



# WNV Typer: A server for genotyping of West Nile viruses using an alignment-free method based on a return time distribution



Pandurang Kolekar<sup>a</sup>, Nilesh Hake<sup>a</sup>, Mohan Kale<sup>b</sup>, Urmila Kulkarni-Kale<sup>a,\*</sup>

<sup>a</sup> Bioinformatics Centre, University of Pune, Pune 411007, India

<sup>b</sup> Department of Statistics, University of Pune, Pune 411007, India

## ABSTRACT

### Article history:

Received 16 July 2013

Received in revised form

27 November 2013

Accepted 17 December 2013

Available online 31 December 2013

### Keywords:

West Nile virus

Genotyping

Return time distribution

Alignment-free

Phylogeny

Bioinformatics

West Nile virus (WNV), genus *Flavivirus*, family *Flaviviridae*, is a major cause of viral encephalitis with broad host range and global spread. The virus has undergone a series of evolutionary changes with emergence of various genotypic lineages that are known to differ in type and severity of the diseases caused. Currently, genotyping is carried out using molecular phylogeny of complete coding sequences and genotype is assigned based on proximity to reference genotypes in tree topology. Efficient epidemiological surveillance of WNVs demands development of objective criteria for typing. An alignment-free approach based on return time distribution (RTD) of  $k$ -mers has been validated for genotyping of WNVs. The RTDs of complete genome sequences at  $k = 7$  were found to be optimum for classification of the known lineages of WNVs as well as for genotyping. It provides time and computationally efficient alternative for genome based annotation of WNV lineages. The development of a WNV Typer server based on RTD is described (<http://bioinfo.net.in/wnv/homepage.html>). Both the method and the server have 100% sensitivity and specificity.

© 2014 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

West Nile virus (WNV; genus *Flavivirus*, family *Flaviviridae*) is a mosquito-borne *Flavivirus* widely distributed across all the continents except Antarctica. It consists of positive sense single stranded RNA genome (~11 kb), enclosed in a ~40–60 nm spherical envelope. The translated polyprotein gives rise to three structural (capsid, pre-membrane and envelope) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5). The genomic organization and translation mechanism of WNV is similar in principle to other flaviviruses (Brinton, 2001; Westaway et al., 2002; Evans et al., 2011; Pesko and Ebel, 2012). *Passeriformes* birds and culicine mosquitoes are the major hosts in enzootic cycle of WNV and are responsible for the enormity of transmission (Bernard et al., 2001). The large variety of hosts is a key factor in wide dispersal of WNV in the tropical and temperate worlds. Understanding the

relative host specificity exhibited by WNVs is still an open challenge in virology research.

The phylodynamics of WNV lineages has been the area of interest and research for epidemiologist to understand their evolutionary emergence and spread in the last two decades due to the significant health burden posed by it (May et al., 2011; McMullen et al., 2011). Several studies have led to an understanding of the mechanisms of evolutionary emergence and transmission of these viruses across the globe (Anderson et al., 2001; Ebel and Dupuis, 2001; Beasley et al., 2003; Ebel et al., 2004; Davis et al., 2005; Bertolotti et al., 2007; Maes et al., 2009; Armstrong et al., 2011; McMullen et al., 2011).

Based on serological cross-reactivity, WNV is classified into Japanese encephalitic complex of flaviviruses (Calisher, 1989). Based on the molecular phylogeny and analysis of their complete coding sequences, two major lineages (I and II) and two additional lineages (III and IV) of WNV are currently accepted. One putative Spanish lineage of WNV has been recently proposed based on the analysis of NS5 region (Vazquez et al., 2010; McMullen et al., 2011; Pesko and Ebel, 2012). Lineage I is further classified into three clades (a–c) of which clade ‘a’ is further subdivided into 6 clusters. The Indian WNV isolates, which were earlier assigned to putative lineage V (Bondre et al., 2007), have been assigned to the lineage Ic (Lineage I, clade c) in recent publications (May et al., 2011). It may be noted that the terms genotype and lineages are used interchangeably. The members of lineage I are distributed worldwide

\* Corresponding author. Tel.: +91 20 2569 0195; fax: +91 20 2569 0087.

E-mail addresses: [pandurang@bioinfo.net.in](mailto:pandurang@bioinfo.net.in) (P. Kolekar), [mmkale@stats.unipune.ac.in](mailto:mmkale@stats.unipune.ac.in) (M. Kale), [urmila@bioinfo.net.in](mailto:urmila@bioinfo.net.in), [urmila.kulkarni.kale@gmail.com](mailto:urmila.kulkarni.kale@gmail.com) (U. Kulkarni-Kale).

and are being studied intensively. The severity and type of disease caused by different lineages of WNVs are dissimilar; for example, NY99 (clade Ia) shows enhanced pathogenesis in birds, whereas the Kunjin virus (clade Ib) exhibits attenuated infection and low neuroinvasion (Brault et al., 2007; Daffis et al., 2011). On the other hand, the severity of the disease caused by the members of lineage II is less and frequency of neuroinvasion is also low. Encephalitis infection caused by lineage II strains in both humans and horses has been reported in South Africa (Venter et al., 2009; Venter and Swanepoel, 2010). Other WNV lineages do not cause mortality but are associated with lower virulence (Davis et al., 2005; Bondre et al., 2007; Hubalek et al., 2010). The differences in processes of migration, adaptation, virulence and neuroinvasiveness in natural hosts and vectors have given rise to worldwide diversity of observed WNV lineages; hence the virus has been termed as an ecological generalist (Pesko and Ebel, 2012). These processes have led to emergence of new lineages, globalization and invasion of WNV (Kilpatrick, 2011). In view of such a wide diversity of WNV, there is need of a web server for rapid identification of the circulating lineages of WNVs for use in diagnosis, control, epidemiological surveillance and further research on the virus.

The present study reports the development of a server for genotyping of WNVs using an alignment-free method developed in-house (Kolekar et al., 2010, 2011, 2012). The method is based on return time distribution (RTD) and has been developed for molecular phylogeny analyses and its applications for genotyping of Mumps viruses and Dengue viruses (Kolekar et al., 2010, 2011, 2012). The current study reports advancement in the RTD-based method to further reduce computations by identifying informative  $k$ -mers and thereby enhancing the performance of the method. It also demonstrates applications of the method in large scale whole genome phylogeny, for in silico lineage typing and annotations of WNV genomes.

## 2. Materials and methods

### 2.1. Datasets

Following datasets were compiled and curated to benchmark the performance of RTD-based alignment-free method for the genotyping of WNVs. List of entries in the respective datasets with GenBank accession numbers is provided in Appendix A.

#### 2.1.1. Reference dataset

The reference dataset consisted of complete genome sequences (~10.3 kb each) of 61 WNVs. The genotype or lineage information for these sequences was available from literature reported earlier (May et al., 2011; McMullen et al., 2011).

#### 2.1.2. True positive dataset

All the available (627) complete genome sequences of WNVs were extracted from West Nile virus database available at <http://www.ncbi.nlm.nih.gov/genomes/VirusVariation/Database/select.cgi?taxid=11082> together with their lineage information to the level of clade (as on June 30, 2013). The entries in the reference dataset were excluded. Thus, the true positive dataset consists of 566 entries. The cluster level annotations however were not available for lineage I clade 'a'.

In order to assign the cluster level information to the members of clade 'Ia' in true positive dataset, independent alignment based molecular phylogeny analysis was carried out. The cluster level annotations were assigned based on proximity with the entries of reference dataset. The phylogenetic analysis was carried out using distance based Neighbor-joining method (100 bootstrap replicates) implemented in MEGA 5 package (Saitou and Nei, 1987; Tamura et al., 2011). The multiple sequence alignment (MSA) was obtained using MUSCLE program in the MEGA 5 (Edgar, 2004). The

lineage/clade/cluster information for the entries in this dataset is made available in Appendix A.

#### 2.1.3. Negative dataset

Complete genome sequences of 57 representative members of the family *Flaviviridae* excluding WNV were retrieved from GenBank to compile the negative dataset. True positive and negative datasets were used to evaluate accuracy of the RTD-based genotyping of WNVs using receiver operating characteristic (ROC) analysis.

### 2.2. Method

Development of an alignment-free RTD-based method and its applications for molecular phylogeny and genotyping of viruses have been reported (Kolekar et al., 2010, 2011, 2012). The method converts the sequence data into return times. Sample sequence and calculation of return time and its parameters is available in the Author's preprint of Kolekar et al. (2012) provided on the server <http://bioinfo.net.in/wnv/homepage.html> under "How RTD works?" pull down. The steps in RTD-based method for molecular phylogeny are (1) compilation of sequence dataset, (2) choice of the value of  $k$ , (3) computation of RTDs and their parameters in each sequence at chosen value of  $k$ , (4) deriving the pair-wise Euclidian distance matrix and (5) building the phylogenetic tree. The RTD-based genotyping method demands optimization of value of  $k$  that leads to genotyping with high sensitivity and specificity (Kolekar et al., 2011, 2012). A brief account of optimization of the value of  $k$  for genotyping of WNVs is given below.

#### 2.3. Optimization of value of $k$ for genotyping of WNVs using return time distribution

Initially, the statistical parameters viz., mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of each of the  $4^k$  possible RTDs of  $k$ -mers in each of the nucleotide sequences of reference dataset were computed separately for varying values of  $k$  ranging from 1 to 8 as described earlier (Kolekar et al., 2011). Thus each genomic sequence was represented as the  $2 \times 4^k$  dimensional numeric vector. The Euclidian type pair-wise distance matrix was then derived for each value of chosen  $k$  from the respective numeric vectors to infer the Neighbor-joining (NJ) phylogenetic tree (Saitou and Nei, 1987; Kolekar et al., 2012). The value of  $k$  for which the WNVs were accurately classified as per the known genotype information was chosen as the optimum value of  $k$  for subsequent analysis.

#### 2.4. Selection of informative $k$ -mers at optimized value of $k$

As the value of  $k$  increases, demand on computation and memory also increases. In order to select the informative RTDs of  $k$ -mers, the parameters of their RTDs ( $\mu$  and  $\sigma$ ) obtained at optimized value of  $k$  (7 in case of WNV) were given as an input to Infogain attribute evaluator with Ranker method assembled in WEKA software (settings as recommended) (Mark et al., 2009). The lineages of WNVs (to the clade level) were used as class labels (Ia, Ib, Ic, II, III, IV). Two sets of informative  $k$ -mers, those whose parameters were ranked among the top 100 and the top 200 respectively, were tested independently for their efficacy to resolve the lineages of WNVs.

