibi (Genbank accession AY640589), and compared to the *de novo* Asibi sequence generated in parallel. For the entire set of 17D vaccine and seed strains considered, relative to the WT Asibi strain, a set of 24 coding substitutions was recovered in common between all datasets (M-L36F, E-G52R, E-A170V, E-T173I, E-K200T, E-M299I, E-S305F, E-K331R, E-T380R, E-A407V, NS1-L79F, NS1-I307V, NS2A-M118V, NS2A-T167A, NS2A-T172A, NS2A-S183F, NS2B-I37L, NS2B-I109L, NS3-D485N, NS4A-V146A, NS4B-I95M, NS4B-Y231H, NS5-E836K, NS5-P901L) and two

nucleotide substitutions in the 3'UTR (U10418C, A10847C) [Table 4.2]. The current WHO standard for YFV vaccines lists 20 coding substitutions and four nucleotide changes in the 3'UTR (3'UTR) in reference to the parental strain Asibi that are conserved between all substrains of 17D [Table 1.1] (World Health Organization 2010). All of these sites were recovered in the vaccines and seeds considered in the study, with substrain-specific, incomplete fixation of sites in the 3'UTRs. Nucleotide substitutions U10367C and U10550C were exclusively conserved in 17D-204 and 17D-213 substrain viruses. Amino acid substitutions relative to the Asibi strain are summarized in Table 4.2.

Fixation across substrains was observed at all coding sites and 3 of four untranslated sites (U10418C, G10800A, A10847C) described by the WHO standard; excepting the presence of WT identity in the vaccine strain 17D-204/USA/2012 at nucleotide position A10800G. Thus, in addition to the previously identified 20 common amino acid substitutions, the amino acid substitutions E-K331R, NS1-L79F, NS2B-I37L, and NS4B-Y232H were conserved across all vaccine strains considered. Note E-K331R had been previously reported as a vaccine-specific residue, but the lysine residue in Asibi is unique among sequenced WT YFV strains, and in addition other WT strains also carry the arginine residue. Using an alignment of all publicly available full-length YFV sequences, the arginine residue was observed in all vaccine sequences; tyrosine was conserved in all African WT sequences except that of the West African FVV, and not in any South American WT strains [Figure 4.2,4.3,4.4,4.5].

#### 4.3.5.1 Consensus Analysis of 17D-204 Substrain

The secondary seed 17D-204/UK/YFS/11 differed from Asibi at 67 nucleotides, 30 of which were coding substitutions. The secondary seed 17D-204/South Africa/5114.15

differed from Asibi at 67 nucleotides, 30 of which were coding substitutions. The Connaught vaccine strain 17D-204/USA/1995 differed from Asibi at 67 nucleotides, of which 31 were coding substitutions. The Sanofi-Pasteur vaccine strain 17D-204/USA/2012 differed from Asibi at 66 nucleotides, 31 of which were coding substitutions. The NDC secondary seed 17D-204/USA/NDC-6676 differed from Asibi at 65 nucleotides, of which 30 were coding substitutions. The NDC vaccine strain 17D-204/USA/NDC-6677 differed from Asibi at 65 nucleotides, of which 30 were coding substitutions. Thus, a set of 30 coding substitutions was recovered in common for sequences of the 17D-204 substrain [Table 4.3].

#### **4.3.5.2 Consensus Analysis of 17D-213 Substrain**

The WHO primary seed lot 17D-213/WHO/213-77 differed from Asibi at 66 nucleotides, 31 of which were coding substitutions. The vaccine strain 17D-213/WHO/213-B2-81 differed from Asibi at 65 nucleotides, 31 of which were coding substitutions. The vaccine strain 17D-213/Nigeria differed from Asibi at 63 nucleotides, of which 30 were coding substitutions. The vaccine strain 17D-213/Russia/SW13 differed from Asibi at 67 nucleotides, of which 31 were coding substitutions. Thus, a set of 31 coding substitutions was recovered in common for sequences of the 17D-213 substrain in reference to public 17D-213 sequences, excepting reversion to WT in the vaccine strain 17D-213/Nigeria at sites E-N153T (E protein N-linked glycosylation site), and NS5-Q22R [Table 4.4] (C. N. d. Santos et al. 1995). The WT residue NS4B-Y232L was identified in the newly sequenced vaccines, while being absent in comparator Genbank sequences YFU17067 and JN628279 [Table 4.4].

#### 4.3.5.3 Consensus Analysis of 17DD Substrain

The secondary seed 17DD/Senegal/SLII-75-1 differed from Asibi at 65 nucleotides, 27 of which were coding substitutions. The vaccine strain 17DD/Columbia, differed from Asibi at 66 nucleotides, 27 of which were coding substitutions. Thus, a set of 27 coding substitutions was recovered in common for sequences of the 17DD substrain [Table 4.5]. Coding substitutions were identical to those described by dos Santos and colleagues, in reference to Genbank accession YFU17066, except the observation of WT nucleotide identity at position T10367C in the 3'UTR for both strains (C. N. d. Santos et al. 1995) [Table 4.5].

#### 4.3.6 Diversity of Strains

Summary values of NSE were similar for each virus, along multiple scales. [Table 4.6]. Using a Kruskal-Wallis rank-sum test factored by substrain, a significant effect was observed ( $X^2=23.71$ , degrees of freedom (d.f.)=2,  $p<0.001$ ). Using a single-factor Kruskal-Wallis rank-sum test factored by individual strain, a significant effect was observed ( $X^2=237.72$ ,  $df=11$ ,  $p<0.001$ ). Post-hoc Dunn's tests for pairwise effects revealed differences between 17DD and both 17D-204 and 17D-213 substrains using Bonferroni correction (17DD/17D-204:  $Z=-3.57$ ,  $p=.001$ ) (17DD/17D-213:  $Z=4.77$ ,  $p<.001$ ). Significant differences were not observed between 17D-204 and 17D-213 substrains. Significant differences were observed at the virus strain level for 34 of 66 possible pairs [Figure 4.6] To investigate potential bias of high-entropy artifacts arising from low coverage at the 5' and 3' termini of the genomes, the test was repeated excluding the terminal 5' and 3'UTRs, and one less pairwise comparison (17D-204/USA/NDC-6676 and 17D-204/USA/NDC-6677, 33 total) differed significantly in the dataset [Figure 4.7].

### 4.3.7 Divergence of Strains and Stability of Vaccine Genotypes

For each virus strain, genetic distance of each nucleotide position from the WT strain Asibi was represented as RMSD, with each vaccine strain showing considerable stability of selection for the conserved vaccine genotype [Figure 4.8]. Phylogenetic relationships calculated using median RMSD across the viral open reading frame did not fully resolve the subculture divergence of the vaccines (that is already known by serial passage histories), however basal substrain groupings were resolved. [Figure 4.9].

### 4.3.8 Identification of Variant Structure and Substitution Ratios *In Situ*

For all vaccine strains considered, variant nucleotide sites were recovered by modeled criteria. Relative to the consensus sequence of each strain considered, the  $dN/dS$  ratio was computed for the entire open reading frame. For the vaccine strain 17D-213/WHO/213-B2-81, 138 of 244 variants recovered were coding, with reversion at 18 coding sites to WT Asibi ( $dN/dS=0.651$ ). For the secondary seed 17DD/Senegal/SLII-75-1, 126 of 201 variants recovered were coding, with 23 WT coding revertants ( $dN/dS=0.432$ ). For vaccine strain 17DD/Columbia, 231 of 377 variants recovered were coding, with 14 WT coding revertants ( $dN/dS=0.403$ ). For the primary seed lot 17D-213/WHO/213-77, 242 of 673 variants recovered were coding, with 8 WT coding revertants ( $dN/dS=0.611$ ). For the vaccine strain 17D-204/USA/2012 556 of 300 variants recovered were coding, with 8 WT coding revertants ( $dN/dS=0.572$ ). For the vaccine strain 17D-204/USA/1995 643 of 349 variants recovered were coding, with 10 WT coding revertants ( $dN/dS=0.564$ ). For the secondary seed 17D-204/UK/YFS/11, 175 of 103 variants recovered were coding, with 3 WT coding revertants ( $dN/dS=0.358$ ). For the secondary seed 17D-204/South Africa/5114.15 99 of 165 variants recovered were coding, with 1 WT

revertant ( $dN/dS=0.443$ ). For the vaccine strain 17D-213/Nigeria, 393 of 303 variants were coding, with 14 WT coding revertants ( $dN/dS=0.508$ ). For the secondary seed 17D-204/USA/NDC-6676, 452 of 156 variants were coding, with 6 WT coding revertants ( $dN/dS=0.533$ ). For the vaccine strain 17D-204/USA/NDC-6677, 693 of 339 variants were coding, with 14 WT coding revertants ( $dN/dS=0.561$ ). For the vaccine strain 17D-213/Russia/SW13, 60 of 77 variants were coding, with 3 WT coding revertants ( $dN/dS=0.746$ ). SNVs and revertants are depicted in Figure 4.10. In reference to the reprocessed Asibi strain, a minority of WT coding revertants were present in the vaccine at very low frequencies; the presence of these coding revertants was frequently repeated between members arising from the same substrain lineage [Table 4.7].  $dN/dS$  ratios were summarized along each gene segment for all strains sequenced in the study [Table 4.8, Figure 4.11]. Identified  $dN/dS$  peaks greater than 1 (positive selection) were infrequently observed in the 17D strain alignments, and did not occur equivalently [Figure 4.12].

#### **4.3.9 Traversal of 17D Lineage Using Sequentially Passaged Strains: Seed-lot Relationships**

##### **4.3.9.1 Transition of secondary seed 17D-204/USA/NDC-6676 to vaccine strain 17D-204/USA/NDC-6677: 1 Passage Distance**

No consensus differences were observed between the strains. Using the paired-test model, 58 sites increased in frequency between 17D-204/USA/NDC-6676 and 17D-204/USA/NDC-6677. Of these, 30 were amino acid substitutions (mean frequency= 0.097) relative to the 17D-204/USA/NDC-6676 consensus sequence. A WT revertant at nucleotide position 1431 was observed at a frequency of 0.01, coding for positive selection of the amino acid substitution E-T153N, an N-linked glycosylation site ( $p<0.001$ ).

**4.3.9.2 Transition of primary seed 17D-213/WHO/213-77 to vaccine strain 17D-213/Nigeria (>1 Passage Distance; exact passage of vaccine is unknown):**

Four consensus differences were identified, with one amino acid substitution E-T153N a reversion to the WT identity, encoding for acquisition of an N-linked glycosylation site. Using the paired-test model, 53 variants increased in frequency between 17D-213/WHO/213-77 and 17D-213/Nigeria. 21 of these (mean frequency= 0.012) were amino acid substitutions relative to the 17D-213/WHO/213-77 consensus sequence.

**4.3.9.3 Transition of primary seed 17D-213/WHO/213-77 to vaccine strain 17D-213/WHO/213-B2-81: 2 passage distance**

No consensus differences were observed between the strains. Using the paired-test model, 54 variants increased in frequency between 17D-213/WHO/213-77 and 17D-213/WHO/213-B2-81. 29 of these (mean frequency= 0.097) were amino acid substitutions relative to the 17D-213/WHO/213-77 consensus sequence.

**4.3.9.4 Transition of primary seed 17D-213/WHO/213-77 to vaccine strain 17D-213/Russia/SW13: >1 passage distance**

No consensus differences were observed between the strains. Using the paired-test model, 24 variants increased in frequency between 17D-213/WHO/213-77 and 17D-213/Russia/SW13. 6 of these (mean frequency= 0.017) were amino acid substitutions relative to the 17D-213/WHO/213-77 consensus sequence.

**4.3.9.5 Transition of secondary seed 17DD/Senegal/SLII-75-1 to vaccine strain 17DD/Columbia: 1 passage distance**

No consensus differences were observed between the strains. Using the paired-test model, 6 variants increased in frequency between 17D-17DD/Senegal/SLII-75-1

and 17DD/Columbia. One of these (frequency = 0.015) was an amino acid substitution relative to the 17DD/Senegal/SLII-75-1 consensus sequence.

#### **4.3.10 Comparison of 17D Lineage Strains At The Same Passage Level**

##### **4.3.10.1 Secondary seed 17D-204/UK/YFS-11 and secondary seed 17D-204/South Africa/5114.15**

One silent nucleotide difference was observed in the consensus sequence of the South African seed at position U4873G, which was identified as WT in the comparison of the parental Asibi strain with an ATCC collection strain of 17D-204 (Hahn et al. 1987). Using the paired-test model, 38 variants increased in frequency between 17D-204/UK/YFS-11 and 17D-204/South Africa/5114.15. Eight of these (mean frequency= 0.084) were amino acid substitutions relative to the 17D-204/UK/YFS-11 consensus sequence.

##### **4.3.10.2 YF-Vax<sup>®</sup> strains 17D-204/USA/1995 and 17D-204/USA/2012**

No consensus differences were observed between the strains. Using the paired-test model, 5 variants increased in frequency between the 17D-204/USA/1995 and 17D-204/USA/2012 lots of the vaccine, of which one (frequency=0.005) was an amino acid substitution relative to the 17D-204/USA/1995 consensus sequence.

## **4.4 Discussion**

### **4.4.1 17D Seeds and Vaccines Represent A Phenotypically Stable Framework To Investigate Population-Level Effects**

The maintenance of attenuating sequence features for live-attenuated vaccines is necessary to ensure consistency of the safe and efficacious properties of production batches. Of particular concern is the potential susceptibility of the virus to selection or drift at these sites, which may produce undesirable change in the



vaccine phenotype once administered. The consistency of YFV vaccine seeds has been reported in both nonclinical and clinical trials, in which vaccine produced from replacement seeds is assessed for relative immunogenicity and tolerability compared to previous seeds and/or lots. To date, limited evidence has been put forth in the literature to suggest any meaningful divergence in protection offered by 17D vaccines as currently produced from substrains 17D-204, 17D-213, and 17DD, however, the 17DD substrain has been reported to induce a slightly higher pro-inflammatory profile in the innate immune response compared to 17D-204 and 17D-213 viruses. In a clinical trial comparing vaccine lots produced from substrains 17DD and 17D-213, seroconversion rates were equivalent for cohorts receiving either vaccine; immunogenicity was higher for recipients of vaccines produced from the 17D-213 substrain (Camacho, M. d. S. Freire, et al. 2004). Similar results were reported for a comparison of 17D-213 and 17D-204 substrain-derived vaccines. Though all licensed YFV vaccines originate from seeds meeting or exceeding standard immunogenicities (and thus all are considered protective), mechanisms influencing substrain differences in trial cohorts are unknown. Substrain-comparative data on safety is limited, however equivalent rates of adverse events following administration of 17DD and 17D-213 substrain vaccines have been reported for both adult and pediatric cohorts (Camacho, Aguiar, et al. 2005; Ripoll et al. 2014). The safety profile of 17D vaccines is considerable, with the few reported serious adverse events attributed to host factors (Silva et al. 2010). Selective pressure of functional innate immune responses are presumably required to prevent dissemination of the vaccine virus in host tissues, consequently a state of immunosuppression often contraindicates administration of the vaccine (Erickson and Pfeiffer 2013).

#### 4.4.2 Previously Unobserved Sequence Content of the Vaccines Is Revealed by Deep Sequencing

Consensus sequences of the vaccines were reconstructed *de novo* from population nucleotide frequencies, demonstrating strong similarity of 17D seed substrains studied here to those already published. In reference to the WT parental strain Asibi, 20 amino acid substitutions are contained in the WHO standard for the 17D vaccine sequence; these sites were recovered in our data and are conserved across all vaccine viruses (World Health Organization 2010).

We observed four additional conserved substitutions E-K331R, NS1-L79F, NS2B-I37L, NS4B-Y232H [Figure 4.2,4.3,4.4,4.5]. Though the substitution E-K331R was observed in all vaccine strains considered in the study, the arginine residue is found in other WT strains, and was specifically observed in an Asibi strain adapted by passage to productively infect hamster liver; any contribution to vaccine attenuation is therefore unclear (McArthur et al. 2003). The substitution NS1-L79F possibly reflects ecological niche adaptation of the parental virus; for WT African YFV strains, the leucine residue was observed in West African strains (the parental Asibi strain was isolated in Ghana in 1929 and is a member of this taxonomic group); phenylalanine was instead observed at the codon in East African strains. Again, observation of the vaccine residue at this position in WT strains would likely indicate little influence on attenuation of 17D strains.

The WT isoleucine residue of the substitution NS2B-I37L is conserved in all African and South American strains presently sequenced at this region; again, the substitution is conservative, and has not been tested in an infectious clone system. The amino acid substitution NS4B-Y232H is a nonconservative substitution located in the transmembrane domain 5 of the protein that has been suggested to be involved in suppression of RNA interference (Kakumani et al. 2013), any conferred phenotype

has not been tested. Functions of NS4B mutations in attenuation of the vaccine is not understood, however evidence has been put forth to suggest roles for NS4B in both interferon antagonism and as a component of the replication complex (J. Zou, X. Xie, L. T. Lee, et al. 2014; Muñoz-Jordán et al. 2005).

The WT parental strain Asibi sequence generated in the study was compared to that of Genbank sequence AY640589, and no confounding instability was identified at these four sites that would otherwise be attributed to laboratory adaptation of Asibi. It is therefore likely that NS2B-I37L and NS4B-Y232H are selected in the vaccine preparations themselves. As reported previously, the 17DD and 17D-213 substrain viruses were observed to have acquired a predicted N-linked glycosylation site in the envelope protein at residue 153, which is not carried in vaccines or seeds of the 17D-204 substrain (C. N. d. Santos et al. 1995). Acquisition of the residue was associated with loss of monkey neurotropism for the WHO reference lot 168-73, but this was not confirmed by an infectious clone (Moulin et al. 2013). The predicted glycosylation was observed in a strain recovered from a fatal case of YEL-AND, however the virus (which was of greater neurotropism in a mouse model than the vaccine itself) did not differ at the asparagine residue from that of the vaccine lot administered (Jennings et al. 1994).

#### **4.4.3 Differences in Vaccine Diversity Are Associated With Passage Distance**

The statistically significant differences in diversity profile between pairs of strains should be interpreted in context of passage history trends. Nonparametric comparison of normalized entropy shows that differences in the nucleotide diversity profiles of the strains is potentially influenced by passage distance, indicating that some random subpopulation drift occurs over extended passage histories. When all the vaccine viruses are considered in aggregate, there is some evidence for de-

viation of diversity profiles between the vaccines, although again, clinical data has demonstrated no loss of protection across the substrains. It is notable that significant differences in diversity profiles were observed when comparing strains that were characterized nonclinically, but not manufactured. In particular, the seed lot 17D-204/USA/NDC-6676 was produced by the NDC in 1969 from Rockefeller Foundation lot 145-3A, in an attempt to produce by filtration a seed stock free of ALV. The vaccine was characterized for equivalence to ALV-containing materials, demonstrating equivalent immunogenicity and neurotropism to ALV-containing reference materials (Tauraso, Spector, Jahnes, et al. 1968; Tauraso, Spector, and Trimmer 1972). The seed 17D-204/USA/NDC-6676 failed a standard monkey neurovirulence test, and was never used in production (Monath, Kinney, et al. 1983). Paradoxically, analysis of 17D-204/USA/NDC-6676 by T1 oligonucleotide fingerprint found no difference in electrophoretic patterns between the seed and a 17D-204 reference strain but this technique is only sensitive to 10% of the genome (Monath, Kinney, et al. 1983). The seed lot was observed to contain a relatively high number of windowed regions with  $dN/dS > 1.0$ ; this instability may reflect the undesirable properties of the strain [Figure 4.12], although this has not yet been shown experimentally. Under laboratory conditions, 17D is observed to possess an inherent level of diversity with respect to both monoclonal antibody epitope antigenicity and neurotropism *in vivo* (Gould et al. 1989). Also, there are noise effects from the sequencing chemistry and PCR reactions that obscure diversity effects for extremely low nucleotide frequencies.

#### **4.4.4 Purifying Selection Contributes to Maintenance of the Vaccine Genotype**

Consistency of small-frequency variant population is observed across the 17D lineage as a general pattern, however. For single mutations identified with a probabilistic model, the distribution of variant frequencies is similar between strains

[Figure 4.13]. For the comparison vaccine strains arising from the same manufacturer (17D-204/USA/1995 and 17D-204/USA/2012), the observation of very low nucleotide frequency change, coupled with no consensus change is certainly an indicator of high stability for the vaccine over a local production history. For all comparisons of strains using the paired-test model, it is especially significant that, even in cases of changing nucleotide frequency between strains, the frequencies involved in both sides of the transition are very low.  $dN/dS$  ratios recovered for whole open reading frame and for gene segments are less than 1.00 in the most of cases, supporting a plausible hypothesis that negative selection is a predominant influence on the highly adapted vaccine genomes.  $dN/dS$  is less accurate and artificially smoothed over long genomic distances, and is naïve to the localized selection pressures that are domain specific, however, this likely represents a resolution of tolerable mutational diversity in the vaccine strains.

#### **4.4.5 Naturally Arising Diversity In Flaviviral Vaccines Influences Manufacture**

Natural diversity in live-attenuated preparations suggests that population structure is likely to be valuable when evaluating the safety and lot-to-lot consistency of vaccine produced by novel technologies. Adaptation of 17D derivatives to novel production methods is, in some cases, likely to promote instability of the vaccine. In Vero cell culture, adaptation of a 17D-204 strain was associated with consensus mutations at genome regions observed to be unstable in other contexts (Beasley et al. 2013). When a 17DD seed was adapted to CEF, phenotypic alterations were observed; standard neurovirulence scores in monkeys were elevated, although not to statistical significance (M. S. Freire et al. 2005). Paradoxically, the sequence of CEF-adapted virus was identical to that of the input by consensus Sanger sequencing. Therefore, it is hypothesized that elucidation of key population structures for live

vaccines will permit the association of phenotype with the variable presence of WT alleles.

In summary, the findings demonstrate a remarkable level of stability in the 17D-derived vaccines across a broad subculture history. Patterns of subpopulation structure in the vaccine occur repeatably along small passage scales in the history of 17D; little evidence has been found in this dataset to identify any pervasive destabilizing effects on the vaccine over time in response to egg production methods. Most significantly, the presented findings substantiate previous observations of small-scale diversity in the vaccine as evidenced by extremely low WT sequence content. Therefore, based on these findings, it is suggested that NGS technology has potential application to differentiate levels of diversity in live attenuated vaccine preparations that may be of consequence to the monitoring and prediction of vaccine strain phenotype.

## **4.5 Conclusions**

The presented results represent the first comprehensive analysis by deep sequencing of a live-attenuated vaccine strain that directly incorporates lineage and passage history. The vaccine was largely found to have a stable population structure over the passage scales represented by the collected panel of vaccine strains. The observation of low-frequency WT sequence content is expected due to random effects, however these observations are not expected to be consequential to the safety profile of the vaccine as most of the seeds used in the study were used to generate production lots. Significantly, the use of rare 17D seed strains indicates that the results are representative of the stability of vaccine production for lots numbering in the millions of doses.

**Table 4.1:** Origin information for the seeds used in the study.

Substrain	Strain	Type	Origin and Lot Number	Subculture Level
17D-204	5114.15	Secondary Seed	National Institute for Virology, South Africa	233
17D-204	NDC-6676	Secondary Seed	National Drug Company, USA	233
17D-204	NDC-6677	Vaccine	National Drug Company, USA	234
17D-204	YFS-11	Secondary Seed	Wellcome Research Laboratories, UK	234
17D-204	YF-Vax/2012	Vaccine	Sanofi-Pasteur USA	237
17D-204	YF-Vax/1995	Vaccine	Lot: UH356AA Connaught Laboratories USA	237
17D-213	Russia/SW13	Vaccine	Lot: 4L51152 Institute for Poliomyelitis and Viral Encephalitis, Russia	239
17D-213 <sup>a</sup>	Nigeria	Vaccine	Vaccine Production Laboratories, Nigeria Lot:625	Unknown
17D-213	213/B2/81	Vaccine	World Health Organization	239
17D-213	213/77	Primary Seed	Robert Koch Institute/WHO	237
17DD	SLII-75-1	Secondary Seed	Institute Pasteur, Senegal	286
17DD	Colombia	Vaccine	Instituto Nacional de Salud, Colombia	287

**Table 4.2:** Amino acid and 3'UTR substitutions between Asibi and all 17D strains sequenced in the study. Pos = Nucleotide Position, Cod = Local Codon, AA = Amino Acid, UTR=Untranslated Region.

Asibi		17D-204			17D-213					17DD				
Pos	Gene	Cod	AA	YF-Vax® 2012	YF-Vax® 1995	YFS-11	5114.15NDC 6676	NDC 6677	213-77	213-B2-81	Russia SW13	Nigeria	SLII-75-1	Colombia
854	M	36	L	F	F	F	F	F	F	F	F	F	F	F
1127	E	52	G	R	R	R	R	R	R	R	R	R	R	R
1140	E	56	A	V	V	V	V	V	V	V	V	V		
1431	E	153	N	T	T		T		T	T	T			
1482	E	170	A	V	V	V	V	V	V	V	V	V	V	V
1491	E	173	T	I	I	I	I	I	I	I	I	I	I	I
1572	E	200	K	T	T	T	T	T	T	T	T	T	T	T
1692	E	240	A		V		V							
1870	E	299	M	I	I	I	I	I	I	I	I	I	I	I
1887	E	305	S	F	F	F	F	F	F	F	F	F	F	F
1946	E	325	P	S	S	S	S	S	S	S	S	S		
1965	E	331	K	R	R	R	R	R	R	R	R	R	R	R
2112	E	380	T	R	R	R	R	R	R	R	R	R	R	R
2193	E	407	A	V	V	V	V	V	V	V	V	V	V	V
2219	E	416	A	T	T	T	T	T	T	T	T	T		
2220	E	416	A										V	V
2687	NS1	79	L	F	F	F	F	F	F	F	F	F	F	F
3371	NS1	307	I	V	V	V	V	V	V	V	V	V	V	V
3860	NS2A	118	M	V	V	V	V	V	V	V	V	V	V	V
4007	NS2A	167	T	A	A	A	A	A	A	A	A	A	A	A
4013	NS2A	169	L	F	F	F	F	F	F	F	F	F		
4022	NS2A	172	T	A	A	A	A	A	A	A	A	A	A	A
4056	NS2A	183	S	F	F	F	F	F	F	F	F	F	F	F
4289	NS2B	37	I	L	L	L	L	L	L	L	L	L	L	L
4505	NS2B	109	I	L	L	L	L	L	L	L	L	L	L	L



[illegible]

**Table 4.3:** 17D-204 strain amino acid and 3'UTR substitutions using the newly sequenced strains, in reference to publicly available sequences. Pos = Nucleotide Position in Asibi, Cod = Local Codon, AA = Amino Acid, UTR = Untranslated Region.

Pos	Gene	Cod	AA	YF- Vax <sup>®</sup> 1995	YF- Vax <sup>®</sup> 2012	YFS/11	5114.15	NDC 6676	NDC 6677	Stamaril <sup>®</sup> ATCC X15062 <sup>a</sup> X03700 <sup>a</sup>	Tiantan <sup>®</sup> FJ654700 <sup>a</sup> Vax <sup>®</sup>	YF- UF795AA
237	C	40	P								L	
854	M	36	L	F	F		F	F	F	F	F	F
1127	E	52	G	R	R		R	R	R	R	R	R
1140	E	56	A	V	V		V	V	V	V	V	V
1431	E	153	N	T	T		T	T			T	T
1482	E	170	A	V	V		V	V	V	V	V	V
1491	E	173	T	I	I		I	I	I	I	I	I
1572	E	200	K	T	T		T	T	T	T	T	T
1692	E	240	A				V				V	
1870	E	299	M	I	I		I	I	I	I	I	I
1887	E	305	S	F	F		F	F	F	F	F	F
1946	E	325	P	S	S		S	S	S	S	S	S
1965	E	331	K	R	R		R	R	R	R	R	R
2112	E	380	T	R	R		R	R	R	R	R	R
2193	E	407	A	V	V		V	V	V	V	V	V
2219	E	416	A	T	T		T	T	T	T	T	T
2687	NS1	79	L	F	F		F	F	F	F	F	F
3371	NS1	307	I	V	V		V	V	V	V	V	V
3740	NS2A	78	T								A	
3860	NS2A	118	M	V	V		V	V	V	V	V	V
4007	NS2A	167	T	A	A		A	A	A	A	A	A
4013	NS2A	169	L	F	F		F	F	F	F	F	F
4022	NS2A	172	T	A	A		A	A	A	A	A	A
4056	NS2A	183	S	F	F		F	F	F	F	F	F
4289	NS2B	37	I	L	L		L	L	L	L	L	L

4505	NS2B	109	I	L	L	L	L	L	L	L	L	L	L
4559	NS2B	127	G									R	
5153	NS3	195	I	V		V	V	V	V	V	V	V	V
6023	NS3	485	D	N	N	N	N	N	N	N	N	N	N
6758	NS4A	106	I									V	
6876	NS4A	146	V	A	A	A	A	A	A	A	A	A	A
7171	NS4B	95	I	M	M	M	M	M	M	M	M	M	M
7580	NS4B	232	Y	H	H	H	H	H	H	H	H	H	H
7701	NS5	22	Q	R	R	R	R	R	R	R	R	R	R
8153	NS5	173				S	S	S					
9605	NS5	657	N								D		
10142	NS5	836	E	K	K	K	K	K	K	K	K	K	K
10338	NS5	901	P	L	L	L	L	L	L	L	L	L	L
10367	3'UTR	-	T	C	C	C	C	C	C	C	C	C	C
10418	3'UTR	-	T	C	C	C	C	C	C	C	C	C	C
10550	3'UTR	-	T	C	C	C	C	C	C	C	C	C	C
10555	3'UTR	-	A	A		C	G						
10722	3'UTR	-	G					A					
10800	3'UTR	-	G		A	A	A	A	A	A	A	A	A
10815	3'UTR	-	G									A	
10847	3'UTR	-	A	C	C	C	C	C	C	C	C	C	C
10854	3'UTR	-	A		M								
10860	3'UTR	-	A									T	

<sup>a</sup> Genbank accession for publicly available sequence.

**Table 4.4:** 17D-213 strain amino acid and 3'UTR substitutions using the newly sequenced strains, in reference to publicly available sequences. Pos = Nucleotide Position, Cod = Local Codon, AA = Amino Acid, UTR=Untranslated Region.

Pos	Gene	Cod	AA	Nigeria	Russia SW13	213-77	213-B2- 81	213-77 YFU17067 <sup>a</sup>	RK1/Crucell JN628279 <sup>a</sup>
854	M	36	L	F	F	F	F	F	F
1127	E	52	G	R	R	R	R	R	R
1140	E	56	A	V	V	V	V	V	V
1431	E	153	N	T	T	T	T	T	T
1482	E	170	A	V	V	V	V	V	V
1491	E	173	T	I	I	I	I	I	I
1572	E	200	K	T	T	T	T	T	T
1870	E	299	M	I	I	I	I	I	I
1887	E	305	S	F	F	F	F	F	F
1946	E	325	P	S	S	S	S	S	S
1965	E	331	K	R	R	R	R	R	R
2112	E	380	T	R	R	R	R	R	R
2193	E	407	A	V	V	V	V	V	V
2219	E	416	A	T	T	T	T	T	T
2687	NS1	79	L	F	F	F	F	F	F
3371	NS1	307	I	V	V	V	V	V	V
3860	NS2A	118	M	V	V	V	V	V	V
4007	NS2A	167	T	A	A	A	A	A	A
4013	NS2A	169	L	F	F	F	F	F	F
4022	NS2A	172	T	A	A	A	A	A	A
4056	NS2A	183	S	F	F	F	F	F	F
4289	NS2B	37	I	L	L	L	L	L	L
4505	NS2B	109	I	L	L	L	L	L	L
5153	NS3	195	I	V	V	V	V	V	V
6023	NS3	485	D	N	N	N	N	N	N
6876	NS4A	146	V	A	A	A	A	A	A
7171	NS4B	95	I	M	M	M	M	M	M



**Table 4.5:** 17DD strain amino acid and 3'UTR substitutions using the newly sequenced strains, in reference to publicly available sequences.

Position	Gene	Codon	Amino Acid	SLII-75-1	Columbia	17DD YFU17066 <sup>b</sup>
854	M	36	L	F	F	F
1127	E	52	G	R	R	R
1436	E	155	D			S <sup>a</sup>
1437	E	155	D			S <sup>a</sup>
1482	E	170	A	V	V	V
1491	E	173	T	I	I	I
1572	E	200	K	T	T	T
1870	E	299	M	I	I	I
1887	E	305	S	F	F	F
1965	E	331	K	R	R	R
2112	E	380	T	R	R	R
2193	E	407	A	V	V	V
2220	E	416	A	V	V	V
2687	NS1	79	L	F	F	F
3371	NS1	307	I	V	V	V
3860	NS2A	118	M	V	V	V
4007	NS2A	167	T	A	A	A
4022	NS2A	172	T	A	A	A
4056	NS2A	183	S	F	F	F
4289	NS2B	37	I	L	L	L
4505	NS2B	109	I	L	L	L
5115	NS3	182	Q	R	R	R
6023	NS3	485	D	N	N	N
6876	NS4A	146	V	A	A	A
7171	NS4B	95	I	M	M	M
7580	NS4B	232	Y	H	H	H
8808	NS5	391	N	S	S	S
10142	NS5	836	E	K	K	K
10338	NS5	901	P	L	L	L
10367	3'UTR	-	T			C
10418	3'UTR	-	T	C	C	C
10800	3'UTR	-	G	A	A	A

<sup>a</sup> Substitutions arising from the same codon.

<sup>b</sup> Genbank accession for publicly available sequence.

**Table 4.6:** SE and RMSD estimations for the 17D strains considered in the study. RMSD is expressed as genetic distance from Asibi.

Strain	Normalized En-tropy				RMSD from Asibi			
	N	Mean	Median	s.d.	N	Mean	Median	s.d.
213/B2/81	10862	0.109	0.105	0.056	10850	0.012	0.005	0.066
Columbia	10862	0.107	0.102	0.059	10853	0.012	0.005	0.068
SLII-75-1	10862	0.108	0.104	0.058	10850	0.012	0.005	0.068
213/77	10862	0.107	0.102	0.059	10852	0.012	0.004	0.067
YF-	10862	0.107	0.103	0.058	10850	0.012	0.005	0.066
Vax/2012								
YFS/11	10862	0.109	0.105	0.057	10847	0.012	0.005	0.064
5114.15	10862	0.108	0.104	0.058	10853	0.012	0.005	0.065
Nigeria	10862	0.104	0.099	0.064	10850	0.013	0.005	0.065
NDC/6676	10862	0.103	0.098	0.064	10844	0.013	0.005	0.065
NDC/6677	10862	0.105	0.102	0.061	10850	0.013	0.005	0.065
YF-	10862	0.105	0.101	0.061	10846	0.011	0.004	0.066
Vax/1995								
Russia/SW	10862	0.102	0.097	0.062	10853	0.012	0.004	0.066
Substrain								
17D-213	32586	0.106	0.102	0.059	32555	0.012	0.004	0.066
17DD	21724	0.108	0.103	0.059	21703	0.012	0.005	0.068
17D-204	76034	0.106	0.102	0.060	75940	0.012	0.005	0.065

<sup>a</sup> ORF = Open reading frame.

**Table 4.7:** Wild-type revertant variants observed in the vaccine strains.

Position	Gene-Codon	Vaccine Amino Acid	Wild-Type (Revertant) Amino Acid	Revertant Percent	Strain Observed
854	M-36	F	L	0.2523	17D-204/USA/1995
				0.2688	17D-204/USA/NDC-6677
				0.629	17D-213/Nigeria
1140	E-56	V	A	0.3238	17D-204/USA/1995
				0.5028	17D-204/USA/NDC-6676
				0.2411	17D-204/USA/NDC-6677
				0.383	17D-213/Nigeria
				0.3468	17D-213/WHO/213-77
				0.6654	17D-213/WHO/213-B2-81
1431	E-153	T	N	1.418	17D-213/WHO/213-77
1436	E-155	S	G	0.4812	17DD/Colombia
1437	E-155	S	N	0.1495	17DD/Colombia
1482	E-170	V	A	0.2973	17D-204/USA/NDC-6677
				0.3044	17DD/Colombia
1887	E-305	F	S	4.968	17D-204/UK/YFS/11
1946	E-325	S	P	0.3126	17D-204/UK/YFS/11
				0.387	17D-204/USA/1995
				0.4051	17D-204/USA/2012
				0.4105	17D-213/Nigeria
				0.4537	17D-213/WHO/213-77
				0.3973	17D-213/WHO/213-B2-81
2193	E-407	V	A	0.8189	17D-204/USA/NDC-6676
2219	E-416	T	A	0.2426	17D-213/WHO/213-77
				0.1773	17D-213/WHO/213-B2-81
2687	NS1-79	F	L	0.1304	17D-204/USA/NDC-6677
3371	NS1-307	V	I	1.116	17D-213/WHO/213-B2-81
				1.14	17DD/Senegal/SLII-75-1
4013	NS2A-169	F	L	0.2675	17D-204/USA/1995
				0.3635	17D-204/USA/2012
				0.2417	17D-204/USA/NDC-6676
				0.2467	17D-213/Nigeria
				0.5885	17D-213/WHO/213-B2-81
4022	NS2A-172	A	T	0.5079	17D-213/WHO/213-B2-81
				0.7709	17DD/Senegal/SLII-75-1
4056	NS2A-183	F	S	0.6359	17DD/Senegal/SLII-75-1
4505	NS2B-109	L	I	0.4611	17D-213/WHO/213-B2-81
				0.4884	17DD/Senegal/SLII-75-1
5115	NS3-182	R	Q	0.2883	17DD/Colombia

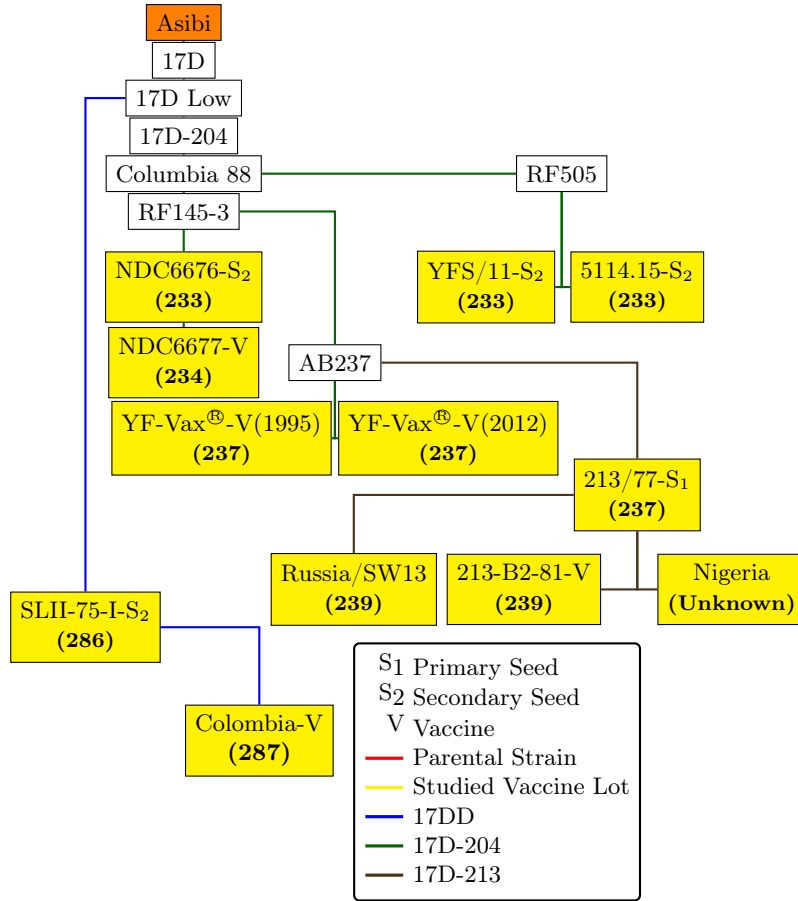


				1.279	17DD/Senegal/SLII-75-1
5153	NS3-195	V	I	9.748	17D-204/South Africa/5114.15
				1.244	17D-204/USA/1995
				0.7215	17D-204/USA/2012
6023	NS3-485	N	D	0.2114	17D-204/USA/2012
				0.8056	17D-204/USA/NDC-6676
				0.2517	17D-213/WHO/213-77
				0.2958	17D-213/WHO/213-B2-81
				0.3898	17DD/Senegal/SLII-75-1
6876	NS4A-146	A	V	0.888	17D-204/UK/YFS/11
7171	NS4B-95	M	I	0.3887	17D-204/USA/NDC-6677
10142	NS5-836	K	E	0.5452	17D-213/WHO/213-B2-81
				1.637	17DD/Senegal/SLII-75-1
10338	NS5-901	L	P	0.1849	17D-213/WHO/213-77
				0.6494	17D-213/WHO/213-B2-81
				1.047	17DD/Senegal/SLII-75-1

**Table 4.8:**  $dN/dS$  summary for all strains considered in the study, considering the YFV open reading frame and each gene segment individually.

