Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mutation mapping and evolutionary dynamics of Bundibugyo Ebolavirus (BDBV-2026) genomes reported from Uganda

David Kelvin

David.Kelvin@dal.ca

Dalhousie University

Anuj Kumar

Dalhousie University, Halifax, Canada https://orcid.org/0000-0002-5023-7618

Andrew Nsawotebba

Central Emergency Surveillance and Response Laboratory, Ministry of Health

Barnabas Bakamutumaho

Uganda Virus Research Institute

Bruce Kirenga

Interdisciplinary Consortium for Epidemics Research and Response (ICER)

Mansi Dutt

Dalhousie University

Gustavo Sganzerla Martinez

Dalhousie University https://orcid.org/0000-0002-7656-0579

Winters Muttamba

Interdisciplinary Consortium for Epidemics Research and Response (ICER)

Wilber Sabiiti

University of St Andrews https://orcid.org/0000-0002-4742-2791

Misaki Wayengera

Interdisciplinary Consortium for Epidemics Research and Response (ICER)

Henry Bosa

Interdisciplinary Consortium for Epidemics Research and Response (ICER)

Bernard Lutwama

lwabula@gmail.com

Susan Nabadda

Central Public Health Laboratories of the Republic of Uganda

Charles Olaro

Ministry of Health

Diana Atwiine

Ministry of Health

Charles Ayume

Ministry of Health

Chris Baryomunsi

Ministry of Health

Atek Kagirita

Article

Keywords: Ebola, Bundibugyo, Outbreak, Uganda, and Mutations

Posted Date: July 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10310109/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. AK, MD, GSK, and DJK are shareholders of the company BioForge Canada Limited. BioForge Canada Limited is a company that uses bioinformatics in immunological approaches to the monitoring, prevention, and treatment of infectious diseases. The authors disclose that the interests of BioForge Canada Limited had no impact on this study. DJK is a scientific advisory board member for Emergent BioSolutions.

Mutation mapping and evolutionary dynamics of Bundibugyo Ebolavirus (BDBV-2026) genomes reported from Uganda

Anuj Kumar[1#], Andrew Nsawotebba[2#], Barnabas Bakamutumaho[3,4], Bruce Kirenga[3,5], Mansi Dutt[1], Gustavo Sganzerla Martinez[1,6], Winters Muttamba[3], Wilber Sabiiti[3,7*], David J. Kelvin[1,6,8,9*], Misaki Wayengera[3,11,12], Henry Kyobe Bosa[3,10,13*], Bernard Lutwama[11], Atek Kagirita[11], Suzan Nabadda[2], Charles Olaro[11], Diana Atwiine[11], Charles Ayume[11], Chris Baryomunsi[11]

1. Department of Microbiology and Immunology, Dalhousie University, Halifax, Canada
2. Central Emergency Surveillance and Response Laboratory, Ministry of Health, Kampala, Uganda
3. Interdisciplinary Consortium for Epidemics Research and Response (ICER), Kampala, Uganda
4. Uganda Virus Research Institute, Entebbe, Uganda
5. School of Medicine, College of Health Sciences, Makerere University, Kampala, Uganda
6. Department of Immunology, Shantou University Medical College, Shantou, China
7. Division of Infection and Global Health, School of Medicine, University of St Andrews, Scotland, UK
8. Canadian Center for Vaccinology, Dalhousie University, Izaak Walton Killam (IWK) Health Centre, and the Nova Scotia Health Authority, Halifax, NS, Canada
9. Department of Pediatrics, IWK Health Center, Canadian Center for Vaccinology, Halifax, NS, Canada
10. Incident Management Team, Ministry of Health, Kampala, Uganda
11. Ministry of Health, Kampala, Uganda
12. Departments of Immunology, Molecular Biology & Pathology, School of Biomedical Sciences, College of Health Sciences, Makerere University, Kampala, Uganda
13. Rear Public Health Laboratory, Kampala, Uganda

#Co-first authors

***Corresponding authors:** David J. Kelvin, PhD (Email: David.Kelvin@dal.ca), Henry Kyobe-Bosa, PhD (hskyobe@gmail.com), Wilber Sabiti, PhD (ws31@st-andrews.ac.uk).

Abstract

The ongoing outbreak of Ebola Virus Disease (EVD), caused by the Bundibugyo virus (BDBV) 2026, is evolving rapidly with an increasing number of cases, geographical spread, and cross-border transmission between the Democratic Republic of the Congo (DRC) and Uganda. To identify novel mutations, annotate, and understand the evolutionary dynamics of BDBV-2026, we analysed a total of 19 genomes isolated from Ugandan cases during the ongoing BDBV-2026 outbreak using a comparative genomics approach. We observed a substantial number of amino acid substitutions with lower single nucleotide polymorphisms and complete deletion of 3' leader sequences (61 nucleotides and truncation of 5' trailer sequences (73 nucleotides) in all queried BDBV-2026 genomes. Furthermore, hotspots for unique amino acid substitutions were observed in nucleocapsid (NP), large (L) and glycoprotein (GP) open reading frames. Molecular dynamic simulations showed that changes in the NP protein had a stabilizing effect when compared with older wild-type BDBV NPs. Phylogenetic analysis demonstrated diversity within the BDBV-2026 genomes and the emergence of two sub-groups, I and II, of BDBV-2026 viruses. These findings suggest that the ongoing BDBV-2026 outbreak might have originated from a novel zoonotic spillover. This study provides a high-resolution genomic perspective on BDBV-2026 and highlights key features useful to designing vaccine and anti-viral therapeutics.

Keywords: Ebola; Bundibugyo; Outbreak, Uganda, and Mutations

Main

The Bundibugyo virus (BDBV) is a member of the *Orthoebolavirus* genus (formerly known as *Ebolavirus*) and belongs to the *Filoviridae* family[1]. Previously, it has been defined as one of six species of the Ebola virus[2]. This enveloped and single-stranded RNA virus is known to cause viral haemorrhagic fever in humans. Taxonomically, BDBV is distinct from the Zaire ebolavirus species which have been responsible for multiple outbreaks in Africa, including the 2014–2016 West African epidemic with a case fatality rate of 63%[3]. Symptoms of Ebolavirus Disease (EVD) caused by contagious BDBV generally appear suddenly, i.e., 2 to 21 days after exposure, with an average of 8 to 10 days[4]. Following an outbreak of the rare BDBV strain in the Democratic Republic of the Congo (DRC) and Uganda on 17 May 2026, the World Health Organization (WHO) declared this contagious disease a Public Health Emergency of International Concern (PHEIC) based on the recommendations of the Emergency Committee. In the ongoing outbreak, the first confirmed case was reported from the Ituri province of the DRC. BDBV is now spreading to neighbouring provinces (North Kivu, South Kivu, and Haut-Uele) in the eastern part of the country. As of 9 July 2026, BDBV has led to a total of 1759 confirmed cases and 600 confirmed deaths in DRC[5]. In the neighbouring country, Uganda, 20 confirmed cases have been reported with two deaths as

of the 9 July 2026 Ministry of Health epidemiological update. Out of the 20 confirmed cases, 15 were imported from DRC while five were local transmission cases (four health workers and one taxi driver)[6]. By the same date, seventeen cases had been successfully treated and discharged while one was receiving treatment[5,6]. On 24 June 2026, France also confirmed its first case of BDBV-2026, highlighting the first spread outside Africa. The infected individual was identified as a doctor who was on a humanitarian mission in DRC and contracted the virus[7].

