

Vital Surveillances

Molecular Evolution in the Second Hypervariable Region of the G Gene for Human Respiratory Syncytial Virus Lineage A.D — Worldwide, 2010–2024

Yuqing Shi¹; Zhen Zhu¹; Jinhua Song¹; Naiying Mao¹; Jie Jiang¹; Ke Lu¹; Na Wang¹; Liwei Sun²; Linqing Zhao³; Yonghao Guo⁴; Ruiqiu Zhao⁵; Jiahui Xie⁶; Chao Liu⁷; Ying Chen⁸; Feng Zhang⁹; Junrui Chen¹⁰; Lei Cao¹; Hongqiao Hu¹; Baicheng Xia¹; Yan Zhang^{1,*}

ABSTRACT

Introduction: Human respiratory syncytial virus subgroup A (HRSV-A) lineage A.D is characterized by a unique 72-nucleotide duplication in the second hypervariable region of the G-gene (G-HVR2) and has become the predominant global circulating strain since 2010. Molecular evolutionary features of this region should be monitored.

Methods: A global HRSV-A G-HVR2 sequence dataset (2010–2024) was assembled that included 144 Chinese sequences and the molecular characteristics and evolutionary patterns in lineage A.D were analyzed.

Results: Molecular clock analysis based on G-HVR2 estimated that HRSV-A lineage A.D originated in 2006, with subsequent diversification into four distinct lineages. A.D.4 and A.D.5 were grouped as A.D.4/A.D.5. Lineages circulate globally with temporal fluctuations and sequential dominance shifts. The evolutionary rate of A.D was high, with the highest rate observed in A.D.4/A.D.5. Amino acid analysis indicated five shared, lineage-specific mutations in A.D.1 and A.D.4/A.D.5 that coincided with high-frequency mutation sites. Nine positively-selected sites were predicted across all lineages, three of which resided within the 72-nucleotide duplication. A.D.1 and A.D.4/A.D.5 lost conserved N- and O-glycosylation sites. A 72-nucleotide duplication in lineage A.D introduced two extra β -strands and two unique α -helical motifs relative to the prototype strain.

Conclusion: Adaptive evolution in the G-HVR2 region of HRSV-A lineage A.D was evident, likely facilitating the rapid transmission of this lineage. Sustained monitoring of lineage-specific evolution in this region is critical for targeted prevention and control of HRSV infections.

Human respiratory syncytial virus (HRSV) is the leading cause of acute lower respiratory tract infections worldwide, and it poses a substantial burden in terms of hospitalization and mortality in young children, older adults, and immunocompromised individuals (1). Currently, HRSV infection has become a major global public health concern.

HRSV is considered a single serotype but is divided into two antigenic subgroups: A and B (2). HRSV subgroup A (HRSV-A) exhibits high genetic diversity and has diversified into multiple lineages, of which lineage A.D has increased rapidly in global prevalence since its initial detection in 2010 (3). Whole genome analyses have classified A.D into five major lineages (A.D.1–A.D.5). A defining molecular feature of this lineage is a 72-nucleotide duplication (position: aa284–307) in the second hypervariable region of the G gene (G-HVR2) (position: aa212–321), a region characterized by high genetic variability and substantial clinical immune selection pressure (4).

In recent years, accumulation of G-HVR2 sequences has made it possible to comprehensively investigate their molecular characteristics and evolutionary dynamics. Therefore, we assembled a global G-HVR2 dataset for the HRSV-A lineage A.D from 2010–2024, and conducted a phylogenetic analysis, evolutionary assessment, genetic variation profiling, and structural prediction to characterize the adaptive evolution of the G-HVR2 region containing the 72-nucleotide duplication. The study aimed to provide scientific data for the targeted prevention and control of HRSV infection, as well as future investigations into its potential links with viral biological properties.

METHODS

G-HVR2 Dataset Construction

A total of 52 representative G-HVR2 sequences (National Microbiology Data Center reference

numbers: NMDCN0009ROK–NMDCN0009RQ7) of HRSV-A lineage A.D strains from nine provinces in China from 2012–2024 were included in this study (Supplementary Table S1, available at <https://weekly.chinacdc.cn>). After excluding sequences with incomplete targets or ambiguous temporal and geographic information, 2,402 representative G-HVR2 sequences belonging to lineage A.D from 30 countries (2010–2024) were downloaded from GenBank (version 264.0; <https://www.ncbi.nlm.nih.gov/genbank/>; December 2024; National Center for Biotechnology Information, Bethesda, MD, USA) based on sequence divergence and spatiotemporal characteristics. The integration of these sequences yielded a comprehensive database comprising 2,454 G-HVR2 sequences of the HRSV-A lineage A.D for subsequent analysis, which included 144 Chinese sequences from 16 provinces (2010–2024).

Phylogenetic and Molecular Evolutionary Analysis

Sequences were aligned using MAFFT (version 7.490; Osaka University, Osaka, Japan). A maximum likelihood (ML) phylogenetic tree was constructed using MEGA (version 11; Pennsylvania State University, State College, PA, USA) (GTR+I+R4 model, 1,000 bootstraps) and visualized using the online tool, iTOL (version 7.5.1; <https://itol.embl.de/>; April 2025; European Molecular Biology Laboratory, Heidelberg, Germany). Furthermore, 199 representative A.D lineage sequences from 28 countries were selected for molecular evolutionary analysis. Bayesian phylogenetic inference was conducted using BEAST (version 1.10.4; University of Auckland, Auckland, New Zealand) (Strict clock model, coalescent-based Bayesian skyline plot). Temporal signals were assessed using root-to-tip regression in TempEst (version 1.5.3; University of Edinburgh, Edinburgh, UK), and the best-fit model was selected using jModelTest (version 2.1.6; University of Vigo, Vigo, Spain). Evolutionary rates and time to the most recent common ancestor (tMRCA) were estimated, and population dynamics were reconstructed using Bayesian skyline plot (BSP) analysis in Tracer (version 1.7.2; University of Auckland, Auckland, New Zealand).

