Virology Journal

BRIEF REPORT

Open Access

Phylogenomic insights into West Nile virus lineage 2 circulating in Croatia

Barbara Anđelić Dmitrović[1], Kristian Bodulić[2] and Ivan-Christian Kurolt[1*]

Abstract

West Nile Virus (WNV) has been circulating in Croatia for over a decade, as well as in numerous countries across the Mediterranean Basin and Central Europe. Using high-throughput Illumina sequencing, we successfully sequenced the WNV genome from serum and urine samples of three Croatian patients infected in 2016, 2023 and 2024. We employed an amplicon-based approach, with optimized primers designed using previously published WNV lineage 2 sequences. This work represents the first report of complete WNV genome sequences from Croatia. Sequenced genomes from 2023 to 2024 exhibited 99% sequence identity with Hungarian genomic sequences, while the genome from 2016 showed 99% identity with genomic sequences from Austria and Czech Republic. Phylogenetic analysis supported these findings, placing the genomes from 2023 to 2024 in subclade D.2.1 and the genome from 2016 in subclade E.2, thereby providing evidence for potential viral dissemination routes. The approach employed in this study performed well in whole-genome assembly of WNV and in identifying nucleotide changes specific for WNV subtypes circulating in Croatia. This study offers deeper genomic insights into the phylogenetic position of WNV in Croatia, bridging a significant knowledge gap and contributing to future studies on WNV transmission dynamics.

Keywords Molecular phylogenetics, West Nile virus, Whole genome sequencing, Croatia

Introduction

Mosquito borne flaviviruses represent an important group of emerging arboviruses posing a significant threat to public health. West Nile virus (WNV), species *Orthoflavivirus nilense* (*Orthoflavivirus*, *Flaviviridae*) is likely the most widely distributed encephalitic arbovirus, causing severe and recurrent outbreaks in North America and Europe over the past 25 years [1]. Of the nine proposed WNV lineages, only two have been associated with the ongoing WNV outbreaks. This includes lineage 1a, present in Europe, Africa, and North America, and lineage 2, mainly circulating in Europe [1, 2]. Since its first detection in Europe, lineage 2 has become dominant, spreading throughout Central and Southern Europe from Hungary in 2004 [1]. Phylogenetic analyses have identified four distinct clades within lineage 2, termed A, B, C, and F. Despite WNV's relatively slow mutation rate, due to its replication cycle involving mosquitoes and birds, with humans and horses representing dead end hosts, the European clade F further subdivided into subclades D and E after its introduction around 25 years ago [3]. Subclade D includes strains circulating in Central Europe, the Western Balkans, and Southern Europe, while subclade E encompasses strains from Western and Central Europe [3].

*Correspondence:
Ivan-Christian Kurolt
ikurolt@bfm.hr
[1]Department for Dangerous Pathogens, University Hospital for Infectious Diseases "Dr. Fran Mihaljević", Mirogojska 8, Zagreb 10000, Croatia
[2]Department for Bioinformatics and Statistics, University Hospital for Infectious Diseases "Dr. Fran Mihaljević", Mirogojska 8, Zagreb 10000, Croatia

The first cases of neuroinvasive disease in humans caused by WNV have been observed in Croatia in 2012, with new cases reported annually [4]. While previous molecular characterizations in Croatia identified lineage 2 strains, they were limited to a 163 bp NS5 fragment, hindering detailed phylogenetic analysis [5]. Although WNV has been circulating in Croatia for over a decade, the lack of complete genomic data has impeded clade-level characterization. Given that the previously published short WNV sequences from Croatia are uninformative and provide limited phylogenetic insight, complete genomic sequences obtained in this study build upon prior research on West Nile Virus in Croatia, enabling more robust phylodynamic and phylogeographic analyses. In this context, the present study provides the first complete and near-complete WNV genome sequences from Croatia, derived from three patients infected in 2016, 2023, and 2024. Additionally, we conducted a detailed phylogenetic analysis of the obtained sequences within the WNV lineage 2, offering insights into the genetic relationships and evolutionary history of the virus within the region.

