SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
VIRAL INFECTION
PHYLOGENETICS

Received
3 July 2014

Accepted
3 December 2014

Published
23 January 2015

Evolutionary analysis of rubella viruses in mainland China during 2010–2012: endemic circulation of genotype 1E and introductions of genotype 2B

Zhen Zhu[1]*, Pierre Rivailler[2]*, Emily Abernathy[2], Aili Cui[1], Yan Zhang[1], Naiyin Mao[1], Songtao Xu[1], Shujie Zhou[3], Yue Lei[4], Yan Wang[5], Huanying Zheng[6], Jilan He[7], Ying Chen[8], Chongshan Li[9], Fang Bo[10], Chunfang Zhao[11], Meng Chen[12], Peishan Lu[13], Fangcai Li[14], Suyi Gu[15], Hui Gao[16], Yu Guo[17], Hui Chen[18], Daxing Feng[19], Shuang Wang[20], Xiaomin Tang[21], Yake Lei[22], Yan Feng[23], Lili Deng[24], Tian Gong[25], Lixia Fan[26], Wenbo Xu[1], Joseph Icenogle[2] & Rubella Virology Surveillance Working Group#

Correspondence and requests for materials should be addressed to W.B.X. (wenbo_xu1@aliyun.com) or J.I. (jci1@cdc.gov)

* These authors contributed equally to this work.

A comprehensive list of authors and affiliations appear at the end of the paper.

[1]WHO WPRO Regional Reference Measles/Rubella Laboratory and Key Laboratory of Medical Virology Ministry of Health, National Institute for Viral Disease Control and Prevention, Beijing, People's Republic of China, [2]Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, [3]Anhui Provincial Centers for Disease Control and Prevention, People's Republic of China, [4]Tianjin Provincial Centers for Disease Control and Prevention, People's Republic of China, [5]Liaoning Provincial Centers for Disease Control and Prevention, People's Republic of China, [6]Guangdong Provincial Centers for Disease Control and Prevention, People's Republic of China, [7]Sichuan Provincial Centers for Disease Control and Prevention, People's Republic of China, [8]Gansu Provincial Centers for Disease Control and Prevention, People's Republic of China, [9]Shanghai Provincial Centers for Disease Control and Prevention, People's Republic of China, [10]Heilongjiang Provincial Centers for Disease Control and Prevention, People's Republic of China, [11]Chongqing Provincial Centers for Disease Control and Prevention, People's Republic of China, [12]Beijing Provincial Centers for Disease Control and Prevention, People's Republic of China, [13]Jiangsu Provincial Centers for Disease Control and Prevention, People's Republic of China, [14]Hunan Provincial Centers for Disease Control and Prevention, People's Republic of China, [15]Neimeng Provincial Centers for Disease Control and Prevention, People's Republic of China, [16]Shanxi Provincial Centers for Disease Control and Prevention, People's Republic of China, [17]Hebei Provincial Centers for Disease Control and Prevention, People's Republic of China, [18]Ningxia Provincial Centers for Disease Control and Prevention, People's Republic of China, [19]Henan Provincial Centers for Disease Control and Prevention, People's Republic of China, [20]Jilin Provincial Centers for Disease Control and Prevention, People's Republic of China, [21]Guizhou Provincial Centers for Disease Control and Prevention, People's Republic of China, [22]Hubei Provincial Centers for Disease Control and Prevention, People's Republic of China, [23]Zhejiang Provincial Centers for Disease Control and Prevention, People's Republic of China, [24]Guangxi Provincial Centers for Disease Control and Prevention, People's Republic of China, [25]Jiangxi Provincial Centers for Disease Control and Prevention, People's Republic of China, [26]Qinghai Provincial Centers for Disease Control and Prevention, People's Republic of China.