Genetic Variation, Selection Pressure Analysis, and Glycosylation Site Prediction

Amino acid variations were analyzed using MEGA

11, with lineage-specific mutations defined as those with >80% frequency relative to the A.D prototype strain (GenBank accession number: JN257693). High-frequency mutation sites (>60%) were identified with entropy analysis using BioEdit (version 7.2.5; North Carolina State University, Raleigh, NC, USA). Selection pressures were assessed using FUBAR, MEME, and SLAC on Datamonkey (version 2.0; <https://www.datamonkey.org/>; Temple University, Philadelphia, PA, USA), with positive selection (dN/dS >1) defined as $P < 0.05$ or posterior probability >0.95, requiring support from at least two methods. Potential N- and O-glycosylation sites were predicted using NetNGlyc 1.0 and NetOGlyc 4.0 (<https://services.healthtech.dtu.dk/services>; Technical University of Denmark, Lyngby, Denmark) servers, respectively.

Protein Structural Prediction

Using the A.D prototype strain as a reference, the identified lineage-specific mutations and positive selection sites were mapped onto the predicted secondary structure of the G protein using PSIPRED (version 4.0; <https://bioinf.cs.ucl.ac.uk/psipred/>; University College London, London, UK) to evaluate their potential impacts.

RESULTS

Phylogenetic Classification of Lineage A.D Based on G-HVR2

Phylogenetic analysis showed that 2,454 A.D sequences could be subdivided into four distinct lineages based on G-HVR2, largely consistent with the novel genotype classification (A.D.1–A.D.5) based on whole genome sequences (5). However, the highly genetically related strains, A.D.4 and A.D.5, as defined in the whole genome genotyping method, were topologically indistinguishable in G-HVR2 analysis and thus grouped as 'A.D.4/A.D.5' in the current study. The proportions of the four A.D lineages in the database were similar: A.D.1 (763 strains, 31.1%), A.D.2 (501 strains, 20.4%), A.D.3 (698 strains, 28.4%), and A.D.4/A.D.5 (492 strains, 20.1%) (Supplementary Table S2, available at <https://weekly.chinacdc.cn>). Based on the G-HVR2 region, the average genetic distance across all A.D lineages was 0.04. The inter-lineage distances were 0.037–0.048, and the maximum intra-lineage distance was 0.035 (Supplementary Table S3, available at <https://weekly.chinacdc.cn>).

Prevalence Patterns of Four A.D Lineages

A global epidemic pattern analysis indicated that four A.D lineages co-circulated from 2010 to 2024, each exhibiting a broad geographical distribution. However, their detection rates varied across the years. Lineage A.D.1 emerged as the dominant strain from 2018 onwards, peaking at 46.52% in 2020. Lineage A.D.2 accounted for a substantial proportion (12.50%–58.57%) from 2013–2020 and became dominant in 2017. Lineage A.D.3 was predominant from 2010–2016 (42.11%–100%) and re-emerged as the dominant lineage in 2024 (43.59%). Lineage A.D.4/A.D.5 dominated between 2021 and 2023,

with detection rates of 35.91%–50.35% (Figure 1).

All four lineages were detected in China, each showing a distinct spatiotemporal pattern. Lineages A.D.2 (16.7%, six provinces) and A.D.3 (73.6%, 11 provinces) were detected from 2012–2020 and 2012–2024, respectively. Since 2018, lineages A.D.1 (5.5%, six provinces, 2019–2023) and A.D.4/A.D.5 (4.2%, five provinces, 2018–2023) were successively identified. Consistent with global patterns, A.D.2 (46.67%–50.00%) and A.D.3 (50.00%–100%) dominated from 2011–2017, with A.D.3 being the most frequent lineage during this period. Since 2018, A.D.3 has been predominant (57.14%–100%), except

