Infection, Genetics and Evolution 9 (2009) 709–715

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Evolution and dispersal of St. Louis encephalitis virus in the Americas

Albert J. Auguste [a], Oliver G. Pybus [b], Christine V.F. Carrington [a,*]

[a] Department of Preclinical Sciences, Faculty of Medical Sciences, University of the West Indies, St. Augustine, Trinidad and Tobago
[b] Department of Zoology, University of Oxford, OX1 3PS, UK

ARTICLE INFO

Article history:
Received 29 March 2008
Received in revised form 15 July 2008
Accepted 16 July 2008
Available online 29 July 2008

Keywords:
St. Louis encephalitis virus
Flavivirus
Molecular evolution
Substitution rate
Relaxed molecular clock
Most recent common ancestor
Gene flow
Phylogeography

ABSTRACT

Using a Bayesian coalescent approach on a dataset of 73 envelope gene sequences we estimated substitution rates and dates of divergence for St. Louis encephalitis virus (SLEV) in the Americas. We found significant rate heterogeneity among lineages, such that "relaxed" molecular clock models were much better supported than a strict molecular clock. The mean substitution rate estimated for all SLEV was 4.1×10^{-4} substitutions/site/year (95% HPD 2.5–5.7)—higher than previous estimates that relied on the less well-suited strict clock. Mean substitution rates for individual lineages varied from 3.7×10^{-4} to 7.2×10^{-4} substitutions/site/year. For the first time we also assessed the magnitude and direction of viral gene flow within the Americas. The overall direction of gene flow during the period represented by the phylogeny is from South to North, and the region between 15°N and 30°N latitude appears to be the major source of virus for the rest of North America, which is consistent with migratory birds returning to their northern breeding grounds having acquired infection while wintering in the region of the Gulf of Mexico.

1. Introduction

St. Louis encephalitis virus (SLEV) is a single-strand positive sense RNA virus (family *Flaviviridae*) and the etiological agent of St. Louis encephalitis, a febrile illness in humans that can be fatal when neurological complications arise. The virus is distributed throughout the Americas and the Caribbean islands (Spence et al., 1977; Spence, 1980), but is currently only a major health problem in North America, where small outbreaks have occurred within recent years (Reisen, 2003). In contrast, in Central and South America, human disease is infrequently reported and an outbreak in Córdoba Province of Argentina in 2005 (with 47 laboratory-confirmed cases, 9 of which were fatal) was the first outbreak attributed to this virus outside of North America (Diaz et al., 2006).

Like West Nile virus (WNV), to which it is antigenically closely related, SLEV is transmitted between wild birds by the *Culex* mosquito, and humans serve only as dead-end hosts. Birds (primarily passeriformes and columbiformes) are recognised as the primary vertebrate host in North America, but transmission cycles appear to vary geographically, and in South America isolates have been collected from rodent species such as *Calomys Musculinus*, *Mus Musculus* and *Akodon sp.* and from the common opposum *Didelphis marsupialis* (Mitchell et al., 1980, 1983). A variety of other bird species, including herons, egrets, and cormorants may also act as viral reservoirs in South America (Reisen, 2003; Spence, 1980). However, it is still unclear if any of these mammalian and avian species are amplification or dead-end hosts.

Previous phylogenetic studies demonstrated seven largely geographically defined SLEV lineages (Kramer and Chandler, 2001). Lineages I and II are predominantly North American, while lineages III–VII exist in South and Central America. Such geographic clustering suggests that SLEV is mainly maintained locally, although there is evidence of occasional migration among areas, both within the USA and between the USA and other regions of the Americas (Kramer and Chandler, 2001). This phylogeographical pattern suggests that migrating birds are responsible for the long distance movement both within the USA, and between the USA and the rest of the Americas (Kramer and Chandler, 2001). Here, we use a parsimony approach to determine, for the first time, the magnitude and direction of viral gene flow across the Americas. We also used a Bayesian coalescent approach to estimate the ages of currently circulating SLEV lineages, as well as their rates of evolutionary change. Previous work on RNA viruses has shown that a molecular clock is frequently rejected, and that there is high rate variation among lineages (Jenkins et al., 2002; Drummond et al., 2006; Lemey et al., 2006). Therefore in contrast to a recent study that assumed a strict molecular clock (Baillie et al., 2008), we use

* Corresponding author. Tel.: +1 868 662 1873; fax: +1 868 662 1873.
E-mail address: christine.carrington@sta.uwi.edu (Christine V.F. Carrington).

1567-1348/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2008.07.006

more realistic relaxed molecular clock models that allow for rate variation among branches in the tree (Drummond et al., 2006).

