

Preplanned Studies

Emergence and Global Circulation of the HRSV BA Variant with a 7-AA Extension in the G Protein Under Persistent BA Predominance — Global, 1996–2025

Jinhua Song¹; Ke Lu¹; Huiwen Liu¹; Baicheng Xia¹; Pengbo Yu²; Liwei Sun³; Linqing Zhao⁴; Xiangpeng Chen⁵; Zhaoguo Wang⁶; Bing Zhu⁷; Yonghao Guo⁸; Huan Wei⁹; Xin Li¹⁰; Lei Cao¹; Huiling Wang¹; Yan Zhang^{1, #}

Summary

What is already known about this topic?

Human respiratory syncytial virus variant with a 7-amino-acid extension in the G protein of the BA genotype (BA-7AA) is associated with severe neonatal pneumonia outbreak, high intensive care unit occupancy and death. However, limited data are available on the epidemiology and transmission dynamics of the BA-7AA variants.

What is added by this report?

This study integrated publicly available sequences of the second hypervariable region of the G gene of the BA-7AA variant with those sequenced in this study to clarify the emergence and transmission dynamics of the BA-7AA variant.

What are the implications for public health practice?

By revealing the circulation of the BA-7AA variant, this study established a basis for forecasting and warning of the reemergence of BA-7AA variants associated with severe cases and outbreaks.

v2.3.6 and Bayesian stochastic search variable selection methods in BEAST v10.5.

Results: The study included 501 BA-7AA HVR2-G sequences (37 sequenced variants and 464 sequences downloaded from GenBank). BA-7AA variants were initially detected in 2013 and spread gradually to 33 countries by 2024, peaking in 2016, when 27.0% of the BA genotypes were BA-7AA variants. According to the phylogenetic tree, no distinct clusters circulating over time or by region were found during the past decade.

Conclusion: BA-7AA variants circulated worldwide from 2013 to 2024. This work will allow for extensive surveillance to identify BA-7AA variants and provide timely warning of a re-emergence.

ABSTRACT

Introduction: Human respiratory syncytial virus (HRSV) BA genotype variant with a 7-amino-acid extension in the G protein (BA-7AA) has been associated with severe neonatal pneumonia outbreaks, high intensive care unit occupancy, and death. It is vital to understand BA-7AA variant emergence and global circulation.

Methods: The second hypervariable region of the G gene (HVR2-G) sequence dataset was integrated with HVR2-G and whole-genome sequences downloaded from GenBank prior to September 17, 2025, together with the sequences obtained in this study from 2013 to 2024 in China. Phylogenetic analysis and transmission pathways analyses were performed using IQTREE

Human respiratory syncytial virus (HRSV) is a leading cause of lower respiratory tract infections in infants and young children worldwide (1). Eighty-two HRSV outbreaks were reported worldwide between 1962 and 2025. The highest outbreak rates were observed during the winter in neonatal intensive care units (ICUs) (30.49%), with children accounting for the highest percentage of cases (42.29%). The case fatality rate was the highest among older adults across all age groups. The HRSV-B subgroup is the most frequently reported (2).

In 2019, an outbreak of severe neonatal pneumonia in a postpartum care center in China caused by the HRSV-B BA9 genotype was reported after phylogenetic analysis of the second hypervariable region of the G gene (HVR2-G) (3). A 7-amino-acid (Q-R-L-Q-S-Y-A) extension mutation after the stop codon of the G protein was identified (BA-7AA variant). A comparison of the biological properties and virulence of infectious clones of the wild-type strain and a clone with a 7-amino-acid deletion revealed that

the mutated strain exhibited increased adaptability in mice and enhanced viral virulence (4). In addition, strains with a 7-amino-acid extension have also been reported to be associated with a high ICU occupancy rate and fatalities (5–6).

While BA genotypes have been well established as the dominant genotype since 2004–2005, limited data are available regarding the epidemiology and transmission dynamics of the BA-7AA variant (5–8). To clarify the molecular epidemiology of the BA-7AA variant, strains collected in China from 2013 to 2024 were sequenced, and phylogenetic analyses were performed together with sequences downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/>).

Total RNA was extracted from nasopharyngeal or throat swabs of hospitalized patients with acute respiratory tract infection from eight regions (Jilin, Shaanxi, Gansu, Henan, Liaoning, Guangdong, Shandong, and Beijing) under the Chinese HRSV surveillance system, using the QIAamp RNA Mini Kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. Reverse transcription–polymerase chain reaction (RT-PCR) and sequencing were performed as previously described (3).