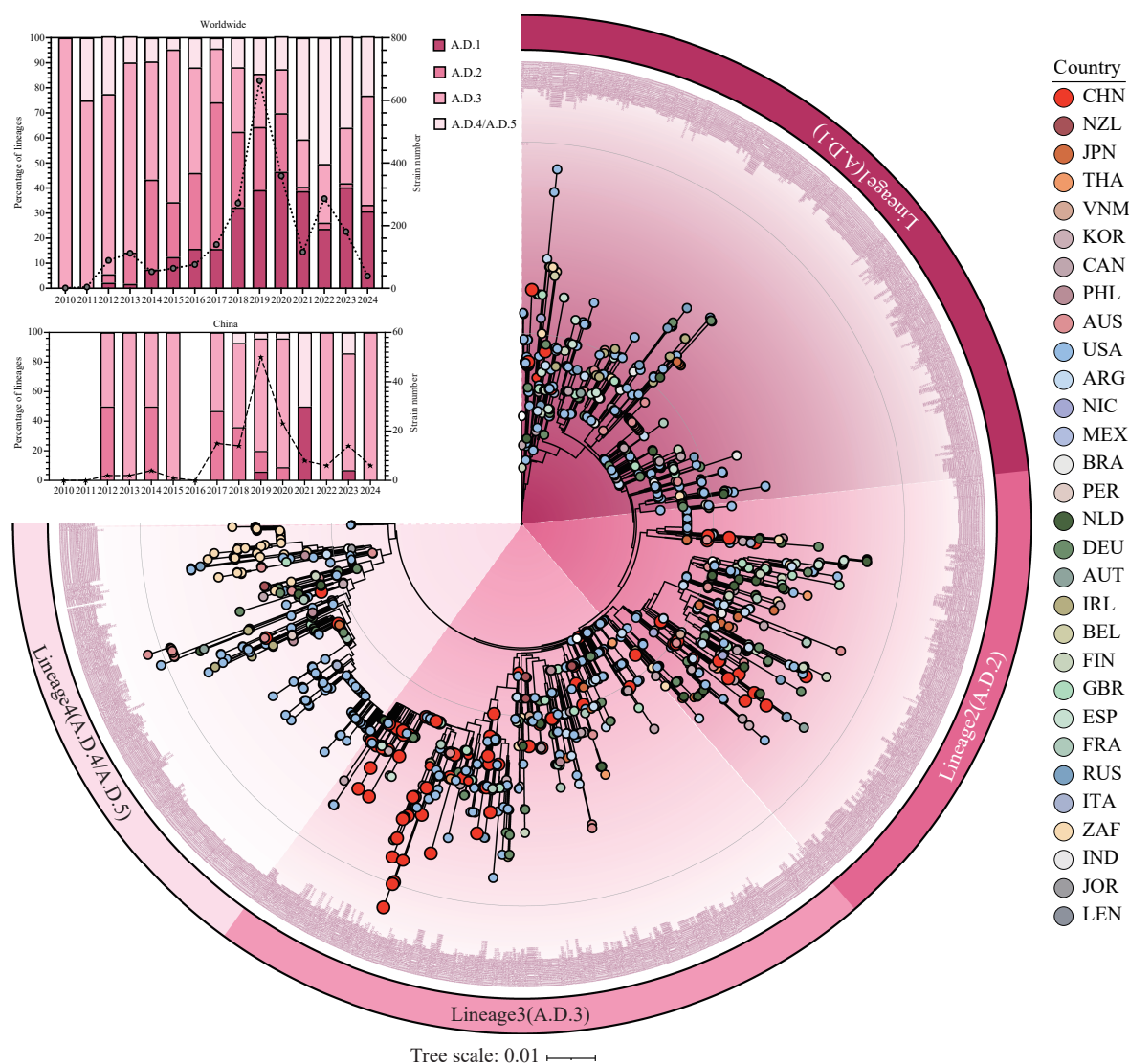


FIGURE 1. Phylogenetic and spatiotemporal distribution of 2,454 HRSV-A lineage A.D sequences.

Note: The upper-right panel presents the maximum likelihood phylogenetic tree constructed based on G-HVR2. The outer ring indicates the lineage classification (Lineages A.D.1–A.D.4/A.D.5). The upper-left panel illustrates the prevalence of the four lineages in China and worldwide.

Abbreviation: HRSV-A=human respiratory syncytial virus subgroup A; G-HVR2=second hypervariable region of the G gene.

in 2021 when only A.D.1 (50.00%) and A.D.4/A.D.5 (50.00%) were detected and aligned with global lineages circulating at that time (Figure 1).

Molecular Evolution of Lineage A.D

Based on the representative sequences, root-to-tip regression analysis indicated a strong temporal signal with a correlation coefficient = 0.9987 and $R^2 = 0.9974$ (Figure 2). The overall estimated evolutionary rate for lineage A.D was 5.63×10^{-3} substitutions/site/year (95% highest posterior density (HPD): 1.00×10^{-3} to 11.28×10^{-3}), with tMRCA estimated to be approximately 2006 (95% HPD: 2005–2010). The evolutionary rates of the four lineages varied from 1.92×10^{-3} (95% HPD: 4.18×10^{-6} to 6.36×10^{-3}) to 5.26×10^{-3} substitutions/site/year (95% HPD: 1.93×10^{-5} to 13.01×10^{-3}), with the highest rate in lineage A.D.4/A.D.5. The tMRCAs for these lineages were similar, and they occurred between 2008 and 2009 (Supplementary Table S4, available at <https://weekly.chinacdc.cn>). Furthermore, BSP analysis indicated that lineage A.D underwent a rapid and pronounced expansion in effective population size between 2010 and 2012, followed by a period of relative stability (Figure 2).

Amino Acid Variation and Selective Pressure Analysis of A.D Lineages

Compared to the A.D prototype strain, lineage-specific mutations in G-HVR2 were identified: A.D.1 had four sites (L274P, L298P, Y304H, T320A), and A.D.4/A.D.5 had two mutations (Y304H and L310P), with Y304H being a shared mutation between the two lineages. No lineage-specific mutations were detected in A.D.2 or A.D.3 (Figure 3A). Entropy analysis indicated five high-frequency mutations in G-HVR2 (aa274, aa298, aa304, aa319, and aa320). Four of these (excluding aa319) corresponded to lineage-specific mutations and two (aa298 and aa304) were located within a 72-nucleotide duplication. Selection pressure analysis indicated nine sites under positive selection across A.D lineages: three in A.D.1 (aa214, aa255, aa314), five in A.D.2 (aa273, aa297, aa298, aa309, and aa314), three in A.D.3 (aa274, aa297, aa298), and three in A.D.4/A.D.5 (aa273, aa284, aa297). Notably, aa284, aa297, and aa298 were located in the 72-nucleotide duplication (Figure 3B).

Glycosylation Site Prediction

Glycosylation site prediction identified two N-

glycosylation sites (aa237 and aa318) in the G-HVR2 of A.D lineages. A.D.1 and A.D.4/A.D.5 lost their N-glycosylation site at aa318. No N-glycosylation sites were found in the 72-nucleotide duplications. Additionally, the A.D lineages contained 43 O-glycosylation sites in G-HVR2, 10 of which were located in the 72-nucleotide duplication. However, A.D.1 lost one O-glycosylation site at aa320, and A.D.4/A.D.5 had deletions at aa305, aa319, and aa320 (Figure 3C).

Structural Prediction

The DMPFold structural prediction results indicated that the 72-nucleotide duplication at the C-terminus of G protein induced the formation of a coiled conformation, introducing two extra β -strands and two unique α -helical motifs relative to the prototype strain of HRSV-A (Figure 4A). However, the identified lineage-specific mutations did not lead to observable changes in the protein conformation (Figure 4B).

DISCUSSION

The HRSV-A lineage A.D, characterized by a 72-nucleotide duplication in G-HVR2, has rapidly become the predominant strain circulating globally over the past decade (6). Molecular clock analyses estimated that lineage A.D originated in 2006 (7). This lineage subsequently diversified into multiple distinct lineages (A.D.1–A.D.4/A.D.5) occurring between 2008 and 2009. After a period of covert transmission, a rapid and substantial expansion in the effective population size of lineage A.D was observed between 2010 and 2012 using BSP analysis, suggesting lineage-specific adaptive evolution of A.D within the human population. This process may be driven by multiple factors, such as amino acid mutations, glycosylation site alterations, and conformational changes in G-HVR2.