2. Methods

2.1. Data sets

Previously published partial envelope (E) gene sequences for SLEV were downloaded from GenBank, manually aligned using the Se-Al program (http://tree.bio.ed.ac.uk/software/seal/) and then trimmed to a common length of 1503 bp. All sequences were confirmed as non-recombinants using the software package Recombination Detection Program, version 2 (Martin et al., 2005; http://darwin.uvigo.es/rdp/rdp.html) and a data set consisting of 73 taxa representing 16 distinct geographic regions and 72 years was prepared. Taxa used, their Genbank accession numbers, geographic origin and year of isolation are listed in Table S1.

2.2. Phylogenetic analyses

Maximum Likelihood (ML) phylogenies were estimated using the General Time Reversible (GTR + Γ_4 + I) model, which estimates the relative rate of each substitution type (GTR), the proportion of invariant sites (I), and the shape parameter of a gamma among-site rate heterogeneity model with four rate categories (Γ_4). This model was identified as the best-fit model of nucleotide substitution using MODELTEST (Posada and Crandall, 2001). Bootstrapping was subsequently performed to assess the robustness of tree topologies using 1000 replicate neighbor-joining trees under a ML substitution model. All maximum likelihood analyses were performed with PAUP* (Swofford, 2003).

2.3. Bayesian inference of substitution rates and divergence dates

The BEAST (v1.4.7) package (Drummond and Rambaut, 2007) was used to infer demographic histories, rates of nucleotide substitution and times to common ancestry from the SLEV sequences analyzed here. This package uses a Bayesian Markov Chain Monte Carlo method to analyse serially sampled gene sequence data, that is, sequences sampled at evolutionarily distinct points in time (Drummond et al., 2002). Analyses were carried out using the Bayesian Skyline Plot (Drummond et al., 2005), which does not use a pre-specified model of demographic history, and using three different molecular clock models: the strict clock, the uncorrelated exponential (UCED) relaxed clock, and the uncorrelated lognormal (UCLN) relaxed clock. The latter two models allow rate variation among lineages to be estimated (Drummond et al., 2006). A chain length of 50 million states, sampling every 1000 states was used to ensure sufficiently high effective sample sizes (ESS). The program Tracer version 1.4 (http://tree.bio.ed.ac.uk/software/tracer/) was used to confirm convergence for each chain and to perform model comparison via Bayes Factor calculations (Suchard et al., 2001). Statistical uncertainty in the data is illustrated by the 95% highest probability density (HPD) values. These analyses were performed on the complete data set (all SLEV; 73 taxa) and subsets representing specific lineages, then repeated on corresponding data sets from which a Brazilian 1968 isolate sequence (Bra68) was removed (see below for explanation).

2.4. Inference of magnitude and direction of gene flow

The pattern of gene flow within the Americas was inferred using the parsimony method implemented in the MacClade program (Maddison and Maddison, 2000), based on the maximum likelihood tree estimated as described above and rooted with the closely related Rocio virus (GenBank accession no. AY632542). In this analysis, each isolate was assigned a specific "character state" based on its geographic origin (i.e. country, state or latitude; see Table S1). We considered movement amongst (i) individual countries and states within the USA, (ii) between geographic latitudes and (iii) within the region bounded by latitudes 15°N and 30°N (major source population). Given the tree phylogeny and the assigned isolates' states, the minimum number of "state changes" necessary to give rise to the observed distribution of states was estimated. To calculate the expected number of state changes under the null hypothesis of panmixis (no significant clustering of sequences according to location), the isolates' states were randomized 500 times, and for each randomization the number of state changes was calculated as before. Expected changes were then summed across all 500 randomizations and divided by the number of replicates. To account for polytomies, all calculations were performed across 1000 randomly resolved trees. The difference between the mean number of observed and expected changes for each pair of states indicates the magnitude of genetic isolation (difference <0) or migration (difference >0). A *P*-value was then calculated for each category of analysis, which represents the probability that the difference between observed and expected values arose by chance alone.

3. Results

3.1. Phylogenetic relationships of SLEV

The maximum likelihood tree inferred (Fig. 1) contains six well-supported clades (bootstrap values greater than 95%). Five of these (labeled 1–3, 5 and 6) correspond to lineages previously defined as lineages I, II, III, V and VII (Kramer and Chandler, 2001). Four isolates from Panama 1973 and 1977 formed the sixth well-supported group (labeled lineage 4). These four isolates were previously designated as clade IV and distinguished as separate from the Panama 1983 isolate, which was the sole isolate in lineage VI (Kramer and Chandler, 2001). However, we found very low bootstrap support for the node separating the latter from the other Panamanian isolates. Hence, this tree reveals a clear geographic structure in this data, the details of which are considered in more detail below.

