

<https://doi.org/10.1038/s42003-025-08159-9>

Development of a novel amplicon based whole-genome sequencing framework for improved surveillance of Toscana virus



Martina Brandolini ^{1,2}✉, Ludovica Ingletto², Giorgio Dirani¹, Silvia Zannoli¹, Andreea Maria Lapusneanu¹, Massimiliano Guerra ¹, Alessandra Mistral De Pascali^{1,2}, Davide Lelli ³, Francesco Defilippo ³, Antonio Lavazza³, Mattia Calzolari ³, Michele Dottori³, Laura Dionisi², Claudia Colosimo², Giulia Gatti^{1,2}, Anna Marzucco¹, Maria Sofia Montanari^{1,2}, Laura Grumiro^{1,2}, Monica Cricca^{1,2}, Alessandra Scagliarini ² & Vittorio Sambri^{1,2}

Toscana virus (TOSV), a Phlebovirus transmitted by sandflies, is a leading cause of aseptic meningitis in the Mediterranean region. Despite its clinical significance, underreporting and limited availability of complete genomic data hinder a thorough understanding of its genetic diversity and evolution. This study presents a novel amplicon-based whole-genome sequencing (WGS) method using Illumina library preparation kits and proprietary software to optimize workflows and enhance bioinformatic analyses. Primers targeting TOSV lineage A genomes were designed with PrimalScheme to generate 400 bp amplicons, incorporating degenerate bases to improve coverage. Library preparation utilized Illumina Microbial Amplicon Prep (iMAP) kits, followed by de novo assembly using BaseSpace DRAGEN Targeted Microbial software. The method's sensitivity was tested on viral propagates at various RNA concentrations (10^4 to 10 copies/ μ L), demonstrating robust performance at concentrations above 10^2 copies/ μ L. Validation on high-titre viral propagates ($n = 7$), low-titre clinical samples ($n = 15$), and phlebotomine pools ($n = 5$) confirmed its reproducibility and ability to comprehensively cover coding regions. Cerebrospinal fluid samples yielded the most consistent results compared to urine and sandfly pools. This innovative WGS approach represents a significant advancement in TOSV genomic surveillance, enabling large-scale studies of its genetic diversity and evolutionary dynamics, which are critical for improving diagnostics and public health strategies.

Toscana virus (TOSV) is a neurotropic sandfly-transmitted virus belonging to the *Bunyavirales* order, *Phenuiviridae* family, and *Phlebovirus* genus. Its genome is represented by a tri-segmented, negative-sense RNA and is divided into three segments known as small (S), medium (M), and large (L), named according to their nucleotide length, with a total length of ~12 kb¹. TOSV was first isolated in 1971 from *Phlebotomus* sandflies in Tuscany (Italy). Its distribution encompasses various countries within the Mediterranean basin, from Southern Europe to Northern Africa^{2,3}. Its estimated seroprevalence was assessed between 1% and 45%^{4–7}, with peaks of 77%⁸ in occupationally exposed populations. Clinical manifestations are often mild and self-limiting, with febrile symptoms generally not requiring medical attention, while severe neuroinvasive manifestations are rare. Such vague

symptoms might lead to misdiagnosis, resulting in underreporting and underestimation of the total number of cases and of the disease-associated burden^{9,10}. Nonetheless, TOSV, along with Enterovirus and Herpesvirus, represents one of the most frequent causes of aseptic meningitis in Southern Europe². TOSV is transmitted to humans (and other vertebrates) by sandflies, specifically *Phlebotomus perniciosus* and *Phlebotomus perfiliewi*^{11–13}, although different species within the genus may also act as competent vectors¹⁴. Consequently, most cases are diagnosed during the warm season, with a peak between June and September, corresponding to the highest vector circulation and activity period. Sporadic off-season cases may also be detected in May, October, and November. To date, the natural reservoir responsible for TOSV maintenance in nature is not known. The

¹Unit of Microbiology, The Greater Romagna Area Hub Laboratory, Cesena, Italy. ²Department of Medical and Surgical Sciences (DIMEC), University of Bologna, Bologna, Italy. ³Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna "Bruno Ubertini" (IZSLER), Brescia, Italy.

✉ e-mail: martina.brandolini3@unibo.it

hypothesis is that the virus is primarily sustained in vectors, with documented instances of vertical transmission^{15,16}. It has been ruled out that it may be represented by humans, given the low viraemic levels reached during infection. Other mammals, such as dogs, were considered but later excluded for the same reason^{17,18}.

TOSV is considered a significant public health concern in Southern Mediterranean countries, many of which have implemented surveillance and notification systems for neuroinvasive cases. In Italy, TOSV is included in some surveillance plans for human arboviral infections, which was first implemented in 2018^{19,20}. Since then, a total of 615 cases of neuroinvasive TOSV infections have been recorded, with an average of 88 cases per year and peaks well above 100 notified cases in 2022 and 2023 (157 and 128, respectively). Emilia-Romagna region, in Northeastern Italy, has the highest incidence of annual TOSV neuroinvasive cases, accounting for an average of 58% of the total number of cases registered in Italy¹⁹, possibly due to an under-notification in other regions.

To correctly investigate and eventually mitigate the impact of this virus on human health, it is crucial to gain a deeper understanding beyond the anecdotal and fragmented evidence that currently exists, especially from a genomic point of view. This study aims to establish a streamlined workflow for amplicon-based whole-genome sequencing of TOSV using the Illumina technology, extensively developed and optimized during the COVID-19 pandemic to facilitate rapid genomic surveillance and variant tracking of SARS-CoV-2, and successfully repurposed in the post-pandemic period for the study of other pathogens, enabling high-throughput genomic analyses to address broader infectious disease challenges. This approach will leverage commercially available customizable library preparation kits, specifically the Illumina Microbial Amplicon Prep (iMAP), and use Illumina-licensed software for data analysis. By integrating these resources, the study aims to facilitate an efficient and user-friendly method for the comprehensive genomic characterization of TOSV.

Results

Primers set

A set of 45 oligonucleotide primer pairs was designed based on TOSV lineage A reference sequences, generating 26 primer pairs for segment L, 13 for segment M, and 6 for segment S capable of amplifying overlapping sequences spanning the entire TOSV genome. The primer set for TOSV lineage A amplicon sequencing is summarized in Table 1.

Phylogenetic trees were constructed for the S, M, and L segments to assess the sensitivity of the designed primers for circulating TOSV lineage A strains. Phylogenetic inference revealed that sequences selected for primer design are distributed across distinct clusters within the trees, representing a diverse set of viral isolates spanning different years and geographic regions (Fig. 1 and Supplementary Table 1). By incorporating degenerate bases on primers based on the chosen genomic sequences, we conceivably managed to enhance the sensitivity by maximizing the binding efficacy of the primers to multiple strains, ultimately mitigating the risk of amplification failure across diverse strains. Consequently, the strategic degeneration of primers, based on the selection of phylogenetically informative sequences, optimized amplicon-based sequencing by maintaining high specificity while accounting for genetic variability. This approach ultimately enhanced genomic surveillance and improved the detection of diverse viral strains.

Sensitivity test

For sensitivity evaluation, four serial 10-fold dilutions (corresponding to 10^4 – 10 copies/ μ L) of two viral propagates were processed in duplicate for amplicon-based sequencing (Fig. 2).