Gene	Codons	Mean	Median	SD
Capsid	121	0.553	0.548	0.117
prM	164	0.549	0.588	0.162
Envelope	493	0.533	0.519	0.200
NS1	352	0.519	0.565	0.135
NS2A	224	0.506	0.550	0.167
NS2B	130	0.557	0.637	0.277
NS3	623	0.498	0.489	0.123
NS4A	149	0.565	0.603	0.261
NS4B	250	0.574	0.626	0.114
NS5	905	0.478	0.476	0.116
ORF <sup>a</sup>	3412	0.483	0.500	0.105

<sup>a</sup> ORF = Open reading frame.



**Figure 4.1:** Tree of seeds and vaccines used in the study, organized by substrain and passage level relative to the parental strain Asibi. Passage levels are listed numerically, with the seed identity of the strain (Primary, Secondary, or Vaccine) marked, if known. Substrain lineage of each strain (17D-204, 17DD, 17D-213) is marked by branching color.

	325
	326
	327
	328
	329
	330
	331
	332
	333
	334
	335
Asibi - AY640589	PKGAPCKIPVI
Asibi*	PKGAPCKIPVI
Asibi NGS Beck - KF769016	PKGAPCKIPVI
French Viscerotropic - YFU21056	PKGAPCRIPVI
85-82 Ivory Coast- YFU54798	PKGAPCKIPVI
Ivory Coast 1999 - AY603338	PKGAPCKIPVI
Angola 1971 - AY968064	PKGAPCKIPVI
Gambia 2001 - AY572535	PKGAPCKIPVI
Uganda 1948 - AY968065	PKGAPCKIPVI
Ethiopia 1961 - DQ235229	PKGAPCKIPVI
Uganda 2010 - JN620362	PKGAPCKIPVI
Brazil 1980 - JF912179	PKGAPCRIPVM
Brazil 1981 - JF912180	PKGAPCRIPVM
Brazil 1983 - JF912181	PKGAPCRIPVM
Brazil 1984 - JF912182	PKGAPCRIPVM
Brazil 1984 - JF912183	PKGAPCRIPVM
Brazil 1987 - JF912184	PKGAPCRIPVM
Brazil 1992 - JF912185	PKGAPCRIPVM
Brazil 1994 - JF912186	PKGAPCRIPVM
Brazil 2000 - JF912187	PKGAPCRIPVM
Brazil 2000 - JF912188	PKGAPCRIPVM
Brazil 2001 - JF912189	PKGAPCRIPVM
Brazil 2002 - JF912190	PKGAPCRIPVM
17D-204 YF-Vax NGS - KF769015	SKGAPCRIPVI
17D-204 ATCC - X03700	SKGAPCRIPVI
17D-204 Stamaril - X15062	SKGAPCRIPVI
17D-204 YF-Vax 1995*	SKGAPCRIPVI
17D-204 YF-Vax 2012*	SKGAPCRIPVI
17D-204 UK S2 YFS/11*	SKGAPCRIPVI
17D-204 SA S2 5114.15*	SKGAPCRIPVI
17D-204 NDC/6676 S2*	PKGAPCRIPVI
17D-204 NDC/6677 Vaccine*	PKGAPCRIPVI
17D-213 - YFU17067	SKGAPCRIPVI
17D-213/77 WHO S1*	SKGAPCRIPVI
17D-213 Russia/SW13*	SKGAPCRIPVI
17D-213 Vaccine Lot 213-B2-81*	SKGAPCRIPVI
17D-213 Nigeria*	SKGAPCRIPVI
17DD Brazil - YFU17066	PKGAPCRIPVI
17DD Brazil- DQ100292	PKGAPCRIPVI
17DD Dakar S2 SLII-71-1*	PKGAPCRIPVI
17DD Columbia*	PKGAPCRIPVI

**Figure 4.2:** Alignment for newly sequenced and publicly available sequences showing vaccine divergence at E331. Residues numbers (above) are listed from the N-terminus of the YFV envelope protein. \*=Sequenced in this study.

	23747576777879808182
Asibi - AY640589	ADEINAILFEEN
Asibi*	ADEINAILFEEN
Asibi NGS Beck - KF769016	ADEINAILFEEN
French Viscerotropic - YFU21056	ADEINAILFEEN
85-82 Ivory Coast- YFU54798	ADEINAILFEEN
Ivory Coast 1999 - AY603338	ADEINAILFEEN
Angola 1971 - AY968064	ADEINAIFFEEN
Gambia 2001 - AY572535	ADEINAILFEEN
Uganda 1948 - AY968065	ADEINAIFFEEN
Ethiopia 1961 - DQ235229	ADEINAIFFEEN
Uganda 2010 - JN620362	ADEINAIFFEEN
Brazil 1980 - JF912179	ADEINAILFEEN
Brazil 1981 - JF912180	ADEINAILFEEN
Brazil 1983 - JF912181	ADEINAILFEEN
Brazil 1984 - JF912182	ADEINAILFEEN
Brazil 1984 - JF912183	ADEINAILFEEN
Brazil 1987 - JF912184	ADEINAILFEEN
Brazil 1992 - JF912185	ADEINAILFEEN
Brazil 1994 - JF912186	ADEINAILFEEN
Brazil 2000 - JF912187	ADEINAILFEEN
Brazil 2000 - JF912188	ADEINAILFEEN
Brazil 2001 - JF912189	ADEINAILFEEN
Brazil 2002 - JF912190	ADEINAILFEEN
17D-204 YF-Vax NGS - KF769015	ADEINAIFFEEN
17D-204 ATCC - X03700	ADEINAIFFEEN
17D-204 Stamaril - X15062	ADEINAIFFEEN
17D-204 YF-Vax 1995*	ADEINAIFFEEN
17D-204 YF-Vax 2012*	ADEINAIFFEEN
17D-204 UK S2 YFS/11*	ADEINAIFFEEN
17D-204 SA S2 5114.15*	ADEINAIFFEEN
17D-204 NDC/6676 S2*	ADEINAIFFEEN
17D-204 NDC/6677 Vaccine*	ADEINAIFFEEN
17D-213 - YFU17067	ADEINAIFFEEN
17D-213/77 WHO S1*	ADEINAIFFEEN
17D-213 Russia/SW13*	ADEINAIFFEEN
17D-213 Vaccine Lot 213-B2-81*	ADEINAIFFEEN
17D-213 Nigeria*	ADEINAIFFEEN
17DD Brazil - YFU17066	ADEINAIFFEEN
17DD Brazil- DQ100292	ADEINAIFFEEN
17DD Dakar S2 SLII-71-1*	ADEINAIFFEEN
17DD Columbia*	ADEINAIFFEEN

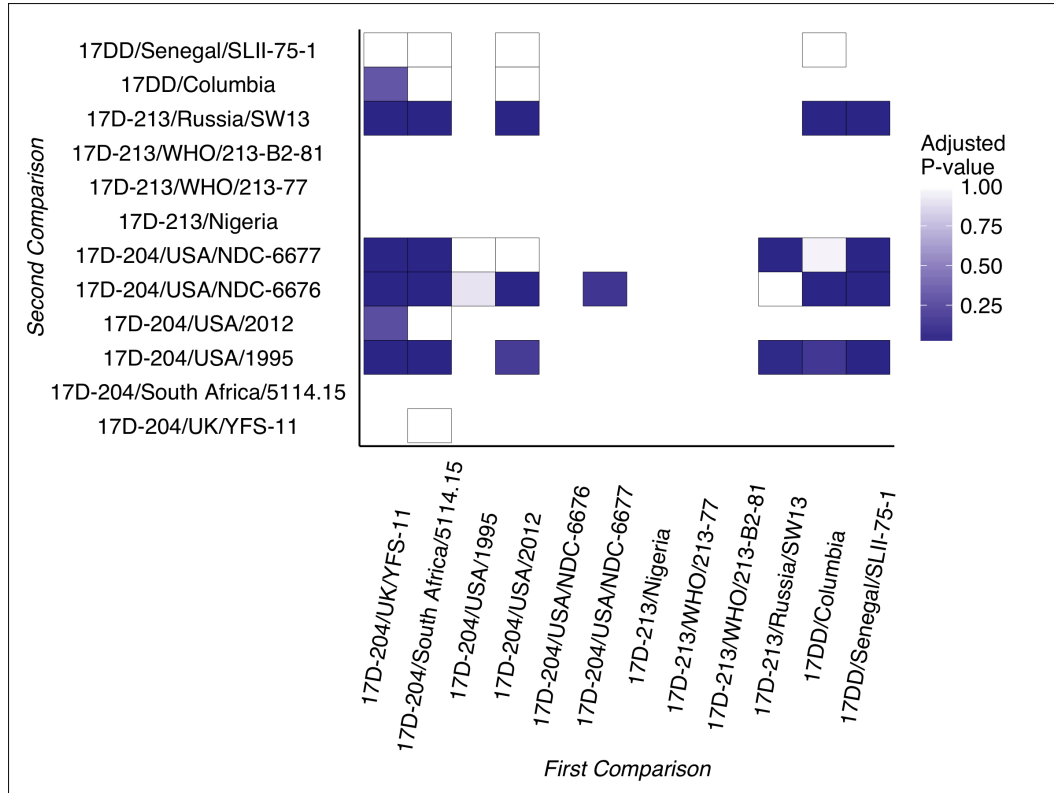
**Figure 4.3:** Alignment for newly sequenced and publicly available sequences showing vaccine divergence at NS179. Residues numbers (above) are listed from the N-terminus of the YFV NS1 protein. \*=Sequenced in this study.

	3132333435363738394041
Asibi - AY640589	PIAVGGILMML
Asibi*	PIAVGGILMML
Asibi NGS Beck - KF769016	PIAVGGILMML
French Viscerotropic - YFU21056	PIAVGGILMML
85-82 Ivory Coast- YFU54798	PIAVGGILMML
Ivory Coast 1999 - AY603338	PIAVGGILMML
Angola 1971 - AY968064	PIAVGGILMML
Gambia 2001 - AY572535	PIAVGGILMML
Uganda 1948 - AY968065	PIAVGGILMML
Ethiopia 1961 - DQ235229	PIAVGGILMML
Uganda 2010 - JN620362	PIAVGGILMML
Brazil 1980 - JF912179	PVAVGGILMML
Brazil 1981 - JF912180	PVAVGGILMML
Brazil 1983 - JF912181	PVAVGGILMML
Brazil 1984 - JF912182	PVAVGGILMML
Brazil 1984 - JF912183	PVAVGGILMML
Brazil 1987 - JF912184	PVAVGGILMML
Brazil 1992 - JF912185	PVAVGGILMML
Brazil 1994 - JF912186	PVAVGGILMML
Brazil 2000 - JF912187	PVAVGGILMML
Brazil 2000 - JF912188	PVAVGGILMML
Brazil 2001 - JF912189	PVAVGGILMML
Brazil 2002 - JF912190	PVAVGGILMML
17D-204 YF-Vax NGS - KF769015	PIAVGGILMML
17D-204 ATCC - X03700	PIAVGGILMML
17D-204 Stamaril - X15062	PIAVGGILMML
17D-204 YF-Vax 1995*	PIAVGGILMML
17D-204 YF-Vax 2012*	PIAVGGILMML
17D-204 UK S2 YFS/11*	PIAVGGILMML
17D-204 SA S2 5114.15*	PIAVGGILMML
17D-204 NDC/6676 S2*	PIAVGGILMML
17D-204 NDC/6677 Vaccine*	PIAVGGILMML
17D-213 - YFU17067	PIAVGGILMML
17D-213/77 WHO S1*	PIAVGGILMML
17D-213 Russia/SW13*	PIAVGGILMML
17D-213 Vaccine Lot 213-B2-81*	PIAVGGILMML
17D-213 Nigeria*	PIAVGGILMML
17DD Brazil - YFU17066	PIAVGGILMML
17DD Brazil- DQ100292	PIAVGGILMML
17DD Dakar S2 SLII-71-1*	PIAVGGILMML
17DD Columbia*	PIAVGGILMML

**Figure 4.4:** Alignment for newly sequenced and publicly available sequences showing vaccine divergence at NS2B37. Residues numbers (above) are listed from the N-terminus of the YFV NS2B protein. \*=Sequenced in this study.

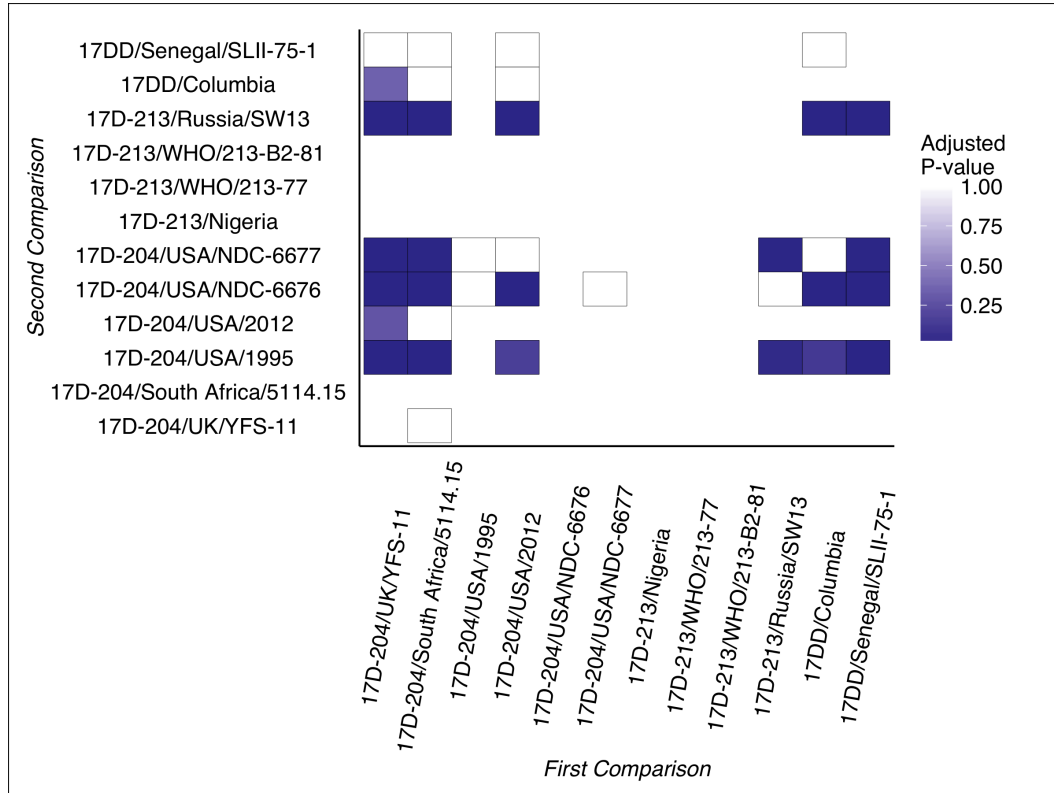
	88 229 230 231 232 233 234 235 236 237 238
Asibi - AY640589	MRGNYYAFVGV
Asibi*	MRGNYYAFVGV
Asibi NGS Beck - KF769016	MRGNYYAFVGV
French Viscerotropic - YFU21056	MRGNYYAFVGV
85-82 Ivory Coast- YFU54798	MRGNYYAFVGV
Ivory Coast 1999 - AY603338	MRGNYYAFVGV
Angola 1971 - AY968064	MRGNYYAFVGV
Gambia 2001 - AY572535	MRGNYYAFVGV
Uganda 1948 - AY968065	MRGNYYAFVGV
Ethiopia 1961 - DQ235229	MRGNYYAFVGV
Uganda 2010 - JN620362	MRGNYYAFVGV
Brazil 1980 - JF912179	MRGNYYAFVGV
Brazil 1981 - JF912180	MRGNYYAFVGV
Brazil 1983 - JF912181	MRGNYYAFVGV
Brazil 1984 - JF912182	MRGNYYAFVGV
Brazil 1984 - JF912183	MRGNYYAFVGV
Brazil 1987 - JF912184	MRGNYYAFVGV
Brazil 1992 - JF912185	MRGNYYAFVGV
Brazil 1994 - JF912186	MRGNYYAFVGV
Brazil 2000 - JF912187	MRGNYYAFVGV
Brazil 2000 - JF912188	MRGNYYAFVGV
Brazil 2001 - JF912189	MRGNYYAFVGV
Brazil 2002 - JF912190	MRGNYYAFVGV
17D-204 YF-Vax NGS - KF769015	MRGNHYAFVGV
17D-204 ATCC - X03700	MRGNHYAFVGV
17D-204 Stamaril - X15062	MRGNHYAFVGV
17D-204 YF-Vax 1995*	MRGNHYAFVGV
17D-204 YF-Vax 2012*	MRGNHYAFVGV
17D-204 UK S2 YFS/11*	MRGNHYAFVGV
17D-204 SA S2 5114.15*	MRGNHYAFVGV
17D-204 NDC/6676 S2*	MRGNHYAFVGV
17D-204 NDC/6677 Vaccine*	MRGNHYAFVGV
17D-213 - YFU17067	MRGNHYAFVGV
17D-213/77 WHO S1*	MRGNHYAFVGV
17D-213 Russia/SW13*	MRGNHYAFVGV
17D-213 Vaccine Lot 213-B2-81*	MRGNHYAFVGV
17D-213 Nigeria*	MRGNHYAFVGV
17DD Brazil - YFU17066	MRGNHYAFVGV
17DD Brazil- DQ100292	MRGNHYAFVGV
17DD Dakar S2 SLII-71-1*	MRGNHYAFVGV
17DD Columbia*	MRGNHYAFVGV

**Figure 4.5:** Alignment for newly sequenced and publicly available sequences showing vaccine divergence at NS4B232. Residues numbers (above) are listed from the N-terminus of the YFV NS4B protein. \*=Sequenced in this study.\*=Sequenced in this study.

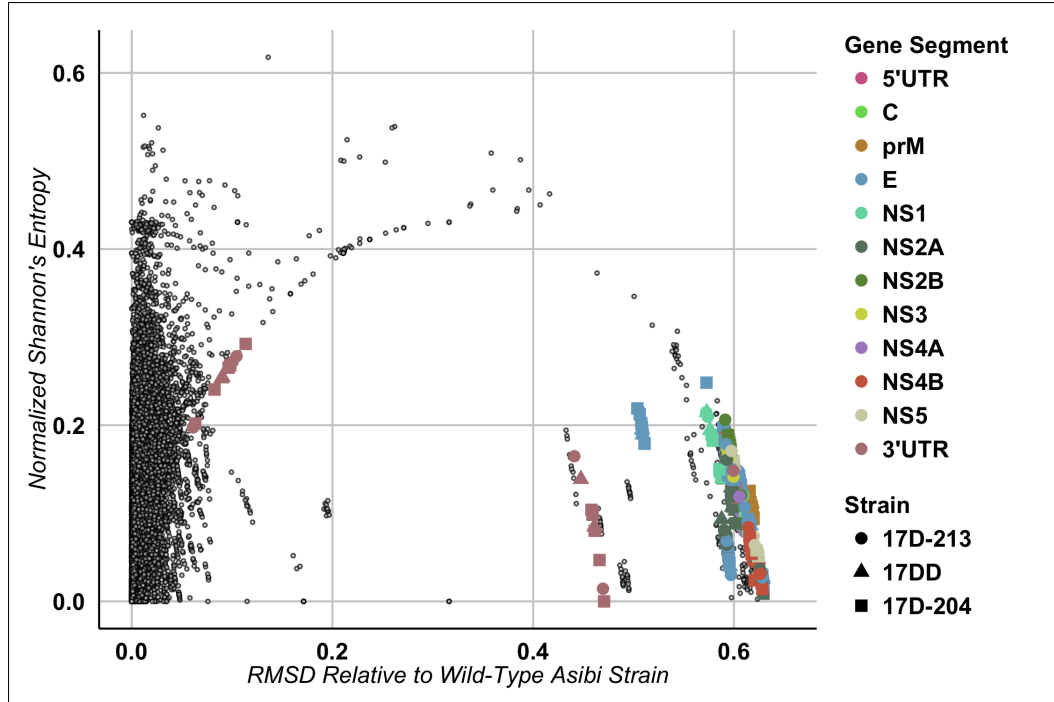


**Figure 4.6:** Heat map of Kruskal-Wallis p-values for all pairwise comparisons of NSE, considering values along the whole genome extent of each strain. Dunn's post-hoc correction was used to adjust for multiple tests.

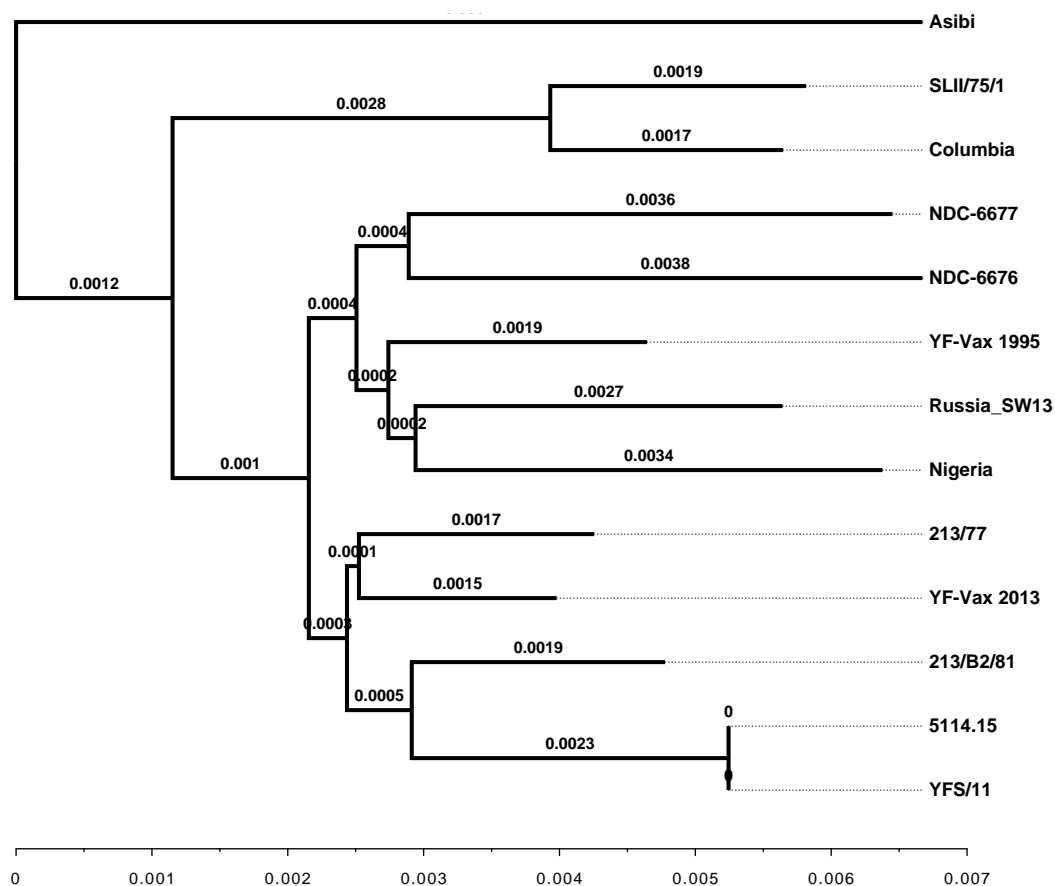




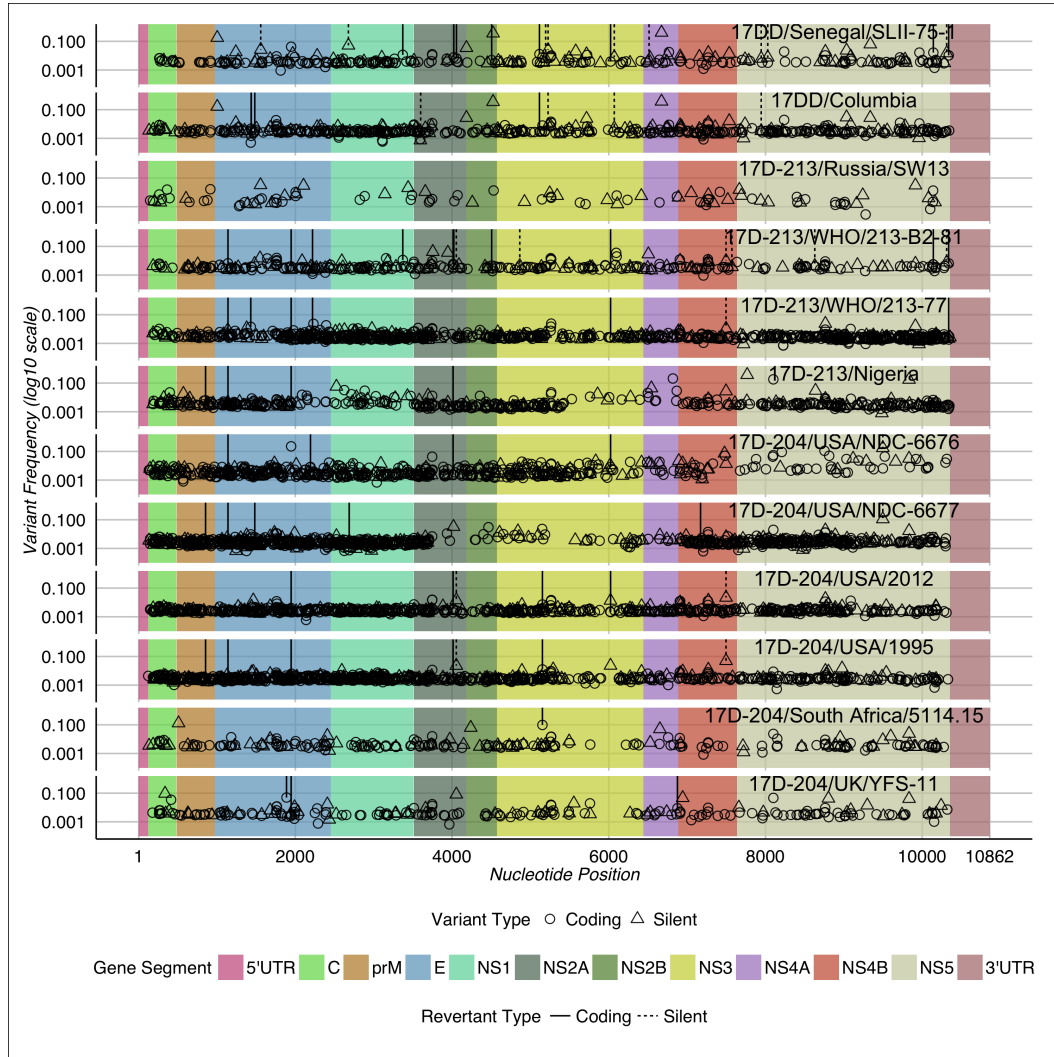
**Figure 4.7:** Heat map of Kruskal-Wallis p-values for all pairwise comparisons of NSE, considering values only along the open reading frame of each strain. Dunn's post-hoc correction was used to adjust for multiple tests.



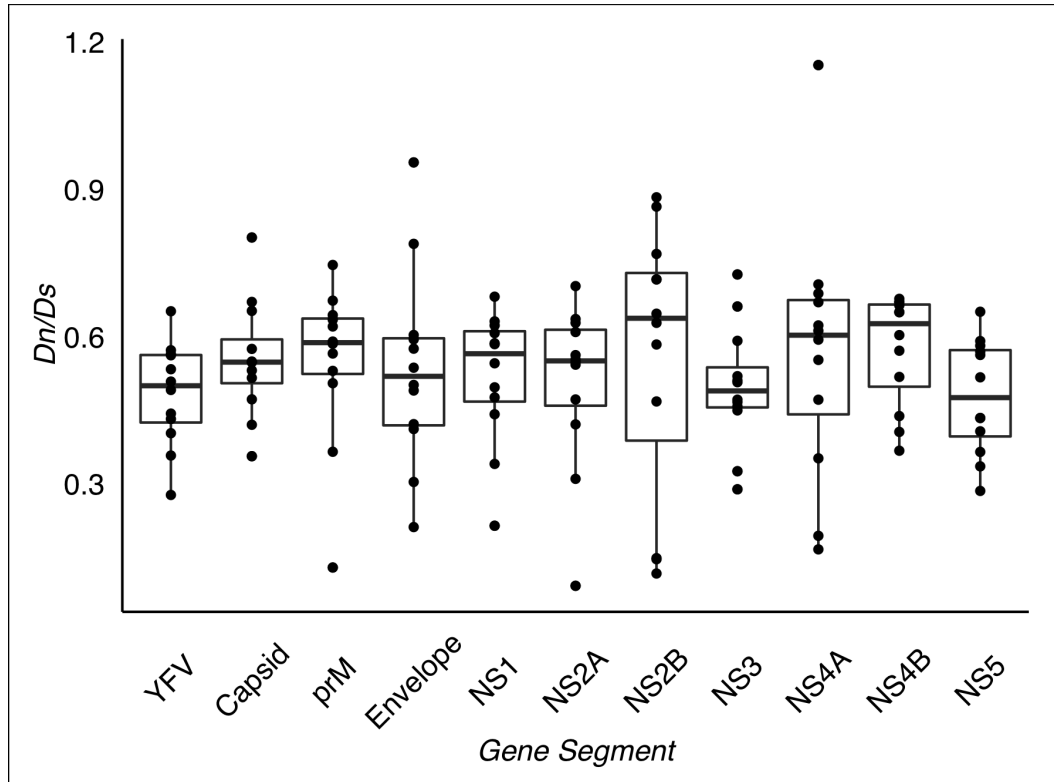
**Figure 4.8:** Scatter plot of NSE versus RMSD relative to Asibi for all nucleotide sites recovered in each strain. Colored glyphs are contained in the WHO standard genotype for YFV, and specifically indicate that at the nucleotide positions described for the vaccine genotype, the nucleotide subpopulations of those sites are equivalently selected from the parental genotype for all vaccine strains.



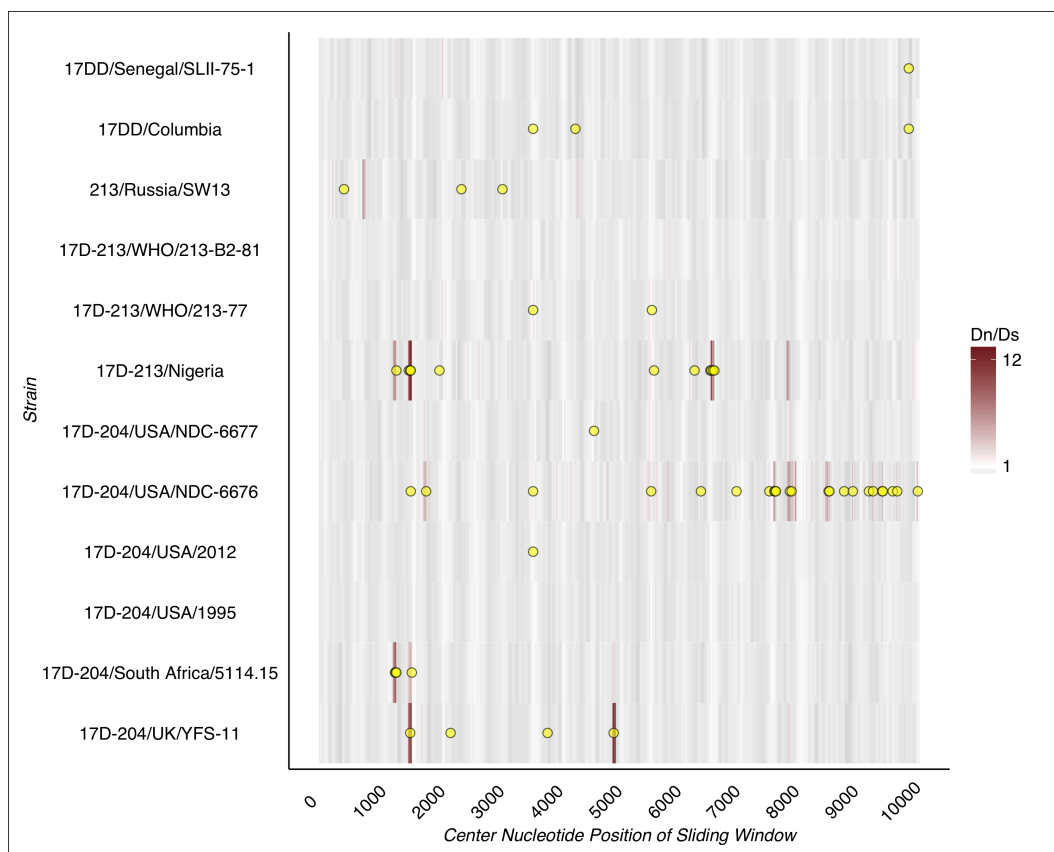
**Figure 4.9:** Neighbor-joining phylogeny calculated by pairwise mean RMSD for all seeds considered in the study. A pairwise matrix of RMSD was constructed and used as the distance measurement from which the branch lengths were derived.



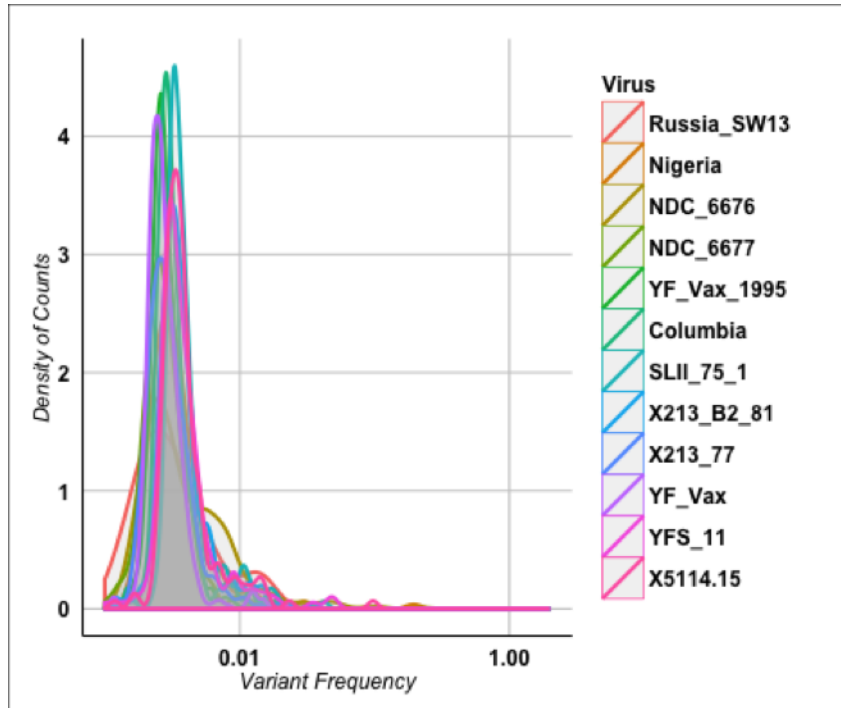
**Figure 4.10:** Modeled variants for the strains, with coding/silent classifications and wild-type revertants indicated. SNVs are plotted as frequency against nucleotide position in the read alignment. Coding and silent variants were classified against the consensus sequences of each respective strain, and are differentiated by glyph shape. Vertical lines indicate SNVs that represent subpopulation revertants to the wild-type genotype, and are likewise classified as coding or silent.



**Figure 4.11:** Boxplots of summarized  $dN/dS$  values along each gene segment. All box plots depict the median, first quartile, third quartile, with whiskers showing  $\pm 1.5$  times the interquartile distance.



**Figure 4.12:** Heatplot of  $dN/dS$  values for all strains considered. Color scale indicates increasing  $dN/dS$  for 20-codon windows, generated with a sliding window step size of single-codon distance.



**Figure 4.13:** Density plot for counts of variants, derived from modeled SNVs in the population. The curves show a typical pattern for low-diversity viral populations, which is that of a large population of small-frequency variants (the high density area at left in the plot), with a low population count of higher-frequency variants in the tail of the distribution. Curves overlap considerably, visualizing that all of the vaccine strains analyzed exhibit this type of similarity.

## Chapter 5

# Deep Sequencing of French Neurotropic Yellow Fever Vaccine Strains Reveals A Conserved Genotype and Dissimilarity of Population Structure

### 5.1 Abstract

The live-attenuated FNV vaccine strain of YFV was derived empirically in 1930 by passage of WT FVV in mouse brain. The vaccine was administered extensively in French-speaking Africa until discontinuance in 1980, due to high rates of post-vaccination encephalitis in children. The first deep sequencing study of FNV was performed to explore competing hypotheses previously reported to explain the safety profile of the vaccine, using four vaccine strains from different institutional collections. Diversity and specific population structures were compared in reference to the WT parental FVV strain, and between the vaccines themselves. Lower abundance of polymorphism content was observed for FNV strains to each other relative to the parental strain. Although the vaccines were of lower diversity than FVV, heterogeneity between the vaccines was observed. Presence of WT identity was variable and included revertants to WT identity, which was inversely associated with neurovirulence phenotypes previously described *in vivo* for the vaccine strains. The analysis is the first by deep sequencing of an adverse-reactive vaccine, providing novel sequence information and evidence that adaptation of FNV to mouse brain was a contributor to the neurotropic adverse phenotype.



## 5.2 Introduction

Phenotypic adaptation of viruses arising from serial culture and animals is ubiquitous, and constitutes the fundamental property of empirical derivation of live-attenuated vaccines. Of particular significance to the study of live-attenuated vaccines are those administered to prevent infection by YFV, which are among the oldest live-attenuated vaccines. YFV was a considerable obstruction to economic development until the empiric derivation of two live-attenuated vaccine strains in the early twentieth century. The vaccines arose from separately isolated parental strains, and distinct attenuating subculture methods. The vaccine strain 17D was developed by passage of WT strain Asibi in mouse and chicken tissue in 1936 by the Rockefeller Foundation (New York, USA); derivatives of this strain remain in production today and are administered by the subcutaneous route (Theiler and H. H. Smith 1937).

Concurrent efforts by the Institut Pasteur (Dakar, Senegal) produced the FNV, which was administered extensively in French-speaking Africa starting in 1940, more than ten years after the initial adaptation of the virus. FNV was attenuated from the WT FVV, isolated in 1927 from the Lebanese patient Françoise Mayali, by 128 passages in mouse brain, yielding a strain that was attenuated for both vicerotropism and vector competence; however, neurotropic properties were enhanced (W. Lloyd and Penna 1933; Davis, W. Lloyd, and Frobisher 1932; Durieux 1956b). FNV was administered by scarification from a reconstitution of infected and desiccated mouse brain, with desirable properties of thermostability and high immunogenicity (Stuart 1953; Durieux 1956b). Commercial lots were administered between passages 250 through 260, relative to the parental strain (Durieux 1956b). Deployment of the FNV is credited as a key intervention in the suppression of YF disease in French West Africa between 1940 and 1954, with over 84 million doses produced during that time (Durieux 1956a; Durieux 1956b). FNV was noted for unacceptable

rates of post-vaccinal encephalitis in children (3:1000), constituting a rationale for restricting administration of the vaccine to those above 10 years of age in 1961, and later discontinuance of the vaccine in 1980 due to the availability of the safer 17D strain (which is routinely given to those aged 9 months and older) (Vainio and Cutts 1998).

Mechanisms of attenuation for both YFV vaccines are poorly understood, however information on FNV is particularly obscure. YFV strains are variably neurotropic when directly introduced to central nervous tissues of mice, a property that is exploited in lethality and serological protection assays (Theiler 1930; Sawyer and W. Lloyd 1931; Fitzgeorge and Bradish 1980; A. D. T. Barrett and Gould 1986). Rhesus and cynomolgous macaques are susceptible to neurotropic disease when challenged by the i.c. route with vaccine strains, of which FNV is observed to cause death in monkeys within 5 days; WT strains however are not neurotropic and cause viscerotropic disease even when directly introduced in to the brain. In contrast to FNV, the widely administered 17D strain vaccines are mildly virulent to monkeys when administered i.c., causing a transient fever, some evidence of lesions in the brain and survivable encephalitis. Currently, loss of neurovirulence for monkeys is a safety criterion for YFV vaccine seeds as required by WHO technical specifications; 17D strain derivatives are the only YFV vaccine lineage in current production (World Health Organization 2010).

Two alternative hypotheses have been reported to explain the neurotropic adverse phenotype of FNV. First, it is possible that the attenuating passage of FVV in mouse brain selected for enhanced capacity of the virus to invade or to productively infect the human central nervous system. FNV was reported to be especially susceptible to binding inhibition by treatment with monkey brain homogenates when compared to other prototypical YFV strains FVV (WT parent of FNV), Asibi (WT

parent of 17D vaccine), and 17D-204 (derived by passage of Asibi strain without nervous tissue) (Ni, Ryman, et al. 2000). Second, it is possible that FNV is genetically unstable in the host, leading to WT reversion events. Instability of the FNV genome was reported after a short cell culture passage series, with concomitant alterations to mouse neurotropism, appearance of putative WT genotypes, and greater clonal sequence variation than a 17D strain vaccine strain used as comparator (Holbrook et al. 2000). FNV genome instability under laboratory conditions is unique amongst studied YFV strains. WT YFV sequences were compared to those of dengue virus by phylogenetic methods, observing relatively low mutation rates for YFV (Sall et al. 2010). The 17D vaccine strain was reported by multiple studies to incorporate very few mutations over time under cell culture passage, using Sanger methods (H. Xie, Cass, and A. D. Barrett 1998; Pugachev, Guirakhoo, et al. 2003).