3.2. Evolutionary dynamics of SLEV

Estimates for rates of nucleotide substitution and times to common ancestry were derived using the complete data sets and data sets from which the Brazilian sequence from 1968 (Bra68; lineage 2) was omitted (see Table 1). The analyses were performed on data sets with and without Bra68 because this isolate differed from a 1933 isolate from Missouri (MO33) at only one nucleotide position in the region sequenced despite the two isolates being separated by 35 years. Given a Poisson distribution, to have a reasonable probability ($\geq$5%) of observing less than two nucleotide changes in 1503 nucleotides over this time period, the substitution rate would have to be no more than 0.9×10^{-4} nucleotide substitutions/site/year (subs/site/yr). While this rate is plausible for an RNA virus it is at the lower end of the range for flaviviruses (Jenkins et al., 2002) thus raising the possibility of sequencing, labeling or annotation error. Since the 1933 sequence was recently resequenced by Baillie et al. this was assumed to be correct and was retained.

Using the modified data set, the estimated mean substitution rate for the affected lineage 2 was lower (3.7 subs/site/yr vs. 4.7×10^{-4} subs/site/yr) while the rates for the other lineages and the overall rate (4.5×10^{-4} subs/site/yr) were all slightly higher

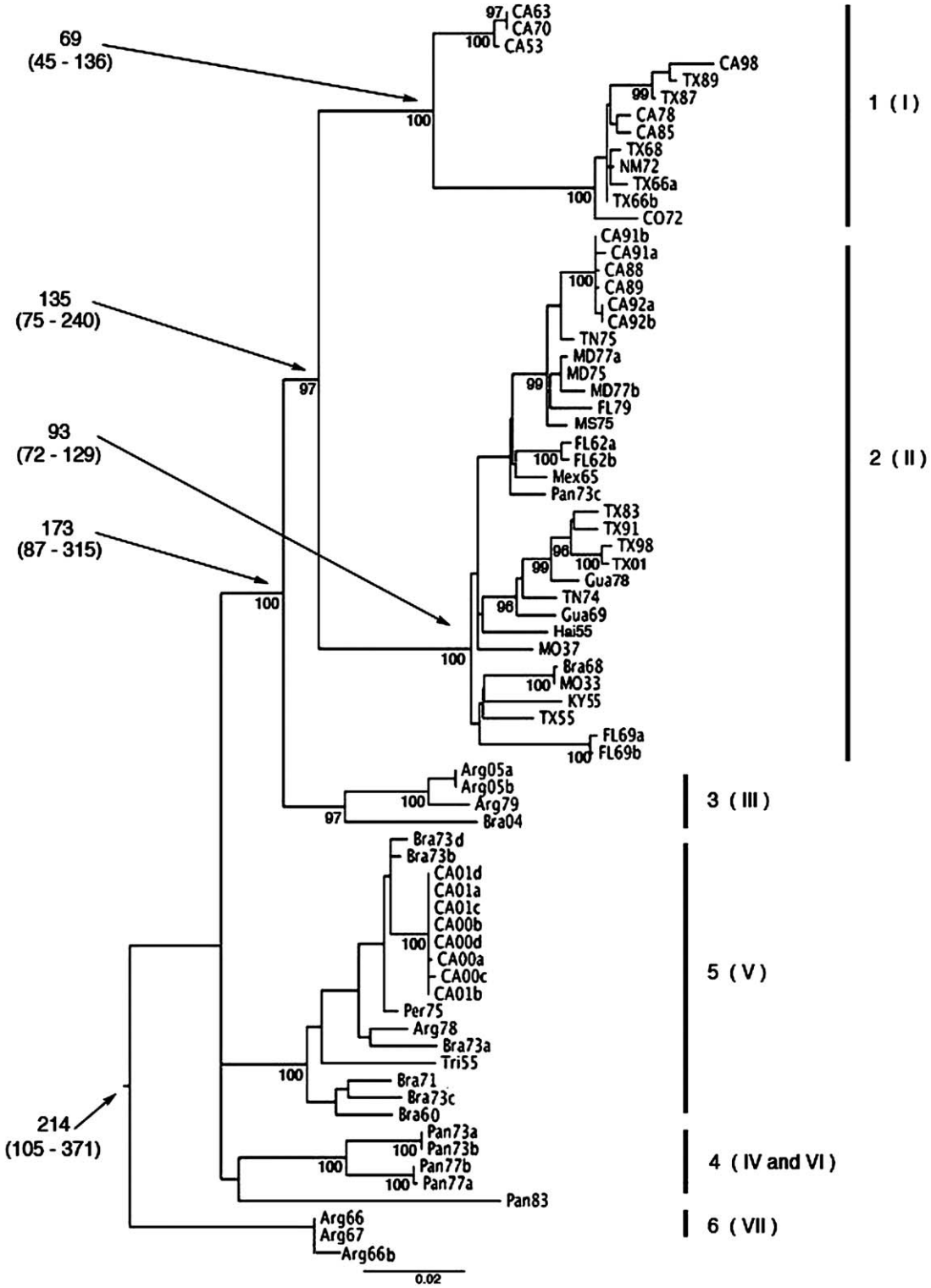

Fig. 1. Maximum likelihood phylogeny based on 1503bp fragment of the E gene of SLEV. Country of origin and dates of isolation are included in the taxon names. Accession numbers are listed in Table S1. The ages of nodes with 95% HPDs are indicated at relevant nodes. Bootstrap values above 95% are also indicated and the associated clades have been numbered 1–6 with the names for these lineages as previously published by Kramer and Chandler (2001) in brackets. Percentage divergence per site is indicated by the scale bar.

than when Bra68 was included (Table 1). Estimated root heights varied correspondingly such that the mean estimate for the age of the entire tree was 214 years (95% HPD 105–371) prior to 2005 if all 73 taxa were included, and 196 years (95% HPD 99–339) when Bra68 was excluded (Table 1). Both indicate a relatively recent origin for the currently circulating lineages of SLEV in the Americas. Ages of nodes estimated using the complete data sets are indicated on the maximum likelihood tree in Fig. 1.