Prevalence trend analysis showed that the A.D lineages (A.D.1–A.D.4/A.D.5) subsequently circulated worldwide with marked temporal fluctuations and successive dominance shifts. This sequential dominance pattern possibly reflects the combined effects of interlineage competition, immune-mediated selection, and coronavirus disease 2019-related public health interventions (8). Circulation patterns in China largely paralleled global patterns but exhibited distinct regional features, with A.D.3 maintaining a

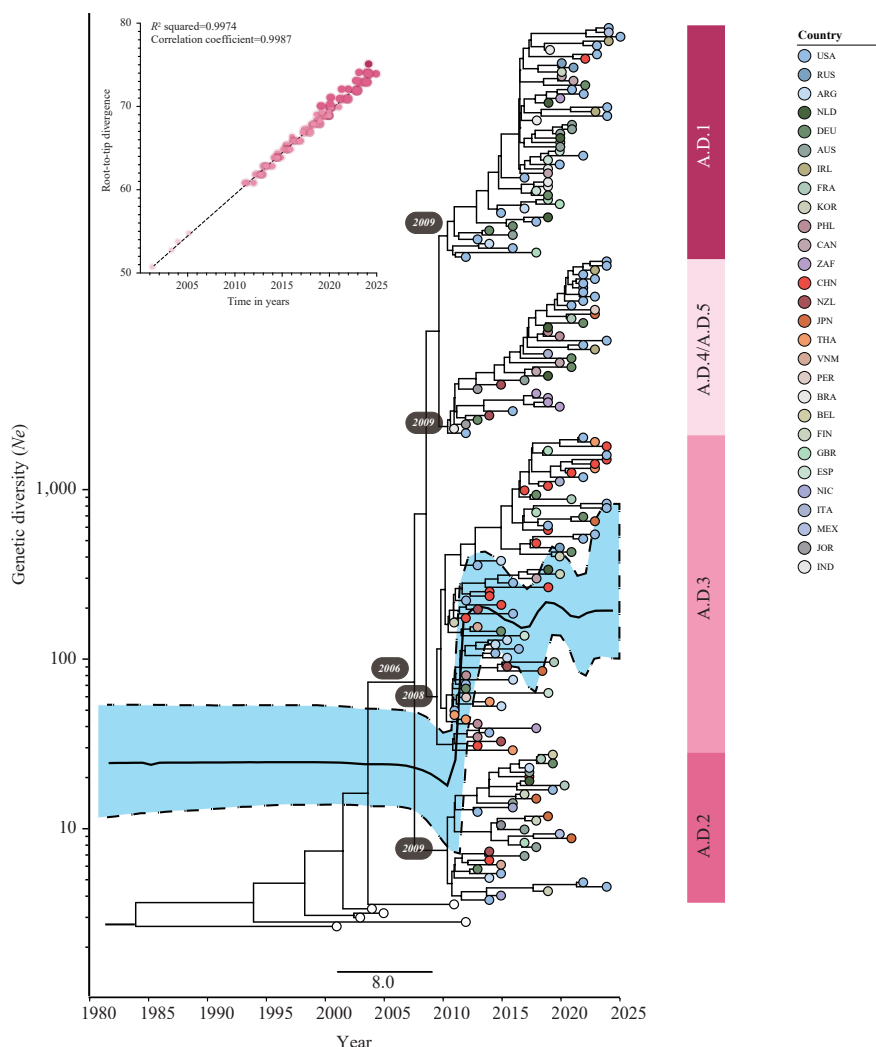


FIGURE 2. Time-scaled phylogeny and evolutionary dynamics of 199 HRSV-A lineage A.D and six reference sequences from other lineages (A.1–A.3).

Note: A maximum clade credibility tree was constructed based on G-HVR2. The scatter plot on the upper left shows the results of the root-to-tip regression analysis. The diagram embedded in the tree represents a Bayesian skyline plot. The vertical axis indicates the effective population size, and the horizontal axis represents the year. The representative strains corresponding to the unlabeled branches in the figure are A.1–A.3 representative strains.

Abbreviation: HRSV-A=human respiratory syncytial virus subgroup A; G-HVR2=second hypervariable region of the G gene.

consistently high prevalence across most years, highlighting regional heterogeneity in lineage adaptation and transmission dynamics. Therefore, continuous lineage-related surveillance is crucial to understand the global prevalence and evolutionary patterns of HRSV-A (9).

Molecular evolutionary analysis based on G-HVR2 indicated a high overall evolutionary rate for A.D lineages (5.63×10^{-3} substitutions/site/year), which was consistent with previous findings (10). Of the four lineages, A.D.4/A.D.5 exhibited the highest rate. Moreover, lineage-specific mutations in A.D.1 and A.D.4/A.D.5 overlapped substantially with high-

frequency mutation sites, suggesting that these genetic changes might be attributed to the adaptive evolution associated with its recent emergence as a dominant strain (11).

Selection pressure analysis identified nine positively-selected sites across all lineages, three of which (aa284, aa297, and aa298) were located within the 72-nucleotide duplication, indicating that lineage-defining residues in G-HVR2 were strongly targeted by positive selection. Several duplication sites exhibited concurrent high-frequency variations and lineage-specific changes. For example, residue 298 was positively selected in A.D.2 and A.D.3, whereas in A.D.1, it represented a

