Comparison of Genotypes I and III in Japanese Encephalitis Virus Reveals Distinct Differences in Their Genetic and Host Diversity

Na Han, James Adams, Ping Chen, Zhen-yang Guo, Xiang-fu Zhong, Wei Fang, Na Li, Lei Wen, Xiao-yan Tao, Zhi-ming Yuan, Simon Rayner

Bioinformatics Group, Key Laboratory of Agricultural and Environmental Microbiology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei, China;
State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

ABSTRACT

Japanese encephalitis (JE) is an arthropod-borne disease associated with the majority of viral encephalitis cases in the Asia-Pacific region. The causative agent, Japanese encephalitis virus (JEV), has been phylogenetically divided into five genotypes. Recent surveillance data indicate that genotype I (GI) is gradually replacing genotype III (GIII) as the dominant genotype. To investigate the mechanism behind the genotype shift and the potential consequences in terms of vaccine efficacy, human cases, and virus dissemination, we collected (i) all full-length and partial JEV molecular sequences and (ii) associated genotype and host information comprising a data set of 873 sequences. We then examined differences between the two genotypes at the genetic and epidemiological level by investigating amino acid mutations, positive selection, and host range.

We found that although GI is dominant, it has fewer sites predicted to be under positive selection, a narrower host range, and significantly fewer human isolates. For the E protein, the sites under positive selection define a haplotype set for each genotype that shows striking differences in their composition and diversity, with GIII showing significantly more variety than GI. Our results suggest that GI has displaced GIII by achieving a replication cycle that is more efficient but is also more restricted in its host range.

IMPORTANCE

Japanese encephalitis is an arthropod-borne disease associated with the majority of viral encephalitis cases in the Asia-Pacific region. The causative agent, Japanese encephalitis virus (JEV), has been divided into five genotypes based on sequence similarity. Recent data indicate that genotype I (GI) is gradually replacing genotype III (GIII) as the dominant genotype. Understanding the reasons behind this shift and the potential consequences in terms of vaccine efficacy, human cases, and virus dissemination is important for controlling the spread of the virus and reducing human fatalities. We collected all available full-length and partial JEV molecular sequences and associated genotype and host information. We then examined differences between the two genotypes at the genetic and epidemiological levels by investigating amino acid mutations, positive selection, and host range. Our results suggest that GI has displaced GIII by achieving a replication cycle that is more efficient but more restricted in host range.
with most isolates classified as genotype I (GI) or genotype III (GIII) (15). GIII was previously dominant, but surveillance data from the last 20 years, in the form of nucleotide sequences from collected isolates, have revealed a gradual replacement of GIII by GI such that the latter is now dominant or cocirculating with GIII in many JEV regions of endemicity and epidemicity (16–19). In general, lineage replacement is a cause for concern as it can impact disease control measures, resulting in the eruption of new outbreaks (20, 21). Thus, understanding the reasons behind the current genotype replacement event can yield valuable insight into the future control and eradication of JE diseases.

In an earlier study, we observed that the displacement of GIII was accompanied by a corresponding order-of-magnitude increase in the relative genetic diversity of GI versus that of GIII (22). However, compared to many other viruses exhibiting multiple genotypes, there is relatively high conservation at the amino acid level within and between the JE genotype (with a nucleotide divergence of 11.6% and amino acid divergence of 2.8% between GI and GIII).

Previous epidemiology studies on JEV commonly focused on the E protein or on small numbers of genome sequences and, as such, did not permit a comprehensive comparative analysis of different lineages. However, a recent study of E protein gene sequences revealed that genotype I could be subdivided into two clades, GI-a and GI-b, with the majority of isolates placed in GI-b. Furthermore, viral multiplication and temperature sensitivity analyses revealed that the GI-b isolate had significantly higher infectivity titers in mosquito cells from 24 to 48 h postinfection, providing new insight into how GI emerged as the dominant lineage (23). Nevertheless, the recent publication of a large number of whole-genome sequences of JEV and the availability of a very large number of sequences from other genes make further examination of the GII-to-GI displacement event worthwhile. In this study, we collected all available sequence data from all genomic regions and investigated the differences between the two genotypes not only from a genetic but also from an epidemiological perspective. Specifically, we investigated amino acid mutations, positive selection, and host range and considered the impact on the viral transmission cycle and ultimately on susceptible human populations in an attempt to identify determinants driving the observed lineage replacement.