Materials and methods

Whole-genome sequencing

Viral RNA was isolated using the BioMagPure Viral Nucleic Acid Extraction Kit (Biosan, Riga, Latvia). Double-stranded cDNA was synthesized with the SuperScriptTM III Reverse Transcriptase (Thermo Fisher Scientific-Invitrogen, Waltham, MA, USA). The primer panel for the multiplex PCR amplification of the WNV lineage 2 genome was designed with PrimalScheme [5], utilizing WNV genomes listed in Supplement Table1. Each of the designed primer pairs was tested in a separate PCR reaction to confirm binding and specificity. The multiplex reactions were performed with two primer pools, ensuring that neighboring amplicons do not overlap within the same pool. Briefly, 25 µl multiplex reaction was prepared using 2.5 µL of double-stranded cDNA, 10 µM primer pool, 5x Q5 Reaction Buffer, 10 mM dNTPs, and Q5® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA). Reaction conditions were as described in Quick et al. [5]. Prior to library preparation, two multiplex reactions were combined into a single, unique sample based on their concentrations. The NEBNext UltraExpress® FS DNA Library Prep Kit and the NEBNext® Multiplex Oligos for Illumina® (New England Biolabs, Ipswich, MA, USA) were used in library preparation. The Qubit dsDNA HS Assay Kits on the Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Agilent High Sensitivity DNA Kit on the Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) were used to quantify DNA concentration and library size, respectively. Equimolar libraries were sequenced in a 100-bp long paired-end configuration on MiSeq (Illumina, San Diego, CA, USA) using the MiSeq Reagent Kit v3 600-cycle.

Bioinformatic analysis and molecular phylogeny

Read quality control was performed using FastQC (version 0.12.0.). Adapter and quality trimming were performed using Trim Galore (version 0.6.4) [6] with the following parameters: -q 5 -e 0.1 --stringency 10 --length 20. All reads were mapped to the reference genome (KM203860.1) using Bowtie2 (version 2.5.4) [7]. Mapped reads were extracted using samtools (version 1.20) and de-novo assembled into a consensus genomic sequence using SPades (version 3.15.4) with default parameters and the --rnaviral option [8]. Generated sequences were deposited in GenBank database (Table 1). Obtained sequences were compared with previously published WNV-2 sequences using the Basic Local Alignment Search Tool (BLAST) available through NCBI [9, 10]. Multiple DNA and protein sequence alignments were performed with Clustal Omega [11] and maximum-likelihood tree was inferred using IQ-TREE [12] with an ultrafast bootstrap approximation approach [13, 14], considering the best substitution model determined using ModelFinder [15]. Tree visualization was performed using iTOL [16]. Pairwise distances were calculated in MEGA-11 [17]. Nonsynonymous amino acid substitutions were analyzed and plotted using the ggmsa package [18] in R (version 4.4.1) [19].

Results

The complete WNV genome was sequenced from a urine sample collected in 2016 (11086 bp) and a serum sample collected in 2024 (10986 bp) (Table 1). Additionally, a urine sample collected in 2023 yielded a partial genome sequence (9808 bp) (Table 1). All sequences were deposited to GenBank under the accession numbers PQ738243, PQ738244, and PQ778820. Comparative analysis revealed that the 2016 genome (PQ738244) shared 99% sequence identity with genomic sequences reported from Austria and the Czech Republic, while the 2023 (PQ778820) and 2024 (PQ738243) genomes shared

Table 1 Molecular characteristics and additional data of viral samples from Croatia

Year	Location	Sample type	Diagnostic Cq value	Genome completeness (%)	WNV-2 sublineage	Accession number
2016	Zagreb	Urine	31,6	99,4	E2	PQ738244
2023	Slavonski Brod	Urine	25,9	87	D2.1	PQ778820
2024	Slavonski Brod	Serum	30,5	99,3	D2.1	PQ738243