Investigation of the FNV adverse phenotype is frustrated by the rarity of reference strains that exist in very few institutional collections; passage histories are not generally accessible for any collection strain. Only one example of FNV has been fully sequenced in reference to the WT parent FVV, and 77 nucleotide changes were observed, with 33 encoding amino acid substitutions (E. Wang et al. 1995). Although this study reported one full sequence of the vaccine, neurovirulence was compared between four strains of FNV obtained from institutional collections; the vaccines were of unequal neurovirulence in both mice and monkeys [Table 5.1] (E. Wang et al. 1995). Specifically, the four examples of FNV were obtained from collections at Porton Down, UK (FNV-NT), Yale University (FNV-Yale), Institut Pasteur, Paris, France (FNV-IP) and Centers for Disease Control and Prevention (CDC), Fort Collins, CO (FNV-FC). FNV-IP and FNV-FC were lethal for cynomolgus monkeys following i.c. inoculation while FNV-NT was not; FNV-Yale was not tested. FNV-NT, FNV-Yale and FNV-IP were all lethal following intranasal (i.n.) inoculation of

weanling mice and had short average survival times (ASTs) ( $< 6$  days) following IC inoculation of weanling mice while FNV-FC was not lethal by the i.n. route and had a long AST following i.c. inoculation ( $> 11$  days). Although it cannot be excluded that the differences in neurovirulence between the strains may have been the result of passaging in individual laboratories, these rare strains of FNV represent a context by which to investigate continuous divergence of the central adverse-reactive phenotype that led to discontinuance of the vaccine.

The reconstruction of population structure for live-attenuated vaccines offers considerable opportunity to exploit viral diversity, an element of quasispecies theory, as a means to understand attenuation and relative stability of viral phenotypes that arose by empiric means. To this end, a deep sequencing approach was previously used to analyze population structure of the commercial YFV vaccine strain 17D-204, showing that the vaccine was of lower diversity than the WT parental strain Asibi (Beck et al., 2014). From these results, it was postulated that low diversity may contribute to the considerable safety record of 17D vaccine derivatives. For FNV reference strains considered in the study, this pattern is recapitulated *in silico* in reference to the 17D strain that is highly stable. Heterogeneity was however observed between FNV strains themselves. Furthermore, increase of heterogeneity was associated with previously described loss of neurovirulence *in vivo* for the strains considered. The findings demonstrate that adverse reactions to FNV were likely caused by the attenuating passage in nervous tissue, as opposed to reversion in the host.

## 5.3 Results

### 5.3.1 FNV Strain Coverages

For FVV, median coverage was 8338.0 (mean=8409.0, s.d.=1660.3). For FNV-IP, median coverage was 8470.5 (mean=8573.5, s.d.=2912.6). For FNV-Yale, median coverage was 8848.0 (mean=8667.7, s.d.=2111.6). For FNV-FC, median coverage was 9865.0 (mean = 8527.8, s.d.= 3771.8). For FNV-NT, median coverage was 8235.0 (mean=8008.3, s.d.=1965.7). For Asibi, median coverage was 7513.0 (mean=7483.4, s.d.=1647.7). For 17D-204, median coverage was 8392.0 (mean=8113.4, s.d.=1298.4)[Figure 5.1].

### 5.3.2 Consensus Sequences

#### 5.3.2.1 In Reference to Wild-Type Strain FVV

For FNV-IP, 60 nucleotide differences were observed compared to WT parent FVV, of which 27 were amino acid substitutions. For FNV-Yale, 63 nucleotide differences were observed, of which 28 were amino acid substitutions. For FNV-FC, 60 nucleotide differences were observed, of which 27 were amino acid substitutions. For FNV-NT, 122 nucleotide differences were observed, of which 49 were amino acid substitutions. Overall, 19 amino acid substitutions (C-V80A, C-S90G, premembrane (prM)-V25M, M-L36F, E-A54V, E-T200K, E-N249K, E-K331R, NS2A-I82V, NS2A-T105A, NS2B-S60A, NS3-M192K, NS4A-F29L, NS4A-Y135F, NS4B-I95M, NS4B-V98I, NS5-I285T, NS5-R335K, NS5-F531L) are conserved between all four FNV strains considered when compared to FVV [Table 5.2]. In addition, seven amino acid substitutions (E-T7A, E-N153K, E-M457I, NS3-I13V, NS4B-L19S, NS4B-A165V, NS5-R2T) are only conserved between FNV-IP, FNV-Yale, and FNV-FC, but are not observed in FNV-NT, i.e., three of the four FNV strains shared 26 amino acid substitutions. A total of 15 possible substitutions were observed in the 3' untrans-

lated regions of the FNV strains, however none were conserved [Table 5.2]. In a parallel comparison, WT strains FVV and Asibi differ by 11 nucleotide differences encoding 2 amino acid substitutions located at FVV nucleotide positions 1572 (E-K200T) and 5408 (NS3-V279I), and 1 nucleotide difference in the 3'UTR at position 10800 (A->G). Sequence comparisons including silent mutations are shown in appendix table 1.

### **5.3.2.2 In Reference to The Originating FNV-Dakar Vaccine Strain**

The originating FNV-Dakar strain was sequenced at low coverage (<1000 reads/nucleotide site). When compared to the newly sequenced FVV strain, 169 nucleotide substitutions were observed, of which 54 were amino acid substitutions [Table 5.2]. Of the 25 amino acid substitutions conserved at a percentage of at least 75 percent in the FNV-IP, FNV-Yale and FNV-FC strains sequenced at deep coverage, 4 were not observed in the FNV-Dakar strain, NS4A-F29L and NS4B-V98I that were members of the core 19 substitutions shared by the four FNVs and NS3-I13V and NS4B-L19S that were members of the seven substitutions shared by three FNVs (Yale, FC and IP). No phenotypic studies have been performed on FNV-Dakar.

### **5.3.2.3 In Reference to Previously Sequenced FVV and FNV-IP Strains**

The sequence of FNV-IP was previously reported by Wang et al., and compared to FVV; 75 nucleotide differences were reported, of which 33 coded for non-synonymous amino acid substitutions (E. Wang et al. 1995). By contrast, the comparison of FVV and FNV-IP strains in this study revealed 68 nucleotide differences, of which 27 coded for amino acid substitutions. Six coding substitutions reported by Wang et. al. (E-Q142R, E-G227E, NS1-P228L, NS3-S25Y, NS4B-E246K) were

not observed in in the current FNV-IP sequence.

It was hypothesized that the balance of these differences between the earlier FNV-IP genotype and the sequence generated in this study were the result of differences in the sequence of the parental strains analyzed. A comparison of the previously reported FVV (Genbank accession number YFU21056) sequence to that of the newly sequenced FVV strain revealed 17 nucleotide differences, of which 7 (*in silico* FVV genotype in bold: E-**Q**142R, E-**K**200T, E-**G**227E, E-**R**331K, NS1-**P**228L, NS3-**S**25Y, NS4B-**E**246K) coded for nonsynonymous amino acid substitutions. Thus, differences in the current and previously reported FNV-IP genotypes are attributed to differences in the parental reference sequence used.

### 5.3.3 Diversity of Strains and Stability of Vaccine Strain Genotypes

#### 5.3.3.1 Summary Measurements of Diversity

For FVV, full-genome median NSE was 0.131 (mean=0.137, s.d.=0.059). For FNV-IP, full-genome median NS was 0.010 (mean=0.104, s.d.=0.061). For FNV-Yale, full-genome median NSE was 0.095 (mean=0.102, s.d.=0.066). For FNV-FC, full-genome median NSE was 0.095 (mean=0.102, s.d.=0.066). For FNV-NT, full-genome median NSE was 0.010 (mean=0.104, s.d.=0.061) [Figure 5.2]. For Asibi, full-genome median NSE was 0.098 (mean=0.104, s.d.=0.064). For 17D-204, full-genome median NSE was 0.102 (mean=0.106, s.d.=0.059). A single-factor Kruskal-Wallis test comparing open reading frame NSE between the four FNV strains, FVV and 17D-204 was statistically significant ( $X^2=2773.51$ ,  $df=5$ ,  $p<.001$ ). Using error rate, quantile-quantile plots and bootstrapped Kolmogorov-Smirnov (KS) tests reveal significant divergence between all possible pairs of strains, except between pairs FNV-IP/FNV-NT and FNV-Yale/FNV-FC [Figure 5.3].

### **5.3.3.2 Measurements of Diversity and Divergence for Individual Nucleotide Positions**

Sitewise genetic distance from FVV was measured by RMSD, and was not equivalent for the four FNV vaccine strains considered, demonstrating that the putative conserved genotype differentiating FVV from FNV is not stably selected in some strains [Figure 5.4]. Considering the 26 nucleotide positions containing amino acid substitutions between FVV and the vaccine strains individually, low diversity was observed for all four FNV relative to FVV, and greater diversity was associated with FNV-NT, which has previously described to be of low monkey neurovirulence [Table 5.3, Figure 5.5].

### **5.3.4 Single Nucleotide Variant Structure**

For FVV, 667 variant sites were observed, of which 414 encode amino acid substitutions. For FNV-IP, 182 variant sites were observed, of which 91 encode amino acid substitutions [Figure 5.6]. For FNV-Yale, 194 variant sites were observed, of which 105 encode amino acid substitutions. For FNV-FC, 182 variant sites were observed, of which 57 encode amino acid substitutions. For FNV-NT, 396 variant sites were observed, of which 234 encode amino acid substitutions. Concordance of sites with variant content was observed between strains [Figure 5.7 and Figure 5.8]. For Asibi, 321 variant sites were observed, of which 167 were amino acid substitutions. For 17D-204, 969 variant sites were observed, of which 618 were amino acid substitutions.

### **5.3.5 Reversion of Amino Acid Residues from Vaccine to Wild-Type**

For FNV-IP, 22 WT revertants were observed, of which 9 were amino acid substitutions. For FNV-Yale, 23 WT revertants were observed, of which 8 were



amino acid substitutions. For FNV-FC, 53 WT revertants were observed, of which 12 were amino acid substitutions. For FNV-NT, 84 WT revertants sites were observed, of which 29 were amino acid substitutions. For 17D-204, 22 WT revertants were observed, of which 14 were amino acid substitutions. [Table 5.4 and Table 5.5]

## **5.4 Discussion**

### **5.4.1 Complex Populations in FNV Are Consistent With Observations Using Other YFV Strains**

Population diversity of YFV vaccines, though presently conceived by the quasispecies paradigm, was originally hypothesized in 1935 to explain the efficient adaptation of FNV to hepatic passage in rhesus macaques, which selected for WT viscerotropism (Findlay and Clarke 1935; Vignuzzi et al. 2006). Conversely, monkey neurotropism is efficiently rescued by passage of the 17D vaccine strain in mouse brain, showing increased morbidity after a single passage cycle (Meers 1959). Limited diversity is observed in commercial 17D vaccines as evidenced by rare neurotropic variants, which are recoverable at very low frequency from the 17D vaccine strain by plaque purification, and antigenic variants by monoclonal immunoassay (Liprandi 1981; Gould et al. 1989). These studies and others suggest a paradigm by which adverse variability in empirically derived YFV vaccines would be enhanced by a poorly controlled adaptive process.

### **5.4.2 Divergence of FNV Strains in Negative Association With Neurovirulent Phenotype**

FNV lots were produced from seeds of restricted passage level, controlling for sterility, potency, and immunogenicity in monkeys (Durieux, 1956b). In this study FNV population structure was examined from four different sources (FNV-IP, FNV-Yale, FNV-FC and FNV-NT), plus an example of the originating FNV “Dakar”

strain of the vaccine was only sequenced at the consensus level due to quality of the sample. Relative to FVV, conservation of 26 amino acid substitutions was observed for the four FNV examples sequenced [C (2), prM (1), M (1), E (7), NS2A (2), NS2B (1), NS3 (2), NS4A (2), NS4B (4) and NS5 (4)], suggesting that these are the particular amino acids involved in the derivation of FNV from FVV; however, we cannot exclude that other mutations in the original FNV have been lost on passage of the four FNV studied here. In particular, six amino acid substitutions (E-T7A, E-N135K, E-M457I, NS3-I13V, NS4B-A165V, NS5-R2T) were conserved between FNV-IP, FNV-Yale, and FNV-FC, but are not observed in FNV-NT [Table 5.2]. Thus, it is likely that the attenuation of FVV to produce FNV involved 26 amino acid substitutions.

It is reasonably hypothesized from these data that the low abundance of variant amino acid substitutions in FNV-Yale, both by consensus sequence and by variant reconstruction, reflects a short and protected passage history from the original vaccine, as FNV-Yale has been maintained in a low passage virus collection (WRCEVA). A comparison of FNV strains in mice showed that FNV-Yale produced the shortest AST following both i.c. and i.n. inoculation compared to FNV-IP and FNV-NT. [Table 5.1] (E. Wang et al. 1995). Likewise the FNV-NT strain, observed in this study to be the most divergent of YFV vaccines, was previously recognized by Wang et al to be nonlethal in cynomolgus monkeys (AST > 30 days) whereas FNV-IP and FNV-FC were lethal; FNV-Yale was not tested (E. Wang et al. 1995). Therefore it is likely that FNV-IP and FNV-Yale are of least divergence from FNV production lots, representing a genotype with high relative fixation to nervous tissue; this genotype would likely confer the greatest adverse neurotropic properties in humans.

Two distinct patterns arise in the diversity profiles of the vaccines when

only the 26 amino acid substitutions between FVV and FNV are considered [Table 5.3]. First, a set of positions were characterized by low diversity in FNV-IP and FNV-Yale that is progressively greater in FNV-FC and then greatest for FNV-NT (example: M-L36F, FVV: 0.182, FNV-IP: 0.014, FNV-Yale: 0.004, FNV-FC: 0.014, FNV-NT: 0.038). Broadly, these sites are represented in structural regions of the FNV genomes, an expected result in consideration of evidence that envelope gene residues exert particular influence over YFV neurovirulence (Ni, Ryman, et al. 2000). Second, a distinct set of positions were also characterized by low diversity in FNV-IP and FNV-Yale, however the greatest relative entropies are observed for FNV-NT, and then second greatest for FNV-FC (example: NS4B-V98I, FVV: 0.128, FNV-IP: 0.011, FNV-Yale: 0.020, FNV-FC: 0.126, FNV-NT: 0.033). Because FNV-FC was previously reported in mice to be both less neurovirulent and also less mouse neuroinvasive than the other vaccines (by the i.n. and i.c. routes), it is possible that the sites with high relative diversity in FNV-FC are determinants of mouse neuroinvasion. By RMSD, consensus WT revertants cluster at limited genetic distance from the parental strain, this is particularly evident for the FNV-NT strain [Figure 5.4]. This finding suggests that particular residues in the vaccine are selected to revert to WT under laboratory conditions.

#### **5.4.3 Inference of a Compact Genotype Underlying Neurotropism of FNV**

Using the consensus data in concert with sitewise diversity estimations, it is possible to hypothesize a set of residues that contribute to neurovirulence in both monkeys and mice. In particular, the destabilization of entropy for 9 amino acid substitutions in the M [2], E [4], NS4B [2] and NS5 [1] proteins (prM-V25M, M-L36F, E-T7A, E-A54V, E-N153K, E-N249D, NS4B-L19S, NS4B-A165V, and NS5-R2T in FNV-NT indicates that this smaller set of residues contributes to the attenuated phe-

notype in monkeys. Conversely, the destabilization of five amino acid substitutions in NS3 [1], NS4B [1] and NS5 [3] (NS3-I13V, NS4B-V98I, NS5-I285T, NS5-R335K, and NS5-F531L) in FNV-FC indicates that this smaller set of residues contributes to the neuroinvasive and neurovirulent phenotypes in mice.

#### **5.4.4 Amino Acid Substitutions are Similar Between FNV and 17D, Although Attenuated By Different Passage Tissues**

Notably, the substitutions M-L36F, E-K200T, E-K331R, and NS4B-I95M were observed in common for the attenuation processes between FVV to FNV, and Asibi to 17D [Table 5.2]. Flavivirus M protein lies contains a pro-apoptotic domain that, upon transfection and overexpression, is attenuated in phenotype when phenylalanine is substituted at residue 36 (Catteau 2003). Because the substitution at E-K200T is shared by Asibi and all FNV strains, it is likely either a cell culture adaptation or compensatory, and not consequential to the attenuated phenotype. Despite being associated with mouse neuroinvasion the residue did not singularly affect the phenotype of a cloned 17D virus (Lee and Lobigs, 2008). The residue E-K331R was associated with both attenuation of Asibi to yield 17D and experimental adaptation of the Asibi strain to hamster liver (McArthur et al. 2003), so the contribution to vaccine attenuation is complicated by the paucity of animal models and the difficulty to parse neurotropic from viscerotropic determinants of pathogenicity. The function of NS4B mutations in attenuation of the vaccine is not understood, however evidence has been put forth to suggest roles for NS4B in both interferon antagonism and replication complex associations (J. Zou, X. Xie, L. T. Lee, et al. 2014; Muñoz-Jordán et al. 2005).

#### **5.4.5 Reversion to Wild-Type is Associated with Loss of Neurotropism for FNV**

Reversion to WT is observed for FNV strains not only at the consensus level, but also for small population variants. In terms of both raw variant counts and revertants, the strains varied by virtue of hypothesized divergence from the original vaccine, e.g. FNV-IP and FNV-Yale alignments, revealed the relatively low counts of both variants and revertants [Table 5.4]. Conversely, counts of variants and revertants were higher in FNV-FC and FNV-NT, which are hypothesized to be divergent, presumably due to the passage history of these viruses in their institutions. Though genome-scale estimations of diversity are similar between all FNV strains and 17D-204; as expected, diversity of FVV was greater than for the vaccines. The processing and alignment methods used in these analyses were more stringent than previously used for the Asibi strain (Chapter 3), reducing observable differences in estimated diversity between Asibi and 17D-204 (A. Beck et al. 2014). However, density comparison shows that 17D-204 variants are fixed to lower aggregate frequencies than those of Asibi, FVV, or the FNV vaccines, and also that Asibi contains a small coding and silent variant population at higher frequencies than the other viruses [Figure 5.9 and Figure 5.10]. This pattern may contribute to the superior safety record of 17D vaccine compared to FNV, and supports a hypothesis that diversity influencing the vaccine phenotype arises from a limited number of sites, and that some tolerable level of low-frequency variants are contained in otherwise fixed vaccine preparations.

#### **5.4.6 Parental Strains FVV and Asibi Are Less Divergent than Previously Described and Likely Originated from the Same Epidemic Event**

An alignment of current reference consensus sequences for Asibi (Genbank accession AY640589.1) and FVV (Genbank accession YFU21056.1) shows 21 nu-

cleotide differences, of which 8 are coding changes (not shown) (E. Wang et al. 1995). The current study indicates lesser divergence of the parental strains than was previously described by Sanger methods undertaken 20 years ago, revealing instead 11 consensus nucleotide differences and two amino acid substitutions in this study. FVV and Asibi were both isolated in 1927 at considerable geographic separation (Senegal and Ghana, respectively); historical and research literature indicates that the epidemic was far-reaching in West Africa and encompassed both territories where the parental strains were collected (Stokes, Bauer, and Hudson 1928; Bazin 2011, p. 415). The limited divergence observed in this study by NGS methods supports a conclusion that both parental strains originated from the same YF epidemic event or reservoir population. This finding is consistent with previous observations of the low mutation rates of YFV strains in nature, suggesting that Asibi and FVV would be genetically similar if concurrently sampled, even at distance (Sall et al. 2010).

## 5.5 Conclusions

A hypothesis of low diversity for the live-attenuated vaccine strains compared to WT parental viruses was confirmed in these data. Because the reconstruction of highly diverse virus populations are ultimately a proxy for the inspection of low-abundance genotypes, strategies of this type can be advanced in a predictive context. Significantly, viral population structures have been reconstructed for FNV that reveal divergence of an empiric FNV vaccine from a putative original, adapted genotype. In this context, the extent of adaptation and fixation to substrate tissue by a vaccine strain was measured, revealing instability in FNV strain population structures. Reasonably, these features demonstrate an essential dependence upon culture conditions for the fixed genotype to be retained in the vaccine. The presented data support previous findings that employed classical techniques, including

evidence of instability for the FNV neurotropic phenotype *in vivo* (Holbrook et al. 2000; E. Wang et al. 1995). The findings provide further evidence that viral population structure should be considered in safety assessments of live-attenuated vaccine technologies.

**Table 5.1:** Source laboratories and *in vivo* lethality measurements for the FNV strains considered in the study.

Strain	Source	Monkey IC AST (d)	Mouse IC AST (d)	Mouse IN LD50 (log10PFU)
FNV-Yale	Yale Arbovirus Collection	NT	5.8	4.1
FNV-IP	Institut Pasteur, Paris	6	5.8	4.1
FNV-FC	CDC, Fort Collins	9	11.1	>5.6
FNV-NT	Porton Down (Public Health England)	>30	5.8	4.1
FVV	WRCEVA, UTMB	NT <sup>a</sup>		

<sup>a</sup> Not tested in monkeys or mice in Wang, et. al. 1995. When tested in Dunster et. al. 1999, FVV was 100 percent lethal in suckling mice at 21days post-infection (DPI), using a dose of 100 plaque-forming units (PFU) i.p..



**Table 5.2:** Amino acid and 3'UTR substitutions for FNV strains, relative to the parental FVV.

Nucleotide Position	Gene	Codon	FVV	Asibi	FNV-Dakar	FNV-IP	FNV-Yale	FNV-FC	FNV-NT
275	C	53	F	-	-	-	-	-	L
310	C	64	H	-	Q	-	-	-	-
357 <sup>a</sup>	C	80	V	-	A	A	A	A	A
386 <sup>a</sup>	C	90	S	-	G	G	G	G	G
491	prM	4	V	-	M	-	-	-	-
554 <sup>a</sup>	prM	25	V	-	M	M	M	M	M
596	prM	39	Y	-	-	-	-	-	H
701	prM	74	A	-	T	-	-	-	-
854 <sup>ab</sup>	M	36	L	-	F	F	F	F	F
989	E	6	I	-	V	-	-	-	-
992 <sup>a</sup>	E	7	T	-	A	A	A	A	- <sup>c</sup>
1134 <sup>a</sup>	E	54	A	-	V	V	V	V	V
1184	E	71	N	-	D	-	-	-	-
1392	E	140	R	-	K	-	-	-	K
1395	E	141	A	-	-	-	-	-	V
1424	E	151	N	-	D	-	-	-	D
1432 <sup>a</sup>	E	153	N	-	K	K	K	K	- <sup>c</sup>
1437	E	155	D	-	-	-	-	-	A
1440	E	156	I	-	-	-	-	-	T
1461	E	163	A	-	V	-	-	-	-
1506	E	178	A	-	V	-	-	-	-
1572 <sup>ab</sup>	E	200	T	K	K	K	K	K	K
1718 <sup>a</sup>	E	249	N	-	D	D	D	D	D
1965 <sup>ab</sup>	E	331	K	-	R	R	R	R	R
2105	E	378	V	-	I	-	-	-	-
2193	E	407	A	-	-	-	-	-	V
2344 <sup>a</sup>	E	457	M	-	-	I	I	I	- <sup>c</sup>
3234	NS1	261	K	-	-	-	-	-	R
3465	NS1	338	R	-	M	-	-	-	-
3469	NS1	339	K	-	-	-	-	-	N
3596	NS2A	30	M	-	V	-	-	-	-
3752 <sup>a</sup>	NS2A	82	I	-	V	V	V	V	V
3821 <sup>a</sup>	NS2A	105	T	-	A	A	A	A	A
3947	NS2A	147	T	-	A	-	-	-	-
4212	NS2B	11	A	-	-	-	-	-	V
4358 <sup>a</sup>	NS2B	60	S	-	A	A	A	A	A
4464	NS2B	95	D	-	-	-	-	-	G
4550	NS2B	124	H	-	-	-	-	-	Y
4559	NS2B	127	G	-	R	-	-	-	-
4607 <sup>a</sup>	NS3	13	I	-	-	V	V	V	- <sup>c</sup>
4754	NS3	62	R	-	W	-	-	-	W

4878	NS3	103	V	-	A	-	-	-	A
5145 <sup>a</sup>	NS3	192	M	-	K	K	K	K	K
5361	NS3	264	A	-	V	-	-	-	-
5408	NS3	280	I	V	-	-	-	-	-
6123	NS3	518	T	-	I	-	-	-	-
6451	NS4A	4	E	-	D	-	-	-	-
6459	NS4A	7	V	-	-	-	-	-	A
6527 <sup>a</sup>	NS4A	29	F	-	-	L	L	L	L
6602	NS4A	54	M	-	-	-	-	-	L
6611	NS4A	57	I	-	L	-	-	-	-
6761	NS4A	107	M	-	-	-	-	-	L
6843 <sup>a</sup>	NS4A	135	Y	-	F	F	F	F	F
6875	NS4A	145	V	-	-	-	-	-	L
6942 <sup>a</sup>	NS4B	19	L	-	-	S	S	S	- <sup>c</sup>
6953	NS4B	23	S	-	G	-	-	-	G
7171 <sup>ab</sup>	NS4B	95	I	-	M	M	M	M	M
7178	NS4B	98	V	-	-	I	I	I	I
7380	NS4B	165	A	-	-	V	V	V	-
7601	NS4B	239	M	-	V	-	-	-	-
7602	NS4B	239	M	-	T	-	-	-	-
7629	NS4B	248	G	-	A	-	-	-	-
7641 <sup>a</sup>	NS5	2	R	-	T	T	T	T	- <sup>c</sup>
7642	NS5	2	R	-	-	-	-	-	S
7694	NS5	20	D	-	N	-	-	-	N
8030	NS5	132	I	-	V	-	-	-	V
8090	NS5	152	S	-	-	-	-	-	P
8409 <sup>a</sup>	NS5	258	I	-	T	T	T	T	T
8471	NS5	279	E	-	K	-	-	-	-
8618	NS5	328	I	-	L	-	-	-	-
8640 <sup>a</sup>	NS5	335	R	-	K	K	K	K	K
8890	NS5	418	E	-	D	-	-	-	-
8928	NS5	431	K	-	-	-	-	-	R
8939	NS5	435	L	-	-	-	-	-	M
9227 <sup>a</sup>	NS5	531	F	-	L	L	L	L	L
9605	NS5	657	N	-	D	-	-	-	D
9615	NS5	660	K	-	R	-	R	-	R
10074	NS5	813	M	-	R	-	-	-	-
10120	NS5	828	M	-	-	-	-	-	I
10232	NS5	866	H	-	N	-	-	-	-
10268	NS5	878	I	-	V	-	V	-	V
10338	NS5	901	P	-	-	L	-	L	-
10357	3'UTR	-	C	-	U	-	U	-	U
10358	3'UTR	-	A	-	G	-	G	-	G
10363	3'UTR	-	C	-	-	-	-	-	U
10367	3'UTR	-	U	-	-	C	-	C	-
10389	3'UTR	-	C	-	U	-	-	-	-
10394	3'UTR	-	G	-	A	-	-	-	-
10404	3'UTR	-	G	-	-	-	A	-	-

10411	3'UTR	-	C	-	A	-	-	-	-
10418	3'UTR	-	U	-	-	C	-	C	-
10528	3'UTR	-	G	-	C	-	-	-	-
10531	3'UTR	-	G	-	A	-	-	-	-
10550	3'UTR	-	U	-	-	C	-	C	-
10554	3'UTR	-	C	-	-	-	-	-	U
10798	3'UTR	-	A	-	G	-	G	-	G
10800	3'UTR	-	A	G	-	-	-	-	-

**Table 5.3:** Diversity estimates by normalized Shannon’s entropy, considering the consensus sequences re-computed by NGS methods. Table considers each of 26 nucleotide positions containing amino acid substitutions conserved at greater than 75 percent in the FNV strains analyzed, relative to FVV.

Nucleotide Position	FVV	FNV-IP	FNV-Yale	FNV-FC	FNV-NT
357	0.278	0.005	0.004	0.010	0.066
386	0.173	0.025	0.014	0.028	0.097
554	0.062	0.054	0.018	0.009	0.080
854	0.182	0.014	0.004	0.014	0.038
992	0.148	0.007	0.001	0.010	0.131
1134	0.057	0.025	0.009	0.011	0.094
1432	0.133	0.003	0.005	0.021	0.123
1572	0.038	0.147	0.140	0.147	0.184
1718	0.091	0.003	0.007	0.013	0.068
1965	0.039	0.021	0.020	0.020	0.024
2344	0.121	0.020	0.007	0.006	0.085
3752	0.153	0.029	0.018	0.009	0.005
3812	0.126	0.065	0.057	0.057	0.061
4358	0.097	0.028	0.010	0.009	0.003
4607	0.076	0.036	0.022	0.234	0.053
5145	0.180	0.033	0.023	0.017	0.014
6527	0.169	0.008	0.004	0.035	0.006
6843	0.133	0.005	0.004	0.033	0.016
6942	0.106	0.031	0.020	0.099	0.166
7171	0.049	0.010	0.021	0.017	0.020
7178	0.128	0.011	0.020	0.126	0.033
7380	0.253	0.041	0.038	0.158	0.144
7641	0.168	0.013	0.012	0.113	0.279
8409	0.140	0.007	0.014	0.180	0.035
8640	0.054	0.024	0.018	0.128	0.038
9227	0.211	0.001	0.006	0.367	0.012

**Table 5.4:** Counts of WT revertants recovered under the heuristic method, using a one percent frequency cutoff.

Strain	Source	Monkey IC AST (d)	Median Entropy	All Revertants	Coding Revertants
FNV-Yale	Yale Arbovirus Collection	Not Tested	0.095	23	8
FNV-IP	Institut Pasteur, Paris	6	0.100	22	9
FNV-FC	CDC Fort Collins	9	0.095	53	12
FNV-NT	Porton Down (Public Health England)	>30	0.100	84	29
FVV	French Viscerotropic		0.131		
17D-204	17D Vaccine		0.102	22	14
Asibi	Asibi Strain		0.098		

**Table 5.5:** WT SNVs observed in the FNV strain alignments. Each nucleotide position bearing wild-type (FVV, parental strain) identity is indicated with nucleotide or amino acid identity and percentage frequency.

FVV		FNV-IP		FNV-Yale		FNV-FC		FNV-NT		17D-204	
Position	Nucleotide	Amino Acid	Nucleotide	Amino Acid	Percent	Nucleotide	Amino Acid	Percent	Nucleotide	Amino Acid	Percent
247	C	G	-	-	-	-	U	G	1.673	-	-
275	C	L	-	-	-	-	U	F	2.016	-	-
304	A	T	-	-	-	-	-	-	-	G	0.5641
337	U	D	-	-	-	-	C	D	2.249	-	-
357	C	A	-	-	-	-	U	V	3.218	-	-
373	A	K	-	K	0.4662	G	K	K	4.036	-	-
386	G	G	-	-	-	A	S	S	2.94	-	-
455	C	L	-	-	-	-	U	L	2.016	-	-
554	A	M	-	G	0.7743	-	G	V	2.949	-	-
556	U	I	-	-	-	-	G	M	3.359	-	-
562	G	T	-	-	-	-	A	T	3.163	-	-
596	C	H	-	-	-	-	U	Y	2.26	-	-
694	C	V	-	-	-	U	V	-	-	-	-
844	C	I	-	-	-	-	U	I	2.88	-	-
854	U	F	-	-	-	-	-	-	-	C	0.4738
1006	U	F	-	-	-	C	F	F	2.017	-	-
1036	C	V	-	-	-	U	V	V	1.755	-	-
1042	C	A	-	-	-	U	A	A	1.584	-	-
1127	A	R	-	-	-	-	-	-	-	G	0.8417
1134	U	V	-	-	-	C	A	-	-	-	-
1306	A	G	-	G	0.1964	G	G	G	1.846	-	-
1392	A	K	-	-	-	-	-	R	3.53	-	-
1395	U	V	-	-	-	-	C	A	3.06	-	-

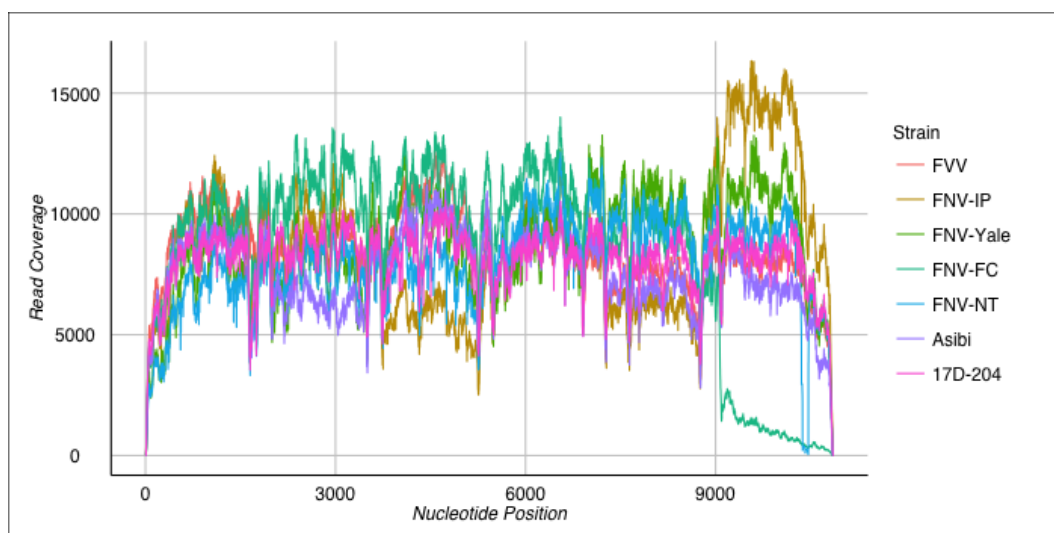




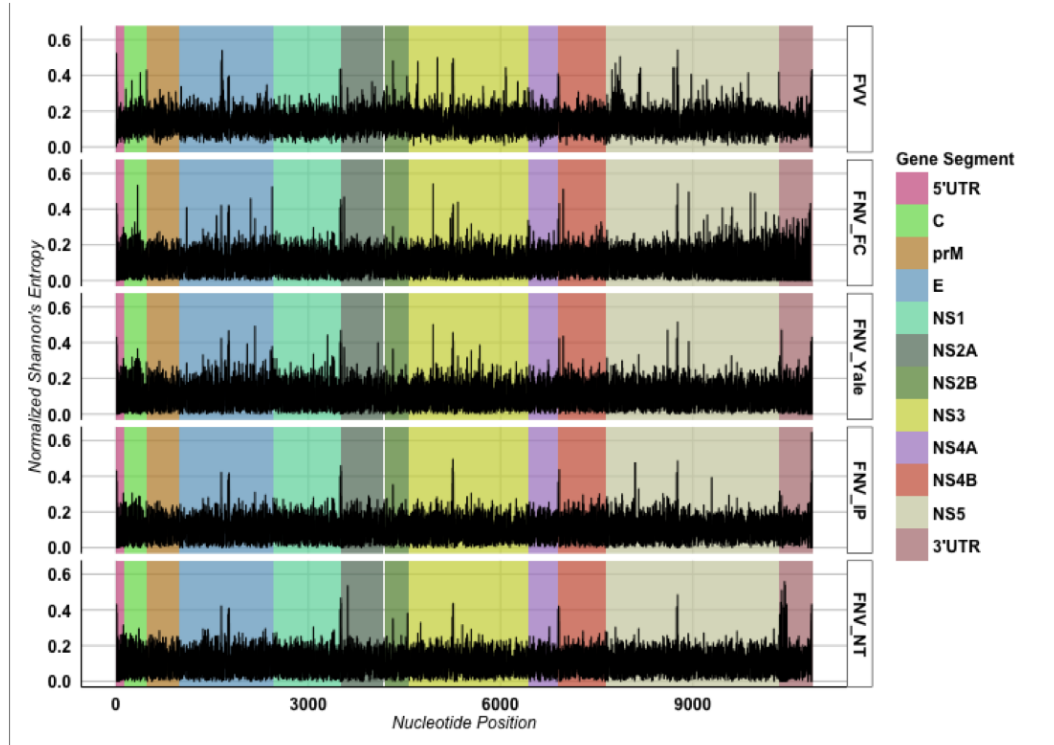




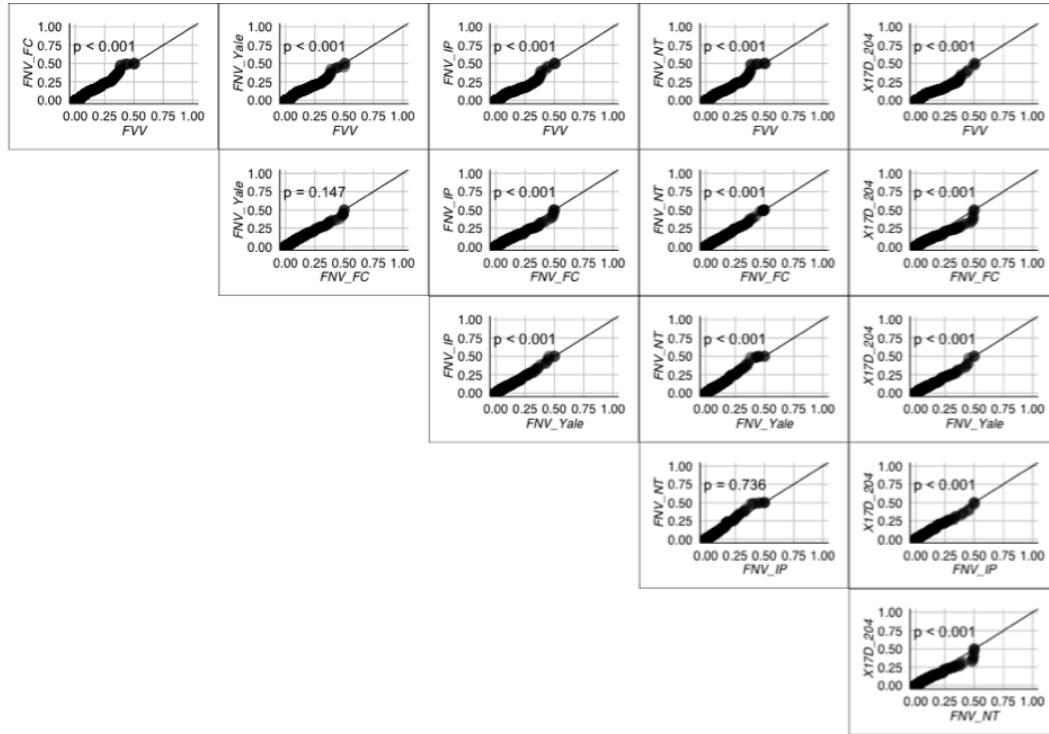




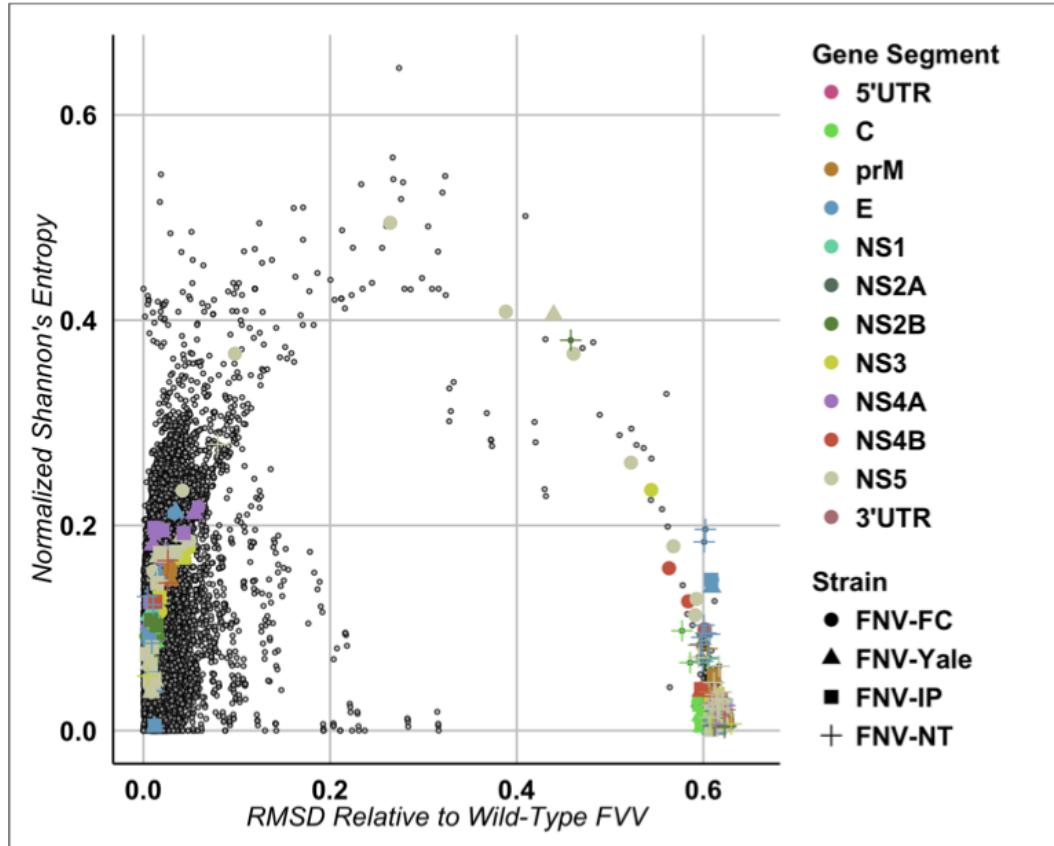
**Figure 5.1:** Plot of read coverages for each virus strain. Coverages were normalized to total read counts of the lowest coverage run, except the Asibi strain. Duplicate removal was performed on all alignments. For all strains considered in the study, mean coverage was  $>7000$  reads per nucleotide position.