### MATERIALS AND METHODS

**Phylogenetic analysis of JEV genome and E protein data sets.** Phylogenetic analyses were performed on a total of 121 full-length genome sequences and 570 complete E protein sequences (315 GI and 255 GIII) retrieved from the GenBank database (http://www.ncbi.nlm.nih.gov/). Vaccine or derivative strains, in vitro-cultured isolates, and isolates with sequences containing significant numbers of ambiguous nonstandard nucleotide or amino acid characters were excluded. The genome sequences were isolated from a variety of hosts (mosquitoes, mides, bats, pigs, horses, and humans) with isolation dates ranging from 1949 to 2009 and a geographic distribution spanning China, Japan, South Korea, Thailand, Viet Nam, India, and Taiwan. Although the earliest GIII strain (the Nakayama strain) was isolated in Japan in 1935, this sample was removed from the genome data set as it contains several ambiguous nucleotides. The E protein sequences from the genome data set were combined with the full-length E data set, and duplicate entries were removed. Only GI and GIII E sequences with minimal associated background information (genotype, year of isolation, and country of isolation) were retained to achieve a final data set of 347 GI and 275 GIII isolates. Background details of the genome and E sequences are available from us on request. Nucleotide and deduced amino acid sequences were aligned using ClustalW v2.0 (24) and manually edited using MEGA.

**Phylogenetic analysis was conducted for the 10,299-nucleotide (nt) complete coding region and for the individual GI and GIII E protein gene data sets using the neighbor-joining (NJ) method and the maximum composite likelihood nucleotide substitution model implemented in MEGA v.5.05 (25). The robustness of phylogenograms was evaluated using 1,000 bootstrap replicates.**

**Mutation analysis and investigation of selection pressures.** To identify site mutations that characterized the differences between GI and GIII at the amino acid level, we generated a consensus alignment for the GIII sequences and compared the complete polypeptide sequences in order to identify sites with mutational differences between the two genotypes that were present in at least 95% of the GIII isolates. The presence of positive selection was assessed using the Codeml program implemented in PAML software package version 4.0 (http://abacus.genetics.ucl.ac.uk/software/paml .html) (26). We examined six different codon-based substitution models to estimate the ratio of nonsynonymous and synonymous nucleotide substitutions ($\omega = dN/dS$) at each codon site: M0 (ratio of 1), M1a (nearly neutral), M2a (positive selection), M3 (discrete $\omega$), M7 ($\omega$ values approximated by a $\beta$ distribution), and M8 ($\omega$ values approximated by $\beta$ distribution and a fraction of sites with $\omega = 1$). As appropriate to each model, a maximum-likelihood approach was used to estimate $\omega$ and the proportion of sites in the alignment that fall into a specific predefined category. A Bayes empirical Bayes (BEB) approach was used to calculate the posterior probabilities that each codon belongs to a particular site class and for identifying sites under positive selection (27). Then, using the likelihood-ratio test (LRT), three independent pairwise comparisons (of M0 and M3, of M1a and M2a, and of M7 and M8) were performed to select the best-fitting models for the data. Amino acid sites on the E protein predicted to be under positive selection in each genotype (together with host, date, and location information) were compared for the GI and GIII sequences in turn to investigate correlations among these parameters. These relationships on the respective E protein gene trees were presented using the online iTOL software package (http://itol.embl.de/).

**Visualization of sites within the JEV NS5 protein structure under positive selection.** Many GIII-to-GI mutations and sites under positive selection were located in the NS5 coding regions. In an attempt to investigate the significance of these mutations, the structure of the NS5 protein recently solved for Japanese encephalitis virus (4K6M) was downloaded from the protein database (http://www.pdb.org/pdb/home/home.do), and sites of interest were located on the structure and visualized using PyMol v1.5 (http://www.pymol.org/).

