HOSTED BY
ELSEVIER

Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Medicine

journal homepage: http://ees.elsevier.com/apjtm

Original research http://dx.doi.org/10.1016/j.apjtm.2016.07.028

The phylogenetic and evolutionary history of Kokobera virus

Aletheia Blasi[1], Alessandra Lo Presti[2], Eleonora Cella[2,3], Silvia Angeletti[1]✉, Massimo Ciccozzi[1,2]

[1]*Unit of Clinical Pathology and Microbiology, University Campus Bio-Medico of Rome, Italy*

[2]*Department of Infectious Parasitic and Immunomediated Diseases, Istituto Superiore di Sanità, Rome, Italy*

[3]*Public Health and Infectious Diseases, Sapienza University, Rome, Italy*

ARTICLE INFO

Article history:
Received 16 May 2016
Received in revised form 17 Jun 2016
Accepted 16 Jul 2016
Available online 10 Aug 2016

Keywords:
Kokobera virus
Phylogeny
Date of origin
Evolution

ABSTRACT

Objective: To estimate the genetic diversity of Kokobera virus, the date of origin and the spread among different viruses in the endemic regions of Australia.
Methods: Two datasets were built. The first consisting of 29 sequences of the NS5/3′ UTR region of Kokobera group downloaded from GenBank, the second including only 24 sequences of Kokobera viruses, focus is on this group.
Results: Bayesian time analysis revealed two different entries in Australia of Kokobera virus in the 50s years with the dated ancestor in 1861 year. Clades A and B showed a clear separation of the Kokobera sequences according to the geographic region.
Conclusions: Data from the study showed as Kokobera virus, despite of its ancient origin and its circulation before the European colonization, remained limited to the Australian country and nowadays limited mostly to the regions were Australian marsupials are mostly found.

1. Introduction

The genus *Flavivirus* comprises more than 50 RNA virus species that include Yellow fever virus, Dengue virus, Japanese encephalitis virus, and the Tick-borne encephalitis virus complex. Many of these arthropod-borne viruses represent dangerous threats to human health and have been subjected to intensive research to unravel their molecular and virological properties [1]. Flaviviruses have a positive (+) sense RNA genome and replicate in the cytoplasm of the host cells. In general, the genome encodes 3 structural proteins (Capsid, prM, and Envelope) and 8 non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5 and NS5B). The Kokobera group of Flaviviruses (family, Flaviviridae; genus, *Flavivirus*) currently includes 5 candidate species: Kokobera (KOKV), Stratford (STRV), Bainyik (previously strain MK7979), Torres (previously strain TS5273), and New Mapoon (NMV) viruses [2].

KOKV is a mosquito-borne flavivirus that has been isolated from mosquitoes throughout Australia [3]. Originally it was isolated from *Culex annulirostris* (*Cx. annulirostris*) mosquitoes collected at Kowanyama (Mitchell River Mission) in northern Queensland (QLD) in 1960 and was named after a local Aboriginal tribe [4]. It was also isolated from mosquitoes collected in widely separated areas of Australia, including New South Wales (NSW), Western Australia (WA), and the Northern Territory (NT), as well as from Papua New Guinea (PNG) [3–5]. STRV was isolated in 1961 from Cairns [4], the Bainyik virus in 1966 from PNG [6], and the Torres virus in 2000 from Saibai Island in the Torres Strait, QLD [7]. NMV was isolated in northern QLD in 1998 [7].

This study aimed to clarify the relationships between the viruses in the Kokobera group through the comparison of partial sequences of the NS5/3′ UTR region. Furthermore, the genetic diversity of Kokobera group, the date of origin and of the spread among the different viruses in the endemic regions was investigated.

2. Materials and methods

Two dataset were built. The first dataset consisted in 29 sequences of NS5/3′ UTR region (24 of these were KOKV, 2 were STRV, 1 was Bainyik virus, 1 was Torres virus and 1 was NMV) downloaded from GenBank (http://www.ncbi.nlm.nih.

First author: Aletheia Blasi, Unit of Clinical Pathology and Microbiology, University Campus Bio-Medico of Rome, Italy.
E-mail: a.blasi@unicampus.it
✉Corresponding author: Silvia Angeletti, Unit of Clinical Pathology and Microbiology, University Campus Bio-Medico, Via Alvaro del Portillo 200, 00128 Rome, Italy.
Tel: +39 06 225411461
E-mail: s.angeletti@unicampus.it
Peer review under responsibility of Hainan Medical University.

gov/genbank/). The second dataset was generated including only the 24 sequences of KOKV, to make a focus on this group (Table 1).

