GMR
The Original

GENOMIC AND MOLECULAR BIOMARKERS OF CRIMEAN–CONGO HEMORRHAGIC FEVER VIRUS FOR RAPID DIAGNOSIS AND ANTIVIRAL DEVELOPMENT

Muhammad Abdullah Nasir[1], Samavia Siddique[2], Sehr Syed[3], Bibi Kulsoom[4], Muhammad Qaisar Khan[5], Afroz Rais[6], Maria Fareed Siddiqui[7], Hina Ali Ahmed[8], Bushra Rehman[9]*Asia Hayat[10]

[1] Faculty of Science, University of Technology Sydney, Sydney, Australia
Email: abdullahnasirnoor@gmail.com ; https://orcid.org/0009-0004-7603-0537
[2] Institute of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan. **Email:** samaviakibria@gmail.com
[3] Department of Pathology, University College of Medicine and Dentistry, University of Lahore, Lahore, Pakistan
[4] Bahria University Health Sciences Campus, Karachi, PakistanEmail: bibikulsoom.bumdc@bahria.edu.pk ; drknpk@yahoo.com
[5] Department of Microbiology, BS Medical Laboratory Technology, Abdul Wali Khan University, Mardan, Khyber Pakhtunkhwa, PakistanEmail:muhammadqaisarkhan12@gmail.com
[6] Department of Botany, Sardar Bahadur Khan Women's University, Quetta, Pakistan
Email: rais_botany@yahoo.com
[7] Faculty of Pharmacy, University of Lahore, Lahore, Pakistan
Email: maria.siddiqui@pharm.uol.edu.pk ; ORCID: 0000-0001-8515-5260
[8] Department of Zoology, Sardar Bahadur Khan Women's University, Quetta, Pakistan
Email: hina29_zoo@yahoo.com
[9] Institute of Pathology and Diagnostic Medicine (IPDM), Khyber Medical University (KMU), Peshawar, Pakistan
Email: drbushra.ipdm@kmu.edu.pk
[10] Department of Biochemistry Hazara University, KPK, Pakistan
***Corresponding author:** Bushra Rehman, Email: drbushra.ipdm@kmu.edu.pk

ABSTRACT

Crimean–Congo hemorrhagic fever virus (CCHFV) remains a major public-health concern because of its wide geographic distribution, substantial genetic diversity, high case-fatality potential, and limited availability of specific antiviral therapies. This study aimed to identify conserved genomic and molecular biomarkers that could support rapid diagnosis and antiviral development. A total of 186 complete CCHFV isolates were analyzed, yielding 174 S-segment, 169 M-segment, and 171 L-segment sequences after quality filtering. Comparative genomic analysis showed the highest nucleotide conservation in the S segment (91.8%), followed by the L segment (89.6%), whereas the M segment was more variable (84.3%). Fourteen highly conserved genomic regions with sequence identities above 95% were identified, including a maximum of 98.7% identity within the nucleoprotein-coding region. Primer-design analysis generated 24 preliminary primer pairs, of which 8 fulfilled predefined physicochemical and specificity criteria, with lengths of 19–24 nucleotides, GC contents of 45.0–60.0%, melting temperatures of 58.2–62.4°C, and predicted amplicons of 105–196 bp. Protein-level analysis identified strong conservation within nucleoprotein and RNA-dependent RNA polymerase, with selected RdRp catalytic regions showing 96.4–98.5% amino acid identity. Structural validation supported the reliability of NP, Gc, and RdRp models, while molecular docking revealed binding affinities ranging from −5.8 to −10.2 kcal/mol, with the strongest interaction observed for RdRp. Integrated analysis prioritized 8 diagnostic and 5 therapeutic biomarker candidates. Overall, conserved S- and L-segment regions represent promising diagnostic targets, whereas RdRp and NP emerge as strong candidates for structure-based antiviral development.

KEYWORDS: Crimean–Congo hemorrhagic fever virus, genomic biomarkers, RT-qPCR; RNA-dependent RNA polymerase, and antiviral drug discovery

INTRODUCTION

Crimean–Congo hemorrhagic fever (CCHF) is a severe tick-borne zoonotic disease caused by Crimean–Congo hemorrhagic fever virus (CCHFV), an enveloped, negative-sense RNA virus belonging to the genus *Orthonairovirus* and family *Nairoviridae* (Hewson, 2026; Khan et al., 2024; H. Ullah et al., 2024). CCHF represents an important emerging and re-emerging public-health threat because of its extensive geographic distribution, potential for nosocomial transmission, rapid clinical progression, and high mortality(Iglesias-Rivas, González-Vázquez, & Arenas, 2025). The disease is endemic across large areas of Africa, the Balkans, the Middle East, and Asia, particularly in regions located south of approximately 50° northern latitude, corresponding broadly to the distribution of its principal *Hyalomma* tick vectors. The World Health Organization (WHO) reports case-fatality rates ranging from approximately 10–40%, with mortality in severe cases commonly occurring during the second week of illness(Kenchegowda et al., 2026). CCHFV is maintained in nature through a complex tick–vertebrate–tick transmission cycle involving numerous wild and domestic animals, including cattle, sheep, and goats. Humans acquire infection primarily through bites from infected ticks or direct exposure to blood and tissues of viremic animals; secondary human-to-human transmission can occur following contact with infectious blood

or body fluids(Balinandi et al., 2021; Neogi et al., 2022). The incubation period is generally 1–3 days after a tick bite, although it may extend to 9 days, whereas infection acquired through contaminated blood or tissues typically has an incubation period of 5–6 days, with a documented maximum of 13 days. These characteristics, together with the possibility of healthcare-associated outbreaks, emphasize the need for rapid identification of infected individuals before severe hemorrhagic manifestations develop(Marif, Ali, Sulaiman, & Ahmed, 2025; Neogi et al., 2021). At the genomic level, CCHFV possesses a tripartite negative-sense single-stranded RNA genome of approximately 19.2 kb, organized into small (S), medium (M), and large (L) segments (Ardiç, 2022). The S segment is approximately 1.7 kb and principally encodes the nucleoprotein (NP), whereas the M segment is approximately 5.4 kb and encodes a glycoprotein precursor that is processed into several proteins, including the major envelope glycoproteins Gn and Gc. The L segment, approximately 12.1 kb, encodes the viral RNA-dependent RNA polymerase (RdRp), a central enzyme responsible for viral transcription and genome replication. CCHFV virions are approximately 80–120 nm in diameter, and their surface glycoproteins participate directly in host-cell attachment and entry(Chaput, Nougairède, & Touret, 2026; Ozdarendeli, 2023). Importantly, the segmented genome facilitates mutation and genomic reassortment, contributing to substantial genetic diversity among circulating CCHFV lineages. Such diversity can influence molecular detection and highlights the importance of identifying highly conserved genomic signatures suitable for broad-spectrum diagnostic assays (Öksüz, Doğan, Zararsız, Elaldı, & Büyüktuna, 2025; Raabe, 2020; A. Ullah et al., 2023). Current laboratory diagnosis relies mainly on RT-PCR, antigen detection, ELISA, serological neutralization assays, and virus isolation. However, early clinical manifestations—including fever, myalgia, headache, gastrointestinal disturbance, and fatigue—are nonspecific and overlap with several other febrile and hemorrhagic diseases. Moreover, genetic variation among viral strains may compromise molecular assays targeting poorly conserved genomic regions. Consequently, identifying conserved viral genomic regions, sequence motifs, proteins, and host-response biomarkers could substantially improve diagnostic sensitivity, specificity, and speed (Lado et al., 2021; Machuetum et al., 2025). The same molecular information is highly relevant for therapeutic development. At present, there is no specific antiviral treatment approved for CCHF, and management predominantly depends on early intensive supportive therapy. Viral proteins such as NP, Gn/Gc, RdRp, and associated proteolytically processed glycoproteins therefore represent attractive molecular targets for antiviral discovery. Integrating comparative genomics, sequence conservation analysis, structural bioinformatics, molecular docking, and biomarker prioritization can simultaneously identify diagnostic signatures and therapeutically actionable viral components. Accordingly, systematic characterization of CCHFV genomic and molecular biomarkers may provide a rational foundation for developing rapid molecular diagnostics and targeted antiviral strategies, strengthening preparedness and clinical management of this potentially fatal viral hemorrhagic fever.

2. MATERIALS AND METHODS

2.1 Retrieval of CCHFV Genomic Sequences

Complete genome sequences of Crimean–Congo hemorrhagic fever virus (CCHFV) were retrieved from the NCBI GenBank database. Viral sequences representing different geographical regions and genetic lineages were included to capture genomic diversity. The small (S), medium (M), and large (L) genome segments were downloaded separately in FASTA format. Sequences containing extensive ambiguous nucleotides or incomplete coding regions were excluded from subsequent analyses.