**Investigation of host range and genotype shift.** To investigate the host range and variations in JEV GI and GIII isolates over time, a third data set (distinct from the genome and E protein data sets described above) was created. This data set comprised the full E sequences collected for the phylogenetic analysis combined with additional partial E, partial C, C-prM, and NS5 gene sequences (data available on request) that were also downloaded from GenBank. Derivative and genetically modified strains, as well as duplicate entries, were removed, and NJ trees were constructed to confirm the genotype of each sequence. GI and GIII sequences with minimal background information (genotype, year of isolation, and country of isolation) were selected to achieve a final data set of 893 isolates comprising 622 complete E protein gene sequences, 41 partial E and NS5 gene sequences, and 230 isolates for the partial C and C-prM sequences. The combined E/partial E/C-prM/NS5 data set was then used to classify the presence of GI and GIII isolates by year and country. Few isolates were obtained from horses, bats, mides, and birds, so only mosquito, pig, and human isolates were considered in the analysis. A 2-by-3 contingency table was generated based on the two genotypes and three hosts, and Pearson’s chi-square test ($\chi^2$) implemented in the R statistical package (http://www.r-project.org) was performed to investigate the relationship

---

**Harvested from V1027 2017 by guest [http://jvi.asm.org] Downloaded from**
between host distribution and genotype. Significance was assigned at $P < 0.01$.

**RESULTS**

**Genome data set and phylogenetic analysis.** A summary of the GI and GIII genome sequences is available on request. The data corresponding to the NJ tree for the 121 whole-genome sequences for all JEV genotypes, sampled between 1949 and 2009, are available on request. The topology of the tree is consistent with previous reports (22) and is divided into five major groups corresponding to genotypes I, II, III, IV, and V. The majority of isolates belong to GI (54 sequences) and GII (60 sequences). There are three and two subclades predicted for GI and GII, respectively. For GII, these subclades have strong bootstrap support; for GI, the bootstrap support is much weaker, but the strains in the two clades can be distinguished by the amino acid at site 2296 in the polyprotein (within NS4B), which is predicted to be under positive selection in both genotypes (see below). Schuh et al. (28) predicted the presence of two distinct G-I clades, G-Ia and G-Ib, on the basis of E protein data, but none of the G-Ia samples have full-length genome sequences available, so the tree represents only G-Ib isolates. Of the remaining genotypes, GV includes just two isolates; one was isolated in Malaysia in 1952 (29), and the other was collected recently (in 2009) in the Tibet autonomous region (30). There is a further report of an isolate from South Korea collected in 2010 (31), but no genome sequence is available. The GII isolates are from Australia and the south Pacific, and the single GIV isolate is from Indonesia. All of these GII and GIV isolates represent earlier strains collected 17 years ago. There are no recent isolates for either of these genotypes.

**Mutations with amino acid properties.** The data with respect to the consensus amino acid changes in GIII versus GI are available on request. In terms of mutation density (i.e., number of mutations/number of sites for each coding region), the C protein is the most variable, with 5 changes across the 127 amino acids (aa) encoding the protein, followed by NS1, NS2A, and NS2B, with 1 mutation for every 40 amino acids, and NS3, NS4B, and NS5, with 1 mutation for every 80 amino acids. The most highly conserved proteins are E and NS4A.