For both samples, the method demonstrated high sequencing efficiency at 10^4 copies/ μ L, with replicate coverage values of 98.2% and 98.5% for IZSLER_181135/14 (mean = 98.35%, standard deviation [SD] = 0.2) and replicates of 96.2% and 96% for RN_050724 (mean = 96.1% and SD = 0.1). At the highest concentrations, the median depth is consistently high across replicates of both propagates, reaching mean values well above 10^3 for all segments. Similarly, at 10^3 copies/ μ L, the performance remained

robust, yielding replicate values of 96.9% and 98.4%, with an average of 97.65% (SD = 1.1) for IZSLER_181135/14. Similar results were obtained for RN_050724, with a coverage of 95.6% and 93.8% for the two replicates (mean = 94.7, SD = 1.3). Depth is maintained consistently high and above 10^3 . At 10^4 and 10^3 copies/ μ L, the consensus length for all segments is consistently full-length for both propagates, suggesting excellent assembly and complete genome recovery at these concentrations. Across all segments, the number of callable bases corresponds closely to the total consensus length. At 10^2 copies/ μ L, a slight decline in sensitivity was observed, with replicate results of 91.4% and 87.2% (mean = 89.3% and SD = 3) for IZSLER_181135/14 and 93.7% and 87.3% (mean = 90.5% and SD = 4.5) for RN_050724. At this concentration, although reduced, depths are still adequate for generating a reliable consensus. At 10^2 copies/ μ L, the consensus lengths are slightly shorter, and callable bases remain high, underlining a good performance, though with a slight reduction in efficiency, highlighting initial limitations in accuracy or assembly at intermediate viral loads. The lowest concentration, 10 copies/ μ L, exhibited notable variability between replicates, with values of 59.9% and 79.1% (mean = 69.5%, SD = 13.6) for IZSLER_181135/14 and 75.3% and 62.6% (mean = 69%, SD = 9) for RN_050724, suggesting that the sequencing efficiency decreases at lower target concentrations, accompanied by greater variability between measurements. At the lowest concentration, the median depth dropped significantly, resulting in suboptimal coverage, which affected assembly quality: the consensus lengths became more variable, and the number of callable bases decreased significantly, indicating that the method failed to produce complete and accurate assemblies.

Considering sensitivity separately for three viral genomic segments (L, M, and S) across the serial dilutions, all segments showed excellent sequencing efficiency at the highest concentration (10^4 and 10^3 copies/ μ L). At 10^2 copies/ μ L, differences between the propagates became more pronounced, and while segment L and S exhibited high performance for both propagates, segment M achieved consistent results for IZSLER_181135/14 but showed significant variability between replicates for RN_050724. At 10 copies/ μ L, detection efficiency decreased substantially for both propagates. Segment L showed comparable performance for IZSLER_181135/14 and RN_050724, while segment M yielded the lowest coverage for both propagates. Segment S showed stronger performance at the lowest concentration.

Validation on viral propagates

The sequence analysis of seven TOSV propagated in Vero E6 cells revealed consistently high-quality output. Total reads ranged from just over 1 million to nearly 3.3 million, with mapped reads consistently exceeding 95% of the total. Sequencing depth was high for all samples ($\sim 10^3$). For all samples, mean coverage exceeded 96% (range 96.3%–99%, SD = 0.8); among the three segments, L and M gave the more consistent results, with coverage above 99% (range 98.7%–100%, SD = 0.45) for the former and averaging 97.8% (range 97.3%–99.7%, SD = 0.85) for the latter. S yielded more variable genome coverage, averaging 96.8% (range 91.9%–98.4%, SD = 2.25). Reconstructed consensus sequence length averaged 6362 nt for segment L (range 6323–6403, SD = 26), 4122 nt for segment M (range 4108–4203, SD = 33), and 1814 nt for segment S (range 1717–1840, SD = 43). For all obtained sequences, the percentage of callable bases in the reconstructed consensus reached 100%. Quantitative output metrics from validation on cell culture isolates are reported in Supplementary Table 3.

Validation on clinical samples

For clinical validation, 17 human samples were sequenced, including 8 cerebrospinal fluid (CSF) and 9 urine samples.

For CSF samples, the Ct values, indicative of viral load, ranged from 29.01 (FO_100724) to 36.54 (RA_030724), with higher Ct values generally correlating with fewer mapped reads and lower coverage. Total reads per sample varied between 1.88 million and 3.11 million. Still, the proportion of mapped reads on the target genome was low overall, ranging from 0.87% (RN_140724) to 22.63% (FO_100724), with only one sample reaching

Table 1 | Primers set used for TOSV amplicon sequencing

Segment	Start	Finish	Name	Degenerate sequence	Pool
S	27	50	S_1_F	ACTTCTAGACAGCCATGCAATCY	1
	392	415	S_1_R	TCTARAGTRGGRATCCCTGTKGG	1
	306	328	S_2_F	AGCCAACYATGAGRGARCTCTA	2
	689	711	S_2_R	TCATYAGATGGGTGTCTCTRGT	2
	618	640	S_3_F	TCCTHGGAGWYATCTGYGTGGT	1
	987	1008	S_3_R	AATCTGAGGGGTGGGATCAYA	1
	868	890	S_4_F	CACCTTTTGGGCTAAGGACTGG	2
	1244	1272	S_4_R	TCTCAAAGACYATAAATCCATCTCTGAG	2
	1180	1200	S_5_F	TCCACTCATGGCGGCCATRT	1
	1550	1572	S_5_R	AGAAGGGTGCTTYTATTGTGGC	1
	1332	1356	S_6_F	CCAGRTCAATGATYCCAGCRAAGC	2
	1847	1869	S_6_R	TGCATGGGTGAATGAGTTTGCT	2
M	14	36	M_1_F	AGGGAACGGCACTAGAACTTCT	1
	462	484	M_1_R	AYAATTGGTGGTGGGATMACTG	1
	398	420	M_2_F	CGGAARCCATTYTGSAAGAGT	2
	775	797	M_2_R	YTCACGCTCCTTCTTYTTGAGC	2
	709	731	M_3_F	SGACYTRGCAAAAGAAATAGCG	1
	1092	1114	M_3_R	GCCTCHARCATGCTGTACTTGT	1
	1020	1042	M_4_F	ATGGCTCARAGTGCAAATCCTG	2
	1406	1428	M_4_R	AGTACCTTGTRYTTGGGSTGGS	2
	1333	1355	M_5_F	CTCAAAAACCYTGAGCGGYCAG	1
	1714	1738	M_5_R	ACATTTCAGTYTCACASGGRTAC	1
	1632	1655	M_6_F	AGAAGGTYAARGTTGAGTCATGC	2
	2026	2048	M_6_R	AACAATCTCGCTGTTTCCTGGG	2
	1962	1984	M_7_F	CRTGTAGCCATGGKTCTTGCAA	1
	2347	2369	M_7_R	GCCCCACTYTCYTTYTCCAGCTT	1
	2260	2283	M_8_F	GGGRGCTAARGATCTTGTYAAGA	2
	2641	2663	M_8_R	ACCCTTCARAATYAGGCAGGAC	2
	2557	2578	M_9_F	CATGCAATGYACAACCTCRGG	1
	2940	2962	M_9_R	TGGACRTAGAGACAAGAAGGGT	1
	2880	2904	M_10_F	TGCATGAGAATARRTGCTTTGAGC	2
	3258	3280	M_10_R	GCCCCGAATAGCTGTYTCTTCTG	2
	3200	3222	M_11_F	GATGAYCCATATTCYCCAGAGC	1
	3564	3589	M_11_R	GRRCCCTCATCATGARTAACAGC	1
	3492	3516	M_12_F	CATTRGAYGACTATGAYGTGGTGT	2
	3884	3906	M_12_R	AYGAGGTYAGTCCAGACACCCA	2
	3799	3819	M_13_F	YACTGCCCCGCATGATGAYA	1
	4192	4215	M_13_R	ACACAAAGACCGGTGCTTTAAAG	1
L	16	40	L_1_F	RAAAACAACCWGCTCCTGTAAGAG	1
	397	419	L_1_R	AGATCTGCTGAGTTTGCATCVM	1
	333	359	L_2_F	CCTGACATGATCATAGACAGAGAC	2
	708	731	L_2_R	AAGCTCGGRAATTTCTTCTCGGT	2
	607	637	L_3_F	GAGAAATTAATCCAGCTCTRTTTGACTCAG	1
	981	1003	L_3_R	TTGTCACCCAGGGAGGRAACTG	1
	862	890	L_4_F	CCAATAAAGAGAGATTTGAGCTRAACTC	2
	1243	1265	L_4_R	CAAACCTCCYAGGACTGCAGCAT	2
	1157	1179	L_5_F	TGCAATGTCWGGGTCAACAGAG	1
	1540	1561	L_5_R	CCATTTGCGAGCCAGCTTCTTA	1
	1464	1486	L_6_F	CAGCCTACTCTGATCCAYATGG	2
	1854	1874	L_6_R	GCCCAGAARCAAGAAGCTGC	2
	1768	1789	L_7_F	ATGCWGGTGAYCTCCTGGTYA	1