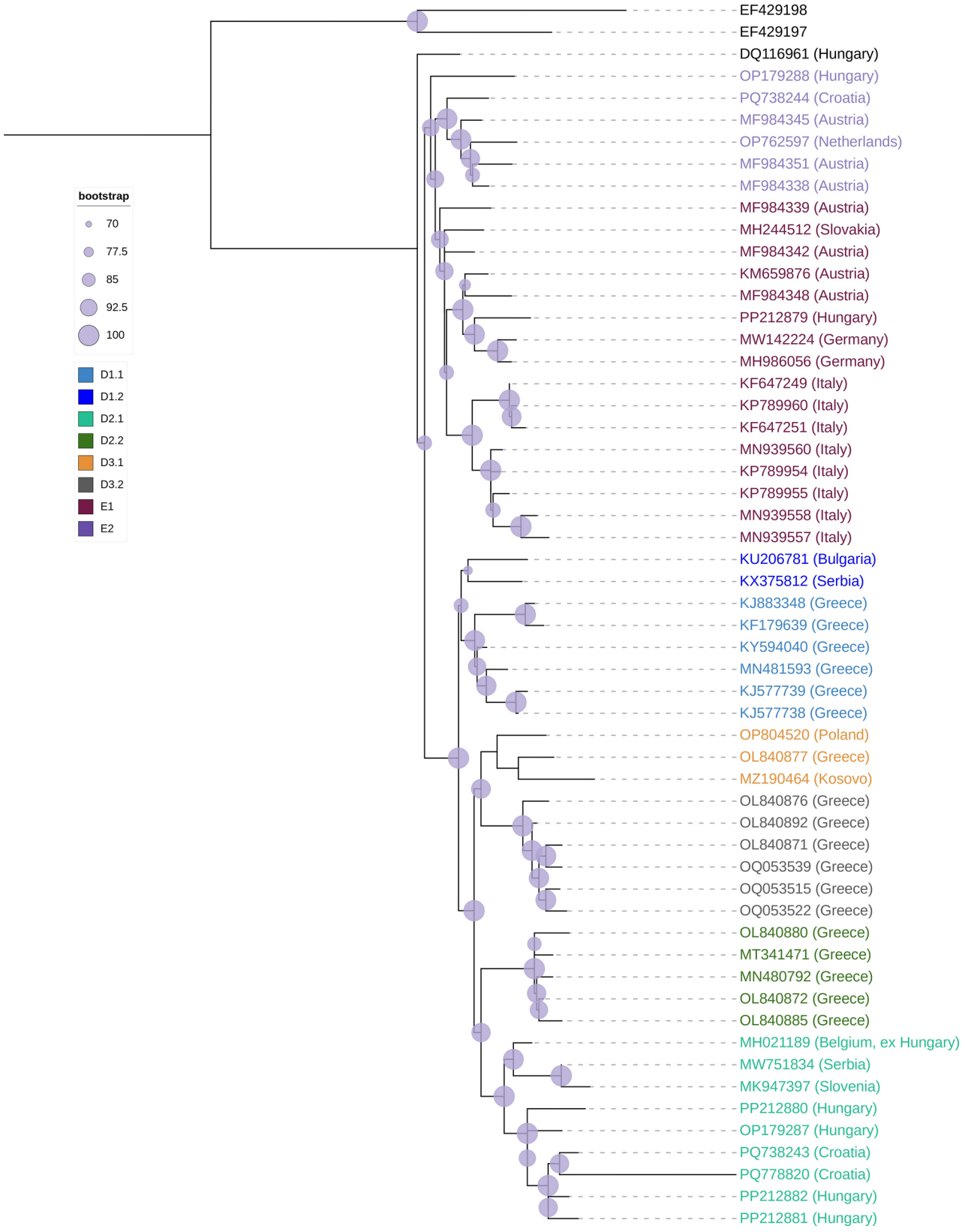

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 The phylogenetic analysis of WNV lineage 2 genomic sequences. The maximum-likelihood tree was inferred using IQ-Tree [12]. The numbers located on the branches represent ultrafast bootstrap values (≥ 0.7) [13, 14]. The TN + F + G4 model was applied as the optimal nucleotide substitution model. The dataset included sequences from Croatia (generated in this study), WNV lineage 2 sequences from the BV-BRC database isolated from human samples from Europe and sequences from GenBank for specific lineages and clades. The tree was rooted using WNV genomes EF429198 and EF429197 as outgroups. WNV genome sequences are annotated with GenBank accession numbers, country of origin, clades and sublineages [20, 21].