The BDBV genome encodes seven structural proteins, including nucleoprotein (NP), polymerase complex protein (VP35), matrix protein (VP40), spike glycoprotein (GP), minor nucleoprotein (VP30), membrane-associated protein (VP24), and RNA-dependent RNA polymerase (L)[8]. The 3' leader and 5' trailer sequences are characterized as highly conserved short, non-coding flanking regions at the very beginning and end regions of the genome, respectively[9]. These sequences possess the highly conserved *cis*-acting signals that are required for the viral life cycle and likely regulate transcription, replication, and progeny genome packaging of Ebola virus[10]. The 3' leader sequence (located at the 3' beginning of the genome) acts as an entry site and primary promoter for the L (large protein) to initiate the transcription of the viral mRNA and full-length synthesis of positive-sense antigenome[11,12]. The 5' trailer sequence harbours the replication promoter required for generating new negative-sense genomes from the positive-sense antigenomic template[11,12]. In Ebola viruses, the 3' leader sequence has been reported with a total length of 55 nt (ranging from 1 to 55 nt) and the 5' trailer sequence with a total length of 676 nt (ranging between 18,284 and 18,959 nt)[13]. Recently, we reported a clinical and genomic presentation of the first travel-related case of a 59-year-old patient who travelled from DRC to Kampala, Uganda[14]. In another more recent study, we reported BDBV-2026 as a new variant based on the evolutionary analysis of the first three genomes (02 DRC and 01 Uganda)[15]. To date, no comprehensive mutation mapping, annotation, and evolutionary analysis-based studies have been performed for the bulk BDBV-2026 genomes from Uganda.

In the present study, we utilised a total of 19 genomes of BDBV-2026 from 19 different infected individuals to perform in-depth mutational mapping and annotation, followed by phylogeny using an integrated comparative genomics, evolutionary, and molecular modelling approach.

Results

The quality check performed by Nextclade revealed that out of 19 genomes, three genomes (PP_0076RPB.1, PP_0076RVZ.1, and PP_0076RXV.1) showed a large number of N or ambiguous elements. These three genomes were found to have 3360, 781, and 849 N or ambiguous elements, respectively. The mutational analysis revealed single nucleotide

polymorphisms (SNPs), amino acid (aa) substitutions, and stop codons in the BDBV-2026 genomes.

A substantial number of amino acid mutations with a lower number of SNPs in BDBV-2026

The 19 BDBV-2026 genomes considered in this analysis shared 98.79 to 99.04 percentage identity with the reference genome (NC_014373.1) (**Figure 1c**). As shown in **Figure 1c**, when compared to the reference genome, which represents the BDBV-2007 outbreak, the BDBV-2026 genomes were found to have a total of 158 to 221 SNPs (**Supplementary Table 1**). Compared to the BDBV-2012 sequences (average 238 SNPs) of human isolates, BDBV-2026 genomes showed a significantly lower number of SNPs. The lower number of SNPs screened in the BDBV-2026 genomes is consistent with the outcome of our recent molecular clock study, which suggested an alternative evolutionary process for BDBV-2026[16]. Having more protein mutations or substitutions with few SNPs describes a situation where very few changes at the DNA level (SNPs) result in widespread changes at the protein level. This typically happens through frameshift mutations or when the DNA changes alter how RNA is processed but may also be associated with premature stop codon functionality.[17,18]

Deletion of 3' leader and truncation of 5' trailer sequences in BDBV-2026 genomes

In the present study, genome-wide sequence alignment between 19 BDBV-2026 genomes and 17 genomes from previous outbreaks (BDBV-2007 and 2012) reveals that all 19 BDBV-2026 Uganda genomes have a deletion of the 3' leader sequence (61 nt, C1 to A61) and truncation of the 5' trailer sequence (73 nt, T18,867 to A18,940) (**Figure** 1b and **Figure** 2). Like BDBV-2026, two BDBV-2012 genomes reported from the DRC also showed a similar pattern of deletions in the 3' leader sequence. A total of eight DRC sequences from the BDBV-2012 outbreak were also found to have similar or large deletions in the 5' trailer sequence. However, no deletion was observed in all five BDBV-2007 sequences. Previously, the impact of the deletion of the 3' leader and 5' trailer sequences has been evaluated[10,19,20]. Deletions in either the 3' leader or 5' trailer disrupt the promoter regions due to failure of the L protein to effectively bind to the BDBV genome. This process may disrupt viral replication or inhibit the synthesis of viral proteins. Previous reverse genetics studies reported that significant deletions in the 3' leader and 5’ trailer regions can produce a non-viable virus or a heavily “attenuated” (weakened) Ebola strain. Due to deletions in these regions, viruses replicate more slowly and may be more amenable to host immune clearance [10,20]. From an epidemiological point of view, these deletions may result in lower numbers of case fatalities. Early data suggest that signs and symptoms of the ongoing BDBV-2026 are mostly consistent with the infection descriptions of previous outbreaks of BDBV and other species of filovirus[21].

Glycoprotein stop codons and amino acid substitutions

The GP protein, which is critical for the attachment, fusion, and entry of target cells[22], was found to have a total of sixteen amino acid substitutions (**Supplementary Figure 1**). When we compared GP BDBV 2026 to the GP protein of the BDBV-2012 (**PP_006X6W0.1**), we observed two unique substitutions, Y387H and R506S. Based on the domain information available from UniProt (B8XCN0_9MONO), Y387H belongs to the GP1 subunit: mucin-like domain (313–501 aa), while R506S is associated with the GP2 subunit: base interface (502–676 aa). GP_003, a small non-structural secreted glycoprotein, demonstrated a total of four aa changes. Among them, Y151F and E188K, were observed as common mutations within BDBV-2012 outbreak sequences, while T320I and P328L were found to be unique in GP_003 of BDBV-2026. GP_003 protein was also found to have a stop codon at the 367th position.

Hotspots of unique 5-6aa substitutions in NP and L proteins of BDBV-2026

A total of eight proteins from Nextclade, including NP, VP35, VP40, GP, GP_003, VP30, VP24, and L, were found to possess amino-acid (aa) substitutions based on proteome-wide mutation analysis (**Figure 1d and Supplementary Figure 1**). Overall, six proteins, GP, GP_003, L, NP, VP30, and VP35, were found to harbour unique mutations specific to the BDBV-2026 outbreak. However, two key proteins, namely NP and L, contained hotspots of unique mutations with five and six aa unique substitutions, respectively (**Figure 1e-i**).

The NP protein which encodes the nucleoprotein, was found to have hotspots for mutations with a total of nine aa substitutions (**Figure 1 e-f**). Among these nine aa substitutions, five substitutions, N496D, R537K, V544I, D546N, and A569V, were observed to be unique to the BDBV-2026 NP protein. Interestingly, compared to the BDBV-2007 NP protein, the NP protein of BDBV-2012 was found to have a D546S substitution at the 546th position; however, in BDBV-2026, we observed a further substitution (S546N) at the same position. All five unique mutations were found located in the C-terminal domain (NP-Ct) (451–739 aa) of the NP protein (**Table 1**). Most of the genomes from the previous outbreaks lack mutations in this region of the L and NP proteins. Previous reports tend to note conservative substitutions rather than hydrophobic ones[23]. The C-terminal tail and domain of the NP protein play an important role in protein-protein interactions in Ebola viruses[24]. The C-terminal tail is key for binding matrix proteins (VP35 and VP40); therefore, mutations in this region, such as V544I and S546N, may disrupt protein interactions and alter viral fitness[24,25].