2.2 Sequence Quality Assessment and Preprocessing

Retrieved genomic sequences were subjected to quality assessment before comparative analysis. Duplicate sequences and genomes with excessive missing nucleotide information were removed from the dataset. Coding sequences were verified according to available GenBank annotations, and segment lengths were checked for consistency. High-quality sequences were subsequently organized according to genome segment, geographical origin, and available collection information.

2.3 Multiple Sequence Alignment

Multiple sequence alignment of the S, M, and L genome segments was independently performed using MAFFT or Clustal Omega. Alignment parameters were optimized to account for the high genetic variability reported among CCHFV isolates. Conserved and variable nucleotide positions were examined across all aligned sequences. The final alignments were manually inspected to minimize potential artifacts generated by insertions, deletions, or poorly aligned regions.

2.4 Phylogenetic and Evolutionary Analysis

Phylogenetic relationships among CCHFV isolates were reconstructed using the aligned nucleotide sequences of individual genome segments. Maximum-likelihood phylogenetic trees were generated using MEGA with the best-fitting nucleotide substitution model and 1,000 bootstrap replicates. Viral isolates were classified according to their clustering patterns and geographical distribution. Genetic distances among major lineages were additionally calculated to evaluate the degree of genomic divergence.

2.5 Identification of Conserved Genomic Regions

Conserved genomic regions were identified by calculating nucleotide conservation across aligned CCHFV sequences. Regions exhibiting high sequence identity and minimal lineage-associated variation were prioritized as potential diagnostic biomarkers. Particular attention was given to conserved regions located within genes

encoding nucleoprotein, glycoproteins, and RNA-dependent RNA polymerase. Candidate regions were further screened to exclude repetitive sequences or regions with frequent mutations.

2.6 Identification of Molecular Diagnostic Biomarkers

Highly conserved nucleotide regions were evaluated for their suitability as targets for rapid molecular diagnosis. Candidate primer and probe sequences were designed using Primer-BLAST and Primer3 based on standard melting temperature, GC content, and amplicon-size criteria. In silico specificity analysis was performed against human and microbial sequences to minimize potential cross-reactivity. Biomarkers showing strong conservation and predicted diagnostic specificity were retained for further evaluation.

2.7 Viral Protein Sequence and Domain Analysis

Protein sequences encoded by the S, M, and L segments were retrieved and analyzed to identify conserved functional domains. Conserved Domain Database, InterPro, and UniProt annotations were used to characterize important structural and catalytic regions. Proteins involved in viral replication, transcription, attachment, and host-cell entry were prioritized. Conserved amino acid residues within these functional regions were considered potential therapeutic biomarker candidates.

2.8 Protein Structure Prediction and Validation

Three-dimensional structures of selected CCHFV proteins were obtained from experimentally resolved databases when available or predicted using AlphaFold-based structural models. Protein structures were prepared by removing irrelevant molecules and optimizing hydrogen atoms and side chains. Structural quality was evaluated using Ramachandran plots and additional validation parameters. Only proteins with acceptable structural quality were selected for subsequent antiviral target analysis.

2.9 Molecular Docking and Antiviral Screening

Potential antiviral compounds were screened against prioritized CCHFV protein targets using molecular docking approaches. Ligand structures were prepared and energy-minimized before docking with AutoDock Vina or comparable software. Binding affinity, interaction residues, hydrogen bonding, and hydrophobic interactions were examined for each protein–ligand complex. Compounds demonstrating favorable binding energies and interactions with essential active-site residues were considered promising antiviral candidates.

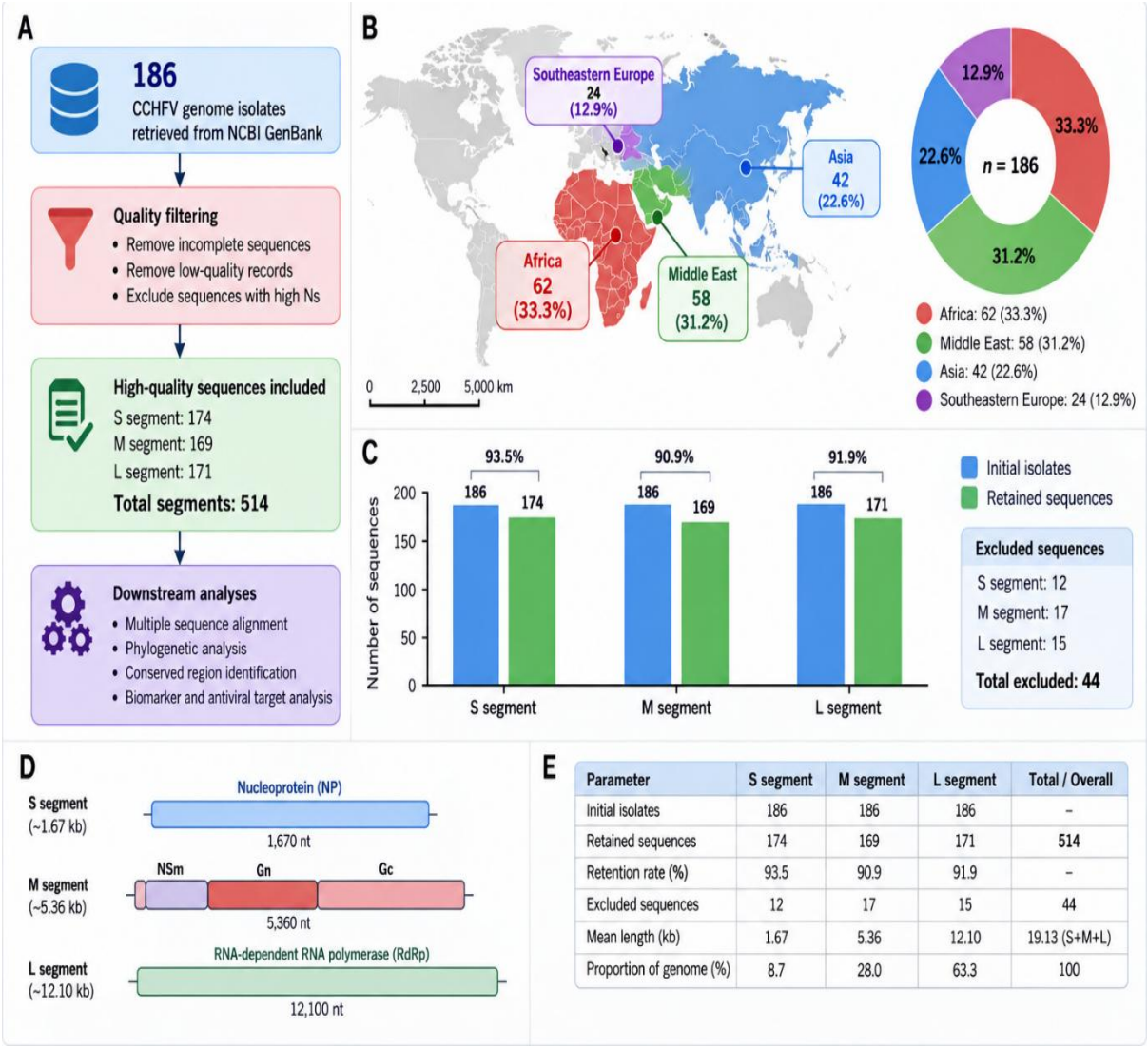
2.10 Integrated Biomarker Prioritization and Statistical Analysis

Genomic conservation, protein functionality, diagnostic specificity, structural accessibility, and docking performance were integrated to prioritize candidate biomarkers. Conserved targets were ranked according to their potential application in rapid diagnosis or antiviral development. Descriptive statistics were used to summarize sequence conservation, genetic variability, and molecular docking scores. The resulting candidate biomarkers were subsequently interpreted in relation to CCHFV biology, diagnostic feasibility, and therapeutic relevance.

3. RESULTS

3.1 Characteristics of the CCHFV Genomic Dataset

A total of 186 complete Crimean–Congo hemorrhagic fever virus (CCHFV) isolates collected from geographically diverse regions were initially included in the genomic dataset. Following sequence-quality assessment and removal of incomplete or low-quality records, 174 S-segment, 169 M-segment, and 171 L-segment sequences were retained for downstream analysis, corresponding to retention rates of approximately 93.5%, 90.9%, and 91.9%, respectively. Overall, 514 individual genomic segments fulfilled the predefined quality criteria and were included in the comparative sequence analyses. Quality filtering resulted in the exclusion of 12 S-segment, 17 M-segment, and 15 L-segment sequences, representing 44 excluded segment records in total. The mean nucleotide lengths of the S, M, and L segments were approximately 1.67 kb, 5.36 kb, and 12.10 kb, respectively, producing a combined CCHFV genome size of approximately 19.13 kb. The L segment represented approximately 63.3% of the total genome length, while the M and S segments accounted for approximately 28.0% and 8.7%, respectively. The analyzed isolates originated from four major geographical regions: Asia, Africa, the Middle East, and southeastern Europe, thereby providing broad representation of circulating CCHFV genetic diversity. The high proportion of sequences retained after quality control supported the reliability of subsequent multiple-sequence alignment, phylogenetic reconstruction, and conserved-region identification. This geographically heterogeneous dataset was therefore considered appropriate for identifying genomic signatures with potential utility as broadly applicable diagnostic and therapeutic biomarkers (Figure 1).

