Comparison of the Live Attenuated Yellow Fever Vaccine 17D-204 Strain to Its Virulent Parental Strain Asibi by Deep Sequencing

Andrew Beck,1,2 Robert B. Tesh,1,2 Thomas G. Wood,3 Steven G. Widen,3 Kate D. Ryman,4 and Alan D. T. Barrett1,2

1Department of Pathology, 2Sealy Center for Vaccine Development, and 3Molecular Genomics Core Facility, University of Texas Medical Branch, Galveston; and 4Center for Vaccine Research, University of Pittsburgh, Pennsylvania

(See the editorial commentary by Tangy and Desprès on pages 318–20.)

Background. The first comparison of a live RNA viral vaccine strain to its wild-type parental strain by deep sequencing is presented using as a model the yellow fever virus (YFV) live vaccine strain 17D-204 and its wild-type parental strain, Asibi.

Methods. The YFV 17D-204 vaccine genome was compared to that of the parental strain Asibi by massively parallel methods. Variability was compared on multiple scales of the viral genomes. A modeled exploration of small-frequency variants was performed to reconstruct plausible regions of mutational plasticity.

Results. Overt quasispecies diversity is a feature of the parental strain, whereas the live vaccine strain lacks diversity according to multiple independent measurements. A lack of 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 wild-type population.

Conclusions. Relative quasispecies structure is a plausible correlate of attenuation for live viral vaccines. 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 vaccines and the risk of reversion to virulence.

Keywords. Yellow Fever Virus; Vaccine; Quasispecies; Flavivirus; 17D, Viral Population; Variants; Deep Sequencing; Parallel Sequencing.

All of the live attenuated vaccines in use today were derived empirically, and our understanding of the 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 [1–3]. 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 [4, 5]. In this context, the yellow fever virus (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 [6]. 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 [7]. 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 [8]. Our results offer the first comprehensive coverage depth (>1000), permitting direct comparison of low-frequency variants between the rare, archived parental strain and commercial vaccine derivative.

YFV is a mosquito-borne flavivirus that is the prototype member of the family Flaviviridae. It has a
single-stranded, positive-sense RNA genome of 10.8 kb, which encodes 10 genes (C, prM, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). Wild-type strains of YFV cause a febrile syndrome that may proceed to a toxic phase featuring liver damage and hemorrhage, termed viscerotropic disease. The jaundice that occurs in advanced human cases is the yellow color from which the disease derives its name. Mortality rates for human YFV infection are estimated to be 20%–44% for patients entering the toxic phase [9, 10].

The vaccine was derived from the wild-type 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 [11]. 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 [10]. The 17D-204 vaccine strain was derived from the original 17D strain at passage 204 and is now presently manufactured in the United States, France, Senegal, Russia, and China [12]. Standardization of 17D-derived vaccine is achieved by a seed-lot system, ensuring that all production lots are derived from seed lots by 1 passage in embryonated chicken eggs [12]. The substitutions contributing to the attenuated phenotype of the 17D-204 vaccine are unknown. By consensus, comparison of the Asibi and 17D-204 genomes identified 67 nucleotide differences, encoding for 31 amino acid substitutions [13, 14]. Reasonably, it is hypothesized that the historical data set of mutations differentiating Asibi and 17D-204 represent determinants of mammalian pathogenicity [15]. However, to date, the specific amino acids contributing to the attenuation of 17D have not been elucidated.

METHODS

Viruses
The Asibi strain of YFV was obtained as lyophilized cell culture supernatant from the World Reference Center for Emerging Viruses and Arboviruses (Galveston, TX). The passage history of Asibi strain virus was as follows from the original isolate: 6 times in rhesus macaques (Macaca mulatta) [16], 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, lot UF795AA, Sanofi Pasteur), using the provided injection diluent. Asibi virus was reconstituted in sterile, deionized water.