Table 1 (continued) | Primers set used for TOSV amplicon sequencing

Segment	Start	Finish	Name	Degenerate sequence	Pool
	2150	2171	L_7_R	AARGGGGTGGCTGATATTCGC	1
	1921	1946	L_8_F	CTGTGRRGTTTTYWGATGCATGGTT	2
	2307	2329	L_8_R	TGGCTGATAATGCTGAWGGCTC	2
	2289	2311	L_9_F	AACAARGARGAAGAYACYGAGC	1
	2658	2681	L_9_R	GTCTTCCYCTATGRGCAAACTG	1
	2592	2616	L_10_F	TGGTACACTTTCAAGGAMGTGAAA	2
	2964	2988	L_10_R	CCACTTTCTAGCATC RT CTGAGGT	2
	2922	2944	L_11_F	AGAGCCA ARARA AGGTGTGGTC	1
	3303	3328	L_11_R	TR GCTGTAGTCTTTACAACTYCTG	1
	3067	3090	L_12_F	GRGGTTGCTCGATGTTCAAAAC	2
	3438	3466	L_12_R	CCAGYTTGTAYCTCATYTTAATCTTCTC	2
	3464	3486	L_13_F	GGTYGCTGCTATGTGTTTYAGG	1
	3834	3856	L_13_R	ACAAGAAGA ARCCA AGCCAGG	1
	3575	3600	L_14_F	GTTCTTCTTYCATTCTCAGCATGTG	2
	3947	3970	L_14_R	CCCCYTTYACCCCTCTTCATGAAA	2
	3870	3890	L_15_F	TCTGCKGGTCTGTGGGATT	1
	4242	4267	L_15_R	CKGAAGATGCCATGACTTTACAGAC	1
	4179	4206	L_16_F	CTAAGAATAGCTGARAAGGTGCATAGT	2
	4548	4568	L_16_R	TCTGGRGCAACTCTGGCTGA	2
	4481	4503	L_17_F	TGTTATYCCAAGAGCAAGCCAG	1
	4855	4877	L_17_R	TCRTCTCCAGCCYTGTAAGA	1
	4753	4775	L_18_F	GTAGGCC R AGGGTCATCAAAGT	2
	5136	5156	L_18_R	TGGYTAACTGCTGTGCCTT	2
	5073	5098	L_19_F	AGTGAAATAGGCATWGCAACACTCA	1
	5454	5478	L_19_R	GCCTAGATTTGT RAT GT CR TGGCG	1
	5310	5332	L_20_F	CAGAGTTTGA ARG CATGGTGTG	2
	5684	5706	L_20_R	C ARG GGCTGGCAYAACATCCAT	2
	5613	5639	L_21_F	AGYTCAA ARC ACATAATGGAGGATGA	1
	6000	6028	L_21_R	AGAGATTGAGACTATCATCATCAAAGCT	1
	5882	5909	L_22_F	GCCTATTGACATGAGTGAAATGTTTGA	2
	6381	6404	L_22_R	GCAGTTAGCCAAGAAGTGCATCA	2

For each oligonucleotide sequence, degenerations in ambiguous sites are highlighted in bold. Pool assignment for multiplex amplification is reported for each primer pair.

97.67% (overall mean of 20.6% and SD = 32). Sequencing depth varied greatly among samples, with a mean of 1446 (range 0–4842, SD = 1523) for segment L, 1056 (range 6–3179, SD = 961) for segment M and 556 (range 58–2258, SD = 713) for segment S. Mean percentage coverage averaged 85.3% (range 55–93, SD = 13.4), with mean segment coverage maintained above 80% for all sequenced samples. Reconstructed consensus sequence length averaged 6048 nt for segment L (range 4539–6266, SD = 609), 4030 nt for segment M (range 3795–4109, SD = 144), and 1656 nt for segment S (range 1423–1704, SD = 97). In the reconstructed consensus sequences, percentage of callable bases averaged 89.6% (range 58.3–98.2, SD = 13) for L, 83.7% (range 55.1–94.4, SD = 15) for M, and 97.7% (range 92–100, SD = 3) for S. While six out of eight CSF samples yielded optimal and consistent sequencing output in terms of genome and segment percentage coverage, two samples, RN_140724 and RN_190724, demonstrated lower performance metrics (Supplementary Table 4).

For urine samples, Ct values ranged from 32.76 (RA_270624) to 39.51 (RN_150624), reflecting a varying viral load among samples. Total reads also varied significantly, from 1.11 million (RA_280723) to 17.47 million (RA_290724). Mapped reads percentages differed widely, from a low 5.91% (RN_150624) to a high of 98.05% (RA_280723), illustrating variations in sequencing efficiency, with an overall mean of 46.2% and SD = 40. Sequencing depth varied greatly among samples, with a mean of 1009 (range 0–2452,

SD = 954) for segment L, 1843 (range 0–3378, SD = 1547) for segment M and 514 (range 0–1890, SD = 599) for segment S. Mean percentage coverage averaged 74.2% (range 26.7–94, SD = 22). Reconstructed consensus sequence length averaged 5965 nt for segment L (range 5637–6265, SD = 297), 3595 nt for segment M (range 1645–4109, SD = 815), and 1525 nt for segment S (range 386–1704, SD = 430). In the reconstructed consensus sequences, the percentage of callable bases averaged 80.1% (Supplementary Table 4).

Validation on phlebotomine pools

Five phlebotomine pools with Ct values ranging from 18.73 to 38.47 were sequenced. The total number of reads generated ranged from ~2.5 to 2.9 million across samples. However, the percentage of reads mapped to the reference genome varied markedly, reflecting differences in viral RNA quantity and quality, from 1.44% to 97.13% (mean = 50.9% and SD = 44.2). Notably, samples 333605/3 and 366877/34 exhibited exceptional mapping efficiencies of 97.05% and 97.13%, respectively, indicative of abundant and high-quality viral genetic material. Conversely, 390419/4B demonstrated a drastically reduced mapping percentage of 1.44%, despite its Ct, lower than other samples, suggesting poor RNA quality rather than low abundance. Sequencing depth varied markedly across samples with a mean of 2943 (range 0–5801, SD = 2635) for segment L, 2557 (range 0–6346, SD = 2921) for M, and 1343 (range 0–5875, SD = 2543) for S. Mean percentage coverage