As shown in **Figure 1f**, a total of fourteen substitutions were observed in the L protein which encodes the RdRp. Out of these fourteen substitutions, a total of six substitutions (L1011F, N1342H, K1738N, Q1770R, N1771D, and R1954G) were found unique in the BDBV-2026 genome when compared to the BDBV-2012 genomes. Among these six unique mutations,

two unique substitutions (L1011F and N1342H) were found in the range of the capping (Cap) domain (1010–1390 aa). The K1738N substitution was found located in the active site catalytic core (1730 to 1750), and Q1770R and N1771D substitutions were found located in the SAM-binding loop cradle (1765-1785 aa) of the methyltransferase domain. The R1954G substitution was also found to belong to the methyltransferase domain (1600–2000 aa) (**Table 1 and Figure 1 g-i**). As acidic and basic residues are involved in controlling the cap-dependent and cap-independent methylation, mutations in this region may result in a defective viral polymerase, which may prevent the virus from replicating or escaping the host's immune response during host-pathogen interactions[26]. At the structural level, multiple mutations in the acidic and basic residues-enriched C-terminal domain may destabilize its structural dynamics and cause misfolding of the L protein and can disrupt the methylation process leading to altered protein function[26].

Other proteins, including VP24, VP30, VP35, and VP40, were also found to harbour aa substitutions (Supplementary **Figure** 1). The multifunctional VP24 protein, essential for evading the host immune response and assembling infectious viral particles[27,28], was found to have one substitution, S24N, which is also common within the BDBV-2012 outbreak. The VP30 protein was found to have one consensus substitution (H5Y); this substitution is unique to BDBV-2026. Additionally, one substitution, T52I, was also observed in the VP30 protein, encoded by only one of the 19 genomes (PP_0076RXV.1). VP30 a well-characterized viral transcription factor and key component of the viral nucleocapsid, is required to kickstart the transcription of viral genes[23]. VP35 protein was found to possess three substitutions (N79S, G187E, and M192I). Among these three substitutions, N79S was identified as unique to the BDBV-2026 outbreak. One of 19 BDBV-2026 genomes (PP_0076RT3.1) was found to have an additional substitution (V88E) in this protein. Previously, VP35 has been defined as a multifunctional virulence factor; it is important for viral replication and functions as a shield against the host's innate immune response during infection[29,30]. The VP40 protein[31], which is defined as a master regulator of viral assembly, budding, and structural integrity, was found to possess two common substitutions (I14M and A31S) shared with the BDBV-2012 outbreak. Out of the 19 genomes, one genome (PP_0076RWX.1) was found to have one additional unique mutation, F73S, in this protein. The unique substitutions found in these proteins belong to the important functional domains that play an important role in viral replication, transcription, host-pathogen interactions, etc (**Table 1**).

Mutations in the NP protein affect its thermodynamics

Apo simulation was performed for wild-type (BDBV-2007) and mutant (BDBV-2026) models of NP and L proteins (hotspots of unique mutations) to evaluate the impact of the mutations

on protein stability at the atomic level over a standard 500 ns. The calculated root mean square deviation (RMSD) and root mean square fluctuation (RMSF) plots of simulated trajectories showed that novel BDBV-2026 mutations in the NP protein significantly affected protein stability, while no significant differences in protein stability were observed between the wild-type BDBV-2007 L protein and the mutant L protein of BDBV-2026 (**Figure** 3).

The mean RMSD values for the wild-type and mutant models of the NP protein were 2.56 and 1.80 nm, respectively. As evident from **Figures** 3a and b, the wild-type (blue) NP protein model showed fluctuations up to approximately 150 ns, while no major fluctuations were observed in the mutant NP protein model (red) except for the fluctuation at the starting point. The simulated L mutant model showed a stable pattern with lower nm values throughout the simulations on a 500 ns time scale. The calculated RMSD plot of the L protein models reflects that both wild-type (blue) and mutant L protein (red) had similar patterns of fluctuations throughout the simulations. The mutant L protein of BDBV-2026 showed a slightly lower mean score of 0.60 nm compared to the wild L protein, which had a mean score of 0.65 nm (**Figure** 3 d and e). As shown in **Figure** 1c, the calculated RMSF plot of the NP protein model depicts higher fluctuations in the wild-type NP model, and lower fluctuations in the mutant NP model throughout the simulations. Like RMSD, the measured RMSF plot of the L protein models also showed a close pattern of stability in the wild-type NP and mutant NP models throughout the simulations (**Figure** 1f). Altogether, the MD simulations show that mutations in the NP protein of BDBV-2026 yield a more stable dynamic pattern compared to wild-type BDBV-2007 NP, while no significant changes are observed in L protein wild-type and mutant L protein models.

Comparison between cell culture sequences and natural sequences

We compared the number of total substitutions, deletions, insertions, and amino acid substitutions in the BDBV sequence of the BDBV 2026 active outbreak, sequences from past human outbreaks (BDBV-2007 and BDBV-2012), and cell culture sequences (**Figure 4**). Out of the four tested variables, total deletions and total insertions exhibited statistically significant differences across the groups (Kruskal-Wallis $H = 8.03$ and 11.85, $p = 0.01$ and 0.002, $\eta^2 = 0.1131$ and 0.2142, respectively). Pairwise comparisons revealed differences between groups. First, the variable total substitutions were significantly lower in contemporary outbreak sequences than and they were in cell culture sequences (Cliff's $\delta = -0.4251$). Next, the variable total deletions were significantly higher in past outbreak sequences when compared to contemporary outbreak sequences and cell culture sequences (Cliff's $\delta = 0.4118$ and 0.3348, respectively). Variations in the remaining pairwise comparisons were negligible ($\eta^2 \leq 0.05$, $p > 0.05$) with weak pairwise effect magnitudes (all $|\delta| \leq 0.33$).

Phylogeny showed distinct sub-clades of BDBV-2026 with two major sub-groups

The phylogenetic analysis was performed to investigate the evolutionary dynamics of the BDBV-2026 sequence. A calculated bootstrap neighbour-joining phylogenetic tree of the human isolate sequences representing the BDBV-2007 (n=05), BDBV-2012 (n=12), and ongoing BDBV-2026 (n=19) outbreaks from the DRC and Uganda was divided into three sub-clades (**Figure 5**). Interestingly, the newly sequenced genomes from the BDBV-2026 outbreak showed a separate subclade, indicating an emerging novel subclade for BDBV 2026 (**Figure 5**). Furthermore, the Ugandan genomes from BDBV-2026 genomes were found to be divided into two sub-groups (I and II). The PP_0076RGS.1 showed a separate branch within the emerging subclade of BDBV 2026. Compared to the BDBV-2012 sequence, the BDBV-2026 cluster showed a closer association with BDBV-2007 (earlier reported from Uganda in 2007). The phylogenetic tree calculated using Nextclade also showed the same evolutionary pattern as the NJ tree (**Supplementary Figure 2**). This phylogenetic tree also indicates a distinct group of the BDBV-2026 Ugandan sequences with multiple sub-groups and validates the outcome of theNJ tree (**Supplementary Figure 2**). An alternative evolutionary hypothesis of the BDBV-2026 genomes was suggested as the BDBV-2026 genomes did not fit the assumptions for clock-like evolution, a pattern that was consistent in other *Orthoebolavirus* species including BDBV genomes from 2007 and 2012[25], indicating BDBV-2026 may have evolutionary origins or pathways in separate reservoir or intermediate host species.

In the present study, we provide the mutational landscape and evolutionary dynamics of the 19 Ugandan genomes from emerging novel BDBV-2026. Findings of the present study, especially in terms of the complete deletion of the 3' leader sequence and truncation of the 5' trailer sequence, and more mutations in BDBV-2026 genomes despite fewer SNPs, serve as a foundation for additional research efforts in East Africa, especially in DRC and Uganda, to uncover the transmission and mechanisms of the emerging BDBV-2026 and develop potential medical countermeasures. We acknowledge that this study is based on comparative genomics and structural modelling and needs more *in-vitro* and *in-vivo* experiments to evaluate if identified deletions in non-coding 3' leader and 5' trailer sequences and substitutions in the proteins alter viral pathogenicity, transmissibility, or viral fitness of BDBV-2026. We also acknowledge that more global efforts are required for increased supplies, resources, and infrastructure in the DRC and Uganda for diagnostics, sequencing, and in-depth genomic surveillance to develop effective treatments against BVD.