Rubella remains a significant burden in mainland China. In this report, 667 viruses collected in 24 of 31 provinces of mainland China during 2010–2012 were sequenced and analyzed, significantly extending previous reports on limited numbers of viruses collected before 2010. Only viruses of genotypes 1E and 2B were found. Genotype 1E viruses were found in all 24 provinces. Genotype 1E viruses were likely introduced into mainland China around 1997 and endemic transmission of primarily one lineage became established. Viruses reported here from 2010–2012 are largely in a single cluster within this lineage. Genotype 2B viruses were rarely detected in China prior to 2010. This report documents a previously undetected 2B lineage, which likely became endemic in eastern provinces of China between 2010 and 2012. Bayesian analyses were performed to estimate the evolutionary rates and dates of appearance of the genotype 1E and 2B viral linages in China. A skyline plot of viral population diversity did not provide evidence of reduction of diversity as a result of vaccination, but should be useful as a baseline for such reductions as vaccination programs for rubella become widespread in mainland China.

Rubella virus (RV) infection in children and adults usually presents as a mild self-limited illness. However, infection during pregnancy, especially in the first trimester, often leads to serious consequences such as fetal death and various birth defects collectively known as congenital rubella syndrome (CRS)[1]. In China, national rubella and CRS surveillance has not been established; however, since 2009 several smaller scale surveillance projects have been initiated by the Ministry of Health (MOH) and World Health Organization (WHO)[2,3].

Rubella vaccination has proved to effectively prevent and control rubella and CRS[4]. The live attenuated RA27/3 strain (WHO name: RVi/Pennsylvania.USA/0.64/VAC), developed from a genotype 1a virus, is currently the most widely used vaccine worldwide[4]. In China, in addition to RA27/3, a domestically developed rubella vaccine known as BRD-II (WHO name: RVi/Beijing.CHN/0.80/VAC), derived from a 1979 genotype 2A virus, has also been used in vaccination campaigns. The BRD-II vaccine was shown to be highly effective against rubella[2]. Before 2008, rubella vaccines were available in the private sector, but the BRD-II vaccine was included in the national immunization program in 2008[5]. The national immunization strategy involved two doses of vaccine including the first dose (MR) for infants at 8 months of age and the second dose (MMR) for 18–24 months of age.

Rubella virological surveillance is important to identify circulating viruses, track importation of new viruses, and monitor disappearance of specific wild-type RV lineage[6–8]. Vaccine introduction may change transmission dynamics of endemic viruses; it is therefore important to monitor changes in the epidemic pattern of rubella viruses (since 2008 in this case). Rubella virological surveillance has greatly improved in mainland China since 1999, with targeted virological surveillance sites originally in 3 provinces in 1999 and extended to 27 provinces by 2012[5,9].

RV belongs to the *Rubivirus* genus within the *Togaviridae* family and is an enveloped virus with a positive-sense, single-stranded RNA genome. The 9762-nucleotides genome contains two open reading frames (ORFs) and encodes three structural proteins (SP) (C, E1 and E2) and two non-structural proteins (NSP) (P150 and P90)[1,10]. The envelope glycoprotein E1, with 481 amino acids, is a class 1 transmembrane protein with some important functional domains including antigenic sites, hemagglutination inhibition activity and a neutralization epitope[1]. According to WHO recommendation, a 739 nucleotide window within the E1 coding region is required for genotype identification[7].There are currently 12 recognized RV genotypes (1B, 1C, 1D, 1E, 1F, 1G, 1H, 1I, 1J, 2A, 2B, and 2C) and 1 provisional genotype (1a)[8]. Four genotypes, 1E, 1G, 1J and 2B, have recently been reported most frequently and two genotypes, 1E and 2B, have widespread geographic distribution[8].

In China, rubella virological surveillance was initiated in 1999; from 1999 to 2009, four RV genotypes (1E, 1F, 2A, and 2B) were identified. Genotype 1F was last reported in 2002 and is considered inactive and likely extinct[5,9,11]. Genotype 2A viruses were found in vaccine-related cases in 1999[9]. Genotype 1E has been identified as the predominant vial genotype, circulating continuously throughout China since it was first identified in 2001[5,9]. In contrast, genotype 2B viruses have been sporadically detected in 2000 and 2008[8].