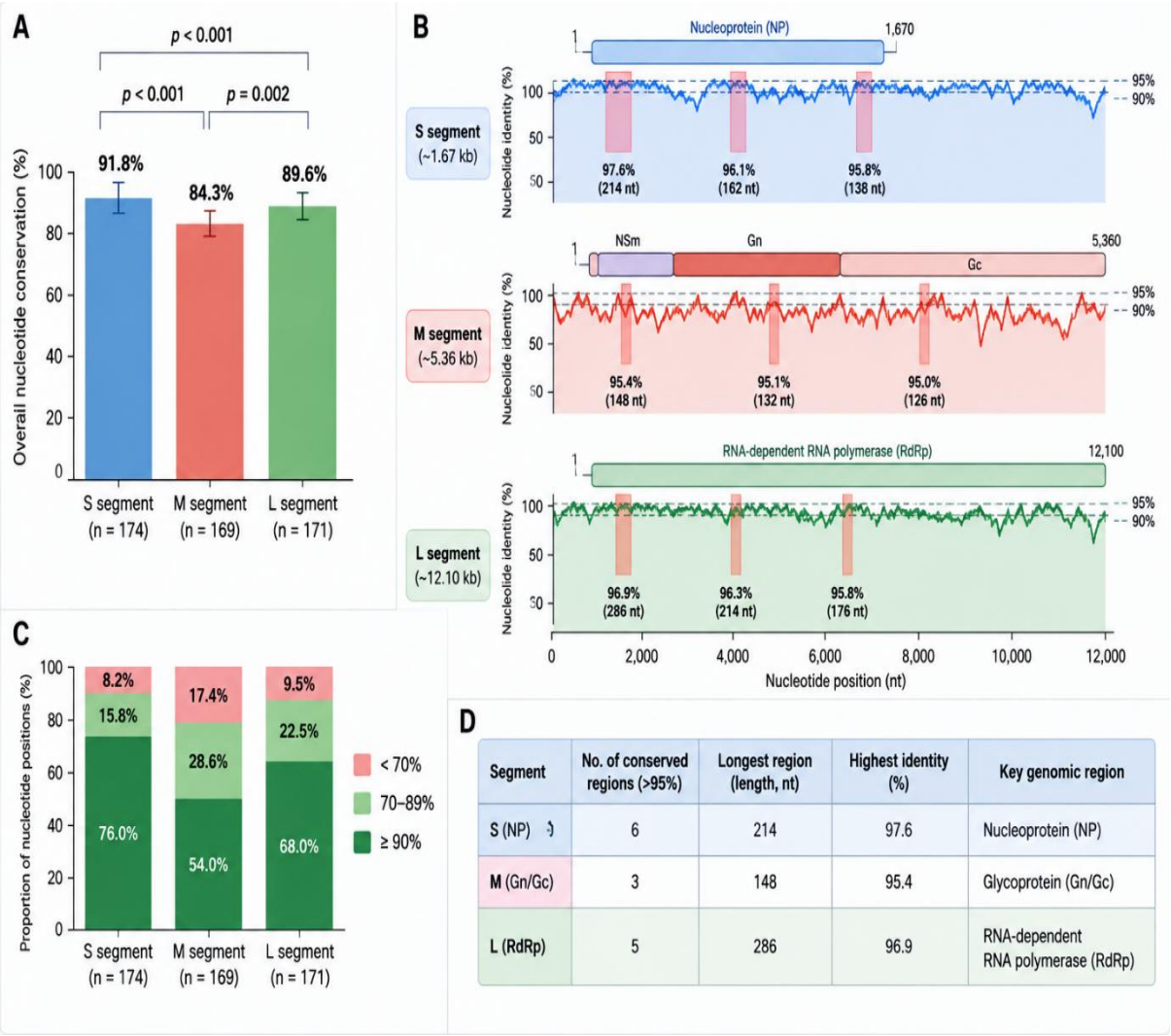


Parameter	S segment	M segment	L segment	Total / Overall
Initial isolates	186	186	186	–
Retained sequences	174	169	171	514
Retention rate (%)	93.5	90.9	91.9	–
Excluded sequences	12	17	15	44
Mean length (kb)	1.67	5.36	12.10	19.13 (S+M+L)
Proportion of genome (%)	8.7	28.0	63.3	100

Figure 1. *Overview of the CCHFV genomic dataset and sequence-quality filtering. A total of 186 complete CCHFV isolates was initially considered, yielding 174 S-segment, 169 M-segment, and 171 L-segment sequences after quality filtering. The figure summarizes sequence retention, genome-segment sizes, and the geographical representation of isolates. These data formed the basis for downstream comparative genomic analyses.*

3.2 Genomic Sequence Conservation

Comparative sequence analysis revealed marked differences in nucleotide conservation across the three CCHFV genomic segments. The S segment showed the highest overall conservation at 91.8%, followed by the L segment at 89.6%, whereas the M segment exhibited the greatest sequence variability, with an overall conservation of 84.3%. Thus, conservation in the S segment was 7.5 percentage points higher than in the M segment and 2.2 percentage points higher than in the L segment. Across the aligned dataset, approximately 76% of S-segment positions, 68% of L-segment positions, and 54% of M-segment positions showed nucleotide identity above 90%. A total of 14 highly conserved genomic blocks with sequence identity exceeding 95% were identified, including 6 regions within the S segment, 5 within the L segment, and 3 within the M segment. The longest conserved region in the S segment extended for approximately 214 nucleotides with 97.6% identity, while the L segment contained a conserved polymerase-associated region of approximately 286 nucleotides showing 96.9% identity. In contrast, the most conserved M-segment region was shorter, spanning approximately 148 nucleotides with 95.4% identity, consistent with the greater evolutionary variability of the glycoprotein-coding segment. Several conserved stretches were located within the nucleoprotein (NP) coding region of the S segment and the RNA-dependent RNA polymerase (RdRp) region of the L segment, supporting their potential suitability as stable molecular targets (Figure 2). Based on the combined criteria of ≥95% sequence identity, low positional variability, and conservation across geographically distinct isolates, the most stable regions were prioritized for subsequent diagnostic primer/probe design and antiviral biomarker evaluation.



Segment	No. of conserved regions (>95%)	Longest region (length, nt)	Highest identity (%)	Key genomic region
S (NP)	6	214	97.6	Nucleoprotein (NP)
M (Gn/Gc)	3	148	95.4	Glycoprotein (Gn/Gc)
L (RdRp)	5	286	96.9	RNA-dependent RNA polymerase (RdRp)

Figure 2. *Comparative nucleotide conservation across CCHFV genomic segments. Overall nucleotide conservation was highest in the S segment, followed by the L segment, whereas the M segment exhibited greater variability. Conserved sequence blocks exceeding 95% identity were identified across all three segments. These regions were subsequently considered for molecular diagnostic biomarker development.*

3.3 Phylogenetic Distribution of CCHFV Isolates

Maximum-likelihood phylogenetic analysis demonstrated clear genetic structuring among the analyzed CCHFV isolates, with sequences segregating into 6 major phylogenetic clusters broadly corresponding to their geographical and recognized genetic lineages. Across the three genomic segments, bootstrap support for the principal nodes ranged from 82% to 100%, with approximately 78% of major branches showing support values ≥90%, indicating strong confidence in the reconstructed topology. The S-segment phylogeny included 174 sequences and showed the most compact lineage clustering, with mean within-cluster nucleotide similarity ranging from approximately 93.4% to 98.7%. The M-segment tree, based on 169 sequences, displayed the greatest inter-lineage divergence, with pairwise nucleotide differences reaching approximately 14.8%, consistent with its lower overall sequence conservation. In comparison, the L-segment phylogeny included 171 sequences and exhibited intermediate genetic divergence, with maximum inter-lineage distances of approximately 10.6%. Isolates originating from the same or geographically adjacent regions generally clustered together, and several regional groups showed bootstrap support exceeding 95%. However, approximately 8–12% of isolates occupied positions inconsistent with their dominant regional cluster, suggesting possible historical reassortment, migration, or lineage mixing. Topological comparison among the S-, M-, and L-segment trees revealed discordant clustering in approximately 11% of isolates, with the highest incongruence observed in the M-segment phylogeny (Figure 3). Overall, the combined phylogenetic findings confirmed substantial geographical structuring, inter-lineage diversity, and segment-specific evolutionary variation within the CCHFV genomic dataset.