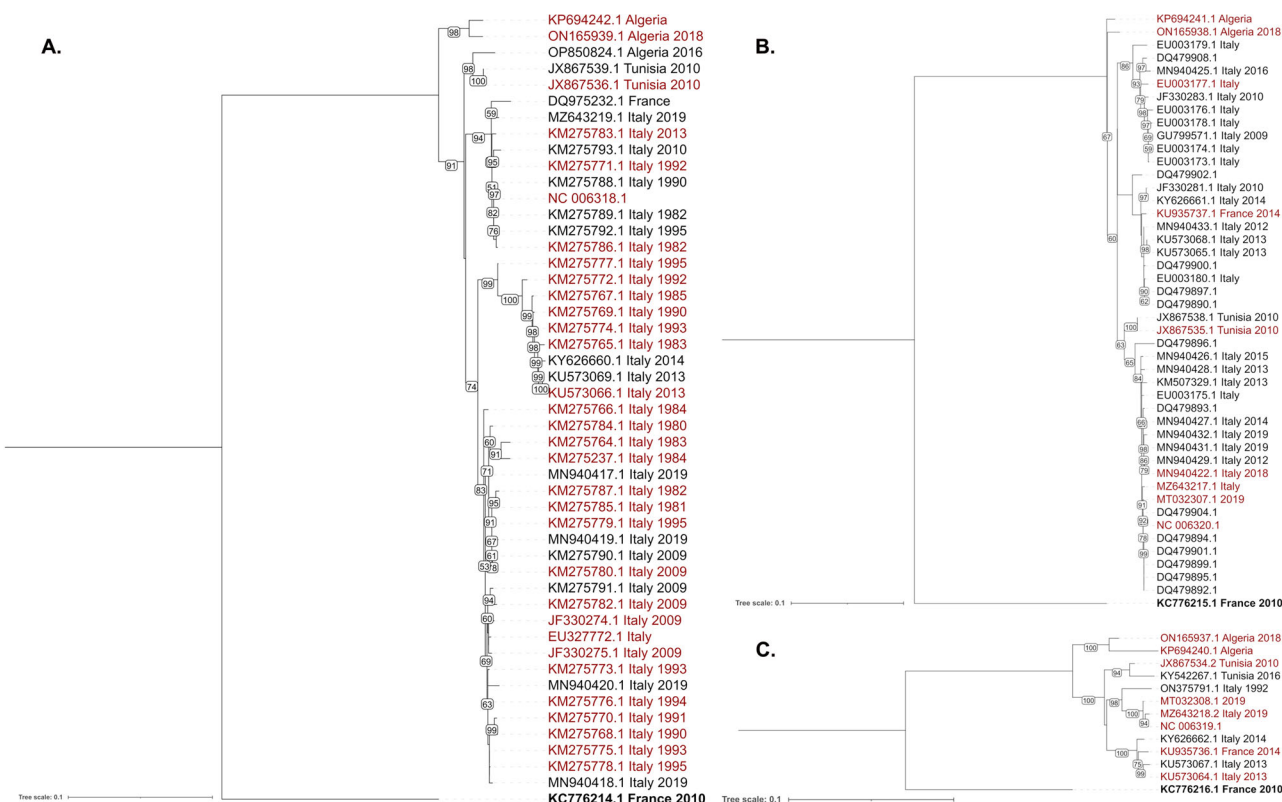


Fig. 1 | Maximum Likelihood phylogenetic tree of segment S, M and L of Toscana virus. This figure illustrates the phylogenetic relationships among the sequences originally used for primer design and degeneration (marked in red) and other TOSV lineage A annotated as complete or near-complete segment sequences in GenBank. Phylogenetic trees were rooted using TOSV lineage B sequences: KC_776214.1 for segment S, KC_776215.1 for segment M, and KC_776216.1 for segment L, with the

outgroup marked in bold. Bootstrap values ranging from 50 to 100 are shown. Where available, isolation location and isolation date are reported in the tree. The tree was inferred using IQ-TREE v2.1.4, with automated model selection (TVM+I+R2 for segment S (A), TVM+F+R2 for segment M (B), and GTR+F+G4 for segment L (C)) and 10,000 ultrafast bootstrap replicates.

averaged 66.6% (range 13.3–96, SD = 35.7), with two samples reaching 95.5% (366877/34) and 96% (333605/3). Reconstructed consensus sequence length averaged 5697 nt for L (range, 3644–6365, SD = 1158), 3238 nt for M (range 390–4109, SD = 1615), and 1097 nt for S (range 0–1702, SD = 840). Percentage of callable bases averaged 82.9% for L (range 53.9–100, SD = 19), 86.5% for M (range 60.7–100, SD = 18.9), and 80% for S (range 0–100, SD = 44.7). Sequencing and consensus generation of the S segment failed in sample 300419/4B, potentially due to low-viral genome quality (Supplementary Table 5). Figure 3 illustrates the relationship between number of genome copies per μL and mean percentage genome coverage across various sample types, providing insights into how viral load, sample matrix, and RNA quality influence sequencing performance and depth.

Sequencing of negative samples (CSF, urine, a phlebotomine pool, a culture supernatant aliquot, and a water sample) did not produce any contigs that mapped to the TOSV genome.

Discussion

TOSV is a major public health threat in Southern European and Northern African countries. It has gained increasing attention due to its clinical relevance, particularly as a cause of meningitis and meningoencephalitis. Despite its clinical importance, TOSV remains under-reported, and knowledge of its epidemiology, ecology, and genetic diversity is limited compared to other arboviruses. From a genomic standpoint, three lineages (A, B, and C) have been described, displaying distinct geographical distribution in countries overlooking the Mediterranean^{21–23}, presumably but not unequivocally linked to competent vector populations and environmental factors. A comprehensive understanding of TOSV's genetic variability is crucial for gaining a deeper insight into its epidemiology and evolution. From this point of view, a significant body of research has focused

on sequencing portions of the TOSV genome (nucleoprotein gene on S segment and glycoprotein gene on M segment) due to the relative ease and lower cost. These partial sequencing efforts have been particularly beneficial in characterizing TOSV's lineages, identifying phylogenetic relationships, and tracking the virus's geographical spread²⁴. However, this approach does not offer a comprehensive understanding of its full genetic complexity and evolutionary dynamics, potentially altering viral replication and infectivity, as well as virulence and transmissibility, may go unnoticed. This is particularly relevant in a context in which environmental changes and human activity increase the potential for viral spread into new geographic areas.

Unlike other viruses with high public health impact, TOSV has limited availability of high-throughput sequencing platforms or dedicated resources for whole-genome studies. These limitations are compounded by the virus's relatively low RNA concentration in clinical specimens, necessitating advanced extraction and amplification methods to obtain adequate material for full genome analysis. Additionally, TOSV's segmented genome poses sequencing challenges, as each segment may require unique amplification primers and conditions. The TOSV segmented genome also represents a challenge for assembly and alignment processes or raw sequencing data. Therefore, our study aimed at the development of a standardized and efficient workflow for whole-genome sequencing of TOSV using a commercial library preparation kit, which streamlines the complex steps of library preparation, critical for robustness and reproducibility, with the library preparation process being completed in ~5 h. The integration of proprietary user-friendly software for data analysis, equipped with robust analytical tools that reduce the complexity of raw data interpretation, further enhanced this workflow by providing user-friendly, tailored bioinformatics solutions optimized for virus-specific data sets, enabling researchers to focus on biological insights rather than technical challenges.

