Importation and circulation of rubella virus lineages 1E-L2 and 2B-L2c between 2018 and 2021 in China: Virus evolution and spatial–temporal transmission characteristics

Ying Liu,1,†,‡ Zhenzhi Han,1,2,† Chuyun Kang,3 Aili Cui,3 Yan Zhang,1 Li Liu,4 Ying Chen,5 Lili Deng,6 Hua Zhao,7 Jun Zhou,8 Fangcai Li,9 Shuji Zhou,10 Daxing Feng,11 Xiaoling Tian,12 Yan Feng,13 Xiaoxian Cui,14 Yue Lei,15 Yan Wang,16 Fang Yuan,17 Lixia Fan,18 Xiaomin Tang,19 Meng Chen,20 Xiaofang Peng,21 Yu Guo,22 Hui Gao,23 Suting Wang,24 Liqun Li,25 Ting Zhang,26 Xiuying Deng,27 Haiyun Chen,28 Shuang Wang,29 Yu Ma,30 Zhen Zhu,31 and Wenbo Xu1,3

1WHO WPRO Regional Reference Measles/Rubella Laboratory, NHK Key Laboratory of Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, No. 155, Changbai Road, Changping District, Beijing 102206, China, 2Laboratory of Virology, Beijing Key Laboratory of Etiology of Viral Diseases in Children, Capital Institute of Pediatrics, Beijing, China, 3Department of Maternal and Child Health, School of Public Health, Peking University, Beijing, China, 4Institute of Microbiology, Sichuan Provincial Center for Disease Control and Prevention, Chengdu City, Sichuan Province China, 5Department of Immunization Program, Gansu Provincial Center for Disease Control and Prevention, Lanzhou City, Gansu Province China, 6Department of Expanded Programme on Immunization, Guangxi Provincial Center for Disease Control and Prevention, Nanning City, Guangxi Province, China, 7Department of Microbiological Testing, Chongqing Provincial Center for Disease Control and Prevention, Chongqing, China, 8Institute of Virology, Jiangxi Provincial Center for Disease Control and Prevention and Prevention, Nanchang City, Jiangxi Province, China, 9Department of Microbiological Testing, Hunan Provincial Center for Disease Control and Prevention, Changsha City, Hunan Province, China, 10Department of Expanded Programme on Immunization, Anhui Provincial Center for Disease Control and Prevention, Hefei City, Anhui Province, China, 11Department of Expanded Programme on Immunization, Henan Provincial Center for Disease Control and Prevention, Zhengzhou City, Henan Province, China, 12Department of Immunization Program, Neimeng Provincial Center for Disease Control and Prevention, Huhehaote City, Neimeng Province, China, 13Department of Immunization Program, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou City, Zhejiang Province, China, 14Division of Microbiology, Shanghai Provincial Center for Disease Control and Prevention, Shanghai, China, 15Department of Pathogenic Microbiology, Tianjin Provincial Center for Disease Control and Prevention, Tianjin, China, 16Department of Immunization Program, Liaoning Provincial Center for Disease Control and Prevention, Shenyang City, Liaoning Province, China, 17Department of Virology, Ningxia Provincial Center for Disease Control and Prevention, Yinchuan City, Ningxia Province, China, 18Inspection and Testing Center, Qinghai Provincial Center for Disease Control and Prevention, Xining City, Qinghai Province, China, 19Department of Virology, Guizhou Provincial Center for Disease Control and Prevention, Guiyang City, Guizhou Province, China, 20Immunization Prevention Institute, Beijing Provincial Center for Disease Control and Prevention, Beijing, China, 21Institute of Immunization, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou City, Guangdong Province, China, 22Institute of Immunization, Hebei Provincial Center for Disease Control and Prevention, Shijiazhuang City, Hebei Province, China, 23Department of Disease Inspection, Shanxi Provincial Center for Disease Control and Prevention, Taiyuan City, Shanxi Province, China, 24Department of Expanded Programme on Immunization, Shandong Provincial Center for Disease Control and Prevention, Jinan City, Shandong Province, China, 25Department of Immunization Program, Yunnan Provincial Center for Disease Control and Prevention, Kunming City, Yunnan Province, China, 26Virus Detection Department, Institute of Inspection and Testing, Hubei Provincial Center for Disease Control and Prevention, Wuhan City, Hubei Province, China, 27Department of Expanded Programme on Immunization, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing City, Jiangsu Province, China, 28Microbiological Laboratory, Testing and Inspection Institute, Hainan Provincial Center for Disease Control and Prevention, Haikou City, Hainan Province, China, 29Department of Viral Disease Control and Prevention, Jilin Provincial Center for Disease Control and Prevention, Changchun City, Jilin Province, China, 30Immunization Planning Institute, Shaanxi Provincial Center for Disease Control and Prevention, Xi’an City, Shaanxi Province, China

1These authors contributed equally to this work.
2https://orcid.org/0000-0003-4030-2778
3https://orcid.org/0000-0003-1772-7946
*Corresponding authors: E-mail: zhuzhen76@163.com; wenbo_xu1@aliyun.com

Abstract

To better understand the importation and circulation patterns of rubella virus lineages 1E-L2 and 2B-L2c circulating in China since 2018, 3,312 viral strains collected from 27 out of 31 provinces in China between 2018 and 2021 were sequenced and analyzed with the representative international strains of lineages 1E-L2 and 2B-L2c based on genotyping region. Time-scale phylogenetic analysis revealed that the global lineages 1E-L2 and 2B-L2c presented distinct evolutionary patterns. Lineage 1E-L2 circulated in relatively limited geographical areas (mainly Asia) and showed geographical and temporal clustering, while lineage 2B-L2c strains circulated widely throughout the world and exhibited a complicated topology with several independently evolved branches. Furthermore, both lineages showed extensive international transmission activities, and phylogeographic inference provided evidence that lineage 1E-L2 strains circulating in China possibly originated from Japan, while the source of lineage 2B-L2c isolated since 2018 is still unclear. After importation into China in 2018, the spread of lineage 1E-L2 presented a three-stage transmission pattern from southern to northern China, whereas lineage 2B-L2c spread from a single point in western China to all the other four regions. These two transmission patterns allowed both
imported lineages to spread rapidly across China during the 2018–9 rubella epidemic and eventually established endemic circulations. This study provides critical scientific data for rubella control and elimination in China and worldwide.