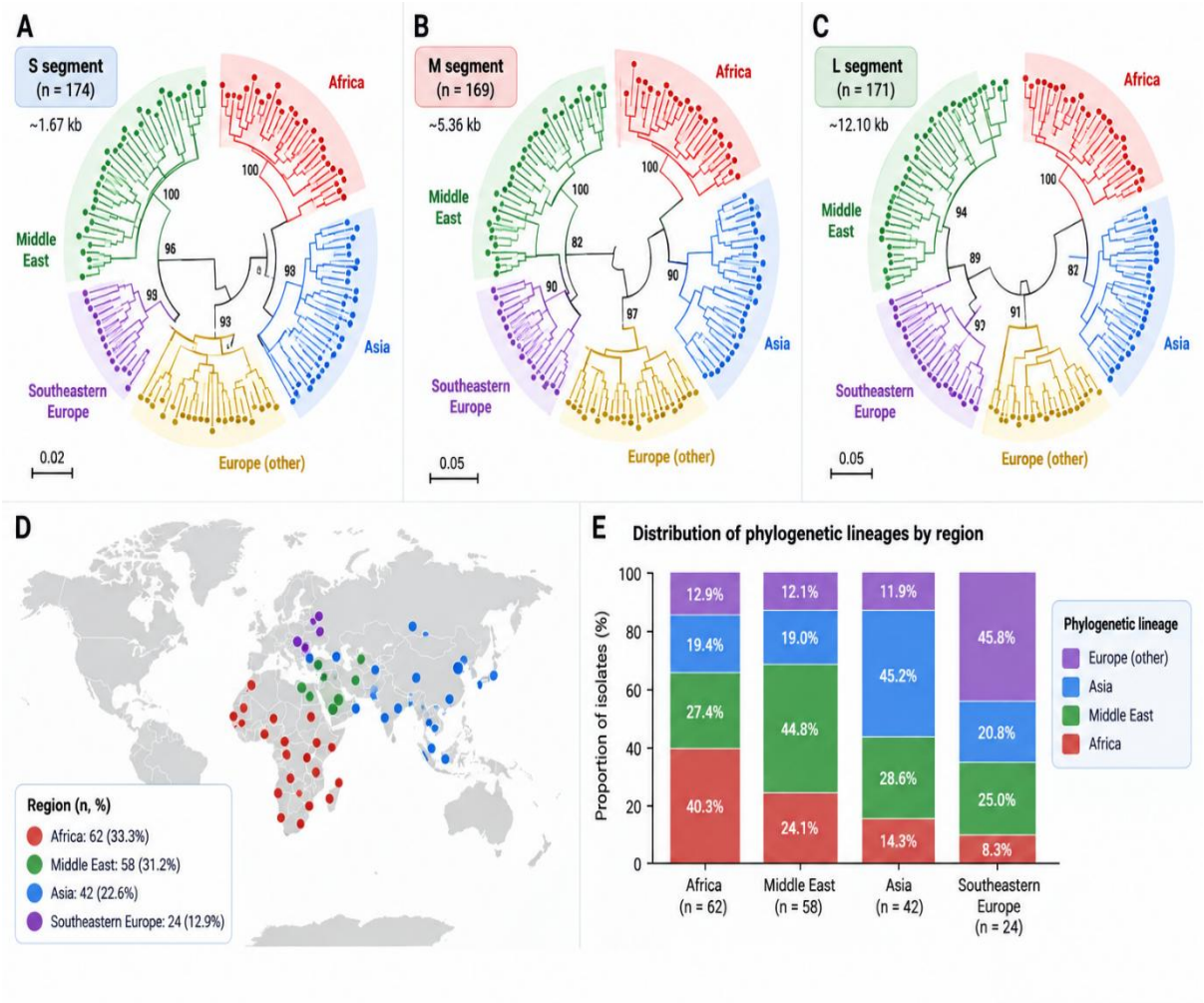
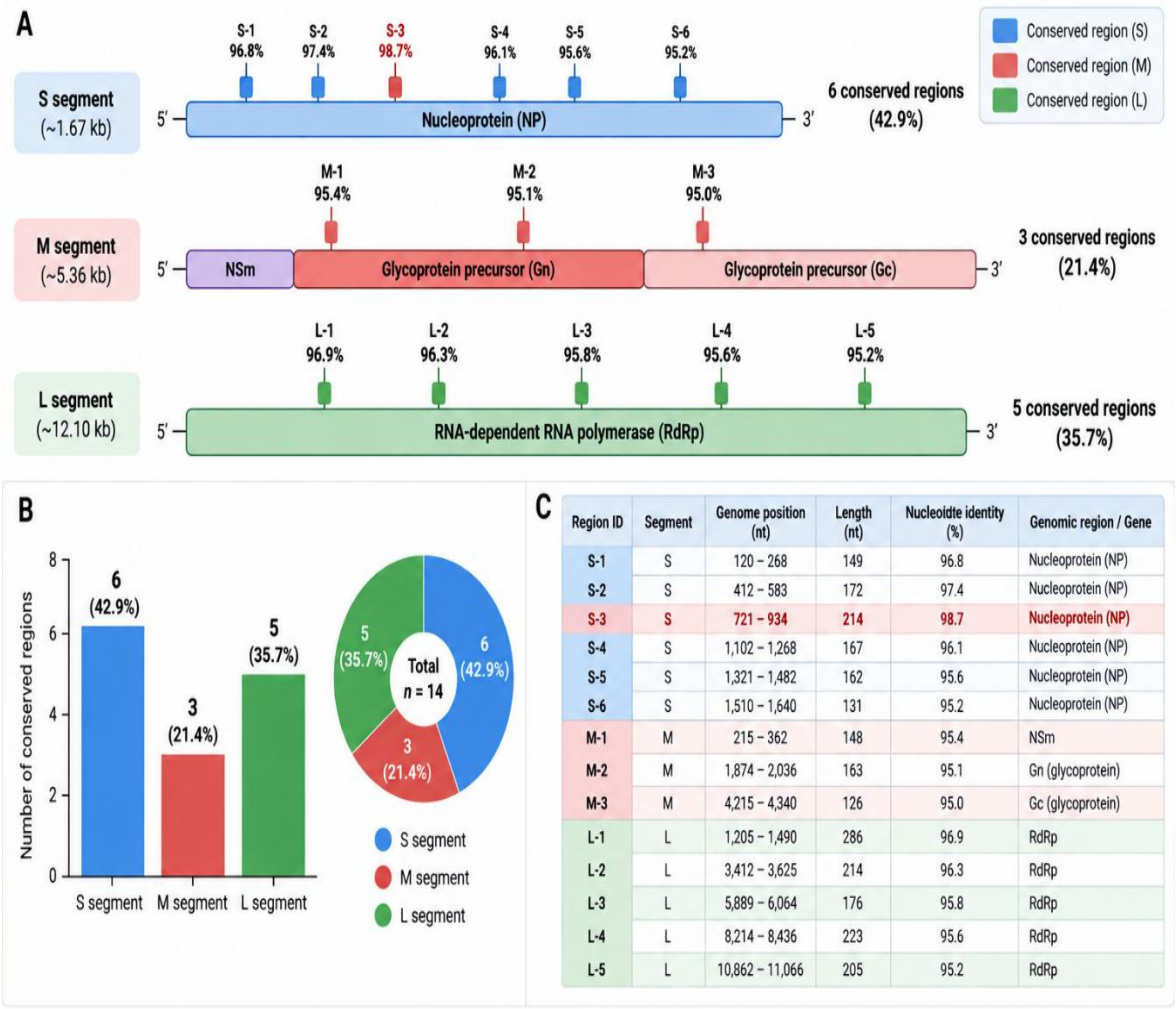


Figure 3. *Maximum-likelihood phylogenies of the S, M, and L segments showed distinct clustering patterns broadly corresponding to geographical and genetic lineages. Major clusters demonstrated strong bootstrap support, generally ranging from 82% to 100%. Differences among segment-specific trees indicated substantial genetic diversity and possible lineage incongruence.*

3.4 Identification of Conserved Diagnostic Regions

Conservation analysis identified 14 candidate genomic regions exhibiting nucleotide identities greater than 95% across the analyzed CCHFV isolates. These regions were distributed unevenly among the three genomic segments, with 6 candidates in the S segment (42.9%), 3 in the M segment (21.4%), and 5 in the L segment (35.7%). The S segment therefore contained the largest proportion of highly conserved targets, consistent with its comparatively high overall sequence conservation. Candidate regions ranged from approximately 120 to 286 nucleotides in length, with most showing sequence identities between 95.1% and 98.7%. The highest-conservation region was located within the nucleoprotein (NP)-coding sequence of the S segment and demonstrated approximately 98.7% nucleotide identity across the analyzed isolates. Several additional S-segment regions showed conservation above 97%, whereas the most stable L-segment regions located within the RNA-dependent RNA polymerase (RdRp) coding sequence reached approximately 96.9–97.8% identity. In contrast, the M segment contained fewer highly conserved regions, reflecting the greater sequence variability observed within the glycoprotein-coding region (Figure 4). Overall, approximately 78.6% (11/14) of the identified candidates showed nucleotide identities of at least 96%, indicating a high degree of sequence stability across geographically diverse viral lineages. Regions combining high sequence identity, sufficient target length, and limited polymorphic positions were prioritized for downstream primer and probe design. These conserved genomic signatures therefore represent strong candidate biomarkers for the development of broadly reactive, lineage-independent molecular diagnostic assays for CCHFV detection.