Methods

Annotation and structural modelling

Recently, we deposited a total of 19 full-length genomes to the Pathoplexus database (https://pathoplexus.org/) from Uganda. These genomes were sequenced with the methodology described in Nsawotebba et al.,[14]. In the present analysis, these newly sequenced genomes from the ongoing BDBV-2026 outbreaks were considered for sequence analysis and annotation, followed by structural modelling. To identify the mutations in the BDBV-2026 sequence, mapping to the reference genome (NC_014373.1) was performed in a genome- and proteome-wide manner. The quality check was performed by Nextclade[32]. The sequence alignment was performed using MAFFT[33] (Multiple Alignment using Fast Fourier Transform) version 7 and Snipit[34] tools. Mutation analysis, including nucleotide substitutions and amino acid substitutions, etc., was performed using the Nextclade platform[32] utilizing NC_014373.1 as the reference genome. The three-dimensional (3D) structural models of the BDBV proteins were modelled using the state-of-the-art Google DeepMind-based AlphaFold tool[35]. Subsequently, mutative and superimposed models were generated using the PyMOL package (https://www.pymol.org/). ChimeraX[36] (https://www.rbvi.ucsf.edu/chimerax/) and PyMOL (https://www.pymol.org/) packages were employed to render and highlight the functional domains in the modelled structures based on the annotations available from UniProt[37].

Mutation mapping and quantification

We analysed the median of total substitutions, total insertions, total deletions, and total amino acid substitutions in BDBV genomes from past outbreaks, the current outbreak, and cell culture. Due to the small and unbalanced sample sizes in the current outbreak group (n=19), exact p-values were calculated for the Kruskal-Wallis omnibus test and subsequent post-hoc analyses to improve robustness under limited sample conditions. We note that although the omnibus test can detect large group-level effects, post-hoc comparisons involving the current outbreak group are inherently underpowered and might fail to achieve conventional significance thresholds despite substantial effect magnitudes. Effect sizes were emphasized alongside p-values. The magnitude of the omnibus effect was quantified using eta-squared (η^2) obtained from the Kruskal-Wallis *H* statistic. For variables showing a significant omnibus effect ($p<0{\cdot}05$), pairwise comparisons were performed on independent samples using Mann-Whitney U tests with exact inference. Pairwise effect sizes were calculated using Cliff's Delta (δ) to account for unequal sample sizes. Statistical analyses were executed in Python (version 3.12) using scipy.stats and statannotations libraries.

Phylogenetics

To investigate the evolutionary pattern of the BDBV-2026 genome, a phylogenetic analysis was performed using well-established methods. Other than the 19 new sequences of BDBV-2026 reported from Uganda, a total of 17 other sequences of human isolates representing

the BDBV-2007 and the BDBV-2012 outbreaks were also downloaded from the Pathoplexus platform (https://pathoplexus.org/) in the FASTA format. These sequences were further utilized for multiple sequence alignments (MSA) employing MAFFT[33] version 7 with the auto (FFT-NS-1, FFT-NS-2, FFT-NS-i or L-INS-i) method. Subsequently, after getting the MSA, a bootstrap phylogenetic tree (1000 bootstrap sampling) was calculated applying the Neighbour-joining method utilizing the Jukes-Cantor model. To validate the computed NJ tree, an additional tree was calculated using the Nextclade[32] pipeline. The final tree was visualized and annotated using iTOL[38] (https://itol.embl.de/).

Molecular dynamics simulations for 500 nanoseconds

Molecular dynamics simulations were performed for the apo form of wild and mutant models of the NP and L proteins for 500 nanoseconds. The MD simulation was performed with a primary goal of evaluating the impact of the mutations on protein stability at the atomic level. GROMACS 2024.4[39] package, installed on the high-performance computing (HPC) facility provided Digital Research Alliance of Canada (https://ccdb.alliancecan.ca/), was utilized to perform the MD simulation, following the steps earlier described by Gajula et al.,[40].

Data availability

The newly sequenced genomes of the human isolates used in this study are available on https://pathoplexus.org/. Other sequence analysis data are presented in the form of the Figures and supplementary data.

Acknowledgments

The authors are thankful to Nikki Kelvin for editing the final manuscript. This study is supported by awards from the Canada Research Chair program, Moderna Global Fellowship 2024 (2024-MGF-000316 [91353]), and the Li-Ka Shing Foundation. AK is a Moderna Global Fellow, and DJK is the Canada Research Chair in Translational Vaccinology and Inflammation. Authors from Uganda are supported by the Ministry of Health, Government of Uganda, that provided all laboratory resources for case identification, diagnosis, and whole genome sequencing.

Contributions

Conceptualization and study design: A.K, M.W., H.K.B., W.S., and D.J.K., M.W. Sequencing: A.N., S.N., M.W., and H.K.B. Data collection: B.L., and A.K. Data analysis: A.K., A.N., G.S.M., and M.D. Annotation, evolution, and phylogeny: A.K. Data visualization: A.K., A.N., M.D., and G.S.M. Drafting the manuscript: A.K. All authors contributed to the review and editing and

approval of the final manuscript. Resources: C.O., D.A., C.A., and C.B. Supervised the study: D.J.K.

Competing interests

AK, MD, GSK, and DJK are shareholders of the company BioForge Canada Limited. BioForge Canada Limited is a company that uses bioinformatics in immunological approaches to the monitoring, prevention, and treatment of infectious diseases. The authors disclose that the interests of BioForge Canada Limited had no impact on this study. DJK is a scientific advisory board member for Emergent BioSolutions.

References

1. Kuhn, J. H. *et al.* New filovirus disease classification and nomenclature. *Nature Reviews Microbiology 2019 17:5* **17**, 261–263 (2019).
2. Letafati, A., Salahi Ardekani, O., Karami, H. & Soleimani, M. Ebola virus disease: A narrative review. *Microb. Pathog.* **181**, 106213 (2023).
3. The current Ebola outbreak is a public health emergency of international concern. *Nature Communications 2026 17:1* **17**, 4603- (2026).
4. Ebola Disease Basics | Ebola | CDC. https://www.cdc.gov/ebola/about/index.html.
5. Ebola Outbreak: Current Situation | Ebola | CDC. https://www.cdc.gov/ebola/situation-summary/index.html.
6. Ebola Updates. https://evd-daily.health.go.ug/.
7. France reports Ebola case in doctor returning from Congo | Reuters. https://www.reuters.com/business/healthcare-pharmaceuticals/france-reports-ebola-case-doctor-returning-humanitarian-mission-congo-2026-06-24/.
8. Levine, C. B., Mire, C. E. & Geisbert, T. W. Comparison of Zaire and Bundibugyo Ebolavirus Polymerase Complexes and Susceptibility to Antivirals through a Newly Developed Bundibugyo Minigenome System. *J. Virol.* **95**, e00643-21 (2021).
9. Bach, S. *et al.* Identification and characterization of short leader and trailer RNAs synthesized by the Ebola virus RNA polymerase. *PLoS Pathog.* **17**, e1010002 (2021).
10. Sztuba-Solinska, J. *et al.* A small stem-loop structure of the Ebola virus trailer is essential for replication and interacts with heat-shock protein A8. *Nucleic Acids Res.* **44**, 9831–9846 (2016).