#### 2.5. Development of server for genotyping of WNVs

The server for genotyping of WNVs, named as "WNV Typer", is developed by using RTD-based alignment-free method with the help of Apache, CGI, PHP architecture and is available in public domain at <http://bioinfo.net.in/wnv/homepage.html>.

### 3. Results

Since its introduction in New York in 1999, the *West Nile virus* has diversified extensively, which is very evident from the phylogeographic and population genetics studies carried out recently (May et al., 2011; Pesko and Ebel, 2012). In view of these studies, WNVs is expected to continue to evolve into additional genotypic lineages. Therefore, the effective diagnosis and control of WNV epidemics demands methods for genotyping. Classical genotyping approaches requires the identification of complete genome sequences and alignment-based molecular phylogeny analyses (MPA). However, MPA-based genotyping is at times open to subjective interpretation of tree topology. Furthermore, multiple sequence alignment of whole genomes in case of large-scale phylogeny analysis demands high-end computational facility and excessive amount of time. Hence, a new RTD-based alignment-free method for MPA has been developed (Kolekar et al., 2012). Its application for genotyping was demonstrated for mumps and dengue viruses, where empirical cut-off criteria for genotyping were developed using the reference datasets for respective viruses (Kolekar et al., 2011, 2012). The importance of development of pre-computed empirical cut-off for geno/serotyping has resulted in achieving two objectives. The first one is elimination of MSA and MPA while typing new isolates of known geno/serotypes and/or newly emerging novel geno/serotypes. The second objective is development of a user-friendly server to facilitate MPA and typing of viral populations. In case of emergence of new genotype/s, classical MPA requires repeating of the exercise of MSA of all sequences whereas RTD based alignment-free method does not require such compute intensive, time consuming step and only compares the new sequences to pre-computed RTDs of reference dataset. Therefore, to facilitate epidemiological surveillance and to monitor WNV genotypes circulating in the population, an RTD-based server has been developed.

In order to develop the RTD-based genotyping server, a WNV reference dataset was compiled to include representative members of all the four major lineages (I–IV). The lineage I has 3 clades (a–c) where clade 'I-a' is further subdivided into 6 clusters (1–6). These lineages have been determined using extensive molecular phylogeny analysis of complete genome sequences of WNVs. These genome sequences of WNVs are found to differ from each other by 5–25% (Pesko and Ebel, 2012). Recently, an additional putative lineage of WNV from Spain (GenBank: GU047875) has been proposed using molecular phylogeny analysis of NS5 gene (Vazquez et al., 2010). This putative lineage was not included in the reference dataset at present. It could be considered for inclusion in the reference dataset once its lineage status is validated and its complete genome sequence becomes available. Thus, the reference dataset includes all confirmed lineages of WNVs. It was used as a training dataset to optimize value of  $k$  for the RTD-based genotyping of WNVs and to validate applicability of the method for WNV genotyping.

The optimization of value of  $k$  is the first step in RTD-based genotyping analysis as described and reported earlier in the case of genotyping of *Mumps virus* (Kolekar et al., 2011). In case of WNV, the optimum value of  $k$  was found to be 7. It was observed that all the known genotypic lineages and clades of WNV in the reference dataset clustered accurately. Being the only member, Ia6, was found to cluster at the root of Ia5 group. Subsequent availability of multiple strains of Ia6 cluster will help to resolve evolutionary relationships between Ia6 and Ia5 (see Fig. 1).

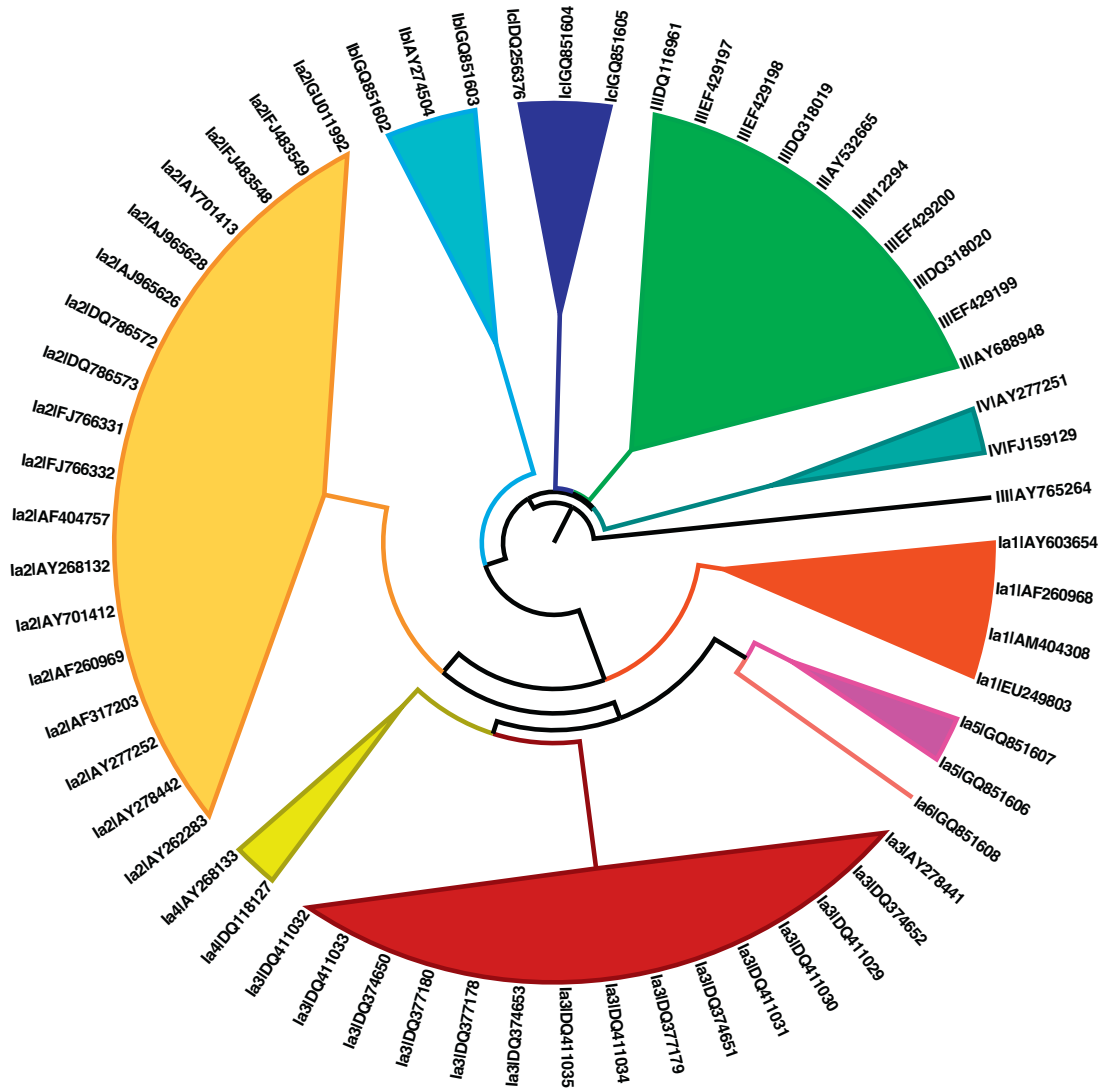
At  $k=7$ , the number of RTDs (heptamers) increases to 16,384 (i.e.  $4^7$ ). Since each RTD was summarized using two statistical parameters viz. ( $\mu$  and  $\sigma$ ), each complete genome sequence was represented as the numeric vector of size  $2 \times 16,384$ . The pair wise distances between these numeric vectors in

reference dataset were calculated using Euclidian type distance function described and reported earlier (Kolekar et al., 2011, 2012). The resultant distance matrix was used as an input to NJ method implemented in Neighbor program of PHYLIP package [<http://evolution.genetics.washington.edu/phylip.html>] to infer the phylogenetic tree as shown in Fig. 1. It took  $\sim 154$  s to infer this phylogenetic tree using RTD-based method, implemented in Perl script on a desktop configuration with 32-bit Windows operating system, 4 GB RAM and 2.80 GHz processor.

It was observed that out of 16,384 possible  $k$ -mers (at  $k=7$ ), means of RTDs of  $\sim 14,048$   $k$ -mers were accounting to zero indicating the absence of "return times" of those  $k$ -mers in the complete genome sequence of *West Nile virus* strain EthAn4766 (GenBank: AY603654), as an example. Therefore, there could be many  $k$ -mers, which may or may not reappear in all the WNV genomes. Thus, in order to further save the time of analysis while retaining the accuracy of genotyping, it was decided to explore if number of RTDs could be reduced based on their information content. As described in method section, the Infogain feature selection method available in WEKA software was used with recommended settings to score the RTD parameters. Infogain calculates the information gain for parameters of each of the RTDs and rank them based on mutual information content for the purpose of classification. We, thus used  $k$ -mers whose RTD based parameters were ranked among top 100 and 200. These two groups of  $k$ -mers were independently evaluated using RTD-based approach for their ability to cluster WNVs in known genotypic lineages using the reference dataset. It was observed that the tree topology obtained using the RTDs of top 100 informative  $k$ -mers did not corroborate with the known lineage structure of WNVs. However, the parameters of RTDs of  $k$ -mers, which were ranked among top 200, were found to be sufficient to accurately resolve all the genotypic lineages of WNVs as shown in Fig. 1. This analysis resulted in significant reduction in the number of computations in RTD-based analysis. Selection of informative RTDs also help in reduction of computational complexity and in saving time that would have been otherwise required for analysis involving all possible  $4^k$   $k$ -mers. It was observed that the time for the analysis decreased from  $\sim 154$  s to  $\sim 3$  s without compromising accuracy. Thus, the modification involving use of only informative RTDs in the RTD based phylogeny analyses proved to be successful and resulted in gain with respect to time while maintaining accuracy.