Amplification and Sequencing
Viral RNA was extracted using the QIAGen viral RNA column isolation kit (Qiagen, Gaithersburg, MD). Reverse-transcription polymerase chain reaction (RT-PCR) amplicons were generated using the listed primers and amplification strategy (Supplementary Table 1), which were designed in reference to Genbank accession X03700.1, the earliest entry for 17D [13]. Overlapping RT-PCR amplicons were generated from extracted viral RNA directly, without intervening passage of either virus. Six amplicons were designed to be at least 2 kb in size, covering the entire viral genome, with 250-b overlap. The amplification program was as follows: reverse transcription at 50°C for 30 minutes, initial denaturation at 94°C for 2 minutes, and 40 cycles of denaturation at 94°C for 10 seconds, annealing at a primer-dependent temperature (Supplementary Table 1) for 30 seconds, and extension at 68°C for 2 minutes. Final extension was at 68°C for 7 minutes. Amplicons were diluted to molar equivalence, and indexed complementary DNA libraries were generated. Libraries were sequenced on an Illumina Hi-Seq-1000 instrument, obtaining paired-end, 50-b reads.

Data Pipeline
Trimmed reads were filtered for average quality scores of >20 (Illumina 1.9 encoding). Three bases were trimmed from the 5’ ends of all reads. De novo assembly was performed with ABySS v.1.3.4, using the paired-end function, and k-mer sizes from 20–40. Full-length consensus sequences were chosen from the longest contigs [17]. Primers were trimmed with Trimmomatic v.0.30, using the “Illuminaclip” function [18]. Total read counts were matched by proportional random downsampling of the Asibi strain read set to match read counts returned from the 17D-204 set. Downsampling was repeated 5 times. Realignment were performed with Bowtie2 v.2.1.0, using end-to-end mode and very sensitive presets [19]. File conversions, indexing, and pileup file generation were performed with Samtools v.0.1.19, with base-alignment-quality filtering disabled [20].

Conservative Variant Exploration
Alignments for Asibi and 17D-204 read sets were explored for variant populations using a heuristic variant frequency cutoff of 1%, excluding variants with a strand bias of >90% and retaining bases with quality scores of >30. Variant calls were made using Varscan v.2.3.5 [21]. Each of the 5 downsammpled Asibi strain alignments were analyzed by the same procedure, retaining variants that were recovered from all data sets. Reverse variant exploration of Asibi and 17D-204 alignments was performed for positions already described by consensus to differentiate the strains [13,14].

Intersection of Viral Population Identity
Intersection of virus alignment identity between Asibi and 17D-204 was estimated in R v.3.0.1, using adaptations of libraries deepSNV v.1.6.0, and Biostrings v.2.28.0, which are components of the Bioconductor library collection [22]. Intersection was expressed as the frequency of a variant matching the greatest frequency nucleotide identity of the opposing alignment.
Measurement of Absolute Diversity

Diversity indices were calculated with custom R scripts over the possible nucleotide character set \{A,U,C,G,-\}. Shannon entropy was calculated using the method described by Nishijima et al [23]. Simpson’s diversity index (1-D) was calculated using the method described by Hunter et al [24]. Indices for Asibi and 17D-204 strains were compared nonparametrically along 3 scales: along the full genome length, along discrete genes, and at the consensus mutation sites observed by Hahn et al [13]. Error rates were computed along all sites for both alignments as the total frequency of all base calls not matching the consensus. Quantile-quantile plots were prepared to assess differences in extent and distribution of error rates along the full genomes and along discrete genes and as a control for low-end coverages.

Paired-Samples Variant Calls

Variants were called under a paired-samples design, using Asibi reads as the control case and 17D-204 reads as the test case. Asibi and 17D-204 reads were aligned to the de novo-assembled Asibi strain genome consensus. Mutant frequencies were called using an adaptation of the R library deepSNV v.1.6.0 [25]. Frequencies were computed using base calls with quality scores of >30. Variants with frequency changes returning Bonferroni-corrected \( P \) values of <.05 were considered statistically significant. Synonymous, synonymous-degenerate, and nonsynonymous substitutions for each significantly represented variant were called using the Perl script SNPdat v.1.0.5 [26]. A sliding window average (window size = 10) of the recovered \( P \) values from the paired test was calculated with the R library Zoo v.1.7-10. Venn diagrams were generated from set inclusion comparisons of returned mutations to those listed in the parent and attenuated mutation sets of the French neurotropic vaccine [27] and the HeLa p6 virus [28].