11. Weik, M., Enterlein, S., Schlenz, K. & Mühlberger, E. The Ebola Virus Genomic Replication Promoter Is Bipartite and Follows the Rule of Six. *J. Virol.* **79**, 10660–10671 (2005).

12. Deflubé, L. R. *et al.* Ebolavirus polymerase uses an unconventional genome replication mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8535–8543 (2019).

13. Volchkov, V. E. *et al.* Characterization of the L gene and 5' trailer region of Ebola virus. *J. Gen. Virol.* **80 (Pt 2)**, 355–362 (1999).

14. Nsawotebba, A. *et al.* Clinical Profile and Genomic Characterization of the 2026 Bundibugyo Virus Index Case in Uganda. *Nature Medicine 2026* 1–1 (2026) doi:10.1038/S41591-026-04510-7.

15. Ayitewala, A. *et al.* Bundibugyo ebolavirus of the 2026 Ebola outbreak in Uganda and DR Congo is a new variant. *The Lancet* **407**, (2026).

16. Sganzerla Martinez, G. *et al.* Emerging Problems in Infectious Diseases Clock-like behavior across Orthoebolavirus species suggests alternative evolution of the 2026 Bundibugyo virus outbreak. https://doi.org/10.3855/jidc.23736 (2026) doi:10.3855/jidc.23736.

17. Wang, Z. & Moult, J. SNPs, protein structure, and disease. *Hum. Mutat.* **17**, 263–270 (2001).

18. Prabantu, V. M., Naveenkumar, N. & Srinivasan, N. Influence of Disease-Causing Mutations on Protein Structural Networks. *Front. Mol. Biosci.* **7**, 620554 (2021).

19. Hoenen, T., Watt, A., Mora, A. & Feldmann, H. Modeling The Lifecycle Of Ebola Virus Under Biosafety Level 2 Conditions With Virus-like Particles Containing Tetracistronic Minigenomes. *J. Vis. Exp.* 52381 (2014) doi:10.3791/52381.

20. Fang, J. *et al.* Spatial and functional arrangement of Ebola virus polymerase inside phase-separated viral factories. *Nature Communications 2023 14:1* **14**, 4159- (2023).

21. Akilimali, P. *et al.* Clinical Characteristics of Patients Infected with Bundibugyo Virus, DRC 2026. *New England Journal of Medicine* https://doi.org/10.1056/NEJMC2608070 (2026) doi:10.1056/NEJMC2608070.

22. Lee, J. E. & Saphire, E. O. Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol.* **4**, 621 (2009).

23. Ayub, G. & Waheed, Y. Sequence analysis of the L protein of the Ebola 2014 outbreak: Insight into conserved regions and mutations. *Mol. Med. Rep.* **13**, 4821–4826 (2016).

24. Noda, T., Hagiwara, K., Sagara, H. & Kawaoka, Y. Characterization of the Ebola virus nucleoprotein–RNA complex. *J. Gen. Virol.* **91**, 1478 (2010).

25. Watanabe, S., Noda, T. & Kawaoka, Y. Functional Mapping of the Nucleoprotein of Ebola Virus. *J. Virol.* **80**, 3743 (2006).

26. Valle, C. *et al.* The C-Terminal Domain of the Sudan Ebolavirus L Protein Is Essential for RNA Binding and Methylation. *J. Virol.* **94**, e00520-20 (2020).

27. Xu, W. *et al.* Ebola virus VP24 targets a unique NLS-binding site on karyopherin5 to selectively compete with nuclear import of phosphorylated STAT1. *Cell Host Microbe* **16**, 187 (2014).

28. Zhang, A. P. P. *et al.* The ebolavirus VP24 interferon antagonist: Know your enemy. *Virulence* **3**, 440 (2012).

29. Fujita-Fujiharu, Y. *et al.* Structural basis for Ebola virus nucleocapsid assembly and function regulated by VP24. *Nature Communications 2025 16:1* **16**, 2171- (2025).

30. Leung, D. W., Prins, K. C., Basler, C. F. & Amarasinghe, G. K. Ebolavirus VP35 is a multifunctional virulence factor. *Virulence* **1**, 526 (2010).

31. Johnson, K. A. *et al.* PI(4,5)P2 binding sites in the Ebola virus matrix protein VP40 modulate assembly and budding. *J. Lipid Res.* **65**, 100512 (2024).

32. Aksamentov, I., Roemer, C., Hodcroft, E. B. & Neher, R. A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **6**, 3773 (2021).

33. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

34. O'Toole, Á., Aziz, A. & Maloney, D. Publication-ready single nucleotide polymorphism visualization with snipit. *Bioinformatics* **40**, btae510 (2024).

35. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature 2024 630:8016* **630**, 493–500 (2024).

36. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70 (2020).

37. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

38. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).

39. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

40. Gajula, M. P., Kumar, A. & Ijaq, J. Protocol for Molecular Dynamics Simulations of Proteins. *Bio. Protoc.* **6**, 2051 (2016).

41. Kirchdoerfer, R. N., Moyer, C. L., Abelson, D. M. & Saphire, E. O. The Ebola Virus VP30-NP Interaction Is a Regulator of Viral RNA Synthesis. *PLoS Pathog.* **12**, e1005937 (2016).

42. Zinzula, L. *et al.* Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species. *Structure* **27**, 39-54.e6 (2019).

43. West, B. R. *et al.* Structural Basis of Broad Ebolavirus Neutralization by a Human Survivor Antibody. *Nat. Struct. Mol. Biol.* **26**, 204 (2019).

44. Martínez, M. J. *et al.* Role of Ebola Virus VP30 in Transcription Reinitiation. *J. Virol.* **82**, 12569 (2008).

45. Martin, B. *et al.* The methyltransferase domain of the Sudan ebolavirus L protein specifically targets internal adenosines of RNA substrates, in addition to the cap structure. *Nucleic Acids Res.* **46**, 7902–7912 (2018).

46. Yuan, B. *et al.* Structure of the Ebola virus polymerase complex. *Nature* **610**, 394 (2022).

Table

Table 1. List of the BDBV proteins based on the Nextclade annotation, highlighting their functions and amino acid substitutions. Description, function, and domain information are based on data available from the GenBank and Uniprot repositories. Domain functions are based on the literature survey.

S.No	Protein	Description	Function	Common (shared substitutions with BDBV-2012)	Unique (substitutions In BDBV-2026	Domains harbour unique substitutions	Domains function
1.	NP	Nucleoprotein	Encapsidation of genomic RNA	Q517R; G524S; V531A;	N496D; R537K; V544I; D546N; A569V	C-terminal Domain (NP-Ct) (451–739 aa)	Drives cytoplasmic inclusion body formation; recruits transcription factor VP30.[25,41]
2.	VP35	Polymerase complex protein	RNA-dependent RNA polymerase cofactor	G187E; M192I	N79S	N-terminal Domain (NTD) (1–82 aa)	Anchors multi-protein complexes; manages early structural assembly steps.[42]
3.	VP40	Matrix protein	Coalesce nucleocapsids and cell membranes in virion assembly (budding)	I14M; A31S	N/A	N/A	N/A
4.	GP	Spike glycoprotein	Receptor binding and fusion	Y151F; E188K; P365S; P367L; P389L; R394G; T458I; P481L; T486I; A489V	Y387H; R506S	GP1 Subunit: Mucin-like Domain (313–501 aa)	Creates a dense glycan shield to hide the receptor-binding site from antibodies.[43]