#### 3.1. "WNV Typer" server for genotyping of WNVs using RTD-based method

The successful genotyping of WNVs in reference dataset allowed us to develop the RTD-based genotyping server for WNVs, named here as "WNV Typer". The server accepts the complete genome sequence(s) ( $\sim 11$  kb) of the WNVs in FASTA format through interactive sequence submission form. Upon submission of sequence(s) the CGI script at the backend computes the statistical parameters of RTDs of 200 informative heptamers as reported in the previous section. The Euclidian type distance between query sequences and reference dataset is then calculated. A two-step procedure was adopted to assign the lineage to query sequence. In the first step, the closest reference genotype from reference dataset was identified using the distance measure. In the second step, it was checked whether the distance of the query sequence lies in the pre-computed range of distances. In case of failure to meet these criteria, the message "No lineage can be assigned to this query sequence using RTD-based method" will be returned on the 'WNV Typer' server. Thus, the RTD-based method not only helps in substantially reducing time for typing of WNVs but also brings in objectivity in the assignment of lineages. The steps involved in RTD-based genotyping of WNV are shown in Fig. 2. It is to be noted that WNV Typer

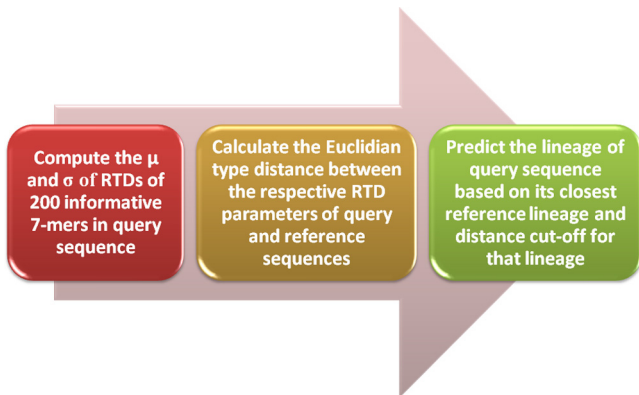


**Fig. 1.** The phylogenetic tree obtained for the reference dataset of complete genome sequences of *West Nile Virus* using RTD-based alignment-free method at  $k=7$ . Note: Tip labels in tree are divided into two parts by '|' (pipe) character. First part indicates the genotype of sequence wherein Roman and Arabic numerals respectively indicate lineages and clusters while lower case alphabets indicate the clades as per WNV nomenclature. Second part indicates the GenBank accession number of the sequence. The tree was drawn using FigTree program [<http://tree.bio.ed.ac.uk/software/figtree/>].

predicts the lineage of query sequence up to lineage/clade/cluster level as applicable.

### 3.2. Validation of performance of the server

In order to assess the performance of WNV Typer, true positive and negative datasets as mentioned in the section of datasets were compiled. The true positive dataset consisted of 566 complete genome sequences of WNVs with known lineage information (only to the clade level) as provided in *West Nile virus* database at National Center for Biotechnology Information (NCBI), Bethesda, USA. The cluster level information was assigned using alignment based molecular phylogenetic analysis. The lineage-clade-cluster data of WNV isolates in reference and true positive datasets are given in Table 1. As listed in Table 1, genomic sequences of lineage Ia outnumber other WNV lineages. The RTD based WNV Typer server predicted accurately all the entries in the true positive dataset in accordance with their lineage-clade-cluster level, thus achieving 100% sensitivity. The RTD-based prediction results corroborate with the alignment based molecular phylogeny. However,



**Fig. 2.** The steps involved in RTD based method for genotyping of *West Nile viruses* using complete genome sequences.

**Table 1**  
The number of entries in the reference and true positive data sets.

Lineage-clade-cluster	Reference dataset	True positive dataset
Ia1	4	7
Ia2	18	11
Ia3	15	2
Ia4	2	527
Ia5	2	0
Ia6	1	0
Ib	3	6
Ic <sup>a</sup>	3	1
II	10	10
III	1	0
IV	2	2
Total	61	566

<sup>a</sup> Putatively known as lineage 5.

it must be mentioned that the true positive sequences for lineage Ia5, Ia6 and III were not available for current round of evaluation.

The negative dataset includes the complete genome sequences of representative members of the family *Flaviviridae*. None of the entries in the negative dataset was assigned to any WNV lineage resulting in 100% specificity of the RTD-based alignment-free WNV Typer server.

#### 4. Discussion

The availability of an increasing number of genome sequences of WNV has facilitated the understanding of WNV evolution in greater detail. In the view of recent epidemics and increasing genomic data of WNVs, it is important to keep track of diverse lineages and their pattern of evolution. In order to achieve this task it is very important to have an efficient mechanism in place to identify and annotate the genomic sequences with lineage/clade/cluster information. Such a mechanism would permit epidemiological surveillance, as well as understanding the phylodynamics and pattern of lineage specific disease progression and diagnosis of WNVs. As of now, the lineage specific study of interaction of WNV with Human Leukocyte Antigen (HLA) and Killer immunoglobulin-like receptors (KIR) does not seem to have been done but recently it has been reported that specific KIR genotypes could be connected with WNVs infection (Spiroski et al., 2013). The lineage specific information of WNV would help to investigate whether there is any association between the specific interaction mechanisms adopted by virus and host systems. On the other hand, a strong evidence of correlation has been found between the neuroinvasive virulence phenotype of the WNV infection and the genotype/lineage of the WNV when studied using

mouse model (Beasley et al., 2002). The identification of WNV lineages thus may enable the prediction of the virulence pattern and its consequences, which in turn would assist in diagnosis and treatment of the infection.

The available alignment-based MPA methods, though useful, have a number of limitations. Complexity of analysis is directly proportional to the number and size of genomic sequences. Furthermore, whole analysis starting with MSA needs to be repeated even when a single sequence is to be annotated. In this context, RTD-based alignment-free method proves to be an efficient and accurate option for annotating lineage data of WNV genomic sequences. The development of server using reference dataset adds an advantage by eliminating several steps in conventional MSA based MPA. The molecular phylogeny analysis of 627 complete genomes in reference and true positive datasets using alignment-free RTD approach using 200 informative heptamers required just 10 min for successful completion. Further, the WNV Typer server requires about only 1 s to assign the lineage/clade/cluster per genome sequence providing a distinctive advantage over alignment-based phylogenetic analysis.

#### 5. Conclusions

The return time distribution based alignment-free method was suitably modified and successfully applied for the genotyping of *West Nile viruses*. A web server is made available online and is expected to play a role on lineage/clade/cluster assignment and thus may help in epidemiological surveillance of WNVs. The method provides a robust, scalable and computationally efficient alternative for large-scale genotyping over the conventional MPA analysis.

#### Acknowledgements

This work is supported by the Department of Biotechnology (DBT), Government of India, New Delhi, under Centre of Excellence (COE) grant to the Bioinformatics Centre, University of Pune, Pune (India). UKK acknowledges DBT for financial assistance. PSK acknowledges the Bioinformatics National Certification (BINC) fellowship awarded by the DBT. Part of this project was carried out by NH under supervision of Mr. Sanjay Londhe, Information Scientist, Bioinformatics Centre, University of Pune toward fulfillment of the requirements for M.Sc. degree in Bioinformatics. The authors acknowledge Dr. A.N. Latey for useful discussions and enhancing the readability of the manuscript.

## Appendix A. GenBank accession numbers and information about the datasets used in the study.

### 1. Reference dataset.

Sr_no.	Accession_no.	Length	Strain\isolate	Genotype: lineage\clade\cluster
1	AY603654	11,029	EthAn4766	Ia1
2	AF260968	11,029	Eg101	Ia1
3	AM404308	10,979	PTROxo	Ia1
4	EU249803	11,029	68856	Ia1
5	GQ851607	10,962	IBAN7019	Ia5
6	GQ851606	10,951	ArD27875	Ia5
7	GQ851608	10,964	ArB310/67	Ia6
8	AY262283	10,984	KN3829	Ia2
9	AY278442	10,842	LEIV-Vlg00-27924	Ia2
10	AF317203	10,972	VLG-4	Ia2
11	AY277252	10,845	LEIV-Vlg99-27889	Ia2
12	AF260969	11,029	LEIV-Vlg99-27889	Ia2
13	AY701412	10,945	LEIV-Vlg99-27889	Ia2
14	FJ766332	10,961	GE-2o/V	Ia2
15	FJ766331	10,960	GE-1b/B	Ia2
16	AF404757	11,029	WN Italy 1998-equine	Ia2
17	DQ786573	10,938	France 407/04	Ia2
18	DQ786572	10,938	France 405/04	Ia2
19	FJ483548	11,030	France 405/04	Ia2
20	FJ483549	11,030	France 405/04	Ia2
21	GU011992	11,030	France 405/04	Ia2
22	AJ965626	10,979	France 405/04	Ia2
23	AJ965628	10,977	France 405/04	Ia2
24	AY701413	10,945	France 405/04	Ia2
25	AY268132	10,989	France 405/04	Ia2
26	AY278441	10,998	Ast99-901	Ia3
27	DQ374652	10,789	Ast04-2-824A	Ia3
28	DQ411029	10,811	Ast01-66	Ia3
29	DQ411030	10,811	Ast01-182	Ia3
30	DQ411031	10,811	Ast01-187	Ia3
31	DQ374651	10,811	Ast02-3-570	Ia3
32	DQ377179	10,811	Ast02-2-298	Ia3
33	DQ411032	10,811	Ast02-3-146	Ia3
34	DQ411033	10,811	Ast02-3-165	Ia3
35	DQ411034	10,811	Ast02-2-691	Ia3
36	DQ411035	10,811	Ast02-2-692	Ia3
37	DQ374653	10,811	Ast02-2-25	Ia3
38	DQ377178	10,811	Ast02-2-26	Ia3
39	DQ377180	10,811	Ast02-3-208	Ia3
40	DQ374650	10,811	Ast02-3-717	Ia3
41	DQ118127	10,969	goose-Hungary/03	Ia4
42	AY268133	10,989	goose-Hungary/03	Ia4
43	GQ851602	10,981	goose-Hungary/03	Ib
44	AY274504	11,022	goose-Hungary/03	Ib
45	GQ851603	10,983	goose-Hungary/03	Ib
46	DQ256376	11,014	goose-Hungary/03	Ic/V <sup>a</sup>
47	GQ851604	10,489	goose-Hungary/03	Ic/V
48	GQ851605	10,997	goose-Hungary/03	Ic/V
49	DQ116961	11,028	goshawk-Hungary/04	II
50	DQ318019	11,038	goshawk-Hungary/04	II
51	M12294	10,962	goshawk-Hungary/04	II
52	EF429197	11,052	goshawk-Hungary/04	II
53	EF429200	11,051	goshawk-Hungary/04	II
54	AY532665	11,038	goshawk-Hungary/04	II
55	EF429198	11,052	goshawk-Hungary/04	II
56	AY688948	11,057	goshawk-Hungary/04	II
57	DQ318020	11,048	goshawk-Hungary/04	II
58	EF429199	11,052	goshawk-Hungary/04	II
59	AY765264	10,972	97-103	III
60	AY277251	10,741	LEIV-Krnd88-190	IV
61	FJ159129	10,845	101.5-06-Uu	IV

<sup>a</sup> Lineage Ic is putatively known as lineage V.