All the sequences were aligned using ClustalX software followed by manual editing using the Bioedit program v7.2.5, as already described [8].

The phylogenetic signal of the first and second dataset was investigated by means of the likelihood mapping analysis of 10 000 random quartets by using TreePuzzle program as already described [9]. In this analysis, groups of four randomly chosen sequences (quartets) were evaluated using Maximum Likelihood. For each quartet, the three possible unrooted trees were reconstructed under the selected substitution model. The likelihoods of each tree were then plotted on a triangular surface, so that fully resolved trees fall into the corners and the unresolved quartets in the centre of the triangle (indicating a star-like signal). When using this strategy, if more than 30% of the dots fall into the centre of the triangle, the data are considered unreliable for the purposes of phylogenetic inference.

The evolutionary model was chosen, as the best-fitting nucleotide substitution model in accordance with the results of the hierarchical likelihood ratio test implemented in MODELTEST software (version 3.7) [10].

The Bayesian phylogenetic tree was reconstructed by means of Mr Bayes using the HKY + G model of nucleotide substitution for the first dataset and the HKY + I + G for the second dataset.

The evolutionary rate was estimated on the first dataset by using a Bayesian Markov Chain Monte Carlo (MCMC) approach (Beast v. 1.8.2, http://beast.bio.ed.ac.uk) implementing the evolutionary model selected by ModelTest [11,12].

In order to investigate also the demographic history, independent MCMC runs were carried out enforcing both a strict and relaxed clock with an uncorrelated log normal rate distribution and one of the following coalescent priors: constant population size, exponential growth, non-parametric smooth skyride plot Gaussian Markov Random Field, and non-parametric Bayesian skyline plot [11,13,14]. Marginal likelihoods estimates for each demographic model were obtained using path sampling and stepping stone analyses [15–17]. Uncertainty in the estimates was indicated by 95% highest posterior density (95% HPD) intervals, and the best fitting model for each dataset was by calculating the Bayes Factors [16,18]. In practice, any two models can be compared to evaluate the strength of evidence against the null hypothesis (H_0), defined as the one with the lower marginal likelihood: $2lnBF < 2$ indicates no evidence against H_0; 2–6, weak evidence; 6–10: strong evidence, and >10 very strong evidence. Chains were conducted for at least 50×10^6 generations, and sampled every 5000 steps for each molecular clock model. Convergence of the MCMC was assessed by the ESS for each parameter. Only parameter estimates with ESS's of >250 were accepted. Maximum clade credibility trees were obtained from the trees posterior distributions with the Tree-Annotator software v 1.8.2, included in the Beast package [11,12]. Statistical support for specific monophyletic clades was assessed by posterior probability.

Table 1
Kokobera virus group isolates.

Sequence strain	Virus	Year	Place of isolation
OR666	Kokobera virus	1975	Kununurra, WA
SW12	Kokobera virus	1994	South-west, WA
SW10	Kokobera virus	1994	South-west, WA
SW9	Kokobera virus	1994	South-west, WA
SW8	Kokobera virus	1994	South-west, WA
SW7	Kokobera virus	1994	South-west, WA
SW6	Kokobera virus	1994	South-west, WA
SW5	Kokobera virus	1994	South-west, WA
SW3	Kokobera virus	1994	South-west, WA
SW2	Kokobera virus	1994	South-west, WA
SW1	Kokobera virus	1994	South-west, WA
CSIRO31	Kokobera virus	1975	Beatrice Hill, NT
CSIRO37	Kokobera virus	1975	Beatrice Hill, NT
K1077	Kokobera virus	1986	Kimberley, WA
K1059	Kokobera virus	1986	Kimberley, WA
K1004	Kokobera virus	1986	Kimberley, WA
OR408	Kokobera virus	1974	Kimberley, WA
NL26075	Kokobera virus	1981	Lake Narran, NSW
NG516	Kokobera virus	1980	Sandy Camp, NSW
WD26547	Kokobera virus	1981	Wandoona, NSW
WD26383	Kokobera virus	1981	Wandoona, NSW
NL26072	Kokobera virus	1981	Lake Narran, NSW
WD26632	Kokobera virus	1981	Wandoona, NSW
AusMRM32	Kokobera virus	1960	Mitchell River, QLD
TS5273complete	Torres virus	2000	Saibai Island, QLD
MK7979complete	Bainyik virus	1966	Papua, New Guinea
C338complete	Stratford virus	1961	Cairns, QLD
Stratford23759	Stratford virus	1995	Bateman's Bay, NSW
NewMapoonCY1014	New Mapoon virus	1998	New Mapoon, QLD