For the above-mentioned analyses, with the exception of lineage 1, all parameters were estimated using the relaxed UCED

Table 1
Evolutionary rates and dates of divergence for SLEV

Data set (n)	Substitution rate ($\times 10^{-4}$ subs/site/yr)		Root height (years)		Date MRCA existed	
	Full data sets	Bra68 excluded	Full data sets	Bra68 excluded	Full data sets	Bra68 excluded
All data (73/72)	4.1 (2.5–5.7)	4.5 (2.7–6.3)	214 (105–371)	196 (99–339)	1791 (1634–1900)	1809 (1666–1905)
Lineages 1–3 (48/47)	4.0 (2.3–5.7)	4.7 (2.7–6.6)	173 (87–315)	144 (78–250)	1832 (1690–1918)	1861 (1755–1927)
Lineages 1 and 2 (44/43)	4.7 (2.8–6.6)	5.4 (3.4–7.5)	135 (75–240)	122 (72–207)	1866 (1761–1926)	1879 (1794–1929)
Lineage 1 (13)	7.2 (2.2–12.4)	–	69 (45–136)	–	1929 (1862–1953)	–
Lineage 2 (31/30)	4.7 (2.8–6.5)	3.7 (2.6–4.7)	93 (72–129)	89 (71–119)	1908 (1872–1929)	1912 (1882–1930)

MRCA, most recent common ancestor; subs/site/yr, substitution per site per year. Values shown are mean estimates with upper and lower 95% HPDs in brackets.

clock model, since this model was supported over both the strict and UCLN clock models and was either decisive (ln BF $\gg$ 4.6) or strongly supported (lineage 2 minus Bra68; BF $>$ 2.3) (Table 2). In the case of lineage 1, estimates were derived using the relaxed UCLN clock model which was the model of best-fit [coeff. of variation = 1.45 (95% HPD 0.59–2.37)]. The relaxed clock models allow for variation in substitution rates amongst branches, with the rate on each branch independently drawn from an exponential distribution or a lognormal distribution (Drummond et al., 2006). In the case of SLEV we found that the mean rate estimated for individual lineages ranged from 3.7 to 7.2 $\times$ 10^{-4} subs/site/yr.

3.3. Patterns of geographical dispersal in SLEV

SLEV exhibits clear population subdivision within the Americas as demonstrated by the large extent of geographic clustering in the phylogeny (Fig. 1). This was also reflected in our analysis of substructure and migration patterns, which revealed that the total number of lineage movement events was less than expected under the null hypothesis of panmixis ($P < 0.002$; Table 3 and S2). However, despite this statistically significant geographic isolation, there is also substantial evidence for gene flow from the 15°N and 30°N latitude regions to the rest of North America. There is also