## 2. True positive dataset.

Sr.no.	Accession no.	Strain\isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
1	AY274505	–	lb	lb	lb
2	D00246	MRM61C	lb	lb	lb
3	JX276662	CH16532	lb	lb	lb
4	JX503084	WNV/USA/BID-G15493	la	la4	la4
5	JX503092	WNV/USA/BID-G15501	la	la4	la4
6	HM147824	–	II	II	II
7	HM147823	–	II	II	II
8	HM147822	–	II	II	II
9	AY646354	–	la	la4	la4
10	AB185914	NY99-6922	la	la4	la4
11	AB185915	NY99-6922	la	la4	la4
12	AB185917	BC787	la	la4	la4
13	AB185916	BC787	la	la4	la4
14	GQ507468	007WG-TX05EP	la	la4	la4
15	GQ507469	009WG-NM05LC	la	la4	la4
16	GQ507470	011WG-TX06EP	la	la4	la4
17	GQ507471	013WG-TX07EP	la	la4	la4
18	GQ507472	024WG-CA03OR	la	la4	la4
19	DQ431696	03-104WI	la	la4	la4
20	DQ431697	03-113FL	la	la4	la4
21	DQ431698	03-120FL	la	la4	la4
22	DQ431699	03-124FL	la	la4	la4
23	DQ431693	03-20TX	la	la4	la4
24	DQ431694	03-22TX	la	la4	la4
25	DQ431695	03-82IL	la	la4	la4
26	DQ431700	04-213CA	la	la4	la4
27	DQ431701	04-214CO	la	la4	la4
28	DQ431702	04-216CO	la	la4	la4
29	DQ431703	04-218CO	la	la4	la4
30	DQ431704	04-219CO	la	la4	la4
31	DQ431705	04-233ND	la	la4	la4
32	DQ431706	04-236NM	la	la4	la4
33	DQ431707	04-237NM	la	la4	la4
34	DQ431708	04-238CA	la	la4	la4
35	DQ431709	04-240CA	la	la4	la4
36	DQ431710	04-244CA	la	la4	la4
37	DQ431711	04-251AZ	la	la4	la4
38	DQ431712	04-252AZ	la	la4	la4
39	GQ507473	080WG-CA04LA	la	la4	la4
40	GQ507474	091WG-CA04SB	la	la4	la4
41	GQ507475	099WG-CA05SB	la	la4	la4
42	GQ507476	101WG-CA05SB	la	la4	la4
43	GQ507477	103WG-CA05LA	la	la4	la4
44	GQ507478	116WG-CA05LA	la	la4	la4
45	GQ507479	124WG-AZ05PI	la	la4	la4
46	GQ507480	132WG-CA05LA	la	la4	la4
47	GQ507481	142WG-NE06DO	la	la4	la4
48	GQ507482	144WG-AZ06PI	la	la4	la4
49	GQ507483	148WG-CA07LA	la	la4	la4
50	GQ507484	149WG-CA07LA	la	la4	la4
51	AF206518	Connecticut 1999	la	la4	la4
52	FJ159130	5_50-05-Uu	IV	IV	IV
53	FJ159131	8_1-05-Uu	IV	IV	IV
54	DQ080051	A-AZ-03-1623	la	la4	la4
55	AY795965	ARC10	la	la4	la4
56	DQ164201	AZ 2004	la	la4	la4
57	DQ080052	B-AZ-03-1681	la	la4	la4
58	DQ666451	BSL13-2005	la	la4	la4
59	DQ666452	BSL2-2005	la	la4	la4
60	DQ666448	BSL5-2004	la	la4	la4
61	AY712945	Bird 1153	la	la4	la4
62	AY712946	Bird 1171	la	la4	la4
63	AY712947	Bird 1461	la	la4	la4
64	GU827998	Bird114	la	la4	la4
65	GU828000	Bird1175	la	la4	la4
66	GU828004	Bird1519	la	la4	la4
67	GU827999	Bird1576	la	la4	la4
68	GU828003	Bird1881	la	la4	la4
69	DQ080053	C-AZ-03 03-1799	la	la4	la4
70	JF703162	CA-03 COAV997	la	la4	la4
71	JF703164	CA-03 IMPR116	la	la4	la4
72	JF703161	CA-04 COAV689	la	la4	la4
73	JF703163	CA-05 COAV2900	la	la4	la4
74	DQ164204	1	la	la4	la4
75	DQ164203	2	la	la4	la4
76	JN716371	COL524/08	la	la4	la4

Sr.no.	Accession no.	Strain\isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
77	JN716372	COL9835/08	la	la4	la4
78	DQ080054	E-CA-03 GRLA-1260	la	la4	la4
79	DQ080055	F-CA-03 IMPR 102	la	la4	la4
80	DQ005530	FDA-BSL5-2003	la	la4	la4
81	DQ080072	FL232	la	la4	la4
82	DQ080071	FL234	la	la4	la4
83	DQ080056	G-CA-03 IMPR-1075	la	la4	la4
84	DQ164196	1	la	la4	la4
85	DQ164197	2	la	la4	la4
86	DQ666449	GCTX1-2005	la	la4	la4
87	DQ666450	GCTX2-2005	la	la4	la4
88	DQ080057	I-CA-03 Arcadia-S0331532	la	la4	la4
89	DQ164200	IN 2002	la	la4	la4
90	JF719065	Italy/2008/J-242853	la	la2	la2
91	JF719066	Italy/2008/M-203204	la	la2	la2
92	JF719067	Italy/2009/G-223184	la	la2	la2
93	JF719068	Italy/2009/J-225677	la	la2	la2
94	JN858069	Italy/2011/AN-1	la	la2	la2
95	JN858070	Italy/2011/AN-2	II	II	II
96	DQ080058	J-CA-03 Arcadia-S0334814	la	la4	la4
97	DQ080059	L-CA-04 SAC-04-7168	la	la4	la4
98	JF415914	M12214	la	la4	la4
99	JF415919	M19433	la	la4	la4
100	JF415928	M20122	la	la4	la4
101	JF415926	M20140	la	la4	la4
102	JF415927	M20141	la	la4	la4
103	JF415922	M37012	la	la4	la4
104	JF415923	M37906	la	la4	la4
105	JF415925	M38488	la	la4	la4
106	JF415930	M6019	la	la4	la4
107	AY660002	TM171-03	la	la4	la4
108	AY712948	Mosquito v4369	la	la4	la4
109	JN887352	NSW2011	lb	lb	lb
110	DQ164194	NY 2001 Suffolk	la	la4	la4
111	DQ164187	NY 2002 Broome	la	la4	la4
112	DQ164193	NY 2002 Clinton	la	la4	la4
113	DQ164195	NY 2002 Nassau	la	la4	la4
114	DQ164186	NY 2002 Queens	la	la4	la4
115	DQ164189	NY 2003 Albany	la	la4	la4
116	DQ164191	NY 2003 Chautauqua	la	la4	la4
117	DQ164192	NY 2003 Rockland	la	la4	la4
118	DQ164190	NY 2003 Suffolk	la	la4	la4
119	DQ164188	NY 2003 Westchester	la	la4	la4
120	FJ411043	NY99-6922	la	la4	la4
121	HQ537483	Nea Santa-Greece-2010	II	II	II
122	DQ164202	OH 2002	la	la4	la4
123	GQ903680	Q3574-5	II	II	II
124	FJ425721	Reb.VLG.07.H	II	II	II
125	EU068667	SPU116-89	II	II	II
126	JF719069	Spain/2010/H-1b	la	la2	la2
127	DQ080070	TVP9115	la	la4	la4
128	DQ080069	TVP9117	la	la4	la4
129	DQ080068	TVP9218	la	la4	la4
130	DQ080067	TVP9219	la	la4	la4
131	DQ080066	TVP9220	la	la4	la4
132	DQ080065	TVP9221	la	la4	la4
133	DQ080064	TVP9222	la	la4	la4
134	DQ080063	TVP9223	la	la4	la4
135	DQ080062	TWN165	la	la4	la4
136	DQ080061	TWN496	la	la4	la4
137	DQ164198	1	la	la4	la4
138	DQ164205	2	la	la4	la4
139	DQ164199	TX 2003	la	la4	la4
140	DQ164206	4	la	la4	la4
141	JF415929	TX5058	la	la4	la4
142	JF415915	TX5810	la	la4	la4
143	JF415916	TX6276	la	la4	la4
144	JF415917	TX6647	la	la4	la4
145	JF415918	TX6747	la	la4	la4
146	JF415920	TX7191	la	la4	la4
147	JF415921	TX7558	la	la4	la4
148	JF415924	TX7827	la	la4	la4
149	AF404753	WN MD 2000-crow265	la	la4	la4
150	AF404754	WN NJ 2000 MQ5488	la	la4	la4
151	AF404756	WN NY 2000-crow3356	la	la4	la4
152	AF404755	WN NY 2000-grouse3282	la	la4	la4
153	JN819305	WNV-1/BID-V5030	la	la4	la4