In this study, we expand current knowledge by analyzing sequences from a large number of RVs collected during 2010–2012 to precisely understand current circulation patterns and examine evolutionary changes. The entire dataset used in this study was representative of the viral population across mainland China as data from 27 of 31 provinces representing all 6 Chinese administrative regions (Dongbei, Huadong, Zhongnan, Xinan, Xibei, and Huabei). We report a predominant cluster (Cluster A) within the previously reported lineage of genotype 1E viruses found in China as well as a new cluster (Cluster Cα) within genotype 2B viruses; we also report the evolutionary rates for viruses of both genotypes, estimate the dates of appearance of genotype 1E and 2B viral linages in China, and analyses viral population diversity.

Methods

Ethics statement. This study did not involve human participants or human experimentation; the only human materials used were throat swab and urine specimens collected from the clinically suspected rubella patients for the purpose of public health and disease control. This study was approved by the second session of the Ethics Review Committee of the National Institute for Viral Disease Control and Prevention in the Chinese Center for Disease Control (CDC) and the methods were carried out in accordance with the approved guidelines. Written informed consent for the use of the clinical specimens was obtained from all patients involved in this study.

Rubella incidence data. Rubella is classified as a category C notifiable infectious disease by the Ministry of Health of China. The number of rubella cases and the annual incidence rates were taken directly from the National Notifiable Disease Reporting System (NNDRS), which covers all hospitals in mainland China and is the official source of data since 2004[5]. Briefly, the data were collected monthly in the county CDCs and then submitted through the prefectural and provincial CDC and finally reached NNDRS in the China CDC. The map was generated using Mapinfo Professional software (Version 11.0).

Specimen collection and virus isolation. Throat swab or urine specimens were collected from persons with febrile maculopapular rash in 24 of 31 provinces of mainland China from 2010 to 2012. Clinical samples were inoculated onto Vero cells expressing the signaling lymphocytic activation molecule (Vero/SLAM) as previously described[12]. Although, the SLAM receptor is not necessary for RVs, Vero/SLAMs cells are used due to a combined measles/rubella isolation protocol. The culture supernatant was harvested and used to inoculate fresh Vero/SLAM cells for up to 2 additional passages.

Virus identification. Viral RNA was extracted from cultured viruses using a QIAamp Viral RNA Extraction Mini Kit (Qiagen, Valencia, CA) according to the manufacturer instructions. Viral RNA was detected by real-time reverse transcription-polymerase chain reaction (real-time RT-PCR) as previously described[3,13]. Briefly, a 185-bp amplicon was generated and detect by a TaqMan probe labeled with FAM and Blackhole Quencher. The amplification was carried out using an ABI 7500 real-time system (Life Technologies, NY) and the results analyzed with the instrument SDS software (version 2.0.1).

RT-PCR amplification and sequence determination of WHO standard genotyping window and complete E1 gene. For the WHO standard genotyping window amplification, RT-PCR was performed using a One Step RT-PCR kit (TaKaRa Biotechnology Dalian, China) to amplify a 945-nt (nt 8633-9577) product containing the 739-nt genotyping window as described previously[3].

To obtain the complete E1 gene sequence, RT-PCR was performed with the Qiagen One-step RT-PCR kit following the manufacturer instructions. The primer pair (forward primer: 5′-CTA TGG CGA GGA GGC TTT CAC-3′; reverse primer: 5′-CTA TAC AGC AAC AGG TGC-3′) was designed to obtain a 1512-nt product which included the 1443-nts of the E1 coding region.