A phylogenetic tree was constructed using IQTREE (version 2.3.6, <https://github.com/iqtree/iqtree2>) with the best-fit model TN+I+R3 chosen according to the Bayesian information criterion. Branch support values were estimated using an ultrafast bootstrap with 1000 replicates and an approximate likelihood ratio test with 1000 replicates. Tree visualization and editing were performed using TVBOT online (version 2.6.1, <https://www.chiplot.online/>, Shantou University, Shantou, China). To understand the transmission pathways of the BA-7AA variant, the Bayesian stochastic search variable selection (BSSVS) method in BEAST software (version 10.5, University of Auckland, Auckland, New Zealand) was used. The migration routes were summarized using Spread3 (version 0.9.7, <https://beast.community/spread3>) (9). A Bayes factor ≥ 3 and a posterior probability value > 0.5 were considered to significantly support the pairwise diffusion pathways. Pathways were visualized using Gephi software (version 0.10.1, <https://github.com/gephi>).

The HVR2-G and whole-genome sequences of HRSV were downloaded from GenBank on September 17, 2025, and the HVR2-G region was extracted from the whole genome to identify the BA-7AA variant. All the HVR2-G sequences were aligned with the reference HVR2-G sequence. Each sequence was

truncated to the stop codon of the G protein. Sequences with terminal deletions at either the 5' or 3' end exceeding 10% of the sequence length or with $>2.5\%$ ambiguous bases were excluded. The BA-7AA variant was used for phylogenetic analyses (Figure 1). To perform BSSVS analyses, one hundred thirty-three representative BA-7AA variant sequences reported in each country were collected. The BA-7AA variants sequenced in this study were submitted to the National Microbiology Data Center (NMDC) under accession numbers–NMDCN000ATPQ–NMDCN000ATQU (<https://nmdc.cn/resource/genomics/sequence>).

A total of 8,705 sequences were downloaded from GenBank, including 173 sequences without the 60-nt insertion that were excluded. Among the remaining 8,532 HVR2-G sequences, representing the dominant BA genotype between 1996 and 2025, there were 464 BA-7AA variants, which were combined with the 37 BA-7AA variants sequenced in our study to yield a

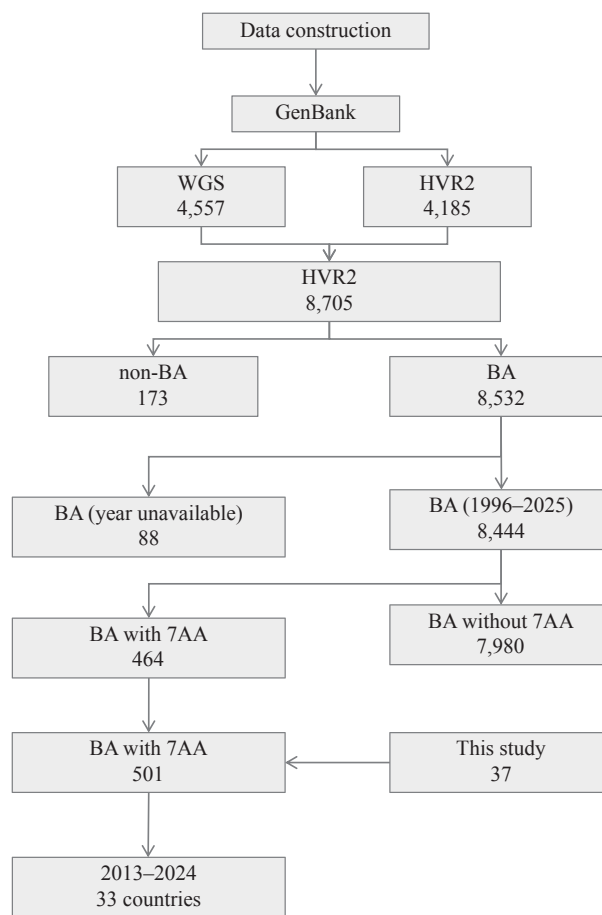


FIGURE 1. Dataset construction of HVR2-G of HRSV-B strains.

Abbreviation: HVR2=second hypervariable region; HRSV=human respiratory syncytial virus; WGS=whole genome sequence; BA=BA genotype.

total of 501 BA-7AA variants (Figure 1).