Sr.no.	Accession no.	Strain/isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
154	JN819306	WNV-1/BID-V5031	la	la4	la4
155	JN819307	WNV-1/BID-V5032	la	la4	la4
156	JN819309	WNV-1/BID-V5036	la	la4	la4
157	JN819310	WNV-1/BID-V5037	la	la4	la4
158	JN819311	WNV-1/BID-V5038	la	la4	la4
159	JN819312	WNV-1/BID-V5039	la	la4	la4
160	JN819313	WNV-1/BID-V5040	la	la4	la4
161	JN819315	WNV-1/BID-V5042	la	la4	la4
162	JN819316	WNV-1/BID-V5043	la	la4	la4
163	JN819317	WNV-1/BID-V5044	la	la4	la4
164	JN819318	WNV-1/BID-V5045	la	la4	la4
165	JN819319	WNV-1/BID-V5047	la	la4	la4
166	JN819320	WNV-1/BID-V5048	la	la4	la4
167	JN819323	WNV-1/BID-V5050	la	la4	la4
168	JN819321	WNV-1/BID-V5051	la	la4	la4
169	JN819324	WNV-1/BID-V5052	la	la4	la4
170	HM488122	WNV-1/CTFS/BID-V4111/2006	la	la4	la4
171	HM488123	WNV-1/CTFS/BID-V4112/2006	la	la4	la4
172	HM488124	WNV-1/CTFS/BID-V4113/2006	la	la4	la4
173	HQ671687	WNV-1/Culex/BID-V4161/legs	la	la4	la4
174	HQ705670	WNV-1/Culex/BID-V4163/legs	la	la4	la4
175	HQ671688	WNV-1/Culex/BID-V4165/legs	la	la4	la4
176	HQ705671	WNV-1/Culex/BID-V4166/midgut	la	la4	la4
177	HQ705672	WNV-1/Culex/BID-V4168/legs	la	la4	la4
178	HQ705673	WNV-1/Culex/BID-V4169/midgut	la	la4	la4
179	HQ671731	WNV-1/Culex/BID-V4171/midgut	la	la4	la4
180	HQ705674	WNV-1/Culex/BID-V4173/midgut	la	la4	la4
181	HQ671689	WNV-1/Culex/BID-V4174/legs	la	la4	la4
182	HQ671690	WNV-1/Culex/BID-V4180/midgut	la	la4	la4
183	HQ705675	WNV-1/Culex/BID-V4184/legs	la	la4	la4
184	JN183895	WNV-1/Culex/BID-V5768/midgut	la	la4	la4
185	JN183896	WNV-1/Culex/BID-V5776/midgut	la	la4	la4
186	JN183897	WNV-1/Culex/BID-V5808/midgut	la	la4	la4
187	JN183893	WNV-1/Gallus/BID-V4954/kidney	la	la4	la4
188	JN183894	WNV-1/Gallus/BID-V4955/kidney	la	la4	la4
189	HQ671691	WNV-1/Gallus/BID-V4958/kidney	la	la4	la4
190	HQ671692	WNV-1/Gallus/BID-V4959/kidney	la	la4	la4
191	HQ705677	WNV-1/Gallus/BID-V4960/kidney	la	la4	la4
192	HQ671693	WNV-1/Gallus/BID-V4961/kidney	la	la4	la4
193	HQ705678	WNV-1/Gallus/BID-V4962/kidney	la	la4	la4
194	HQ671694	WNV-1/Gallus/BID-V4963/kidney	la	la4	la4
195	JF357960	WNV-1/Gallus/BID-V5109/skin	la	la4	la4
196	HQ671695	WNV-1/Gallus/BID-V5112/skin	la	la4	la4
197	HQ671668	WNV-1/Mus/BID-V4728/spleen	la	la4	la4
198	JF730041	WNV-1/Mus/BID-V4729/spleen	la	la4	la4
199	HQ671669	WNV-1/Mus/BID-V4730/spleen	la	la4	la4
200	HQ671670	WNV-1/Mus/BID-V4731/spleen	la	la4	la4
201	HQ671671	WNV-1/Mus/BID-V4920/spleen	la	la4	la4
202	HQ671672	WNV-1/Mus/BID-V4921/spleen	la	la4	la4
203	JF899530	WNV-1/Mus/BID-V4923/spleen	la	la4	la4
204	HQ671673	WNV-1/Mus/BID-V4924/spleen	la	la4	la4
205	HQ671674	WNV-1/Mus/BID-V4925/spleen	la	la4	la4
206	HQ671675	WNV-1/Mus/BID-V4926/spleen	la	la4	la4
207	HQ705663	WNV-1/Mus/BID-V4936/spleen	la	la4	la4
208	HQ671676	WNV-1/Mus/BID-V4937/spleen	la	la4	la4
209	HQ671677	WNV-1/Mus/BID-V4938/spleen	la	la4	la4
210	HQ671678	WNV-1/Mus/BID-V4940/spleen	la	la4	la4
211	HQ671679	WNV-1/Mus/BID-V4969/serum	la	la4	la4
212	HQ705676	WNV-1/Mus/BID-V4971/serum	la	la4	la4
213	JF784158	WNV-1/Mus/BID-V4974/serum	la	la4	la4
214	JF899531	WNV-1/Mus/BID-V4979/serum	la	la4	la4
215	HQ891009	WNV-1/Mus/BID-V4980/serum	la	la4	la4
216	HQ671680	WNV-1/Mus/BID-V4981/serum	la	la4	la4
217	JF730040	WNV-1/Mus/BID-V4982/serum	la	la4	la4
218	HQ671681	WNV-1/Mus/BID-V4985/spleen	la	la4	la4
219	HQ671682	WNV-1/Mus/BID-V4986/spleen	la	la4	la4
220	HQ891010	WNV-1/Mus/BID-V4987/spleen	la	la4	la4
221	JF899532	WNV-1/Mus/BID-V4992/brain	la	la4	la4
222	HQ891011	WNV-1/Mus/BID-V4993/brain	la	la4	la4
223	JF899533	WNV-1/Mus/BID-V4996/brain	la	la4	la4
224	JF357958	WNV-1/Mus/BID-V4997/brain	la	la4	la4
225	JF357959	WNV-1/Mus/BID-V4999/brain	la	la4	la4
226	HQ671732	WNV-1/Mus/BID-V5000/brain	la	la4	la4
227	JF899534	WNV-1/Mus/BID-V5002/brain	la	la4	la4
228	HQ671733	WNV-1/Mus/BID-V5004/brain	la	la4	la4
229	HQ671683	WNV-1/Mus/BID-V5006/brain	la	la4	la4
230	HQ671684	WNV-1/Mus/BID-V5007/brain	la	la4	la4

Sr.no.	Accession no.	Strain\isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
231	HQ671685	WNV-1/Mus/BID-V5008/brain	la	la4	la4
232	HQ671686	WNV-1/Mus/BID-V5017/brain	la	la4	la4
233	HQ891012	WNV-1/Mus/BID-V5118/serum	la	la4	la4
234	HQ891013	WNV-1/Mus/BID-V5122/serum	la	la4	la4
235	JF899535	WNV-1/Mus/BID-V5142/brain	la	la4	la4
236	JF899536	WNV-1/Mus/BID-V5145/brain	la	la4	la4
237	JF899537	WNV-1/Mus/BID-V5146/brain	la	la4	la4
238	HM488199	WNV-1/US/BID-V4090/2007	la	la4	la4
239	HM488200	WNV-1/US/BID-V4092/2007	la	la4	la4
240	HM488201	WNV-1/US/BID-V4093/2007	la	la4	la4
241	HM488202	WNV-1/US/BID-V4094/2007	la	la4	la4
242	HM756678	WNV-1/US/BID-V4095/2007	la	la4	la4
243	HM488203	WNV-1/US/BID-V4096/2008	la	la4	la4
244	HM756660	WNV-1/US/BID-V4097/2008	la	la4	la4
245	HM488204	WNV-1/US/BID-V4098/2008	la	la4	la4
246	HM488205	WNV-1/US/BID-V4099/2008	la	la4	la4
247	HM488206	WNV-1/US/BID-V4100/2008	la	la4	la4
248	HM488207	WNV-1/US/BID-V4101/2008	la	la4	la4
249	HM488114	WNV-1/US/BID-V4102/2002	la	la4	la4
250	HM488115	WNV-1/US/BID-V4103/2005	la	la4	la4
251	HM488116	WNV-1/US/BID-V4104/2005	la	la4	la4
252	HM488117	WNV-1/US/BID-V4105/2005	la	la4	la4
253	HM488118	WNV-1/US/BID-V4107/2005	la	la4	la4
254	HM488119	WNV-1/US/BID-V4108/2005	la	la4	la4
255	HM488120	WNV-1/US/BID-V4109/2005	la	la4	la4
256	HM488121	WNV-1/US/BID-V4110/2005	la	la4	la4
257	HM488125	WNV-1/US/BID-V4186/1999	la	la4	la4
258	HM488126	WNV-1/US/BID-V4187/1999	la	la4	la4
259	HM488127	WNV-1/US/BID-V4188/1999	la	la4	la4
260	HM488128	WNV-1/US/BID-V4189/1999	la	la4	la4
261	HM488129	WNV-1/US/BID-V4191/2000	la	la4	la4
262	HM488130	WNV-1/US/BID-V4192/2000	la	la4	la4
263	HM488131	WNV-1/US/BID-V4193/2000	la	la4	la4
264	HM488132	WNV-1/US/BID-V4194/2000	la	la4	la4
265	HM488133	WNV-1/US/BID-V4195/2001	la	la4	la4
266	HQ671696	WNV-1/US/BID-V4196/2001	la	la4	la4
267	HQ671697	WNV-1/US/BID-V4197/2001	la	la4	la4
268	HM488134	WNV-1/US/BID-V4198/2001	la	la4	la4
269	HM488135	WNV-1/US/BID-V4199/2001	la	la4	la4
270	HM488136	WNV-1/US/BID-V4200/2001	la	la4	la4
271	HM488137	WNV-1/US/BID-V4202/2002	la	la4	la4
272	HQ671698	WNV-1/US/BID-V4203/2002	la	la4	la4
273	HM488208	WNV-1/US/BID-V4204/2002	la	la4	la4
274	HM756648	WNV-1/US/BID-V4205/2002	la	la4	la4
275	HQ671699	WNV-1/US/BID-V4206/2002	la	la4	la4
276	HM488138	WNV-1/US/BID-V4207/2003	la	la4	la4
277	HM488139	WNV-1/US/BID-V4208/2003	la	la4	la4
278	HQ705659	WNV-1/US/BID-V4209/2003	la	la4	la4
279	HM488140	WNV-1/US/BID-V4210/2003	la	la4	la4
280	HM488141	WNV-1/US/BID-V4212/2003	la	la4	la4
281	HM488142	WNV-1/US/BID-V4214/2004	la	la4	la4
282	HM488143	WNV-1/US/BID-V4215/2004	la	la4	la4
283	HM488144	WNV-1/US/BID-V4216/2004	la	la4	la4
284	HM488145	WNV-1/US/BID-V4217/2004	la	la4	la4
285	HM488146	WNV-1/US/BID-V4218/2004	la	la4	la4
286	HM488147	WNV-1/US/BID-V4219/2004	la	la4	la4
287	HM488148	WNV-1/US/BID-V4220/2004	la	la4	la4
288	HM488149	WNV-1/US/BID-V4223/2005	la	la4	la4
289	HM488150	WNV-1/US/BID-V4224/2005	la	la4	la4
290	HM488151	WNV-1/US/BID-V4225/2005	la	la4	la4
291	HM488152	WNV-1/US/BID-V4226/2005	la	la4	la4
292	HM488153	WNV-1/US/BID-V4227/2005	la	la4	la4
293	HM488154	WNV-1/US/BID-V4228/2005	la	la4	la4
294	HM488155	WNV-1/US/BID-V4229/2006	la	la4	la4
295	HM488156	WNV-1/US/BID-V4230/2006	la	la4	la4
296	HM488157	WNV-1/US/BID-V4231/2006	la	la4	la4
297	HM488158	WNV-1/US/BID-V4232/2006	la	la4	la4
298	HM488159	WNV-1/US/BID-V4233/2006	la	la4	la4
299	HM488177	WNV-1/US/BID-V4336/2002	la	la4	la4
300	HM488178	WNV-1/US/BID-V4337/2002	la	la4	la4
301	HM488179	WNV-1/US/BID-V4338/2002	la	la4	la4
302	HM488180	WNV-1/US/BID-V4339/2002	la	la4	la4
303	HM488181	WNV-1/US/BID-V4340/2002	la	la4	la4
304	HM488182	WNV-1/US/BID-V4341/2002	la	la4	la4
305	HQ705669	WNV-1/US/BID-V4342/2002	la	la4	la4
306	HQ671742	WNV-1/US/BID-V4343/2002	la	la4	la4
307	JN183891	WNV-1/US/BID-V4344/2002	la	la4	la4