**Key words:** rubella virus; importation; spatial–temporal transmission; virus evolution

## 1. Introduction

Rubella, also known as ‘German measles’, presents as a mild and self-limiting disease (Lambert et al. 2015). However, threats to public health arise due to its teratogenic effects during pregnancy, which can lead to miscarriage or congenital rubella syndrome (CRS) in the fetus (Kaushik, Verma, and Kumar 2018). Although much progress has been made toward rubella and CRS elimination worldwide with ninety-three countries and regions declaring successfull control of rubella as of 2020 (Zimmerman 2022), the disease remains a global concern. It has been estimated that over 100,000 neonates are born with CRS every year (Vynnycky et al. 2017). The suboptimal coverage of population immunity in select countries has led to its intermittent resurgence and outbreak (Mori et al. 2017; Ujiie 2019).

Rubella virus (RuV), the identified pathogen that causes rubella, is a recognized member of the *Rubivirus* genus within the *Mataonaviridae* family (Bennett et al. 2020). It contains a single-stranded positive-sense RNA genome that is approximately 9,762 nt in length (Frey 1994). The coding region of structural protein, located at the 3′-end position, produces three structural proteins, C, E2, and E1. E1 glycoprotein, which includes 481 amino acids, is a Type I transmembrane protein that plays a decisive role in viral recognition, cellular attachment, and membrane fusion (DuBois et al. 2013; Mangala Prasad, Klose, and Rossmann 2017). Given its key role in RuV, a 739-nt fragment within the E1 gene (nt8731–9469, E1-739) was recommended by the World Health Organization (WHO) as a target gene for genotyping identification and regional molecular surveillance (WHO 2005). To date, two clades, including twelve genotypes and one provisional genotype, have been classified and only two genotypes, 1E and 2B, have been detected worldwide since 2018, according to the global rubella nucleotide surveillance (RubNeS) database (Brown et al. 2019). With the gradual increase in the genetic diversity of genotypes 1E and 2B, the genetic distance threshold for lineage division greater than 1.5 per cent is considered as the standard for lineage classification in order to improve the resolution of sequence data for molecular epidemiology (Rivailler, Abernathy, and Icenogle 2017).

In China, rubella incidence has gradually declined, reaching a record low in 2017 following the introduction of an effective rubella-containing vaccine as part of the expanded immunization program in 2018 (Su et al. 2018). Furthermore, with high vaccination coverage in infants (>95 per cent), the endemic circulations of lineages 1E-L1 and 2B-L1 were successfully interrupted in 2016 and 2019, respectively. However, due to the importation of lineages 1E-L2 and 2B-L2c and the accumulation in the number of susceptible teenagers aged 10–19 years, a rubella resurgence and outbreak occurred in China between 2018 and 2019 (Zhu et al. 2021). Therefore, to better understand the importation and circulation of lineage 1E-L2 and 2B-L2c in China, the virus sequences corresponding to the 2018–21 period, which were obtained from the Chinese Measles and Rubella Laboratory Network (CMRLN), were summarized. Thereafter, the characteristics of virus evolution and the spatial–temporal transmission were analyzed. The results of this study would contribute to optimizing public health intervention measures for eliminating rubella in China.

## 2. Materials and methods

### 2.1 Workflow of rubella virological surveillance in China

Rubella virological surveillance in China, which can be traced back to 1999, has greatly improved with support from the three-tiered CMRLN. With the official integration of rubella into the national case-based measles surveillance system in 2014, the CMRLN provided high-quality laboratory support for measles and rubella surveillance (Xu et al. 2017). In this study, according to surveillance guidelines (Xu et al. 2017), prefecture laboratories collected throat swabs from suspected rubella cases within 5 days after rash onset and verified the presence of rubella RNA via real-time reverse transcription–polymerase chain reaction (RT-PCR). Furthermore, provincial laboratories were responsible for virus isolation and genotyping-positive samples according to the previously described standard procedures (Zhu et al. 2007). For further confirmation and analysis, virus strains and their epidemiological information as well as genotyping sequences were referred to the National Laboratory.

### 2.2 Genotyping and lineage division

Viral RNA was extracted from the RuV strains using the QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA, USA). After the amplification of two overlapping fragments (F1, ntc8633–9112; F2, ntc8945–9577) using a one-step RT-PCR kit (Qiagen, Hilden, Germany), bidirectional sequencing was performed using an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Thereafter, sequences were edited and assembled using Sequencher software version 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA) to obtain the E1-739 region. The sequences obtained for this study were aligned with the sequences of WHO reference strains, which represented thirteen RuV genotypes (WHO 2013) and reported lineage reference strains of genotypes 1E and 2B (Zhu et al. 2021) using Multiple alignment program for amino acid or nucleotide sequences software (MAFFT) software version 7.490 (Katoh and Standley 2013). Finally, to determine the genotypes and lineages of the virus strains, a phylogenetic tree was constructed via the neighbor-joining method using Molecular Evolutionary Genetics Analysis software (MEGA) software version 7.0 (Kumar, Stecher, and Tamura 2016) based on the E1-739 region.