After purification of the amplification product with a QIA Gel Extraction Kit (Qiagen, Valencia, CA), the sequences were determined bi-directionally using the Sanger dideoxy terminator sequencing method with a BigDye Terminator Version 3.1 Cycle Sequencing kit (Life Technologies,NY, USA) and ABI PRISM™ 3100 Genetic Analyzer (Life Technologies, Japan). Sequencher software version 5.0 (Gene Codes Corporation) was used to edit and assemble the raw sequence data in order to obtain the 739-nt and complete E1 gene sequences, respectively. Six hundred and sixty-seven 739-nt sequences and 72 full-length E1 sequences were generated. Representative sequences were submitted to GenBank (accession numbers are listed in Supplementary Table 1).

Datasets. A total of 6 datasets were generated, 3 datasets for each genotype, 1E (datasets 1 to 3) and 2B (datasets 4 to 6). Datasets 1, 2, 4 and 5 concern 739-nt of E1 gene whereas datasets 3 and 6 concern the full length sequence of E1 gene. For dataset 1, 802 739-nt genotype 1E sequences were assembled by combining 543 newly generated sequences, 127 previously reported sequences from viruses collected in mainland China and 132 sequences of viruses collected outside China. Identical sequences were discarded leaving 486 unique sequences; sequences of viruses collected at the earliest time were included in the dataset. In order to avoid oversampling, based on the chronological and geographical distribution of the sequences, similar sequences (less than 1% nucleotide divergence) from viruses collected at the same year and same province were discarded leaving 158 sequences. Dataset 2 (174 sequences) is dataset 1 plus 16 non-1E sequences of WHO reference viruses for BEAST analysis. Dataset 3 (75 sequences) consist of 54 newly sequenced E1 genes of genotype 1E viruses, combined with 1 previously reported Chinese sequence, 5 non-Chinese 1E sequences and 15 E1 sequences of WHO reference viruses. For dataset 4, 124 newly generated 739-nt genotype 2B sequences were combined with 4 previously reported sequences from viruses collected in China and 132 sequences of viruses collected outside China generating a dataset of 260 sequences. Identical sequences were discarded leaving 221 unique sequences. Similar sequences from viruses collected at the same year and same province were deleted, as described above, further decreasing the dataset to 101