Table 2
Comparison of molecular clock models

Trace	Ln P(model/data)	Standard error	Ln Bayes factor		
			UCED	UCLN	STRICT
All data (73 taxa)					
UCED	−7808.371	±2.443	–	**14.743**	**39.006**
UCLN	−7823.114	±3.88	−14.743	–	**24.264**
STRICT	−7847.378	±1.189	−39.006	−24.264	–
Lineage 1 (13 taxa)					
UCED	−2873.872	±1.823	–	−2.766	**9.421**
UCLN	−2871.106	±0.59	**2.766**	–	**12.188**
STRICT	−2883.294	±1.721	−9.421	−12.188	–
Lineage 2 (31 taxa)					
UCED	−4090.439	±0.391	–	**5.01**	**10.831**
UCLN	−4095.449	±0.488	−5.01	–	**5.821**
STRICT	−4101.27	±0.845	−10.831	−5.821	–
Lineages 1 & 2 (44 taxa)					
UCED	−5152.152	±0.543	–	**7.47**	**27.112**
UCLN	−5159.622	±2.13	−7.47	–	**19.642**
STRICT	−5179.264	±1.055	−27.112	−19.642	–
Lineages 1 & 2 & 3 (48 taxa)					
UCED	−5727.88	±1.36	–	**8.828**	**28.246**
UCLN	−5736.708	±1.303	−8.828	–	**19.418**
STRICT	−5756.126	±0.619	−28.246	−19.418	–
All data minus Bra68 (72 taxa)					
UCED	−7798.736	±1.717	–	**8.833**	**29.655**
UCLN	−7807.568	±0.959	−8.883	–	**20.823**
STRICT	−7828.391	±0.546	−29.655	−20.823	–
Lineage 2 minus Bra68 (30 taxa)					
UCED	−4081.814	±0.761	–	**2.671**	2.086
UCLN	−4084.485	±0.961	−2.671	–	−0.585
STRICT	−4083.9	±0.725	−2.086	0.585	–
Lineages 1 & 2 minus Bra68 (43 taxa)					
UCED	−5140.465	±0.675	–	**9.837**	**17.799**
UCLN	−5150.302	±1.242	−9.837	–	**7.963**
STRICT	−5158.265	±0.736	−17.799	−7.963	–
Lineages 1 & 2 & 3 minus Bra68 (47 taxa)					
UCED	−5719.345	±0.798	–	**8.882**	**18.615**
UCLN	−5728.226	±1.652	−8.882	–	**9.733**
STRICT	−5737.959	±0.8	−18.615	−9.733	–

Ln Bayes factor (BF) of 2.3–3.4 is strong evidence of supporting the preferred model, ln BF 3.4–4.6 is very strong evidence and ln BF of >4.6 is decisive. BF values greater than 2.3 are shown in bold.

Table 3
Difference between observed and expected number of state changes (gene flow between latitudes)

Origin	Destination			
	Above 30°N	15–30°N	0–15°N	Less 0°N
Above 30°N	–	−5.30	−3.31	−5.48
15–30°N	**2.50**	–	**0.50**	−0.37
0–15°N	−0.04	−0.07	–	−0.07
Less 0°N	**0.45**	−0.52	**0.68**	–
Total				−11.03 $P < 0.002$

°N, degrees North Latitude. Differences greater than zero are in bold.

evidence of lineage migration (albeit to a much lesser extent) from South America (below the equator) into North America (above 30°N) and the southern half of Central America, and from northern to southern regions of Central America (Table 3; summarized in Fig. 2). In an analysis of movement among individual states and/or countries, Texas, Florida, and Brazil were identified as source populations (Table 4; total gene flow from each to all others >0). When we assessed movement within the region bounded by latitudes 15–30°N, which was the major source of virus to other regions (i.e. the Gulf of Mexico [Southern Texas, Florida, Mexico], Guatemala and Haiti; Table 5), we found that the difference between the expected and observed number of state changes was about zero, indicating that there is no significant geographic structuring within this major source population.

4. Discussion

Previous phylogenetic analysis of SLEV identified seven lineages (Kramer and Chandler, 2001). However, using a more robust maximum likelihood approach we identified only six lineages with high bootstrap support. In the aforementioned studies, five Panamanian isolates were divided into lineage IV, containing two 1973 and two 1977 isolates, and lineage VI containing a single isolate from 1983 despite relatively low bootstrap support (71%) and the fact that the latter was loosely associated with lineage IV in the neighbor-joining tree (Kramer and Chandler, 2001). Our analyses provide support for grouping the lineage IV isolates together (bootstrap value of 100%) but there was no support (27%) for assigning the Panama 1983 isolate to an independent lineage. We therefore suggest that the SLEV lineages be renumbered using Arabic numerals as shown in Fig. 1 with the five Panamanian isolates together in lineage 4. A phylogeny based on whole genomes of 23 SLEV isolates including the Panama 1983 isolate and one of the 1973 isolates (Baillie et al., 2008), also supports this grouping, although the proposed lineage 4 is referred to as lineage VI in that publication.