The BA-7AA variants were mostly detected between 2016 and 2018, with the highest numbers reported in Germany, Spain, Kenya, and China (Figure 2A). The BA-7AA variant was first detected in Shaanxi Province, China in 2013 (1.1% of the BA genotype). It accounted for 2.1% of BA cases reported in China, Germany, and Panama in 2014. By 2015, its prevalence had increased to 9.5% of BA cases in eight countries and peaked in 2016, accounting for 27.0% of the BA genotype, with reports from 12 countries. In 2017, it accounted for 19.6% of BA cases across 17 countries, and in 2018, it accounted for 11.3% of the BA genotypes across 15 countries. Finally, in 2019, it accounted for 6.4% of BA genotypes across 12 countries. Thereafter, the incidence decreased annually, although the BA-7AA variant continues to be detected in new countries. From 2022 to 2024, as the number of detected strains decreased, no new country reported the variant, and it was not detected in countries where it had previously been prevalent. By 2024, only one sequence was detected in Spain and China. As of September 2025, no BA-7AA variant had been identified in GenBank (Figure 2B and 2C).

A phylogenetic tree was constructed using the 501 BA-7AA variant sequences and 21 reference genotype strains. The results demonstrated that all BA-7AA sequences were grouped into a single cluster with the BA (BA9) genotype reference strains. Sequences from different periods and geographic regions were intermixed and no discernible temporal or geographic patterns were observed. Furthermore, after 11 years of circulation, no distinct branches with high bootstrap support values were identified in the phylogenetic tree (Figure 3, the sequence information used for phylogenetic analysis is available at <https://doi.org/10.57760/sciencedb.011nd>).

Based on the phylogenetic tree results, we attempted to predict the transmission pathways of BA-7AA variants using BSSVS analysis. Nine transmission pathways were revealed with a posterior probability greater than 0.5 using 133 representative variants (The sequences used for BSSVS analysis are available at <https://doi.org/10.57760/sciencedb.011nd>). Among these pathways, those from Germany to South Africa had the highest Bayes factor with a posterior probability of 0.997. The pathway from China to Australia received the second highest level of support, with a posterior probability of 0.992. The Bayes factors

and posterior probabilities for the remaining seven pathways were all less than 200 and 0.9, respectively (Figure 4).

DISCUSSION

As one of the major membrane surface glycoproteins, the G protein mediates the adhesion of viral particles to the host cell membrane, facilitating efficient viral binding, cell fusion, and viral release. The G protein contains a CX3C chemokine mimic motif, alters leukocyte migration patterns, and inhibits type I interferon production and antiviral immunity (10–11). Moreover, HVR2-G was chosen as the target for virus surveillance because of its highly variable characteristics, such as a 60-nt duplicated insertion in HRSV-B (BA genotype) and a 72-nt duplicated insertion in HRSV-A (ON1 genotype). In 2024, the HRSV genotyping consensus consortium defined the criteria for HRSV genetic typing and renamed these two genotypes, as A.D. and B.D. based on whole-genome sequences (12), which have evolved into six more lineages according to Nextstrain (<https://nextstrain.org/>).

Studies have shown that the duplicated region in the BA strain G protein augments viral attachment and fitness, and that the 72-nt insertion in the ON1 strain increases the efficiency of pseudotype lentiviral particle infection and enhances attachment function (13–14). Although the BA-7AA variant has a broad geographic distribution, it exhibits limited genetic variation during circulation. Nevertheless, the 7-AA insertion in the BA genotype improved viral adaptation in mice and enhanced viral virulence, which may contribute to its low prevalence and relatively low evolutionary divergence. Given the cost-effectiveness and practical utility of the HVR2-G region for virological surveillance, it is particularly valuable for studies investigating viral functions and characteristics in resource-limited regions.