3. Results

Phylogenetic noise of the first and second datasets was investigated by means of likelihood mapping by using the evolutionary model selected with Modeltest (HKY + G). The percentage of dots falling in the central area of the triangle was 8.9% and 7.2% for the first and second datasets respectively; the dataset didn't show more than 30% of noise and contained sufficient phylogenetic signal (data not shown).

Bayesian phylogenetic tree reconstructed by Mr Bayes on the first and second datasets are shown in Figures 1 and 2, respectively. Phylogenetic relationships among the different viruses were supported by posterior probability >80%.

The Bayesian phylogenetic tree of the first dataset (Figure 1) revealed two main statistically supported clades. In the first clade, the sequence of Torres virus and two sequences of Stratford virus were found. In the second clade, all the sequences of KOKV clustered together and separate from the Bainyik virus sequence,

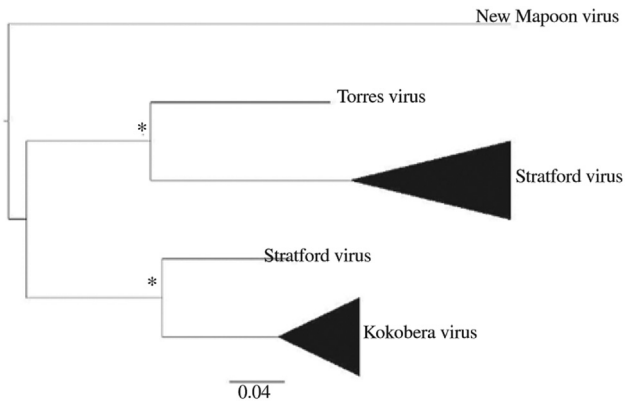

Figure 1. Bayesian phylogenetic tree of first dataset.
24 sequences of KOKV, 2 of STRV, 1 of Bainyik virus, 1 of Torres virus and 1 of NMV. The tree was rooted by using the midpoint rooting. Asterisks (*) along the branches represent significant statistical support (posterior probability $P > 0.90$) for the clade subtending that branch. The scale bar at the bottom indicates 0.04 nucleotide substitutions per site.