99% sequence identity with genomic sequences previously reported in Hungary. Phylogenetic analysis shown in Fig. 1 further supported these findings. The sequence from 2016 was assigned to clade E, sublineage E2, clustering with sequences from Austria and Hungary. In contrast, the sequences from 2023 to 2024 were placed within clade D, sublineage D2.1, grouping with sequences from Hungary and other neighboring countries.

The nucleotide differences between the prototype Hungary/04 strain and the Croatian WNV-2 genomes were 0.46% for the 2016 genome and 0.82% for the 2024 genome (Supplement Table 2). The nucleotide divergence between the 2016 Croatian genome and the previously published E2 sublineage genome from Austria was 0.29%, while the divergence between the 2024 genome and the D2.1 sublineage genome from Serbia was 0.51%.

Analysis of the WNV-2 polyprotein revealed fixed nonsynonymous substitutions that can serve as distinctive markers for genome-based phylogenetic clades, including sites such as E159, NS4B15, NS4B53, and RdRp302 (Fig. 2). Such substitutions were observed across the whole polyprotein, encompassing structural and nonstructural protein genes. Two amino acid substitutions were present in the two Croatian sequences in clade D. However, the sequence from clade E exhibited eight amino acid substitutions compared to the 2023 Croatian clade D sequence and ten substitutions compared to the 2024 Croatian clade D sequence. Notably, three of these substitutions also differed from the other analysed sequences within clade E-sublineage E2. We recorded a total of five substitutions between the Croatian E2 sequence and the previously reported E2 sequence from Austria.

Discussion

The complete WNV genomes sequenced from urine and serum samples collected in 2016 and 2024, along with a partial genome from 2023 urine sample, provide insights into the WNV genetic diversity in Croatia. The phylogenetic analysis grouped the 2016 genome with clade E, sublineage E2 while the 2023 and 2024 genomes were assigned to clade D, sublineage D2.1. While these findings imply at least two possible introductions of WNV in Croatia, a more comprehensive dataset is required to confirm this. Furthermore, the results presented in this study corroborate previous research demonstrating that Croatia, like other Western Balkan and Central Europe countries, is characterized by the intermixing of WNV clades, with multiple clades co-circulating in this region over the past decade [20, 21].

Considering that the earliest detections of clade E were reported in Austria in 2008 and in Serbia and Hungary in 2012 [20], the placement of the 2016 Croatian sample within this clade suggests a potential spread of clade E into Croatia from neighboring countries.

Lu et al. [22] highlight the importance of improving West Nile virus sampling in European countries identified as potential hotspots for viral spread. Additionally, the reconstruction of WNV's complete phylogeographic transmission history remains constrained by the limited availability of publicly accessible whole genome sequences, which is crucial for countries like Croatia where WNV has been detected in both humans and animals [5, 16]. The genomic data generated in this study provide the first insights into WNV's phylogenetic placement and clade distribution in Croatia. This is particularly significant considering Croatia's proximity to Serbia and Hungary, both recognized as WNV hotspots [15]. Moreover, the WNV population in Europe exhibits epidemic cycles every three to five years, highlighting the importance of continuous virus surveillance combined with whole genome sequencing. Expanding such efforts, particularly in regions lacking comprehensive genomic data, is essential for understanding the epidemiology and transmission dynamics of WNV [17].

Within the WNV-2 polyprotein, we observed several nonsynonymous substitutions that highlight the significance of specific amino acid changes as markers for phylogenetic classification and adaptation. The existence of these fixed substitutions points towards the possibility of adaptive alterations that are beneficial to viral fitness or virulence [23]. In clade E sequences, we identified a previously described substitution in the RNA-dependent RNA polymerase (RdRp-A302T), important for enzyme folding and stability [23, 24]. For clade D, the observed substitution NS3-H253P, associated with increased virulence and pathogenicity in WNV lineage 1 [22], has also been previously reported [23]. Interestingly, the 2024 genome that grouped in D2.1 sublineage showed two additional amino acid changes compared to the 2023 genome and previously known sequences from this sublineage (E-V324A, NS4A-V33M), suggesting ongoing adaptation within the sublineage. The continuous identification of new sequences in this WNV clade will shed more light on the key drivers of WNV adaptation and the possibility of further diversification within this lineage.