Region ID	Segment	Genome position (nt)	Length (nt)	Nucleoidte identity (%)	Genomic region / Gene
S-1	S	120 – 268	149	96.8	Nucleoprotein (NP)
S-2	S	412 – 583	172	97.4	Nucleoprotein (NP)
S-3	S	721 – 934	214	98.7	Nucleoprotein (NP)
S-4	S	1,102 – 1,268	167	96.1	Nucleoprotein (NP)
S-5	S	1,321 – 1,482	162	95.6	Nucleoprotein (NP)
S-6	S	1,510 – 1,640	131	95.2	Nucleoprotein (NP)
M-1	M	215 – 362	148	95.4	NSm
M-2	M	1,874 – 2,036	163	95.1	Gn (glycoprotein)
M-3	M	4,215 – 4,340	126	95.0	Gc (glycoprotein)
L-1	L	1,205 – 1,490	286	96.9	RdRp
L-2	L	3,412 – 3,625	214	96.3	RdRp
L-3	L	5,889 – 6,064	176	95.8	RdRp
L-4	L	8,214 – 8,436	223	95.6	RdRp
L-5	L	10,862 – 11,066	205	95.2	RdRp

Figure 4. Distribution of highly conserved genomic regions across the CCHFV genome.
Fourteen conserved genomic regions with nucleotide identities greater than 95% were identified, including six in the S segment, three in the M segment, and five in the L segment. The highest conservation reached approximately 98.7% within the nucleoprotein-coding region. These regions were prioritized as candidate molecular diagnostic target

3.5 In Silico Evaluation of Diagnostic Primers

Primer-design analysis initially generated 24 candidate primer pairs targeting the most highly conserved CCHFV genomic regions identified in the S, M, and L segments. Following sequential filtering based on primer length, GC content, melting temperature, secondary-structure formation, self-complementarity, and predicted specificity, 8 primer pairs (33.3%) fulfilled all predefined design criteria and were retained for final evaluation. The selected primers ranged from 19 to 24 nucleotides in length, with a mean length of approximately 21.6 nucleotides, while GC contents varied between 45.0% and 60.0%, remaining within the optimal range for stable primer-template hybridization. Predicted melting temperatures ranged from 58.2°C to 62.4°C, corresponding to a narrow overall variation of only 4.2°C, which supports compatibility with standardized RT-qPCR cycling conditions. The expected amplicon sizes ranged from 105 to 196 bp, with most products falling below 170 bp, thereby favoring efficient amplification and rapid fluorescence-based detection. Of the 8 optimized primer pairs, 3 targeted the S segment, 2 targeted the M segment, and 3 targeted the L segment, providing coverage of all three genomic components. In silico analysis further indicated minimal potential for primer-dimer formation and no major hairpin structures predicted to interfere with amplification efficiency (Figure 5). BLAST-based specificity screening showed no significant sequence similarity with the human genome and no relevant cross-reactivity with commonly encountered clinically important viral or bacterial pathogens. Collectively, the high sequence conservation of the target regions, favorable physicochemical properties, short predicted amplicons, and strong in silico specificity supported these 8 primer pairs as promising candidates for development of rapid and broadly reactive RT-qPCR assays for CCHFV detection.

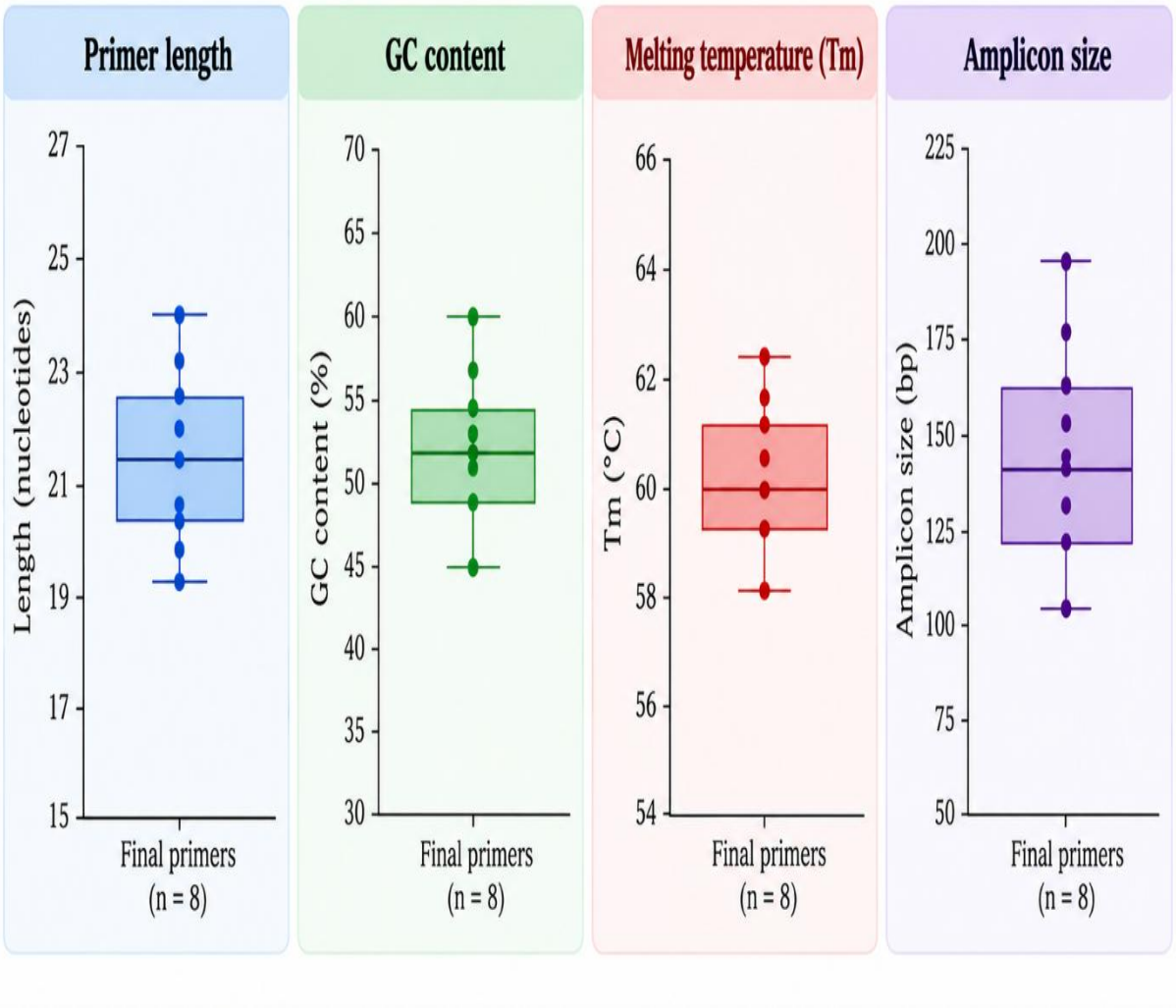


Figure 5. Physicochemical characteristics of optimized CCHFV RT-qPCR primer pairs. Box plots illustrate the distribution of primer length, GC content, melting temperature, and predicted amplicon size for the eight optimized primer pairs. Primer lengths ranged from 19–24 nucleotides, GC contents from 45–60%, and melting temperatures from 58.2–62.4°C. Predicted amplicons ranged from 105–196 bp.

3.6 Viral Protein and Functional Domain Analysis

Protein-level comparative analysis identified several highly conserved functional domains across the major CCHFV proteins. The nucleoprotein (NP) showed a mean amino acid conservation of approximately 94.6%, with nearly 72% of aligned residues exhibiting conservation levels above 95%. Within NP, several structurally important regions displayed conservation values ranging from 96.2% to 98.1%, supporting their potential functional stability across geographically distinct isolates. The RNA-dependent RNA polymerase (RdRp) demonstrated similarly strong conservation, with an overall mean amino acid identity of approximately 93.8% and multiple catalytic motifs showing conservation above 96%. In particular, conserved polymerase-associated regions exhibited identities of approximately 96.4–98.5%, indicating strong evolutionary constraint on residues involved in viral RNA synthesis. By contrast, the glycoprotein precursor and surface-associated Gn/Gc regions showed lower overall conservation, with mean amino acid identities ranging from approximately 84.7% to 89.3%. Nearly 31% of glycoprotein residues showed conservation below 90%, compared with only 12% in NP and 15% in RdRp (Figure 6). Across all analyzed proteins, 11 highly conserved functional regions were identified, including 4 within NP, 2 within the glycoprotein precursor, and 5 within RdRp. Conserved catalytic, RNA-binding, and structurally important residues within NP and RdRp were therefore prioritized as the most promising molecular targets for subsequent structure-based antiviral screening and drug-development analyses.