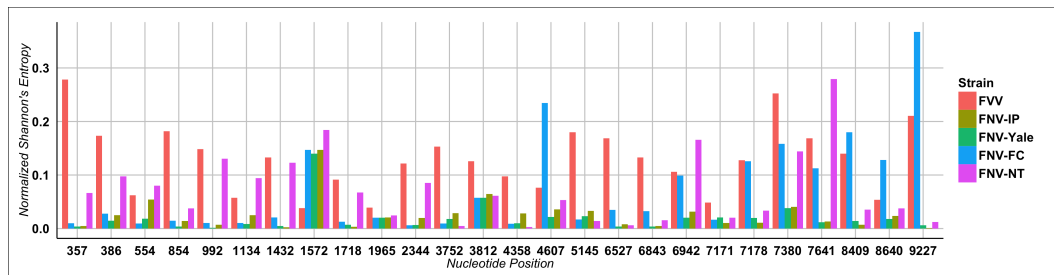
**Figure 5.2:** Normalized Shannon's entropy along the full genomes of FVV and all FNV strains. Entropy is shown for summed counts between sense and antisense reads, in order to exclude primer bias if present. In this context, higher relative diversity is repeatedly clustered across the viral genomes.



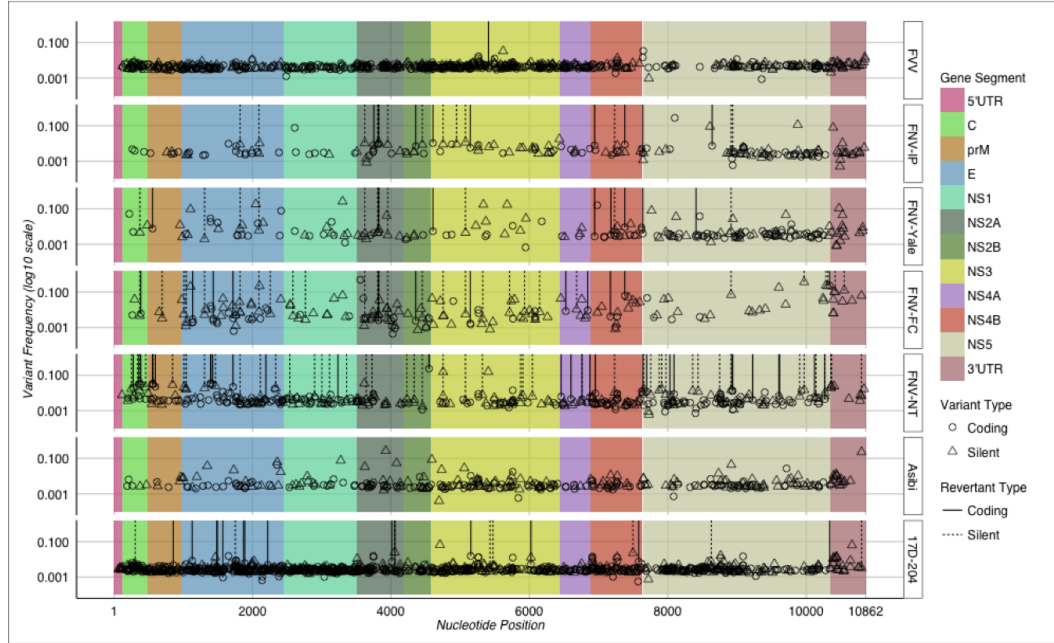
**Figure 5.3:** Pairwise quantile-quantile plots of sequencing error rate for FVV and FNV strains. Divergence of non consensus values is observed by aggregation of points away from a slope of 1.00. P-values were generated by a bootstrapped Kolmogorov-Smirnov test, using a Bonferroni corrected alpha of 0.008. Under these criteria, all pairings differ significantly except those of FNV-NT/FNV-Yale and FNV-NT/FNV-IP, demonstrating both the divergence of the vaccines from the parental strain, and between individual strains.



**Figure 5.4:** Scatter plot for all nucleotide positions of NSE (diversity) versus *RMSD* (increasing genetic distance) relative to the parental strain FVV. Sites at which any amino acid change were observed relative to FVV are shown for all vaccine strains (color). All other sites are depicted in grey. The divergence of genotypes for FNV strains is visualized, revealing a bimodal trend for some substitution sites at which reversion to wild-type is observed.



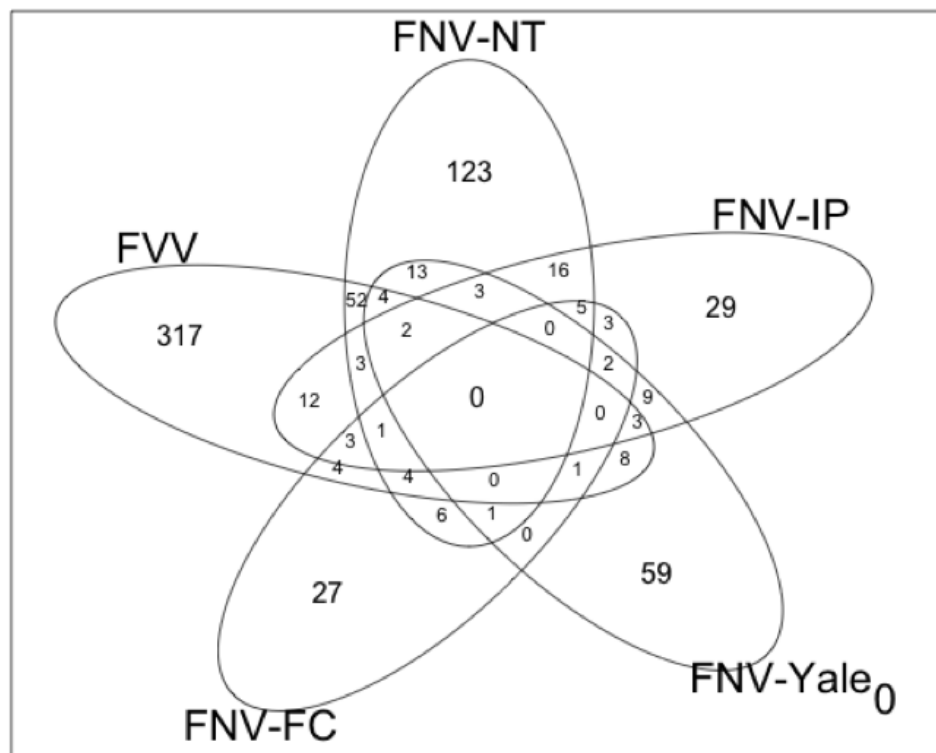
**Figure 5.5:** Diversity estimate by normalized Shannon's entropy for each of 26 nucleotide positions containing amino acid substitutions in the four FNV strains relative to FVV. Substitutions are conserved in the vaccine strains at minimum 75 percent (3 of four strains).



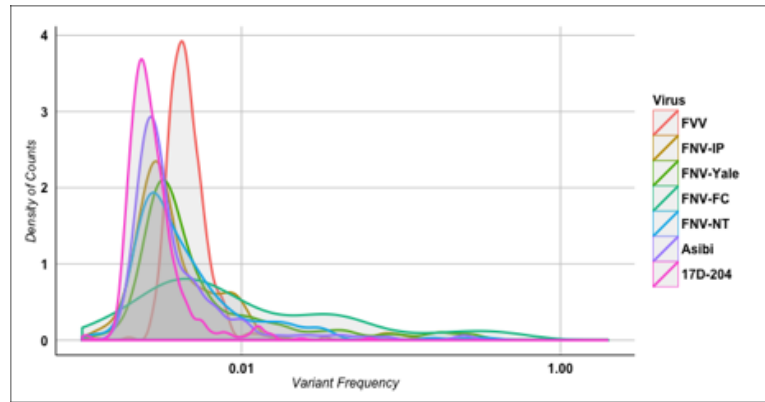
**Figure 5.6:** Scatter plots of nucleotide polymorphisms recovered in each virus strain, using a probabilistic model. Variants were classified against newly computed consensus sequences, showing variable abundance of population variants in the vaccine strains, relative to FVV. Low-frequency SNVs coding for reversion to the WT identity are shown, indicating greater relative abundance of revertants for vaccine strains previously observed to be poorly neurotropic.



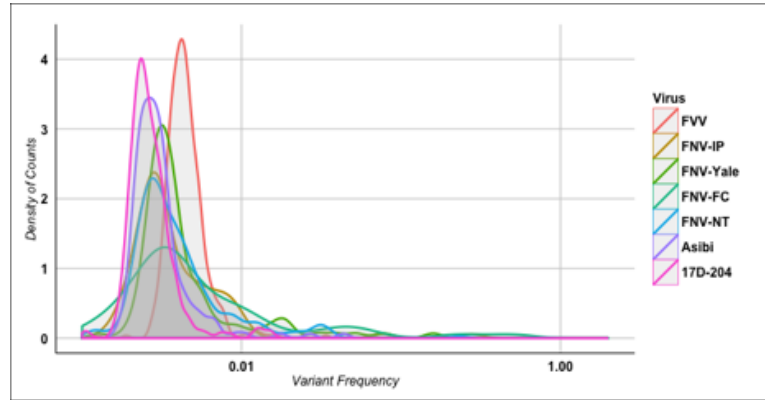




**Figure 5.8:** Venn diagram showing the intersection of count for only coding variant sites for FVV and FNV strains. Again, the greatest abundance is seen in the wild-type strain FVV.



**Figure 5.9:** Density plots showing relative accumulations of all SNVs variants as recovered by modeled criteria. Densities are of bimodal shape for the FNV vaccine strains and for the 17D parental strain Asibi. Plot includes both coding and silent variants, as classified against respective consensus sequences.



**Figure 5.10:** Density plots showing relative accumulations of coding SNVs as recovered by modeled criteria. Densities are of bimodal shape for the FNV vaccine strains and the 17D parental strain Asibi. Variants are classified against respective consensus sequences.

## Chapter 6

### Intrahepatic Passage of YFV Asibi Strain Selects a Strain With Reduced Population Complexity

#### 6.1 Abstract

Severe cases of YFV infection cause a fulminant hepatitis, which may lead to hemorrhage and death. This liver infection, called viscerotropic disease, is the most poorly understood phase of infection due to the effectiveness of available vaccines and the paucity of animal models that reflect this particular phase of the disease course in humans. However reported examples exist in which the viscerotropic phenotype is rapidly altered by passage. The hamster adapted strain hamster/P7, adapted by seven serial passages of the Asibi strain in hamster liver, produces viscerotropic disease and was observed to have a population of low complexity. The results indicate that infection of the liver is influenced by a selected genotype which in WT transmission is of low abundance, and that the role of viral diversity in the progression of YFV infection is to permit traversal of intermediate tissue barriers.

#### 6.2 Introduction

Empiric adaptation of viruses, e.g. gain-of-function studies are a common practice that has been used both to develop animal models of disease and to develop live-attenuated vaccines. For YFV, studies of viral adaptation represent a number of early examples of techniques used to select desired viral phenotypes under laboratory conditions. Wild-type YFV infection in humans and some non-human primate species is viscerotropic, referring to infection of the liver with consequent jaundice

and hemorrhagic manifestations in severe infections. Mortality has been estimated to be as high as 47 percent for patients entering this severe phase of YF disease, identified by an observation of jaundice less than two weeks following acute onset of fever (Johansson, Vasconcelos, and Staples 2014; World Health Organization 2003). Consequently, alterations to viscerotropic outcomes of YF disease under laboratory conditions have been closely studied. Recapitulation of viscerotropic disease in non-human primate models has been historically difficult. Of the small mammalian models featuring viscerotropic disease, the first to be characterized was developed by adaptation of the WT Panamanian Jimenez strain to Syrian golden hamsters (*Mesocricetus auratus*) by 10 i.p. passages of liver homogenate (Tesh et al. 2001; Xiao et al. 2001). The second, and the subject of this study, was similarly derived in *M. auratus* by 7 i.h. passages of the prototype West African Asibi strain, which was isolated in 1927 and is the parental strain of the 17D vaccine lineage [Figure 6.1] (McArthur et al. 2003). A mouse model with viscerotropic features was later described using non-adapted Asibi virus in IFN- $\alpha/\beta$  (preventing transduction of type I IFN signals) or  $\alpha/\beta/\gamma$  (preventing transduction of type I and type II IFN signals) receptor knockout mice (A129 and AG129) (Meier et al. 2009).

The Jimenez strain model has been used successfully as a challenge virus in nonclinical development of an inactivated 17D candidate vaccine (Monath, C. K. Lee, et al. 2010), however, models producing viscerotropy from challenge with the Asibi strain are of specific and particular use to the study of attenuation for live-attenuated 17D vaccines, as the passage relationship between the parent and vaccine is well-characterized (World Health Organization 2010; Theiler and H. H. Smith 1937). The Asibi hamster-adapted model described by McArthur et al. also recapitulates features of human viscerotropic disease, including histological features in the liver of hepatocyte necrosis, microvesicular steatosis, and inflammation (McArthur et al.

2003). High viremia was also observed for hamsters challenged with the adapted, “hamster/P7” strain. This feature of the hamster model is significant, as viremia is considered a correlate of viscerotropism, the absence of which is used as a criterion to qualify the loss of viscerotropism for YFV vaccine seeds in nonhuman primates.

Determinants of viscerotropism are poorly understood for YFV, the elucidation of which is complicated by a lack of experimental models reflecting perturbation of this phenotype. When sequenced by consensus, hamster/P7 genome differed from the WT non-hamster-passage input strain by seven amino acid substitutions (E-Q27H, E-D28G, E-D155A, E-K323R, E-F331R, NS2A-T48A), NS4B-V98I), predominately clustered in the envelope protein. Three clones of the hamster/P7 strain were sequenced by Sanger methods, revealing some population complexity in one (E-D28G, and one silent mutation at position 802A/G were observed to be mixed at 2/3) of the seven coding mutations acquired by i.h. passage. Frequently, passage-adapted viruses are identified by mixed clonal populations associated with intermediate phenotypes; the identification of a quasispecies effect is therefore considered likely for the hamster/P7 strain.

Because hamster/P7 represents divergence of the viscerotropic phenotype from the same parental source virus as the 17D vaccine, it represents an ideal candidate by which to investigate population-level effects that may influence virulence as compared to attenuation for 17D virus. Wild-type arboviruses are thought to mount complex population structures, or quasispecies, due to the need to replicate in both arthropod and mammalian hosts in such a manner as to permit the progress of infection through multiple tissue barriers, or the traversal of multihost transmission cycles. A previous comparison by deep sequencing of the Asibi strain to a commercial lot of 17D vaccine (which is attenuated for viscerotropism) showed that viral attenuation of the strain may be characterized by a lack of diversity (A. Beck

et al. 2014). Consequently, it is hypothesized that the divergence of attenuated or viscerotropic phenotype for the hamster/P7 strain will be reflected by attendant divergence of population structures *in silico*.

## 6.3 Results

### 6.3.1 Consensus Sequences

For the Asibi strain, the consensus sequence differed from the published reference (Genbank accession AY640589) at two silent nucleotide substitutions A2704G (NS1) and A10847C (3'UTR). For the hamster/P7 strain, 10 nucleotide substitutions were observed relative to Asibi virus sequenced in this study, of which 6 (E-Q27H, E-D155A, E-K323R, E-K331R, NS2A-T48A, and NS4B-V98I) were amino acid substitutions [Table 6.1].

### 6.3.2 Diversity of Strains

Mean whole-genome NSE for the WT strain Asibi was 0.104 (median=0.098 s.d.=0.063). Mean normalized entropy for the Asibi Hamster/P7 adapted strain was 0.105 (median=0.101, s.d.=0.059). When parsed by individual gene segment, very little change in diversity was observed between the two viruses [Table 6.2, Figure 6.2].

Selection of the hamster/P7 genotype was observed by a simultaneous increase in genetic distance from the parental Asibi genotype (RMSD) with a reduction in diversity for those sites [Figure 6.3]. The pattern was observed for both the 10 individual nucleotide substitutions and the selected genotype en masse. For nucleotide position 887, Asibi strain NSE was 0.209, hamster/P7 NSE was 0.014 (RMSD=0.604). For nucleotide position 1000, Asibi strain NSE was 0.094, hamster/P7 NSE was 0.014 (RMSD=0.619). For nucleotide position 1054, Asibi strain



NSE was 0.296, hamster/P7 NSE was 0.006 (RMSD=0.570). For nucleotide position 1437, Asibi strain NSE was 0.081, hamster/P7 NSE was 0.006 (RMSD=0.624). For nucleotide position 1941, Asibi strain NSE was 0.147, hamster/P7 NSE was 0.008 (RMSD=0.612). For nucleotide position 1965, Asibi strain NSE was 0.149, hamster/P7 NSE was 0.003 (RMSD=0.597). For nucleotide position 2779, Asibi strain NSE was 0.180, hamster/P7 NSE was 0.342 (RMSD=0.489). For nucleotide position 3821, Asibi strain NSE was 0.018, hamster/P7 NSE was 0.006 (RMSD=0.629). For nucleotide position 7178, Asibi strain NSE was 0.089, hamster/P7 NSE was 0.002 (RMSD=0.622). For nucleotide position 8971, Asibi strain NSE was 0.239, hamster/P7 NSE was 0.051 (RMSD=0.584). When the consensus mutations were considered in aggregate (n=10), mean NSE for Asibi was 0.150 (median=0.148, s.d.=0.083), and mean NSE for hamster/P7 was 0.045 (median=0.007, s.d.=0.105, mean RMSD=0.594, median RMSD=0.608). When only considering in aggregate the nucleotide positions containing nonsynonymous mutations between Asibi and hamster/P7 (n=6), mean NSE for Asibi was 0.130 (median=0.118, s.d.=0.095), and mean NSE for hamster/P7 was 0.005 (median=0.006, s.d.=0.002, mean RMSD=0.609, median RMSD=0.617).

### **6.3.3 Single Nucleotide Variant and Wild-Type Revertant Content of Strains**

For the WT strain Asibi, 340 variants were observed using probabilistic model recovery, which was not previously done, of which 192 coded for amino acid substitutions. For the hamster/P7adapted strain, 592 variants were observed, of which 383 coded for amino acid substitutions [Figure 6.4]. The variants are broadly distributed across the genomes of both strains, however few were of high frequency, or were coding in reference to the consensus sequence used. A minority of coding variants exceeded one percent frequency for either Asibi (5/192, or 2.6%) or hamster/P7

(3/383; or 0.8%) strains [Table 6.3]. High frequency coding variants for the Asibi strain were E-A407V (9.247% V), E-A459V (4.355% V), NS3-I195V (1.053% V), NS3-R397K (1.754% K), and NS5-D694N (2.686% N). High frequency coding variants for the hamster/P7 strain were NS4A-T157A (1.271% A), NS4B-L19S (1.170% S), and NS5-F202Y (1.402% Y).

Three subpopulation variants with partial reversion to the Asibi genotype were observed in the hamster/P7 strain, two of which were silent: A1000G (0.3638% G) and C2779U (20.30% U). One revertant coded for an amino acid substitution: NS4B-I98V (0.2817% V).

#### 6.3.4 Ratio of Nonsynonymous to Synonymous Mutations

When considered across the entire reading frame (codons = 3412) for each virus,  $dN/dS$  ratio for Asibi was 0.387, and for hamster/P7 was 0.554. When considered each gene, all  $dN/dS$  values for Asibi and hamster/P7 were below 1.000 [Figure 6.5]; thus, all summary values for  $dN/dS$  demonstrate negative selection across the considered ranges, whereas some inequality in the ratios for Asibi and hamster/P7 was observed for specific genes [Table 6.2]. When 20-codon sliding window analysis was used to resolve local trends for  $dN/dS$  ratios, maxima were observed in the Asibi genome in the envelope, NS1 and NS5 genes [Figure 6.6, upper]. For hamster/P7, local maxima were observed across the nonstructural regions of the genome [Figure 6.6, lower]. When the 20 codon windows were ranked, the highest  $dN/dS$  value for Asibi was 4.605 for the window spanning residues E-406 through E-426; the highest  $dN/dS$  value for hamster/P7 was 1.800 for the window spanning residues NS2A-182 through NS2A-202. Highest-ranked windows are classified in Table 6.4, with full tables in Appendix 1.

## 6.4 Discussion

### 6.4.1 Perturbation of Viral Diversity Is Frequently Associated With Change of Phenotype

Although perturbations in the genomic diversity of viruses are frequently identified in response to passage or drug treatment (Nishijima et al. 2012), the influence of viral diversity on disease pathogenesis in complex WT arbovirus transmission cycles is poorly understood. In cases such as with YFV, where close study of the central, WT disease manifestation (in this case viscerotropism) is hampered by a lack of animal models accurately recapitulating human disease, the use of population structure studies provides an association of fine virus complexity to *in vivo* outcomes. Significantly, Findlay and Clarke hypothesized a similar paradigm to the quasispecies model in 1935, following a study in which it was observed that viscerotropic and neurotropic phenotypes could be independently selected by *in vivo* passage of YFV in either mouse brain or monkey liver; he speculated that certain unlinked populations of virus would be of restricted competence to infect either the liver or brain (Findlay and Clarke 1935).

### 6.4.2 Intersection of Consensus and Subpopulation Sequence Content With Other Models of YFV Phenotype Adaptation

The consensus sequence observed in this study confirm those reported by McArthur, but differ by four nucleotides [Table 6.1, footnote a]; previously reported mutations A802G, A1056G (E-D28G), G3274A, and G4864A were not observed in the current findings (McArthur et al. 2003). However, the *in silico* approach used to produce the current findings provides rationale for the previously observed clonal variability at the silent substitution A802G and the coding substitution E-D28G. Modeled variant analysis revealed a subpopulation of the E-28 glycine residue in the Asibi strain at a frequency of 0.247 percent.

As shown previously, the selected consensus substitutions intersect with several other models of YFV phenotype differences by passage, indicating a limited sequence space available to the virus. Coding substitutions E-Q27H, E-D155A, E-K331R, NS2A-T48A, and NS4B-V98I intersect the attenuated “Hela/P6” strain, which was generated by six serial passages of Asibi virus in HeLa cells (Dunster, Gibson, et al. 1990). Coding substitution E-K331R is observed in 17D strain vaccines, although the arginine/lysine residues are geographically associated with South American and African WT strains, respectively; it is not likely that this codon contributes meaningfully to the phenotype (Hahn et al. 1987; C. N. d. Santos et al. 1995). Substitutions NS2A-T48A, and NS4B-V98I both intersect with the genotype of FNV, however, neither residue has been characterized for phenotype in an infectious clone system (E. Wang et al. 1995).

Recovery of WT revertants in hamster/P7 virus is limited, but indicates that selection pressures encountered by Asibi virus in serial hepatic passage in hamsters are not equivalent along the genome. The only nonsynonymous coding revertant NS4B-I98V, was incompletely fixed to isoleucine in the hamster/P7 population. The greatest subpopulation variability by frequency in the Asibi strain is observed in the envelope gene, whereas in hamster/P7 the presence of variant structure is more limited by frequency and is localized to the nonstructural regions of the genome. This pattern indicates that the envelope residues are more heavily influenced by positive selective pressure under i.h. passage of Asibi, with fixation occurring more rapidly in the envelope than for residues in the nonstructural genes. This may suggest that adaptation to hamster liver is in part due to the ability of Asibi virus to bind to cell receptors and infect hamster hepatocytes.

#### **6.4.3 Transition from Asibi to Hamster/P7 Indicates Unique Local Selection Patterns Across Parent and Selected Genomes**

Arboviral genomes are broadly subject to negative selective pressures, owing to the necessity to productively infect multiple hosts in WT transmission. Across multiple scales,  $dN/dS$  ratios for both Asibi and hamster/P7 strains substantiate this expected pattern and indicate some level of conservation for the strains. As demonstrated in the single nucleotide variant analysis, local resolution of  $dN/dS$  by 20 codon windows reveals that some level of positive selection is ongoing in the nonstructural regions of the hamster/P7 genome, whereas the structural genes are fixed by the seventh passage in liver. High  $dN/dS$  ratios in the envelope gene of Asibi virus may indicate selection of a WT quasispecies that has the ability to infect hamster hepatocytes, however this has not been tested using a controlled viral population. This pattern was further confirmed by the reduction in diversity at the consensus mutation sites differentiating Asibi and hamster/P7 viruses.

#### **6.4.4 A Rationale for the “Iceberg Effect” By Observed Relaxation of Evolutionary Tradeoffs**

In the case of arboviruses, the paradigm of “trade-off” hypothesizes that the overall fitness is compromised of a virus that must productively infect both arthropod and vertebrate hosts in nature. In addition, the “iceberg” paradigm is frequently used to explain the low probability for a particular virus to cause symptomatic infection or rarely an advanced or severe infection of the host. This property is also observed for YFV, in which a recent modeling study estimated that 12% of infections cause severe disease featuring jaundice (Johansson, Vasconcelos, and Staples 2014). Although less diverse than the parental input strain, the change in consensus and diversity of hamster/P7 strain, reflects an expected result as a WT viral population would not be expected to reliably infect the hamster liver unless introduced to an ideal

host, and this is observed by hamster passage 1 to hamster passage 5 viruses. The presence of hamster viscerotropic alleles in WT YFV would therefore in most cases be expected to be rare.

## 6.5 Conclusions

The working hypothesis that the adapted hamster/P7 would contain a unique population structure relative to the unpassaged Asibi strain was confirmed in this dataset. In this study, an example of rapid phenotype selection for YFV were analyzed by *in silico* methods, counterintuitively revealing that the adapted and viscerotropic hamster/P7 strain was of low diversity, with a narrowly selected genotype that was selected to produce liver infection of hamsters. The sensitive detection by these methods of shifts in the viral quasispecies in association with changes to the viscerotropic phenotype suggests the use of deep sequencing to resolve the presence of virulent subpopulation alleles for viruses lacking established or reliable animal models.

**Table 6.1:** Amino Acid Substitutions in the Hamster/P7 virus, relative to the unpassaged (P0) Asibi strain from which it was derived.

Position	Nucleotide		Gene	Codon	Amino Acid	
	Asibi/P0	Hamster/P7			Asibi/P0	Hamster/P7
802 <sup>a</sup>	A	G				
887	C	U				
1000	G	A				
1054	A	C	E	27	Q	H
1056 <sup>a</sup>	A	G	E	28	D	G
1437	A	C	E	155	D	A
1941	A	G	E	323	K	R
1965	A	G	E	331	K	R
2779	U	C				
3274 <sup>a</sup>	G	A				
3821	A	G	NS2A	48	T	A
4864 <sup>a</sup>	G	A				
7178	G	A	NS4B	98	V	I
8917	C	U				

<sup>a</sup> Substitution was described by McArthur et al, but was not observed in this study; these results recapitulate by NGS methods the results of the McArthur study, in which clonal variability was observed at nucleotide positions 802 and 1056 (McArthur et al. 2003).

**Table 6.2:** Summary of  $dN/dS$  values for gene segments and complete open reading frame in both Asibi and Hamster/P7 strains.

Gene	Asibi			Hamster/P7		
	Mean	Median	S.D.	Mean	Median	S.D.
5'UTR	0.097	0.075	0.093	0.104	0.083	0.099
C	0.102	0.090	0.067	0.104	0.097	0.063
prM	0.103	0.099	0.058	0.104	0.101	0.054
E	0.104	0.097	0.065	0.105	0.102	0.059
NS1	0.103	0.096	0.066	0.107	0.103	0.058
NS2A	0.105	0.100	0.062	0.106	0.104	0.055
NS2B	0.102	0.099	0.060	0.102	0.100	0.053
NS3	0.104	0.098	0.060	0.104	0.100	0.055
NS4A	0.105	0.104	0.058	0.107	0.109	0.052
NS4B	0.106	0.099	0.068	0.106	0.099	0.068
NS5	0.103	0.098	0.062	0.103	0.098	0.060
3'UTR	0.108	0.098	0.070	0.109	0.105	0.065



**Table 6.3:** Coding SNVs in Asibi and Hamster/P7 strains, in reference to the respective consensus of each strain.

Strain	Position	Gene	Codon	Nucleotide		Amino Acid		Percent
				Cons <sup>a</sup>	Var <sup>b</sup>	Cons	Var	
Asibi	2193	E	407	C	U	A	V	9.247
	2349	E	459	C	U	A	V	4.355
	5153	NS3	195	A	G	I	V	1.053
	5760	NS3	397	G	A	R	K	1.754
	9716	NS5	694	G	A	D	N	2.686
Ham/P7	6911	NS4A	157	A	G	T	A	1.271
	6942	NS4B	19	U	C	L	S	1.170
	8241	NS5	202	U	A	F	Y	1.402

<sup>a</sup> Consensus nucleotide or residue in the newly computed consensus sequence.

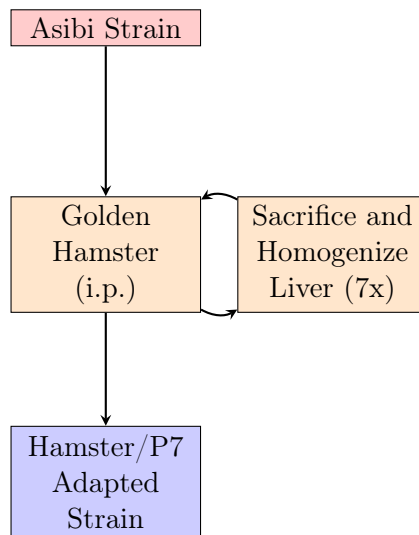
<sup>b</sup> Variant nucleotide or residue in the alignment, which is observed below the consensus frequency.

**Table 6.4:** 20-codon sliding windows with high  $dN/dS$  ratios in the strains considered.

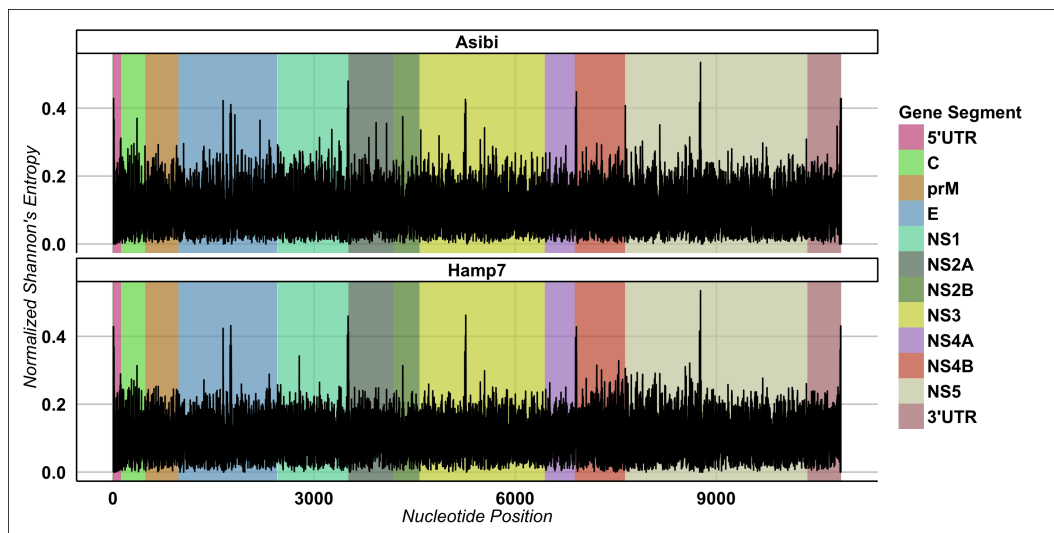
Strain	Gene	Nucleotide Window <sup>a</sup>		Codon Window <sup>b</sup>		$dN/dS$
		Start	End	Start	End	
Asibi	E	2189	2249	406	426	4.605
	E	2180	2240	403	423	3.947
	E	2171	2231	400	420	3.693
	E	2165	2225	398	418	3.329
	E	2156	2216	395	415	3.305
	E	1013	1073	14	34	2.887
	E	1028	1088	19	39	2.655
	NS1	2456	2516	2	22	2.641
	E	1001	1061	10	30	2.501
Hamster/P7	NS1	2465	2525	5	25	2.485
	NS2A	4052	4112	182	202	1.800
	NS2A	4025	4085	173	193	1.792
	NS5	8312	8372	226	246	1.740
	NS5	8294	8354	220	240	1.722
	NS1	2813	2873	121	141	1.688
	NS5	8300	8360	222	242	1.649
	NS5	8228	8288	198	218	1.648
	NS2A	3746	3806	80	100	1.641
	NS3	5033	5093	155	175	1.638
	NS4B	7499	7559	205	225	1.622

<sup>a</sup> Sixty nucleotides, inclusive.

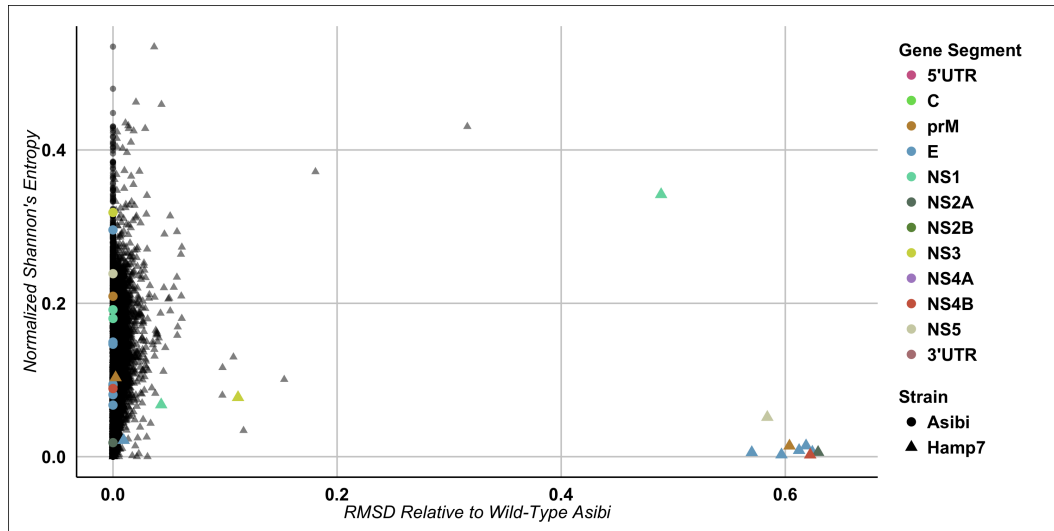
<sup>b</sup> Twenty residues, inclusive.



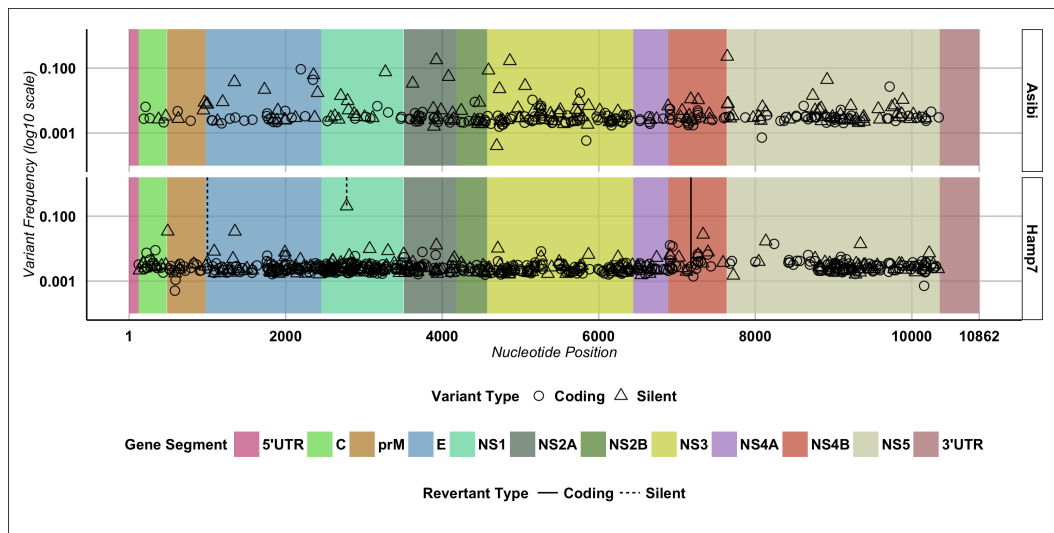
**Figure 6.1:** Derivation of Hamster/P7 strain.



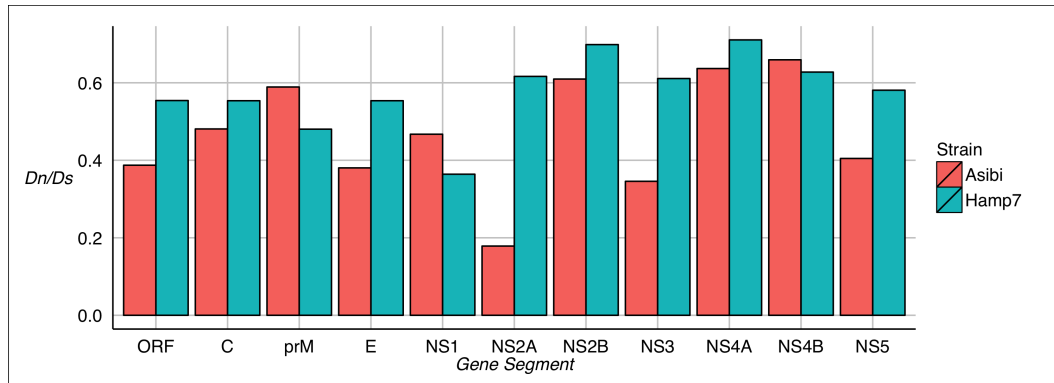
**Figure 6.2:** Line chart of diversity estimates for each nucleotide site in the genomes of Asibi and hamster/P7 strains, expressed as normalized Shannon's entropy for each nucleotide position recovered in the read alignments.



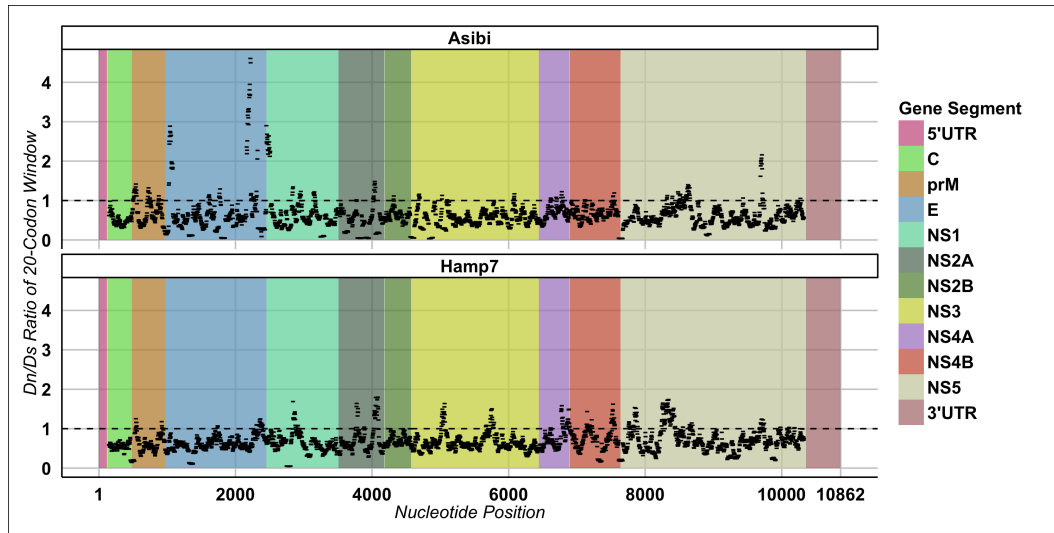
**Figure 6.3:** Scatter plot considering each nucleotide position recovered by deep sequencing in Asibi and hamster/P7 strains; normalized Shannon's entropy (diversity index) versus RMSD (genetic distance) from Asibi is shown for each position. Hamster/P7 sites reveal complete selection from the parental genotype; or a pattern of both low diversity and high genetic distance from the parental strain. RMSD for Asibi positions = zero, and are included for reference. Color indicates the gene of the observed mutation, and indicates that the nucleotide position was recovered in the panel of consensus mutations differentiating Asibi and hamster/P7 [Table 1].



**Figure 6.4:** Variant populations identified by a model (V-Phaser 2.0) in the sub-population structure of the strains Asibi, and hamster/P7. With respect to the consensus sequence of each strain, SNVs were classified as either coding or silent. Points indicated by vertical lines represent revertants to the parental (Asibi) genotype that are incompletely selected in the population.