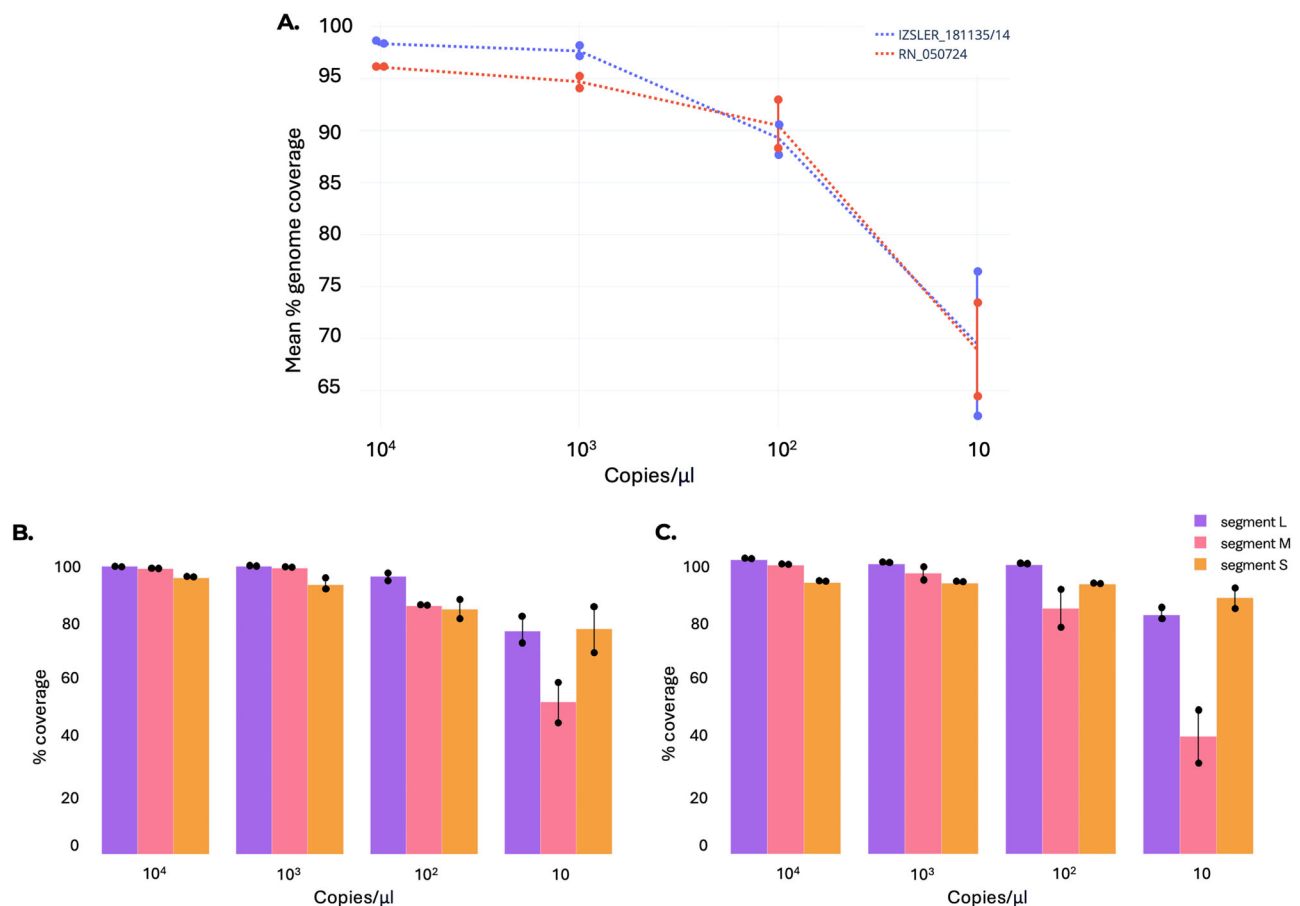


Fig. 2 | Graphical representation of mean % coverage and segment % coverage across dilutions. **A** Comparison of mean percentage genome coverage at varying RNA input concentrations (copies/ μ L) for two TOSV isolates, IZSLER_181135/14 (blue) and RN_050724 (red), with error bars representing standard deviation. Percentage genome coverage of TOSV segments (L, M, S) across varying RNA input

concentrations (copies/ μ L) for IZSLER_181135/14 (**B**) and RN_050724 (**C**), highlighting segment-specific performance with error bars indicating standard deviation. Images are representative of two independent experiments. Error bars represent standard deviations.

Overall, the method demonstrated high sensitivity and robustness at higher viral concentrations with minimal variability in depth, consensus length, and callable bases. At intermediate to low concentration, the method demonstrated reduced consistency and more significant challenges, including incomplete assemblies and reduced callable base percentages, providing insights into its practical limitations for low-abundance targets. Nonetheless, this approach has proven effective in generating high-quality outputs from high-titer viral isolates and samples with low-viral loads, which are commonly encountered in clinical settings such as CSF or urine specimens. By employing a sensitive combination of library preparation and sequencing technologies, this method ensured reliable genome recovery, minimizing the challenges posed by limited nucleic acid availability in low-titer samples. This capacity to work effectively across varying viral loads significantly enhances its utility in both research and diagnostic contexts, enabling robust analysis of clinical samples that would otherwise yield suboptimal results in traditional workflows.

More specifically, sequencing of high-viral-load TOSV propagates yielded remarkable sequencing output metrics. Despite differences in total reads and minor variations in mapping percentages, the consistency in genome coverage across all segments was notable, yielding a mean 98% coverage with 100% of callable bases for every sequenced propagate. This discrepancy can be attributed to the primer design employed for the multiplex amplification. Specifically, the high genetic variability in the untranslated regions made these areas challenging to target effectively, leading to their partial or total exclusion from primer design in most cases. This strategic exclusion ensured more efficient and accurate amplification of

the highly conserved coding regions. Despite such limitations, this strategy provided comprehensive coverage of the entire coding region of the viral genome by capturing the most biologically relevant portions of the genome while mitigating the challenges posed by variability in non-coding regions.

Regarding sequencing validation on real low-viral-load clinical samples, CSF samples generally produced fewer total reads compared to urine samples (average of 2.3 million for CSF against 3.9 million for urine). However, the percentage of mapped reads varied widely for both matrices. CSF samples range from 0.87% to 22.63%, with only one sample achieving exceptionally high values (97.67%). In contrast, urine samples showed a more evenly distributed broader range, from as low as 6.89% to as high as 98.05%. CSF samples tended to achieve consistently higher coverage across all genomic segments compared to urine samples, with a non-negligible difference of about 10 percentage points for every segment, and most CSF samples approaching or exceeding 90% coverage. Additionally, CSF samples generally had a higher percentage of bases included in the consensus, likely due to better amplification consistency, while urine samples showed greater variability in consensus quality. Median sequencing depth is comparable between CSF and urine samples. All in all, CSF samples generally displayed more uniform and reliable sequencing results, with high coverage and consensus quality across all segments for most samples, despite their lower total reads. This consistency likely reflects the less complex nature of CSF as a sample matrix and the higher integrity of the viral genome in CSF samples compared to urine. On the other hand, urine samples showed a wider range of performance outcomes: while some samples achieved high coverage and consensus quality comparable to CSF samples, others performed

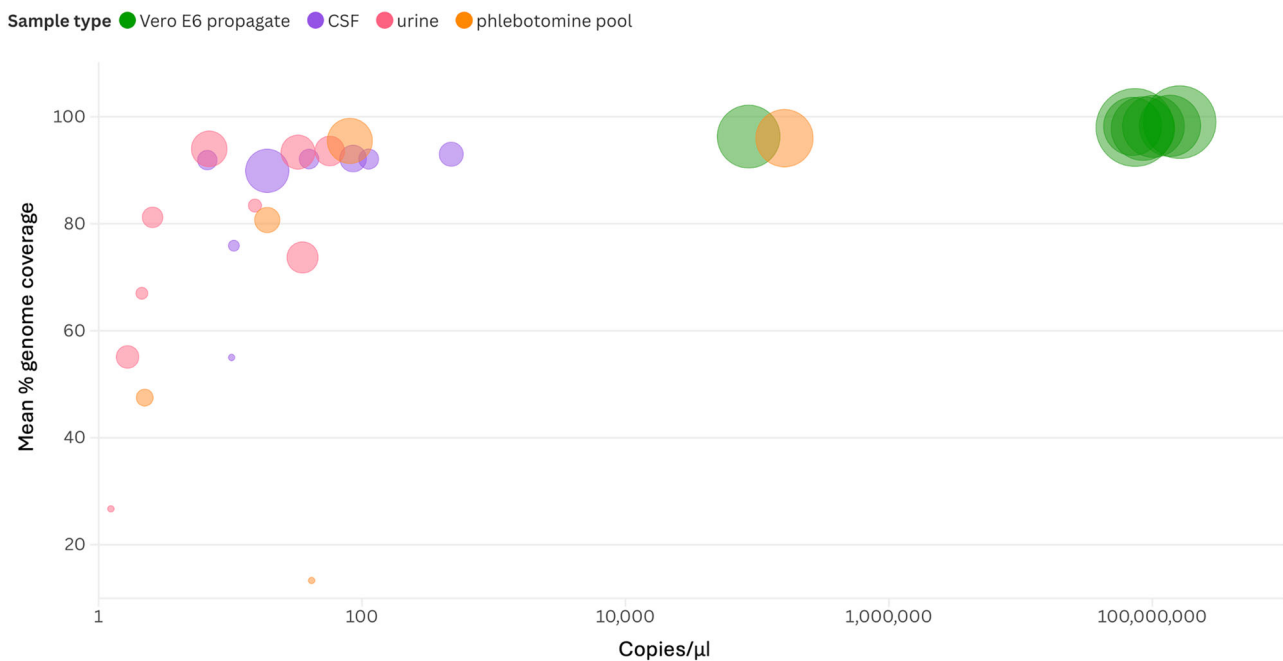


Fig. 3 | Correlation between viral genome copies/μL and mean percentage genome coverage across different sample types (Vero E6 propagates, CSF, urine, and phlebotomine pools). The size of the circles represents sequencing depth, highlighting the relationship between viral load (indicated by copies/μL), sequencing performance (mean genome coverage), and consensus sequence reliability (depth). Higher viral loads are associated with higher coverage and sequencing depth, with

Vero E6 samples consistently achieving near-complete coverage. Variability is observed in clinical (CSF and urine) and vector (phlebotomine pool) samples, reflecting differences in matrix complexity and RNA quality, leading to different sequencing outputs. Vero E6 propagates are represented in green, CSF samples are represented in purple, urine samples are represented in pink, and phlebotomine are represented in orange.

significantly worse, with lower coverage, fewer callable bases, and poorer consensus quality. These differences might stem from the higher complexity of urine as a matrix, with more significant variability in viral RNA integrity.