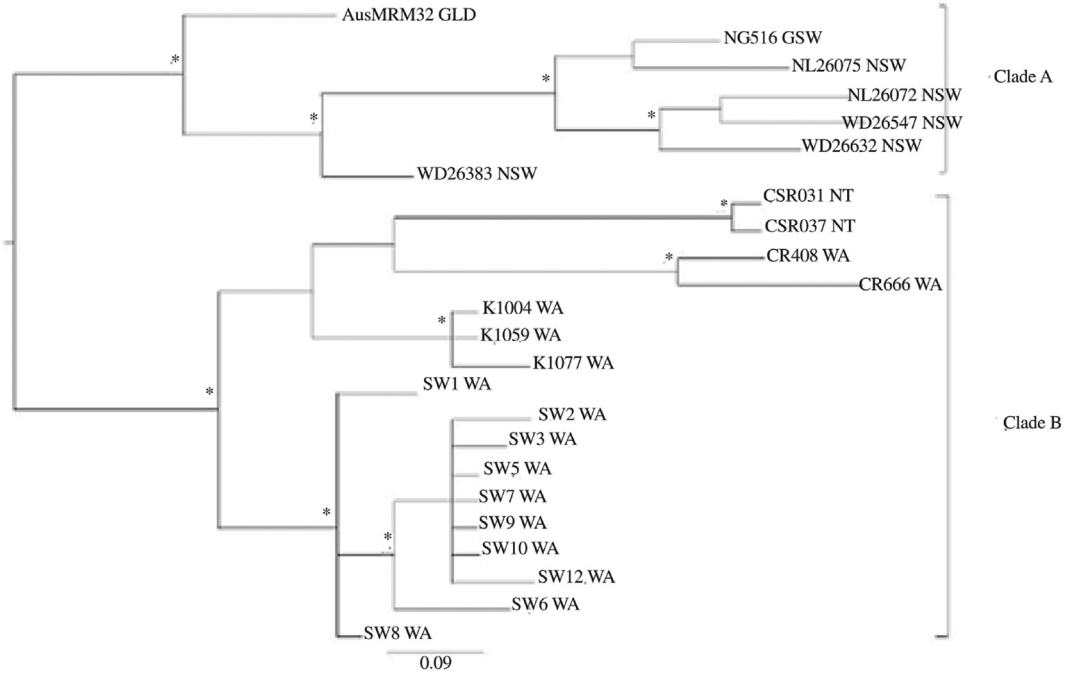

Figure 2. Bayesian phylogenetic tree of the 24 KOKV isolates (second dataset).
The tree was rooted by using the midpoint rooting. Asterisks (*) along the branches represent significant statistical support (posterior probability $P > 0.90$) for the clade subtending that branch. The scale bar at the bottom indicates 0.09 nucleotide substitutions per site. The acronyms NSW, NT, QLD and WA mean New South Wales, Northern Territory, Queensland and Western Australia in the Australia Continent, respectively.

which represented the outgroup of this clade. The New Mapoon sequence was more distantly related to these two clades.

The Bayesian phylogenetic tree of the second dataset (Figure 2) showed that almost all clades were supported. There were two separated clades, A and B respectively. In the clade A there were AusMRM32, that was the first strain isolated in the Mitchell River Mission in the 1960 in Queensland and other six sequences isolated between 1980 and 1981 in the South-East of the Australia just under the Queensland. Instead, in the clade B there were seventeen sequences of KOKV isolated from 1974 to 1994 in the Nord, West of the Australia.

The exponential growth demographic model with a relaxed molecular clock was selected as the most appropriate to describe the evolutionary history of Kokobera group. The estimated mean