**Figure 6.5:**  $dN/dS$  estimation summarized along each gene for all viruses considered in the study. Ratio for all gene segments in the open reading frames of Asibi and hamster/P7 strains were below 1, indicating no evidence of positive selection relative to the consensus sequence.



**Figure 6.6:**  $dN/dS$  estimation along the full genome for each strain considered in the study, using a sliding window. Ratio is calculated along a sliding 20-codon window, using a step size of 1.0 codon distance. Each window with a value greater than 1 (horizontal dashed line) is considered to be subject to positive selection, as the ratio of nonsynonymous substitutions per possible nonsynonymous site is greater than that for silent substitutions.



## Chapter 7

### Comparison of the Live-Attenuated Japanese Encephalitis Vaccine SA14-14-2 to Its Virulent Parental Strain SA14 By Deep Sequencing

#### 7.1 Abstract

Japanese encephalitis (JE) is a mosquito-borne flavivirus disease that is controlled by vaccination. Live attenuated JEV vaccine strain SA14-14-2 was derived by serial passage of WT strain SA14 in hamster kidney, licensed in China in 1989 and 8 other countries thereafter. It has been used as a component of childhood immunization programs throughout Asia with approximately 400 million doses distributed. Although the vaccine is considered highly safe and effective, and some progress has been made to elucidate the discrete mechanisms of attenuation, the complexity of the attenuating passage history for SA14-14-2 has prevented full resolution of the genotype for the vaccines. We have previously investigated another live attenuated flavivirus vaccine, yellow fever strain 17D, by deep sequencing and shown that the WT parent Asibi has a typical quasispecies structure of a RNA virus while 17D has a homogeneous structure suggesting that attenuation is in part due to lack of diversity of the 17D genome. Using a similar deep sequencing approach, the subpopulation structures of two examples of SA14-14-2 vaccine were compared to those of the WT parental strain SA14. SA14-14-2 virus was found to have lower diversity than the parental strain; this pattern is highly localized, along with limited observation of WT sequence content in the viral subpopulation. The findings demonstrate that similar lack of diversity for two live attenuated flavivirus vaccines may indicate a

characteristic of live-attenuated RNA viruses generally. These observations indicate that diversity of live viral vaccines should be considered in assessments of vaccine safety and quality.

## 7.2 Introduction

JEV is the etiologic agent of JE, which is a major threat to public health across a range comprising 24 countries in east and southeast Asia. JEV is a member of the family *Flaviviridae*, genus *flavivirus*, and has a 50nanometer (nm) enveloped particle, containing a single-stranded, positive-sense RNA genome of 11kb, which encodes 3 structural genes (C, prM/M, E) and 7 nonstructural genes (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5). The polyprotein sequence is flanked by 5' and 3' UTRs. JEV is the prototype member of a same-named serogroup that includes a number of arthropod-borne pathogens causing infection of the central nervous system in humans, including WNV, Murray Valley encephalitis virus (MVEV), Usutu virus (UsV), and others. The ecology of JEV transmission is complex, in which the virus is principally vectored by mosquitoes of the genus *Culex*, in particular *Cx. tritaeniorhynchus*, which is ubiquitously found across the range of the virus and associated with human cases (R. H. Miller et al. 2012). Amplification hosts include domestic pigs and wading birds; humans are considered dead-end hosts, encountering the virus niche by proximity to rice cultivation or swine husbandry (Zhao et al. 2014).

Annual case burden is estimated to be 68,000 in total across the endemic and epidemic ranges, mostly in children, despite widespread access to vaccination (Campbell et al. 2011). Although the pathogenesis of JE disease is poorly understood, the virus is observed histologically to invade the mammalian central nervous system, a finding observed in multiple animal models of JEV infection (F. Li et al.

2015; Myint et al. 2014; World Health Organization 2003). In severe cases, JEV infection may cause permanent neurological sequelae, including paralysis, weakness, or impaired cognition (Yin et al. 2015). Owing to the geographic ubiquity of JEV in Asia, severity of JE disease, and economic impacts of long-term care, a series of vaccines has been developed, beginning with formalin-inactivated mouse brain preparations using the Nakayama-NIH or Beijing-1 strain in the 1950s and subsequently replaced by formalin-inactivated cell-culture derived vaccines. The live-attenuated vaccine strain SA14-14-2, produced since 1988 by the Chengdu Institute of Biological Products in China, was the only live-attenuated JEV vaccine using the entire JEV genome. SA14, the parental WT strain of SA14-14-2, was prepared from a pool of *Cx pipiens* mosquitoes in Xi'an, China. The resulting preparation was passaged 100 times in primary hamster kidney (PHK) cells, then in a complex series of mammalian tissues, from which plaque picks were periodically taken.

The full passage history of SA14-14-2 is shown in Figure 7.1, with reference to other live-attenuated JEV vaccine strains studied or developed along other points on the same passage series. The vaccine is greatly attenuated with respect to both direct mouse neurovirulence and neuroinvasion (entry to the central nervous tissues from periphery) to the extent that intracerebral inoculation of one million PFU causes no clinical or histopathologic signs of infection in mice and monkeys (World Health Organization 2014). RNA viruses encode a RNA-dependent-RNA polymerase, which unlike DNA polymerases lacks proof-reading activity and, consequently, RNA virus genomes exist as a quasispecies of related RNA sequences. Varying levels of naturally-occurring subpopulation diversity arises during the lifecycle of RNA viruses, a consequence of large population sizes and error-prone replication. In the case of empirically-derived vaccine strains such as SA14-14-2, the population structure of the virus is likely to carry informative signals of passage history,

providing an association of specific elements of viral diversity with the controlled vaccine phenotype. Current WHO technical specifications for live-attenuated JEV vaccines does not prescribe a standard genotype, owing to a multiplicity of reported genotypes for the vaccine, and the acceptable safety margins afforded by empiric, *in vivo* qualification of the vaccine (Ni and A. D. Barrett 1995; Nitayaphan et al. 1990; Song et al. 2012; Aihara et al. 1991; World Health Organization 2014). In the case of arthropod-borne viruses such as JEV, changes in population structure in the empiric transition to vaccine phenotype are likely to be especially meaningful, owing to the complex multiple-host niches occupied by the virus in nature to derive a live vaccine strain that is attenuated in mammalian hosts and not mosquito competent. In this paradigm, the properties of the viral “quasispecies” population may effect pathogenesis or conversely attenuation; low-abundance populations in a virus may exert wild type effects or otherwise determine the course of pathogenesis (Vignuzzi et al. 2006). In a strict sense, this model of the virus lifecycle predicts that some level of population diversity is necessary to traverse infection barriers or host tissues. Because the infection caused by the JEV vaccine is attenuated both for neurotropic disease and characterized by low viremia, a reasonable hypothesis arises that SA14-14-2 is of lower diversity than the WT strain from which it was derived.

NGS studies represent a family of *in silico* techniques with potential to assess and predict phenotype of vaccines, particularly by the resolution of population substructure and detection of contaminants (Prachi et al. 2013; Luciani, Bull, and A. R. Lloyd 2012; Victoria et al. 2010). Previous work on this topic using NGS techniques has observed that the related flaviviral live-attenuated YFV vaccine strain 17D is of lower diversity than the WT parental Asibi strain from which the vaccine was derived (A. Beck et al. 2014). Like SA14-14-2, 17D was passaged to attenuate in a series of multiple, alternating tissues, leading to the occurrence of viral population

bottlenecks or purifying selection for the vaccine genotype (Theiler and H. H. Smith 1937). Because SA14-14-2 represents a closely related live attenuated vaccine to YFV, JEV offers a unique opportunity to confirm that patterns of low diversity are a feature of live-attenuated strains, and to contribute to a framework by which this information is used to guide expectations of live vaccine safety and stability.

## **7.3 Results**

### **7.3.1 Read Coverages**

Mean read coverage of the WT strain SA14 was 2262.337 (median=2188, s.d.=702.368). Mean read coverage of the vaccine strain SA14-14-2/P1 was 2258.343 (median=2299, s.d.=501.016). Mean read coverage of the vaccine strain SA14-14-2/P2 was 2263.687 (median=2028, s.d.=1020.042) [Figure 7.2].

### **7.3.2 Consensus Sequences**

#### **7.3.2.1 Comparison of Newly Sequenced SA14 to Previously Reported SA14 Sequences**

Three other strains of SA14 have been previously sequenced at the consensus level, and were compared in reference to the newly sequenced SA14 strain by consensus [Table 7.1]. Genbank accession M55506 (Nitayaphan et al. 1990) differed from SA14 at 11 amino acids (E-E244G, E-A315V, E-K439R, NS1-G292S, NS1-R339M, NS2A-N2K, NS2A-A40V, NS2A-V106I, NS3-K73R, NS5-K328E, NS5-N644T). Genbank accession D90194 (Aihara et al. 1991) differed from SA14 at five amino acids (E-E244G, E-P334S, NS3-K73R, NS3-A235V, NS5-G741D). Genbank accession JEU14163 (Ni, Burns, et al. 1994) differed from SA14 at four amino acids (NS1-G292S, NS2B-M102T, NS3-K73R, NS4A-K79R). Only one of the 16 amino acid substitutions (NS3-K73R) was conserved between all of the previously reported strains, in reference to SA14.

### **7.3.2.2 Comparison of SA14 to SA14-14-2/P1 and SA14-14-2/P2**

In comparison to the newly sequenced parental WT strain SA14, SA14-14-2/P1 and SA14-14-2/P2 were identical to each other, each containing 53 nucleotide differences, of which 22 were amino acid substitutions [Table 7.5]. With respect to viral genes, one amino acid substitution was observed in C (C-L66S), eight E (E-L107F, E138K, I176V, T177A, E244G, Q264H, A315V, K439R), three in NS1 (NS1-G292S, R339M, D351H), two in NS2A (NS2A-N2K, A40V), two in NS2B (NS2B-E63D, N65G), three in NS3 (NS3-M59V, A105G, A235V), one in NS4B (NS4B-I106V), and three in NS5 (NS5-H386T, Q639H, V671A).

### **7.3.3 Diversity and Genetic Distance of Vaccines, Relative to the Wild-Type Strain**

#### **7.3.3.1 Summary of Diversity for Entire Strain Genomes and For Gene Segments**

Diversity was estimated at individual nucleotide positions across the genomes of SA14, SA14-14-2/P1, and SA14-14-2/P2 [Figure 7.3]. For all nucleotide positions in entire open reading frames (n=10296), mean normalized entropy for the WT strain SA14 was 0.103 (median=0.100, s.d.=0.049). Mean normalized entropy for the vaccine strain SA14-14-2/P1 was 0.103 (median=0.100, s.d.=0.051). Mean normalized entropy for the vaccine strain SA14-14-2/P2 was 0.100 (median=0.094, s.d.=0.057). Sets differed significantly by the Kruskal-Wallis rank-sum test ( $X^2=24681.59$ , d.f.=2,  $p<0.001$ ). Between the strains, the rank-order of estimated diversity for each gene segment was variable [Table 7.3].

#### **7.3.3.2 Summary of Diversity for Proposed SA14-14-2 Genotype**

When considering the set of nucleotide positions containing the consensus amino acid substitutions between SA14 and SA14-14-2 strains only (n=22), mean

normalized entropy for the WT strain SA14 was 0.121 (median=0.093, s.d.=0.099). Mean normalized entropy for the vaccine strain SA14-14-2/P1 was 0.097 (median=0.097, s.d.=0.043). Mean normalized entropy for the vaccine strain SA14-14-2/P2 was 0.095 (median=0.090, s.d.=0.049). Strains differed significantly by the Kruskal-Wallis rank-sum test ( $X^2=56.524$ , d.f.=2,  $p < 0.001$ ).

### 7.3.3.3 Individual Diversity of Amino Acid Substitution Sites for Proposed SA14-14-2 Genotype

When considering the set of nucleotide positions containing the consensus amino acid substitutions between SA14 and SA14-14-2 strains only, differences in summarized diversity were unequally localized between discrete genes [Table 7.4]. Genes containing 14 of the 22 sites were demonstrated to have greatest relative normalized entropy compared to the vaccine strains; mean values for these genes were as follows: E:(n=8 SA14=0.169, SA14-14-2/P1=0.090, SA14-14-2/P2=0.092), NS2A:(n=2, SA14=0.143, SA14-14-2/P1=0.101, SA14-14-2/P2=0.111) NS3:(n=3, SA14=0.106, SA14-14-2/P1=0.093, SA14-14-2/P2=0.093), and NS4B:(n=1, SA14=0.080, SA14-14-2/P1=0.032, SA14-14-2/P2=0.051).

Differences in mean entropy were further localized to a subset of positions within the 14 sites identified to have high relative diversity in SA14 [Table 7.5, Figure 7.4]. These positions were, respectively, 1296-E-L102F:(SA14=0.137, SA14-14-2/P1=0.097, SA14-14-2/P2=0.110), 1506-E-T177A:(SA14=0.175, SA14-14-2/P1=0.084, SA14-14-2/P2=0.140), 1708-E-E244G:(SA14=0.489, SA14-14-2/P1=0.081, SA14-14-2/P2=0.108), 1921-E-A315V:(SA14=0.112, SA14-14-2/P1=0.076, SA14-14-2/P2=0.083), 2293-E-K439R:(SA14=0.193 SA14-14-2/P1=0.037, SA14-14-2/P2=0.062), 3652-NS2A-A40V:(SA14=0.222, SA14-14-2/P1=0.127, SA14-14-2/P2=0.175), and 4782-NS3-M59V:(SA14=0.195, SA14-14-2/P1=0.142, SA14-14-2/P2=0.160). Genetic distance (RMSD) of these sites in the vaccine strains relative to SA14 was computed using the

nucleotide frequencies in the alignments, and revealed a pattern of stable selection for the vaccine genotype [Figure 7.5].

#### 7.3.4 Variant and Wild-Type Revertant Structure of SA14-14-2 Strains

For the WT strain SA14, 62 SNVs were observed in total; of these, 35 were coding changes relative to the consensus sequence. For the vaccine strain SA14-14-2/P1, 92 SNVs were observed in total; of these, 44 were coding relative to the consensus. For the vaccine strain SA14-14-2/P2, 72 SNVs were observed in total; of these, 35 were coding relative to the consensus [Figure 7.6]. Variants were consistently observed between SA14-14-2/P1 and SA14-14-2/P2; of the 121 possible unique SNVs observed between either of the two vaccine strains, 43 (35.5%) were common. 20 of the 43 variants were observed that coded for amino acid substitutions with respect to the [identical] consensus sequences of the vaccines [Table 7.6]. Two WT revertant amino acids were observed in common between the vaccines, NS1-V392G (0.9572% in P1, 1.365% in P2) and NS3-V235A (1.693% in P1, 1.707% in P2).

#### 7.3.5 Selection Analysis and $dN/dS$ Ratios

Over the open reading frame of each strain,  $dN/dS$  for SA14 was 0.413, for SA14-14-2/P1 was 0.474, and for SA14-14-2/P2 was 0.419. When considering individual genes,  $dN/dS$  ratios were not equal between genes for both the WT and vaccine strains [Table 7.7]. Broadly,  $dN/dS$  ratios for all strains did not exceed 1.000, indicative of negative or purifying selection. Of the segments with  $dN/dS$  ratios equal to or greater than 1, indicating neutral or positive selection, two were observed in the SA14 alignment (prM=1.073, NS1=1.246), two were observed in SA14-14-2/P1 (E=1.023, NS4B=1.710), and one was observed in SA14-14-2/P2 (NS4B=1.871). When analyzed using a 20-codon sliding window, regions of the JEV genome were



observed to have values greater than 1.000, [Table 7.8, Figure 7.7] indicating the possibility that these particular domains are less conserved in the genotype or are under positive selection. In SA14, windows of high-ranking  $dN/dS$  were clustered in E, NS1, and NS3. In both SA14-14-2 strains, high-ranking  $dN/dS$  values were clustered in the E gene.

## 7.4 Discussion

### 7.4.1 Resolution of Consensus for SA14 and SA14-14-2 Strains Using Limited Sequence Information

Resolution of the genotypes of SA14 and S14-14-2 are complicated by both a lack of publicly available sequences and dissimilarity of passage history for those reported strains. When comparing the SA14 strain sequenced in this study to those publicly available, very limited divergence was observed between SA14 and the two most recently sequenced strains of SA14: Genbank accession D90194 (5 amino acid differences) (Aihara et al. 1991) and Genbank accession JEU14163 (4 amino acid differences) [Table 7.1]. Greater divergence was observed in reference to Genbank accession M55506 (11 amino acid differences) (Nitayaphan et al. 1990). Because the SA14 sequence used in this study was a single passage from a low-passage institutional collection stock, it is likely to closely represent the originating SA14 genotype, and therefore is an appropriate reference by which to infer the genotype of SA14-14-2 vaccine strains. SA14-14-2/P1 and SA14-14-2/P2 differ by a single passage in Vero cells, and were identical in reference to the SA14 [Table 7.5]. The observed substitutions confirmed, with limited exceptions, the genotype of the vaccine as reported by Gromowski et. al., in which a infectious clone of SA14-14-2 was constructed and used to investigate vaccine determinants of mouse neurovirulence, however, the SA14-14-2 genotype reported in this study does not contain substitutions in either untranslated region of the virus genome as previously reported in reference to a recombinant in-

fectious clone of WT strain India/78 (Gromowski, Firestone, Bustos-Arriaga, et al. 2015; Gromowski, Firestone, and Whitehead 2015). Additionally, the amino acid substitutions E-E244G, E-A315V, E-K439R, NS1-G292S, NS1-R339M, NS1-D351H, NS2A-N2K, NS2A-A40V and NS5-Q639H, conserved across SA14-14-2 strains and were not previously hypothesized to represent the vaccine genotype. Because these substitutions were observed relative to a low-passage stock of the parent SA14, it is plausible that the substitutions represent a more accurate SA14-14-2 genotype than previously investigated.

#### **7.4.2 Previous Investigations of Neurovirulence Determinants**

Neurovirulence studies on discrete genomic determinants of SA14-14-2 attenuation have been investigated by infectious clone systems, although very limited work of this type has been reported. In an experiment to investigate the stability of the vaccine, SA14-14-2 was passaged 11 times in suckling mouse brain, and this increased the neurovirulence phenotype of the strain for mice (D. Yang et al. 2014). In the same study, Yang et al. observed by plaque picking and conventional Sanger sequencing a set of intermediate, quasi-genotypes that were selected during in mouse brain passage, including the WT revertants E-F107L, E-K138E, E-G244E (D. Yang et al. 2014). Construction of an infectious clone of WT JEV strain JaOArS982 was reported by Sumiyoshi et. al, in which the envelope substitution E-F107L was observed to reduce mouse neurovirulence (Sumiyoshi, Hoke, and Trent 1992; Sumiyoshi, Tignor, and Shope 1995). Residues E-107L and E-138E were also observed by Gromowski et al. to be potential determinants of attenuation for mouse neurovirulence in a study using a JEV infectious clone system to study each whole-gene set of vaccine substitutions *en masse*, although in this study the substitutions were not assayed individually (Gromowski, Firestone, and Whitehead 2015). Significantly, this study reported that the entire envelope genotype was sufficient to

attenuate the mouse neurovirulence phenotype for the vaccine.

#### **7.4.3 Features of Stability and Low Diversity for SA14-14-2 Vaccine Strains**

*In silico* techniques were previously used to compare the YFV WT Asibi strain with live-attenuated vaccine strain 17D [Chapter 3]. These studies showed that WT Asibi had a typical quasispecies structure of a RNA virus while 17D vaccine had a relatively homogeneous population and it was hypothesized that the lack of diversity contributes to the attenuated phenotype of the 17D vaccine (A. Beck et al. 2014). This paradigm for live attenuated RNA viruses was hypothesized to be conserved for serially-passaged viral vaccines, specifically WT JEV strain SA14 and its live attenuated vaccine derivative SA14-14-2. Although live attenuated SA14-14-2 has been found to be both effective and safe, derivation of the vaccine utilized a complex passage process in multiple cell types and hosts. SA14-14-2 was derived from SA14-5-3, a strain that was attenuated but not sufficiently immunogenic in clinical trials [Figure 7.1]. SA14-5-3 was passaged five times in mouse skin, and plaque-picked to generate SA14-14-2; the short adapting passage series produced a strain with increased immunogenicity and seroconversion rates in children (C. H. Huang 1982).

It was hypothesized that the complex selection pressures exerted upon the vaccine by serial passage would be resolvable in the population structure of the vaccine. These hypotheses were confirmed for the JEV strains considered in the study, however diversity effects were highly localized. Low diversity was a general feature of the vaccine strains in comparison to SA14. When restricting the analysis to discrete genes, the effect was most apparent for gene segments in which putative attenuating mutations are clustered, and as such was especially apparent in the envelope protein gene. As expected, loss of diversity is most apparent for the 22

sites shown to carry amino acid substitutions that distinguish SA14 from SA14-14-2. Notably, a restricted set of these sites were observed in which loss of diversity for the vaccine is especially great; in this set are included E-F107L and E-G244E that were previously observed to revert under passage in mouse brain at the population level. A later infectious clone study demonstrated that substitution E-G244E (relative to SA14-14-2) conferred a lethal neurovirulence phenotype in mice for infectious clone derived SA14-14-2 strains carrying the WT glutamic acid residue (Yun et al. 2014). Also notably, both the coding variant populations [Table 7.6] and the diversity profiles [Tables 7.3, 7.4, 7.5] were similar between SA14-14-2/P1 and SA14-14-2/P2, indicating negligible fluctuation in population frequencies between the two vaccine strains on single Vero cell passage [Figure 7.1]. These findings are consistent with those of Yang et al., in which Vero passage did not appear to perturb the phenotype of the vaccine (D. Yang et al. 2014).

#### **7.4.4 Patterns of Selective Pressure Are Typical For Arboviruses**

The findings are consistent with previous comparison of population structure for YFV strains Asibi and 17D-204 (A. Beck et al. 2014). Similarly by consensus, the attenuation of 17D produced a vaccine genotype with a number of amino acid substitutions in the envelope gene (8 of 20 substitutions [Table 1.1]), although the attenuation of YFV is presumed to be influenced by both structural effects in the envelope as well as a phenomenon of IFN antagonism; this has been proposed based on functional studies of both the YFV NS4B gene and envelope residues (E. Lee and Lobigs 2008). By population, diversity in the YFV genome is also localized to the envelope protein in the same proportion as in JEV, with 3 of 9 major coding variants present reported in the Asibi genome [Chapter 3, Table 7.8]. The direct comparison of population structure change for YFV and JEV vaccines is however complicated by the different WT manifestations of YF (viscerotropism) and JE dis-

ease (neurotropism), and as well the different ecological niches in which the viruses are encountered in nature. Selection pressures along SA14 and SA14-14-2 genomes were broadly observed to be negative ( $dN/dS < 1$ ), which is an expected result for arboviruses, and is presumably due to the complex trade-offs in fitness required for the virus to productively infect multiple WT hosts.

However counter-intuitively, some regions were observed to be influenced by positive selection ( $dN/dS > 1$ ); for the vaccines, this was localized to the envelope and NS4B genes [Table 7.8]. It is possible that the signal of selection at these regions shows localization of particular instability for the vaccine genotype under cell culture, however the stability of SA14-14-2 under Vero cell passage has been attested experimentally (D. Yang et al. 2014). Also counter-intuitively, some genes in the vaccine strains were of greater diversity in the vaccine strains than in SA14, however, for amino acid substitutions known to influence the phenotype of SA14-14-2, loss of diversity in the vaccines was readily apparent.

The significance of these persistent subpopulations to the phenotype of the vaccine, if any, is not known. However, cases in which low-frequency viral populations are associated with phenotype in a continuous manner have been studied. A nonclinical characterization YFV 17DD substrain vaccine was performed after adapting the seed virus to CEFs (17D vaccines are produced in embryonated eggs). The resulting vaccine, when tested using a standard monkey neurovirulence assay, was of greater neurovirulence in monkeys by histological assessment, however, this was not statistically significant (M. S. Freire et al. 2005). The consensus sequences were identical, suggestive of a population-level effect.

## 7.5 Conclusions

Previous studies comparing the population structures of another live-attenuated viral vaccine, YFV 17D, revealed that fluctuations in viral subpopulations frequently accompany the complex selection pressures encountered in attenuation by serial passage. It is especially significant that the present findings on JEV are in agreement with clonal sequencing studies of SA14-14-2 that were previously reported. Considerable support is presented in these findings to suggest that changes in phenotype for SA14-14-2 vaccines may be predicted based on features of the viral quasispecies, which in this case were repeatably observed between two vaccine strains. Patterns of diversity change in transition to the SA14-14-2 genotype are localized to regions of the genome previously shown to influence the attenuation of neurovirulence for the vaccine in infectious clone studies. Accordingly, the presented data represents a confirmation of previously observed patterns for YFV, and establishes a framework by which to construct further studies on the pathogenic significance of flaviviral vaccine subpopulations.

**Table 7.1:** Amino acid substitutions differentiating the SA14 strain sequenced in this study to that of three others previously reported.

Nucleotide Position	Gene	Codon	SA14 <sup>a</sup>	M55506 <sup>b</sup>	D90194 <sup>c</sup>	JEU14163 <sup>d</sup>
1708	E	244	E	G	G	-
1921		315	A	V	-	-
1977		334	P	-	S	-
2293		439	K	R	-	-
3351	NS1	292	G	S	-	S
3493		339	R	M	-	-
3539	NS2A	2	N	K	-	-
3652		40	A	V	-	-
3849		106	V	I	-	-
4519	NS2B	102	M	-	-	T
4825	NS3	73	K	R	R	R
5311		235	A	-	V	-
6700	NS4A	79	K	-	-	R
8658	NS5	328	K	E	-	-
9607		644	N	T	-	-
9898		741	G	-	D	-

<sup>a</sup> Sequenced in this study.

<sup>b</sup> (Nitayaphan et al. 1990)

<sup>c</sup> (Aihara et al. 1991)

<sup>d</sup> (Ni, Burns, et al. 1994)

**Table 7.2:** Amino acid and 3'UTR substitutions observed in SA14-14-2 vaccine strains, in reference to the newly-sequenced, WT SA14 strain.

Position	Gene	Codon	SA14 <sup>a</sup>	P1 <sup>a</sup>	P2 <sup>a</sup>	AF315119 <sup>b</sup>	D90195 <sup>c</sup>	JN604986 <sup>d</sup>
292	C	66	L	S	S	S	S	S
1296	E	107	L	F	F	F	F	F
1389		138	E	K	K	K	K	K
1503		176	I	V	V	V	V	V
1506		177	T	A	A	A	A	A
1708		244	E	G	G	G	G	G
1769		264	Q	H	H	H	H	H
1813		279	K	-	-	M	M	M
1921		315	A	V	V	V	V	V
2293		439	K	R	R	R	R	R
2317		447	G	-	-	D	-	-
3351	NS1	292	G	S	S	S	S	S
3493		339	R	M	M	M	M	M
3528		351	D	H	H	H	H	H
3539	NS2A	2	N	K	K	K	K	K
3652		40	A	V	V	V	V	V
4403	NS2B	63	E	D	D	D	D	D
4408		65	D	G	G	G	G	G
4475		87	L	-	-	F	-	-
4782	NS3	59	M	V	V	V	V	V
4921		105	A	G	G	G	G	G
5311		235	A	V	V	V	V	-
5634		343	R	-	-	-	W	-
6634	NS4A	57	I	-	-	-	T	-
7227	NS4B	106	I	V	V	V	V	V
7768	NS5	31	A	-	-	G	-	-
7809		45	R	-	-	S	-	-
8261		195	M	-	-	I	-	-
8832		386	H	Y	Y	Y	Y	Y
9593		639	Q	H	H	H	-	H
9688		671	V	A	A	A	A	A
9954		760	A	-	-	P	-	-
9978		768	L	-	-	V	-	-



10428	3'UTR	-	-	-	-	C	C	C
10574		-	-	-	-	U	-	-
10701		-	-	-	-	-	-	-
10784		-	-	-	-	-	U	-

<sup>a</sup> Sequenced in this study.

<sup>b</sup> Not published.

<sup>c</sup> (Aihara et al. 1991)

<sup>d</sup> (Song et al. 2012)

**Table 7.3:** Diversity summary for SA14, SA14-14-2/P1, and SA14-14-2/P2 using normalized Shannon’s entropy, split by gene segment. Within each segment, strains are sorted by descending order of mean entropy value. n = number of nucleotide positions per segment.

Gene	Strain	n	Mean	Median	S.D.
5’UTR	SA14-14-2/P2	95	0.088	0.060	0.097
	SA14	95	0.087	0.071	0.087
	SA14-14-2/P1	95	0.085	0.067	0.090
C	SA14-14-2/P1	381	0.105	0.100	0.057
	SA14	381	0.104	0.101	0.053
	SA14-14-2/P2	381	0.102	0.089	0.061
prM	SA14-14-2/P1	501	0.103	0.100	0.047
	SA14	501	0.102	0.098	0.048
	SA14-14-2/P2	501	0.099	0.093	0.054
E	SA14-14-2/P1	1500	0.104	0.098	0.054
	SA14	1500	0.103	0.099	0.051
	SA14-14-2/P2	1500	0.100	0.091	0.063
NS1	SA14-14-2/P1	1056	0.103	0.099	0.049
	SA14	1056	0.103	0.100	0.049
	SA14-14-2/P2	1056	0.101	0.096	0.054
NS2A	SA14-14-2/P1	681	0.105	0.101	0.053
	SA14	681	0.104	0.103	0.045
	SA14-14-2/P2	681	0.102	0.098	0.055
NS2B	SA14	393	0.101	0.099	0.049
	SA14-14-2/P1	393	0.098	0.096	0.047
	SA14-14-2/P2	393	0.097	0.091	0.054
NS3	SA14-14-2/P1	1857	0.104	0.099	0.054
	SA14	1857	0.103	0.101	0.050
	SA14-14-2/P2	1857	0.101	0.096	0.057
NS4A	SA14-14-2/P1	447	0.104	0.103	0.049
	SA14	447	0.103	0.101	0.053
	SA14-14-2/P2	447	0.100	0.095	0.055
NS4B	SA14-14-2/P1	765	0.104	0.104	0.052

	SA14	765	0.102	0.103	0.047
	SA14-14-2/P2	765	0.100	0.097	0.058
NS5	SA14-14-2/P1	2718	0.102	0.098	0.049
	SA14	2718	0.102	0.098	0.047
	SA14-14-2/P2	2718	0.100	0.093	0.055
3'UTR	SA14-14-2/P1	582	0.095	0.092	0.074
	SA14	582	0.092	0.095	0.076
	SA14-14-2/P2	582	0.088	0.086	0.081

**Table 7.4:** Summary of diversity by NSE for nucleotide positions bearing SA14-14-2 amino acid substitutions, by gene segment. Within each segment, strains are sorted by descending order of mean entropy value. n = count of nucleotide positions in each segment.

Gene	Strain	n	Mean	Median	S.D.
C	SA14-14-2/P1	1	0.117	0.117	-
	SA14	1	0.088	0.088	-
	SA14-14-2/P2	1	0.084	0.084	-
E	SA14	8	0.169	0.124	0.136
	SA14-14-2/P2	8	0.092	0.096	0.033
	SA14-14-2/P1	8	0.090	0.082	0.033
NS1	SA14-14-2/P1	3	0.120	0.099	0.049
	SA14-14-2/P2	3	0.095	0.069	0.061
	SA14	3	0.053	0.049	0.032
NS2A	SA14	2	0.143	0.143	0.111
	SA14-14-2/P2	2	0.111	0.111	0.091
	SA14-14-2/P1	2	0.101	0.101	0.036
NS2B	SA14-14-2/P1	2	0.155	0.155	0.011
	SA14-14-2/P2	2	0.128	0.128	0.017
	SA14	2	0.117	0.117	0.002
NS3	SA14	3	0.106	0.062	0.076
	SA14-14-2/P2	3	0.093	0.096	0.069
	SA14-14-2/P1	3	0.093	0.095	0.050
NS4B	SA14	1	0.080	0.080	-
	SA14-14-2/P2	1	0.051	0.051	-
	SA14-14-2/P1	1	0.032	0.032	-
NS5	SA14-14-2/P2	3	0.096	0.123	0.078
	SA14-14-2/P1	3	0.075	0.100	0.056
	SA14	3	0.070	0.072	0.063

**Table 7.5:** Summary of diversity by NSE for nucleotide positions with SA14-14-2 amino acid substitutions, by gene segment.

<b>Nucleotide Position</b>	<b>Gene</b>	<b>SA14</b>	<b>SA14-14-2/P1</b>	<b>SA14-14-2/P2</b>
292	C	0.088	0.117	0.084
1296	E	0.137	0.097	0.110
1389		0.062	0.141	0.123
1503		0.087	0.074	0.046
1506		0.175	0.084	0.140
1708		0.489	0.081	0.108
1769		0.099	0.132	0.065
1921		0.112	0.076	0.083
2293		0.193	0.037	0.062
3351	NS1	0.024	0.084	0.069
3493		0.049	0.099	0.051
3528		0.087	0.176	0.164
3539	NS2A	0.065	0.076	0.046
3652		0.222	0.127	0.175
4403	NS2B	0.116	0.147	0.116
4408		0.118	0.162	0.140
4782	NS3	0.195	0.142	0.160
4921		0.062	0.041	0.023
5311		0.062	0.095	0.096
7227	NS4B	0.080	0.032	0.051
8832	NS5	0.131	0.113	0.157
9593		0.006	0.010	0.008
9688		0.072	0.100	0.123

**Table 7.6:** Amino acid variants concordantly observed in both SA14-14-2/P1 and SA14-14-2/P2, relative to the newly sequenced SA14 strain.

Position	Gene	Codon	Nucleotide		Amino Acid		Variant Percent	
			Cons <sup>a</sup>	Var <sup>b</sup>	Cons	Var	SA14-14-2/P1	SA14-14-2/P2
948	M	66	C	U	L	F	1.026	0.8972
1116	E	47	A	G	N	D	22.87	36.18
1453		159	U	A	V	D	10.59	11.38
1512		179	A	G	K	E	36.48	42.28
1911		312	A	G	K	E	1.538	0.9815
2646	NS1	57	G	A	V	I	0.9662	0.386
3652		392	U	G	V	G <sup>c</sup>	0.9572	1.365
3669		398	C	U	L	F	6.045	8.313
3695		406	G	C	L	F	3.907	4.841
5311	NS3	235	U	C	V	A <sup>c</sup>	1.693	1.707
5781		392	G	A	D	N	0.6344	1.04
5835		410	G	A	D	N	11.18	8.654
7167	NS4B	86	G	A	V	M	2.13	3.166
7657		249	A	G	D	G	20.12	15.99
7774	NS5	24	U	C	I	T	1.901	1.767
8791		372	C	U	A	V	1.077	2.365
9292		539	G	C	G	A	0.736	1.795
10069		798	C	U	T	I	0.8492	0.8072
10092		806	G	C	G	R	1.096	0.7716
10329		906	G	A	V	I	2.132	1.709

<sup>a</sup> Consensus nucleotide or residue in the newly computed consensus sequence.

<sup>b</sup> Variant nucleotide or residue in the alignment, which is observed below the consensus frequency.

<sup>c</sup> Variant residue is a revertant to wild-type SA14 genotype.

**Table 7.7:**  $dN/dS$  ratios for each gene segment and complete open reading frame for newly-sequenced SA14, SA14-14-2/P1, and SA14-14-2/P2 strains.

<b>Gene</b>	$dN/dS$		
	<b>SA14</b>	<b>SA14-14-2/P1</b>	<b>SA14-14-2/P2</b>
ORF	0.413	0.474	0.419
Capsid	0.715	0.557	0.417
prM	1.073	0.310	0.329
Envelope	0.778	1.023	0.865
NS1	1.246	0.275	0.285
NS2A	0.856	0.327	0.339
NS2B	0.071	0.515	0.498
NS3	0.501	0.359	0.264
NS4A	0.032	0.403	0.329
NS4B	0.612	1.710	1.871
NS5	0.404	0.227	0.156

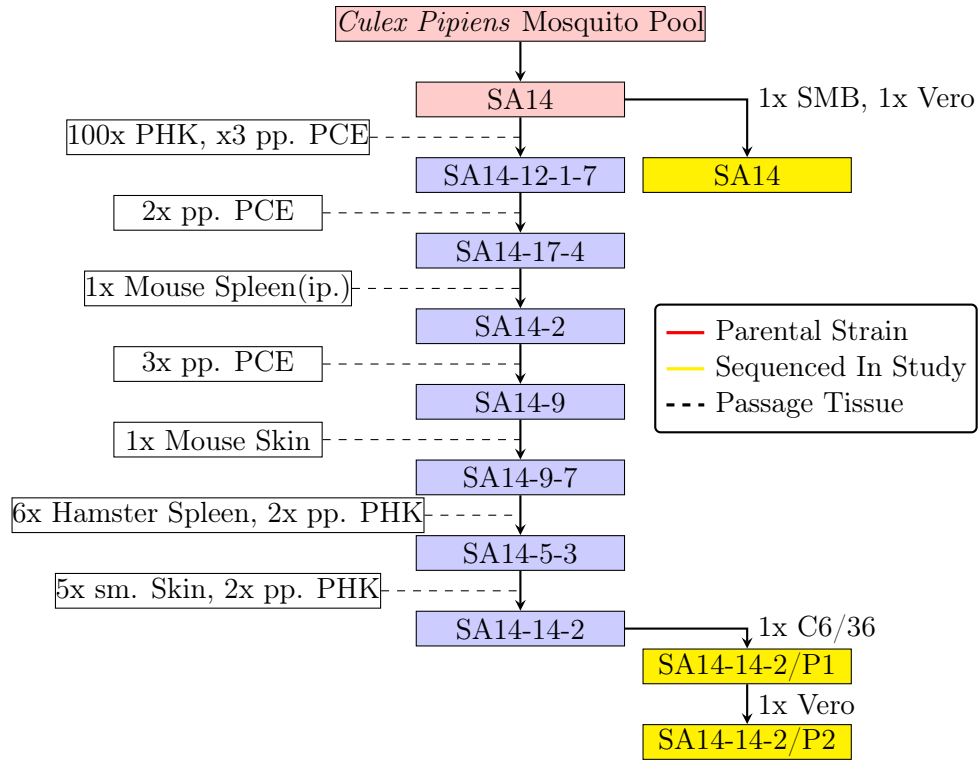
**Table 7.8:** 20-codon sliding windows with high  $dN/dS$  ratios in the open reading frames of SA14, SA14-14-2/P1, and SA14-14-2/P2.

Strain	Gene	Nucleotide Window <sup>a</sup>		Codon Window <sup>b</sup>		$dN/dS$
		Start	End	Start	End	
SA14	Envelope	162	222	23	43	86.494
	Envelope	1068	1128	31	51	74.207
	NS1	1080	1140	35	55	46.638
	NS1	1089	1149	38	58	38.702
	Envelope	1104	1164	43	63	37.590
	Envelope	1116	1176	47	67	35.728
	Envelope	1395	1455	140	160	34.724
	NS1	1401	1461	142	162	28.143
	NS1	1458	1518	161	181	27.996
	NS3	1470	1530	165	185	27.171
SA14-14-2/P1	Envelope	3624	3684	31	51	31.444
	Envelope	3630	3690	33	53	29.072
	Envelope	3639	3699	36	56	21.508
	Envelope	3651	3711	40	60	20.507
	Envelope	3657	3717	42	62	19.234
	Envelope	3669	3729	46	66	10.751
	Envelope	3681	3741	50	70	9.590
	NS4B	3687	3747	52	72	8.647
	NS4B	4290	4350	26	46	8.462
	Envelope	5781	5841	392	412	6.831
SA14-14-2/P2	Envelope	1668	1728	231	251	52.537
	Envelope	1692	1752	239	259	51.716
	Envelope	1704	1764	243	263	48.745
	Envelope	2145	2205	390	410	47.104
	Envelope	2160	2220	395	415	46.162
	Envelope	2919	2979	148	168	45.595
	Envelope	2940	3000	155	175	45.458
	Envelope	3123	3183	216	236	44.278
	Envelope	3132	3192	219	239	20.804
	Envelope	3147	3207	224	244	20.048

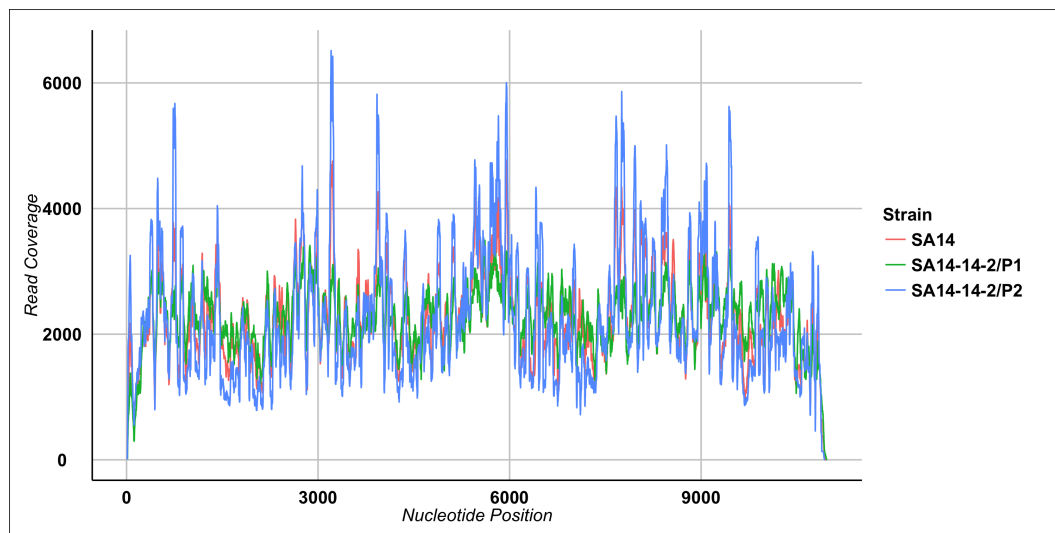
<sup>a</sup> Sixty nucleotides, inclusive.