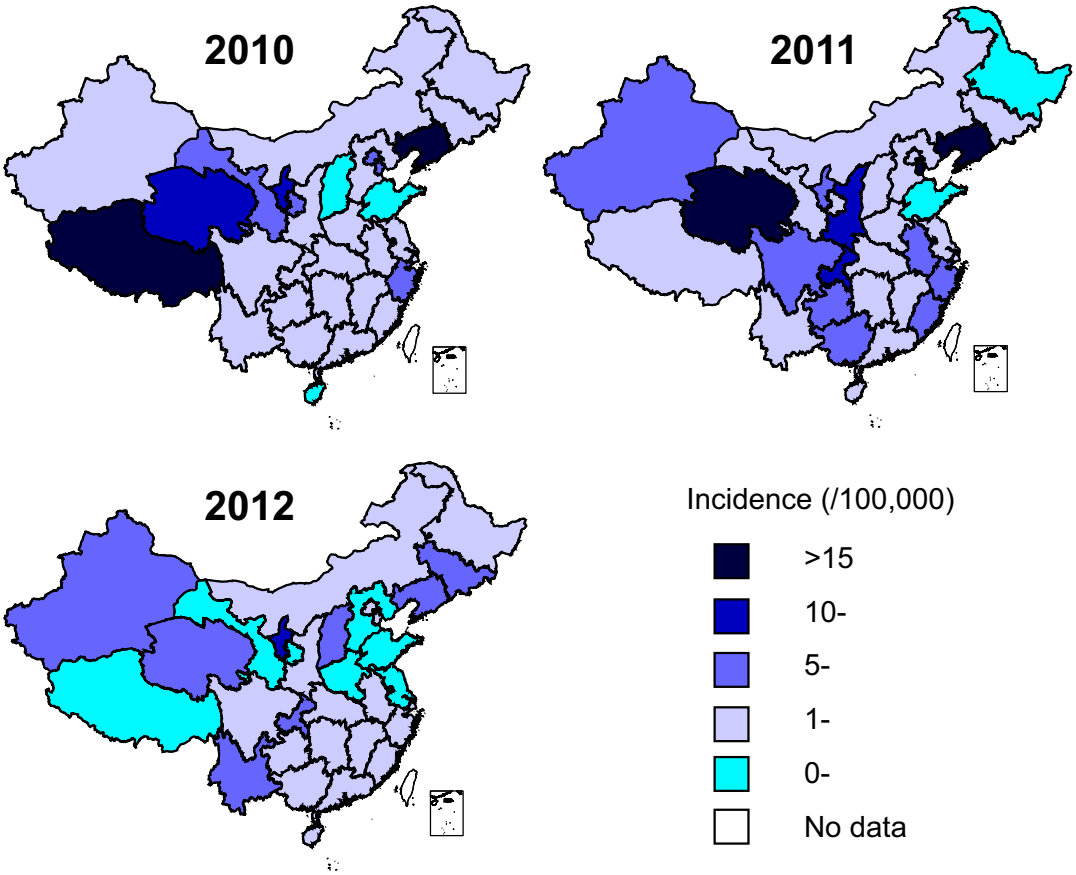


Figure 1 | Average rubella incidence by province of mainland China during 2010-2012. Average incidence is color coded according to the legend. The data was taken directly from NNDRS. The map was generated using Mapinfo Professional software (Version 11.0).

sequences (Dataset 4). Dataset 5 (116 sequences) is dataset 4 plus 15 739-nt sequences of WHO reference viruses. Dataset 6 (39 sequences) consist of 18 newly generated E1 sequences of genotype 2B combined with one previously reported sequence from China, 6 non-Chinese 2B sequences and 14 E1 sequences of WHO reference viruses. All datasets are listed in Supplementary Table 1.

Phylogenetic analysis. Sequence alignments and phylogenetic trees were performed using the MEGA program (version 5.03)[14]. Phylogenies were inferred using the neighbor-joining method and a maximum composite likelihood nucleotide model[15–17]. The reliability of phylogenetic inference at each branch node was estimated by the bootstrap method with 1000 replicates, using the MEGA program. Genotypes were determined according to WHO recommendations[8]. Distances between and within clusters were calculated using the MEGA program.

Estimation of the Time of the Most Recent Common Ancestor and evolutionary rate. Times of the Most Recent Common Ancestor (TMRCA) and evolutionary rates were estimated using a coalescent-based Bayesian method implemented in BEAST (Version 1.7.4)[18]. jModelTest (Version 0.1)[19] was used to assess the nucleotide substitution model for each dataset (Dataset 2, 3, 5 and 6) and the GTR model (general time reversible model) + I (proportion of invariable sites) + G (gamma distribution shape parameter) was determined as the best-fit nucleotide substitution model for all the datasets. Three different molecular clock models were implemented in the analysis: strict clock, uncorrelated exponential relaxed clock and the uncorrelated log-normal relaxed clock. The uncorrelated exponential relaxed clock model was considered the best model using the Bayes factor test. After fixing the clock model, three independent analyses were run and combined using LogCombiner software included in the BEAST package. The Markov Chain Monte Carlo method was run for 50 to 100 million generations and sampled so that 10,000 trees were saved for further analysis. After 10% burnin and reaching an effective sample size of at least 200 for relevant criteria, a Maximum Clade Credibility tree was created using TreeAnnotator and edited in FigTree (Version 1.3.1). This study is a molecular epidemiology study, and the techniques used (RT-PCR, sequencing, and bioinformatics analysis) are widely used in this research field.