Sr.no.	Accession no.	Strain isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
308	HM488183	WNV-1/US/BID-V4345/2002	la	la4	la4
309	HM488184	WNV-1/US/BID-V4346/2002	la	la4	la4
310	HM488185	WNV-1/US/BID-V4347/2003	la	la4	la4
311	HM756676	WNV-1/US/BID-V4349/2003	la	la4	la4
312	HM488186	WNV-1/US/BID-V4350/2003	la	la4	la4
313	HM488187	WNV-1/US/BID-V4351/2003	la	la4	la4
314	HM488188	WNV-1/US/BID-V4353/2004	la	la4	la4
315	HM756649	WNV-1/US/BID-V4354/2006	la	la4	la4
316	HM488160	WNV-1/US/BID-V4355/2006	la	la4	la4
317	HM488161	WNV-1/US/BID-V4356/2007	la	la4	la4
318	HM488162	WNV-1/US/BID-V4357/2007	la	la4	la4
319	HM488163	WNV-1/US/BID-V4359/2007	la	la4	la4
320	HM488164	WNV-1/US/BID-V4360/2007	la	la4	la4
321	HM488165	WNV-1/US/BID-V4361/2007	la	la4	la4
322	HM488166	WNV-1/US/BID-V4362/2008	la	la4	la4
323	HM488167	WNV-1/US/BID-V4363/2008	la	la4	la4
324	HM488168	WNV-1/US/BID-V4364/2008	la	la4	la4
325	HM488169	WNV-1/US/BID-V4365/2008	la	la4	la4
326	HM488170	WNV-1/US/BID-V4366/2008	la	la4	la4
327	HM488189	WNV-1/US/BID-V4367/2004	la	la4	la4
328	HM488190	WNV-1/US/BID-V4368/2004	la	la4	la4
329	HM488191	WNV-1/US/BID-V4369/2004	la	la4	la4
330	HM488192	WNV-1/US/BID-V4371/2005	la	la4	la4
331	HM488193	WNV-1/US/BID-V4373/2005	la	la4	la4
332	HM488194	WNV-1/US/BID-V4374/2005	la	la4	la4
333	HM488195	WNV-1/US/BID-V4375/2005	la	la4	la4
334	HM488196	WNV-1/US/BID-V4376/2005	la	la4	la4
335	HM488197	WNV-1/US/BID-V4377/2005	la	la4	la4
336	HM488198	WNV-1/US/BID-V4378/2005	la	la4	la4
337	JN183892	WNV-1/US/BID-V4379/2005	la	la4	la4
338	HM756677	WNV-1/US/BID-V4530/2005	la	la4	la4
339	HM488253	WNV-1/US/BID-V4553/2006	la	la4	la4
340	HM488254	WNV-1/US/BID-V4559/2007	la	la4	la4
341	HM488171	WNV-1/US/BID-V4560/2003	la	la4	la4
342	HM488172	WNV-1/US/BID-V4561/2003	la	la4	la4
343	HM488173	WNV-1/US/BID-V4562/2003	la	la4	la4
344	HM488174	WNV-1/US/BID-V4563/2003	la	la4	la4
345	HM488209	WNV-1/US/BID-V4564/2003	la	la4	la4
346	HM488210	WNV-1/US/BID-V4565/2003	la	la4	la4
347	HM488211	WNV-1/US/BID-V4566/2003	la	la4	la4
348	HM488212	WNV-1/US/BID-V4567/2003	la	la4	la4
349	JF920728	WNV-1/US/BID-V4568/2003	la	la4	la4
350	HM488175	WNV-1/US/BID-V4569/2003	la	la4	la4
351	HM488213	WNV-1/US/BID-V4571/2003	la	la4	la4
352	HM488214	WNV-1/US/BID-V4572/2003	la	la4	la4
353	HM488215	WNV-1/US/BID-V4573/2003	la	la4	la4
354	HM488216	WNV-1/US/BID-V4574/2003	la	la4	la4
355	HM488176	WNV-1/US/BID-V4575/2003	la	la4	la4
356	HQ671700	WNV-1/US/BID-V4576/2003	la	la4	la4
357	HM488217	WNV-1/US/BID-V4581/2003	la	la4	la4
358	HM756650	WNV-1/US/BID-V4582/2003	la	la4	la4
359	HM488218	WNV-1/US/BID-V4583/2003	la	la4	la4
360	HM756651	WNV-1/US/BID-V4584/2003	la	la4	la4
361	HM488219	WNV-1/US/BID-V4585/2003	la	la4	la4
362	HM488220	WNV-1/US/BID-V4586/2003	la	la4	la4
363	HM756652	WNV-1/US/BID-V4587/2003	la	la4	la4
364	HM756653	WNV-1/US/BID-V4588/2003	la	la4	la4
365	HQ671701	WNV-1/US/BID-V4590/2003	la	la4	la4
366	HM488221	WNV-1/US/BID-V4593/2003	la	la4	la4
367	HQ671702	WNV-1/US/BID-V4595/2003	la	la4	la4
368	JF920306	WNV-1/US/BID-V4597/2003	la	la4	la4
369	HM756654	WNV-1/US/BID-V4598/2003	la	la4	la4
370	HM488222	WNV-1/US/BID-V4599/2003	la	la4	la4
371	HM488223	WNV-1/US/BID-V4603/2003	la	la4	la4
372	HM488224	WNV-1/US/BID-V4604/2003	la	la4	la4
373	HM488225	WNV-1/US/BID-V4605/2003	la	la4	la4
374	HM488226	WNV-1/US/BID-V4607/2003	la	la4	la4
375	HM488227	WNV-1/US/BID-V4608/2003	la	la4	la4
376	HM488228	WNV-1/US/BID-V4609/2003	la	la4	la4
377	HM488229	WNV-1/US/BID-V4610/2003	la	la4	la4
378	HQ671703	WNV-1/US/BID-V4611/2003	la	la4	la4
379	HM488230	WNV-1/US/BID-V4612/2003	la	la4	la4
380	HM488231	WNV-1/US/BID-V4613/2003	la	la4	la4
381	HM488232	WNV-1/US/BID-V4614/2003	la	la4	la4
382	HM756656	WNV-1/US/BID-V4615/2003	la	la4	la4
383	HM488233	WNV-1/US/BID-V4616/2003	la	la4	la4
384	HM488234	WNV-1/US/BID-V4617/2003	la	la4	la4