**Natural selection and adaptive evolution.** To investigate the extent of the selective pressure acting on the JEV genome, we estimated the ratio of nonsynonymous to synonymous substitutions across the entire polyprotein using six different codon substitution models implemented in the PAML software package to represent a range of possible scenarios. Details of the parameter estimates for each model are available on request. Table 1 shows the likelihood-ratio test (LRT) results for pairwise comparison of specific model pairs. The comparison of models M0 and M3 is a test of variable $\omega$ among sites, and the significant differences (supported by the $\chi^2$ test results) for both GI and GIII suggest that, as expected, both genotypes experience significant variation throughout their genomes. Comparison of models M1a and M2a is a test of neutral versus positive selection. Whereas there is a clear distinction between these two models for GIII, surprisingly, there is no significant difference estimated for GI ($P < 0.05$). A significant difference is achieved only in comparisons of the more complicated M7 and M8 models, which approximate site variation among sites, and the significant differences (sup-
The unrooted neighbor-joining E protein trees for GIII (a) and GI (b) were constructed from the combined E/genome data set alignments of 275 GIII and 347 GI isolates, respectively. Taxons are colored by haplotype as defined by amino acid sites 417, 503, 521, and 702 within the E protein that are predicted to be under positive selection in genotype III. The inner circle shows the hosts of isolates; the outer circle shows the isolation date of the isolates.
rate in mouse neuroblastoma cells and pathogenicity in mice (33), suggesting a similar association with pathogenicity in GI. For site 521 (aa 227 of the E protein), it has been shown that Ser-227-Pro mutants can slightly retard the growth rate in Vero cells (33).

There is growing evidence that indicates that NS4B plays a role in blocking interferon (IFN) signaling in West Nile virus (WNV) and dengue virus (DENV) infections (34). For site 24 in the NS4B protein, all the GIII isolates had a serine residue, with the exception of four human strains with a proline residue. In GI, however, there were approximately equal proportions of isolates with serine and proline residues. A previous study suggested that this site might be correlated with host preference in GI and that a Pro24 residue tends to be associated with human cases (35). However, our data set shows no significant association with hosts, with the proline residue occurring in mosquito, pig, and human isolates. In GI, these proline and serine residues clustered into separate clades in the estimated genome tree (data available on request), suggesting that the proline substitution occurred in a common ancestor and was inherited by present-day virus strains. In GIII, however, the four isolates with a Pro residue were dispersed throughout the three different clades. No clear pattern or functional significance was identified for the remaining GI-specific sites at codons 2525 (NS4B), 2956 (NS5), and 3416 (the C-terminal end of the NS5 RNA-dependent RNA polymerase [RdRp]).

**Host adaptation.** For the 11 sites predicted to be under positive selection in GIII, all of the bat samples (isolated from Yunnan province in China) have 100% identity, but no similar patterns exist for other hosts. For the sites located within the E protein (sites 417, 503, 521, and 702 in the polyprotein and sites 123, 209, 227, and 408 within the E protein), there are 12 haplotypes, with the SKSS and SKPS haplotypes predominant. Figure 1a shows the haplotypes mapped onto the estimated GIII E protein gene tree. Human and mosquito isolates show the greatest diversity, with 10 and 9 haplotypes, respectively (Table 3). Only three haplotypes, SKSS, NKSS, and SKSL, are present in porcine samples, suggesting a possible significance for sites 209K and 227S in this host. The other interesting feature in the tree is that RKSL appears to be an emerging haplotype on the east coast of China, as these isolates form a subclade with strong bootstrap support and the subclade is comprised of recent samples from mosquitoes and pigs. (Fig. 1a).

Once again, the corresponding GI results are strikingly different. The GI isolates are defined by only five haplotypes, with SKSS and NKSS accounting for 99% of the isolates and only single occurrences of the remaining three haplotypes. Figure 1b shows the haplotypes mapped on to the GI E tree. The tree shows a classification of G-I isolates into G-Ia and G-Ib consistent with the findings of Schuh et al. (28). The majority of samples with NKSS sequences were recently (i.e., since 2003) collected from mosquitoes, with three earlier samples isolated from pigs. Thus, the GI isolates are defined by a range of haplotypes that is far narrower than that seen with the GIII isolates, although there is no distinct haplotype for the smaller G-Ia clade.

**Visualization of the sites under selection in the NS5 protein structure.** Given the number of GI-to-GIII mutations and the number of sites under positive selection that are located within the RdRp catalytic domain of the NS5 protein, we next attempted to investigate their locations within the folded protein structure for this domain. The NS5 protein comprises a methyltransferase (MTase) at the N-terminal and an RdRp at the C-terminal end. The recently solved crystal structure of the full-length JEV NS5 protein (36) was downloaded from the Protein Database (PDB code 4K6M), and a schematic representation is shown in Fig. 2, with the GI-to-GIII mutations and sites under positive selection marked.