Furthermore, despite only a small number of Phlebotomine pools being included in the study, our results provided a compelling rationale for the application of the method to this particular matrix. The data highlighted the ability of the whole-genome sequencing approach to recover high-quality consensus sequences for the TOSV genome, with significant coverage and depth observed, particularly in samples with lower Ct values, which reflect higher viral loads. The consistent reconstruction of all three genomic segments (L, M, and S) in several samples underscored the robustness of the methodology, even when applied to complex biological matrices such as phlebotomine homogenates. Despite the challenges associated with sequencing low-quality or scarce genetic material, as seen in a small subset of samples where the assembly of specific segments was incomplete, the overall success rate demonstrated the method's feasibility and adaptability. Encouraged by these preliminary results, extending the application of this sequencing approach to a larger number of phlebotomine pools would provide further evidence of the applicability of the method beyond cell culture isolates and human clinical samples, advocating for its broader application in viral genomic research.

Overall, the correlation between viral load and sequencing coverage demonstrates variability for clinical and vector-derived samples, including CSF, urine, and phlebotomine pools, suggesting that the matrix type may play a role. In particular, CSF samples generally outperformed urine and phlebotomine pools, likely due to the more straightforward matrix and higher quality of RNA present in CSF, which facilitates better genome quality recovery. As viral load decreased to about 20 copies/μL (Ct values approached 35), a decline in genome coverage occurred, and while some samples still achieved moderate coverage with sufficient sequencing depth, others struggled, likely reflecting differences in RNA quality, fragmentation, and matrix effects. At viral load below 10 copies/μL (Ct values above 35), genome coverage declined across all sample types, and sequencing depth mostly became negligible. The detection limit of this sequencing

methodology appeared to fall around a viral load of 20 copies/μL. Beyond this threshold, genome coverage and sequencing depth dropped significantly, indicating that the ability to reconstruct a meaningful and reliable viral genome is compromised. This limitation was most apparent in samples with complex matrices, such as urine and phlebotomine pools, which introduce additional challenges like low RNA recovery and the possible presence of inhibitors. Therefore, samples with Ct values above 35 may require alternative RNA enrichment strategies or complementary techniques (targeted sequencing) to achieve meaningful sequencing results.

From a data analysis perspective, issues were encountered with two propagated samples (RA_190823 and RN_280524A) where the initial analysis with Illumina DRAGEN software failed to generate the consensus sequence for the S segment. This problem was subsequently resolved by instructing the software to perform assembly exclusively against the reference sequence of the S segment rather than the entire genome. This workaround successfully recovered the missing S segment consensus, suggesting the potential occurrence of assembly conflicts, likely arising from sequence similarities or ambiguities in the data. However, such issues would represent a relatively minor concern for TOSV, for which it is implausible for one segment to be completely absent in a propagated sample where the other two are present. Nonetheless, this issue highlights a potentially significant limitation when analyzing samples containing multiple viruses or mixed populations of viral genomes. In such cases, the failure to assemble the genome of a virus, even when its reads are present in the sequencing data, could lead to misleading or incomplete results. For example, if a specific virus in a mixed sample fails to generate a consensus genome due to assembly errors or misalignment during reference-guided contigs assembly, its presence may go undetected or be underestimated. This could lead to erroneous conclusions about the composition of the viral population within the sample, potentially affecting downstream analyses such as epidemiological studies, pathogenicity assessments, or therapeutic decision-making. This bioinformatic challenge underlines the importance of carefully optimizing the workflows. While the TOSV genome structure provided a clue to troubleshoot and recover the missing S segment, this might not always be

the case for less characterized viruses or when analyzing metagenomic samples. In these contexts, ensuring robust assembly pipelines capable of accurately reconstructing all target genomes is essential to avoid incomplete or biased interpretations of the sequencing data.

While primer design and degeneration based on a set of representative TOSV lineage A sequences may ensure broad coverage for lineage A circulating strains, the fact that we exclusively focused on TOSV lineage A, the only lineage circulating in Italy and thus the most relevant to our epidemiological context²¹, may represent a limitation of the study as the effectiveness of the proposed protocol for lineages B and C remains unverified. It is nonetheless plausible that the genetic diversity among the three TOSV lineages could significantly impact the performance of amplification and sequencing protocols, making lineage-specific approaches necessary. This would entail designing distinct sets of primers tailored to each lineage. Future studies should investigate the adaptability of this protocol to lineages B and C, potentially with modifications to account for genetic variability. Another key limitation is the absence of investigation into the potential cross-reactivity of the primer designed for amplification with other phleboviruses. This is a significant issue that merits further exploration, given the possible co-circulation of TOSV and other phleboviruses within the same geographical area. However, the problem of specificity could be addressed by employing species-specific PCR assays, which would direct sequencing efforts exclusively towards samples testing positive for TOSV.

The described approach nonetheless holds the potential to be adopted by other laboratories, thereby enabling a broader range of research institutions to contribute to the global genomic surveillance of TOSV. This expansion of WGS capabilities would enable more coarse studies of viral evolution and interactions with specific sandfly species, potentially uncovering factors driving lineage-specific transmission dynamics and environmental adaptation.

Increased genomic data would not only enhance our knowledge of TOSV evolution and support the development of more effective public health interventions and diagnostic tools. The integration of genomic data with epidemiological and clinical information, together with environmental data, can, in fact, lead to a better understanding of how TOSV spreads and evolves over time, ultimately improving surveillance and public health responses. Another important public health consideration is the potential for TOSV to spread beyond its current endemic regions. Although TOSV is primarily transmitted by *Phlebotomus* sandflies, these vectors are increasingly being found in northern European countries due to rising temperatures associated with climate change. The expansion of sandfly habitats could lead to the introduction of TOSV into new areas, where populations may be immunologically naïve, increasing the risk of outbreaks. Thus, genomic surveillance is essential for early detection of such events and for implementing preventive measures. Moreover, the availability of genomic data can aid in the development and refinement of molecular diagnostic tools by providing a valuable reference for designing primers and probes that are broad-spectrum but also specific, hence guaranteeing lineage inclusivity and avoiding cross-reactivity with other Phleboviruses in different epidemiological settings.

As the world becomes increasingly interconnected and environmental changes impact vector-borne disease transmission, prioritizing whole-genome sequencing initiatives could enhance our preparedness against this and similar emerging pathogens. Expanding WGS in TOSV research aligns with advancing technologies and provides a crucial foundation for managing TOSV and related Phleboviruses to mitigate their public health impact.

Materials and methods

Primer design

Primers were designed on TOSV lineage A S, M, and L segments reference sequences (NCBI accession numbers: NC_006319 for segment L, NC_006320 for segment M, and NC_006318 for segment S) using the Primal Scheme web tool to generate 400 bp tiled amplicons along each segment of the targeted genome. Candidate primers were automatically selected based on primer length, annealing temperature, GC content, 3'

terminus stability, predicted secondary structure, and the likelihood of dimerization. Degenerate bases were automatically introduced at relevant ambiguous sites by Primal Scheme based on a multi-sequence alignment of segment sequences available on NCBI GenBank to ensure coverage even in the presence of genetic variability: 7 sequences for segment L, 8 for segment M and 31 for segment S were retrieved to introduce degenerations (reported in Table S1). To assess whether the variability captured by the designed primers was sufficient to encompass the genetic diversity of circulating TOSV lineage A strains, phylogenetic trees were constructed for each genomic segment using sequences used for primer degeneration and complete and near-complete sequences retrieved from the GenBank NCBI database. Sequence alignment was performed using MAFFT v.7.525 with the *-auto* option²⁵. The resulting multiple sequence alignment was examined and visualized using the Molecular Evolutionary Genetics Analysis software v.11.0.13²⁶. TrimAL v.1.5rev0 was then applied to trim the alignment using the gappyout option²⁷. Finally, phylogenetic analysis was conducted with IQ-TREE v.2.1.4, employing the Maximum Likelihood method with ultrafast bootstrap set to 10,000 and automatic selection of the optimal nucleotide substitution model²⁸. Trees were rooted using KC_776214.1 for segment S, KC_776215.1 for segment M, and KC_776216.1 for segment L (TOSV lineage B sequences). The resulting tree was visualized using iTOL v.1.0. The sequences used to design primers are marked in red.