Although the BA-7AA variant first appeared in the database of this study in 2013, it did not attract substantial attention until it was associated with the 2019 outbreak. This highlights deficiencies in the surveillance system, lack of timely coordination between clinical and public health efforts, and gaps in epidemiological information. Molecular epidemiological surveillance should be strengthened to enable prompt detection of variant strains. Furthermore, although the application of artificial

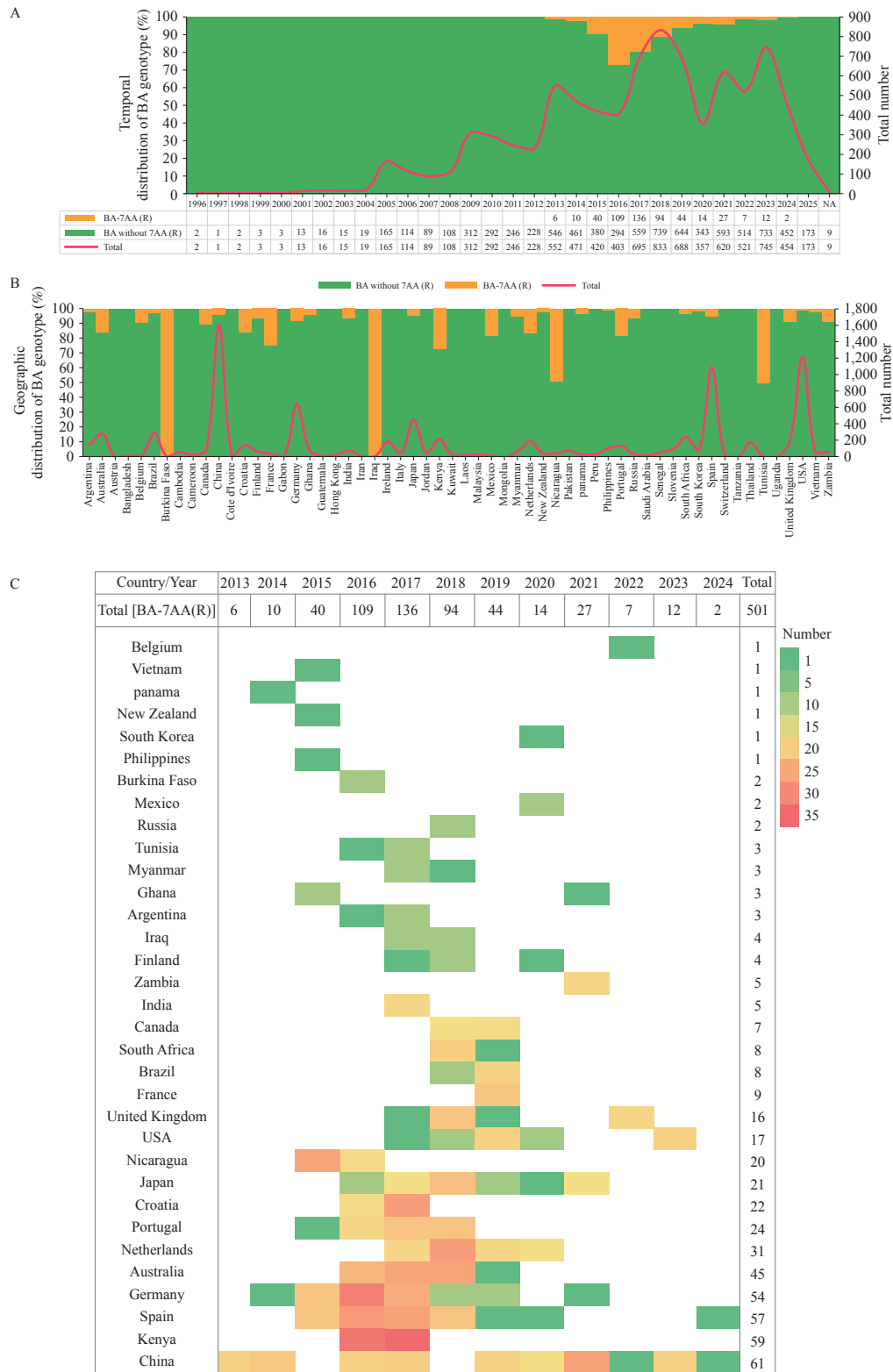


FIGURE 2. Temporal and geographic distributions of BA genotypes of HRSV subgroup B and the BA-7AA variants from 2013 to 2024. The total number of BA sequences is shown using a red line by (A) years or (B) regions; (C) Sequence numbers of BA-7AA by country and year.

Note: The percentages of BA-7AA variant and BA genotype without 7AA are colored with yellow and green bars, respectively, in panels (A) and (B).

Abbreviation: HRSV=human respiratory syncytial virus; NA=not available; USA=United States of America.