Statistical Analysis

Estimation of Nucleotide Variability by Diversity Indices

For each scale of comparison (whole genome, discrete genes, and vaccine mutation sites), Shannon entropy and Simpson’s 1-D were compared nonparametrically, using a Mann-Whitney \( U \) test. Comparisons returning \( P \) values of <.05 were considered to be statistically significant.

Paired Tests

For paired tests of mutation frequency change, mutation frequencies were assessed across each site of the control (Asibi/
Figure 1. 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. Downsampling Asibi read sets were averaged and floor-rounded by counts. A, Graph of complete Shannon entropy and Simpson’s 1-D across both genomes, with consensus vaccine mutation positions for reference (black dots). B, Box plots of diversity indices computed for every nucleotide position along the entire genomes for Asibi and 17D-204 strains, and compared. C, Box plots of diversity indices for the vaccine mutations sites observed by Hahn et al [13], and compared. All box plots depict the median, first quartile, third quartile, with whiskers showing ± 1.5 times the interquartile distance. *P<.05. Abbreviation: NS, not significant.
parental) and test (17D-204/vaccine) viruses, using the algorithm of Gerstung et al [25]. Briefly, P values of sense and antisense strand frequencies from the control and test genome alignments were computed under the expectation that nucleotide frequencies along the genome are binomially distributed, with opposing strands then combined using the method of Fisher. The combined P values underwent Bonferroni correction. Loess smoothing was used to provide correction from biases between the 2 sequencing runs. Changes in mutation frequency between control (Asibi) and test (17D-204) read sets with P values of <.5 were considered statistically significant and were retained for downstream analysis.

RESULTS

Assembly and Inspection of Viral Genomes

Nucleotide counts of 5 randomly downsampled Asibi strain alignments were averaged and floor-rounded along all sites, and the median coverage of this alignment was 5531 (mean [±SD], 5562.2 ± 2054.1). Median coverage depth for the 17D-204 alignment was 5144 (mean [±SD], 5682.1 ± 2434.6).

Consensus sequences generated by plurality from de novo assemblies were compared to reference GenBank sequences for the Asibi and 17D-204 vaccines. Following comparison of our de novo–assembled Asibi strain consensus to that of the published Asibi sequence (gbAY40589), recovery of nucleotide positions 11–10 843 was observed, with the presence of 1 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 (gbJX503529.1). Recovery of nucleotide positions 14–10 850 for the 17D-204 consensus was observed. Relative to a 17D-204 sequence published by Xie et al (gbAF052437.1) [29], 2 nonsynonymous mutations were observed at positions U7496C and C7497U, both independently coding for the same amino acid substitution NS4B S67P.

Heuristic Identification of Variants

For the Asibi strain alignment, 55 variant sites, distributed over the entire genome, meeting the inclusion criteria were observed (Table 1 and Supplementary Table 2). 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 1 and Supplementary Table 3).

Comparison of Quasispecies Diversity

Estimates of diversity were compared at multiple scales (Figure 1A). For Shannon entropy, median values for Asibi and 17D-204 strains were 0.0179 and 0.0170, respectively, and strains were significantly different (U = 1 69 189; P = 1.60e–09; Figure 1B). 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 = 1 69 189; P = 1.60e–09; Figure 1B). For Simpson’s 1-D, median values for the mutation sites and 0.006, respectively, and strains were not significantly different (U = 2611; P = .62; Figure 1C).

Table 2. Comparisons of Diversity Indices for Asibi and 17D-204 Alignments, Using the Mann-Whitney U Test