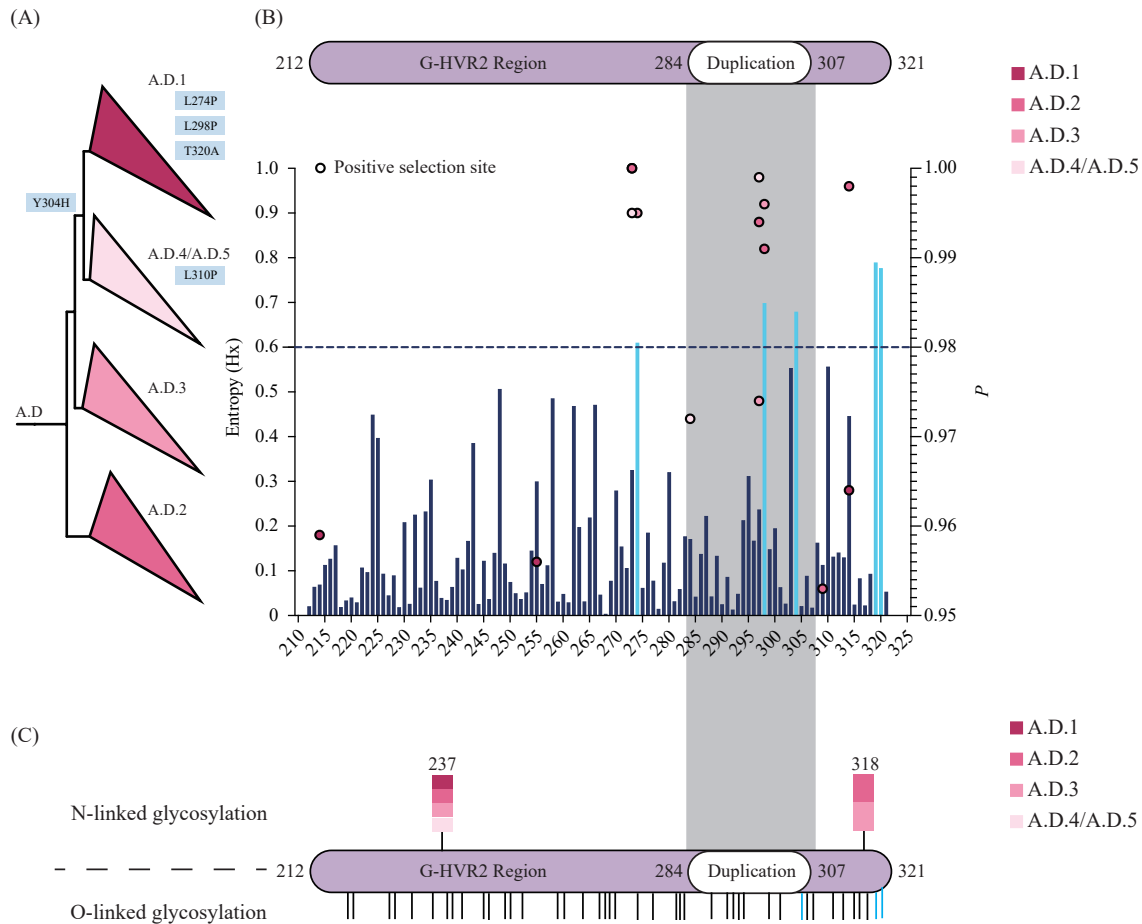


FIGURE 3. Lineage-specific amino acid sites, entropy analysis, selective pressure, and glycosylation of the HRSV-A A.D lineages based on G-HVR2. (A) Lineage-specific amino acid sites. (B) Entropy analysis and predicted positively-selected sites. (C) Predicted N- and O-linked glycosylation sites.

Note: For (A), a schematic phylogenetic tree displaying color-coded lineages and lineage-specific amino acid sites highlighted in blue boxes. For (B), the schematic presents the G-HVR2 region, in which the 72-nucleotide duplication is highlighted with gray shading. Entropy analysis identified high-frequency mutation sites (vertical axis, entropy; horizontal axis, amino acid position); those exceeding an entropy of 0.6 are displayed in light-blue, and predicted positively-selected sites are marked by circles whose colors correspond to different lineages. For (C), sites including predicted N-glycosylation sites are presented in the upper panel, with colors indicating different lineages; predicted O-glycosylation sites are in the lower panel, where vertical lines denote site positions and blue lines show O-glycosylation sites specific to lineages A.D.2 and A.D.3.

Abbreviation: HRSV-A=human respiratory syncytial virus subgroup A; G-HVR2=second hypervariable region of the G gene.

lineage-specific L298P substitution and a high-frequency variant. These findings were consistent with those of previous studies (12–13), supporting the critical role of the 72-nucleotide duplication in G-HVR2 in HRSV-A immune evasion and adaptive evolution.

In addition to the shared Y304H mutation, A.D.1 and A.D.4/A.D.5 exhibited a concurrent loss of conserved N- and O-linked glycosylation sites, indicating a convergent adaptive advantage in these lineages. Altered glycosylation patterns may affect protein conformational stability, immunogenicity, or

receptor-binding properties, potentially shaping the phenotypic and epidemiological profiles of distinct viral lineages (14–15). Given the increasing global detection rates of A.D.1 and A.D.4/A.D.5 in recent years, it has become imperative to elucidate the potential link between glycosylation changes and disease burden or clinical phenotypes.

Secondary structure prediction indicated that, compared to the HRSV-A prototype strain (lineage A.1), the G protein of lineage A.D contained a duplication insertion that introduced two additional β -strands and two unique α -helical motifs. This