<sup>b</sup> Twenty residues, inclusive.

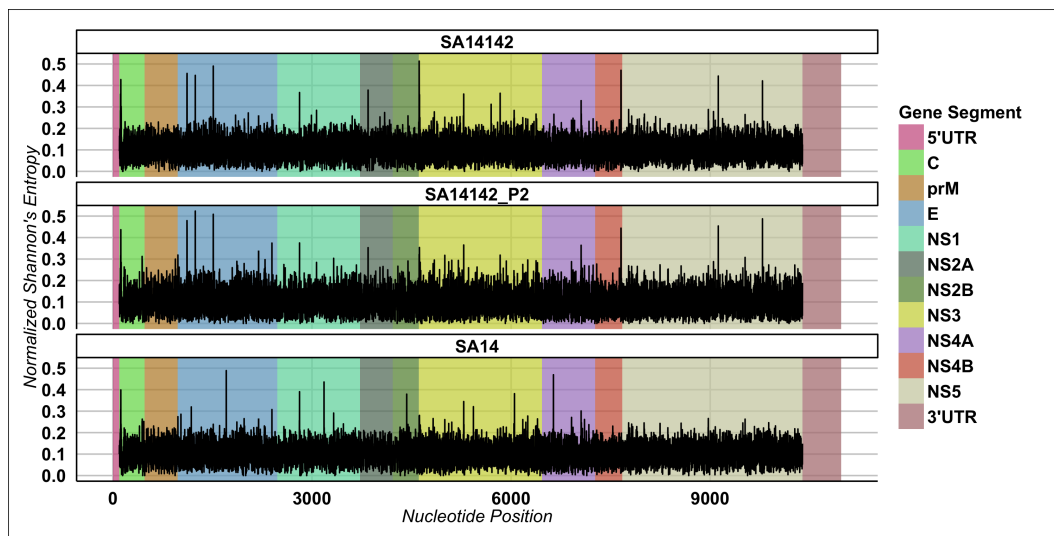




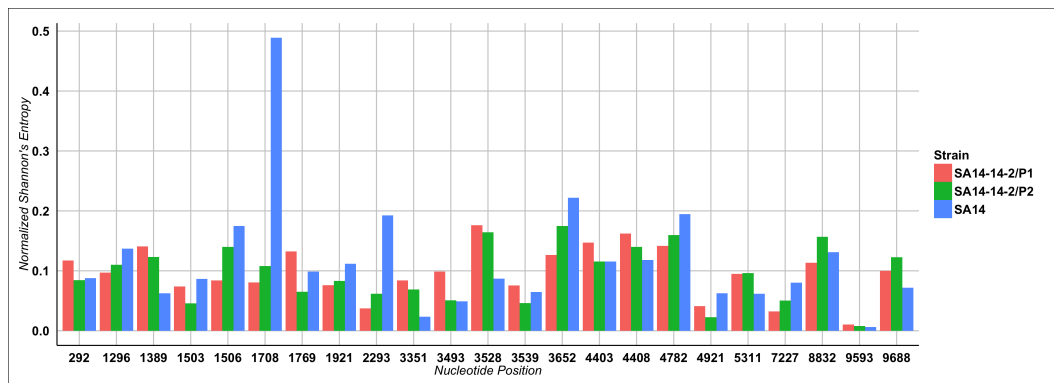
**Figure 7.1:** Derivation of JEV Vaccine, showing the passage history of newly analyzed strains. PCE = Primary Chick Embryo, PHK = Primary Hamster Kidney, pp. = Plaque Pick, SMB = Suckling Mouse Brain.



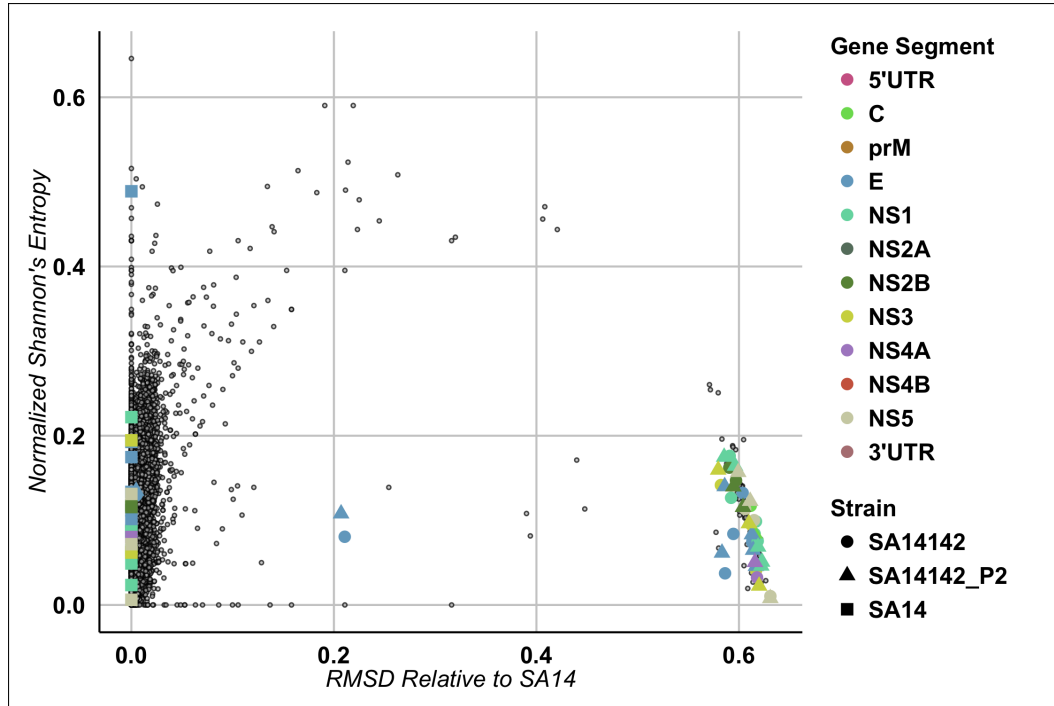
**Figure 7.2:** Read coverage along the length of genomes, for each virus analyzed in the study.



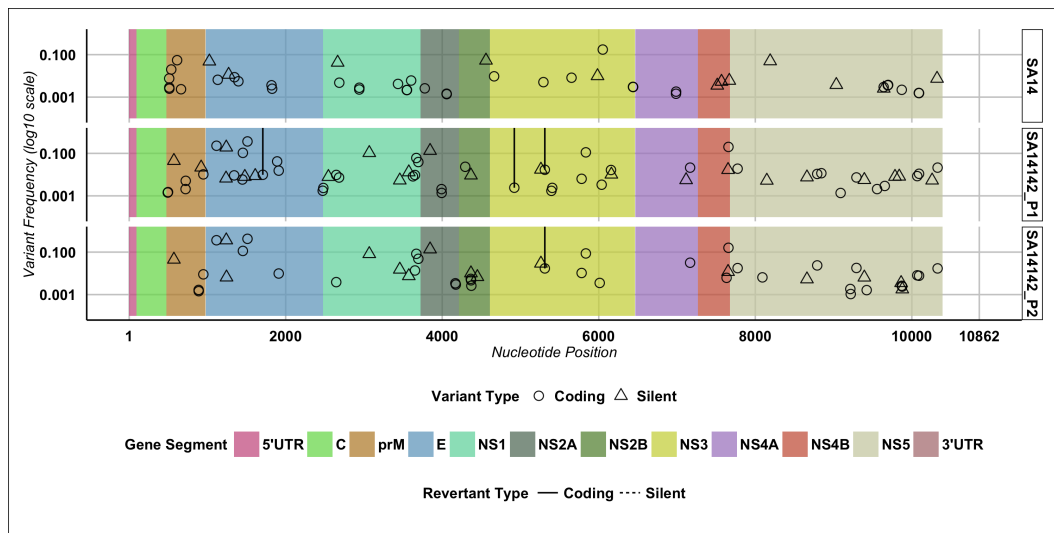
**Figure 7.3:** Sitewise diversity estimation, measured by NSE, for each strain considered in the study. Diversity was estimated at each nucleotide position recovered in the read alignment for the JEV strains SA14, SA14-14-2/P1, and SA14-14-2/P2.



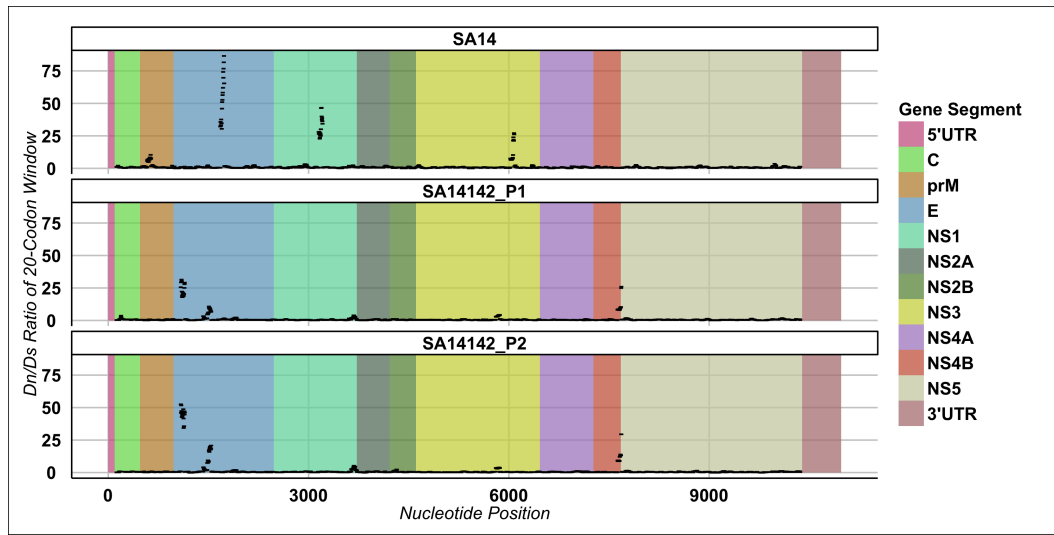
**Figure 7.4:** Summary of diversity, measured by NSE, for each nucleotide position containing an amino acid substitution between the newly sequenced wild-type strain SA14 and vaccine strains SA14-14-2/P1 or SA14-14-2/P2.



**Figure 7.5:** A plot of normalized Shannon's entropy versus RMSD relative to the wild-type parental strain SA14, for each nucleotide position recovered in the three JEV strains. SA14 is shown to control, and are clustered at implied zero distance and relatively higher diversity than the vaccines. Colored points represent the nucleotide positions that with amino acid substitutions in the vaccine strains relative to SA14. Clustering of points these at high genetic distance from the parental strain is shown, and is evidence for selection of the vaccine genotype.



**Figure 7.6:** A plot of the frequency of each SNV recovered from the JEV strains studied by a probabilistic model in V-Phaser v. 2.0, in order of nucleotide position. SNVs are classified as either silent(triangle glyph) or coding (circle glyph) relative to the consensus of the strain considered; as well the SNVs are classified for reversion to the wild-type genotype identified by vertical line segments.



**Figure 7.7:** 20-codon sliding window analysis of  $dN/dS$ , along the open reading frames of JEV strains SA14, SA14-14-2/P1, and SA14-14-2/P2. Broadly, ratios along the length of the JEV genomes are low, indicative of purifying selection and a stable population. A small number of regions are influenced by apparent positive selection, in which  $dN/dS$  exceeds 1.000.

## Chapter 8

### Discussion

#### 8.1 Rationale for Studies Performed in the Dissertation

The studies performed in the dissertation were designed to give insight to the discrete attenuation mechanisms for the YFV 17D strain vaccines, which are poorly understood to date, despite widespread and longstanding use of the vaccine exceeding 600 million doses administered (Monath, M. Gershman, et al. 2013). Very limited work has been performed on this topic by means of reverse genetics systems as these studies are complicated by a lack of immunocompetent small animal models in which to screen putative attenuating alleles for viscerotropic disease. The quasispecies paradigm offer advantages by which to first explore the selection pressures encountered by the WT and vaccine strains of YFV, and by doing so narrowly identify a stable genotype under selection in the vaccine. At the start of this work, viral population structure was completely unexplored for YFV in the laboratory setting, and the few datasets in open literature were of insufficient coverage or quality to observe the deep viral populations at any reasonable level of complexity (Victoria et al. 2010). Studies of viral diversity are frequently presented *ex nihilo*, or without proper biological context for the features or structures described therein. Consequently, the central motivating insight for development of the presented studies was that, by using standardized vaccine strains of well-described properties, the association of standard phenotype with *in silico* datasets is accomplished with rigorous, naturally-arising controls.



Historically, the precedent for a diversity paradigm of YFV predates the widespread use of 17D or French vaccine types. Before standardization of 17D and the creation of the seed-lot system, stability of YFV vaccine strains was frequently investigated by serial passage studies in animals where the properties of viscerotropism and neurotropism were observed to be partially unlinked, meaning that each phenotype could be selected independently by passage *in vivo* (Findlay and Clarke 1935). The low viremia and mild infection caused by 17D was later shown to be associated with restriction of the vaccine virus from tissue dissemination, which is a consequence of innate immunity (Erickson and Pfeiffer 2013); conversely, productive WT YFV infection in nonhuman primates results in high viremia and eventual tropism for the liver (Monath, Brinker, et al. 1981). If tissue tropism and compartment traversal in some manner is associated with WT infection (e.g., viscerotropism and neurotropism phenotypes), the mechanisms underlying this are almost certainly exerted at the sub-population level, meaning that some alleles are carried in the population may occupy divergent fitness niches within a single host or infection event.

Because many of the vaccine strains available in the study arise from controlled or standardized preparations for which precise information is available, a number of open research questions that derive insight from classical vaccinology of YFV or rational understanding of attenuation mechanisms were then rendered amenable to experimentation. The **overall objective** of the studies was to elucidate viral diversity and population structure for the YFV 17D vaccine, and in doing so to provide insight to the mechanism(s) of attenuation of the vaccine. The **central hypothesis** was that specific features of population structure would be observed to differentiate WT YFV from the live-attenuated vaccine strain 17D.

## 8.2 Main Findings and Future Directions

### 8.2.1 Chapter 3

In chapter 3, a deep sequencing study was performed to compare the population structures of the YFV 17D-204 vaccine with the WT parental strain Asibi, from which the vaccine was derived. This study was performed to establish a baseline comparative relationship between the two strains related by precise documentation of serial attenuating passage, under a hypothesis that, besides providing confirmation of the known consensus genotype of the vaccine (which is well-described and used in production standards), sequence subpopulations would be resolved that uniquely classify and differentiate 17D from the parental strain. By a number of measurements of diversity and by specific sequence content, this hypothesis was supported by the dataset. First, when variant subpopulations in the strains were analyzed, greater numbers of SNVs were apparent in the WT strain, and resolved SNVs were observed in the parental strain in overall greater frequency than those in the vaccine. Second, a subset of identified variant populations were observed in the Asibi strain bearing overlapping (intersecting) identity with the vaccine whereas the converse was not observed in 17D, indicating that it is possible that some portion of the vaccine genotype is maintained naturally in the WT virus, either by random effects or by selection in nature. Third, these patterns were observed by numerous independent measurements of diversity for the two strains. In the light of these findings, it was considered likely that a lack of diversity and attendant lack of subpopulation content would be a typical feature of live-attenuated YFV vaccine strains, and in theory other live-attenuated viral vaccines would contain similar features. Some of these questions were addressed in subsequent chapters, especially those of repeatability for subpopulations observed in the vaccines. Selection pressures maintaining the apparent population structure of Asibi in the preparation used (or of other WT

strains) were not addressed; Asibi was chosen exclusively by virtue of the known and direct relationship between this strain and the vaccine derivatives, in addition to the low passage history for the isolate as maintained in an institutional collection (WRCEVA, UTMB). However, further the investigation of these features across a wide panel of WT YFV isolates would be of great use to establish a rigorous baseline for WT YFV diversity; this would ideally be performed in such a manner as to associate viral populations with particular host ecologies, tissue tropisms, or disease outcomes. The principal context of this study was, however, only limited to the properties of the vaccine.

### 8.2.2 Chapter 4

In chapter 4, a deep sequencing study was performed to resolve the population structures of a set of phenotypically stable 17D vaccine strains, in this case a set of rare isolates for the YFV vaccine 17D. The panel of strains used in the study represents the most complete assembly of genetic relationships for the 17D lineage to date. Lineage and passage relationships for 17D are well described, forming the basis of a seed-lot system that is used to regulate the manufacture of vaccine lots. Although some previous efforts have been undertaken to describe the divergence of 17D vaccines, to this point this has been performed using a combination of consensus sequencing, Bayesian phylogenetics, and classical, *in vitro* methods such as plaque size, or classification of vaccine-specific epitope content (Barban et al. 2007; Stock et al. 2012). 17D is considered highly stable both *in vitro* and *in vivo* methods, and is (with very few benign exceptions between substrains) phenotypically stable with respect to immunogenicity and reactogenicity. Reasonably, it was hypothesized that the previously observed features of stability would be associated with stable population structure of the strains considered. This hypothesis was supported by several findings. First, the diversity profiles of the vaccines were largely similar; sta-

tistical divergence of strain diversity was typically observed for strains with distant relationships along the passage lineage, an expected result that is suggestive of some marginal genetic drift of the vaccine; selection analysis by  $dN/dS$  revealed the expected pattern of broad negative selection across the vaccine genomes, meaning that marginal differences in subpopulation diversity are most likely attributed to random effects at noncoding sites. Second, the genetic divergence was computed based on the relative frequency of nucleotides in the alignments for each vaccine strain (RMSD); however, the use of RMSD to generate a fine-scale pairwise divergence measurement did not result in the complete resolution of a typical distance phylogeny for the vaccines, a result that was roughly similar to that produced by consensus sequences. Third, the variant populations of the vaccines and seeds contained marginal levels of WT content at low frequencies, which in some cases was observed in repeatable patterns between closely related strains. Fourth, one secondary seed strain (NDCs 6676), while included in the comparative analyses, was never used for manufacture of vaccine, as the seed failed a monkey neurovirulence test (Tauraso, Spector, Jahnes, et al. 1968; Monath, Kinney, et al. 1983). This particular strain contained dissimilar population content in reference to the others, in particularly sister strain 6677, with greater evidence of positive selection; although the consensus sequence of this seed strain was unremarkable, this is likely evidence of mishandling leading to subpopulation divergence of sequence content. In this light of these findings, the presented data represents a reasonable framework by which low-level sequence differences of newly produced 17D vaccine seeds or lots could be assessed against a dataset representing the full lineage of standard strains. Reverse genetics of the particular SNVs in the 6676 strain populations would be useful to understand the contribution of quasispecies to monkey neurovirulence. Similarly, chimeric Asibi/17D viruses generated by reverse genetics to examine the influence of one or more of the amino acid substitutions that separate Asibi and 17D would provide excellent material to

investigate the localized affects of diversity due to specific amino acid substitutions. In particular, these studies would be motivated by a reasonable hypothesis that the rates of nucleotide error afforded by the YFV polymerase are a significant contributor to the population dynamics of YFV, and so mutagenesis studies on the NSs5 protein would likely be fruitful.

### 8.2.3 Chapter 5

In chapter 5, a deep sequencing study was performed using a panel of live attenuated vaccines that were developed and fielded in parallel with the 17D strain aqueous preparations. These “French neurotropic” strains were developed by the Institut Pasteur (Dakar, Senegal) and widely used in West Africa between 1940 and 1980; they were ultimately discontinued by the Institut Pasteur for reasons of safety. *In silico* methods were used to investigate competing hypotheses on the mechanisms influencing the adverse neurotropic phenotype of the vaccine. Significantly, the strains had been previously assayed for neurovirulence in both mice and monkeys; the strains were unequally neurovirulent in both model systems. Primarily, the results show that for FNV, high neurovirulence *in vivo* is associated with a homogenous viral subpopulation structure. Strains with low neurovirulence exhibited complex populations, with considerable WT sequence content by consensus and by variant subpopulation. As presented, the findings represent the first analysis of population structure for a set of vaccine strains known to have an unacceptable safety profile. Although some previous efforts had indicated that highly virulent strains would exhibit a greater level of diversity relative to poorly virulent strains, the converse was observed in this study. In the light of these findings, it is likely that the neurotropic adverse reactions caused by FNV were the result of the initial adaptation and production of the vaccine in mouse brain, leading to greater propensity of the vaccine strain to enter or productively infect human central nervous tissues. Also

most significantly, it is demonstrated that subtle divergence of population structure in YFV vaccine preparations may be associated as a continuous quantity with other key measurements of phenotype. Further work will parse the effects of discrete population destabilization for FNV, leading to a definitive model of attenuation for YFV in nervous tissue.

#### 8.2.4 Chapter 6

In chapter 6, a deep sequencing study was performed to compare the prototype strain Asibi (identical to that used in chapter 3) to that of a strain that was adapted from Asibi by serial passage to productively infect Syrian golden hamsters, causing viscerotropic disease. The adapted strain, called Hamster/P7, was previously reported and as such a consensus sequence comparison of Asibi and Hamster/P7 was already available in open literature, along with histological confirmation of disease outcomes in the model (McArthur et al. 2003). Evidence of selection for the genotype of Hamster/P7 was demonstrated by reduction in diversity and specific identifying variant populations that differentiate input and output strains. In this light of these findings, it is substantiated that the viscerotropic pathogenicity of WT YFV follows an “iceberg” paradigm, in which YFV and other arbovirus are thought to cause fulminant disease in only a small minority of all infections. In this context, there is support offered for the “trade-off” hypothesis, in which a virus that inhabits a particularly complex ecological niche with multiple hosts is poorly adapted to infect any of the hosts encountered in WT transmission. The short [seven] passage series required to produce a pathogenic strain in hamsters substantiates the likelihood that the viscerotropic genotype exists as a naturally occurring subpopulation in the Asibi strain, as opposed to one arising from purely random effects in the host. In further work, the results of this study will be compared to another adapted models arising from the Asibi strain, especially that of an attenuated strain resulting from passage

of Asibi in Hela cells, to provide a common reference point for phenotype divergence (Dunster, H. Wang, et al. 1999).

### 8.2.5 Chapter 7

In chapter 7, a deep sequencing study was performed using several strains of JEV, in which a comparison was made between the live-attenuated vaccine strain SA14-14-2 and the WT parental strain SA14 from which the vaccine was derived by serial passage in PHK cells. The intent for this study was to attempt translation of the findings of chapter 3 into another flavivirus vaccine; in doing so the studies would establish that diversity effects are a common feature in live-attenuated, positive-stranded, RNA virus vaccines. Although the attenuation paths of YFV 17D and JEV SA14-14-2 differ with respect to pathophysiology and tissue tropism, several similarities to YFV were apparent in the JEV dataset. First, SA14-14-2 was of lower diversity than SA14, in the same manner that 17D is low diversity relative to the Asibi strain. Second, the consensus sequences observed for two vaccine strains analyzed were similar, but not identical to, previously reported genotypes for SA14-14-2; it is important to note that there is some dissimilarity between the few strains of both SA14 and SA14-14-2 reported in public databases. Third, variant subpopulations of two SA14-14-2 strains were very similar, despite a difference of once passage in cell culture. Fourth, changes in the population structure in transition from SA14 to the SA14-14-2 vaccine genotype were highly localized; this suggests a similar profile of stability to YFV 17D on limited passage. Changes in both diversity and subpopulation frequencies were clustered almost exclusively in the envelope gene of the vaccine; this substantiates previously reported results that used mutagenesis techniques to demonstrate that envelope residues contain the predominant attenuation determinants of SA14-14-2 (Gromowski, Firestone, and Whitehead 2015; Yun et al. 2014; Sumiyoshi, Tignor, and Shope 1995). This differs from the governing

paradigm of attenuation for YFV 17D, in which the effects are considered to originate with significant mechanistic influence from residues in multiple gene segments.

### **8.3 Critical Methods Analysis for Strategies Used in the Dissertation**

#### **8.3.1 RT-PCR and Isolation of Vaccine Nucleic Acids**

To the greatest extent possible, YFV strains used in the dissertation studies were not passaged. The intent for this was to preserve the resolvable population structures of vaccine strains without the application of undue selection pressure and consequent alterations to nucleotide frequency. Thus in many cases, sequencing was performed from RT-PCR amplicons. In such cases, RT-PCR amplification may indeed introduce mutational biases, either by mutagenesis or by amplification of (assumed by primer design) majority populations defined by the oligonucleotide set. Several controls were used in the *in silico* procedures to limit the effects of potential PCR biases, and are a typical feature of variant calling procedures. First, the techniques and programs used to identify subpopulations themselves used methods designed to limit these potential biases at they would appear in final results to process downstream in the pipeline. Two of these control methods were used in the dissertation. For example in chapter 3, a heuristic variant analysis is presented [Table 3.3] for YFV strains Asibi and 17D-204, meaning that variant frequencies are retained from the raw percentage of the called nucleotide in the read alignment. In this method, variants are excluded from the final result if biased by greater than a 90 percent divergence in frequency between forward and reverse reads; this procedure offers some protection against both enhancement of particular nucleotides by specific priming and biases in the pool of random hexamers used in the sequencing chemistry itself. Second, it is significant that in chapters for which the working hypothesis was one of population stability between strains, repeatability of variant structure and



WT content between individual strains was very frequently observed. In chapter 7, JEV strains separated by a single *in vitro* passage were observed to have roughly identical major coding variant populations, meaning that in some cases, SNVs are expected to persist under experimental conditions. In chapters 4, 5, 6, and 7, variants are called using a probabilistic criterion in which nucleotide frequencies are resolved on forward and reverse strands separately, and then the frequencies on opposing strands are tested for statistical evidence that the opposing strands (Fisher’s test, false discovery rate). Further work would ideally involve the deployment of some type of nucleic acid purification to remove the complex egg matrix and later permit direct interrogation of the vRNA without intervening RT–PCR amplification.

### 8.3.2 Diversity Indices

The use of diversity indices is a common method of data exploration in NGS studies (Nishijima et al. 2012; A. Beck et al. 2014). In the described findings, SE (Shannon’s index) was deployed to visualize and resolve areas of the viral genomes that may be of greater complexity than others, thus warranting closer inspection. In many of the direct comparisons of strains, the index was useful to estimate raw stability over a passage series, and was especially useful to differentiate strains of divergent phenotype (chapters 3, 5, 6, 7). The significance of a diversity measurement is limited to the extent that the estimate is actually a proxy for variant content, and so the index is strictly meaningful in reference to resolution of the subpopulation genotype contributing the diversity effect at a particular nucleotide position. Typically this value is aggregated across some desired length of the genome in question, or enumerated for a single nucleotide position. In chapter 3, the use of SE permitted an initial visualization of the Asibi and 17D-204 strain genomes, which revealed that greater number of variant populations were present in the WT strain. In chapter 4, diversity was compared between a number of 17D vaccine strains, recovering limited

statistical evidence for divergence of the vaccines. It is significant that in chapter 7, in which the JEV strains were analyzed from vRNAs directly (unbiased analyses), patterns of low diversity that were hypothesized in reference to the findings in YFV datasets were similarly observed. The chapter 7 results are especially significant to question of whether or not 17D and SA14-14-2 are of similar quasispecies structure; the results for the JEV dataset indicate that diversity effects may be extremely localized, or may involve a very small set of nucleotide sites bearing pathogenically significant nucleotide content.

### 8.3.3 *In Silico* Methods for Resolution of Population Structure

As mentioned previously, two methods of variant identification were used, with variant calls performed exclusively by probabilistic methods in later chapters. The selected model (V-Phaser v. 2.0) was originally published in a collaboration of developers at the Broad Institute (Cambridge, MA) and Colorado State University (X. Yang et al. 2013). Because the initial validation of the program took place using a set of West Nile virus isolates (Family: *Flaviviridae*, genus: *flavivirus*) it was thought that the accuracy of the model would easily translate to a use case with YFV, which is identically classified at the genus level. A typical control on the *in silico* procedures would be to perform a plaque-pick followed by Sanger sequencing of the resultant clones, however the usefulness of this procedure would be extremely limited as in the case of the most stable vaccine strains considered in the dissertation (chapter 4), the typical frequencies of rare variants observed in the viral population are relatively low, in such would frustrate attempts to resolve the population clonally. Also, neither heuristic nor probabilistic methods for variant calling as implemented in the studies were designed to resolve linkage between variant populations, meaning the observation of SNVs as simultaneous arising from the same template strand is not implied by any analysis in the dissertation. Later work would ideally elucidate

the groups of SNVs that may arise from the same template by *in vitro* methods, but models to reconstruct template linkage from NGS data are not mature.

#### 8.3.4 Selection Analyses

Selection was estimated by a  $dN/dS$  ratio, which was calculated using a custom Python script, referring to the technique of Morelli (Morelli et al. 2013). For the initial estimation of the possible mutations available,  $N$  and  $S$  were derived from the consensus sequence of the alignment. This technique was not used in every chapter, but was used in chronologically later efforts, following development of the analysis script. The most significant patterns arising from the use of the  $dN/dS$  ratio in the dissertation were foremost that the YFV strains analyzed (chapter 3, 5, 6, 7), summary results for the ratio were typically found to not exceed 1.000, meaning that the genome in question is considered to be predominantly influenced by negative (purifying) selection. This pattern was observed at multiple scales in the YFV vaccine and parental genomes, across the entire open reading frames of the samples, across discrete genes, and in local patterns resolved by 20-codon sliding window analyses. Purifying selection is a biologically plausible observation, which is frequently observed phylogenetically for YFV and other arboviruses (Schuh, Tesh, and A. D. T. Barrett 2011; Holmes 2003). Some cases were observed in which  $dN/dS$  was  $>1.000$ , and this was typically localized, as in the case of both JEV SA14-14-2/P1 and SA14-14-2/P2 (chapter 7) in which  $dN/dS$  of NS4B was 1.710 and 1.871, respectively. Also in chapter 7, short windows in the envelope genes of the two vaccine strains were of high  $dN/dS$ , indicating that some level of positive selection influences the vaccine genotype; it is not known as to whether this represents an essential instability in the vaccine, but passage-stability tests suggest this may not be cause for concern (D. Yang et al. 2014). One potential problem in the ratio generation is that it relies on the accuracy of the called consensus, and in a naïve

manner refers to this sequence as the true, underlying population from which the analyzed strain was selected; the selection of vaccine genotypes from a completely homogenous underlying population is not likely to occur in any WT strain handled for attenuating serial passage in the laboratory. It is significant to observe that For later work, it is likely that by constructing a model of population for multiple, related WT strains, the selection pressures influencing the population genotypes of derived vaccines would be rendered in greater accuracy, however this has not been attempted in the dissertation lab or by any other group.

## 8.4 Statistical Analyses

In chapters 3, 4, 5, 6, and 7, diversity indices are compared for divergence nonparametrically, usually by a Kruskal-Wallis/Mann-Whitney test or equivalent. The rationale behind the choice of test is that entropy values are not normally distributed, instead taking a positive-bounded, long-tailed shape similar in appearance to either the gamma or lognormal distributions. This pattern is likely to arise from a complex process, that involves both the inherent noise of the sequencing chemistry combined with the true diversity effects that are being resolved. Other diversity measurements were deployed in the dissertation to provide confirmation to entropy profiles. In chapter 3, Asibi and 17D-204 strains were triply analyzed by entropy, Simpson's  $1 - D$  (a more stringent diversity index), and error rate. In chapters 3 and 5, error rates for the strains were compared by quantile-quantile plots, which in some cases revealed divergence of error between strains of differing phenotype. It chapters using multiple independent measurements to confirm diversity patterns, it is significant that techniques were typically found to be in agreement. In further work, it will be appropriate to develop a parametric test that reliably compares diversity profiles between NGS datasets, but this is rarely attempted and the statistical tools

available from academic sources are largely focused on resolution of variant structure. In further work, it would be highly useful to attempt a continuous regression of diversity profiles against some other continuous variable, e.g. plaque titer, treatment parameter, or lethality endpoint in an animal model.

## **8.5 Experimental Issues in Flavivirus Vaccinology That May Benefit From A Diversity Framework**

A new and growing body of literature on the topic of viral diversity, while compelling, represents significant challenges to experimental control, namely the presentation of defined viral populations to experimental systems. Studies of this type are typically constructed in such a manner as to assay viral populations when perturbed by some external selection pressure, e.g. pharmaceutical treatment in the host, alternating passage *in vitro*, or comparative rational mutagenesis of an infectious clone strain. Great difficulty arises in the preparation of controlled viral populations for presentation to an experimental system, whether *in vitro* or *in vivo*. A number of findings suggest that *in vivo* population effects would reveal tissue compartmentalization, which is likely for WT YFV based on the unnatural and widespread dissemination achieved by the virus when innate immunity is ablated (Erickson and Pfeiffer 2013). In order to reduce these findings to the fewest number of variables possible, the ideal model system would test an *in vivo* endpoint against a mixed population of a single hypothetical allele. This is sometimes referred to as a “selection” test, in which a single allele with split population frequency is introduced to an experimental system in order to measure selection of allele frequency conferred by that particular system.

### **8.5.1 Reconciling Population-Level Effects With Typical Assessments of YFV Vaccine Attenuation**

Current international technical specifications specify that YFV vaccine seeds be clinically qualified before use in manufacture; this testing may include comparative immunogenicity and safety studies in human subjects. In addition to the clinical qualification, technical specifications indicate that nonclinical qualification of seeds may take place with a diversity of methods, which presumably would admit the utility of NGS to resolve the sequence or population structure of a production seed in development, in conjunction with the typical use of the standard monkey neurovirulence assay (World Health Organization 2010). The findings of these studies indicate a high level of utility for NGS analyses in determining the stability of key attenuating features for these vaccines before manufacturers encounter the expense of further qualification. The results of chapters 3, 4, 5, and 7, particularly indicate that with further effort, NGS would be useful in the prediction of viral phenotype to ensure highest safety of the vaccine before introduction to living systems.

## Appendix

## Appendix 1

### Extended Variant Tables

All tables contained below are extended lists of SNVs, above that which is shown in the chapter text. Data includes both silent and coding SNVs.



**Table 1.1:** Complete variants from the WT Asibi parental strain as an extension of the data shown in Table 3.1. Variants are classified as coding or silent in reference to the newly-computed consensus sequence of Asibi. A 1.00 percent frequency cutoff heuristic is used.

Position	Base	Variant Base	Polyprotein Codon	Gene	Local Codon	Reference Amino Acid	Variant Amino Acid	Total Coverage	Variant Coverage	% Var
434	C	U	106	C	106	L	-	7748	103	1.33
620	U	C	168	prM	47	Y	H	8174	129	1.59
949	C	U	277	prM	156	V	-	7767	103	1.33
964	G	A	282	prM	161	P	-	7058	143	2.03
991	U	C	291	E	7	I	-	6522	100	1.54
1000	G	A <sup>a</sup>	294	E	9	R	-	6191	70	1.13
1054	A	C <sup>a</sup>	312	E	27	Q	H	6743	656	9.74
1195	C	U	359	E	74	C	-	6440	66	1.03
1345	C	U	409	E	124	S	-	6879	677	9.85
1729	C	U	537	E	252	G	-	5852	143	2.44
1819	C	U <sup>c</sup>	567	E	282	S	-	9057	2529	27.9
1993	U	C	625	E	340	L	-	2699	55	2.04
2061	U	C	648	E	363	V	A	3065	39	1.27
2193	C	U	692	E	407	A	V	2640	346	13.1
2349	C	U	744	E	459	A	V	2985	121	4.07
2356	C	U	746	E	461	L	-	3746	617	16.5
2407	C	U	763	E	478	I	-	2604	109	4.21
2481	A	G	788	NS1	10	K	R	3356	420	12.5
2791	U	C	891	NS1	113	Y	-	2924	48	1.67
2803	U	C	895	NS1	117	T	-	2638	30	1.15
3171	A	C	1018	NS1	240	E	A	2440	56	2.32
3274	A	G <sup>ac</sup>	1052	NS1	274	E	-	4007	455	11.3
3622	G	A <sup>b</sup>	1168	NS2A	38	L	-	7795	978	12.5
3817	A	G <sup>c</sup>	1233	NS2A	103	V	-	6532	1475	22.5
3886	A	U	1256	NS2A	126	L	-	6966	171	2.45
3925	A	U	1269	NS2A	139	V	-	7132	1831	25.6
4081	U	C	1321	NS2A	191	T	-	8972	1317	14.6
4177	A	G	1353	NS2A	223	R	-	7048	102	1.45
4416	G	A	1433	NS2B	79	S	N	6477	195	3.01
4474	G	A	1452	NS2B	98	V	-	6406	119	1.86
4591	U	C	1491	NS3	7	D	-	7506	1147	15.2

4732	C	U	1538	NS3	54	V	-	7599	235	3.10
4864	A	G <sup>ac</sup>	1582	NS3	98	Q	-	8192	2120	25.8
4966	C	U	1616	NS3	132	D	-	7038	83	1.19
4993	U	C	1625	NS3	121	P	-	7160	72	1.02
5059	C	U	1647	NS3	163	S	-	6876	633	9.21
5153	A	G <sup>c</sup>	1679	NS3	195	I	V	6042	103	1.70
5248	G	A	1710	NS3	226	L	-	5170	71	1.38
5743	C	U	1875	NS3	391	N	-	4611	53	1.17
5760	G	A	1881	NS3	397	R	K	4590	47	1.03
6886	C	U	2256	NS4A	149	A	-	5629	94	1.67
7066	C	U	2316	NS4B	60	V	-	8728	88	1.01
7175	C	U	2353	NS4B	97	L	-	9027	154	1.71
7211	C	U	2365	NS4B	109	L	-	9352	104	1.11
7411	C	U	2431	NS4B	175	L	-	4805	56	1.18
7642	U	C <sup>bc</sup>	2508	NS5	2	S	-	2320	734	31.6
7645	G	A	2509	NS5	3	A	-	2276	36	1.62
7651	A	G	2511	NS5	5	G	-	2391	32	1.35
7750	U	C	2544	NS5	38	D	-	3719	39	1.07
8354	C	U	2746	NS5	240	L	-	3909	40	1.04
8734	U	C	2872	NS5	366	D	-	3378	92	2.75
8917	C	U	2933	NS5	427	V	-	7692	500	6.51
9410	A	U	3098	NS5	592	M	L	4246	55	1.31
9716	G	A	3200	NS5	692	D	N	3918	118	3.02
10432	U	A	-	3'UTR	NA	-	-	3155	81	2.59
10504	A	G	-	3'UTR	NA	-	-	3016	31	1.05
10800	G	A <sup>c</sup>	-	3'UTR	NA	-	-	1578	310	19.7

<sup>a</sup> Intersection of sequence with the HeLa/P6 strain.

<sup>b</sup> Intersection of sequence with FNV.

<sup>c</sup> Intersection of sequence with the 17D-204 strain.

**Table 1.2:** Complete variants from the 17D vaccine strain as an extension of the data shown in Table 3.2. Variants are classified as coding or silent in reference to the newly-computed consensus sequence of 17D. A 1.00 percent frequency cutoff heuristic is used.

Position	Base	Variant Base	Polyprotein Codon	Gene	Local Codon	Reference Amino Acid	Variant Amino Acid	Total Coverage	Variant Coverage	% Var
263	U	C	49	C	49	F	L	4591	46	1.00
1441	U	C	441	E	156	I	-	9264	164	1.77
1463	C	U	449	E	164	L	-	7518	106	1.41
1990	U	C	624	E	339	D	-	3664	38	1.04
2023	U	C	635	E	350	V	-	6682	124	1.86
3027	A	G	970	NS1	192	K	R	5476	60	1.10
4054	U	C <sup>a</sup>	1312	NS2A	182	N	-	5434	74	1.36
4708	G	A	1530	NS3	46	G	-	4768	461	9.67
4873	G	U <sup>a</sup>	1585	NS3	103	A	-	3282	53	1.61
4891	C	U	1591	NS3	107	N	-	3515	48	1.37
5153	G	A <sup>a</sup>	1679	NS3	195	V	I	5057	72	1.42
5197	C	A	1693	NS3	209	L	-	5025	62	1.23
5290	U	C	1724	NS3	240	F	-	4327	70	1.62
5362	U	C	1748	NS3	264	A	-	6363	185	2.91
5765	U	C	1883	NS3	399	Y	H	3073	31	1.01
5869	U	C	1917	NS3	433	A	-	4076	52	1.28
5872	U	C	1918	NS3	434	F	-	4747	66	1.39
6022	A	G	1968	NS3	484	E	-	4837	148	3.06
6418	U	C	2100	NS3	616	I	-	4254	53	1.25
6916	A	G	2266	NS4B	10	K	-	2828	34	1.20
6943	A	G	2275	NS4B	19	L	-	5064	126	2.49
7496	C	U	2460	NS4B	204	L	-	3914	107	2.73
8578	C	U	2820	NS5	312	T	-	3951	55	1.39
8773	C	U	2885	NS5	379	N	-	1531	33	2.16
8803	A	G	2895	NS5	389	E	-	3251	86	2.65
9093	A	G	2992	NS5	486	E	G	2969	40	1.35
9094	G	A	2992	NS5	486	E	-	2963	50	1.69
9541	U	C	3141	NS5	635	H	-	7400	98	1.32
10367	C	U <sup>ab</sup>	-	3'UTR	-	-	-	2735	43	1.57
10638	C	U	-	3'UTR	-	-	-	1591	19	1.19
10662	C	U	-	3'UTR	-	-	-	1539	101	6.56

10838	U	C	-	-	-	-	688	7	1.02
-------	---	---	---	---	---	---	-----	---	------

<sup>a</sup> Intersection of sequence with the Asibi strain.  
<sup>b</sup> Intersection of sequence with the HeLa/P6 strain.