The JEV NS5 architecture is characteristic of the recognized flavivirus polymerase structure, with a right-hand conformation consisting of fingers, palm, and thumb (37), and shows significant

---

**TABLE 3** GI and GIII haplotypes for E protein by host*

<table>
<thead>
<tr>
<th>Host</th>
<th>GI result</th>
<th>GIII result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mosquito</td>
<td>SKSS 79</td>
<td>SKSS 185</td>
</tr>
<tr>
<td></td>
<td>SKPS 25</td>
<td>NKSS 50</td>
</tr>
<tr>
<td></td>
<td>RKSL 10</td>
<td>KKSS 1</td>
</tr>
<tr>
<td></td>
<td>SRSS 5</td>
<td>NESS 1</td>
</tr>
<tr>
<td></td>
<td>SRPL 3</td>
<td>SKPS 1</td>
</tr>
<tr>
<td></td>
<td>NKSS 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RKSS 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RRPS 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RKPS 1</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>SKSS 45</td>
<td>SKSS 8</td>
</tr>
<tr>
<td></td>
<td>SKPS 38</td>
<td>NKSS 3</td>
</tr>
<tr>
<td></td>
<td>SRSS 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RKPS 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SRPS 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RKSL 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RRSS 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SRPS 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RKPS 1</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>SKSS 34</td>
<td>SKSS 91</td>
</tr>
<tr>
<td></td>
<td>NKSS 3</td>
<td>NKSS 5</td>
</tr>
<tr>
<td></td>
<td>RKSS 4</td>
<td></td>
</tr>
<tr>
<td>Midge</td>
<td>SKPS 2</td>
<td>SKSS 1</td>
</tr>
<tr>
<td>Horse</td>
<td>SKPS 1</td>
<td>SKSS 1</td>
</tr>
<tr>
<td>Bat</td>
<td>RKPS 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RKPS 4</td>
</tr>
</tbody>
</table>

* Haplotypes are defined by amino acid sites 417, 503, 521, and 702 within the E protein that are predicted to show positive selection in GIII.
The majority of GI-to-GIII mutations are located in the region coding for RdRp, but there is a single mutation in the MTase region (E22D), a site that has been shown to affect replication efficiency in dengue virus 4 (DV4) (38). Of the remaining mutations, there are two in the N-terminal extension (K280R and K287R) and one each in the palm (A372V), middle finger (E588G), and thumb (V878I) regions. It is interesting that the four remaining mutations correspond to charge changes in close proximity within the NS5 structure, located in the fingertip region of the finger subdomain (37). Two of the mutations, at sites 429 and 432, are charge complementary, producing charge changes of $\frac{1}{2}$ and $\frac{1}{2}$, respectively. These two sites are in close proximity to motif F (aa 453 to 476) in JEV, which is thought to provide a binding site for incoming nucleoside triphosphates (NTPs) entering through the tunnel from the back of RdRp molecule (39). Two Zn$^{2+}$ binding sites have been identified in the RdRp structure, one of which is coordinated by sites E440, H444, C449, and C452 in JEV. Site D438N (associated with a charge change of $+1$ in G) is in the proximity of this zinc binding site and has been shown to be related to replication efficiency in DV2 (38).