						GP2 Subunit: Base Interface (502–676 aa)	Shapes exterior neutralizing epitopes; regulates entry and fusion kinetics.[43]
5.	GP_003	Small secreted glycoprotein	Receptor binding and fusion	Y151F; E188K;	T320I; P328L	N/A	N/A
6.	VP30	Minor nucleoprotein	Multifunctional protein that acts as a viral transcriptional activator.	N/A	H5Y	N-terminal Domain (NTD) (1–44 aa, IDR)	Acts as a flexible tail adjusting the nearby replication-transcription switch.[44]
7.	VP24	Membrane-associated protein	Evading the host immune response and assembling infectious viral particles.	S24N	N/A	N/A	N/A
8.	L	RNA-dependent RNA polymerase	Catalyzes the replication of viral genomic RNA	V420I; S132A; K1570R; L1595Q; I1664R; I1784V; E1969D; A2191V	L1011F; N1342H; K1738N; Q1770R; N1771D; R1954G	Capping (Cap) Domain (1010–1390 aa)	Catalyzes the 5' capping and structural processing of nascent viral mRNAs.[10]
						Methyltransferase Domain (1600–2000 aa)	Executes ribose 2'-O-methylation via a core K-D-K-E catalytic tetrad.[45]
						Methyltransferase Domain: Active Site Catalytic Core (1730-1750 aa)	Coordinates cap methylation; forms the starting shell of the primary K-D-K-E catalytic pocket.[46]

						Methyltransferase Domain: SAM-Binding Loop Cradle (1765-1785 aa)	Constructs the physical cavity lining that secures the SAM methyl donor co-factor in place.[45]

480

481

482

483

484

Figures

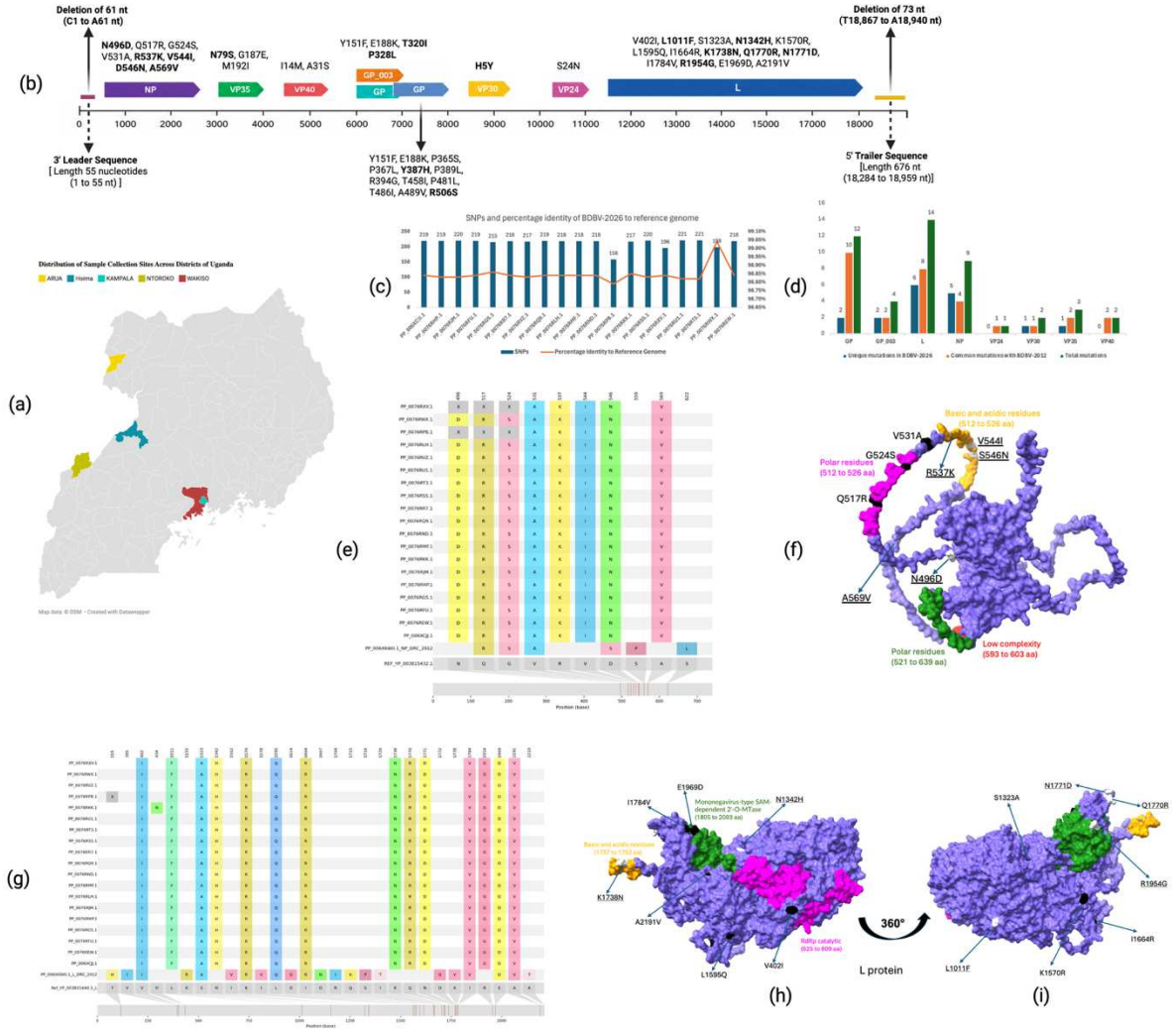


Figure 1. Sequence alignment and mutational profiling of BDBV-2026: (a) Geographical map of Uganda highlighting the sample collection sites. The figure was prepared using Datawrapper software (https://www.datawrapper.de/). (**b)** Cartoon representation of the BDBV genome encoding proteins and 3' leader and 5' trailer regions. Unique substitutions are highlighted in bold. (**c**) Graphical distributions of the SNPs and percentage identity of the BDBV-2026 genomes to the reference genome. (**d**) Graphical representation of the unique, common mutations, and total mutations. Unique, common, and total mutations are highlighted with blue, orange, and green, respectively (x-axis). **(e)** Representation of the multiple sequence alignment (MSA) between the NP protein of BDBV-2026 and BDBV-2012 representative. (**f**) Superimposed 3D structural model of the NP protein of BDBV. Functional domains are indicated with the corresponding colour codes. Unique mutations are highlighted in white, and common mutations are highlighted in black. (**g)** Representation of

MSA between L proteins of BDBV-20226 and BDBV-2012. **(h-i**) Surface representation of the superimposed 3D structural model of the L protein of BDBV. Functional domains are indicated with the corresponding colour codes based on the annotation available from Uniprot. For the structure, common mutations shared with the previous BDBV-2012 outbreak protein sequence are highlighted in black, while unique mutations in the L protein of the ongoing BDBV-2012 outbreak are depicted in white. Superimposed structures were rendered using PyMOL and ChimeraX.