We also provide estimates of rates of evolutionary change in SLEV. The mean rates estimated for all SLEV (4.1 or 4.5×10^{-4} subs/site/yr depending on the data set used) are typical of vector-borne RNA viruses and RNA viruses in general (Jenkins et al., 2002). They are however twice the rate estimated for SLEV by Baillie et al. (2008) on the basis of 23 complete coding regions. While the gene region under investigation may account for some of this difference, the contribution appears to be small

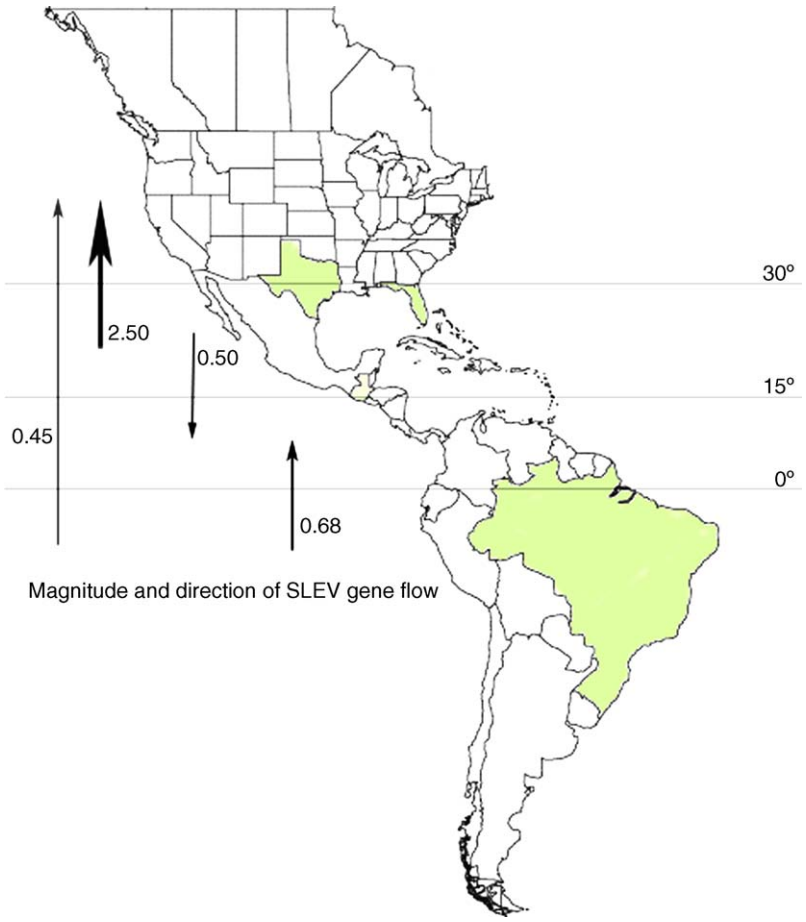

Fig. 2. Pattern of SLEV gene flow among geographic latitudes in the Americas. Direction of gene flow is indicated by arrows with the relative magnitude adjacent.

Table 4
Total difference between observed and expected number of state changes (gene flow from the country/state indicated to all other regions). Total differences greater than zero (i.e. associated with source populations) are highlighted in bold

From	To all other regions
Argentina	−0.63
Brazil	**2.15**
California, USA	−15.33
Colorado, USA	0
Florida, USA	**2.06**
Guatemala	−0.01
Haiti	0
Kentucky, USA	0
Maryland, USA	−0.11
Mexico	0
Mississippi	0
New Mexico, USA	0
Panama	−0.39
Peru	0
Missouri, USA	−0.02
Tennessee, USA	−0.03
Texas, USA	**0.74**
Trinidad	0
Total	−11.56 $P < 0.002$

because when a strict clock was used on our data we obtained rates that were very close to Baillie's estimate [2.3×10^{-4} subs/site/yr (95% HPD 1.7–2.9) for 73 taxa and 2.7×10^{-4} subs/site/yr (95% HPD 2.1–3.4) for 72 taxa vs. Baillie's 2.2×10^{-4} subs/site/yr (95% HPD 1.8–2.6)]. This suggests that the difference observed primarily reflects the use of better fitting relaxed clock models. Our model comparison showed that the relaxed UCED model was significantly better than the strict clock, and in the case of lineage 1 the UCLN clock was about 10^5 times better than the strict clock. Substitution rate variation has been previously demonstrated in a number of RNA viruses (including vector-borne flaviviruses) and suggests that mutation rates, replication rates, or selective constraints vary among lineages (Jenkins et al., 2002; Foster et al., 2004; Drummond et al., 2006 PLoS Biol; Lemey et al., 2006). In the case of SLEV and other arboviruses, alternation between arthropod and vertebrate host species, and variations in extrinsic and intrinsic factors affecting host (and thus virus) population sizes are likely to have a significant effect on the factors mentioned.