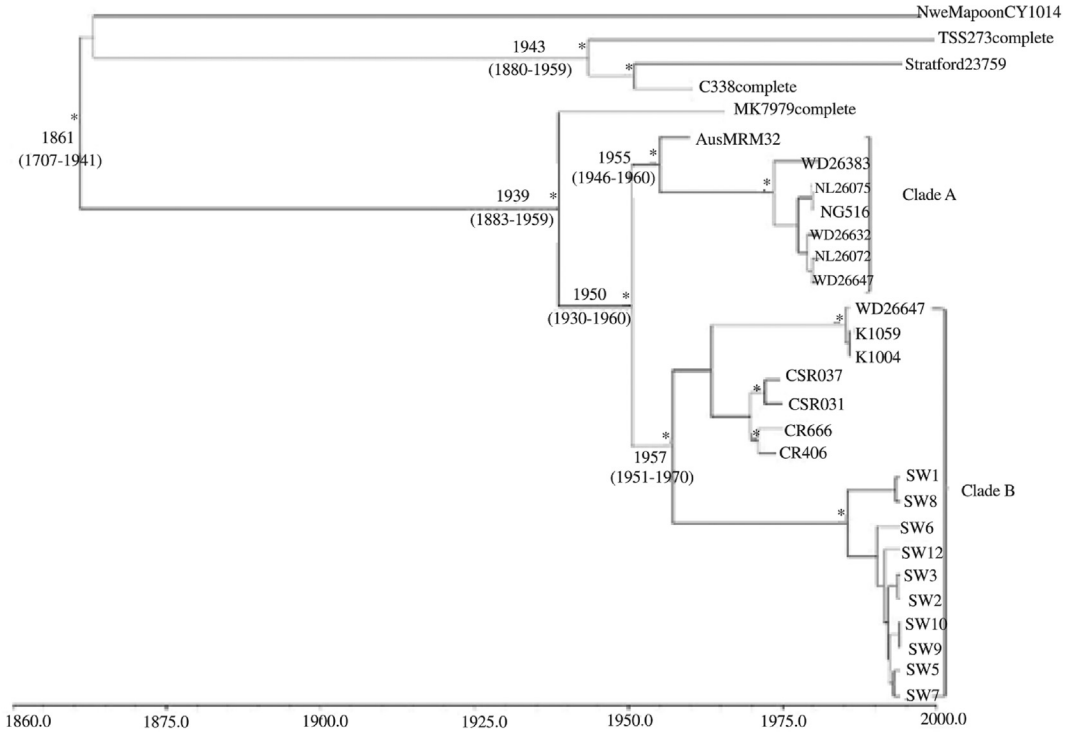

Figure 3. Bayesian maximum clade credibility tree including Kokobera group NS5/3′ UTR region sequences.
Asterisks (*) along the branches represent significant statistical support for the clade subtending that branch (posterior probability $P > 0.90$). The scale at the bottom of the tree represents time in years. Main clades are indicated with letters A and B. Viruses belonging to the Kokobera group other than KOKV (clades A and B) are marked with square.

value of the evolutionary rate for the first dataset was 2.158×10^{-3} substitution/site/year (95% HPD: 9.6×10^{-4}–3.53×10^{-3}). Figure 3 showed the Bayesian maximum clade credibility tree and the most common recent ancestor (tMRCA) estimates conducted on this dataset. The root of the tree had a time of tMRCA corresponding to 1861 (HPD 95%, 1707–1941). The sequences of New Mapoon virus, Torres virus and Stratford virus were in a clade together. The probable tMRCA of Torres and Stratford viruses, was in 1943 (HPD 95%, 1880–1959) and was probably related to when the separation between these two candidate species occurred. The Bainyik virus was outside the KOKV clade as outgroup and the probable tMRCA was 1939 (HPD 95%, 1883–1958). The KOKV sequences had a tMRCA in 1950 (95% HPD 1930–1960). In the 50s there were two different entries in Australia of KOKV forming two major clades: clades A and B respectively. The tMRCA corresponding to 1955 (HPD 95%, 1946–1960) and 1957 (HPD 95%, 1951–1972) for clades A and B, respectively. Interestingly in the clade A there were just sequences from QLD and NSW and in the clade B sequences just from NT and WA, showing a clear separation of the Kokobera sequences according to the geographic region.

4. Discussion

Kokobera virus was first isolated in 1960 from *C. annulirostris* at Kowanyama [4,7]. Together with NMV, isolated in 1998 from *C. annulirostris* mosquitoes, and STRV, isolated from *Aedes vigilax*, KOKV forms a group of strains closely related. Last recent characterization of the KOKV group provides support for the separation of this group into five distinct viruses, KOKV, STRV, NMV, MK7979 and TS5273 [7]. Viruses in the KOKV group are found only in Australia and PNG. Acute polyarticular disease in humans has been attributed to KOKV [19,20]. Only one description of a case of a man who developed encephalitis and myelitis, in whom serological testing suggested KOKV as a cause for his illness was published [21].

Studies on genetic diversity of KOKV have shown isolates from the same geographical area, divided in different cluster [22]. Previous studies showed results probably influenced by isolates from the same year, that can be explained as the result of genetic similarity probably only because from the same outbreak [22].