**Host preference of genotypes.** The combined E/partial E/C/prM/NS5 data set comprising 454 GI strains and 439 GIII isolates was used to investigate the host preference of the two major genotypes. Hosts comprised mosquitoes, midges, humans, and pigs, with a small number of isolates collected from horse, bats, and birds. The majority of strains were isolated from mosquitoes, with 278 and 197 GI and GIII isolates, respectively. **Figure 3** shows the variations in the numbers of GI and GIII isolates collected over time. **Figure 3a** shows the variations for all isolates and highlights how GIII began to be displaced by GI as the dominant genotype in the middle of the 1990s. The variations for mosquitoes (Fig. 3b) and pigs (Fig. 3c) reflect this displacement pattern. However, Fig. 3d shows that the majority of human cases remain associated with GIII. Furthermore, it seems that, overall, GI has a much narrower host preference (Table 4). To determine whether these observed differences are significant, we constructed a contingency table to analyze the relationship among hosts and genotypes. Statistical analysis revealed a highly significant difference (Pearson’s chi-square test, $P$ value = 2.2 $\times 10^{-16}$); i.e., the host composition of GI is significantly different from the host composition of GIII. Panels e and f of Fig. 3 show the geographical composition of the samples collected between 1990 to 2011, the time range over which the genotype shift primarily occurred. China, Taiwan, Japan, South Korea, and Thailand are all well represented across the time interval, suggesting that there is no evidence of sampling bias for any specific country. The full temporal-spatial details of the isolates are available on request.

The two genotypes also showed notable differences in the composition of vector hosts (Table 5). Of the 279 GI isolates and 199 GIII vector isolates, 218 and 128 isolates were able to be resolved to the group or subgroup level. For both genotypes, the dominant species is *Culex tritaeniorhynchus*, which accounts for 94% ($n = 204$) of isolates in GI but for only 74% ($n = 95$) in GIII. The difference also extends to the genus-group-subgroup range of the two genotypes. For GI, the remaining 6% of isolates ($n = 13$) span 9 genera, groups, or
FIG 3 Change in GI versus GIII host shift over time. Data represent variations over time in the number of collected JEV isolates for GI and GIII for all samples and for sample subsets. The red solid line represents GIII, and the blue dashed line represents GI. (a) Variations for all isolates, highlighting how GIII was displaced by GI as the dominant genotype in the middle of the 1990s. (b and c) Variations over time for mosquitoes and pigs, respectively, demonstrating similar displacement patterns. (d) Variations over time for human isolates. Despite GIII being displaced as the dominant genotype, the majority of human cases remain associated with GIII. (e and f) Geographical composition of samples collected over the time period (1990 to 2011) during which genotype replacement appears to have occurred. Panel e represents the sum over the entire period; panel f shows the variation by year. There is no region in which samples appear to have been preferentially sampled, and many regions have been continually represented over the time course.
subgroups; for GIII, the remaining 24% of isolates (n = 31) span 14 different genera, groups, or subgroups. However, the classification of subgroups within Culex families is based on morphological characteristics and is somewhat subjective (40, 41), so the significance of these differences is unclear.

**DISCUSSION**

In this work, we consolidated all sequence data collected from across the JEV incidence region to try and identify differences between “before” and “after” GIII-to-GI shift events. In this way, we created a comprehensive data set describing the impact on JEV in the Asia-Pacific region over the course of the last 30 years. Previous studies have considered the phylogeographic events associated with this genotype shift and compared the differences in genetic diversity by investigating a single protein. Here, by collecting and consolidating all publicly available samples and their associated background data, we have also obtained a perspective on the differences that exist between these two genotypes among the primary hosts and reservoirs in the JEV replication cycle.

Previous studies have shown that the evolution of JEV is driven predominantly by strong purifying selection, and few sites have been predicted to be under positive selection. However, those analyses were conducted on individual genes (E or prM-C) or on a small number of genome sequences (42–44). Here, we reexamined this issue by the use of a much larger data set and, for the first time, investigated the two genotypes separately. Overall, our results are consistent with those from previous studies with respect to predicting a consistent phylogeny, estimating low ratios of non-synonymous substitutions to synonymous substitutions (dN/dS) in both genotypes, and identifying a small number of individual sites under positive selection. However, the differences in the specific predictions for the individual genotypes are surprising. Typically, an emerging strain is considered to have gained some adaptive advantage over other strains with respect to host or environment and this is reflected in the identification of positive selection within its viral genome (45). Based on this, we would have expected to identify stronger effects in GI as the emerging lineage. However, while 11 sites were identified as evolving under diversifying selective pressure in GIII, only 5 sites were identified in GI. While similar differences in the numbers of sites for GI and GIII in the predictions for the E protein were reported in a recent publication by Schuh et al. (23), there are discrepancies between these two studies in the identified sites. However, similar inconsistencies were also observed between their predictions based on the fast, unconstrained Bayesian approximation (FUBAR) and fixed-effects-likelihood (FEL) methods. These differences are likely a consequence of the highly conserved nature of the JEV genome (compared to, for example, that of HIV) and make it difficult to apply complex site variation models to predict sites under positive selection. What is significant among all these results is that there are consistently fewer sites predicted for GI than for GIII. Similarly, while GIII was predicted to show positive rather than neutral selection, there was insufficient support to allow making this distinction for GI.