Primers validation

A serial dilution approach was employed to assess the sensitivity of the sequencing method. Serial 10-fold dilutions of two TOSV cell culture propagates were prepared, and viral RNA extraction was performed with Maxwell® CSC Pathogen Total Nucleic Acid Kit on an automated Maxwell® CSC 48 instrument. Eluates were assayed by RT-PCR (reverse transcription polymerase chain reaction) and dPCR (digital PCR). For RT-PCR, the primers and probes developed by Thirion et al. were employed at the suggested concentrations²⁹. Reactions were run with AgPath-ID™ One-Step RT-PCR Reagents (ThermoFisher, Waltham, USA) with the following thermal cycling protocol: 50 °C for 20 min, 95 °C for 15 min, and 45 cycles of 95 °C for 10 s, 60 °C for 20 s, and 72 °C for 10 s. Dilutions with a number of copies/μL between 10⁴ and 10 (corresponding to RT-PCR Ct [Cycle Threshold] values between ~20 and 35) were subjected to the sequencing protocol. The aim was to determine the lowest viral concentration at which reliable genomic coverage and consensus sequence reconstruction could be achieved, thus establishing the method's detection limit. Each dilution was sequenced in duplicate. Quantitative analysis of sequencing output metrics was used to evaluate performance across dilutions.

To validate the newly developed sequencing method, a set of TOSV-positive Vero E6 cell culture isolates (*n* = 7), clinical specimens (*n* = 17, including *n* = 8 CSF samples and *n* = 9 urine samples), and phlebotomine pools (*n* = 5) were collected and subjected to the sequencing workflow. For all samples, viral RNA extraction was performed with the Maxwell® CSC Pathogen Total Nucleic Acid Kit. Prior to sequencing, all samples were assayed by RT-PCR and dPCR, as previously described. Human samples were collected from patients enrolled in the C.ARBOSQ project, aimed at monitoring arboviruses in the Romagna region (Northeastern Italy). Cell culture isolate IZSLER_181135/14 was kindly provided by the Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna (IZSLER). Sandfly pools (consisting of a maximum of 50 female individuals each) were collected and obtained through the regional plan of surveillance and control of Arboviruses. Prior to RNA extraction, they were homogenized through vortexing with copper-plated 5 mm beads in 300 μL of PBS (Phosphate Buffered Saline) and then centrifuged at 4000 × *g* for 3 min to allow debris sedimentation. Viral RNA extraction and RT-PCR were performed as for human samples and propagates. Sequencing output metrics were analyzed to assess the robustness and reliability of the method across diverse sample types and across a wide range of viral concentrations. Negative controls were included in the sequencing runs, consisting of a TOSV-negative sample from each matrix considered (CSF, urine, a

phlebotomine pool, and a TOSV-negative cell culture supernatant aliquot), as well as a water sample.

Library preparation and sequencing

Amplicon-based library preparation was carried out using the iMAP kit and custom-designed primers. Protocol steps were performed according to the manufacturer's instructions³⁰ with some modifications introduced to enhance sequencing efficiency. Random hexamers (EPH3, Elution Prime Fragment 3HC Mix) were annealed to eluted RNA, combining 8.5 µL of sample and 8.5 µL of EPH3, and incubating at 65 °C for 3 min. Reverse transcription of annealed RNA (17 µL) was carried out with FSM (First Strand Mix, 9 µL) and RVT (Reverse Transcription Mix, 1 µL) at 25 °C for 5 min, 42 °C for 10 min, and 80 °C for 5 min. Template cDNA (5 µL) was used for amplification using IPM (Illumina PCR Mix, 15 µL) and two non-overlapping primer pools at 10 nM concentration (4.3 µL per reaction) to generate ~400 bp amplicons tiled along the genome with the following thermal protocol: 98 °C for 3 min and 35 cycles of 98 °C for 15 s, 58 °C for 1 min and 72 °C for 3 min. After multiplex amplification, for each sample, 10 µL of the two PCR products were pooled and 12 µL of TB1 (Tagmentation Buffer 1), 4 µL of EBLTS (Enrichment Bead-Linked Transposomes), and 20 µL of nuclease-free water were added; tagmentation was carried out at 55 °C for 5 min. Reaction was stopped with 10 µL of ST2 (Stop Tagment Buffer 2), incubated for 5 min. Clean-up was carried out on a magnetic stand with two 3-min sequential washes with 100 µL of TWB (Tagmentation Wash Buffer). Tagmented products were indexed with 24 µL of EPM (Enhanced PCR Mix) and 10 µL of Illumina dual index adapters with the following thermal conditions: 72 °C for 3 min, 98 °C for 3 min and 7 cycles of 98 °C for 20 s, 60 °C for 30 s and 72 °C for 1 min, with a final extension at 72 °C for 3 min. Libraries clean-up was carried out independently for each sample with IPB (Illumina Purification Beads, 45 µL per 50 µL of sample) and two 5-min sequential washes in 80% Ethanol on the magnetic stand. Libraries were eluted in 20 µL of RSB (Resuspension Buffer). Libraries were quantified using Qubit™ 1X dsDNA High Sensitivity Assay (ThermoFisher, Waltham, USA), normalized to 10 nM, and pooled. Pooled libraries were diluted to a final concentration of 10 pM, and 5% PhiX was added as a spike-in control for low-diversity libraries. Paired-end sequencing was carried out on a MiSeq instrument (Illumina) with a 300-cycles V2 Reagents kit.

Data analysis, statistics, and reproducibility

Data analysis was performed with BaseSpace DRAGEN Targeted Microbial (version 4.3.6)³¹ with default parameters. Reads were automatically filtered, and adapter and primer sequences were trimmed using Trimmomatic³². Following dehosting with a modified version of the SRA Human Read Scrubber tool³³, de novo assembly of the scrubbed reads was performed with MEGAHIT³⁴, and similar contigs were clustered using CD-HIT-EST³⁵ to reduce redundancy. The resulting contigs were then mapped to TOSV lineage A segments reference sequences using minimap2³⁶; the best matching reference for each contig was selected for short read mapping and scrubbed reads were aligned to TOSV reference sequences (NCBI accession numbers: NC_006319 for segment L, NC_006320 for segment M and NC_006318 for segment S) using DRAGEN v4.3.6. The following metrics parameters were considered to evaluate the overall quality of the sequencing approach: total reads, mapped reads (and % of mapped reads), % coverage for each segment (and mean % genomic coverage), median depth, consensus length, number of callable bases (and % of callable bases). Sensitivity testing was performed through two independent experiments, whereas validation on clinical and field samples was conducted without technical replicates or independent experiments. Reproducibility was inferred from the consistent results obtained from different samples with similar viral RNA copy numbers, taking into account the variability associated with the analysis of complex biological matrices.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequencing fastq files have been deposited with links to the BioProject accession no. PRJNA1207295 in the NCBI BioProject database. Consensus genomes in FASTA format were deposited in the NCBI database. A table with raw data, SRA accession numbers, and consensus sequences NCBI accession numbers for every isolate can be found in the Supplementary Materials (Supplementary Table 6). Primers definition BED files for primer trimming can be downloaded from <https://doi.org/10.6084/m9.figshare.28869608>³⁷. All data used for graphs are available in the Supplementary Materials (Supplementary Tables 2–5).