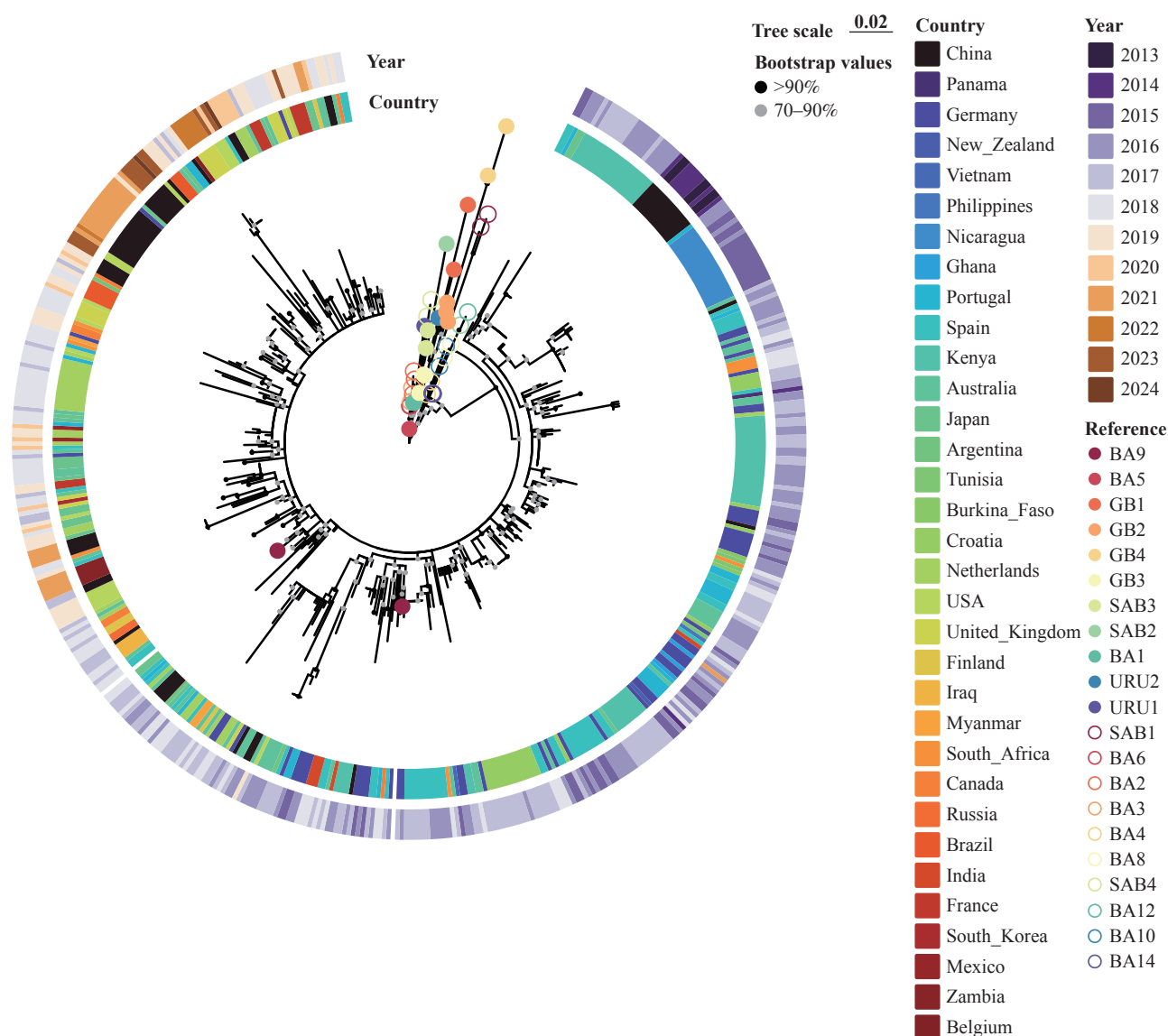


FIGURE 3. Phylogenetic tree of HVR2-G sequences of HRSV BA-7AA variants from 2013 to 2024.

Note: The inner ring shows the collection country of each sequence, and the outer ring shows the collection year of each sequence. Reference strains were colored with dots or rings.

Abbreviation: HVR2-G=second hypervariable region of the G gene; HRSV=human respiratory syncytial virus; USA=United States of America.

intelligence in the field of infectious diseases faces some challenges, it is possible to integrate AI-based predictions using a large accumulation of real-world data to enable an early warning system (15). Therefore, continuous monitoring of HRSV should be conducted to identify highly pathogenic strains, prevent healthcare system overload and outbreaks, and provide foundational support for the prediction, warning, prevention, and control of HRSV infection.