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length, b</th>
<th>Shannon Entropy</th>
<th>Simpson’s 1-D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asibi</td>
<td>17D-204</td>
<td>Mann-Whitney U</td>
</tr>
<tr>
<td>5′ UTR</td>
<td>119</td>
<td>0.020</td>
<td>0.019</td>
</tr>
<tr>
<td>C</td>
<td>363</td>
<td>0.014</td>
<td>0.016</td>
</tr>
<tr>
<td>prM</td>
<td>492</td>
<td>0.014</td>
<td>0.016</td>
</tr>
<tr>
<td>E</td>
<td>1479</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>NS1</td>
<td>1056</td>
<td>0.014</td>
<td>0.019</td>
</tr>
<tr>
<td>NS2A</td>
<td>672</td>
<td>0.022</td>
<td>0.017</td>
</tr>
<tr>
<td>NS2B</td>
<td>390</td>
<td>0.018</td>
<td>0.014</td>
</tr>
<tr>
<td>NS3</td>
<td>1869</td>
<td>0.019</td>
<td>0.017</td>
</tr>
<tr>
<td>NS4A</td>
<td>447</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>NS4B</td>
<td>750</td>
<td>0.022</td>
<td>0.017</td>
</tr>
<tr>
<td>NS5</td>
<td>2715</td>
<td>0.020</td>
<td>0.017</td>
</tr>
<tr>
<td>3′ UTR</td>
<td>510</td>
<td>0.018</td>
<td>0.015</td>
</tr>
</tbody>
</table>

All P values <.05 were considered statistically significant. For both Shannon entropy and Simpson’s 1-D, the observation of significance for specific gene regions comparisons was concordant.

Abbreviation: UTR, untranslated region.
Figure 2. Quantile-quantile comparisons of error rate along the multiple genome scales for Asibi and 17D-204 strains. Downsampeed 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. A, Full genome. B, Local comparisons along gene segments. Abbreviation: UTR, untranslated region.
Significant strain differences were observed for local comparison of 3′ untranslated regions and for viral genes C, prM, E, NS1, NS2A, NS2B, NS3, NS4B, and NS5 (Table 2). 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 2A and 2B).

**Intersection of Quasispecies Identity**

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 3A). Of these 8 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 than vice versa (Figure 3B). Considering the variant frequency at 74 (67 fixed and 7 clonal) sites differentiating Asibi and 17D-204 (Figure 2C) [13], Asibi identity intersects with the consensus 17D-204 identity at 2.89% of total coverage in the parental read set (n = 11 196), whereas 17D-204 identity intersects with the Asibi consensus at 0.42% of total coverage in the vaccine read set (n = 1595; Figure 3B).

![Figure 3. Intersection of variant identity between Asibi and 17D-204 strain quasispecies is unidirectional. Asibi frequencies are averaged for downsampled read sets. A, Bar plot showing sites of overt quasispecies identity at sites that are fixed in the 17D-204 strain. B, Box plot summary of intersection frequency for Asibi and 17D-204 strains, depicting the frequency of sites bearing the quasispecies identity of the opposing viral read set. By frequency, Asibi nucleotide identity intersects that of 17D-204 to a greater extent than the converse.](https://academic.oup.com/jid/article-abstract/209/3/334/843061)
Recovery of Small Variants

A total of 157 mutations were recovered by this criterion, after exclusion of indels and stop codons. One hundred nine mutations were noncoding; of these, 53 represent 4-fold degenerate sites, while 48 represent amino acid substitutions (Figure 3). Of non-synonymous substitutions recovered, 2 were observed in prM, 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 (Figure 4). Significantly represented frequency changes (P < .5) were clustered in specific regions of the viral genome (Figure 5A and 5B). Mutations recovered from the paired test model were compared both to the consensus (50% 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 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 et al (79.8% recovery) [13].

DISCUSSION

The wild-type 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, and very limited evidence was found for the existence of the wild-type nucleotide identity within the vaccine population. Production of 17D-204 occurs with some variability because of 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. Since analysis of viral quasispecies structure is expected to inform predictive models of pathogenicity and attenuation [3], 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 2A and 2B). As expected, these methods demonstrate a restricted pattern of quasispecies variability for both genomes but especially for the 17D-204 vaccine. This admits the possibility that a small number of large entropy values is present in an otherwise predictably distributed set, an expectation frequently used in construction of phylogenetic models of sequence divergence that accommodate among-site differences in mutation rate [30]. This property of the sequencing data set is consistent with a hypothesis that site variants with the greatest mutation rates are localized to a small proportion of tolerated sites.