Analysis of full-length genome sequences revealed that the differences between the two lineages can be represented by only 30 amino changes within the 3,432 amino acids comprising the polyprotein, corresponding to a difference of less than 1%. While the significance of the set of identified GI-to-GIII site mutations and sites predicted to be under positive selection needs to be established, the high similarity of the two lineages means that this set represents a relatively small number of sites, and it seems that they would be worthwhile candidates for experimental studies, as many of them may be associated with host adaptation or play relevant functional roles in the virus life cycle.

Further differences were revealed when we considered the host composition of the more comprehensive set of E/prM-NS5 sequences. The change in the ratio of GI isolates to GIII isolates from pig and mosquito is consistent with the gradual emergence of GI, but, unexpectedly, the majority of human isolates are still GIII. Moreover, these host differences have been shown to be statistically significant with very strong support. The
differences in host composition also extended to the vector host, with the GIII mosquito isolates collected from a wider range of mosquitoes and GI isolates almost exclusively restricted to Culex tritaeniorhynchus.

The highly conserved nature of the JEV genome is reflected in the predicted trees for the genome and the respective E protein gene trees for GI and GIII. Whereas the genome tree predicted two and three subclades, respectively, for GI and GIII, these classifications were not reflected in the tree generated from the shorter E protein gene sequences. However, mapping of the sites predicted to be under positive selection in this protein showed strong clustering by haplotype and, once again, showed distinct differences that highlighted the greater sequence variation within the GIII isolates.

Given the similarities of the two genotypes at the nucleotide and amino acid levels, it is remarkable that GI has displaced GIII so effectively in so many regions. This is in stark contrast to, for example, DENV, which exists as multiple genotypes within each of the four serotypes and exhibits far more complex genotype and serotype emergence patterns (for examples, see references 46 and 47). However, notable distinctions exist between DENV and JEV in terms of their respective host ranges and viral life cycles. The DENV transmission cycle involves humans acting as reservoirs and mosquitoes serving as vectors (48). The predominant vector is Aedes aegypti, with secondary contributions from Aedes albopictus and sporadic reports of outbreaks associated with Aedes hensilli, although only sporadic cases of dengue hemorrhagic fever (DHF) have been associated with Aedes albopictus outbreaks (49, 50). In contrast, in the enzootic JEV cycle, a number of mosquito species and other arthropod insects act as vectors, swine serve as an amplifying host (51), and wading ardeid water birds (e.g., herons and egrets) act as virus reservoirs. While each of these hosts play an important role in the maintenance of JEV in nature (13), humans are considered a dead-end host, developing insufficient viremia to infect feeding mosquitoes (52). Of all the vector hosts, Culex tritaeniorhynchus has been demonstrated to be the most efficient and prevalent (13). However, other Culex species (e.g., Culex annulirostris, Culex vishnui, Culex bitaeniorhynchus, and Culex pipiens) (52–54) and other mosquitoes of different genera (e.g., Armigeres, Aedes, Anopheles, and Mansonia) have demonstrated secondary roles and have been identified as potential viral transmitters (55, 56).