### 2.3 Representative strain selection and dataset construction

A total of 3,312 RuV strains were successfully identified in this study, with 1,947 strains of lineage 1E-L2 and 1,365 strains of lineage 2B-L2c collected through the CMRLN in China from 2018 to 2021 (Supplementary Table S1). To study the evolutionary situation as well as the temporal–spatial transmission characteristics of the strains of RuV lineages 1E-L2 and 2B-L2c, a total of four datasets were generated, including Chinese domestic and global databases for both lineages. The representative strains for all the datasets were randomly selected based on the chain of transmission to ensure the display of virus diversity, and the chronological and geographical distributions of
the virus strains were also considered to avoid bias. First, representative strains of lineages 1E-L2 (n = 315, from twenty-five provinces) and 2B-L2c (n = 292, from twenty-one provinces) in this study were selected to construct Chinese domestic datasets for these two lineages (1E-L2, Dataset 1; 2B-L2c, Dataset 2). Subsequently, the strain numbers for both lineages were further refined based on the temporal and geographical distributions and were included in the global datasets (1E-L2, \( n = 78 \); 2B-L2c, \( n = 52 \)) (Fig. 1).

The international strains, including all available genetic sequences (genotype 1E, 1,364; genotype 2B, 1,261), were downloaded from the GenBank database (data as of February 2022). After genotyping and lineage division, 538 lineage 1E-L2 strains and 158 lineage 2B-L2c strains within the E1-739 region were identified, and the representative global strains for lineages 1E-L2 (\( n = 106 \), from seven countries or regions corresponding to the 2001–20 period) and 2B-L2c (\( n = 154 \), from twenty-eight countries or regions corresponding to the 2005–20 period) were selected. Twenty-six sequences of lineage 2B-L2c in 2012 from China were also included (Supplementary Table S2).

Finally, 184 strains of lineage 1E-L2 (\( n = 78 \) in this study and \( n = 106 \) from GenBank) and 206 strains of lineage 2B-L2c (\( n = 52 \) in this study and \( n = 154 \) from GenBank) constituted the global datasets (namely, Datasets 3 and 4) (Fig. 1). It is also worth noting that when the imported strains in the GenBank database were documented with clear origin-related information, the accurate geographic information for these strains was recorded to improve the accuracy of virus transmission analysis.

### 2.4 Analysis of evolutionary characteristics

The phylogenetic tree generated based on the maximum likelihood phylogenies was constructed using the ultrafast bootstrap method (Hoang et al. 2018) in an efficient software for phylogenomic inference (IQ-TREE) version 1.6.12 (Nguyen et al. 2015). Furthermore, to examine the temporal signals associated with data sampling, root-to-tip regression analysis was performed using TempEst software version 1.5.3, while the ModelFinder model in IQ-TREE software version 1.6.12 was used to select the appropriate nucleotide substitution model for the datasets (Kalyaanamorthy et al. 2017). Bayesian phylogenetic inferences for lineages 1E-L2 and 2B-L2c were then obtained using the Markov Chain Monte Carlo method in BEAST software version 1.10.4, combining different molecular clock models and a coalescent tree prior for each dataset. The analysis was conducted with 300 million generations in each dataset, and sampling was performed at 3,000-step intervals. The convergence and the effective sample size (>200) were tested using Tracer software version 1.7.2 (Drummond et al. 2012). Thereafter, the maximum clade credibility (MCC) trees were summarized using TreeAnnotator software version 1.10.4 with a 10 per cent burn-in value for the sampled tree and finally visualized and edited using Figtree software version 1.4.3.

### 2.5 Phylogeographic inference on virus transmission

To obtain spatial–temporal transmission insights into lineages 1E-L2 and 2B-L2c, Bayesian stochastic search variable selection (BSSVS) in BEAST software version 1.10.4 was implemented by combining the asymmetric substitution model. To facilitate the analysis of virus spread pathways between the provinces of China, which included thirty-one provinces, these provinces were divided into five discrete geographic regions, namely, the central (seven...
provinces), northern (five provinces), southern (six provinces), eastern (six provinces), and western (seven provinces) regions. Different worldwide transmission pathways were also coded at both the country and regional levels. Moreover, we calculated the number of regions for in- and out-migration in China using the Markov jump stochastic mapping techniques implemented in BEAST software version 1.10.4. The Bayesian factors (BFs) and posterior probability (PP) values generated via BSSVS were then used to determine statistically significant outcomes in diffusion links; pairwise diffusion pathways were considered significantly supported when BF ≥ 3, and the PP value was above 0.5. The map of the world and that of China used to show the transmission pathways of RuV strains in this study were provided by Highcharts (grant number 0321912045738052).

2.6 Ethical statement
The study was approved by the Second Ethics Review Committee of the National Institute for Viral Disease Control and Prevention at the China Center for Disease Control and Prevention. All procedures were performed according to relevant guidelines and regulations. For the purpose of public health and disease control, informed consent for clinical sample collection from rubella cases was waived by the ethics committee.

3. Results
3.1 Lineage 1E-L2 and 2B-L2c strains detected in China within the 2018–21 period
Within the 2018–21 period, a total of 3,312 RuV strains were collected during rubella virological surveillance in China, and the results of genotyping and lineage division showed that 1,947 strains were lineage 1E-L2 strains from 25 provinces and 1,365 strains were lineage 2B-L2c strains from 21 provinces (Fig. 2). The temporal distributions of the strains corresponding to the two RuV lineages were consistent with the trend of rubella incidence in China, and a majority of strains (1,640 lineage 1E-L2 strains from 24 provinces and 1,234 lineage 2B-L2c strains from 20 provinces) corresponded to the nationwide rubella epidemic of 2019 (32,520 cases, 2.33/100,000). With the rubella incidence reaching its nadir in 2021 (940 cases, 0.07/100,000) (Fig. 3A), only 15 strains (lineage 1E-L2, 4 strains from two provinces; lineage 2B-L2c, 11 strains from four provinces) were obtained.