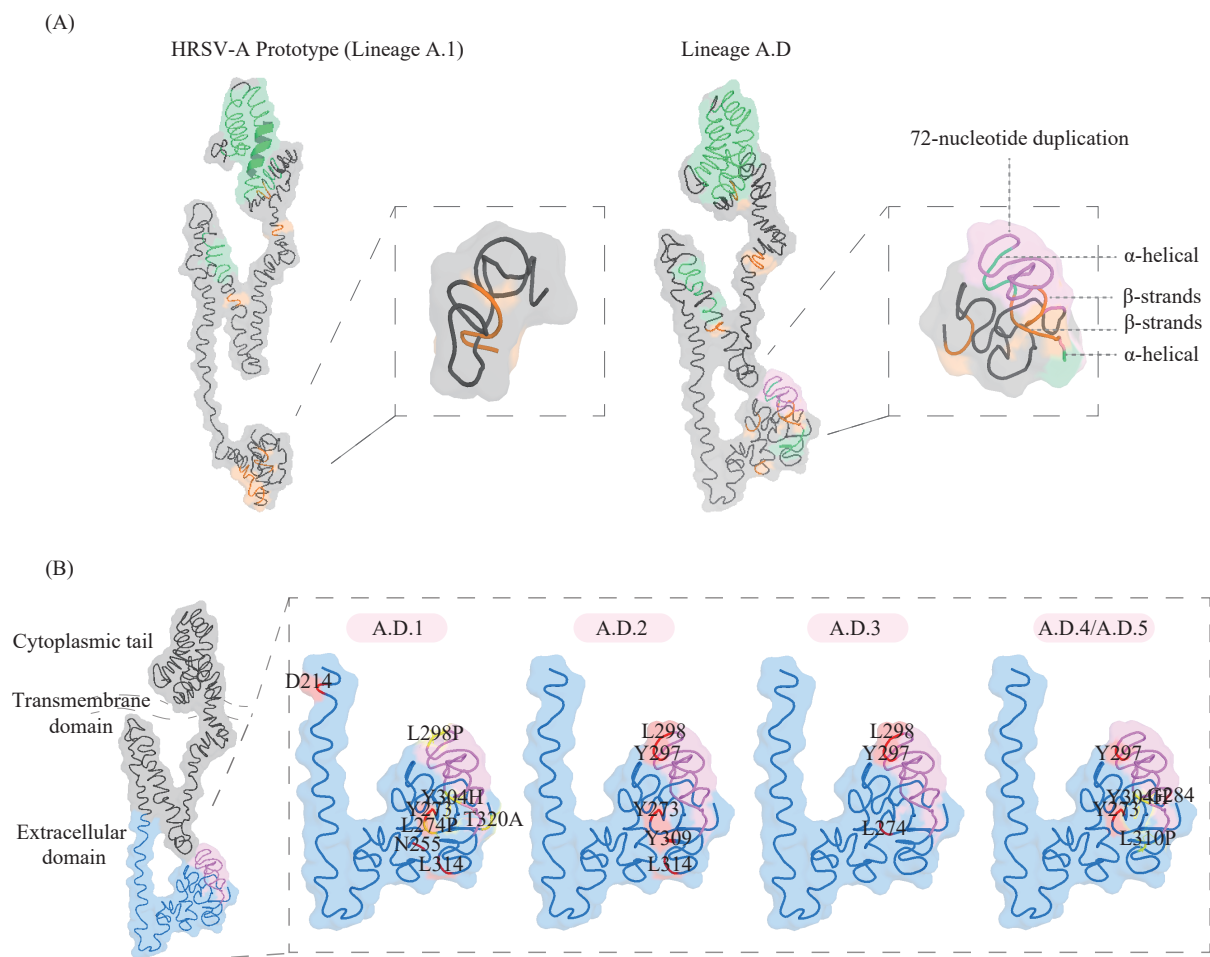


FIGURE 4. Secondary structure prediction of HRSV-A G protein. (A) Comparison of the G protein secondary structure between the HRSV-A prototype strain and lineage A.D. (B) Predicted secondary structures of the G protein in different A.D. lineages.

Note: For (A), on the left is the secondary structure of the G protein from the HRSV-A prototype strain. On the right is the secondary structure of the G protein from A.D. lineages, in which the α-helices and β-sheets introduced by the duplicated insertion are colored green and yellow, respectively. For (B), the G-HVR2 region is highlighted in blue, the 72-nucleotide duplication is shown in pink, lineage-specific sites are marked with a yellow background, and positively selected sites are indicated by a red background.

Abbreviation: HRSV-A=human respiratory syncytial virus subgroup A; G-HVR2=second hypervariable region of the G gene.

structural remodeling may alter G protein binding to host cell receptors and the immunorecognition of conformational epitopes. However, due to the high degree of glycosylation in G proteins, the choice of modeling framework might influence the structural prediction results; thus, such structural changes require further validation.

The findings of this study highlighted that the G-HVR2 occupied a significant position in the sustained adaptive evolution of HRSV-A lineage A.D. This study employed a reverse genetics platform to conduct functional validation of this duplication, enabling direct assessment of its biological characteristics. Preliminary data (unpublished) revealed that the

duplication exerted a substantial influence on viral pathogenicity, underscoring its potential contribution to lineage-specific adaptation.

Limitations of this study: First, compared with whole-genome analysis, lineage division based on the G-HVR2 lacks sufficient resolution to reliably distinguish closely related A.D.4 and A.D.5. Second, the paucity of available Chinese sequence data might lead to underestimation of the circulation of lineages. Third, the findings might be subject to potential influence from insufficient surveillance or sampling bias.

In conclusion, this study delineated the prevalence and transmission dynamics of the currently dominant

HRSV-A lineage A.D. Systematic analysis of genetic polymorphisms, selective pressures, glycosylation patterns, and structural predictions demonstrated significant adaptive evolution in the G-HVR2 region containing a 72-nucleotide duplication, which likely facilitated the lineage's rapid transmission. Accordingly, ongoing monitoring of lineage-specific genetic evolution in this region is critical for targeted prevention and control of HRSV infection.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: Supported by the Beijing Natural Science Foundation (Grant No. L252073).

doi: 10.46234/ccdcw2026.103

* Corresponding author: Zhang Yan, zhangyan@ivdc.chinacdc.cn.

¹ National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, NHC Key Laboratory of Medical Virology and Viral Disease, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China; ² Children's Hospital of Changchun, Changchun City, Jilin Province, China; ³ Laboratory of Virology, Beijing Key Laboratory of Etiology of Viral Diseases in Children, Capital Institute of Pediatrics, Beijing, China; ⁴ Henan Provincial Center for Disease Control and Prevention, Zhengzhou City, Henan Province, China; ⁵ Department of Infectious Diseases, Children's Hospital Affiliated to Chongqing Medical University, Chongqing, China; ⁶ Master's Degree Virus Laboratory, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou City, Guangdong Province, China; ⁷ Shenyang Center for Disease Control and Prevention (Shenyang Health Inspection Institute), Shenyang City, Liaoning Province, China; ⁸ Inner Mongolia Autonomous Region Academy of Preventive Medicine, Inner Mongolia Autonomous Region Center for Disease Control and Prevention, Hohhot City, Inner Mongolia Autonomous Region, China; ⁹ Laboratory of Viral Diseases, Qingdao Municipal Centre for Disease Control and Prevention, Qingdao Institute of Prevention Medicine, Qingdao City, Shandong Province, China; ¹⁰ Jining Medical University, Jining City, Shandong Province, China.