This study had several limitations. First, considerable variation in surveillance and sequencing efforts across countries may artificially inflate or

obscure certain patterns. Therefore, our results represent an exploratory descriptive analysis rather than a definitive account of the global spread of the virus. The absence of the BA-7AA variant by 2025 may be due to a delay in the current surveillance system. The lack of detailed epidemiological data and use of short fragments may introduce bias in the inference of transmission routes. Additionally, a comparative epidemiological analysis between the BA-7AA variants and BA strains without the 7-AA insertion was not feasible due to insufficient clinical metadata in our surveillance database and GenBank. However, the

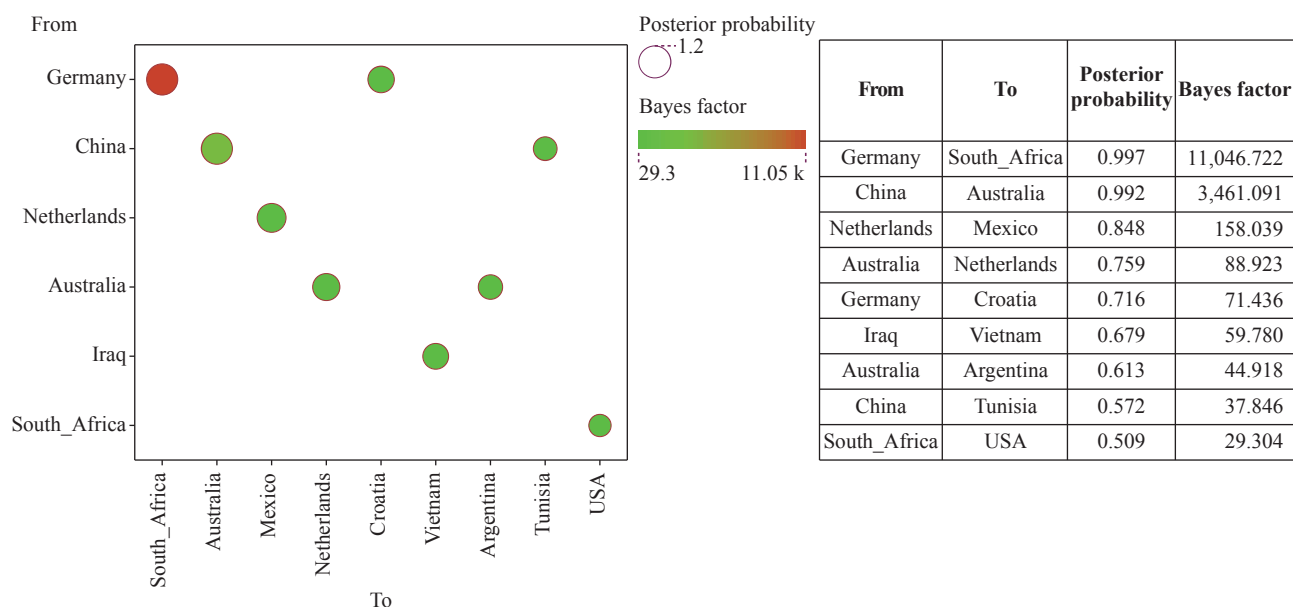


FIGURE 4. Spatial dispersal patterns of the HRSV-B BA-7AA variants, determined by Bayesian phylogeographic inference of HVR2-G sequences.

Note: The size of the circle represents the value of the posterior probability, and the Bayes factor is shown using a gradient from green to red.

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

public health implications of BA-7AA require further investigation using prospective clinical data.

In conclusion, this study demonstrates the widespread circulation of the BA-7AA variant and enriches the HRSV surveillance database. Given its potentially increased pathogenicity, strengthened virological surveillance is warranted to predict the re-emergence of variants associated with severe cases and outbreaks.

Conflicts of interest: No conflicts of interest.

Acknowledgements: The staff of the sentinel hospitals of the HRSV surveillance project and the authors, who originated and submitted the sequences from the GenBank database.