Results

Rubella incidence in mainland China. After a peak number of cases in 2008 (incidence of 9.13/100,000) and a decrease in 2009 (incidence of 5.26/100,000), the average rubella incidence stabilized during 2010–2012, with around 4 cases per 100,000 people. However, the geographic distribution of the incidence across mainland China showed a gradually decreasing trend in most provinces since 2010 (Figure 1). Two provinces (Shandong and Shanxi) had incidences below 1/100,000 in 2010 and 6 provinces (Tibet, Gansu, Henan, Hebei, Jiangsu, and Shandong) reached that level in 2012. In addition, more than half of the provinces had an incidence below 5/100,000. Only Ningxia province had an incidence rate greater than 10/100,000.

Genotyping of RVs collected during 2010–2012. A total of 667 RVs were isolated from throat swabs or urine samples from 24 of 31 provinces of mainland China between 2010 and 2012. Genotyping analysis showed that all viruses were either genotypes 1E (543 strains) or 2B (124 strains). The newly generated sequences were analyzed with previously published sequences to assess the genetic relationship between these new Chinese viruses and other viruses of the same genotype. A total of 153 sequences from viruses collected in mainland China covering all 6 Chinese administrative regions (Dongbei, Huadong, Zhongnan, Xinan, Xibei, and Huabei) were chosen for further analysis (Supplementary Table 1 and supplementary Figure 1).

Evolution of Chinese genotype 1E viruses. Dataset 1, composed of 81 new sequences, 48 previously reported Chinese sequences and 29 international sequences, was used for phylogenetic analysis (Supplementary Table 1A). Figure 2 shows that the sequences of viruses collected in 2010–2012 in mainland China (red) clustered with all but two (black circle) of the previously reported 1E viruses sequences collected in mainland China (blue). In addition to 126 sequences from viruses collected in mainland China, this cluster, designated as 'Chinese lineage', also contains 10 sequences from viruses collected in other countries and regions, namely Taiwan,