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: April 02, 2026

Accepted: May 13, 2026

Issued: May 22, 2026

REFERENCES

- Wang X, Li Y, Shi T, Bont LJ, Chu HY, Zar HJ, et al. Global disease burden of and risk factors for acute lower respiratory infections caused by respiratory syncytial virus in preterm infants and young children in 2019: a systematic review and meta-analysis of aggregated and individual participant data. *Lancet* 2024;403(10433):1241 – 53. [https://doi.org/10.1016/S0140-6736\(24\)00138-7](https://doi.org/10.1016/S0140-6736(24)00138-7).
- Mufson MA, Örvell C, Rafnar B, Norrby E. Two distinct subtypes of human respiratory syncytial virus. *J Gen Virol* 1985;66(Pt 10):2111-24. <http://dx.doi.org/10.1099/0022-1317-66-10-2111>.
- Langedijk AC, Vrancken B, Lebbink RJ, Wilkins D, Kelly EJ, Baraldi E, et al. The genomic evolutionary dynamics and global circulation patterns of respiratory syncytial virus. *Nat Commun* 2024;15(1):3083. <https://doi.org/10.1038/s41467-024-47118-6>.
- Eshaghi A, Duvvuri VR, Lai R, Nadarajah JT, Li AM, Patel SN, et al. Genetic variability of human respiratory syncytial virus A strains circulating in Ontario: a novel genotype with a 72 nucleotide G gene duplication. *PLoS One* 2012;7(3):e32807. <https://doi.org/10.1371/journal.pone.0032807>.
- Goya S, Ruis C, Neher RA, Meijer A, Aziz A, Hinrichs AS, et al. Standardized phylogenetic classification of human respiratory syncytial virus below the subgroup level. *Emerg Infect Dis* 2024;30(8):1631 – 41. <https://doi.org/10.3201/eid3008.240209>.
- Cantú-Flores K, Rivera-Alfaro G, Muñoz-Escalante JC, Noyola DE. Global distribution of respiratory syncytial virus A and B infections: a systematic review. *Pathog Glob Health* 2022;116(7):398 – 409. <https://doi.org/10.1080/20477724.2022.2038053>.
- Yu JM, Fu YH, Peng XL, Zheng YP, He JS. Genetic diversity and molecular evolution of human respiratory syncytial virus A and B. *Sci Rep* 2021;11(1):12941. <https://doi.org/10.1038/s41598-021-92435-1>.
- Guo CY, Zhang Y, Zhang YY, Zhao W, Peng XL, Zheng YP, et al. Comparative analysis of human respiratory syncytial virus evolutionary patterns during the COVID-19 pandemic and pre-pandemic periods. *Front Microbiol* 2023;14:1298026. <https://doi.org/10.3389/fmicb.2023.1298026>.
- Song JJ, Zhu Z, Song JH, Mao NY, Cui AL, Xu WB, et al. Circulation pattern and genetic variation of human respiratory syncytial virus in China during 2008-2021. *J Med Virol* 2023;95(3):e28611. <https://doi.org/10.1002/jmv.28611>.
- Yan B, Wang JF, Xu CD, Liu J. Spatial transmission dynamics of respiratory syncytial virus A in China. *Transbound Emerg Dis* 2025;2025(1):9926198. <https://doi.org/10.1155/tbed/9926198>.
- Zhuang TX, Fall A, Norton JM, Abdullah O, Villafuerte DA, Pekosz A, et al. Whole-genome sequence characterization of respiratory syncytial virus in the Johns Hopkins Health System during the 2024–2025 respiratory season. *Microbiol Spectr* 2025;13(11):e02065 – 25. <https://doi.org/10.1128/spectrum.02065-25>.
- Bergeron HC, Murray J, Arora A, Castrejon AMN, DuBois RM, Anderson LJ, et al. Immune prophylaxis targeting the respiratory syncytial virus (RSV) G protein. *Viruses* 2023;15(5):1067. <https://doi.org/10.3390/v15051067>.
- Botosso VF, De A Zanotto PM, Ueda M, Arruda E, Gilio AE, Vieira SE, et al. Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. *PLoS Pathog* 2009;5(1):e1000254. <https://doi.org/10.1371/journal.ppat.1000254>.
- García-Beato R, Melero JA. The C-terminal third of human respiratory syncytial virus attachment (G) protein is partially resistant to protease digestion and is glycosylated in a cell-type-specific manner. *J Gen Virol* 2000;81(Pt 4):919-27. <http://dx.doi.org/10.1099/0022-1317-81-4-919>.
- Fuentes S, Coyle EM, Golding H, Khurana S. Nonglycosylated G-protein vaccine protects against homologous and heterologous respiratory syncytial virus (RSV) challenge, while glycosylated G enhances RSV lung pathology and cytokine levels. *J Virol* 2015;89(16):8193 – 205. <https://doi.org/10.1128/JVI.00133-15>.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. List of the 52 G-HVR2 sequences in this study.