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 5A and 5B). 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 [15, 31]. The NS2A gene has been implicated as a factor in the assembly of flaviviruses; mutational analysis of Kunjin virus showed influence of the protein over particle egress and intracellular membrane ultrastructure development [32]. NS3

Figure 5. Paired-test model recovers patterns of local plasticity in the yellow fever virus (YFV) genome. 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 value), following Bonferroni correction. B, 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). C, Venn diagrams showing intersection of recovered variant populations with other models of YFV attenuation; this is separately performed for consensus sequences of parent and derivative strains for HeLa p6 virus and the French neurotropic vaccine [27, 28]. Abbreviations: FNV, French neurotropic vaccine; SNP, single-nucleotide polymorphism.
encodes serine protease, helicase, and NTP triphosphatase activities [33, 34]. The NS4B gene of flaviviruses (peak at approximately 7171) has been implicated as an interferon antagonist in vitro following infection by dengue viruses, West Nile virus, and YFV [35]. NS5 encodes methyltransferase and RNA-dependent RNA polymerase functions [36, 37].

It is especially significant that variants in the vaccine overlap the virulent Asibi genotype at low or undetectable frequencies. The relationship is unidirectional, meaning that the vaccine does not contain small populations of wild-type variants at the coverage depths achieved. It is not possible to assess the linkage of variants beyond read length distance (50 b). 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 we cannot exclude mutation of the vaccine virus in vaccinees. However, these data are supportive of the current opinion that serious adverse events giving rise to a wild-type viscerotropic disease phenotype are a consequence of host effects rather than selection pressures exerted by the vaccine virus. We cannot exclude mutation of the vaccine virus in vaccinees.

The relationship is unidirectional, meaning that the vaccine does not contain small populations of wild-type variants at the coverage depths achieved. It is not possible to assess the linkage of variants beyond read length distance (50 b). 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 we cannot exclude mutation of the vaccine virus in vaccinees. However, these data are supportive of the current opinion that serious adverse events giving rise to a wild-type viscerotropic disease phenotype are a consequence of host effects rather than of specific virulent properties of the vaccine [38]. This also supports an interpretation that the genomic stability of YF Vax at the point of use is stably maintained by the seed-lot system.

Intersection of the recovered low-frequency single-nucleotide polymorphism populations with other models of YFV attenuation was observed (Figure 5C). For the Asibi alignment, 12 sites share variant identity with 2 other models of YFV attenuation. The first of these models, the French viscerotropic virus (FVV), was isolated concurrently with Asibi and then passed 128 times in mouse brain to yield the French neurotropic vaccine (FNV), which was distributed in francophone Africa until high rates of post-vaccination encephalitis in recipients <14 years of age prompted discontinuation of FNV use in 1970. Additionally, passage of the Asibi strain in HeLa cells was performed to produce a virus with phenotypic characteristics similar to those of 17D, namely, reduced virulence in both mice and cynomolgus macaques and loss of vector competence in Aedes aegypti [28, 39].

Sequences of these attenuation models were compared to the reconstructed variants, recovering some similarities (Table 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 [28], plus 2 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) [27]. Four silent variants intersect with the FNV strain. For the 17D-204 alignment, 1 silent variant C10367U (1.58% U) was observed in the 3' untranslated region, intersecting the consensus identities of both FNV(U) and Asibi(U).

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 other factors. It is possible that some variants represent tolerable plasticity in the vaccine genotype, composing a multivariate sequence space that is flexibly occupied by the virus [4]. Little is known about the adaptive context of quasispecies distributions in arthropod-borne viruses, but selection for phenotypes that productively infect both arthropod and vertebrate hosts almost certainly propels this feature. Arthropod-borne viruses encounter population stresses during wild-type transmission cycles, which are relieved by single-tissue, in vitro passage at the cost of multihost fitness [40, 41]. 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 that both may be parsed in vivo [42]. 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 [40, 43]. 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.

Supplementary Data

Supplementary materials are available at The Journal of Infectious Diseases online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyrighted. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgment. We thank the World Reference Center for Emerging Viruses and Arboviruses (UTMB, Galveston), for their kind provision of low-passage YFV Asibi virus.

Financial support. This work was supported by the National Institutes of Health (grant RO1 AI081866 to K. D. R., contract HHSN272201000040I/ HHSN27220004/D04 to R. B. T., and training support for A. S. B. from the Biodefense Training Program at the University of Texas Medical Branch, under T32 AI060549).

Potential conflicts of interest. All authors: No reported conflicts.
All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References