## Bibliography

- Afreen, N., I. H. Naqvi, S. Broor, A. Ahmed, and S. Parveen (2015). “Phylogenetic and Molecular Clock Analysis of Dengue Serotype 1 and 3 from New Delhi, India.” English. In: *PLoS ONE* 10.11, e0141628. DOI: 10.1371/journal.pone.0141628.
- Aihara, S., C. M. Rao, Y. X. Yu, T. Lee, K. Watanabe, T. Komiya, H. Sumiyoshi, H. Hashimoto, and A. Nomoto (1991). “Identification of mutations that occurred on the genome of Japanese encephalitis virus during the attenuation process.” English. In: *Virus genes* 5.2, pp. 95–109.
- Akondy, R. S., P. L. F. Johnson, H. I. Nakaya, S. Edupuganti, M. J. Mulligan, B. Lawson, J. D. Miller, B. Pulendran, R. Antia, and R. Ahmed (2015). “Initial viral load determines the magnitude of the human CD8 T cell response to yellow fever vaccination.” English. In: *Proceedings of the National Academy of Sciences* 112.10, pp. 3050–3055. DOI: 10.1073/pnas.1500475112.
- Andino, R. and E. Domingo (2015). “Viral quasispecies.” English. In: *Virology* 479-480C, pp. 46–51. DOI: 10.1016/j.virol.2015.03.022.
- Appaiahgari, M. B. and S. Vрати (2010). “IMOJEV(®): a Yellow fever virus-based novel Japanese encephalitis vaccine.” English. In: *Expert Review of Vaccines* 9.12, pp. 1371–1384. DOI: 10.1586/erv.10.139.
- Barban, V., Y. Girerd, M. Aguirre, S. Gulia, F. Pétiard, P. Riou, B. Barrere, and J. Lang (2007). “High stability of yellow fever 17D-204 vaccine: a 12-year retrospective analysis of large-scale production.” English. In: *Vaccine* 25.15, pp. 2941–2950. DOI: 10.1016/j.vaccine.2006.06.082.
- Barrett, A. D. T. and E. A. Gould (1986). “Comparison of Neurovirulence of Different Strains of Yellow Fever Virus in Mice”. English. In: *The Journal of General Virology* 67.4, pp. 631–637. DOI: 10.1099/0022-1317-67-4-631.
- Barrett, A. D., T. P. Monath, C. B. Cropp, J. A. Adkins, T. N. Ledger, E. A. Gould, J. J. Schlesinger, R. M. Kinney, and D. W. Trent (1990). “Attenuation of wild-type yellow fever virus by passage in HeLa cells.” English. In: *Journal of General Virology* 71 ( Pt 10), pp. 2301–2306.
- Barrett, A. D. T. and D. E. Teuwen (2009). “Yellow fever vaccine - how does it work and why do rare cases of serious adverse events take place?” English.

- In: *Current opinion in immunology* 21.3, pp. 308–313. DOI: 10.1016/j.coi.2009.05.018.
- Bazin, H. (2011). “Yellow Fever Vaccine”. English. In: *Vaccinations: A History*. John Libbey Eurotext, pp. 407–454. ISBN: 9782742007752.
- Beasley, D. W. C., M. Morin, A. R. Lamb, E. Hayman, D. M. Watts, C. K. Lee, D. W. Trent, and T. P. Monath (2013). “Adaptation of yellow fever virus 17D to Vero cells is associated with mutations in structural and non-structural protein genes”. In: *Virus Research*, pp. 1–5. DOI: 10.1016/j.virusres.2013.04.003.
- Beatty, P. R., H. Puerta-Guardo, S. S. Killingbeck, D. R. Glasner, K. Hopkins, and E. Harris (2015). “Dengue virus NS1 triggers endothelial permeability and vascular leak that is prevented by NS1 vaccination”. English. In: *Science Translational Medicine* 7.304, 304ra141–304ra141. DOI: 10.1126/scitranslmed.aaa3787.
- Beck, A. S. and A. D. Barrett (2015). “Current Status and Future Prospects of Yellow Fever Vaccines.” English. In: *Expert Review of Vaccines*, pp. 1–14. DOI: 10.1586/14760584.2015.1083430.
- Beck, A., R. B. Tesh, T. G. Wood, S. G. Widen, K. D. Ryman, and A. D. T. Barrett (2014). “Comparison of the live attenuated yellow fever vaccine 17D-204 strain to its virulent parental strain Asibi by deep sequencing.” English. In: *Journal of Infectious Diseases* 209.3, pp. 334–344. DOI: 10.1093/infdis/jit546.
- Belmusto-Worn, V. E., J. L. Sanchez, K. McCarthy, R. Nichols, C. T. Bautista, A. J. Magill, G. Pastor-Cauna, C. Echevarria, V. A. Laguna-Torres, B. K. Samame, M. E. Baldeon, J. P. Burans, J. G. Olson, P. Bedford, S. Kitchener, and T. P. Monath (2005). “Randomized, double-blind, phase III, pivotal field trial of the comparative immunogenicity, safety, and tolerability of two yellow fever 17D vaccines (Arilvax and YF-VAX) in healthy infants and children in Peru.” English. In: *American Journal of Tropical Medicine and Hygiene* 72.2, pp. 189–197.
- Bera, A. K., R. J. Kuhn, and J. L. Smith (2007). “Functional characterization of cis and trans activity of the Flavivirus NS2B-NS3 protease.” English. In: *The Journal of Biological Chemistry* 282.17, pp. 12883–12892. DOI: 10.1074/jbc.M611318200.
- Best, S. M., K. L. Morris, J. G. Shannon, S. J. Robertson, D. N. Mitzel, G. S. Park, E. Boer, J. B. Wolfenbarger, and M. E. Bloom (2005). “Inhibition

- of Interferon-Stimulated JAK-STAT Signaling by a Tick-borne Flavivirus and Identification of NS5 as an Interferon Antagonist.” English. In: *Journal of Virology* 79.20, pp. 12828–12839. DOI: 10.1128/JVI.79.20.12828-12839.2005.
- Biedenbender, R., J. Bevilacqua, A. M. Gregg, M. Watson, and G. Dayan (2011). “Phase II, randomized, double-blind, placebo-controlled, multicenter study to investigate the immunogenicity and safety of a West Nile virus vaccine in healthy adults.” English. In: *Journal of Infectious Diseases* 203.1, pp. 75–84. DOI: 10.1093/infdis/jiq003.
- Bonaldo, M. C., R. C. Garratt, R. S. Marchevsky, E. S. F. Coutinho, A. V. Jabor, L. F. C. Almeida, A. M. Y. Yamamura, A. S. Duarte, P. J. Oliveira, J. O. P. Lizeu, L. A. B. Camacho, M. S. Freire, and R. Galler (2005). “Attenuation of Recombinant Yellow Fever 17D Viruses Expressing Foreign Protein Epitopes at the Surface”. English. In: *Journal of Virology* 79.13, pp. 8602–8613. DOI: 10.1128/JVI.79.13.8602-8613.2005.
- Bonaldo, M. C., M. Â. Martins, R. Rudersdorf, P. A. Mudd, J. B. Sacha, S. M. Piaskowski, P. C. Costa Neves, M. G. Veloso de Santana, L. Vojnov, S. Capuano, E. G. Rakasz, N. A. Wilson, J. Fulkerson, J. C. Sadoff, D. I. Watkins, and R. Galler (2010). “Recombinant Yellow Fever Vaccine Virus 17D Expressing Simian Immunodeficiency Virus SIVmac239 Gag Induces SIV-Specific CD8+ T-Cell Responses in Rhesus Macaques”. English. In: *Journal of Virology* 84.7, pp. 3699–3706. DOI: 10.1128/JVI.02255-09.
- Bray, M. and C. J. Lai (1991). “Construction of intertypic chimeric dengue viruses by substitution of structural protein genes.” English. In: *Proceedings of the National Academy of Sciences of the United States of America* 88.22, pp. 10342–10346.
- Bredenbeek, P. J., R. Molenkamp, W. J. M. Spaan, V. Deubel, P. Marianneau, M. S. Salvato, D. Moshkoff, J. Zapata, I. Tikhonov, J. Patterson, R. Carrion, A. Ticer, K. Brasky, and I. S. Lukashevich (2006). “A recombinant Yellow Fever 17D vaccine expressing Lassa virus glycoproteins.” English. In: *Virology* 345.2, pp. 299–304. DOI: 10.1016/j.virol.2005.12.001.
- Bres, P. and M. Koch (1987). *Production and Testing of the WHO Yellow Fever Virus Primary Seed Lot 213-77 and Reference Batch 168-73*. Tech. rep. 745. Geneva: World Health Organization.
- Bryant, J. E., E. C. Holmes, and A. D. T. Barrett (2007). “Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas.” English. In: *PLoS Pathogens* 3.5, e75. DOI: 10.1371/journal.ppat.0030075.

- Bryant, J. E., P. F. C. Vasconcelos, R. C. A. Rijnbrand, J. P. Mutebi, S. Higgs, and A. D. T. Barrett (2005). "Size heterogeneity in the 3' noncoding region of South American isolates of yellow fever virus." English. In: *Journal of Virology* 79.6, pp. 3807–3821. DOI: 10.1128/JVI.79.6.3807-3821.2005.
- Burlaud-Gaillard, J., C. Sellin, S. Georgeault, R. Uzbekov, C. Lebos, J.-M. Guillaume, and P. Roingeard (2014). "Correlative scanning-transmission electron microscopy reveals that a chimeric flavivirus is released as individual particles in secretory vesicles." English. In: *PLoS ONE* 9.3, e93573. DOI: 10.1371/journal.pone.0093573.
- Camacho, L. A. B., S. G. d. Aguiar, M. d. S. Freire, M. d. L. F. Leal, J. P. d. Nascimento, T. Iguchi, J. A. Lozana, and R. H. G. Farias (2005). "Reactogenicity of yellow fever vaccines in a randomized, placebo-controlled trial". In: *Revista de saúde pública* 39.3, pp. 413–420.
- Camacho, L. A. B., M. d. S. Freire, M. d. L. F. Leal, S. G. d. Aguiar, J. P. d. Nascimento, T. Iguchi, J. d. A. Lozana, R. H. G. Farias, and Collaborative Group for the Study of Yellow Fever Vaccines (2004). "Immunogenicity of WHO-17D and Brazilian 17DD yellow fever vaccines: a randomized trial." English. In: *Revista de saúde pública* 38.5, pp. 671–678.
- Campbell, G. L., S. L. Hills, M. Fischer, J. A. Jacobson, C. H. Hoke, J. M. Hombach, A. A. Marfin, T. Solomon, T. F. Tsai, V. D. Tsu, and A. S. Ginsburg (2011). "Estimated global incidence of Japanese encephalitis: a systematic review." English. In: *Bulletin of the World Health Organization* 89.10, 766–74– 774A–774E. DOI: 10.2471/BLT.10.085233.
- Campi-Azevedo, A. C., L. P. de Araújo-Porto, M. Luiza-Silva, M. A. Batista, M. A. Martins, R. Sathler-Avelar, D. da Silveira-Lemos, L. A. B. Camacho, R. de Menezes Martins, M. de Lourdes de Sousa Maia, R. H. G. Farias, M. da Silva Freire, R. Galler, A. Homma, J. G. L. Ribeiro, J. A. C. Lemos, M. Auxiliadora-Martins, I. R. Caldas, S. M. Elói-Santos, A. Teixeira-Carvalho, and O. A. Martins-Filho (2012). "17DD and 17D-213/77 yellow fever sub-strains trigger a balanced cytokine profile in primary vaccinated children." English. In: *PLoS ONE* 7.12, e49828. DOI: 10.1371/journal.pone.0049828.
- Capeding, M. R., N. H. Tran, S. R. S. Hadinegoro, H. I. H. J. M. Ismail, T. Chotpitayasunondh, M. N. Chua, C. Q. Luong, K. Rusmil, D. N. Wirawan, R. Nallusamy, P. Pitisuttithum, U. Thisyakorn, I.-K. Yoon, D. van der Vliet, E. Langevin, T. Laot, Y. Hutagalung, C. Frago, M. Boaz, T. A. Wartel, N. G. Tornieporth, M. Saville, A. Bouckennooghe, and CYD14 Study Group (2014). "Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy



- children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial.” English. In: *Lancet* 384.9951, pp. 1358–1365. DOI: 10.1016/S0140-6736(14)61060-6.
- Catteau, A. (2003). “Dengue virus M protein contains a proapoptotic sequence referred to as ApoptoM”. English. In: *Journal of General Virology* 84.10, pp. 2781–2793. DOI: 10.1099/vir.0.19163-0.
- Centers for Disease Control and Prevention/Advisory Committee on Immunization Practices (2015). *Advisory Committee on Immunization Practices (ACIP) -Summary Report*. Tech. rep. Atlanta.
- Coffey, L. L., N. Vasilakis, A. C. Brault, A. M. Powers, F. Tripet, and S. C. Weaver (2008a). “Arbovirus evolution in vivo is constrained by host alternation.” English. In: *Proceedings of the National Academy of Sciences* 105.19, pp. 6970–6975. DOI: 10.1073/pnas.0712130105.
- Coffey, L. L., N. Vasilakis, A. C. Brault, A. M. Powers, F. Tripet, and S. C. Weaver (2008b). “Arbovirus evolution in vivo is constrained by host alternation.” English. In: *Proceedings of the National Academy of Sciences* 105.19, pp. 6970–6975. DOI: 10.1073/pnas.0712130105.
- Coffey, L. L. and M. Vignuzzi (2011). “Host alternation of chikungunya virus increases fitness while restricting population diversity and adaptability to novel selective pressures.” English. In: *Journal of Virology* 85.2, pp. 1025–1035. DOI: 10.1128/JVI.01918-10.
- Davis, N. C., W. Lloyd, and M. Frobisher (1932). “The Transmission of Neurotropic Yellow Fever Virus by Stegomyia Mosquitoes.” English. In: *The Journal of Experimental Medicine* 56.6, pp. 853–865.
- Deardorff, E. R., K. A. Fitzpatrick, G. V. S. Jerzak, P.-Y. Shi, L. D. Kramer, and G. D. Ebel (2011). “West Nile virus experimental evolution in vivo and the trade-off hypothesis.” English. In: *PLoS Pathogens* 7.11, e1002335. DOI: 10.1371/journal.ppat.1002335.
- DeSilva, M., A. Sharma, E. Staples, B. Arndt, W.-J. Shieh, J. Shames, and P. Cieslak (2015). “Notes from the field: fatal yellow fever vaccine-associated viscerotropic disease—Oregon, September 2014.” English. In: *Morbidity and Mortality Weekly Report* 64.10, pp. 279–281.
- Domingo, E., D. Sabo, T. Taniguchi, and C. Weissmann (1978). “Nucleotide sequence heterogeneity of an RNA phage population.” English. In: *Cell* 13.4, pp. 735–744.

- Domingo, E., J. Sheldon, and C. Perales (2012). “Viral Quasispecies Evolution”. English. In: *Microbiology and Molecular Biology Reviews* 76.2, pp. 159–216. DOI: 10.1128/MMBR.05023-11.
- Doran, A. G. and C. J. Creevey (2013). “Snpsnp: easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms.” English. In: *BMC Bioinformatics* 14.1, p. 45. DOI: 10.1186/1471-2105-14-45.
- Dunster, L. M., C. A. Gibson, J. R. Stephenson, P. D. Minor, and A. D. Barrett (1990). “Attenuation of virulence of flaviviruses following passage in HeLa cells.” English. In: *The Journal of General Virology* 71 ( Pt 3).3, pp. 601–607. DOI: 10.1099/0022-1317-71-3-601.
- Dunster, L. M., H. Wang, K. D. Ryman, B. R. Miller, S. J. Watowich, P. D. Minor, and A. D. Barrett (1999). “Molecular and biological changes associated with HeLa cell attenuation of wild-type yellow fever virus.” English. In: *Virology* 261.2, pp. 309–318. DOI: 10.1006/viro.1999.9873.
- Dupuy, A., P. Desprès, A. Cahour, M. Girard, and M. Bouloy (1989). “Nucleotide sequence comparison of the genome of two 17D-204 yellow fever vaccines.” English. In: *Nucleic acids research* 17.10, p. 3989.
- Durieux, C. (1956a). “Mass yellow fever vaccination in French Africa South of the Sahara”. In: *Yellow Fever Vaccination Monograph Series* 30, pp. 115–121.
- Durieux, C. (1956b). “Preparation of yellow fever vaccine at the Institut Pasteur, Dakar”. In: *Yellow Fever Vaccination Monograph Series* 30, pp. 31–32.
- Ebel, G. D., K. A. Fitzpatrick, P.-Y. Lim, C. J. Bennett, E. R. Deardorff, G. V. S. Jerzak, L. D. Kramer, Y. Zhou, P.-Y. Shi, and K. A. Bernard (2011). “Non-consensus West Nile virus genomes arising during mosquito infection suppress pathogenesis and modulate virus fitness in vivo.” English. In: *Journal of Virology* 85.23, pp. 12605–12613. DOI: 10.1128/JVI.05637-11.
- Ellis, B. R. and A. D. Barrett (2008). “The enigma of yellow fever in East Africa.” English. In: *Reviews in Medical Virology* 18.5, pp. 331–346. DOI: 10.1002/rmv.584.
- Erickson, A. K. and J. K. Pfeiffer (2013). “Dynamic Viral Dissemination in Mice Infected with Yellow Fever Virus Strain 17D.” English. In: *Journal of Virology* 87.22, pp. 12392–12397. DOI: 10.1128/JVI.02149-13.

- Findlay, G. M. and L. P. Clarke (1935). "Reconversion of the neurotropic into the viscerotropic strain of yellow fever virus in rhesus monkeys". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 28.6, pp. 579–600.
- Fitzgeorge, R. and C. J. Bradish (1980). "The in vivo differentiation of strains of yellow fever virus in mice." English. In: *The Journal of General Virology* 46.1, pp. 1–13.
- Forrester, N. L., M. Guerbois, R. L. Seymour, H. Spratt, and S. C. Weaver (2012). "Vector-Borne Transmission Imposes a Severe Bottleneck on an RNA Virus Population". English. In: *PLoS Pathogens* 8.9, e1002897–11. DOI: 10.1371/journal.ppat.1002897.
- Freire, M. S., G. F. Mann, R. S. Marchevsky, A. M. Y. Yamamura, L. F. C. Almeida, A. V. Jabor, J. M. N. Malachias, E. S. F. Coutinho, and R. Galler (2005). "Production of yellow fever 17DD vaccine virus in primary culture of chicken embryo fibroblasts: yields, thermo and genetic stability, attenuation and immunogenicity". English. In: *Vaccine* 23.19, pp. 2501–2512. DOI: 10.1016/j.vaccine.2004.10.035.
- Garske, T., M. D. Van Kerkhove, S. Yactayo, O. Ronveaux, R. F. Lewis, J. E. Staples, W. Perea, N. M. Ferguson, and for the Yellow Fever Expert Committee (2014). "Yellow Fever in Africa: Estimating the Burden of Disease and Impact of Mass Vaccination from Outbreak and Serological Data". English. In: *PLoS Medicine* 11.5, e1001638. DOI: 10.1371/journal.pmed.1001638.s015.
- Gaspar, L. P., Y. S. Mendes, A. M. Y. Yamamura, L. F. C. Almeida, E. Caride, R. B. Gonçalves, J. L. Silva, A. C. Oliveira, R. Galler, and M. S. Freire (2008). "Pressure-inactivated yellow fever 17DD virus: implications for vaccine development." English. In: *Journal of Virological Methods* 150.1-2, pp. 57–62. DOI: 10.1016/j.jviromet.2008.03.002.
- Gaucher, D., R. Therrien, N. Kettaf, B. R. Angermann, G. Boucher, A. Filali-Mouhim, J. M. Moser, R. S. Mehta, D. R. Drake, E. Castro, R. Akondy, A. Rinfret, B. Yassine-Diab, E. A. Said, Y. Chouikh, M. J. Cameron, R. Clum, D. Kelvin, R. Somogyi, L. D. Greller, R. S. Balderas, P. Wilkinson, G. Pantaleo, J. Tartaglia, E. K. Haddad, and R. P. Sekaly (2008). "Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses". English. In: *Journal of Experimental Medicine* 205.13, pp. 3119–3131. DOI: 10.1084/jem.20031598.
- Gershman, M. D., J. E. Staples, A. D. Bentsi-Enchill, J. G. Breugelmans, G. S. Brito, L. A. B. Camacho, P. Cottin, C. Domingo, A. Durbin, J. Gascon, F. Guenaneche, E. B. Hayes, Z. Jelenik, A. Khromava, R. d. M. Martins,

- M. M. Wilson, N. Massy, A. Nasidi, M. Niedrig, A. Sherwat, T. Tsai, A. Vilella, M. E. Wilson, K. S. Kohl, and Brighton Collaboration Viscerotropic Disease Working Group (2012). “Viscerotropic disease: case definition and guidelines for collection, analysis, and presentation of immunization safety data.” English. In: *Vaccine* 30.33, pp. 5038–5058. DOI: 10.1016/j.vaccine.2012.04.067.
- Gerstung, M., C. Beisel, M. Rechsteiner, P. Wild, P. Schraml, H. Moch, and N. Beerenwinkel (2012). “Reliable detection of subclonal single-nucleotide variants in tumour cell populations.” English. In: *Nature communications* 3, p. 811. DOI: 10.1038/ncomms1814.
- Global Alliance for Vaccines and Immunization (GAVI) (2015). *Yellow Fever Supply and Procurement Roadmap*. URL: <http://www.gavi.org/library/gavi-documents/supply-procurement/yellow-fever-roadmap-public-summary/> (visited on 06/18/2015).
- Gotuzzo, E., S. Yactayo, and E. Córdova (2013). “Efficacy and duration of immunity after yellow fever vaccination: systematic review on the need for a booster every 10 years.” English. In: *The American Journal of Tropical Medicine and Hygiene* 89.3, pp. 434–444. DOI: 10.4269/ajtmh.13-0264.
- Gould, E. A., A. Buckley, P. A. Cane, S. Higgs, and N. Cammack (1989). “Use of a monoclonal antibody specific for wild-type yellow fever virus to identify a wild-type antigenic variant in 17D vaccine pools.” English. In: *Journal of General Virology* 70 ( Pt 7), pp. 1889–1894.
- Gromowski, G. D., C.-Y. Firestone, J. Bustos-Arriaga, and S. S. Whitehead (2015). “Genetic and Phenotypic Properties of Vero Cell-Adapted Japanese Encephalitis Virus SA14-14-2 Vaccine Strain Variants and a Recombinant Clone, Which Demonstrates Attenuation and Immunogenicity in Mice”. English. In: *American Journal of Tropical Medicine and Hygiene* 92.1, pp. 98–107. DOI: 10.4269/ajtmh.14-0427.
- Gromowski, G. D., C.-Y. Firestone, and S. S. Whitehead (2015). “Genetic Determinants of Japanese Encephalitis Virus Vaccine Strain SA14-14-2 That Govern Attenuation of Virulence in Mice.” English. In: *Journal of Virology* 89.12, pp. 6328–6337. DOI: 10.1128/JVI.00219-15.
- Guy, B., B. Barrere, C. Malinowski, M. Saville, R. Teyssou, and J. Lang (2011). “From research to phase III: preclinical, industrial and clinical development of the Sanofi Pasteur tetravalent dengue vaccine.” English. In: *Vaccine* 29.42, pp. 7229–7241. DOI: 10.1016/j.vaccine.2011.06.094.

- Hahn, C. S., J. M. Dalrymple, J. H. Strauss, and C. M. Rice (1987). "Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it." English. In: *Proceedings of the National Academy of Sciences of the United States of America* 84.7, pp. 2019–2023.
- Holbrook, M. R., L. Li, M. T. Suderman, H. Wang, and A. D. Barrett (2000). "The French neurotropic vaccine strain of yellow fever virus accumulates mutations slowly during passage in cell culture." English. In: *Virus Research* 69.1, pp. 31–39.
- Holland, J. J., J. C. De La Torre, and D. A. Steinhauer (1992). "RNA Virus Populations as Quasispecies". English. In: *Genetic Diversity of RNA Viruses*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–20. ISBN: 978-3-642-77013-5. DOI: 10.1007/978-3-642-77011-1\_1.
- Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol (1982). "Rapid Evolution of RNA Genomes". English. In: *Science (New York, N.Y.)* 215.4540, pp. 1577–1585.
- Holmes, E. C. (2003). "Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus." English. In: *Journal of Virology* 77.20, pp. 11296–11298. DOI: 10.1128/JVI.77.20.11296-11298.2003.
- Höper, D., C. M. Freuling, T. Müller, D. Hanke, V. von Messling, K. Duchow, M. Beer, and T. C. Mettenleiter (2015). "High definition viral vaccine strain identity and stability testing using full-genome population data - The next generation of vaccine quality control." English. In: *Vaccine* 33.43, pp. 5829–5837. DOI: 10.1016/j.vaccine.2015.08.091.
- Huang, C. H. (1982). *Studies of Japanese Encephalitis in China*. Vol. 27. Elsevier, pp. 71–101. ISBN: 9780120398270. DOI: 10.1016/S0065-3527(08)60433-9.
- Huang, Y.-J. S., J. T. Nuckols, K. M. Horne, D. Vanlandingham, M. Lobigs, and S. Higgs (2014). "Mutagenesis analysis of T380R mutation in the envelope protein of yellow fever virus". In: *Virology Journal* 11.1, pp. 1–4. DOI: 10.1186/1743-422X-11-60.
- Jennings, A. D., C. A. Gibson, B. R. Miller, J. H. Mathews, C. J. Mitchell, J. T. Roehrig, D. J. Wood, F. Taffs, B. K. Sil, and S. N. Whitby (1994). "Analysis of a yellow fever virus isolated from a fatal case of vaccine-associated human encephalitis." English. In: *The Journal of Infectious Diseases* 169.3, pp. 512–518.

- Jentes, E. S., G. Pomeroy, M. D. Gershman, D. R. Hill, J. Lemarchand, R. F. Lewis, J. E. Staples, O. Tomori, A. Wilder-Smith, T. P. Monath, and Informal WHO Working Group on Geographic Risk for Yellow Fever (2011). "The revised global yellow fever risk map and recommendations for vaccination, 2010: consensus of the Informal WHO Working Group on Geographic Risk for Yellow Fever." English. In: *The Lancet. Infectious diseases* 11.8, pp. 622–632. DOI: 10.1016/S1473-3099(11)70147-5.
- Johansson, M. A., N. Arana-Vizcarrondo, B. J. Biggerstaff, and J. E. Staples (2010). "Incubation Periods of Yellow Fever Virus". English. In: *The American Journal of Tropical Medicine and Hygiene* 83.1, pp. 183–188. DOI: 10.4269/ajtmh.2010.09-0782.
- Johansson, M. A., P. F. C. Vasconcelos, and J. E. Staples (2014). "The whole iceberg: estimating the incidence of yellow fever virus infection from the number of severe cases." English. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 108.8, pp. 482–487. DOI: 10.1093/trstmh/tru092.
- Johnston, L. J., G. M. Halliday, and N. J. King (2000). "Langerhans cells migrate to local lymph nodes following cutaneous infection with an arbovirus." English. In: *The Journal of Investigative Dermatology* 114.3, pp. 560–568. DOI: 10.1046/j.1523-1747.2000.00904.x.
- Julander, J. G., D. W. Trent, and T. P. Monath (2011). "Immune correlates of protection against yellow fever determined by passive immunization and challenge in the hamster model." English. In: *Vaccine* 29.35, pp. 6008–6016. DOI: 10.1016/j.vaccine.2011.06.034.
- Kakumani, P. K., S. S. Ponia, R. K. S. V. Sood, M. Chinnappan, A. C. Banerjee, G. R. Medigeshi, P. Malhotra, S. K. Mukherjee, and R. K. Bhatnagar (2013). "Role of RNA interference (RNAi) in dengue virus replication and identification of NS4B as an RNAi suppressor." English. In: *Journal of Virology* 87.16, pp. 8870–8883. DOI: 10.1128/JVI.02774-12.
- Kennedy, R. B. and G. A. Poland (2011). "The top five "game changers" in vaccinology: toward rational and directed vaccine development." English. In: *Omics : A Journal of Integrative Biology* 15.9, pp. 533–537. DOI: 10.1089/omi.2011.0012.
- Lancaster, K. Z. and J. K. Pfeiffer (2011). "Mechanisms Controlling Virulence Thresholds of Mixed Viral Populations". English. In: *Journal of Virology* 85.19, pp. 9778–9788. DOI: 10.1128/JVI.00355-11.

- Lang, J., J. Zuckerman, P. Clarke, P. Barrett, C. Kirkpatrick, and C. Blondeau (1999). "Comparison of the immunogenicity and safety of two 17D yellow fever vaccines." English. In: *American Journal of Tropical Medicine and Hygiene* 60.6, pp. 1045–1050.
- Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." English. In: *Nature Methods* 9.4, pp. 357–359. DOI: 10.1038/nmeth.1923.
- Laurent-Rolle, M., J. Morrison, R. Rajsbaum, J. M. L. Macleod, G. Pisanelli, A. Pham, J. Ayllon, L. Miorin, C. Martínez-Romero, B. R. tenOever, and A. García-Sastre (2014). "The Interferon Signaling Antagonist Function of Yellow Fever Virus NS5 Protein Is Activated by Type I Interferon". In: *Cell host & microbe* 16.3, pp. 314–327. DOI: 10.1016/j.chom.2014.07.015.
- Lee, E. and M. Lobigs (2008). "E protein domain III determinants of yellow fever virus 17D vaccine strain enhance binding to glycosaminoglycans, impede virus spread, and attenuate virulence." English. In: *Journal of Virology* 82.12, pp. 6024–6033. DOI: 10.1128/JVI.02509-07.
- Leung, J. Y., G. P. Pijlman, N. Kondratieva, J. Hyde, J. M. Mackenzie, and A. A. Khromykh (2008). "Role of nonstructural protein NS2A in flavivirus assembly." English. In: *Journal of Virology* 82.10, pp. 4731–4741. DOI: 10.1128/JVI.00002-08.
- Li, F., Y. Wang, L. Yu, S. Cao, K. Wang, J. Yuan, C. Wang, K. Wang, M. Cui, and Z. F. Fu (2015). "Viral Infection of the Central Nervous System and Neuroinflammation Precede Blood-Brain Barrier Disruption during Japanese Encephalitis Virus Infection." English. In: *Journal of Virology* 89.10, pp. 5602–5614. DOI: 10.1128/JVI.00143-15.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup (2009). "The Sequence Alignment/Map format and SAMtools." English. In: *Bioinformatics (Oxford, England)* 25.16, pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352.
- Li, K., E. Venter, S. Yooseph, T. B. Stockwell, L. D. Eckerle, M. R. Denison, D. J. Spiro, and B. A. Methé (2010). "ANDES: Statistical tools for the ANALyses of DEep Sequencing." English. In: *BMC Research Notes* 3, p. 199. DOI: 10.1186/1756-0500-3-199.
- Li, S., N. Rouphael, S. Duraisingham, S. Romero-Steiner, S. Presnell, C. Davis, D. S. Schmidt, S. E. Johnson, A. Milton, G. Rajam, S. Kasturi, G. M. Carlone, C. Quinn, D. Chaussabel, A. K. Palucka, M. J. Mulligan, R. Ahmed,

- D. S. Stephens, H. I. Nakaya, and B. Pulendran (2014). “Molecular signatures of antibody responses derived from a systems biology study of five human vaccines.” English. In: *Nature Immunology* 15.2, pp. 195–204. DOI: 10.1038/ni.2789.
- Liprandi, F. (1981). “Isolation of plaque variants differing in virulence from the 17D strain of yellow fever virus.” English. In: *Journal of General Virology* 56.Pt 2, pp. 363–370.
- Lloyd, W., M. Theiler, and N. I. Ricci (1936). “Modification of the virulence of yellow fever virus by cultivation in tissues in vitro”. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 29.5, pp. 481–529.
- Lloyd, W. and H. A. Penna (1933). “Studies on the pathogenesis of neurotropic yellow fever virus in *Macacus rhesus*”. English. In: *The American Journal of Tropical Medicine and Hygiene*.
- Lobigs, M., E. Lee, M. L. Ng, M. Pavy, and P. Lobigs (2010). “A flavivirus signal peptide balances the catalytic activity of two proteases and thereby facilitates virus morphogenesis.” English. In: *Virology* 401.1, pp. 80–89. DOI: 10.1016/j.virol.2010.02.008.
- Lubick, K. J., S. J. Robertson, K. L. McNally, B. A. Freedman, A. L. Rasmussen, R. T. Taylor, A. D. Walts, S. Tsuruda, M. Sakai, M. Ishizuka, E. F. Boer, E. C. Foster, A. I. Chiramel, C. B. Addison, R. Green, D. L. Kastner, M. G. Katze, S. M. Holland, A. Forlino, A. F. Freeman, M. Boehm, K. Yoshii, and S. M. Best (2015). “Flavivirus Antagonism of Type I Interferon Signaling Reveals Prolidase as a Regulator of IFNAR1 Surface Expression.” English. In: *Cell host & microbe* 18.1, pp. 61–74. DOI: 10.1016/j.chom.2015.06.007.
- Luciani, F., R. A. Bull, and A. R. Lloyd (2012). “Next generation deep sequencing and vaccine design: today and tomorrow.” English. In: *Trends in Biotechnology* 30.9, pp. 443–452. DOI: 10.1016/j.tibtech.2012.05.005.
- Luo, D., T. Xu, R. P. Watson, D. Scherer-Becker, A. Sampath, W. Jahnke, S. S. Yeong, C. H. Wang, S. P. Lim, A. Strongin, S. G. Vasudevan, and J. Lescar (2008). “Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein.” English. In: *The EMBO Journal* 27.23, pp. 3209–3219. DOI: 10.1038/emboj.2008.232.
- Ma, L., C. T. Jones, T. D. Groesch, R. J. Kuhn, and C. B. Post (2004). “Solution structure of dengue virus capsid protein reveals another fold.” English. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.10, pp. 3414–3419. DOI: 10.1073/pnas.0305892101.



- Maciel, M., F. d. S. P. Cruz, M. T. Cordeiro, M. A. da Motta, K. M. S. d. M. Cassemiro, R. d. C. C. Maia, R. C. B. Q. de Figueiredo, R. Galler, M. d. S. Freire, J. T. August, E. T. A. Marques, and R. Dhalia (2015). "A DNA Vaccine against Yellow Fever Virus: Development and Evaluation." English. In: *PLoS Neglected Tropical Diseases* 9.4, e0003693. DOI: 10.1371/journal.pntd.0003693.
- McArthur, M. A., M. T. Suderman, J.-P. Mutebi, S.-Y. Xiao, and A. D. T. Barrett (2003). "Molecular characterization of a hamster viscerotropic strain of yellow fever virus." English. In: *Journal of Virology* 77.2, pp. 1462–1468. DOI: 10.1128/JVI.77.2.1462-1468.2003.
- McElroy, K. L., K. A. Tsetsarkin, D. L. Vanlandingham, and S. Higgs (2005). "Characterization of an infectious clone of the wild-type yellow fever virus Asibi strain that is able to infect and disseminate in mosquitoes." English. In: *Journal of General Virology* 86.Pt 6, pp. 1747–1751. DOI: 10.1099/vir.0.80746-0.
- McElroy, K. L., K. A. Tsetsarkin, D. L. Vanlandingham, and S. Higgs (2006). "Manipulation of the yellow fever virus non-structural genes 2A and 4B and the 3'non-coding region to evaluate genetic determinants of viral dissemination from the *Aedes aegypti* midgut." English. In: *American Journal of Tropical Medicine and Hygiene* 75.6, pp. 1158–1164.
- Meers, P. D. (1959). "Adaptation of the 17D yellow fever virus to mouse brain by serial passage". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 53.6, pp. 445–457.
- Meier, K. C., C. L. Gardner, M. V. Khoretonenko, W. B. Klimstra, and K. D. Ryman (2009). "A mouse model for studying viscerotropic disease caused by yellow fever virus infection." English. In: *PLoS Pathogens* 5.10, e1000614. DOI: 10.1371/journal.ppat.1000614.
- Messina, J. P., O. J. Brady, T. W. Scott, C. Zou, D. M. Pigott, K. A. Duda, S. Bhatt, L. Katzelnick, R. E. Howes, K. E. Battle, C. P. Simmons, and S. I. Hay (2014). "Global Spread of Dengue Virus Types: Mapping the 70 Year History." English. In: *Trends in Microbiology* 22.3, pp. 138–146. DOI: 10.1016/j.tim.2013.12.011.
- Miller, B. R. and D. Adkins (1988). "Biological characterization of plaque-size variants of yellow fever virus in mosquitoes and mice." English. In: *Acta Virologica* 32.3, pp. 227–234.

- Miller, R. H., P. Masuoka, T. A. Klein, H.-C. Kim, T. Somer, and J. Grieco (2012). "Ecological niche modeling to estimate the distribution of Japanese encephalitis virus in Asia." English. In: *PLoS Neglected Tropical Diseases* 6.6, e1678. DOI: 10.1371/journal.pntd.0001678.
- Minor, P. D. (2011). "Neurovirulence tests of three 17D yellow fever vaccine strains." English. In: *Biologicals : journal of the International Association of Biological Standardization* 39.3, pp. 167–170. DOI: 10.1016/j.biologicals.2011.04.001.
- Monath, T. P., K. R. Brinker, F. W. Chandler, G. E. Kemp, and C. B. Cropp (1981). "Pathophysiologic correlations in a rhesus monkey model of yellow fever with special observations on the acute necrosis of B cell areas of lymphoid tissues." English. In: *The American Journal of Tropical Medicine and Hygiene* 30.2, pp. 431–443.
- Monath, T. P., R. M. Kinney, J. J. Schlesinger, M. W. Brandriss, and P. Bres (1983). "Ontogeny of yellow fever 17D vaccine: RNA oligonucleotide fingerprint and monoclonal antibody analyses of vaccines produced world-wide." English. In: *The Journal of General Virology* 64 Pt 3, pp. 627–637.
- Monath, T. P., M. S. Cetron, K. McCarthy, R. Nichols, W. T. Archambault, L. Weld, and P. Bedford (2005). "Yellow fever 17D vaccine safety and immunogenicity in the elderly." English. In: *Human Vaccines* 1.5, pp. 207–214. DOI: 10.4161/hv.1.5.2221.
- Monath, T. P., E. Fowler, C. T. Johnson, J. Balser, M. J. Morin, M. Sisti, and D. W. Trent (2011). "An inactivated cell-culture vaccine against yellow fever." English. In: *The New England Journal of Medicine* 364.14, pp. 1326–1333. DOI: 10.1056/NEJMoa1009303.
- Monath, T. P., M. Gershman, J. E. Staples, and A. D. T. Barrett (2013). *Yellow fever vaccine*. Sixth Edition. Elsevier Inc. DOI: 10.1016/B978-1-4557-0090-5.00043-4.
- Monath, T. P., C. K. Lee, J. G. Julander, A. Brown, D. W. Beasley, D. M. Watts, E. Hayman, P. Guertin, J. Makowiecki, J. Crowell, P. Levesque, G. C. Bowick, M. Morin, E. Fowler, and D. W. Trent (2010). "Inactivated yellow fever 17D vaccine: development and nonclinical safety, immunogenicity and protective activity." English. In: *Vaccine* 28.22, pp. 3827–3840. DOI: 10.1016/j.vaccine.2010.03.023.
- Monath, T. P., J. Liu, N. Kanesa-Thasan, G. A. Myers, R. Nichols, A. Deary, K. McCarthy, C. Johnson, T. Ermak, S. Shin, J. Arroyo, F. Guirakhoo, J. S.