According to geographic regions, the RuV strains of both lineages collected during the 2018–21 period covered all the five regions of China. Among them, lineage 1E-L2, which was first detected in Guangxi province located in the southern region of China in January 2018, had a higher detection ratio in eastern (97.3 per cent) and southern China (97.8 per cent) within the 2018–21 period. Comparatively, lineage 2B-L2c, which was first identified in Sichuan province located in the western region of China in August 2018, had a relatively high detection ratio in western China (66.3 per cent) (Fig. 3B). Furthermore, both lineages were detectable in the central and northern regions, with lineage 1E-L2 showing slight dominance (approximately 60 per cent). Moreover, given that most of the strains of these two lineages were collected during the regional outbreaks within the 2018–21 period (Fig. 3C), the E1-739 sequences of the strains investigated in this study showed high similarity with 98.9–100 and 99.2–100 per cent for lineages 1E-L2 and 2B-L2c, respectively.

3.2 Phylogenetic analysis
The results of root-to-tip regression analysis supported the relationships between the temporal signals and root-to-tip divergence of lineage 1E-L2 strains for Dataset 3 (n = 184, correlation = 0.88, R² = 0.77) and lineage 2B-L2c for Dataset 4 (n = 206, correlation = 0.73, R² = 0.53) (Supplementary Fig. S1). Furthermore, Bayesian inference based on the E1-739 region was performed to illustrate the phylogenetic clustering and evolutionary characteristics of the two global lineages. The results corresponding to the MCC trees showed that both lineages were characterized by a high genetic diversity, whereas the detailed clustering situation varied between the two lineages.

The strains in the phylogenetic tree of lineage 1E-L2 showed noticeable geographic and temporal clustering. Furthermore, the topology of the phylogenetic tree showed that the transmission pattern of this lineage could be divided into three major stages (Fig. 4). Within the 2001–11 period, the initial stage predominantly consisted of strains from Taiwan of China. Several sporadic sequences from Malaysia (2001 and 2005), Kazakhstan (2006), Hong Kong of China (2009), and Indonesia (2011) were also observed. Subsequently, within the 2012–7 period, lineage 1E-L2 progressed to the second stage. Virus strains corresponding to this stage originated from Japan, Malaysia, and Hong Kong of China. Furthermore, the imported strains circulating in China within the 2018–21 period constituted an independent branch and clustered with the strains from Japan and Hong Kong of China, within the same rubella epidemic period; thus, these strains were considered to belong to the third stage of the transmission pattern of lineage 1E-L2 (2018–21). These findings indicated that the transmission of lineage 1E-L2 appeared to be limited to only some regions of Asia and presented a gradual evolutionary pattern over time.

Compared with lineage 1E-L2, lineage 2B-L2c has been most frequently detected around the world since 2005, and multiple lineage 2B-L2c viral transmission patterns have been observed over time. Our analysis showed that: (1) Different viral lineage 2B-L2c branches could be detected consecutively in the same country, e.g. in India. Specifically, the sequences detected in different regions of India between 2005 and 2017 were scattered across almost all the branches of the phylogenetic tree. (2) The same viral branch could also be detected in different countries or regions. For example, two branches of lineage 2B-L2c were detected successively in China. One of them was collected during a small-scale outbreak that occurred in Anhui province in 2012. Additionally, this branch clustered with the sequences from Tunisia corresponding to the 2011–12 period and was not detected again in China after the outbreak. The other branch, with sequences that have been circulating in China since 2018, clustered with the strains from Japan and Hong Kong of China within the same period. (3) Several sporadic sequences from certain countries, such as Japan and Hong Kong of China, were detected intermittently in different branches. (4) Individual sequences from other countries or regions, such as Afghanistan, Nigeria, Pakistan, and Yemen, were occasionally detected. Therefore, phylogenetic analysis indicated that different 2B-L2c viruses had been circulating worldwide and that extensive virus importation and transmission occurred between countries (Fig. 5).

3.3 Evolutionary characteristics and genetic diversity
Based on the constructed datasets, the evolutionary rates and the time to the most recent common ancestors (tMRCA) of lineages 1E-L2 and 2B-L2c were determined. The estimated evolutionary rates for these global lineages were 1.35 × 10⁻³ substitutions/site/year (95 per cent highest posterior
density (HPD): $1.05 \times 10^{-3}$ and $1.59 \times 10^{-3}$, respectively. Additionally, the evolutionary rates of 1E-L2 and 2B-L2c circulating in China within the 2018–21 period were $1.39 \times 10^{-3}$ (95 per cent HPD: $0.987 \times 10^{-3}$ and $1.34 \times 10^{-3}$ (95 per cent HPD: $0.77 \times 10^{-3}$), respectively. Furthermore, we speculated that the time of the common ancestor of lineages 1E-L2 and 2B-L2c dated back to 1997 (95 per cent HPD, 1993–9) and 2004 (95 per cent HPD, 2001–3).