In this study, the phylogenetic and evolutionary studies provide different suggestions about the diffusion and the history of Kokobera group through Australian continent. KOKV gives a unique clade different from the other viruses of the same groups as recently published coming from different Australian areas. The genetic distance among members of the Kokobera group of flaviviruses supports theirs separation into distinct clades (data not show). The time-scale analysis of 24 KOKV sequences showed two main different clades divided from areas of isolation. Interestingly in clade B two on seventeen (11.7%) were from NT whereas mostly were from WA. In Clade A one of seven (14%) were from QLD whereas seven from NSW. These great distinctions in specific clades for region of isolation reflect only the probability to have different outbreaks and the main clustering within the clade can due to isolation from the same outbreak. The regional distinction does not reflect any vector's isolation, indeed KOKV has been isolated from mosquitoes in Northern Queensland and the Torres Strait [23,24] so as in the Northern Territory [25], in Western Australia [22,26,27] and New South Wales [22] do not implying any specific virus diffusion in specific vector. On the other hand, not only mosquitoes but macropods too, seems to be involved in KOKV diffusion as vertebrate host [28,29]. Serological studies have, indeed, indicated that the KOKV may utilize land based mammals, as hosts [28,29]. This can be an ulterior explanation about the differences in epidemiological patterns observed indicating mosquitos as principal involved vector but mammalian maybe as reservoir. In this study, the mean evolutionary rate of KOKV has been estimated for the first time. Our time-scaled phylogeny reconstruction showed two main clades, labeled A and B, indicating two distinct epidemic entries of this virus. In the dated tree it is also possible to evidence that sequences from both clades A and B, originated in the years 1950's, have a common progenitor dated back to 1861. This confirming that KOKV is an ancient virus, circulating in Australia since the discovery of this continent, happened in the year 1606. Interestingly, this virus even during the European colonization remained limited to the Australian continent and nowadays its spread outside the country has not been described. Going deeper in the evolution of the KOKV virus we can assume that its presence in Australia could have been in the past, when the continent was populated only by the aboriginal people whose presence has been dated back 40 000 years ago. The evolutionary history of the virus conjugated to its limited spread and to its biological cycle reservoir-vector, let us to suppose that KOKV remained limited to the Australian continent as a consequence of the coexistence of all these factors. Moreover speculating or not, about virus spreading outside the Australian continent, it is possible to assume that the virus reservoir is probably represented mostly by kangaroo and wallaby, marsupials typically living in Australia. The marsupials are mainly distributed in the New South Wales, in the South of Australia, in the Northern Territory, areas, where viral sequences represented in the two main clades of the tree (clades A and B) were isolated. These data could suggest that the virus circulation was most probable in the area of the country where the reservoirs were most frequently distributed being the vector instead found in equal proportion through the continent. This aspect could give an important role to the marsupial reservoirs in the spread of the virus and contribute to the absence of Kokobera virus spread outside Australia, being kangaroo and wallaby marsupials the typical Australian fauna. Moreover the politic of quarantine for flora and fauna protection in force in Australia (Australian Quarantine and Inspection Service (www.daff.org.au)) could have played an important role in limiting the spread of the virus outside from the country.

In conclusion, data from the study contributed to get deeper knowledge on the Kokobera virus history and evolution, and getting why, despite of its ancient origin and its circulation among the aboriginal people before the European colonization, the virus remained limited to the Australian country and nowadays it is limited mostly to the regions were Australian marsupials are mostly found.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgment

This work was supported by Dr. Valerio Ciccozzi for the English language revision of the manuscript.

References

[1] Foundation N. *New treatment strategies for dengue and other flaviviral diseases No. 277*. New York: Wiley; 2006, p. 276.

[2] May FJ, Clark DC, Pham K, Diviney SM, Williams DT, Field EJ, et al. Genetic divergence among members of the Kokobera group of flaviviruses supports their separation into distinct species. *J Gen Virol* 2013; **94**(Pt7): 1462-1467.

[3] Mackenzie JS, Lindsay MD, Coelen RJ, Broom AK, Hall RA, Smith DW. Arboviruses causing human disease in the Australasian zoogeographic region. *Arch Virol* 1994; **136**(3–4): 447-467.

[4] Doherty RL, Carley JG, Mackerras MJ, Marks EN. Studies of arthropod-borne virus infections in Queensland III. Isolation and characterization of virus strains from wild-caught mosquitoes in North Queensland. *Aust J Exp Biol Med Sci* 1963; **41**: 17-39.

[5] Russell RC. Arboviruses and their vectors in Australia: an update on the ecology and epidemiology of some mosquito-borne arboviruses. *Rev Med Vet Entomol* 1995; **83**: 141-158.

[6] Poidinger M, Hall RA, Mackenzie JS. Molecular characterization of the Japanese encephalitis serocomplex of the flavivirus genus. *Virology* 1996; **218**(2): 417-421.

[7] Nisbet DJ, Lee KJ, van den Hurk AF, Johansen CA, Kuno G, Chang GJ, et al. Identification of new flaviviruses in the Kokobera virus complex. *J Gen Virol* 2005; **86**(Pt1): 121-124.

[8] Ciccozzi M, Ciccaglione AR, Lo Presti A, Yalcinkaya T, Taskan ZP, Equestre M, et al. Reconstruction of the evolutionary dynamics of the hepatitis C virus 1b epidemic in Turkey. *Infect Genet Evol* 2011; **11**(5): 863-868.