No. strain	Year	PLAD	NMDC number	No. strain	Year	PLAD	NMDC number
Changchun12-10	2012	Jilin	NMDCN0009ROK	JL-01-2021-934	2021	Jilin	NMDCN0009RPE
Changchun13-12	2013	Jilin	NMDCN0009ROL	JL-01-2021-841	2021	Jilin	NMDCN0009RPF
Changchun14-52	2014	Jilin	NMDCN0009ROM	JL-01-2021-830	2021	Jilin	NMDCN0009RPG
HeN17-21	2017	Henan	NMDCN0009RON	BJ-01-2021-315	2021	Beijing	NMDCN0009RPH
HeN17-18	2017	Henan	NMDCN0009ROO	JL-01-2022-54	2022	Jilin	NMDCN0009RPI
HeN17-15	2017	Henan	NMDCN0009ROP	JL-01-2022-133	2022	Jilin	NMDCN0009RPJ
HeN17-09	2017	Henan	NMDCN0009ROQ	JL-01-2022-103	2022	Jilin	NMDCN0009RPK
HeN17-05	2017	Henan	NMDCN0009ROR	CQ-01-2022-19	2022	Chongqing	NMDCN0009RPL
HeN17-01	2017	Henan	NMDCN0009ROS	CQ-01-2022-10	2022	Chongqing	NMDCN0009RPM
BJ17-29	2017	Beijing	NMDCN0009ROT	BJ-02-2023-139	2023	Beijing	NMDCN0009RPN
BJ17-11	2017	Beijing	NMDCN0009ROU	JL-01-2023-635	2023	Jilin	NMDCN0009RPO
BJ17-08	2017	Beijing	NMDCN0009ROV	HN-01-2023461	2023	Henan	NMDCN0009RPP
BJ17-02	2017	Beijing	NMDCN0009RP0	GD-01-2023-97	2023	Guangdong	NMDCN0009RPQ
HeN18-08	2018	Henan	NMDCN0009RP1	GD-01-2023-210	2023	Guangdong	NMDCN0009RPR
HeN18-03	2018	Henan	NMDCN0009RP2	CQ-01-2023149	2023	Chongqing	NMDCN0009RPS
BJ18-03	2018	Beijing	NMDCN0009RP3	CQ-01-2023129	2023	Chongqing	NMDCN0009RPT
BJ18-02	2018	Beijing	NMDCN0009RP4	CQ-01-2023-072	2023	Chongqing	NMDCN0009RPU
SH19-11	2019	Shanghai	NMDCN0009RP5	CQ-01-2023-034	2023	Chongqing	NMDCN0009RPV
SH19-10	2019	Shanghai	NMDCN0009RP6	CQ-01-2023-031	2023	Chongqing	NMDCN0009RQ0
SH19-08	2019	Shanghai	NMDCN0009RP7	BJ-02-2023-323	2023	Beijing	NMDCN0009RQ1
BJ19-07	2019	Beijing	NMDCN0009RP8	BJ-02-2023-311	2023	Beijing	NMDCN0009RQ2
BJ19-03	2019	Beijing	NMDCN0009RP9	BJ-02-2023-100	2023	Beijing	NMDCN0009RQ3
SD-01-2021-291	2021	Shandong	NMDCN0009RPA	JL-01-2024-67	2024	Jilin	NMDCN0009RQ4
NM01-2021-14	2021	Inner Mongolia	NMDCN0009RPB	JL-01-2024-35	2024	Jilin	NMDCN0009RQ5
LN-01-2021-078	2021	Liaoning	NMDCN0009RPC	JL01-2024-303	2024	Jilin	NMDCN0009RQ6
JL-01-2021-936	2021	Jilin	NMDCN0009RPD	GD-02-2024018	2024	Guangdong	NMDCN0009RQ7

Abbreviation: G-HVR2=second hypervariable region of the G gene; PLAD=provincial-level administrative division; NMDC=national microbiome data collaborative.

SUPPLEMENTARY TABLE S2. The spatiotemporal distribution of four A.D lineages.

Lineages	Global distribution			Distribution in China			
	No. of strains	No. of countries	Period	No. of strains	No. of PLADs	Period	Distribution by PLAD
A.D.1	763	21	2012–2024	8	6	2019; 2021; 2023	Beijing; Shanghai; Shandong; Jilin; Liaoning; Inner Mongolia; Taiwan
A.D.2	501	27	2012–2024	24	6	2012; 2014; 2017–2020	Beijing; Hubei; Henan; Shanghai; Shandong; Zhejiang; Jiangsu
A.D.3	698	25	2010–2024	106	11	2012–2015; 2017–2020; 2021–2024	Beijing; Jilin; Henan; Shanghai; Guangdong; Zhejiang; Taiwan; Chongqing; Gansu; Ningxia; Yunnan
A.D.4/ A.D.5	492	22	2011–2024	6	5	2018; 2021; 2023	Beijing; Henan; Guangdong; Zhejiang; Taiwan

Abbreviation: NMDC=national microbiome data collaborative.

SUPPLEMENTARY TABLE S3. Inter- and intra-lineage nucleotide and amino acid genetic distances for the four lineages of A.D.

Lineages	A.D.1	A.D.2	A.D.3	A.D.4/A.D.5
A.D.1	0.027/aa0.076*	0.086	0.074	0.093
A.D.2	0.037	0.035/aa0.052*	0.084	0.121
A.D.3	0.037	0.040	0.027/aa0.051*	0.111
A.D.4/A.D.5	0.040	0.048	0.045	0.028/aa0.060*
Overall	0.040/aa0.083			

Note: Interlineage genetic distances based on nucleotide sequences are presented in the lower-left region, whereas those based on amino acid sequences are displayed in the upper-right region.

Abbreviation: aa=amino acid.

* The values represent intra-lineage genetic distances.

SUPPLEMENTARY TABLE S4. The estimated evolutionary rate and tMRCA of A.D lineages.

Lineage	Evolutionary rate (95% HPD)	tMRCA (95% HPD)
A.D	5.63×10^{-3} (1.00×10^{-3} , 11.28×10^{-3})	2006 (2005, 2010)
A.D.1	2.66×10^{-3} (3.12×10^{-6} , 7.29×10^{-3})	2009 (2009, 2011)
A.D.2	4.14×10^{-3} (5.01×10^{-5} , 10.89×10^{-3})	2009 (2009, 2011)
A.D.3	1.92×10^{-3} (4.18×10^{-6} , 6.36×10^{-3})	2008 (2008, 2011)
A.D.4/A.D.5	5.26×10^{-3} (1.93×10^{-5} , 13.01×10^{-3})	2009 (2009, 2011)

Abbreviation: HPD=highest posterior density; tMRCA=the most recent common ancestor.