Figure 2. Neighbor-joining tree constructed using the representative sequences of lineage 1E-L2 and 2B-L2c strains circulating in China within the 2018–21 period together with reference sequences for both genotypes and lineage divisions. Different lineages are indicated with different colors, and bootstrap values greater than 80 per cent were highlighted.
The genetic diversity of lineages 1E-L2 and 2B-L2c, which were identified based on PS and SS results, respectively, was analyzed using the Bayesian skyline and GMRF skyline models, respectively (Supplementary Table S3). As estimated, lineage 1E-L2 underwent three periods of genetic diversity conversion (the 2001–11, 2012–17, and 2018–21 periods). This was found to be consistent with the results of the phylogenetic analysis described in Section 3.2. Additionally, the genetic diversity observed within the 2001–11 period was minimal; in 2012, it showed a slight increase; after a rapid decline in 2017, a sharp upward trend was observed within the 2018–19 period; and this trend then remained stable until 2021 (Supplementary Fig. S2A). The average genetic distance of lineage 1E-L2 was 0.013 (Table 1); the distances within the three periods of this lineage were 0.013, 0.012, and 0.004, respectively, and the distance between the three periods was 0.016–0.023. However, lineage 2B-L2c showed a significantly different genetic diversity, showing a gradually increasing trend before 2015 with the peak observed within the 2015–17 period. Furthermore, the genetic diversity of lineage 2B-L2c also showed a sharp decline in 2017 and then gradually recovered after 2018 (Supplementary Fig. S2B). The average genetic distance of lineage 2B-L2c was 0.016 (Table 1). The increasing tendency of the relative genetic diversity values of lineages 1E-L2 and 2B-L2c between 2018 and 2021 might be related to the epidemics caused by both imported lineages in China within this period.

3.4 Inference of the phylogeographic transmission pattern of lineage 1E-L2

To analyze the viral origin and transmission pathways of lineages 1E-L2 and 2B-L2c globally and domestically, the possible spatial transmission patterns were inferred based on discrete geographic information, and the transmission pathways strongly supported by BFs (BF ≥ 3) and PPs (PP > 0.5) are shown. Specifically, to estimate the global transmission trajectory of lineage 1E-L2, eight substantial routes, which showed limited circulation between adjacent regions in Asia, were identified (Fig. 6A). Among them, two countries or regions, including Japan and
Taiwan of China, were identified as the main viral transmission areas for the dissemination of lineage 1E-L2, and four transmission routes were found to have originated from Japan, while three originated from Taiwan of China. Notably, significant supportive evidence was observed for transmission routes from China to Hong Kong of China ($BF = 47,475$), followed by the route from Taiwan of China to Kazakhstan ($BF = 198$). Furthermore, Japan ($BF = 26$) played a major role in the importation of lineage 1E-L2.
Figure 5. MCC tree with temporal phylogenies and evolutionary characteristics for lineage 2B-L2c based on the E1-739 region. The top nine countries and regions (China, Congo, Hong Kong of China, India, Japan, Romania, the UK, Tunisia, and Uganda) with the largest number of sequences are represented by different colors, and then black is used to represent other nineteen countries or regions (Afghanistan, Belarus, Canada, Chile, Côte d’Ivoire, Ethiopia, Germany, Iran, Israel, Italy, Mexico, New Zealand, Nigeria, Oman, Pakistan, Russia, Spain, Thailand, and Yemen). The annotation nodes with a PP greater than 0.75 in the phylogenetic tree are represented by diamonds. For clear display, the tMRCAs of the main clusters were labeled.

The results showed that viral transmission between closer regions was characterized by higher migration rates. In terms of domestic transmission pathways, southern China played a decisive role in the spread of lineage 1E-L2 to the western and eastern regions (BF > 100), while central China served
Table 1. Evolutionary analysis of lineages 1E-L2 and 2B-L2c.

| Dataset | Lineage | Number of strains | Genetic distance | Substitution rate ($\times 10^{-3}$) (95 per cent HPD) | tMRCA (95 per cent HPD) |
|---------|---------|------------------|-----------------|--------------------------------|----------------------|
| China   | 1E-L2   | 315              | 0.002           | 1.39 (0.87–1.80)               | 2016 (2014–8)        |
|         | 2B-L2c  | 292              | 0.002           | 1.34 (0.77–1.83)               | 2016 (2014–8)        |
| Global  | 1E-L2   | 184              | 0.013           | 1.35 (1.05–1.68)               | 1997 (1993–9)        |
|         | 2B-L2c  | 206              | 0.016           | 1.59 (1.27–1.91)               | 2004 (2003–4)        |

Figure 6. Spatial transmission routes of lineages 1E-L2 and 2B-L2c based on Bayesian phylogeographic inference. Different dispersion paths are represented by the curves with arrow marks at the end that connect different regions (statistically supported by BF $\geq 3$). The color of the line indicates the size of the BF, while the thickness of the line represents the distinct migration rate. (A) Transmission routes of lineage 1E-L2 strains circulating worldwide. (B) Transmission routes of lineage 2B-L2c strains circulating worldwide. (C) Spatial transmission routes of lineage 1E-L2 and 2B-L2c strains circulating in the five regions of China. (D) Transmission routes of lineage 1E-L2 strains circulating between provinces in China. (E) Transmission routes of lineage 2B-L2c strains circulating between provinces in China.

As a transit region for the further transmission of the virus to the northern region (Fig. 6C). Therefore, the transmission pathway of lineage 1E-L2 showed a south-to-north transmission trend. Additionally, outward migration assessed using the Markov jump method showed that virus transmission was predominant in southern China, while inward migration corresponding to the central, western, and eastern regions showed relatively high numbers, consistent with the results of the transmission pathway analysis (Supplementary Fig. S3A).

Additionally, the spread of lineage 1E-L2 between provinces in China presented an obvious three-stage transmission pattern (Fig. 6D). As estimated, lineage 1E-L2 strains were first detected in Guangxi province in southern China in January 2018. Thereafter, they spread to two nearby provinces, Jiangxi (BF = 2,183) and Guangdong (BF = 1,447). The second stage of virus transmission initiated in Jiangxi province in eastern China, which served as a secondary viral transmission region, spreading the virus to six other major provinces: Beijing (BF = 1,220), Sichuan (BF = 508), Chongqing (BF = 407), Tianjin (BF = 319), Hunan (BF = 289), and Zhejiang (BF = 123). In the third stage, the virus continued to spread to the surrounding and peripheral provinces. For example, the virus spread from Shandong to Henan (BF = 37) and Neimeng (BF = 30) provinces, and from Chongqing to Hainan (BF = 76) and Yunnan (BF = 64) provinces. This three-stage transmission pattern led to the establishment of an endogenous circulation of lineage 1E-L2 in China after its introduction. Lineage 1E-L2 strains were first detected in January 2018 in Guangxi province; therefore, these surveillance data were consistent with transmission pathway analysis results.