Funding: Supported by grants from the Prevention and Control of Emerging and Major Infectious Diseases, National Science and Technology Major Project (2025ZD01900400).

doi: 10.46234/ccdcw2026.185

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

¹ National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, NHC Key Laboratory of Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China; ² Shaanxi Provincial Center for Disease Control and Prevention, Xi'an City, Shaanxi Province, China; ³ Children's Hospital of Changchun, Jilin Children's Medical Center, Changchun City, Jilin Province, China; ⁴ Beijing Key Laboratory of Etiology of Viral Diseases in Children,

Capital Institute of Pediatrics, Beijing, China; ⁵ Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China; ⁶ Qingdao Municipal Center for Disease Control and Prevention, Qingdao City, Shandong Province, China; ⁷ Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou City, Guangdong Province, China; ⁸ Henan Provincial Center for Disease Control and Prevention, Zhengzhou City, Henan Province, China; ⁹ Gansu Provincial Center for Disease Control and Prevention, Lanzhou City, Gansu Province, China; ¹⁰ Shenyang Center for Disease Control and Prevention, Shenyang City, Liaoning 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 28, 2026

Accepted: August 31, 2026

Issued: September 11, 2026

REFERENCES

- Li Y, Wang X, Blau DM, Caballero MT, Feikin DR, Gill CJ, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis. *Lancet* 2022;399(10340):2047 – 64. [https://doi.org/10.1016/S0140-6736\(22\)00478-0](https://doi.org/10.1016/S0140-6736(22)00478-0).
- Duan QY, Pan JR, Zheng L, Jin ZW, Li SJ, Zhang HJ, et al. Global outbreaks of respiratory syncytial virus infections from 1960 to 2025: a systematic review and meta-analysis. *eClinicalMedicine* 2025;86: 103352. <https://doi.org/10.1016/j.eclinm.2025.103352>.
- Wang B, Song JJ, Song JH, Mao NY, Liang JY, Chen Y, et al. An outbreak of severe neonatal pneumonia caused by human respiratory syncytial virus BA9 in a postpartum care centre in Shenyang, China. *Microbiol Spectr* 2022;10(4):e00974 – 22. <https://doi.org/10.1128/>

- spectrum.00974-22.
4. Wang N, Song JJ, Cao L, Mao NY, Shi YQ, Jiang J, et al. Enhanced pathogenic consequences induced by a seven-amino-acid extension in the G protein of the HRSV BA9 genotype. *Int J Mol Sci* 2025;26(5): 2081. <https://doi.org/10.3390/ijms26052081>.
 5. Xu LL, Gao HM, Zeng JS, Liu J, Lu C, Guan XL, et al. A fatal case associated with respiratory syncytial virus infection in a young child. *BMC Infect Dis* 2018;18(1):217. <https://doi.org/10.1186/s12879-018-3123-8>.
 6. Jiang ML, Wang FJ, Han ZZ, Xu YP, Sun Y, Zhu RN, et al. A new variant of respiratory syncytial virus glycoprotein gene and clinical features of infected children. *Chin J Microbiol Immunol* 2024;44(2): 120 – 7. (In Chinese). <https://doi.org/10.3760/cma.j.cn112309-20240112-00021>.
 7. 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>.
 8. 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>.
 9. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. Spred3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol* 2016;33(8):2167 – 9. <https://doi.org/10.1093/molbev/msw082>.
 10. Anderson L, Jadhao S, Paden C, Tong SX. Functional features of the respiratory syncytial virus G protein. *Viruses* 2021;13(7):1214. <https://doi.org/10.3390/v13071214>.
 11. Bergeron HC, Tripp RA. RSV replication, transmission, and disease are influenced by the RSV G protein. *Viruses* 2022;14(11):2396. <https://doi.org/10.3390/v14112396>.
 12. 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>.
 13. Hotard AL, Laikhter E, Brooks K, Hartert TV, Moore ML. Functional analysis of the 60-nucleotide duplication in the respiratory syncytial virus Buenos Aires strain attachment glycoprotein. *J Virol* 2015;89(16): 8258 – 66. <https://doi.org/10.1128/JVI.01045-15>.
 14. Cui GL, Liu HC, Li XG, Ming L. Preliminary functional and phylogeographic analyses of the 72 nucleotide duplication region in the emerging human respiratory syncytial virus ON1 strain attachment glycoprotein gene. *Biomed Pharmacother* 2020;123:109800. <https://doi.org/10.1016/j.biopha.2019.109800>.
 15. Kamyshnyi O, Halabitska I, Oksenysh V, Kamyshna I, Petakh P, Kainov DE. Forecasting influenza epidemics and pandemics in the age of AI and machine learning. *Rev Med Virol* 2026;36(1):e70107. <https://doi.org/10.1002/rmv.70107>.