Lineages 1 and 2 have been associated with epidemic SLE in North America for several decades. The first epidemic, which occurred in 1933 in St. Louis, Missouri, was caused by a virus later assigned to lineage 2 (Kramer and Chandler, 2001) and involved more than 1000 cases of encephalitis (Lumsden, 1958). Although the incidence of disease has declined in recent years, together lineages 1 and 2 have been responsible for over 4500 reported cases of St. Louis encephalitis (http://www.cdc.gov/ncidod/dvbid/arbor/arbocase.htm). In contrast, prior to a 2005 outbreak in the Cordoba province of Argentina with 47 laboratory-confirmed cases, including 9 fatalities, there were no reports of epidemic disease or even outbreaks of SLE (Diaz et al., 2006). This led to the suggestion that South American strains are less virulent than those from North America. The causative agent belonged to lineage III, which, of the lineages found predominantly in South America, is the most closely related to the North American disease associated strains. As shown in Table 1 the most recent common ancestor for these three lineages is estimated to have existed around the mid-1800s but the 95% HPD is relatively large and spans from the late 17th to early 20th century. The apparently more recent lineages 1 and 2 are estimated to have diverged from lineage 3 about three decades later and from each other around the turn of the 20th century.

Previous phylogenetic and epidemiological data suggest that SLEV is maintained locally from year-to-year in California, Texas and Florida (Reisen et al., 2002). However, it is not clear how the virus is maintained between seasons and it has been suggested that the virus was reintroduced to these areas annually from countries in South and Central America by migrating birds. The approach we have used to infer the magnitude and direction of SLEV gene flow during the period represented by the sequences included is a phylogeny method, and therefore simply describes gene flow among the locations in the sample. A recent study using 1302 complete influenza virus genomes demonstrates that given enough data, hypotheses concerning transmission, persistence and migration can be investigated in much detail (Rambaut et al., 2008). Although further sampling would obviously refine our results, our data set is currently the best possible for this sort of analysis as it includes all published SLEV sequences for which geographic data is currently available.

Our results indicate a pattern of gene flow consistent with northward movement of infected birds with the Gulf of Mexico region (i.e. the region bounded by latitudes 15°N and 30°N) being the major source population for the North American populations sampled. This suggests that birds acquire infection while wintering in the Gulf region and carry it back to their northern breeding grounds in the Spring. The Central, Mississippi and Atlantic migration flyways all traverse this Gulf region, and a number of bird species known to frequent these flyways have been implicated as potential reservoir hosts (Gruwell et al., 2000). As this source population includes Florida and the southern half of Texas (states also identified as source populations in our analysis of gene flow among individual states/countries), we considered the pattern of gene flow within this region (i.e. 15–30°N latitude) to test the

Table 5
Difference between observed and expected number of state changes (gene flow amongst countries/regions between 15°N and 30°N, and between these and the rest of Americas)

Origin:	***Destination***						
	Above 30 °N	**Southern Texas**	**Florida**	**Mexico**	**Guatemala**	**Haiti**	**Below 15 °N**
Above 30 °N	-	-1.27	**0.39**	**0.53**	-0.98	-0.50	-6.25
Southern Texas	-0.02	-	-0.02	0.00	-0.01	0.00	-0.03
Florida	-0.02	-0.01	-	0.00	0.00	0.00	-0.03
Mexico	0.00	0.00	0.00	-	0	0	0
Guatemala	0.00	0.00	0.00	0	-	0	0.0
Haiti	0.00	0.00	0.00	0	0	-	0
Below 15 °N	**0.46**	-0.54	-0.49	-0.13	-0.23	-0.10	-
						TOTAL	**-9.28**
							p < 0.002

°N, degrees North Latitude; gene flow within the region bounded by latitudes 15°N and 30°N is highlighted.

hypothesis that SLEV is regularly reintroduced to Texas and Florida from Central America and/or the Caribbean (specifically Mexico, Guatemala and Haiti; see Table 4). The results indicate that within this major source population the viral population is spatially homogeneous suggesting that there are no barriers to viral gene flow within this geographic region. This has implications for monitoring and control as it suggests that efforts aimed at controlling SLEV may have limited success if they are not co-coordinated across political borders.

Our analyses indicate that California is a sink population that received virus from Brazil and Texas (Table S2). However this influx is insignificant relative to the degree of population subdivision between California and the other regions represented, suggesting a degree of isolation (the total lineage migrations from California to other states is much less than expected; Table 5). This suggests the mechanism of maintenance within California may be different from that in the Gulf region, otherwise one might reasonably expect viral lineages to emanate from this state as well. Alternatively, it is possible that viruses are exported from California, but to regions south and north of California, along the Pacific bird migration flyway, which have not been adequately sampled for virus.