3.5 Inference of the phylogeographic transmission pattern of lineage 2B-L2c

Inference of transmission patterns confirmed the existence of a broad link between the lineage 2B-L2c strains circulating worldwide, and sixteen transmission pathways were identified (Fig. 6B). Based on transmission pattern analysis, we estimated that the virus was transmitted from India, which might serve as the main source of transmission activities, to several other countries, including Japan (BF = 236,692), New Zealand (BF = 450), Chile (BF = 67), the UK (BF = 55), and Congo (BF = 28). Furthermore,
Congo and the UK served as the main areas for the transmission of lineage 2B-L2c strains to other countries in their respective continents. The lineage 2B-L2c strain identified in the Anhui province of China in 2012 was estimated to be imported from Tunisia (BF = 77), which experienced a nationwide rubella outbreak in 2011–2 (Messedi et al. 2014). However, for the lineage 2B-L2c strains circulating in China within the 2018–21 period, the inference suggested that the UK (BF = 8), Hong Kong of China (BF = 5), and India (BF = 4) were the possible sources, while these BF values were not supported by the PP value (<0.5). Therefore, the source of lineage 2B-L2c imported into China in 2018 remains unclear. However, it should be noted that the inferences of the global transmission trajectories of both lineages were generated based on the sequences collected from the GenBank database, which contained limited virological surveillance data for many parts of the world; therefore, the global distribution and transmission patterns of rubella genotypes could not be fully described.

Regarding domestic spread, the spread of lineage 2B-L2c strains was predominant in western China, and from here, the virus was transmitted to the other four regions of China, among which transmission routes to the southern and eastern regions were strongly supported by the BF values (>100) (Fig. 6C). The outward migration of lineage 2B-L2c strains from the western region was overwhelmingly predominant, while the state counts of inward migration in other regions were higher (Supplementary Fig. S3B); this further confirmed the results of domestic pathway analysis. Regarding the spread of lineage 2B-L2c between provinces, Sichuan was primarily responsible for the spread of the virus to nine other provinces in China (with significant BF value support), including Shanghai (BF = 387), Gansu (BF = 379), Chongqing (BF = 105), Ningxia (BF = 97), Guizhou (BF = 71), Shandong (BF = 45), Beijing (BF = 42), Guangxi (BF = 26), and Tianjin (BF = 19) (Fig. 6F). Lineage 2B-L2c strains were first detected in August 2018 in Sichuan province, earlier than in the other provinces; these surveillance data were consistent with transmission pathway analysis results.

4. Discussion

Previous studies on the viral surveillance of RuVs circulating in China indicated the occurrence of three viral genotype or lineage switches over the past few decades, including a shift from lineage 1F to lineage 1E-L1 within the 2001–2 period (Zhu et al. 2012), lineage 1E-L1 to lineage 2B-L1 within the 2015–6 period (Zhu et al. 2016), and lineage 2B-L1 to imported lineages 1E-L2 and 2B-L2 within the 2018–9 period (Zhu et al. 2021), and presently, lineage 1E-L2 and 2B-L2c strains are co-circulating in China. Furthermore, in recent years, these two lineages have been most frequently detected worldwide. Therefore, in this study, we further analyzed their transmission patterns as well as their evolutionary characteristics, aiming to provide important scientific data for rubella control and elimination in China and worldwide.

Time-scale phylogenetic analysis in this study revealed that the global lineages 1E-L1 and 2B-L2c presented distinct evolutionary patterns, indicating that these contemporaneous viruses evolved independently. In this study, the tMRCA of lineage 1E-L2 was estimated at circa 1997, and the first strain was detected in 2001 (strain RV1/MYS01, accession number AY966221), suggesting that circulating lineage 1E-L2 had not been detected for at least 4 years, possibly owing to the existence of a surveillance gap. After emergence, lineage 1E-L2 gradually evolved and circulated in relatively limited geographical areas (mainly Asia). Contrary to lineage 1E-L2 strains, lineage 2B-L2c strains circulated widely throughout the world and exhibited a complicated topology with several independently evolved branches, given that it appeared around 2004 and 2005.

Although rubella elimination has been verified in approximately half of the countries worldwide according to WHO data (Zimmerman et al. 2022), rubella continues to circulate endemically in several countries, and in this study, extensive international transmission activities were observed for both lineages 1E-L2 and 2B-L2c. Therefore, the following issues should be considered: (1) Frequent virus exchange might pose potential threats to countries that had verified or were nearing rubella elimination. For example, rubella was eliminated from the USA in 2004; however, cases and outbreaks persisted because of the importation of strains from countries where it remained endemic (Ali Hammond, Murphy, and Pérez 2018; Robyn et al. 2018). In this regard, joint efforts from all countries are needed to realize the global rubella elimination requirement. (2) International movement of susceptible persons can play an important role in the spread of the virus; thus, effective public health measures, such as providing the opportunity for residents to get rubella vaccination before international travel, are needed. (3) It is noteworthy that rubella importation could only be successful if vaccination coverage was insufficient and the number of susceptible individuals increased. For example, rubella emergence and outbreaks in China within the 2018–9 period coincided with the introduction of imported lineages, 1E-L2 and 2B-L2c, and an increase in the number of susceptible individuals among adolescents and young adults (Zhu et al. 2021). Accordingly, on the basis of sustaining high routine immunization coverage among infants, wide-age range catch-up campaigns are necessary to further close susceptibility gaps.