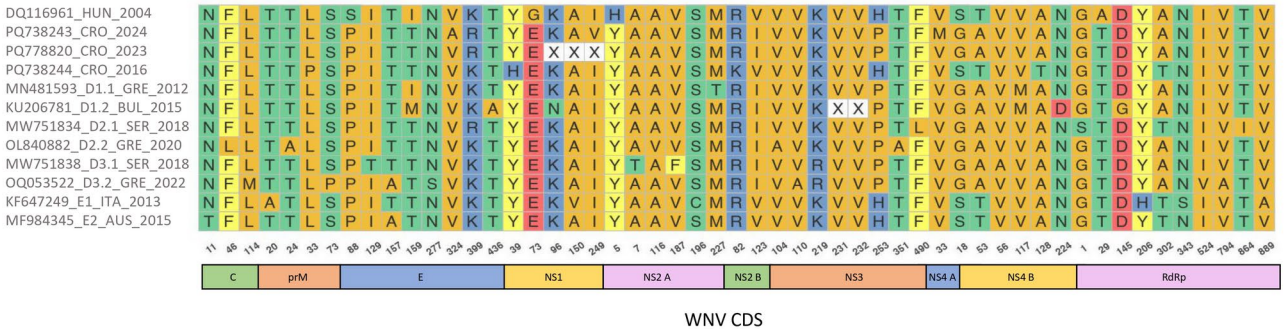

Fig. 2 Nonsynonymous amino acid substitutions within the sequences of European clades and sublineages of West Nile virus. The length of open reading frames (ORFs) of the genomes obtained in this study was 3434 aa. Gaps in the protein sequence are marked with "X" and were a consequence of unsuccessful assembly of the corresponding region. Sequences are labeled with GenBank accession numbers, country of origin, collection year, and clade or sublineage information. Gene positions were based on the West Nile virus lineage 2 reference sequence (NC_001563.2).

Furthermore, the identification of amino acid substitutions is important for better understanding of the WNV molecular epidemiology and may become useful in the development of future therapies [25].

Phylogeographic analysis indicates that WNV transmission is influenced by agricultural activities contributing to ecosystem degradation and biodiversity loss [22]. Agricultural landscapes, such as orchards and irrigated farmland, promote ecological conditions that are highly favorable for WNV vectors and reservoir hosts, potentially leading to higher infection rates in birds, horses, and mosquitoes [26, 27]. In addition, regions with low human population density often feature wide-ranging habitats favorable for the proliferation of *Culex* mosquito larvae, which have been associated with higher WNV incidence [26]. The eastern part of Croatia, which has the highest prevalence of WNV [28] and is the origin of two genomes from this study (PQ778820, PQ738243), exhibits all of the stated characteristics. Altogether, this highlights the complex interplay between agricultural practices, land use, and WNV transmission dynamics, which could be particularly important for Croatia.

Continuous genomic surveillance is important to better understand the WNV evolution and transmission dynamics, particularly in regions influenced by agricultural practices. The capacity of tracking and potentially controlling the spread of WNV across Europe can be improved by continuous collection and analysis of whole-genome sequencing data.

Abbreviations

BLAST	Basic Local Alignment Search Tool
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
RdRp	RNA-dependent RNA polymerase
WNV	West Nile virus

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12985-025-02756-0.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We are thankful to all laboratory members that have provided their help during this study.

Author contributions

B.A.D., and I-C.K contributed to the development of this study protocol and designed the experiments. B.A.D. conducted the molecular analyses. B.A.D. and K.B. performed the data analysis, interpretation, prepared the figures and tables. B.A.D. K.B. and I-C.K prepared and reviewed the manuscript. All authors approved the final version.

Funding

None.