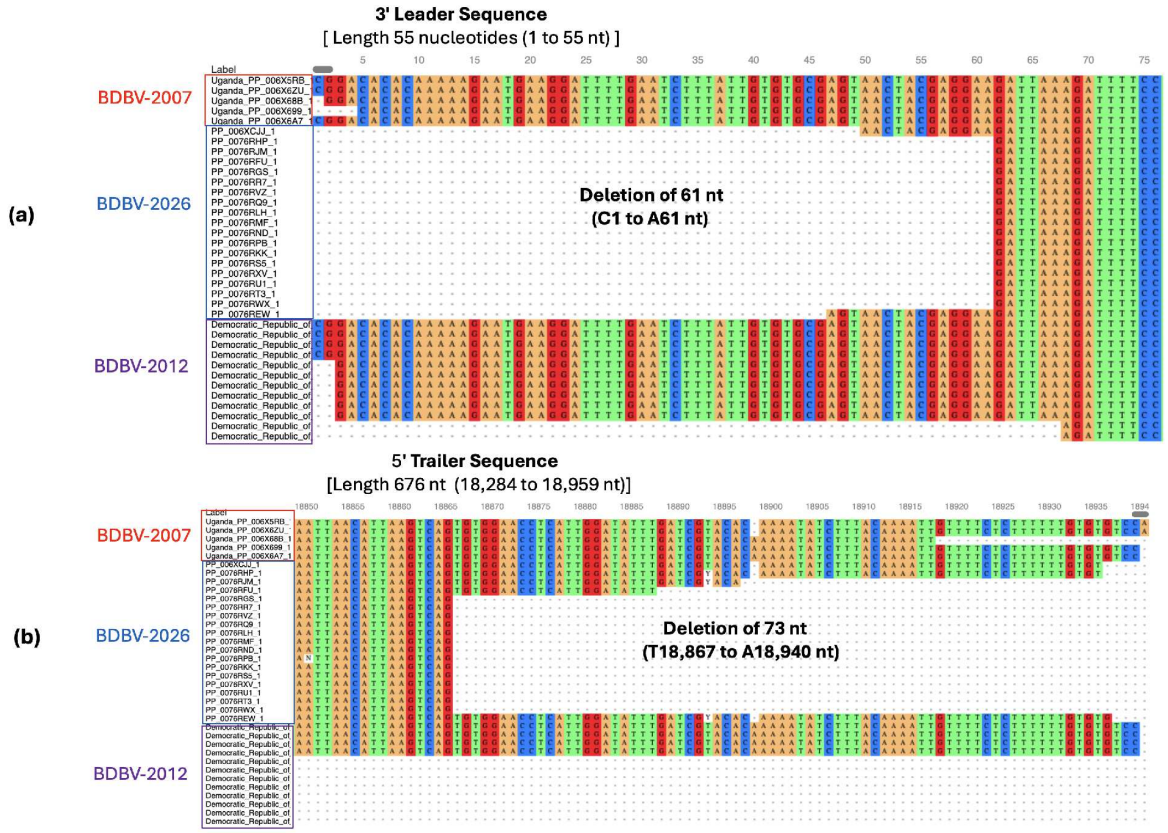


Figure 2. Deletion in 3' leader and 5' trailer sequences: Multiple sequence alignment between BDBV-2007, BDBV-2012, and BDBV-2026 genomes, highlighting the complete deletion of the 3' leader and truncation of the 5' trailer sequences.

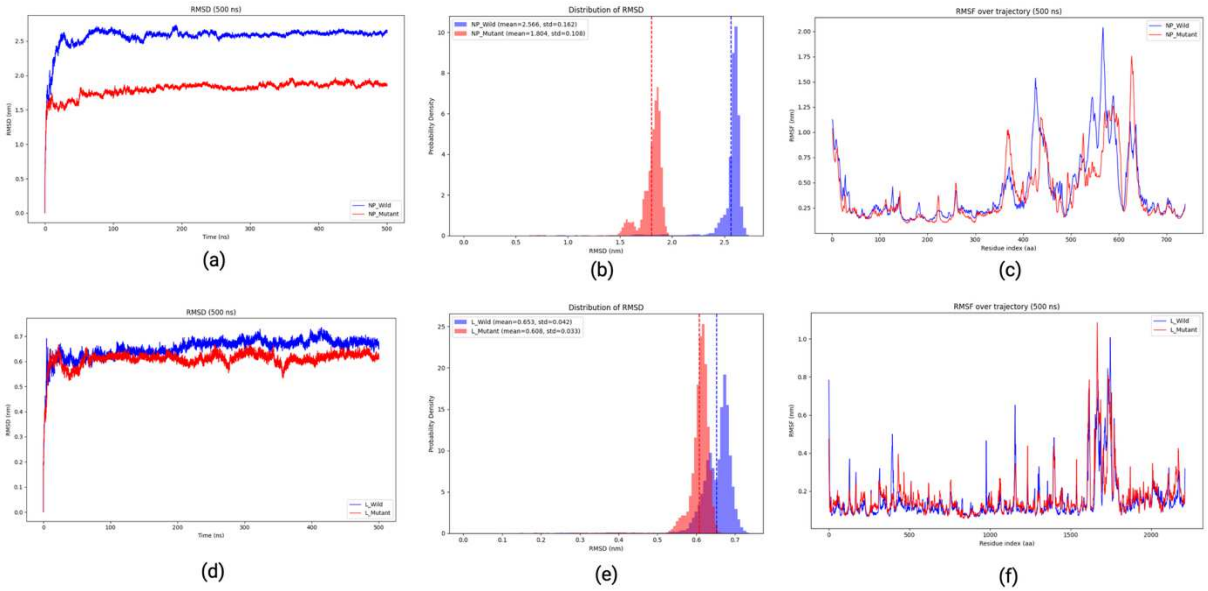
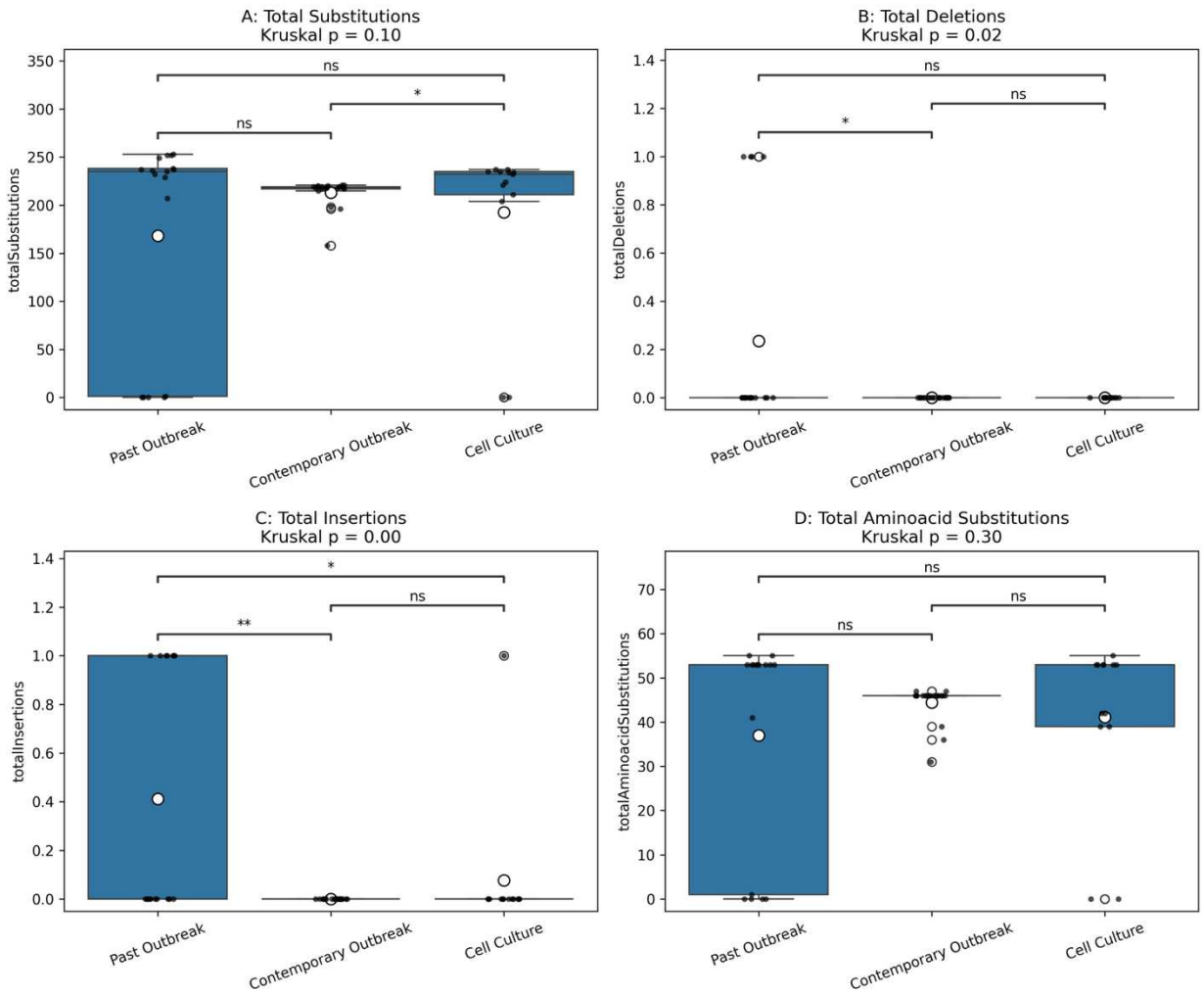