Our analysis confirmed that Japan and Taiwan of China served as the main viral transmission regions for the spread of lineage 1E-L2, which was unsurprising considering the high rates of immigration and international travel in these countries and regions. Additionally, we found evidence for the introduction of lineage 1E-L2 with an unknown origin that transmitted from Japan to China and was further exported to Hong Kong of China within the 2018–9 period, suggesting the occurrence of a synchronized outbreak across multiple countries and regions during this period. Unfortunately, the origin of lineage 2B-L2c strains circulating in China since 2018 is still unclear, even though India was confirmed as the major source. Given that 98 per cent of the sequences in the global RubeNS database originated from China and Japan (Brown et al. 2019), there are limited virological surveillance data for many parts of the world. Thus, the global distribution and transmission patterns of rubella genotypes could not be fully described; this made it difficult to track the source of the imported viruses. Therefore, global virologic surveillance should be improved, particularly in countries with little or no virologic surveillance. Furthermore, the timely sharing of surveillance data worldwide is also essential.

The strains of lineages 1E-L2 and 2B-L2c, which were successively imported into China in 2018, showed different transmission patterns. Among them, the spread of lineage 1E-L2 strains showed a three-stage transmission pattern from southern to northern China, whereas lineage 2B-L2c spread from a single point in western China to all the other four regions of China. These two transmission patterns allowed the imported strains to spread rapidly across China during the 2018–9 rubella epidemic, and eventually, endemic circulations were established. Interestingly, the transmission patterns adopted by the above two lineages were also different from those of the previously observed lineage 2B-L1. After multiple introductions in 2011, lineage 2B-L1
gradually spread from the eastern coastal provinces to the western provinces of China, and instead of lineage 1E-L1, it became the predominant strain within the 2015–7 period (Zhu et al. 2015). Additionally, it is worth noting that Jiangxi and Sichuan provinces played major roles in the spread of lineage 1E-L2 and 2B-L2c strains, respectively. These observations could be related to the movement of the large number of infected individuals from these two provinces. Therefore, timely investigation and early response to importation in such provinces are necessary. Furthermore, province-based serosurvey studies are needed to ascertain population immunity, monitor susceptibility, and identify age groups for supplementary immunization activities (SIAs). Notably, most of the links to importation or transmission patterns in this study were based on the analysis of sequences without supporting epidemiological evidence due to insufficient epidemiological data. To achieve the rubella elimination goal, molecular surveillance data correlated with the detailed epidemiological investigation information are extremely essential to identify the potential sources of virus importation and recognize long-lasting virus transmission chains.

Based on the results of this study, it was estimated that the tMRCA of both imported lineages 1E-L2 and 2B-L2c surfaced around 2015–6 and were introduced into China in 2018. Furthermore, the strains of both lineages were detected immediately after importation, suggesting high sensitivity for the case-based rubella surveillance system in China. However, rubella is typically milder than measles, and 20–50 per cent of infections are asymptomatic (Camejo Leonor and Mendez 2022); this could lead to a low percentage of persons with rubella infection seeking health care. In this case, imported viruses might not be detected in a timely manner. Therefore, other surveillance systems, such as CRS surveillance, which can provide additional opportunities for the identification of the viruses and evidence-based support for rubella elimination, should be introduced.

The estimated evolutionary rates of both lineages 1E-L2 and 2B-L2c globally and domestically were within the range 1.34–1.59 × 10^{-3} substitutions/site/year, which is comparable to the rates that were previously reported for lineages 1E-L1 and 2B-L1, further confirming that the overall substitution rate of RuV is in the order of 10^{-3} substitutions/site/year.

Compared with measles, the elimination of rubella and CRS is considered more technically feasible owing to a lower reproductive number (R0), a more effective vaccine, the good persistence of rubella antibodies, and an apparently lower rate of breakthrough infection (Plotkin 2021). Nonetheless, repeated importation poses great challenges to rubella elimination in China. Therefore, to accelerate success in this regard, the following potential actions and efforts are necessary: developing strategies for SIA for the target population to further close the susceptibility gap, promoting case notification and investigation, strengthening outbreak response capability, maintaining comprehensive surveillance, supported by high-quality laboratory networks, and setting up a CRS surveillance platform. Furthermore, improving the status of global molecular epidemiological surveillance and the timely sharing of surveillance data worldwide is also essential.

**Data availability**

The sequences of 3,312 RuV strains collected during 2018–21 were submitted to global RubeNS database established by the WHO (https://who-gmrln.org/rubens2). Additionally among them, 542 representative sequences used to construct the datasets in this study were submitted to the GenBank database under accession numbers OM899064–OM899605.

**Supplementary data**

Supplementary data are available at Virus Evolution online.

**Acknowledgements**

We gratefully acknowledge all the staff from the Chinese National Measles and Rubella Laboratory Network for clinical sample collection and virus isolation, as well as the epidemiologists for the collection and integration of epidemiological data.

**Funding**

This work was supported by the Key Technologies R&D Program of the National Ministry of Science (grant number 2018ZX10713002).

**Conflict of interest:** None declared.

**References**

Al Hammoud, R., Murphy, J. R., and Pérez, N. (2018) ‘Imported Congenital Rubella Syndrome, United States, 2017’, *Emerging Infectious Diseases*, 24: 800–1.

Baele, G. et al. (2012) ‘Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty’, *Molecular Biology and Evolution*, 29: 2157–67.

Bennett, A. J. et al. (2020) ‘Relatives of Rubella Virus in Diverse Mammals’, *Nature*, 586: 424–8.