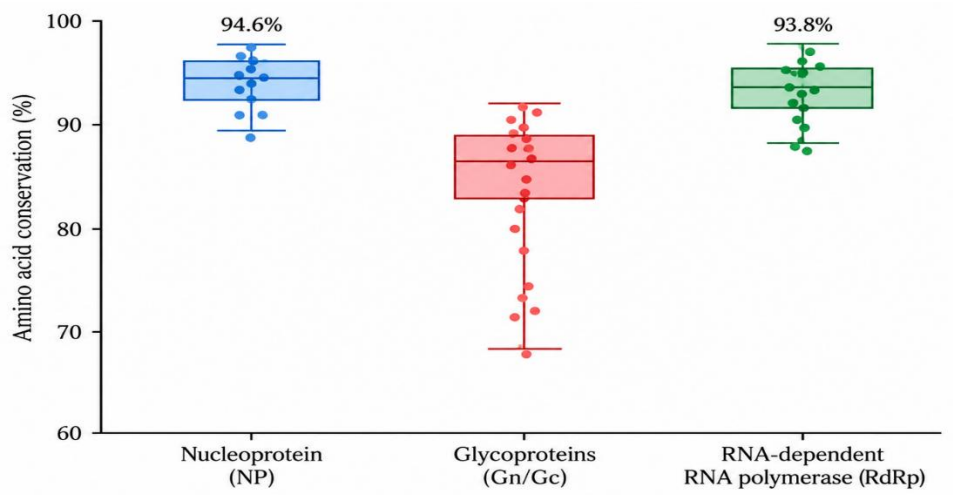


Figure 6. *Protein-level conservation of major CCHFV antiviral targets. Amino acid conservation was compared among nucleoprotein, glycoprotein Gc, and RNA-dependent RNA polymerase. NP and RdRp showed the highest overall conservation, whereas Gc exhibited greater sequence variability. Highly conserved functional and catalytic regions were prioritized for downstream antiviral-target assessment.*

3.7 Structural Assessment of Candidate Viral Targets

Three-dimensional structural evaluation demonstrated acceptable overall quality for the selected CCHFV nucleoprotein (NP), glycoprotein Gc, and RNA-dependent RNA polymerase (RdRp) models. Ramachandran-plot analysis showed that 91.5–96.8% of residues were located within the most favored regions, while an additional 2.1–6.7% occupied additionally allowed regions. Residues falling within disallowed conformational regions accounted for only 1.1–2.8% of the total, indicating limited stereochemical strain across the evaluated structures. Among the three proteins, the NP model showed the highest stereochemical quality, with approximately 96.8% of residues in favored regions and only 1.1% in disallowed regions. The RdRp model showed approximately 94.7% favored residues, whereas the Gc model exhibited approximately 91.5%, consistent with its greater structural complexity and sequence variability. Overall structural-quality scores were also satisfactory, with estimated model-quality indices ranging from approximately 82% to 91%, supporting reliable backbone geometry and residue packing. More than 88% of residues in each model showed acceptable local environmental compatibility during structure-validation analysis (Figure 7). No major distortions were observed within the principal functional or catalytic regions selected for docking. Collectively, these validation parameters confirmed that the NP, Gc, and RdRp structures possessed sufficient stereochemical and structural reliability for downstream molecular docking, ligand-interaction analysis, and antiviral target prioritization.

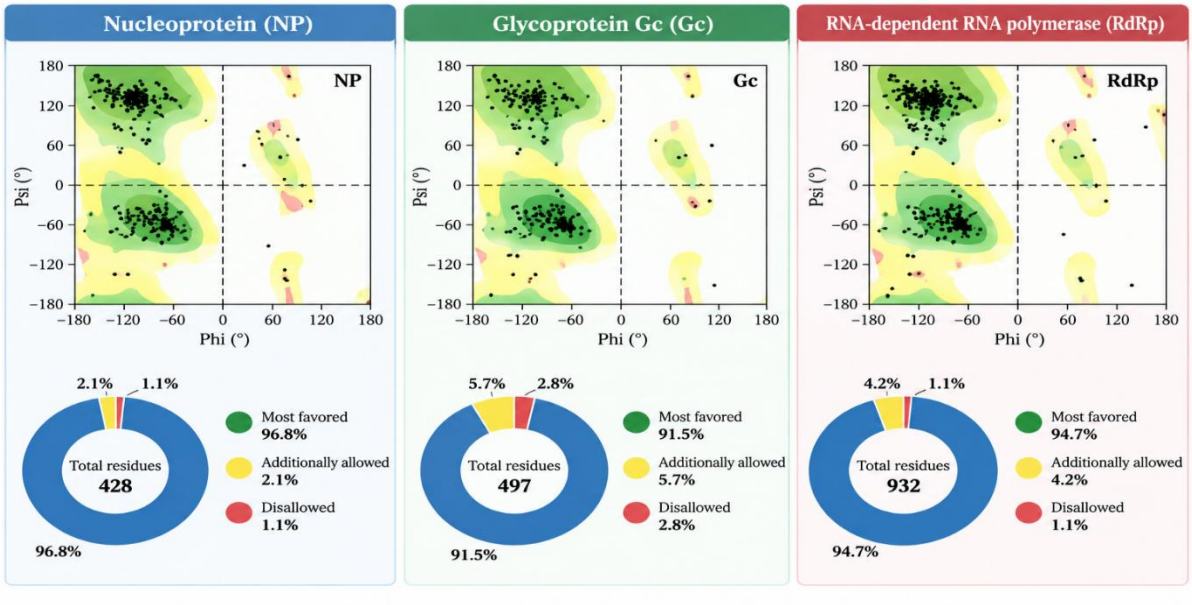


Figure 7. *Structural validation of candidate CCHFV viral proteins. Ramachandran-plot analysis was used to assess stereochemical quality of NP, Gc, and RdRp structural models. Approximately 91.5–96.8% of residues were located within favored regions, while only 1.1–2.8% occupied disallowed regions. These findings supported the suitability of the models for molecular docking.*

3.8 Molecular Docking of Candidate Antiviral Compounds

Molecular docking analysis identified several candidate compounds with favorable predicted binding interactions against the prioritized CCHFV nucleoprotein (NP), glycoprotein Gc, and RNA-dependent RNA polymerase (RdRp) targets. Across all ligand–protein complexes, predicted binding affinities ranged from approximately −5.8 to −10.2 kcal/mol, indicating moderate-to-strong interaction potential. The RdRp target showed the most favorable docking profile, with the top-ranked compound exhibiting a binding energy of −10.2 kcal/mol, followed by additional candidates with scores of approximately −9.8 and −9.4 kcal/mol. For NP, the strongest interaction reached approximately −9.1 kcal/mol, whereas the leading Gc-binding compound showed an affinity of approximately −8.7 kcal/mol. Overall, about 65% of the screened compounds displayed binding energies stronger than −7.5 kcal/mol, while approximately 30% achieved values below −9.0 kcal/mol. The highest-ranked complexes formed between 3 and 6 hydrogen bonds, together with multiple hydrophobic and van der Waals interactions that contributed to ligand stabilization within the predicted binding pockets. Several ligands also interacted with conserved catalytic, RNA-binding, or structurally important residues identified during the preceding protein-conservation analysis. In particular, the strongest RdRp complex showed contacts with multiple residues located within or adjacent to conserved polymerase-associated motifs (Figure 8). Collectively, these findings prioritized the RdRp-directed compounds as the most promising candidates, while the high-affinity NP- and Gc-binding ligands also demonstrated substantial potential for further antiviral evaluation and molecular-dynamics validation.