The contrasting transmission cycles of these two viruses suggest a possible explanation for the differences in our GI and GIII results and the recent displacement of GIII. Our results indicate that GIII shows greater diversity than GI in both sequence variation and host range. In particular, the narrow haplotype range of the GI E protein is associated with a narrower vector range, suggesting that the GI strain has been optimized for transmission to and from Culex tritaeniorhynchus and pigs. This hypothesis is supported by the results from an experimental study by Schuh et al. (23), which showed that the GI-b clade demonstrates an increased in vitro multiplicative ability compared to other sublineages, possibly resulting in increased numbers of mosquito→avian→transmission cycles. Our results support these findings and also suggest a more efficient transmission pattern for the virus (from vector to amplifying host and back to vector). This was achieved by narrowing the host range of the GI-b lineage with the trade-off of reduced numbers of infected human hosts (evidenced by reduced numbers of human cases after the genotype switch). However, since humans are a dead-end host, this does not impact the infection efficiency of the virus.

While our results imply a reduced human risk in regions where GI is dominant, JEV will continue to present a serious threat to public health. The close proximity of pig breeding farms or feral pig populations to human residential areas increases the risks of exposure of human populations to JEV, resulting in significant levels of spillover to humans (57–59). This link was demonstrated in Taiwan by an order-of-magnitude drop in the incident rate in nonvaccinated infants between 1966 and 1980 that coincided with the urbanization of many rural areas (60). In many regions of endemicity, there has been a switch to periurban agricultural methods in recent years in response to growing requirements for pork production that has led to increases in populations of pigs which, in their role as amplifying hosts, may have drastically altered the dynamics of the transmission mode (59, 61, 62). Given the variations in JEV GIII vectors in terms of resting locations (63) and feeding preferences (64–66) as well as seasonal variations in populations (67, 68), the virus would be well positioned to adapt to changes in the infection landscape.

However, there have been two recent reports of human outbreaks associated with GI. South Korea reported JE fatalities (7 of 26 cases) in 2010, and recently published sequences of strains isolated from mosquitoes in this outbreak confirmed that the cases were associated exclusively with GI isolates (69), but no human sequences have been published. Another report from an outbreak in India showed greater numbers of human cases associated with GI, but GIII cases remained dominant (41 of 68 sequences submitted to GenBank [data available on request]). Only the C-prM sequence was published for the majority of the isolates, and the full-length E sequence is available for only two of the GI strains (70). However, they are placed in a distinct clade in our estimated E protein gene tree and contain distinct amino acid mutations that are not present in any other GI isolates, suggesting the possible emergence of another lineage.

While it is tempting to dismiss these differences as merely a consequence of sampling biases, there is no evidence of any striking geographical bias in the samples; many geographically distinct regions, including Japan, China, India, Taiwan, South Korea, and Thailand, have been well represented over the period in which the genotype shift has occurred (Fig. 3e and f), and there is strong statistical support for all these results. Importantly, these are the same data that are used as evidence for genotype replacement. Nevertheless, this highlights not only the need for more surveillance data but also the benefits of some kind of central repository for this information. Having access to a comprehensive data source describing the impact of JEV as well as details of environment and effective control measures could permit more insightful analyses that extend beyond standard phylogenetic investigations. This has already been demonstrated in two recent reports, which investigated the link between environment and the incidence of JEV infection in Nepal (71) and estimated the distribution of JEV in Asia (72) by considering environment and vector distribution based on the integration and analysis of data obtained from WorldClim (http://www.worldclim.org/bioclim) and Mosquito-Map (73). Complementing current vaccination and mosquito control programs with the use of novel analytical techniques applied to comprehensive data collections can lead to a more effective approach for the continued control and eradication of JEV.
REFERENCES

   10.2471/BLT.10.085233. 
   .doi.org/10.1007/978-3-642-59043-8_6. 
   factors. Zootropism and vertical flight of Culex tritaeniorhynchus with  
   emergence and establishment of a newly dominant genotype of dengue virus serotype 2 in Viet Nam: in vivo fitness advantage and  
   /jvi.0.82185-0. 
   Henn MR, Simmons CP. 2010. Emergence of the Asian 1 genotype of dengue virus serotype 2 in Viet Nam: in vivo fitness advantage and  
19. Nga PT, del Carmen Parquet M, Cuong VD, Ma SP, Hasebe F, Inoue  
   encephalitis virus (JEV) genotype circulating in northern Vietnam: implications for frequent introductions of JEV from Southeast Asia to East Asia.  