Brown, K. E. et al. (2019) ‘Genetic Characterization of Measles and Rubella Viruses Detected through Global Measles and Rubella Elimination Surveillance, 2016-2018’, *Morbidity and Mortality Weekly Report*, 68: 587–91.

Camejo Leonor, M., and Mendez, M. D. (2022) ‘Rubella’, StatPearls. Treasure Island, FL: StatPearls Publishing <http://www.ncbi.nlm.nih.gov/books/NBK559040/> accessed 8 Mar 2022.

Drummond, A. J. et al. (2012) ‘Bayesian Phylogenetics with BEAUti and the BEAST 1.7’, *Molecular Biology and Evolution*, 29: 1969–73.

DuBois, R. M. et al. (2013) ‘Functional and Evolutionary Insight from the Crystal Structure of Rubella Virus Protein E1’, *Nature*, 493: 552–6.

Frey, T. K. (1994) ‘Molecular Biology of Rubella Virus’, *Advances in Virus Research*, 44: 69–160.

Hoang, D. T. et al. (2018) ‘UFBoot2: Improving the Ultrafast Bootstrap Approximation’, *Molecular Biology and Evolution*, 35: 518–22.

Kalyaanamoorthy, S. et al. (2017) ‘ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates’, *Nature Methods*, 14: 587–9.

Katoh, K., and Standley, D. M. (2013) ‘MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability’, *Molecular Biology and Evolution*, 30: 772–80.

Kaushik, A., Verma, S., and Kumar, P. (2018) ‘Congenital Rubella Syndrome: A Brief Review of Public Health Perspectives’, *Indian Journal of Public Health*, 62: 52–4.

Kumar, S., Stecher, G., and Tamura, K. (2016) ‘MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets’, *Molecular Biology and Evolution*, 33: 1870–4.

Lambert, N. et al. (2015) ‘Rubella’, *The Lancet*, 385: 2297–307.

Mangala Prasad, V., Klose, T., and Rossmann, M. G. (2017) ‘Assem-
Messedi, E. et al. (2014) 'Clinical Epidemiological and Molecular Aspects of Rubella Outbreak with High Number of Neurological Cases, Tunisia 2011-2012', *Journal of Clinical Virology*, 61: 248–54.

Mori, Y. et al. (2017) 'Molecular Epidemiology of Rubella Virus Strains Detected around the Time of the 2012-2013 Epidemic in Japan', *Frontiers in Microbiology*, 8: 1513.

Nguyen, L. T. et al. (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.

Plotkin, S. A. (2021) 'Rubella Eradication: Not Yet Accomplished, but Entirely Feasible', *The Journal of Infectious Diseases*, 224: S360–6.

Rivailler, P., Abernathy, E., and Icenogle, J. (2017) 'Genetic Diversity of Currently Circulating Rubella Viruses: A Need to Define More Precise Viral Groups', *Journal of General Virology*, 98: 396–404.

Robyn, M. et al. (2018) 'Two Imported Cases of Congenital Rubella Syndrome and Infection-Control Challenges in New York State, 2013-2015', *Journal of the Pediatric Infectious Diseases Society*, 7: 172–4.

Su, Q. et al. (2018) 'Epidemiological Profile and Progress Toward Rubella Elimination in China. 10 Years after Nationwide Introduction of Rubella Vaccine', *Vaccine*, 36: 2079–85.

Ujiie, M. (2019) 'Rubella Resurgence in Japan 2018–2019', *Journal of Travel Medicine*, 26: taz047.

Vynnycky, E. et al. (2016) 'Using Seroprevalence and Immunisation Coverage Data to Estimate the Global Burden of Congenital Rubella Syndrome, 1996-2010: A Systematic Review', *PLoS One*, 11: e0149160.

WHO. (2005) 'Standardization of the Nomenclature for Genetic Characteristics of Wild-Type Rubella Viruses', *Relevé Epidemiologique Hebdomadaire*, 80: 126–32.

———. (2013) 'Rubella Virus Nomenclature Update: 2013', *Relevé Epidemiologique Hebdomadaire*, 88: 337–43.

Xu, W. et al. (2017) 'Global and National Laboratory Networks Support High Quality Surveillance for Measles and Rubella', *International Health*, 9: 184–9.

Zhu, Z. et al. (2016) 'Analysis of Complete Genomes of the Rubella Virus Genotypes 1E and 2B Which Circulated in China, 2000–2013', *Scientific Reports*, 6: 39025.

——— et al. (2012) 'Emergence and Continuous Evolution of Genotype 1E Rubella Viruses in China', *Journal of Clinical Microbiology*, 50: 353–63.

——— et al. (2021) 'Transmission Dynamics of the Rubella Virus Circulating in China During 2010-2019: Two Lineage Switches Between Genotypes 1E and 2B', *Clinical Infectious Diseases*, 73: 1157–1164.

——— et al. (2015) 'Evolutionary Analysis of Rubella Viruses in Mainland China During 2010-2012: Endemic Circulation of Genotype 1E and Introductions of Genotype 2B', *Scientific Reports*, 5: 1–11.

——— et al. (2007) 'Comparison of Four Methods Using Throat Swabs to Confirm Rubella Virus Infection', *Journal of Clinical Microbiology*, 45: 2847–52.

Zimmerman, L. A. et al. (2022) 'Progress Toward Rubella and Congenital Rubella Syndrome Control and Elimination—Worldwide, 2012–2020', *Morbidity and Mortality Weekly Report*, 71: 196–201. <https://www.cdc.gov/mmwr/volumes/71/wr/mm7106a2.htm> accessed 8 Mar 2022.

——— et al. (2022) 'Progress toward Rubella and Congenital Rubella Syndrome Control and Elimination—Worldwide, 2012-2020', *Morbidity and Mortality Weekly Report*, 71: 196–201.