The greater genetic diversity and population subdivision within South America compared to North America may reflect the role of mammals rather than birds as vertebrate hosts in South America and the density of these hosts. Since long distance movement, such as the apparent movement of virus from Brazil to California, is unlikely to have been achieved via an infected terrestrial mammal (humans can engender long distance movement but are dead-end hosts for SLEV), it would seem that birds, or perhaps bats, are the most likely agents of dispersal. However, there is no obvious migration flyway that a bird would follow from Brazil to the Pacific coast of the US. More detailed sampling would be required to ascertain whether this movement was in fact directly between Brazil and California or whether a Brazilian precursor was dispersed to Pacific regions of South America (both Peru and Argentina are represented in this clade), and then carried to California via a bird migrating along the Pacific flyway. The Brazilian isolates in lineage 5 that account for this inferred migration event pre-date the California isolate by about 30 years so an indirect route is highly likely. Likewise the identification of Brazil as a centre of viral diversity (3 out of the 6 well-supported lineages exist in Brazil) may therefore reflect the greater age of SLEV in this location, rather than more recent migration events.

Acknowledgements

We thank the "13th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology" for the training opportunity for AJA. We are also grateful to Orchid Allicock for technical assistance and to Edward Holmes, Jerome Foster and two anonymous reviewers for helpful comments. The Campus Research and Publication Fund of the University of the West Indies, St. Augustine campus, provided financial support for this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2008.07.006.

References

Baillie, G.J., Kolokotronis, S.O., Waltari, E., Maffei, J.G., Kramer, L.D., Perkins, S.L., 2008. Phylogenetic and evolutionary analyses of St. Louis encephalitis virus genomes. Mol. Phylogenet. Evol. 47 (2), 717–728.

Diaz, L.A., Re, V., Almiron, W.R., Farias, A., Vazquez, A., Sanchez-Seco, M.P., Aguilar, J., Spinsanti, L., Konigheim, B., Visintin, A., Garcia, J., Morales, M.A., Tenorio, A., Contigiani, M., 2006. Genotype III Saint Louis encephalitis virus outbreak, Argentina, 2005. Emerg. Infect. Dis. 12 (11), 1752–1754.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Ho, S., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4 (5), e88.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161 (3), 1307–1320.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22 (5), 1185–1192.

Foster, J.E., Bennett, S.N., Carrington, C.V., Vaughan, H., McMillan, W.O., 2004. Phylogeography and molecular evolution of dengue 2 in the Caribbean basin, 1981–2000. Virology 324 (1), 48–59.

Gruwell, J.A., Fogarty, C.L., Bennett, S.G., Challet, G.L., Vanderpool, K.S., Jozan, M., Webb Jr., J.P., 2000. Role of peridomestic birds in the transmission of St. Louis encephalitis virus in southern California. J. Wildl. Dis. 36 (1), 13–34.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54 (2), 156–165.

Kramer, L.D., Chandler, L.J., 2001. Phylogenetic analysis of the envelope gene of St. Louis encephalitis virus. Arch. Virol. 146 (12), 2341–2355.

Lemey, P., Rambaut, A., Pybus, O.G., 2006. HIV evolutionary dynamics within and among hosts. AIDS Rev. 8 (3), 125–140.

Lumsden, L.L., 1958. St. Louis encephalitis in 1933; observations on epidemiological features. Public Health Rep. 73 (4), 340–353.

Maddison, D.R., Maddison, W.P., 2000. MacClade: Analysis of Phylogeny and Character Evolution Version 4. Sinauer Associates Sunderland, MA.

Martin, D., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequences alignments. Bioinformatics 21 (2), 260–262.

Mitchell, C.J., Gubler, D.J., Monath, T.P., 1983. Variation in infectivity of Saint Louis encephalitis viral strains for *Culex pipiens quinquefasciatus* (Diptera: Culicidae). J. Med. Entomol. 20 (5), 526–533.

Mitchell, C.J., Monath, T.P., Sabattini, M.S., 1980. Transmission of St Louis encephalitis virus from Argentina by mosquitoes of the *Culex pipiens* (Diptera: Culicidae) complex. J. Med. Entomol. 17 (3), 282–285.

Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50 (4), 580–601.

Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature 453 (7195), 615–619.

Reisen, W.K., Lothrop, H.D., Chiles, R.E., Cusack, R., Green, E.G., Fang, Y., Kensington, M., 2002. Persistence and amplification of St. Louis encephalitis virus in the Coachella Valley of California, 2000–2001. J. Med. Entomol. 39 (5), 793–805.

Reisen, W.K., 2003. Epidemiology of St. Louis encephalitis virus. Adv. Virus Res. 61, 139–183.

Spence, L., Artsob, H., Grant, L., Th'Ng, C., 1977. St. Louis encephalitis in southern Ontario: laboratory studies for arboviruses. Can. Med. Assoc. J. 116 (1), 35–37.

Spence, L.P., 1980. St Louis encephalitis in tropical America. In: Monath, T.P. (Ed.), St Louis Encephalitis. American Public Health Assoc, Washington, pp. 451–472.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-Time Markov chain evolutionary models. Mol. Biol. Evol. 18, 1001–1013.

Swofford, D.L., 2003. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.