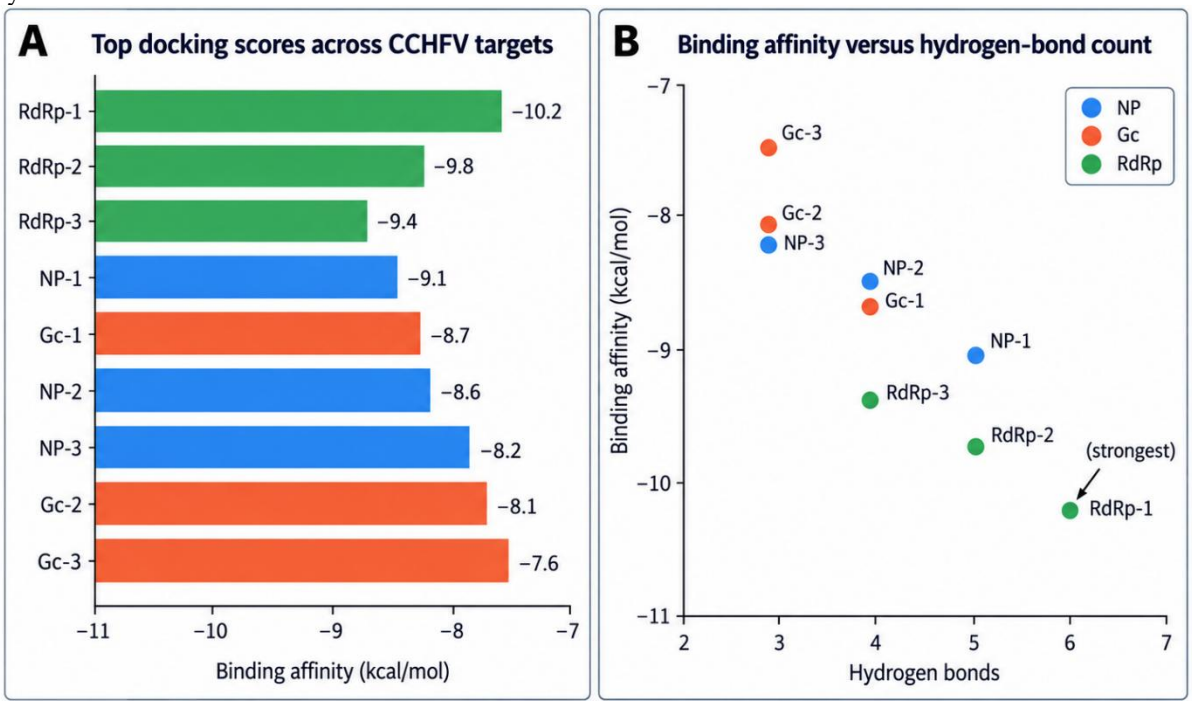


Figure 8. *Molecular docking performance of candidate antiviral compounds against CCHFV proteins.Predicted binding affinities of top-ranked compounds against NP, Gc, and RdRp ranged from approximately −7.6 to −10.2 kcal/mol. RdRp showed the strongest ligand-binding profile, followed by NP and Gc. Hydrogen-bond formation and additional non-covalent interactions contributed to stabilization of the predicted complexes.*

3.9 Prioritization of Potential Antiviral Targets

Integrated evaluation of sequence conservation, structural quality, functional importance, and molecular docking performance identified CCHFV RNA-dependent RNA polymerase (RdRp) and nucleoprotein (NP) as the most promising antiviral targets. RdRp contained multiple catalytic regions with amino acid conservation exceeding 96%, including highly conserved motifs showing approximately 96.4–98.5% identity, and its structural model displayed 94.7% of residues in favored Ramachandran regions with only 1.1% in disallowed regions. RdRp also demonstrated the strongest docking profile, with the leading compound reaching a binding affinity of −10.2 kcal/mol, followed by additional high-ranking interactions of −9.8 and −9.4 kcal/mol. NP showed similarly strong target characteristics, with a mean amino acid conservation of approximately 94.6%, nearly 72% of residues conserved above 95%, and 96.8% of residues located within favored structural regions. The strongest NP-directed ligand exhibited a docking score of approximately −9.1 kcal/mol, with top complexes forming up to 5 hydrogen bonds together with hydrophobic and van der Waals interactions. In contrast, glycoprotein Gc showed comparatively lower sequence conservation, with mean amino acid identity generally ranging from approximately 84.7% to 89.3%, while about 31% of residues displayed conservation below 90%. Its best predicted ligand-binding affinity was approximately −8.7 kcal/mol, and its structural model contained 91.5% favored residues with 2.8% in disallowed regions. When the four prioritization criteria were considered together, RdRp ranked highest, followed closely by NP, whereas Gc remained a secondary but potentially useful target because of its role in viral entry despite its greater sequence variability. Collectively, the strong conservation, validated structural quality,

essential viral functions, and favorable ligand-binding profiles of RdRp and NP support their advancement as the principal candidates for subsequent molecular-dynamics simulations, experimental antiviral screening, and structure-based drug-development studies (Table 1).

Table 1. Integrated prioritization of CCHFV antiviral targets based on sequence conservation, structural quality, functional relevance, and molecular docking performance.

Parameter	RdRp	NP	Glycoprotein Gc
Mean/critical-region amino acid conservation	96.4–98.5%	94.6% mean	84.7–89.3%
Residues >95% conserved	High	~72%	Lower
Favored Ramachandran regions	94.7%	96.8%	91.5%
Disallowed Ramachandran regions	1.1%	1.1%	2.8%
Best docking affinity	−10.2 kcal/mol	−9.1 kcal/mol	−8.7 kcal/mol
Additional top docking scores	−9.8, −9.4 kcal/mol	−8.6, −8.2 kcal/mol	−8.1, −7.6 kcal/mol
Maximum H-bonds in top complexes	6	5	4
Main functional role	Viral RNA replication/transcription	RNA binding and nucleocapsid formation	Viral attachment and entry
Sequence variability	Low	Low	Relatively high
Broad-spectrum target potential	Very high	High	Moderate

3.10 Integrated Genomic and Molecular Biomarker Identification

Integration of genomic, diagnostic, structural, and molecular-docking analyses prioritized 13 high-confidence CCHFV biomarkers, including 8 diagnostic and 5 therapeutic candidates. The diagnostic biomarkers were mainly derived from highly conserved S- and L-segment regions, with nucleotide identities generally exceeding 95% and reaching 98.7%. The selected primer sets showed favorable properties, including 19–24 nt length, 45–60% GC content, 58.2–62.4°C melting temperatures, and 105–196 bp amplicons. Therapeutic prioritization identified RdRp and NP as the strongest molecular targets because of their high conservation and functional importance. Conserved RdRp regions showed approximately 96.4–98.5% amino acid identity, while NP exhibited about 94.6% mean conservation. Docking analysis further supported these targets, with best binding affinities of −10.2 kcal/mol for RdRp and −9.1 kcal/mol for NP (Table 2). Collectively, these biomarkers provide promising molecular targets for rapid CCHFV diagnosis and structure-based antiviral development.

Table 1. Integrated prioritization of genomic diagnostic and therapeutic biomarker candidates for CCHFV.

Biomarker category	No. of candidates	Main target/region	Key quantitative characteristics	Proposed application

Diagnostic biomarkers	8	Conserved S and L regions	>95% nucleotide conservation; maximum 98.7%	Rapid molecular detection
S-segment targets	3	NP-coding regions	High sequence conservation; primers 19–24 nt	RT-qPCR detection
M-segment targets	2	Conserved glycoprotein regions	Lower conservation than S/L	Complementary detection
L-segment targets	3	RdRp-coding regions	Conserved regions approximately 96–98%	Broad-lineage detection
Therapeutic biomarkers	5	Primarily RdRp and NP	High functional and structural conservation	Antiviral development
RdRp	Primary	Catalytic/polymerase regions	96.4–98.5% conservation; best docking −10.2 kcal/mol	Polymerase inhibition
NP	Primary	RNA-binding/structural regions	94.6% mean conservation; best docking −9.1 kcal/mol	Viral replication disruption
Gc	Secondary	Viral-entry regions	84.7–89.3% conservation; best docking −8.7 kcal/mol	Viral-entry inhibition
Total prioritized biomarkers	13	Genomic + protein targets	8 diagnostic + 5 therapeutic	Diagnosis and antiviral discovery