Received: 23 January 2025; Accepted: 2 May 2025;

Published online: 08 May 2025

References

1. Sasaya, T. et al. ICTV Virus Taxonomy Profile: Phenuiviridae. *J. Gen. Virol.* **104**, 1–2 (2023).
2. Charrel, R. N. et al. Emergence of Toscana virus in Europe. *Emerg. Infect. Dis.* **11**, 1657–1663 (2005).
3. Charrel, R. N. Emergence of Toscana virus in the Mediterranean area. *World J. Virol.* **1**, 135 (2012).
4. Alkan, C. et al. Sandfly-borne phleboviruses of Eurasia and Africa: epidemiology, genetic diversity, geographic range, control measures. *Antivir. Res.* **100**, 54–74 (2013).
5. Amodio, E. et al. Immunoglobulin M seropositivity for Toscana virus in a random population sample in Sicily. *Int. J. Infect. Dis.* **16**, e633–e635 (2012).
6. Cusi, M. Toscana virus epidemiology: from Italy to beyond. *Open Virol. J.* **4**, 22–28 (2010).
7. Pugliese, A., Beltramo, T. & Torre, D. Seroprevalence study of Tick-borne encephalitis, *Borrelia burgdorferi*, Dengue and Toscana virus in Turin Province. *Cell Biochem. Funct.* **25**, 185–188 (2007).
8. Valassina, M., Valentini, M., Pugliese, A., Valensin, P. E. & Cusi, M. G. Serological survey of Toscana virus infections in a high-risk population in Italy. *Clin. Vaccin. Immunol.* **10**, 483–484 (2003).
9. The Centers for Disease Control and Prevention. Arboviral Diseases, Neuroinvasive and Non-neuroinvasive Case Definition. <https://ndc.services.cdc.gov/case-definitions/arboviral-diseases-neuroinvasive-and-non-neuroinvasive-2015/>.
10. Braitto, A. et al. Evidence of Toscana virus infections without central nervous system involvement: a serological study. *Eur. J. Epidemiol.* **13**, 761–764 (1997).
11. Verani, P., Nicoletti, L., Ciufolini, M. G. & Balducci, M. Viruses transmitted by sandflies in Italy. *Parassitologia* **33 Suppl.**, 513–518 (1991).
12. Verani, P. et al. Ecological and epidemiological studies of the Toscana virus, an arbovirus isolated from phlebotomines. *Ann. Ist. Super. Sanita* **18**, 397–400 (1982).
13. Es-Sette, N., Nourilil, J., Hamdi, S., Mellouki, F. & Lemrani, M. First detection of Toscana virus RNA from sand flies in the genus *Phlebotomus* (Diptera: Phlebotomidae) naturally infected in Morocco. *J. Med. Entomol.* **49**, 1507–1509 (2012).
14. Ayhan, N., Prudhomme, J., Laroche, L., Bañuls, A. L. & Charrel, R. N. Broader geographical distribution of Toscana virus in the Mediterranean region suggests the existence of larger varieties of sand fly vectors. *Microorganisms* **8**, 114 (2020).
15. Maroli, M., Ciufolini, M. G. & Verani, P. Vertical transmission of Toscana virus in the sandfly, *Phlebotomus perniciosus*, via the second gonotrophic cycle. *Med. Vet. Entomol.* **7**, 283–286 (1993).
16. Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N. & Gradoni, L. Phlebotomine sandflies and the spreading of leishmaniasis and other diseases of public health concern. *Med. Vet. Entomol.* **27**, 123–147 (2013).
17. Dincer, E., Gargari, S., Ozkul, A. & Ergunay, K. Potential Animal Reservoirs of Toscana Virus and Coinfections with Leishmania infantum in Turkey. *Am. Soc. Trop. Med. Hyg.* **92**, 690–697 (2015).

18. Muñoz, C. et al. Experimental infection of dogs with Toscana virus and sandfly fever Sicilian virus to determine their potential as possible vertebrate hosts. *Microorganisms* **8**, 596 (2020).
 19. Istituto Superiore di Sanità. Sistema nazionale di sorveglianza delle arbovirosi: i bollettini periodici. <https://www.epicentro.iss.it/arbovirosi/bollettini>.
 20. Istituto Superiore di Sanità M della S. Piano Nazionale Di Prevenzione, Sorveglianza e Risposta Alle Arbovirosi (PNA) 2020-2025. (2020).
 21. Ayhan, N. & Charrel, R. N. An update on Toscana virus distribution, genetics, medical and diagnostic aspects. *Clin. Microbiol. Infect.* **26**, 1017–1023 (2020).
 22. Collao, X. et al. Genetic diversity of Toscana virus. *Emerg. Infect. Dis.* **15**, 574–577 (2009).
 23. Charrel, R. N. et al. Cocirculation of 2 genotypes of Toscana virus, Southeastern France. *Emerg. Infect. Dis.* **13**, 465–468 (2007).
 24. Marsili, G. et al. Fifty years after the first identification of Toscana virus in Italy: Genomic characterization of viral isolates within lineage A and aminoacidic markers of evolution. *Infect. Genet. Evol.* **122**, 105601 (2024).
 25. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
 26. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
 27. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
 28. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
 29. Thirion, L. et al. Evaluation of a trio Toscana virus real-time RT-PCR assay targeting three genomic regions within nucleoprotein gene. *Pathogens* **10**, 254 (2021).
 30. Illumina. Illumina Microbial Amplicon Prep—instructions for using the Illumina Microbial Amplicon Prep Kit. [https://support-docs.illumina.com/LP/IMAP/Content/LP/IMAP/Protocol\(LibPrep\)_IMAP.htm](https://support-docs.illumina.com/LP/IMAP/Content/LP/IMAP/Protocol(LibPrep)_IMAP.htm).
 31. Illumina. Illumina Infectious Disease Software—DRAGEN v.4.3.6. Accessed 3 December 2024. <https://help.idm.illumina.com/dragen-microbial-amplicon/dragen-targeted-microbial>.
 32. Bolger, A., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics*. Accessed 3 December 2024. <http://www.usadellab.org/cms/?page=trimmomatic> (2014).
 33. GitHub. SRA Human Scrubber (version 2.2.1). Accessed 3 December 2024. <https://github.com/ncbi/sra-human-scrubber>.
 34. GitHub. MEGAHIT (version 1.2.9).
 35. GitHub. cd-hit (version 4.8.1). Accessed 3 December 2024. <https://github.com/weizhongli/cdhit>.
 36. GitHub. minimap2 (version 2.28). Accessed 3 December 2024. <https://github.com/lh3/minimap2>.
 37. Brandolini, M. BED files for primer trimming—amplicon based sequencing of TOSV segments L, M and S (C.ARBO.SEQ project). <https://doi.org/10.6084/m9.figshare.28869608.v1> (2025).
- supervision; S.Z.: conceptualization, methodology, investigation, data curation, supervision; A.M.L.: methodology, investigation, data curation; M.G.: investigation, data curation, writing—review and editing; A.M.D.P.: conceptualization, methodology, writing—original draft; D.L.: methodology, investigation, data curation; F.D.: methodology, investigation, data curation; A.L.: methodology, investigation, data curation; M.C. (Mattia Calzolari): methodology, investigation, data curation; M.D.: methodology, investigation, data curation; L.I.: investigation; L.D.: investigation; C.C.: investigation; G.G.: investigation; A.M.: investigation; M.S.M.: investigation; L.G.: investigation; M.C. (Monica Cricca): project administration, writing—review and editing; A.S.: project administration, writing—review and editing; V.S.: project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

Ethics approval

The study involving human clinical samples was conducted in accordance with the Declaration of Helsinki and approved by the Romagna Local Ethical Board (Comitato Etico della Romagna, CEROM) under protocol code C.ARBO.SEQ of 13 May 2023. Prior to enrollment in the present study, all participants gave their written informed consent for the collection and use of biological samples.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08159-9>.

Correspondence and requests for materials should be addressed to Martina Brandolini.

Peer review information *Communications Biology* thanks Richard Steiner Salvato and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Tobias Goris.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Acknowledgements

This research was supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT). We also acknowledge the CINECA award under the ISCRA initiative for the availability of high-performance computing resources and support for phylogenetic analysis.

Author contributions

M.B.: conceptualization, methodology, formal analysis, writing—original draft; G.D.: conceptualization, methodology, investigation, data curation,