Figure 3. MD simulation for 500 ns: Calculated RMSD (**a** and **d**), RMSD distribution (**b** and **e**), and RMSF plots (**c** and **f**) highlighting the stability pattern of the wild and mutant models of the NP and L proteins. Wild models are highlighted with blue while mutant models are indicated with red.



549

550 **Figure 4. Statistical comparison of mutations in BDBV genomes**. We compared the
551 median of total substitutions (panel A), total deletions (panel B), total insertions (panel C),
552 and total amino acid substitutions (panel D). Omnibus differences were calculated using
553 the Kruskal-Wallis test with η^2 used for measuring omnibus effect size. Independent
554 pairwise comparisons were calculated using Cliff's Delta to estimate pairwise effect sizes.
555 In panel A, no statistical differences were found in omnibus (Kruskal-Wallis $H = 4.6023, p =$
556 $0.1001, \eta^2 = 0.0566$) (past outbreak vs contemporary outbreak $\delta = 0.3127$, past outbreak
557 vs cell culture $\delta = 0.1900$, contemporary outbreak vs cell culture $\delta = -0.4251$). In panel B,
558 statistical differences were found in omnibus (Kruskal-Wallis $H = 8.0314$, $p = 0.02$, $\eta^2 =$
559 0.1311) (past outbreak vs contemporary outbreak $\delta = 0.2353$, past outbreak vs cell culture
560 $\delta = 0.2353$, contemporary outbreak vs cell culture $\delta = 0.00$). In panel C, statistical
561 differences were found in omnibus (Kruskal-Wallis $H = 11.8453, p =< 0.001, \eta^2 = 0.2142$)
562 (past outbreak vs contemporary outbreak $\delta = 0.4118$, past outbreak vs cell culture $\delta =$
563 0.3348, contemporary outbreak vs cell culture $\delta = -0.0769$). In panel D, no statistical

differences were found in omnibus (Kruskal-Wallis $H = 2.4186$, $p = 0.2984$, $\eta^2 = 0.0091$) (past outbreak vs contemporary outbreak $\delta = 0.3127$, past outbreak vs cell culture $\delta = 0.0498$, contemporary outbreak vs cell culture $\delta = -0.1660$).

Figure 5.

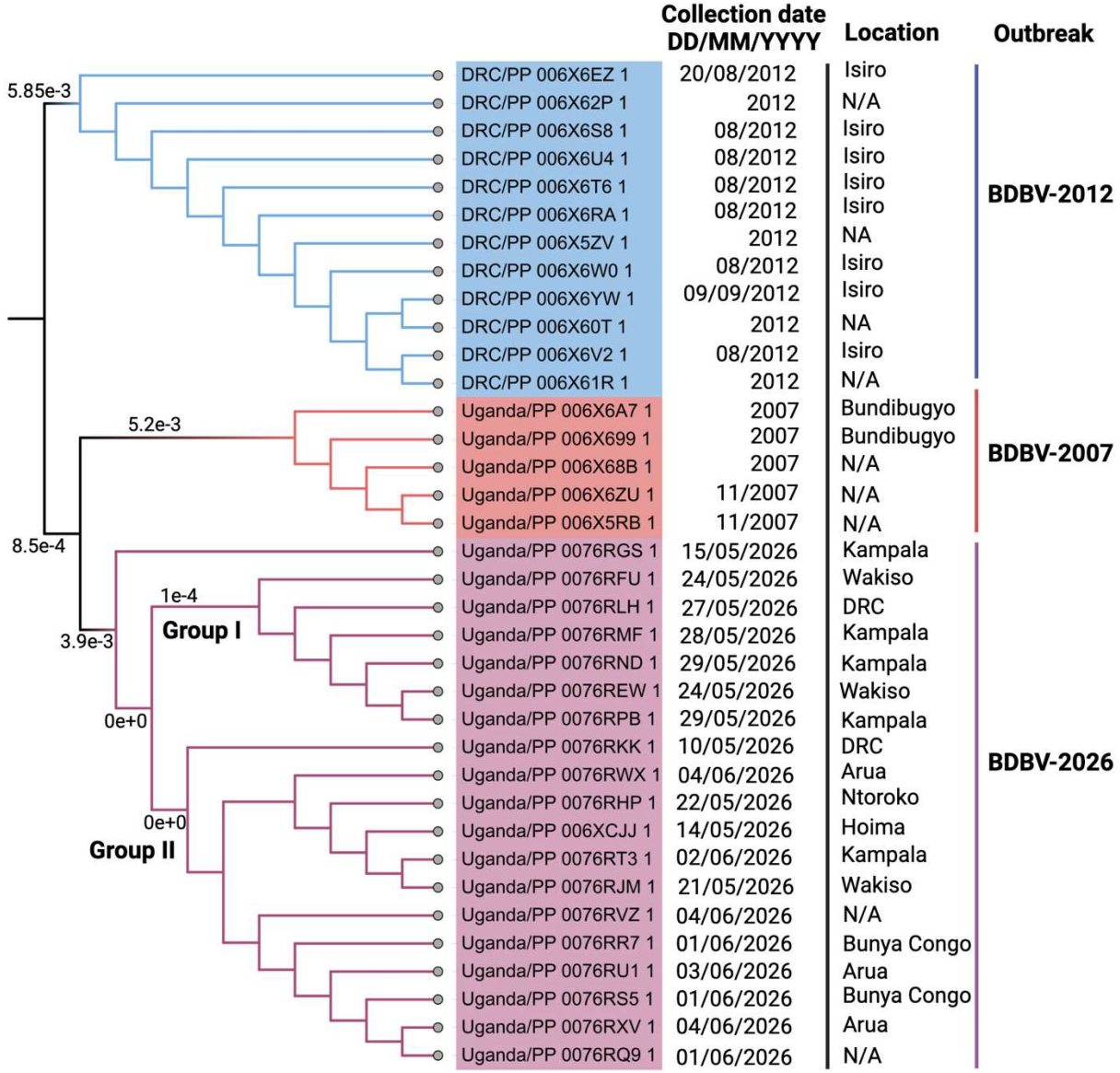


Figure 5. Phylogenetic tree: Rectangular representation of the calculated Neighbour-joining (NJ) phylogenetic tree of the Bundibugyo Ebolavirus (BDBV) sequence derived from the ongoing 2026 outbreak and the previous outbreaks, BDBV-2007 and BDBV-2012. The BDBV-2026 genomes reported from Uganda are classified into two major groups (I and II). The NJ was calculated with 1000 bootstrap values. The metadata regarding collection date and locations are based on the publicly available data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementaryTable1.docx