- Kennedy, F. A. Ennis, S. Green, and P. Bedford (2006). “A live, attenuated recombinant West Nile virus vaccine.” English. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.17, pp. 6694–6699. DOI: 10.1073/pnas.0601932103.
- Monath, T. P. and P. F. C. Vasconcelos (2015). “Yellow fever.” English. In: *Journal of Clinical Virology* 64, pp. 160–173. DOI: 10.1016/j.jcv.2014.08.030.
- Morelli, M. J., C. F. Wright, N. J. Knowles, N. Juleff, D. J. Paton, D. P. King, and D. T. Haydon (2013). “Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts”. English. In: *Veterinary Research* 44.1, pp. 1–1. DOI: 10.1186/1297-9716-44-12.
- Most, R. G. van der, L. E. Harrington, V. Giuggio, P. L. Mahar, and R. Ahmed (2002). “Yellow fever virus 17D envelope and NS3 proteins are major targets of the antiviral T cell response in mice.” English. In: *Virology* 296.1, pp. 117–124. DOI: 10.1006/viro.2002.1432.
- Moulin, J.-C., J. Silvano, V. Barban, P. Riou, and C. Allain (2013). “Yellow fever vaccine: comparison of the neurovirulence of new 17D-204 Stamaril<sup>TM</sup> seed lots and RK 168-73 strain.” English. In: *Biologicals : journal of the International Association of Biological Standardization* 41.4, pp. 238–246. DOI: 10.1016/j.biologicals.2013.04.005.
- Moutailler, S., B. Roche, J.-M. Thiberge, V. Caro, F. Rougeon, and A.-B. Failloux (2011). “Host alternation is necessary to maintain the genome stability of rift valley fever virus.” English. In: *PLoS Neglected Tropical Diseases* 5.5, e1156. DOI: 10.1371/journal.pntd.0001156.
- Mudd, P. A., S. M. Piaskowski, P. C. C. Neves, R. Rudersdorf, H. L. Kolar, C. M. Eernisse, K. L. Weisgrau, M. G. V. de Santana, N. A. Wilson, M. C. Bonaldo, R. Galler, E. G. Rakasz, and D. I. Watkins (2010). “The live-attenuated yellow fever vaccine 17D induces broad and potent T cell responses against several viral proteins in Indian rhesus macaques—implications for recombinant vaccine design.” English. In: *Immunogenetics* 62.9, pp. 593–600. DOI: 10.1007/s00251-010-0461-0.
- Muñoz-Jordán, J. L., M. Laurent-Rolle, J. Ashour, L. Martínez-Sobrido, M. Ashok, W. I. Lipkin, and A. García-Sastre (2005). “Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses.” English. In: *Journal of Virology* 79.13, pp. 8004–8013. DOI: 10.1128/JVI.79.13.8004-8013.2005.
- Myint, K. S. A., A. Kipar, R. G. Jarman, R. V. Gibbons, G. C. Perng, B. Flanagan, D. Mongkolsirichaikul, Y. Van Gessel, and T. Solomon (2014). “Neuropatho-

- genesis of Japanese encephalitis in a primate model.” English. In: *PLoS Neglected Tropical Diseases* 8.8, e2980. DOI: 10.1371/journal.pntd.0002980.
- Neverov, A. and K. Chumakov (2010). “Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines.” English. In: *Proceedings of the National Academy of Sciences* 107.46, pp. 20063–20068. DOI: 10.1073/pnas.1012537107.
- Neves, P. C. C., R. A. Rudersdorf, R. Galler, M. C. Bonaldo, M. G. V. de Santana, P. A. Mudd, M. A. Martins, E. G. Rakasz, N. A. Wilson, and D. I. Watkins (2010). “CD8<sup>+</sup> gamma-delta TCR<sup>+</sup> and CD4<sup>+</sup> T cells produce IFN- $\gamma$  at 5-7 days after yellow fever vaccination in Indian rhesus macaques, before the induction of classical antigen-specific T cell responses.” English. In: *Vaccine* 28.51, pp. 8183–8188. DOI: 10.1016/j.vaccine.2010.09.090.
- Neves, P. C. C., J. R. Santos, L. N. Tubarão, M. C. Bonaldo, and R. Galler (2013). “Early IFN-Gamma Production after YF 17D Vaccine Virus Immunization in Mice and Its Association with Adaptive Immune Responses”. English. In: *PLoS ONE* 8.12, e81953. DOI: 10.1371/journal.pone.0081953.s003.
- Ni, H. and A. D. Barrett (1995). “Nucleotide and deduced amino acid sequence of the structural protein genes of Japanese encephalitis viruses from different geographical locations.” English. In: *The Journal of General Virology* 76 ( Pt 2), pp. 401–407.
- Ni, H., N. J. Burns, G. J. Chang, M. J. Zhang, M. R. Wills, D. W. Trent, P. G. Sanders, and A. D. Barrett (1994). “Comparison of nucleotide and deduced amino acid sequence of the 5’ non-coding region and structural protein genes of the wild-type Japanese encephalitis virus strain SA14 and its attenuated vaccine derivatives.” English. In: *The Journal of General Virology* 75 ( Pt 6), pp. 1505–1510.
- Ni, H., K. D. Ryman, H. Wang, M. F. Saeed, R. Hull, D. Wood, P. D. Minor, S. J. Watowich, and A. D. Barrett (2000). “Interaction of yellow fever virus French neurotropic vaccine strain with monkey brain: characterization of monkey brain membrane receptor escape variants.” English. In: *Journal of Virology* 74.6, pp. 2903–2906.
- Niedrig, M., M. Lademann, P. Emmerich, and M. Lafrenz (1999). “Assessment of IgG antibodies against yellow fever virus after vaccination with 17D by different assays: neutralization test, haemagglutination inhibition test, immunofluorescence assay and ELISA.” English. In: *Tropical Medicine & International Health : TM & IH* 4.12, pp. 867–871.

- Nishijima, N., H. Marusawa, Y. Ueda, K. Takahashi, A. Nasu, Y. Osaki, T. Kou, S. Yazumi, T. Fujiwara, S. Tsuchiya, K. Shimizu, S. Uemoto, and T. Chiba (2012). “Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing.” English. In: *PLoS ONE* 7.4, e35052. DOI: 10.1371/journal.pone.0035052.
- Nitayaphan, S., J. A. Grant, G. J. Chang, and D. W. Trent (1990). “Nucleotide sequence of the virulent SA-14 strain of Japanese encephalitis virus and its attenuated vaccine derivative, SA-14-14-2.” English. In: *Virology* 177.2, pp. 541–552.
- Nogueira, R. T., A. R. Nogueira, M. C. S. Pereira, M. M. Rodrigues, P. C. d. C. Neves, R. Galler, and M. C. Bonaldo (2013). “Recombinant yellow fever viruses elicit CD8+ T cell responses and protective immunity against *Trypanosoma cruzi*.” English. In: *PLoS ONE* 8.3, e59347. DOI: 10.1371/journal.pone.0059347.
- Patkar, C. G., C. T. Jones, Y.-h. Chang, R. Warrier, and R. J. Kuhn (2007). “Functional requirements of the yellow fever virus capsid protein.” English. In: *Journal of Virology* 81.12, pp. 6471–6481. DOI: 10.1128/JVI.02120-6.
- Pereira, R. C., A. N. M. R. Silva, M. C. O. Souza, M. V. Silva, P. P. C. C. Neves, A. A. M. V. Silva, D. D. C. S. Matos, M. A. O. Herrera, A. M. Y. Yamamura, M. S. Freire, L. P. Gaspar, and E. Caride (2015). “An inactivated yellow fever 17DD vaccine cultivated in Vero cell cultures.” English. In: *Vaccine*. DOI: 10.1016/j.vaccine.2015.03.077.
- Pessoa, M. G., N. Bzowej, M. Berenguer, Y. Phung, M. Kim, L. Ferrell, H. Hassoba, and T. L. Wright (1999). “Evolution of hepatitis C virus quasispecies in patients with severe cholestatic hepatitis after liver transplantation.” English. In: *Hepatology* 30.6, pp. 1513–1520. DOI: 10.1002/hep.510300610.
- Pfister, M., O. Kürsteiner, H. Hilfiker, D. Favre, P. Durrer, A. Ennaji, J. L’age Stehr, A. Kaufhold, and C. Herzog (2005). “Immunogenicity and safety of BERNAYF compared with two other 17D yellow fever vaccines in a phase 3 clinical trial.” English. In: *American Journal of Tropical Medicine and Hygiene* 72.3, pp. 339–346.
- Pincus, S., P. W. Mason, E. Konishi, B. A. Fonseca, R. E. Shope, C. M. Rice, and E. Paoletti (1992). “Recombinant vaccinia virus producing the prM and E proteins of yellow fever virus protects mice from lethal yellow fever encephalitis.” English. In: *Virology* 187.1, pp. 290–297.

- Post, P. R., R. Carvalho, and R. Galler (1991). "Glycosylation and secretion of yellow fever virus nonstructural protein NS1." English. In: *Virus Research* 18.2-3, pp. 291–302.
- Prachi, P., C. Donati, F. Masciopinto, R. Rappuoli, and F. Bagnoli (2013). "Deep Sequencing in Pre- and Clinical Vaccine Research". English. In: *Public Health Genomics* 16.1-2, pp. 62–68. DOI: 10.1159/000345611.
- Pugachev, K. V., F. Guirakhoo, S. W. Ocran, F. Mitchell, M. Parsons, C. Penal, S. Girakhoo, S. O. Pougatcheva, J. Arroyo, D. W. Trent, and T. P. Monath (2003). "High fidelity of yellow fever virus RNA polymerase." English. In: *Journal of Virology* 78.2, pp. 1032–1038. DOI: 10.1128/JVI.78.2.1032-1038.2004.
- Pugachev, K. V., S. W. Ocran, F. Guirakhoo, D. Furby, and T. P. Monath (2002). "Heterogeneous nature of the genome of the ARILVAX yellow fever 17D vaccine revealed by consensus sequencing." English. In: *Vaccine* 20.7-8, pp. 996–999.
- Pulendran, B., J. Miller, T. D. Querec, R. Akondy, N. Moseley, O. Laur, J. Glidewell, N. Monson, T. Zhu, H. Zhu, S. Staprans, D. Lee, M. A. Brinton, A. A. Perelygin, C. Vellozzi, P. Brachman Jr, S. Lalor, D. Teuwen, R. B. Eidex, M. Cetron, F. Priddy, C. d. Rio, J. Altman, and R. Ahmed (2008). "Case of Yellow Fever Vaccine–Associated Viscerotropic Disease with Prolonged Viremia, Robust Adaptive Immune Responses, and Polymorphisms in CCR5 and RANTES Genes". English. In: *Journal of Infectious Diseases* 198.4, pp. 500–507. DOI: 10.1086/590187.
- Quaresma, J. A. S., V. L. R. S. Barros, C. Pagliari, E. R. Fernandes, F. Guedes, C. F. H. Takakura, H. F. Andrade, P. F. C. Vasconcelos, and M. I. S. Duarte (2006). "Revisiting the liver in human yellow fever: virus-induced apoptosis in hepatocytes associated with TGF-beta, TNF-alpha and NK cells activity." English. In: *Virology* 345.1, pp. 22–30. DOI: 10.1016/j.virol.2005.09.058.
- Quaresma, J. A. S., M. I. S. Duarte, and P. F. C. Vasconcelos (2006). "Midzonal lesions in yellow fever: A specific pattern of liver injury caused by direct virus action and in situ inflammatory response". English. In: *Medical Hypotheses* 67.3, pp. 618–621. DOI: 10.1016/j.mehy.2006.01.060.
- Querec, T. D., R. S. Akondy, E. K. Lee, W. Cao, H. I. Nakaya, D. Teuwen, A. Pirani, K. Gernert, J. Deng, B. Marzolf, K. Kennedy, H. Wu, S. Bennouna, H. Oluoch, J. Miller, R. Z. Vencio, M. Mulligan, A. Aderem, R. Ahmed, and B. Pulendran (2009). "Systems biology approach predicts immunogenicity of

- the yellow fever vaccine in humans.” English. In: *Nature Immunology* 10.1, pp. 116–125. DOI: 10.1038/ni.1688.
- Querec, T., S. Bennouna, S. Alkan, Y. Laouar, K. Gorden, R. Flavell, S. Akira, R. Ahmed, and B. Pulendran (2006). “Yellow fever vaccine YF-17D activates multiple dendritic cell subsets via TLR2, 7, 8, and 9 to stimulate polyvalent immunity.” English. In: *The Journal of Experimental Medicine* 203.2, pp. 413–424. DOI: 10.1084/jem.20051720.
- Rice, C. M., A. Grakoui, R. Galler, and T. J. Chambers (1989). “Transcription of infectious yellow fever RNA from full-length cDNA templates produced by in vitro ligation.” English. In: *The New biologist* 1.3, pp. 285–296.
- Rice, C. M., E. M. Lenches, S. R. Eddy, S. J. Shin, R. L. Sheets, and J. H. Strauss (1985). “Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution.” English. In: *Science (New York, N. Y.)* 229.4715, pp. 726–733.
- Ripoll, C., A. Ponce, M. M. Wilson, N. Sharif, J. B. Vides, J. Armoni, and D. E. Teuwen (2014). “Evaluation of two yellow fever vaccines for routine immunization programs in Argentina”. English. In: *Human Vaccines* 4.2, pp. 121–126. DOI: 10.4161/hv.4.2.5216.
- Roukens, A. H., D. Soonawala, S. A. Joosten, A. W. de Visser, X. Jiang, K. Dirksen, M. de Gruijter, J. T. van Dissel, P. J. Bredenbeek, and L. G. Visser (2011). “Elderly Subjects Have a Delayed Antibody Response and Prolonged Viraemia following Yellow Fever Vaccination: A Prospective Controlled Cohort Study”. English. In: *PLoS ONE* 6.12, e27753–6. DOI: 10.1371/journal.pone.0027753.
- Roukens, A. H., A. C. Vossen, P. J. Bredenbeek, J. T. van Dissel, and L. G. Visser (2008). “Intradermally Administered Yellow Fever Vaccine at Reduced Dose Induces a Protective Immune Response: A Randomized Controlled Non-Inferiority Trial”. English. In: *PLoS ONE* 3.4, e1993–8. DOI: 10.1371/journal.pone.0001993.
- Roukens, A. H., A. C. Vossen, J. T. van Dissel, and L. G. Visser (2009). “Reduced intradermal test dose of yellow fever vaccine induces protective immunity in individuals with egg allergy”. English. In: *Vaccine* 27.18, pp. 2408–2409. DOI: 10.1016/j.vaccine.2009.02.049.
- Rumyantsev, A. A., A. P. Goncalvez, M. Giel-Moloney, J. Catalan, Y. Liu, Q.-s. Gao, J. Almond, H. Kleanthous, and K. V. Pugachev (2013). “Single-dose vaccine against tick-borne encephalitis.” English. In: *Proceedings of the Na-*

- tional Academy of Sciences* 110.32, pp. 13103–13108. DOI: 10.1073/pnas.1306245110.
- Ryman, K. D., H. Xie, T. N. Ledger, G. A. Campbell, and A. D. Barrett (1997). “Antigenic variants of yellow fever virus with an altered neurovirulence phenotype in mice.” English. In: *Virology* 230.2, pp. 376–380. DOI: 10.1006/viro.1997.8496.
- Sall, A. A., O. Faye, M. Diallo, C. Firth, A. Kitchen, and E. C. Holmes (2010). “Yellow fever virus exhibits slower evolutionary dynamics than dengue virus.” English. In: *Journal of Virology* 84.2, pp. 765–772. DOI: 10.1128/JVI.01738-09.
- Samsa, M. M., J. A. Mondotte, N. G. Iglesias, I. Assunção-Miranda, G. Barbosa-Lima, A. T. Da Poian, P. T. Bozza, and A. V. Gamarnik (2009). “Dengue virus capsid protein usurps lipid droplets for viral particle formation.” English. In: *PLoS Pathogens* 5.10, e1000632. DOI: 10.1371/journal.ppat.1000632.
- Santos, C. N. dos, P. R. Post, R. Carvalho, I. I. Ferreira, C. M. Rice, and R. Galler (1995). “Complete nucleotide sequence of yellow fever virus vaccine strains 17DD and 17D-213.” English. In: *Virus Research* 35.1, pp. 35–41.
- Sawyer, W. A., S. F. Kitchen, and W. Lloyd (1932). “Vaccination Against Yellow Fever with Immune Serum and Virus Fixed for Mice.” English. In: *The Journal of Experimental Medicine* 55.6, pp. 945–969.
- Sawyer, W. A. and W. Lloyd (1931). “The Use of Mice in Tests of Immunity Against Yellow Fever.” English. In: *The Journal of Experimental Medicine* 54.4, pp. 533–555.
- Schäfer, B., G. W. Holzer, A. Joachimsthaler, S. Coulibaly, M. Schwendinger, B. A. Crowe, T. R. Kreil, P. N. Barrett, and F. G. Falkner (2011). “Pre-Clinical Efficacy and Safety of Experimental Vaccines Based on Non-Replicating Vaccinia Vectors against Yellow Fever”. English. In: *PLoS ONE* 6.9, e24505. DOI: 10.1371/journal.pone.0024505.t003.
- Schlesinger, J. J. and M. W. Brandriss (1990). “Cell surface expression of yellow fever virus non-structural glycoprotein NS1: consequences of interaction with antibody”. English. In: *Journal of General Virology* 71.3, pp. 593–599. DOI: 10.1099/0022-1317-71-3-593.
- Schuh, A. J., R. B. Tesh, and A. D. T. Barrett (2011). “Genetic characterization of Japanese encephalitis virus genotype II strains isolated from 1951 to 1978.”



- English. In: *Journal of General Virology* 92.Pt 3, pp. 516–527. DOI: 10.1099/vir.0.027110-0.
- Sellards, A. W. and E. Hindle (1928). “The Preservation of Yellow Fever Virus.” English. In: *British Medical Journal* 1.3512, pp. 713–714.
- Sil, B. K., L. M. Dunster, T. N. Ledger, M. R. Wills, P. D. Minor, and A. D. Barrett (1992). “Identification of envelope protein epitopes that are important in the attenuation process of wild-type yellow fever virus.” English. In: *Journal of Virology* 66.7, pp. 4265–4270.
- Silva, M. L., L. R. Espírito-Santo, M. A. Martins, D. Silveira-Lemos, V. Peruhype-Magalhães, R. C. Caminha, P. de Andrade Maranhão-Filho, M. Auxiliadora-Martins, R. de Menezes Martins, R. Galler, M. da Silva Freire, R. Marcovistz, A. Homma, D. E. Teuwen, S. M. Elói-Santos, M. C. Andrade, A. Teixeira-Carvalho, and O. A. Martins-Filho (2010). “Clinical and immunological insights on severe, adverse neurotropic and viscerotropic disease following 17D yellow fever vaccination.” English. In: *Clinical and Vaccine Immunology* 17.1, pp. 118–126. DOI: 10.1128/CVI.00369-09.
- Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and I. Birol (2009). “ABYSS: a parallel assembler for short read sequence data.” English. In: *Genome Research* 19.6, pp. 1117–1123. DOI: 10.1101/gr.089532.108.
- Sobrinho, F., M. Dávila, J. Ortín, and E. Domingo (1983). “Multiple genetic variants arise in the course of replication of foot-and-mouth disease virus in cell culture.” English. In: *Virology* 128.2, pp. 310–318.
- Song, B.-H., G.-N. Yun, J.-K. Kim, S.-I. Yun, and Y.-M. Lee (2012). “Biological and genetic properties of SA14-14-2, a live-attenuated Japanese encephalitis vaccine that is currently available for humans.” English. In: *Journal of Microbiology (Seoul, Korea)* 50.4, pp. 698–706. DOI: 10.1007/s12275-012-2336-6.
- Soper, F. L. and H. H. Smith (1938). “Vaccination with virus 17D in the control of jungle yellow fever in Brazil”. In: *Third International Congress of Tropical Medicine and Malaria*, pp. 295–313.
- Staples, J. E., M. Gershman, and M. Fischer (2010). “Yellow fever vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP).” English. In: *Morbidity and Mortality Weekly Report* 59.RR-7.
- Stock, N. K., N. Boschetti, C. Herzog, M. S. Appelhans, and M. Niedrig (2012). “The phylogeny of yellow fever virus 17D vaccines.” English. In: *Vaccine* 30.6, pp. 989–994. DOI: 10.1016/j.vaccine.2011.12.057.

- Stoddard, M. B., H. Li, S. Wang, M. Saeed, L. Andrus, W. Ding, X. Jiang, G. H. Learn, M. von Schaewen, J. Wen, P. A. Goepfert, B. H. Hahn, A. Ploss, C. M. Rice, and G. M. Shaw (2015). “Identification, Molecular Cloning, and Analysis of Full-Length Hepatitis C Virus Transmitted/Founder Genotypes 1, 3, and 4”. English. In: *mBio* 6.2, e02518–14–15. DOI: 10.1128/mBio.02518–14.
- Stokes, A., J. H. Bauer, and N. P. Hudson (1928). “Experimental transmission of yellow fever to laboratory animals”. In: *The American Journal of Tropical Medicine and Hygiene* 1.2, pp. 103–164.
- Stones, P. B. and F. N. Macnamara (1955). “Encephalitis following neurotropic yellow fever vaccine administered by scarification in Nigeria: epidemiological and laboratory studies.” English. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 49.2, pp. 176–186.
- Stuart, G. (1953). *The Problem of Mass Vaccination Against Yellow Fever*. English. Tech. rep. WHO/YF/20. Expert Committee on Yellow Fever.
- Sumiyoshi, H., C. H. Hoke, and D. W. Trent (1992). “Infectious Japanese encephalitis virus RNA can be synthesized from in vitro-ligated cDNA templates.” English. In: *Journal of Virology* 66.9, pp. 5425–5431.
- Sumiyoshi, H., G. H. Tignor, and R. E. Shope (1995). “Characterization of a Highly Attenuated Japanese Encephalitis Virus Generated from Molecularly Cloned cDNA”. English. In: *The Journal of Infectious Diseases* 171.5, pp. 1144–1151.
- Szpara, M. L., Y. R. Tafuri, L. Parsons, S. R. Shamim, K. J. Verstrepen, M. Legendre, and L. W. Enquist (2011). “A Wide Extent of Inter-Strain Diversity in Virulent and Vaccine Strains of Alphaherpesviruses”. English. In: *PLoS Pathogens* 7.10, e1002282. DOI: 10.1371/journal.ppat.1002282.t005.
- Tauraso, N. M., S. L. Spector, W. G. Jahnes, and A. Shelokov (1968). “Yellow fever vaccine. I. Development of a vaccine seed free from contaminating avian leukosis viruses.” English. In: *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)* 127.4, pp. 1116–1120.
- Tauraso, N. M., S. L. Spector, and R. W. Trimmer (1972). “Yellow fever vaccine. 3. Antibody response in monkeys inoculated with a vaccine free from contaminating avian leukosis viruses.” English. In: *Archiv für die gesamte Virusforschung* 37.2, pp. 211–217.
- Tesh, R. B., H. Guzman, A. P. da Rosa, P. F. Vasconcelos, L. B. Dias, J. E. Bunnell, H. Zhang, and S. Y. Xiao (2001). “Experimental yellow fever virus infection

- in the Golden Hamster (*Mesocricetus auratus*). I. Virologic, biochemical, and immunologic studies.” English. In: *The Journal of Infectious Diseases* 183.10, pp. 1431–1436. DOI: 10.1086/320199.
- Theiler, M. (1930). “Susceptibility of White Mice to the Virus of Yellow Fever.” English. In: *Science (New York, N.Y.)* 71.1840, p. 367. DOI: 10.1126/science.71.1840.367.
- Theiler, M. and H. H. Smith (1937). “The Effect of Prolonged Cultivation in Vitro Upon the Pathogenicity of Yellow Fever Virus.” English. In: *The Journal of Experimental Medicine* 65.6, pp. 767–786.
- Theiler, M. (1931). “Neutralization tests with immune yellow fever sera and a strain of yellow fever virus adapted to mice”. In: *Annals of Tropical Medicine and Parasitology* 25, pp. 69–77.
- Theiler, M. (1951). “The Virus”. In: *Yellow Fever*. Ed. by G. K. Strode. New York, pp. 45–135.
- UNRAA Expert Commission on Quarantine (1945). “Standards for the Manufacture and Control of Yellow Fever Vaccine”. In: *World Health Organ Epidemiol Bull* 1, pp. 365–370.
- Vainio, J. and F. Cutts (1998). *Yellow Fever*. Tech. rep. Geneva: World Health Organization.
- Victoria, J. G., C. Wang, M. S. Jones, C. Jaing, K. McLoughlin, S. Gardner, and E. L. Delwart (2010). “Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus.” English. In: *Journal of Virology* 84.12, pp. 6033–6040. DOI: 10.1128/JVI.02690-09.
- Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino (2006). “Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population.” English. In: *Nature* 439.7074, pp. 344–348. DOI: 10.1038/nature04388.
- Villar, L., G. H. Dayan, J. L. Arredondo-García, D. M. Rivera, R. Cunha, C. Deseda, H. Reynales, M. S. Costa, J. O. Morales-Ramírez, G. Carrasquilla, L. C. Rey, R. Dietze, K. Luz, E. Rivas, M. C. Miranda Montoya, M. Cortés Supelano, B. Zambrano, E. Langevin, M. Boaz, N. Tornieporth, M. Saville, and F. Noriega (2015). “Efficacy of a Tetravalent Dengue Vaccine in Children in Latin America”. English. In: *The New England Journal of Medicine* 372.2, pp. 113–123. DOI: 10.1056/NEJMoa1411037.

- Vobmann, S., J. Wieseler, R. Kerber, and B. M. Kümmerer (2015). “A basic cluster in the N terminus of yellow fever virus NS2A contributes to infectious particle production.” English. In: *Journal of Virology* 89.9, pp. 4951–4965. DOI: 10.1128/JVI.03351-14.
- Vratskikh, O., K. Stiasny, J. Zlatkovic, G. Tsouchnikas, J. Jarmer, U. Karrer, M. Roggendorf, H. Roggendorf, R. Allwinn, and F. X. Heinz (2013). “Dissection of antibody specificities induced by yellow fever vaccination.” English. In: *PLoS Pathogens* 9.6, e1003458. DOI: 10.1371/journal.ppat.1003458.
- Wang, E., K. D. Ryman, A. D. Jennings, D. J. Wood, F. Taffs, P. D. Minor, P. G. Sanders, and A. D. Barrett (1995). “Comparison of the genomes of the wild-type French viscerotropic strain of yellow fever virus with its vaccine derivative French neurotropic vaccine.” English. In: *Journal of General Virology* 76 ( Pt 11), pp. 2749–2755.
- Wengler, G. (1989). “Cell-associated West Nile flavivirus is covered with E+pre-M protein heterodimers which are destroyed and reorganized by proteolytic cleavage during virus release.” English. In: *Journal of Virology* 63.6, pp. 2521–2526.
- Whitman, L. (2015). “Failure of Aedes Aegypti To Transmit Yellow Fever Cultured Virus (17D)”. In: *American Journal of Tropical Medicine* 19, pp. 19–26.
- Whittembury, A., G. Ramirez, H. Hernández, A. M. Roper, S. Waterman, M. Ticona, M. Brinton, J. Uchuya, M. Gershman, W. Toledo, E. Staples, C. Campos, M. Martínez, G.-J. J. Chang, C. Cabezas, R. Lanciotti, S. Zaki, J. M. Montgomery, T. Monath, and E. Hayes (2009). “Viscerotropic disease following yellow fever vaccination in Peru.” English. In: *Vaccine* 27.43, pp. 5974–5981. DOI: 10.1016/j.vaccine.2009.07.082.
- World Health Organization (1946). “Report to the UNRAA Expert Commission on Quarantine: Dakar Yellow-Fever Vaccine”. In: *Epidemiological Information Bulletin*, pp. 618–637.
- World Health Organization (2003). *WHO-Recommended Standards for Surveillance of Selected Vaccine Preventable Diseases*, pp. 1–52.
- World Health Organization (2006). *International health regulations (2005)*. English. Geneva, pp. 1–74.
- World Health Organization (2010). *WHO Expert Committee on Biological Standardization: Sixty-First Report*. English. Tech. rep. TRS 978, pp. 1–384.

- World Health Organization (2013). “Meeting of the Strategic Advisory Group of Experts on immunization, April 2013 – conclusions and recommendations.” English. In: *Weekly Epidemiological Record* 88.20, pp. 201–206.
- World Health Organization (2014). *WHO Expert Committee on Biological Standardization. Sixty-third report*. English. Tech. rep. TRS 980, pp. 1–489.
- Wu, S. J., G. Grouard-Vogel, W. Sun, J. R. Mascola, E. Brachtel, R. Putvatana, M. K. Louder, L. Filgueira, M. A. Marovich, H. K. Wong, A. Blauvelt, G. S. Murphy, M. L. Robb, B. L. Innes, D. L. Birx, C. G. Hayes, and S. S. Frankel (2000). “Human skin Langerhans cells are targets of dengue virus infection.” English. In: *Nature Medicine* 6.7, pp. 816–820. DOI: 10.1038/77553.
- Xiao, S. Y., H. Zhang, H. Guzman, and R. B. Tesh (2001). “Experimental yellow fever virus infection in the Golden hamster (*Mesocricetus auratus*). II. Pathology.” English. In: *The Journal of Infectious Diseases* 183.10, pp. 1437–1444. DOI: 10.1086/320200.
- Xie, H., A. R. Cass, and A. D. Barrett (1998). “Yellow fever 17D vaccine virus isolated from healthy vaccinees accumulates very few mutations.” English. In: *Virus Research* 55.1, pp. 93–99.
- Xie, H., K. D. Ryman, G. A. Campbell, and A. D. Barrett (1998). “Mutation in NS5 protein attenuates mouse neurovirulence of yellow fever 17D vaccine virus.” English. In: *The Journal of General Virology* 79 ( Pt 8).8, pp. 1895–1899. DOI: 10.1099/0022-1317-79-8-1895.
- Yang, D., X.-F. Li, Q. Ye, H.-J. Wang, Y.-Q. Deng, S.-Y. Zhu, Y. Zhang, S.-H. Li, and C.-F. Qin (2014). “Characterization of live-attenuated Japanese encephalitis vaccine virus SA14-14-2”. English. In: *Vaccine* 32.23, pp. 2675–2681. DOI: 10.1016/j.vaccine.2014.03.074.
- Yang, X., P. Charlebois, A. Macalalad, M. R. Henn, and M. C. Zody (2013). “V-Phaser 2: variant inference for viral populations”. English. In: *BMC Genomics* 14.1, p. 674. DOI: 10.1186/1471-2164-14-674.
- Yang, Z. (1993). “Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites.” English. In: *Molecular Biology and Evolution* 10.6, pp. 1396–1401.
- Yin, Z., X. Wang, L. Li, H. Li, X. Zhang, J. Li, G. Ning, F. Li, X. Liang, L. Gao, X. Liang, and Y. Li (2015). “Neurological Sequelae of Hospitalized Japanese Encephalitis Cases in Gansu Province, China”. English. In: *American Journal of Tropical Medicine and Hygiene* 92.6, pp. 1125–1129. DOI: 10.4269/ajtmh.14-0148.

- Youn, S., R. L. Ambrose, J. M. Mackenzie, and M. S. Diamond (2013). "Non-structural Protein-1 is Required for West Nile Virus Replication Complex Formation and Viral RNA Synthesis." English. In: *Virology Journal* 10.1, p. 339. DOI: 10.1186/1743-422X-10-339.
- Youn, S., T. Li, B. T. McCune, M. A. Edeling, D. H. Fremont, I. M. Cristea, and M. S. Diamond (2012). "Evidence for a Genetic and Physical Interaction Between Nonstructural Proteins NS1 and NS4B that Modulates Replication of West Nile Virus." English. In: *Journal of Virology* 86.13, pp. 7360–7371. DOI: 10.1128/JVI.00157-12.
- Yun, S.-I., B.-H. Song, J.-K. Kim, G.-N. Yun, E.-Y. Lee, L. Li, R. J. Kuhn, M. G. Rossmann, J. D. Morrey, and Y.-M. Lee (2014). "A molecularly cloned, live-attenuated japanese encephalitis vaccine SA14-14-2 virus: a conserved single amino acid in the ij Hairpin of the Viral E glycoprotein determines neurovirulence in mice." English. In: *PLoS Pathogens* 10.7, e1004290. DOI: 10.1371/journal.ppat.1004290.
- Zhao, X., M. Cao, H.-H. Feng, H. Fan, F. Chen, Z. Feng, X. Li, and X.-H. Zhou (2014). "Japanese encephalitis risk and contextual risk factors in southwest China: a Bayesian hierarchical spatial and spatiotemporal analysis." English. In: *International Journal of Environmental Research and Public Health* 11.4, pp. 4201–4217. DOI: 10.3390/ijerph110404201.
- Zhou, Y., D. Ray, Y. Zhao, H. Dong, S. Ren, Z. Li, Y. Guo, K. A. Bernard, P.-Y. Shi, and H. Li (2007). "Structure and function of flavivirus NS5 methyltransferase." English. In: *Journal of Virology* 81.8, pp. 3891–3903. DOI: 10.1128/JVI.02704-06.
- Zou, G., Y. L. Chen, H. Dong, C. C. Lim, L. J. Yap, Y. H. Yau, S. G. Shochat, J. Lescar, and P. Y. Shi (2011). "Functional Analysis of Two Cavities in Flavivirus NS5 Polymerase". English. In: *Journal of Biological Chemistry* 286.16, pp. 14362–14372. DOI: 10.1074/jbc.M110.214189.
- Zou, J., X. Xie, L. T. Lee, R. Chandrasekaran, A. Reynaud, L. Yap, Q.-Y. Wang, H. Dong, C. Kang, Z. Yuan, J. Lescar, and P.-Y. Shi (2014). "Dimerization of flavivirus NS4B protein." English. In: *Journal of Virology* 88.6, pp. 3379–3391. DOI: 10.1128/JVI.02782-13.
- Zou, J., X. Xie, Q.-Y. Wang, H. Dong, M. Y. Lee, C. Kang, Z. Yuan, and P.-Y. Shi (2015). "Characterization of dengue virus NS4A and NS4B protein interaction." English. In: *Journal of Virology* 89.7, pp. 3455–3470. DOI: 10.1128/JVI.03453-14.

## Vita

Andrew Stephen Beck was born in Baltimore, Maryland on February 16th, 1982, the son of Leon Paul Beck, Jr. and Denise Ann Beck. He received the Bachelor of Science degree in 2004 from Loyola College in Baltimore, Maryland, majoring in biology and minoring in classical civilization. Following undergraduate studies, he held a number of technical positions in the defense sector in the Baltimore/Washington area, focusing initially on environmental chemistry and bioremediation, finally taking a position at the United States Army Medical Research Institute of Infectious Diseases (USAMRIID) in Frederick, MD. He then enrolled at the University of Texas Medical Branch in 2009, pursuing doctoral studies through the Experimental Pathology program of the graduate school of biomedical sciences. He married Courtney Moylan Toms on June 7, 2008.

Permanent address: 1414 Church Street  
Upper Apartment  
Galveston, TX 77550

This dissertation was typeset with L<sup>A</sup>T<sub>E</sub>X<sup>†</sup> by the author.

---

<sup>†</sup>L<sup>A</sup>T<sub>E</sub>X is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's T<sub>E</sub>X Program.

## Curriculum Vitae

### Andrew Stephen Beck

andrew.s.beck@gmail.com

1414 Church Street  
Upper Apartment  
Galveston, TX 77550  
(443)865-5372

University of Texas Medical Branch  
Department of Pathology  
Barrett Laboratory  
MS 0609  
(409)772-2547

#### Education

2009–present	<b>PhD Candidate, Experimental Pathology</b> University of Texas Medical Branch Galveston, TX
2000–2004	<b>BS, Biology</b> Loyola College Baltimore, MD

#### Grants and Support

2013–2015	Training support from the University of Texas Medical Branch Biodefense Training Program NIH Grant T32 AI060945
-----------	---

#### Publications

Beck, A., R. B. Tesh, T. G. Wood, S. G. Widen, K. D. Ryman, and A. D. T. Barrett (2014). “Comparison of the live attenuated yellow fever vaccine 17D-204 strain to its virulent parental strain Asibi by deep sequencing.” English. In: *Journal of Infectious Diseases* 209.3, pp. 334–344. DOI: 10.1093/infdis/jit546.



Beck, A., H. Guzman, L. Li, B. Ellis, R. B. Tesh, and A. D. T. Barrett (2013). “Phylogeographic Reconstruction of African Yellow Fever Virus Isolates Indicates Recent Simultaneous Dispersal into East and West Africa”. English. In: *PLoS Neglected Tropical Diseases* 7.3, e1910. DOI: 10.1371/journal.pntd.0001910.s004.

## Reviews

Beck, A. S. and A. D. Barrett (2015). “Current Status and Future Prospects of Yellow Fever Vaccines.” English. In: *Expert Review of Vaccines*, pp. 1–14. DOI: 10.1586/14760584.2015.1083430.

## Extramural Honors

2014	American Society for Virology Travel Award
2014	Finalist, Maurice R. Hilleman Early Career Investigator Award National Foundation for Infectious Diseases
2011	American Society for Virology Travel Award
2011	Who’s Who Among American Students in American Colleges and Universities

## Intramural Honors In The Course of Doctoral Study

2015	Pathology Day Poster Award
2015	James W. McLaughlin Colloquium Travel Award
2015	Sealy Center for Vaccine Development Travel Grant
2014	Sealy Center for Vaccine Development Travel Grant
2014	Pathology Day Poster Award
2013	Center for Tropical Diseases Graduate Student Award
2013	Robert Bennett Tuition Scholarship
2012	Sealy Center for Vaccine Development Travel Grant
2012	James W. McLaughlin Colloquium Travel Award
2012	Pathology Day Poster Award
2011	James W. McLaughlin Colloquium Travel Award

## Abstracts

- Beck, A., S. G. Widen, J. Thompson, and A. D. Barrett (2015). "Massively Parallel Sequencing of 17D Strain Yellow Fever Vaccines Identifies Stable Population Structure and Limited Genetic Divergence Along the Vaccine Seed Lineage". In: *The Annual Conference on Vaccine Research*. National Foundation for Infectious Diseases. Bethesda.
- Beck, A., R. B. Tesh, S. G. Widen, T. G. Wood, J. Thompson, and A. D. T. Barrett (2014). "Deep Sequencing Along the Lineage of the Yellow Fever Vaccine Strain 17D Reveals Low Diversity and Population Stability For Primary and Secondary Seed Lots". In: *Annual Meeting of the International Society of Vaccines*. International Society of Vaccines. Philadelphia.
- Beck, A., R. B. Tesh, S. G. Widen, T. G. Wood, J. Thompson, and A. D. Barrett (2014). "Deep Sequencing Along the Lineage of the Yellow Fever Vaccine Strain 17D Reveals Low Diversity and Population Stability For Primary and Secondary Seed Lots". In: *Annual Meeting of the American Society for Virology*. American Society for Virology. Fort Collins.
- Beck, A., R. B. Tesh, T. G. Wood, S. G. Widen, and A. D. Barrett (2014). "Population Structure and *In Vitro* Genomic Stability of the Live Attenuated Yellow Fever Vaccine 17D-204 by Deep Sequencing". In: *The Annual Conference on Vaccine Research*. National Foundation for Infectious Diseases. Bethesda.
- Beck, A., H. Guzman, L. Li, B. R. Ellis, R. B. Tesh, and A. D. Barrett (2013). "Comparison of the Live Attenuated Yellow Fever Vaccine 17D-204 to its Virulent Parental Strain Asibi by Deep Sequencing". In: *The Annual Conference on Vaccine Research*. National Foundation for Infectious Diseases. Baltimore.
- Beck, A., H. Guzman, L. Li, B. R. Ellis, R. B. Tesh, and A. D. Barrett (2012a). "Bayesian Phylogeographic Reconstruction Using African Yellow Fever Virus Isolates Indicates Recent Simultaneous Dispersal of East and West Lineages". In: *Annual Meeting of the American Society of Tropical Medicine and Hygiene*. American Society of Tropical Medicine and Hygiene. Atlanta.
- Beck, A., H. Guzman, L. Li, B. R. Ellis, R. B. Tesh, and A. D. Barrett (2012b). "Bayesian Phylogeographic Reconstruction Using African Yellow Fever Virus Isolates Indicates Recent Simultaneous Dispersal of East and West Lineages". In: *Annual Meeting of the American Society for Virology*. American Society for Virology. Madison.

## Related Professional Experience

2008–2009	<b>United States Army Medical Research Institute of Infectious Diseases (USAMRIID)</b> <b>Frederick, MD</b> <i>Laboratory Technician, Aerobiology Division</i>
2004–2008	<b>United States Army Edgewood Chemical Biological Center</b> <b>Edgewood, MD</b> <i>Laboratory Technician</i>

## Teaching Experience

2011	<b>University of Texas Medical Branch, School of Medicine</b> <b>Course Title:</b> Pathobiology and Host Defenses 1011 <b>Duties:</b> Laboratory instruction of first-year students, for introductory coursework in microbial pathology.
2008	<b>George Mason University</b> <b>Course Title:</b> BIOL 619, Topics in Aerosol Biology <b>Duties:</b> Laboratory instruction for a graduate-level course on aerosol biology methods. Responsibilities included instruction of small groups and grading of technical proficiencies.

## Professional Affiliations

2011–present	American Society for Virology
--------------	-------------------------------