Sr.no.	Accession no.	Strain\isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
385	HQ671704	WNV-1/US/BID-V4618/2003	la	la4	la4
386	HM488235	WNV-1/US/BID-V4619/2003	la	la4	la4
387	HQ671705	WNV-1/US/BID-V4620/2003	la	la4	la4
388	HM488237	WNV-1/US/BID-V4622/2008	la	la4	la4
389	HM488238	WNV-1/US/BID-V4623/2008	la	la4	la4
390	HM488239	WNV-1/US/BID-V4624/2008	la	la4	la4
391	HQ671721	WNV-1/US/BID-V4625/2008	la	la4	la4
392	JN183885	WNV-1/US/BID-V4626/2008	la	la4	la4
393	HM488240	WNV-1/US/BID-V4627/2008	la	la4	la4
394	HM488241	WNV-1/US/BID-V4628/2008	la	la4	la4
395	JN183886	WNV-1/US/BID-V4629/2008	la	la4	la4
396	HM488242	WNV-1/US/BID-V4631/2008	la	la4	la4
397	HM488243	WNV-1/US/BID-V4632/2008	la	la4	la4
398	HM488244	WNV-1/US/BID-V4634/2008	la	la4	la4
399	HM488245	WNV-1/US/BID-V4635/2008	la	la4	la4
400	HM756657	WNV-1/US/BID-V4685/2003	la	la4	la4
401	HM756658	WNV-1/US/BID-V4686/2003	la	la4	la4
402	HM756659	WNV-1/US/BID-V4687/2003	la	la4	la4
403	HM488246	WNV-1/US/BID-V4689/2001	la	la4	la4
404	HM488247	WNV-1/US/BID-V4691/2001	la	la4	la4
405	HM756661	WNV-1/US/BID-V4692/2001	la	la4	la4
406	HM756662	WNV-1/US/BID-V4693/2001	la	la4	la4
407	HM488248	WNV-1/US/BID-V4694/2001	la	la4	la4
408	HM488249	WNV-1/US/BID-V4696/2001	la	la4	la4
409	HM756663	WNV-1/US/BID-V4697/2001	la	la4	la4
410	JN183890	WNV-1/US/BID-V4699/2003	la	la4	la4
411	HM488236	WNV-1/US/BID-V4700/2003	la	la4	la4
412	HM756664	WNV-1/US/BID-V4701/2002	la	la4	la4
413	HQ671722	WNV-1/US/BID-V4704/2002	la	la4	la4
414	JN183887	WNV-1/US/BID-V4706/2002	la	la4	la4
415	HM756665	WNV-1/US/BID-V4709/2002	la	la4	la4
416	HM756666	WNV-1/US/BID-V4711/2003	la	la4	la4
417	HM756667	WNV-1/US/BID-V4712/2003	la	la4	la4
418	HQ705660	WNV-1/US/BID-V4714/2003	la	la4	la4
419	HQ671723	WNV-1/US/BID-V4715/2003	la	la4	la4
420	HM756668	WNV-1/US/BID-V4716/2003	la	la4	la4
421	HM488250	WNV-1/US/BID-V4717/2003	la	la4	la4
422	HM756669	WNV-1/US/BID-V4718/2003	la	la4	la4
423	HM488251	WNV-1/US/BID-V4719/2003	la	la4	la4
424	HM756670	WNV-1/US/BID-V4720/2003	la	la4	la4
425	HM756671	WNV-1/US/BID-V4798/2004	la	la4	la4
426	HM756672	WNV-1/US/BID-V4799/2004	la	la4	la4
427	JF899528	WNV-1/US/BID-V4800/2004	la	la4	la4
428	HM756673	WNV-1/US/BID-V4801/2004	la	la4	la4
429	JN367277	WNV-1/US/BID-V4803/2004	la	la4	la4
430	HM488252	WNV-1/US/BID-V4805/2005	la	la4	la4
431	HM756675	WNV-1/US/BID-V4806/2005	la	la4	la4
432	JF899529	WNV-1/US/BID-V4808/2005	la	la4	la4
433	HQ671724	WNV-1/US/BID-V4883/2005	la	la4	la4
434	HQ671725	WNV-1/US/BID-V4885/2005	la	la4	la4
435	HQ671726	WNV-1/US/BID-V4887/2005	la	la4	la4
436	HQ671727	WNV-1/US/BID-V4889/2006	la	la4	la4
437	HQ671728	WNV-1/US/BID-V4891/2006	la	la4	la4
438	HQ671729	WNV-1/US/BID-V4892/2006	la	la4	la4
439	JN183888	WNV-1/US/BID-V4896/2006	la	la4	la4
440	HQ671730	WNV-1/US/BID-V4897/2007	la	la4	la4
441	HQ671706	WNV-1/US/BID-V4898/1999	la	la4	la4
442	HQ671707	WNV-1/US/BID-V4899/1999	la	la4	la4
443	HQ671708	WNV-1/US/BID-V4900/2000	la	la4	la4
444	HQ671709	WNV-1/US/BID-V4901/2000	la	la4	la4
445	HQ671710	WNV-1/US/BID-V4902/2000	la	la4	la4
446	HQ671711	WNV-1/US/BID-V4903/2000	la	la4	la4
447	HQ671712	WNV-1/US/BID-V4904/2000	la	la4	la4
448	HQ671713	WNV-1/US/BID-V4905/2001	la	la4	la4
449	HQ671714	WNV-1/US/BID-V4906/2001	la	la4	la4
450	JF920307	WNV-1/US/BID-V4907/2001	la	la4	la4
451	HQ671715	WNV-1/US/BID-V4908/2001	la	la4	la4
452	HQ671716	WNV-1/US/BID-V4909/2001	la	la4	la4
453	HQ671717	WNV-1/US/BID-V4910/2001	la	la4	la4
454	HQ671718	WNV-1/US/BID-V4911/2001	la	la4	la4
455	HQ671719	WNV-1/US/BID-V4912/2001	la	la4	la4
456	HQ671720	WNV-1/US/BID-V4913/2002	la	la4	la4
457	JF730042	WNV-1/US/BID-V5147/2007	la	la4	la4
458	JF488097	WNV-1/US/BID-V5148/2007	la	la4	la4
459	JF488094	WNV-1/US/BID-V5150/2004	la	la4	la4
460	JF488095	WNV-1/US/BID-V5157/2009	la	la4	la4
461	JF488096	WNV-1/US/BID-V5159/2009	la	la4	la4

Sr.no.	Accession no.	Strain/isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
462	JF730043	WNV-1/US/BID-V5170/2002	la	la4	la4
463	JF488086	WNV-1/US/BID-V5176/2004	la	la4	la4
464	JF488087	WNV-1/US/BID-V5177/2004	la	la4	la4
465	JF488088	WNV-1/US/BID-V5178/2004	la	la4	la4
466	JF488089	WNV-1/US/BID-V5179/2004	la	la4	la4
467	JF488090	WNV-1/US/BID-V5180/2004	la	la4	la4
468	JF488091	WNV-1/US/BID-V5181/2004	la	la4	la4
469	JF488092	WNV-1/US/BID-V5182/2004	la	la4	la4
470	JF488093	WNV-1/US/BID-V5188/2005	la	la4	la4
471	JF920729	WNV-1/US/BID-V5196/2006	la	la4	la4
472	JF920730	WNV-1/US/BID-V5197/2006	la	la4	la4
473	JF920731	WNV-1/US/BID-V5201/2006	la	la4	la4
474	JF920732	WNV-1/US/BID-V5202/2006	la	la4	la4
475	JF920733	WNV-1/US/BID-V5203/2006	la	la4	la4
476	JF920734	WNV-1/US/BID-V5204/2006	la	la4	la4
477	JF920735	WNV-1/US/BID-V5205/2006	la	la4	la4
478	JF920736	WNV-1/US/BID-V5206/2006	la	la4	la4
479	JF920737	WNV-1/US/BID-V5207/2006	la	la4	la4
480	JF920738	WNV-1/US/BID-V5208/2007	la	la4	la4
481	JF920739	WNV-1/US/BID-V5209/2007	la	la4	la4
482	JF920740	WNV-1/US/BID-V5210/2007	la	la4	la4
483	JF920741	WNV-1/US/BID-V5212/2007	la	la4	la4
484	JF920742	WNV-1/US/BID-V5213/2007	la	la4	la4
485	JF920743	WNV-1/US/BID-V5214/2007	la	la4	la4
486	JF920744	WNV-1/US/BID-V5215/2007	la	la4	la4
487	JF920745	WNV-1/US/BID-V5216/2007	la	la4	la4
488	JF920746	WNV-1/US/BID-V5217/2007	la	la4	la4
489	JF920747	WNV-1/US/BID-V5218/2008	la	la4	la4
490	JF920748	WNV-1/US/BID-V5219/2008	la	la4	la4
491	JF920749	WNV-1/US/BID-V5220/2008	la	la4	la4
492	JF920750	WNV-1/US/BID-V5222/2008	la	la4	la4
493	JF920751	WNV-1/US/BID-V5223/2008	la	la4	la4
494	JF920752	WNV-1/US/BID-V5224/2008	la	la4	la4
495	JF920753	WNV-1/US/BID-V5225/2008	la	la4	la4
496	JF920754	WNV-1/US/BID-V5226/2008	la	la4	la4
497	JF920755	WNV-1/US/BID-V5227/2008	la	la4	la4
498	JF920756	WNV-1/US/BID-V5228/2008	la	la4	la4
499	JF920757	WNV-1/US/BID-V5229/2008	la	la4	la4
500	JF920758	WNV-1/US/BID-V5230/2008	la	la4	la4
501	JF920759	WNV-1/US/BID-V5233/2009	la	la4	la4
502	JF920760	WNV-1/US/BID-V5234/2009	la	la4	la4
503	JF920760	WNV-1/US/BID-V5235/2009	la	la4	la4
504	DQ080060	WNVcc	la	la4	la4
505	HM152773	WNV_0043h_ISR00	la	la4	la4
506	HM152775	WNV_0304h_ISR00	la	la2	la2
507	HM051416	twN9	la	la1	la1
508	GU828002	v4095	la	la4	la4
509	GU828001	v4380	la	la4	la4
510	AF533540	-	la	la4	la4
511	EF530047	3356.2.1.1	la	la4	la4
512	EF657887	3356 K VP2	la	la4	la4
513	AY848695	385-99	la	la4	la4
514	DQ066423	385-99	la	la4	la4
515	AY848696	385-99	la	la4	la4
516	AY848697	385-99	la	la4	la4
517	AY842931	385-99	la	la4	la4
518	EF571854	385-99	la	la4	la4
519	GQ379160	ArEq001	la	la4	la4
520	GQ379161	ArEq003	la	la4	la4
521	JX041634	Ast-986	la	la3	la3
522	AY490240	Chin-01	la	la1	la1
523	GQ379157	DB080718-14	la	la4	la4
524	EU081844	Egypt 101	la	la1	la1
525	GQ379156	FL2001 crow 67030	la	la4	la4
526	DQ983578	FLO3-FL2-3	la	la4	la4
527	AF202541	HNY1999	la	la4	la4
528	JF707789	HU6365/08	la	la2	la2
529	AF481864	IS-98 STD1	la	la4	la4
530	JX041632	Ig2266	lc	lc	lc
531	JQ928174	Italy/2011/Livenza	la	la2	la2
532	JQ928175	Italy/2011/Piave	la	la2	la2
533	JX556213	Italy/2012/Livenza/31.1	la	la2	la2
534	GQ379159	JPW080813-01	la	la4	la4
535	JX041629	LEIV-1628Az	la	la1	la1
536	JX041630	LEIV-1640Az	la	la1	la1
537	JX041631	LEIV-3266Ukr	ll	ll	ll
538	JX041628	LEIV-72Az	la	la1	la1

Sr.no.	Accession no.	Strain \ isolate	Known lineage/clade as per WNVdb	Lineage/clade/cluster predicted by RTD based method	Lineage/clade/cluster assigned by alignment based Neighbor-joining phylogeny
539	FJ527738	LSU-AR01	la	la4	la4
540	DQ211652	NY99	la	la4	la4
541	FJ151394	NY99-crow-V76/1	la	la4	la4
542	AF260967	NY99-eqhs	la	la4	la4
543	AF196835	NY99-flamingo382-99	la	la4	la4
544	HQ596519	New York 99	la	la4	la4
545	EU155484	OK03	la	la4	la4
546	GQ379158	ORCO0559-07	la	la4	la4
547	JX070655	T-1304	la	la1	la1
548	JN051152	TM171-03	la	la4	la4
549	JN051153	TM171-03	la	la4	la4
550	AY289214	TVP 8533	la	la4	la4
551	DQ176637	TX 2002-HC	la	la4	la4
552	JX015522	TX_AR10-5718	la	la4	la4
553	JX015523	TX_AR10-6572	la	la4	la4
554	JX015515	TX_AR5-2686	la	la4	la4
555	JX015516	TX_AR7-6745	la	la4	la4
556	JX015517	TX_AR8-5947	la	la4	la4
557	JX015518	TX_AR8-6866	la	la4	la4
558	JX015519	TX_AR9-5282	la	la4	la4
559	JX015520	TX_AR9-6115	la	la4	la4
560	JX015521	TX_AR9-7465	la	la4	la4
561	JX123030	V11-03	lb	lb	lb
562	JX123031	V11-07	lb	lb	lb
563	KC601756	1048813	la	la3	la3
564	KC407673	Sad/12	ll	ll	ll
565	KC407667	ABB-B13	la	la4	la4
566	KC407666	NY-99	la	la4	la4

True negative dataset (Members of the family *Flaviviridae*).