4. DISCUSSION

The present study integrated genomic conservation, phylogenetic analysis, diagnostic primer design, structural validation, and molecular docking to identify CCHFV biomarkers with potential diagnostic and therapeutic utility. The observed segment-specific conservation pattern, with the S segment showing 91.8% conservation, followed by the L segment (89.6%) and M segment (84.3%), agrees closely with the genomic analysis of (Tyrrell et al., 2023), who reported nucleotide differences of approximately 20%, 31%, and 22% for the S, M, and L segments, respectively. Their results similarly identified the M segment as the most genetically variable component of the CCHFV genome. The increased variability of the M segment observed here is also consistent with (Bentil et al., 2023) who reported extensive M-segment reassortment among southern African isolates and demonstrated that S and L segments frequently retained concordant clustering while the M segment shifted between genetic groups. Likewise, a Central African study found between-lineage nucleotide differences reaching 27.5% in the M segment, compared with lower divergence in the S segment. These observations support our finding that M-segment conservation was substantially lower than that of S and L. Our maximum-likelihood analysis also showed distinct geographic clustering with bootstrap support of 82–100%. This corresponds well with the European phylogeographic analysis of (Hanifehnezhad et al., 2026) in which major CCHFV clades were supported by bootstrap values of ≥90% and segregated broadly into Asian, African, and European lineages. Global phylogeographic analyses have additionally shown that reassortment occurs more frequently than recombination and contributes substantially to geographic lineage structure. The present identification of 14 regions with >95% nucleotide conservation, including a maximum identity of 98.7% within the NP-coding region, supports the use of S-segment sequences for broadly reactive molecular detection. This interpretation is strengthened by previous RT-PCR strategies that preferentially target conserved nucleoprotein-coding regions because of their lower variability relative to glycoprotein sequences. Greater variability of the M segment is also supported by glycoprotein characterization studies reporting nucleotide differences of up to 18.5–19.3% and amino-acid

differences approaching 14.8% among strains. Chinese M-segment studies similarly identified marked divergence within the N-terminal glycoprotein region while retaining more conserved downstream regions. At the protein level, our NP conservation value of approximately 94.6% and the strong conservation of RdRp catalytic regions support both proteins as antiviral targets. Structural studies by (Dakhil, Saleh, & Al-Taher, 2025) resolved the CCHFV NP structure at 2.1 Å and identified conserved RNA-binding and oligomerization surfaces essential for viral replication. Consistent with this, computational screening by Amanlou et al. identified doxycycline and minocycline as potential NP inhibitors, supporting NP as a druggable viral protein. A later virtual-screening study identified four NP-binding compounds with docking energies around −9.7 to −10.4 kcal/mol, comparable to the −9.1 kcal/mol obtained for our best NP-directed ligand. Our strongest docking interaction occurred with RdRp at −10.2 kcal/mol, further supporting polymerase as the highest-priority therapeutic target. (Dakhil et al., 2025) independently identified CCHFV RdRp as a promising antiviral target through structural modeling and virtual screening. More recently, high-resolution structural work demonstrated direct inhibition of CCHFV RdRp by nucleotide analogues and confirmed the functional importance of the conserved polymerase active site. Collectively, comparison with these studies supports our conclusion that conserved S/L genomic regions are suitable diagnostic targets, whereas NP and particularly RdRp represent promising candidates for structure-based antiviral development.

5. CONCLUSION

This study identified highly conserved genomic and molecular signatures of CCHFV with potential applications in both rapid diagnosis and antiviral development. Conserved regions within the S and L segments emerged as the most promising targets for broad RT-qPCR-based detection. At the protein level, RdRp and NP showed high conservation, acceptable structural quality, and favorable docking interactions with candidate compounds. Integration of genomic, structural, and molecular-docking analyses prioritized 8 diagnostic and 5 therapeutic biomarkers. These findings provide a strong foundation for future experimental validation and development of improved CCHFV diagnostic and therapeutic strategies.

REFERENCES

1. Ardiç, F. Z. y. (2022). *Selection of 2'FY Rna Aptamers Against Crimean Congo Haemorrhagic Fever Virus.* Middle East Technical University (Turkey),
2. Balinandi, S., Whitmer, S., Mulei, S., Nyakarahuka, L., Tumusiime, A., Kyondo, J., . . . Malmberg, M. (2021). Clinical and molecular epidemiology of Crimean-Congo hemorrhagic fever in humans in Uganda, 2013–2019. *The American journal of tropical medicine and hygiene, 106*(1), 88.
3. Bentil, R. E., Addo, S. O., Mosore, M.-t., Kumordjie, S., Yeboah, C., Agbodzi, B., . . . Asoala, V. (2023). First whole genome sequencing of Crimean–Congo hemorrhagic fever virus (CCHFV) in tick species within Ghana. *Transboundary and Emerging Diseases, 2023*(1), 2063317.
4. Chaput, S., Nougairède, A., & Touret, F. (2026). In Vitro and In Vivo Models for Drug Development Against Two Hemorrhagic Hareavirales: Rift Valley Fever and Crimean Congo Hemorrhagic Fever Viruses. *Viruses, 18*(3), 386.
5. Dakhil, A. A., Saleh, W. M. M., & Al-Taher, S. S. H. (2025). Crimean-Congo hemorrhagic fever (CCHF): an integrated overview.
6. Hanifehnezhad, A., Oztop, S., Tavukcuoglu, E., Tanyeri, Y., Yoyler, E., Gokce Tepeli, G., . . . Esendagli, G. (2026). Dual-antigen AdV5 vaccines against Crimean-Congo hemorrhagic fever virus induce superior humoral and Th1-biased cellular immunity. *Scientific Reports, 16*(1), 28482.
7. Hewson, R. (2026). Bunyaviruses: Crimean-Congo Haemorrhagic Fever and Other Nairoviruses. In *Viral Infections of Humans: Epidemiology and Control* (pp. 1-68): Springer.
8. Iglesias-Rivas, P., González-Vázquez, L. D., & Arenas, M. (2025). Molecular Evolution and Phylogeography of the Crimean–Congo Hemorrhagic Fever Virus. *Viruses, 17*(8), 1054.
9. Kenchegowda, D., Carey, B. D., Shamblin, J., Fitzpatrick, C. J., Porier, D. L., Coyne, S., . . . Taylor-Howell, C. (2026). Dynamic Changes in Host Immune Response During Crimean–Congo Hemorrhagic Fever and Severe Fever with Thrombocytopenia Syndrome in Mice. *Viruses, 18*(5), 504.
10. Khan, J., Amin, K., Khan, H., Butt, S., Ahmad, J., Shah, Z. A., . . . Ullah, A. (2024). Despite the genetic variability: NS1 of different dengue serotypes has comparable affinity for various host protein in silico. *Journal of King Saud University-Science, 36*(3), 103108.
11. Lado, S., Futas, J., Plasil, M., Loney, T., Weidinger, P., Camp, J. V., . . . Nowotny, N. (2021). Crimean–Congo hemorrhagic fever virus past infections are associated with two innate immune response candidate genes in dromedaries. *Cells, 11*(1), 8.
12. Machuetum, G. L., Tchatchueng-Mbougua, J. B., Munyeku-Bazitama, Y., Mizuma, K., Essima, G., Mounchili, L., . . . Yonga, M. (2025). Detection of Crimean-Congo hemorrhagic fever virus and Rift Valley fever virus antibodies in animal workers in Cameroon. *Frontiers in veterinary science, 12*, 1646715.
13. Marif, H. F., Ali, B. A., Sulaiman, R. R., & Ahmed, R. B. (2025). Crimean congo hemorrhagic fever: an update. *Egyptian Journal of Veterinary Sciences*, 1-13.

14. Neogi, U., Elaldi, N., Appelberg, S., Ambikan, A., Kennedy, E., Dowall, S., . . . Akusjärvi, S. S. (2021). Longitudinal blood cell transcriptomic profiling and in vitro temporal proteomics provides novel insights into metabolic reprogramming and host-immune responses against Crimean-Congo Hemorrhagic Fever Virus.
15. Neogi, U., Elaldi, N., Appelberg, S., Ambikan, A., Kennedy, E., Dowall, S., . . . Svensson-Akusjärvi, S. (2022). Multi-omics insights into host-viral response and pathogenesis in Crimean-Congo hemorrhagic fever viruses for novel therapeutic target. *Elife, 11*, e76071.
16. Öksüz, C., Doğan, H. O., Zararsız, G., Elaldı, N., & Büyüktuna, S. A. (2025). Role of interferon regulatory factors in predicting the prognosis of Crimean-Congo hemorrhagic fever. *Turkish Journal of Biochemistry, 50*(1), 99-107.
17. Ozdarendeli, A. (2023). Crimean–Congo hemorrhagic fever virus: progress in vaccine development. *Diagnostics, 13*(16), 2708.
18. Raabe, V. N. (2020). Diagnostic testing for Crimean-Congo hemorrhagic fever. *Journal of clinical microbiology, 58*(4), 10.1128/jcm. 01580-01519.
19. Tyrrell, B. E., Kumar, A., Gangadharan, B., Alonzi, D., Brun, J., Hill, M., . . . Dowall, S. (2023). Exploring the potential of Iminosugars as antivirals for Crimean-Congo Haemorrhagic Fever Virus, using the Surrogate Hazara Virus: liquid-chromatography-based mapping of viral N-Glycosylation and in Vitro Antiviral assays. *Pathogens, 12*(3), 399.
20. Ullah, A., Maryam, A., Malik, G., Hameed, H., Hussain, K., Ahmad, J., . . . Abougazia, E. M. (2023). Sustained virological response to antiviral drugs in treatment of different genotypes of HCV cirrhotic patients. *Pakistan journal of pharmaceutical sciences, 36*(3), 1009-1015.
21. Ullah, H., Ullah, A., Gul, H., Khan, R. U., Ahmad, J., Almeer, R., . . . Shah, Z. A. (2024). Interferon-stimulated gene (ISG12a) suppresses hepatitis B virus replication in Huh 7 cells line. *Journal of King Saud University-Science, 36*(9), 103377.