Data availability

The genomic sequences supporting the conclusions of this article are available in the NCBI GenBank repository, under assigned accession numbers: PQ738243, PQ738244, PQ778820.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 January 2025 / Accepted: 22 April 2025
Published online: 08 May 2025

References

1. Rizzoli A, Jiménez-Clavero MA, Barzon L, Cordioli P, Figuerola J, Koraka P, et al. The challenge of West Nile virus in Europe: knowledge gaps and research priorities. Eurosurveillance. 2015;20(20):21135.
2. Pachler K, Lebl K, Berer D, Rudolf I, Hubalek Z, Nowotny N. Putative New West Nile virus lineage in uranotaenia unguiculata mosquitoes, Austria, 2013. Emerging Infectious Diseases journal - CDC. 2014 [cited 2024 Dec 9] 20(12); Available from: https://wwwnc.cdc.gov/eid/article/20/12/14-0921_article

3. Ciccozzi M, Peletto S, Cella E, Giovanetti M, Lai A, Gabanelli E, et al. Epidemiological history and phylogeography of West Nile virus lineage 2. Infect Genet Evol. 2013;17:46–50.
4. Merdić E, Perić L, Pandak N, Kurolt IC, Turić N, Vignjević G, et al. West Nile virus outbreak in humans in Croatia, 2012. Coll Antropol. 2013;37(3):943–7.
5. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for minion and illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6):1261–76.
6. Babraham Bioinformatics - Trim Galore! [cited 2024 May 20]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
7. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
8. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using spades de novo assembler. CP Bioinf. 2020;70(1):e102.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
10. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.
11. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27(1):135–45.
12. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016;44(W1):W232–5.
13. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188–95.
14. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.
15. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
16. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
17. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Battistuzzi FU, editor. Molecular Biology and Evolution. 2021;38(7):3022–7.
18. Zhou L, Feng T, Xu S, Gao F, Lam TT, Wang Q, et al. Ggmsa: a visual exploration tool for multiple sequence alignment and associated data. Brief Bioinform. 2022;23(4):bbac222.
19. R Core Team. A Language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria;; 2024.
20. Šolaja S, Goletić Š, Veljović L, Glišić D. Complex patterns of WNV evolution: a focus on the Western Balkans and Central Europe. Front Vet Sci. 2024 [cited 2024 Dec 9];11. Available from: https://www.frontiersin.org/journals/veterinary-science/articles/10.3389/fvets.2024.1494746/full
21. Srihi H, Chatti N, Ben Mhadheb M, Gharbi J, Abid N. Phylodynamic and phylogeographic analysis of the complete genome of the West Nile virus lineage 2 (WNV-2) in the mediterranean basin. BMC Ecol Evo. 2021;21(1):183.
22. Lu L, Zhang F, Oude Munnink BB, Munger E, Sikkema RS, Pappa S et al. West Nile virus spread in Europe: Phylogeographic pattern analysis and key drivers. Bessen DE, editor. PLoS Pathog. 2024;20(1):e1011880.
23. Barzon L, Papa A, Lavezzo E, Franchin E, Pacenti M, Sinigaglia A, et al. Phylogenetic characterization of Central/Southern European lineage 2 West Nile virus: analysis of human outbreaks in Italy and Greece, 2013–2014. Clin Microbiol Infect. 2015;21(12):e11221–112210.
24. Malet H, Egloff MP, Selisko B, Butcher RE, Wright PJ, Roberts M, et al. Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. J Biol Chem. 2007;282(14):10678–89.
25. Siniavin AE, Dolzhikova IV, Pochtovyi AA, Reshetnikov DA, Gushchin VA. Lethality of West Nile virus strains in a mouse model. 2023 [cited 2024 Dec 23]. Available from: https://doi.org/10.1101/2023.12.22.572994
26. Eisen L, Barker CM, Moore CG, Pape WJ, Winters AM, Cheronis N. Irrigated agriculture is an important risk factor for West Nile virus disease in the hyperendemic Larimer-Boulder-Weld area of North central Colorado. J Med Entomol. 2010;47(5):939–51.
27. Crowder DW, Dykstra EA, Brauner JM, Duffy A, Reed C, Martin E et al. West Nile Virus prevalence across landscapes is mediated by local effects of agriculture on vector and host communities. Wang T, editor. PLoS ONE. 2013;8(1):e55006.
28. Vilibic-Cavlek T, Savic V, Klobucar A, Ferenc T, Ilic M, Bogdanic M, et al. Emerging trends in the West Nile virus epidemiology in Croatia in the 'one health' context, 2011–2020. TropicalMed. 2021;6(3):140.

Publisher's note