Sr. no.	Accession no.	Species
1	U88537	Dengue virus 1
2	M19197	Dengue virus 2
3	AF317645	Dengue virus 3
4	AF326573	Dengue virus 4
5	AF331718	Alkhurma hemorrhagic fever virus
6	AF160193	Apoi virus
7	M91671	Cell fusing agent virus
8	AF311056	Deer tick virus
9	AY898809	Alfuy virus
10	M18370	Japanese encephalitis virus
11	AY149904	Kamiti River virus
12	AY863002	Karshi virus
13	AF253419	Langat virus
14	Y07863	Louping ill virus
15	AJ242984	Modoc virus
16	AJ299445	Montana myotis leukoencephalitis virus
17	AF161266	Murray Valley encephalitis virus
18	AY193805	Omsk hemorrhagic fever virus
19	L06436	Powassan virus
20	AF144692	Rio Bravo virus
21	AF285080	Tamana bat virus
22	U27495	Tick-borne encephalitis virus
23	AY453412	Usutu virus
24	X03700	Yellow fever virus
25	AB114858	Yokose virus
26	AF037405	Border disease virus-X818
27	AJ781045	Bovine viral diarrhea virus 1
28	AF002227	Bovine viral diarrhea virus 2-C413
29	AF326963	Classical swine fever virus strain Eystrup
30	AF144617	Pestivirus giraffe-1 H138
31	AF144618	Pestivirus reindeer-1 V60-Krefeld
32	AF070476	GB virus C variant troglodytes
33	U44402	GB virus C (GBV-HGV A)
34	U94421	GB virus A
35	U36380	GB virus C (GBV-HGV)
36	AY859526	Hepatitis C virus (isolate 6a33)
37	D84264	Hepatitis C virus (isolate VN405)
38	D63822	Hepatitis C virus (isolate JK046)
39	D84263	Hepatitis C virus (isolate VN235)

Sr. no.	Accession no.	Species
40	D84262	Hepatitis C virus (isolate Th580)
41	D84265	Hepatitis C virus (isolate VN004)
42	AB047639	Hepatitis C virus JFH-1
43	AB030907	Hepatitis C virus (isolate JPUT971017)
44	D50409	Hepatitis C virus (isolate BEBE1)
45	AB031663	Hepatitis C virus (isolate VAT96)
46	D17763	Hepatitis C virus (isolate NZL1)
47	D49374	Hepatitis C virus (isolate Tr Kj)
48	D63821	Hepatitis C virus (isolate JK049)
49	AF207767	Hepatitis C virus subtype 1b
50	D14853	Hepatitis C virus (isolate HC-G9)
51	AF009606	Hepatitis C virus subtype 1a
52	Y13184	Hepatitis C virus (isolate EUH1480)
53	Y11604	Hepatitis C virus (isolate ED43)
54	AF179612	Hepatitis GB virus B
55	EU159426	Nounane virus
56	GU566734	GB virus D strain 68
57	GQ165809	Nakiwogo virus strain Uganda08

## References

- Anderson, J.F., Vossbrinck, C.R., Andreadis, T.G., Iton, A., Beckwith, W.H., Mayo, D.R., 2001. A phylogenetic approach to following West Nile virus in Connecticut. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12885–12889.
- Armstrong, P.M., Vossbrinck, C.R., Andreadis, T.G., Anderson, J.F., Pesko, K.N., Newman, R.M., Lennon, N.J., Birren, B.W., Ebel, G.D., Henn, M.R., 2011. Molecular evolution of West Nile virus in a northern temperate region: Connecticut, USA 1999–2008. *Virology* 417, 203–210.
- Beasley, D.W., Li, L., Suderman, M.T., Barrett, A.D., 2002. Mouse neuroinvasive phenotype of West Nile virus strains varies depending upon virus genotype. *Virology* 296 (1), 17–23.
- Beasley, D.W.C., Davis, C.T., Guzman, H., Vanlandingham, D.L., Travassos da Rosa, A., Parsons, R.E., Higgs, S., Tesh, R.B., Barrett, A.D.T., 2003. Limited evolution of West Nile virus has occurred during its southwesterly spread in the United States. *Virology* 309, 190–195.
- Bernard, K.A., Maffei, J.G., Jones, S.A., Kauffman, E.B., Ebel, G., Dupuis, A.P., 2001. West Nile virus infection in birds and mosquitoes, New York State, 2000. *Emerg. Infect. Dis.* 7, 679–685.
- Bertolotti, L., Kitron, U., Goldberg, T.L., 2007. Diversity and evolution of West Nile virus in Illinois and the United States, 2002–2005. *Virology* 360, 143–149.
- Bondre, V.P., Jadhav, R.S., Mishra, A.C., Yergolkar, P.N., Arankalle, V.A., 2007. West Nile virus isolates from India: evidence for a distinct genetic lineage. *J. Gen. Virol.* 88, 875–884.
- Brault, A.C., Huang, C.Y.H., Langevin, S.A., Kinney, R.M., Bowen, R.A., Ramey, W.N., Panella, N.A., Holmes, E.C., Powers, A.M., Miller, B.R., 2007. A single positively

- selected West Nile viral mutation confers increased virogenesis in American crows. *Nat. Genet.* 39, 1162–1166.
- Brinton, M.A., 2001. Host factors involved in West Nile virus replication. *Ann. N. Y. Acad. Sci.* 951, 207–219.
- Calisher, C.A., 1989. Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. *J. Gen. Virol.* 70, 37–43.
- Daffis, S., Lazear, H.M., Liu, W.J., Audsley, M., Engle, M., Khromykh, A.A., Diamond, M.S., 2011. The naturally attenuated Kunjin strain of West Nile virus shows enhanced sensitivity to the host type I interferon response. *J. Virol.* 85, 5664–5668.
- Davis, C.T., Ebel, G.D., Lanciotti, R.S., Brault, A.C., Guzman, H., Siirin, M., Lambert, A., Parsons, R.E., Beasley, D.W.C., Novak, R.J., 2005. Phylogenetic analysis of North American West Nile virus isolates, 2001–2004: evidence for the emergence of a dominant genotype. *Virology* 342, 252–265.
- Ebel, G.D., Carricaburu, J., Young, D., Bernard, K.A., Kramer, L.D., 2004. Genetic and phenotypic variation of West Nile virus in New York, 2000–2003. *Am. J. Trop. Med. Hyg.* 71, 493–500.
- Ebel, G.D., Dupuis, A.P., 2001. Partial genetic characterization of West Nile virus strains, New York State, 2000. *Emerg. Infect. Dis.* 7, 650–653.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 (5), 1792–1797.
- Evans, J.D., Crown, R.A., Sohn, J.A., Seeger, C., 2011. West Nile virus infection induces depletion of IFNAR1 protein levels. *Viral Immunol.* 24, 253–263.
- Hubalek, Z., Rudolf, I., Bakonyi, T., Kazdovı, K., Halouzka, J., ıbesta, O., ıikotovı, S., Juricovı, Z., Nowotny, N., 2010. Mosquito (diptera: culicidae) surveillance for arboviruses in an area endemic for West Nile (Lineage Rabensburg) and Tcaron; ahyna viruses in central Europe. *J. Med. Entomol.* 47, 466–472.
- Kilpatrick, A.M., 2011. Globalization, land use, and the invasion of West Nile virus. *Science* 334, 323–327.
- Kolekar, P.S., Kale, M.M., Kulkarni-Kale, U., 2010. 'Inter-arrival time' inspired algorithm and its application in clustering and molecular phylogeny. In: *AIP Conf. Proc.*, vol. 1298, pp. 307–312.
- Kolekar, P.S., Kale, M., Kulkarni-Kale, U., 2011. Genotyping of Mumps viruses based on SH gene: development of a server using alignment-free and alignment-based methods. *Immunome Res.* 7, 1–7.
- Kolekar, P.S., Kale, M., Kulkarni-Kale, U., 2012. Alignment-free distance measure based on return time distribution for sequence analysis: applications to clustering, molecular phylogeny and subtyping. *Mol. Phylogenet. Evol.* 65 (2), 510–522.
- Maes, P., Matthijnsens, J., Rahman, M., Van Ranst, M., 2009. RotaC: a web-based tool for the complete genome classification of group A rotaviruses. *BMC Microbiol.* 9, 238–241.
- Mark, H., Eibe, F., Geoffrey, H., Bernhard, P., Peter, R., Ian, H.W., 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* 11, 10–18.
- May, F.J., Davis, C.T., Tesh, R.B., Barrett, A.D.T., 2011. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. *J. Virol.* 85, 2964–2974.
- McMullen, A.R., May, F.J., Li, L., Guzman, H., Bueno Jr., R., Dennett, J.A., Tesh, R.B., Barrett, A.D.T., 2011. Evolution of new genotype of West Nile virus in North America. *Emerg. Infect. Dis.* 17, 785–793.
- Pesko, K.N., Ebel, G.D., 2012. West Nile virus population genetics and evolution. *Infect. Genet. Evol.* 12, 181–190.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Spiroski, M., Milenkovic, Z., Petlichkovski, A., Ivanovski, L., Topuzovska, I.K., Djulejic, E., 2013. Killer cell immunoglobulin-like receptor genes in four human West Nile virus infections reported 2011 in the Republic of Macedonia. *Hum. Immunol.* 74 (3), 389–394.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Vazquez, A., Sıınchez-Seco, M.P., Ruiz, S., Molero, F., Hernıandez, L., Moreno, J., Magallanes, A., Tejedor, C.G., Tenorio, A., 2010. Putative new lineage of West Nile virus, Spain. *Emerg. Infect. Dis.* 16, 549–552.
- Venter, M., Human, S., Zaayman, D., Gerdes, G.H., Williams, J., Steyl, J., Leman, P.A., Paweska, J.T., Setzkorn, H., Rous, G., 2009. Lineage 2 West Nile virus as cause of fatal neurologic disease in horses, South Africa. *Emerg. Infect. Dis.* 15, 877–884.
- Venter, M., Swanepoel, R., 2010. West Nile virus lineage 2 as a cause of zoonotic neurological disease in humans and horses in southern Africa. *Vector Borne Zoonotic Dis.* 10, 659–664.
- Westaway, E.G., Mackenzie, J.M., Khromykh, A.A., 2002. Replication and gene function in Kunjin virus. In: *Japanese Encephalitis and West Nile Viruses*. Springer, Berlin, Heidelberg, pp. 323–351.