[9] Zehender G, Ebranati E, Bernini F, Lo Presti A, Rezza G, Delogu M, et al. Phylogeography and epidemiological history of West Nile virus genotype 1a in Europe and the Mediterranean basin. *Infect Genet Evol* 2011; **11**(3): 646-653.

[10] Swofford D, Sullivan J. Phylogeny inference based on parsimony and other methods with PAUP. In: Lemey P, Salemi M, Vandamme AM, editors. *The phylogenetic handbook-a practical approach to phylogenetic analysis and hypothesis*. 2nd ed. New York: Cambridge University Press; 2009, p. 267-312.

[11] Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005; **22**(5): 1185-1192.

[12] Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007; **7**: 214.

[13] Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 2002; **161**(3): 1307-1320.

[14] Minin VN, Suchard MA. Counting labeled transitions in continuous-time Markov models of evolution. *J Math Biol* 2008; **56**(3): 391-412.

[15] Baele G, Lemey P. Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. *Bioinformatics* 2013; **29**(16): 1970-1979.

[16] Baele G, Lemey P, Vansteelandt S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics* 2013; **14**: 85.

[17] Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* 2013; **30**(2): 239-243.

[18] Kass RE, Raftery AE. Bayes factor. *J Amer Stat Assoc* 1995; **90**(430): 773-795.

[19] Boughton CR, Hawkes RA, Naim HM. Illness caused by a Kokobera-like virus in south-eastern Australia. *Med J Aust* 1986; **145**(2): 90-92.

[20] Mein J, O'Grady KA, Whelan P, Merianos A. Dengue or Kokobera? A case report from the top end of the Northern Territory. *Commun Dis Intell* 1998; **22**(6): 105-107.

[21] Warrilow D, Hall-Mendelin S, Hobson-Peters J, Prow NA, Allcock R, Hall RA. Complete coding sequences of three members of the Kokobera group of flaviviruses. *Genome Announc* 2014; **2**(5). e00890–14.

[22] Poidinger M, Hall RA, Lindsay MD, Broom AK, Mackenzie JS. The molecular epidemiology of Kokobera virus. *Virus Res* 2000; **68**(1): 7-13.

[23] Doherty RL, Carley JG, Kay BH, Filippich C, Marks EN, Frazier CL. Isolation of virus strains from mosquitoes collected in Queensland, 1972–1976. *Aust J Exp Biol Med Sci* 1979; **57**(5): 509-520.

[24] Hanna JN, Ritchie SA, Phillips DA, Lee JM, Hills SL, van den Hurk AF, et al. Japanese encephalitis in north Queensland, Australia, 1998. *Med J Aust* 1999; **170**(11): 533-536.

[25] Standfast HA, Dyce AL, St George TD, Muller MJ, Doherty RL, Carley JG, et al. Isolation of arboviruses from insects collected at Beatrice Hill, Northern Territory of Australia, 1974–1976. *Aust J Biol Sci* 1984; **37**(5–6): 351-366.

[26] Liehne PF, Anderson S, Stanley NF, Liehne CG, Wright AE, Chan KH, et al. Isolation of Murray Valley encephalitis virus and other arboviruses in the Ord River Valley 1972–1976. *Aust J Exp Biol Med Sci* 1981; **59**(Pt 3): 347-356.

[27] Van Den Hurk AF, Johansen CA, Zborowski P, Phillips DA, Pyke AT, Mackenzie JS, et al. Flaviviruses isolated from mosquitoes collected during the first recorded outbreak of Japanese encephalitis virus on Cape York Peninsula, Australia. *Am J Trop Med Hyg* 2001; **64**(3–4): 125-130.

[28] Doherty RL, Carley JG, Gorman BM, Buchanan P, Welch JS, Whitehead RH. Further serological investigations of antibodies to group B arboviruses in man and animals (1964). Studies of arthropod-borne virus infections in Queensland IV. *Aust J Exp Biol Med Sci* 1964; **42**: 149-164.

[29] Doherty RL, Standfast HA, Domrow R, Wetters EJ, Whitehead RH, Carley JG. Studies of the epidemiology of arthropod-borne virus infections at Mitchell River Mission, Cape York Peninsula, North Queensland IV. Arbovirus infections of mosquitoes and mammals, 1967–1969. *Trans R Soc Trop Med Hyg* 1971; **65**(4): 504-513.