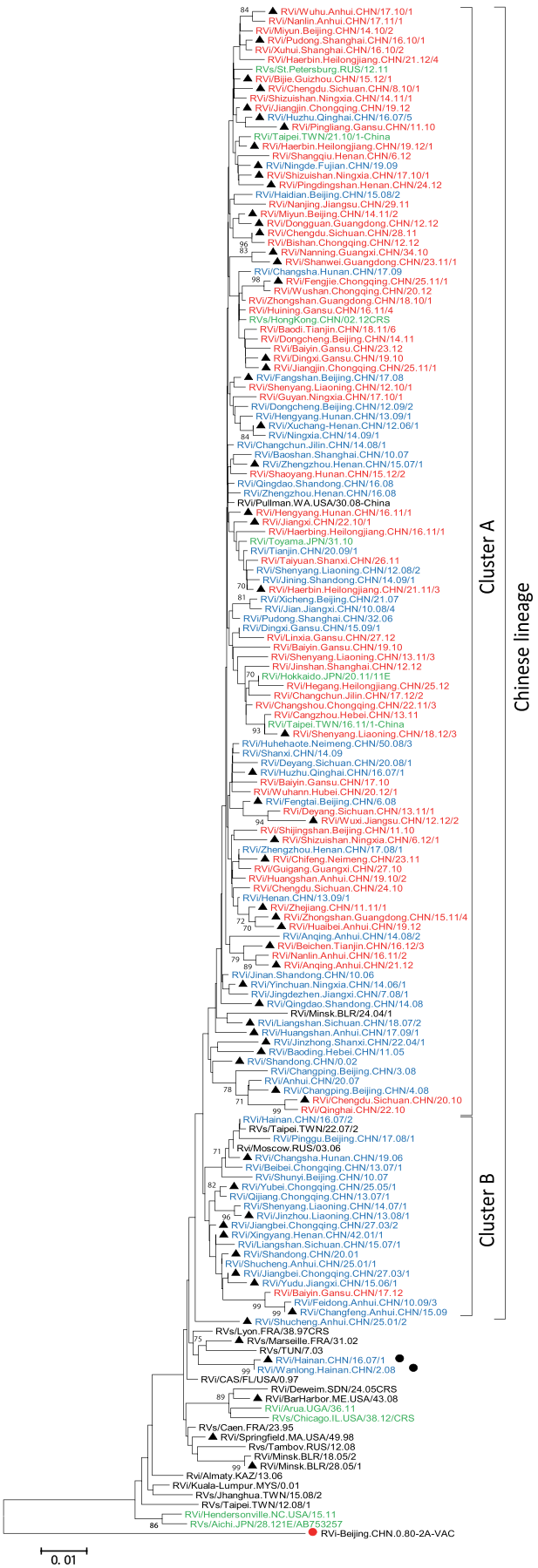


Figure 2 | Neighbor joining phylogenetic tree of 739-nt sequences of genotype 1E rubella viruses. Taxa names are color coded: sequences of viruses collected in mainland China since 2010 are shown in red; those collected in between 2001 and 2009 are shown in blue; those collected outside mainland China since 2010 are shown in green. Rubella viruses selected for whole E1 sequence analysis are indicated by solid black triangles. Two sequences from viruses collected in China that do not fit the Chinese lineage are indicated with a black dot. Bootstrap values greater than 70% are shown. The Chinese vaccine strain BRD II (identified with a red dot) was used as an out-group.

Hong Kong SAR, Russia, USA, Japan, and Belarus (10 viruses collected before 2010 are shown in black, and after 2010 in green in Figure 2). Interestingly, all but one of the 2010–2012 Chinese sequences belong to one group within the Chinese lineage (called Cluster A in this analysis) (Figure 2, Supplementary Table 1A). This cluster comprised 112 sequences with collection dates from 2002–2012; 66 (64%) were collected after 2010. Only one recent sequence (RVi/Baiyin.Gansu.CHN/17.12) belongs to Cluster B, which consists primarily of sequences collected from 2001 to 2009. Cluster B showed little genetic diversity (less than 2% nucleotide diversity) compared to Cluster A. Viruses in both clusters were collected from all 6 Chinese regions (supplementary Table 1A).

In order to better assess Cluster A and Cluster B, but also the rest of genotype 1E viruses, Dataset 2, which includes 739-nt sequences of viruses collected before genotype 1E was identified in China (i.e. 2001), was used for BEAST analysis to determine the TMRCAs and the evolutionary rates for 1E sequences, from the Chinese lineage and Cluster A. The same analyses were run on Dataset 3, which despite having fewer sequences, consists of full-length E1 sequences which increased the number of informative sites. The results from both datasets are relatively similar suggesting that the estimates are likely reliable (Table1). Genotype 1E is estimated to have emerged around 1990 (95% Highest Posterior Density (HPD) interval; 1984–1994), the Chinese lineage around 1997 (95% HPD; 1995–1999), and the Chinese Cluster A around 2000 (95% HPD; 1998–2001) (Figure 3, Table 1). The evolutionary rate is estimated to be around 1.5×10^{-3} substitutions/site/year (95% HPD; 1.19×10^{-3}–1.94×10^{-3}).

Evolution of Chinese genotype 2B viruses. Dataset 4, composed of 24 mainland Chinese sequences and 77 non-Chinese 2B sequences, was used for phylogenetic analysis (Supplementary Table 1B). All but one sequence from viruses collected in mainland China since 2010 (in red, Figure 4) grouped in the same cluster (called Cluster C in this report). This cluster is supported by a bootstrap value of 88%. In addition to viral sequences from mainland China, this cluster also contains sequences from 11 countries, including the Americas, Europe, and Asia, indicating that this cluster has a global distribution. All but one of the recent Chinese sequences further grouped into Cluster Cα which also contained sequences of viruses collected in Argentina, Hong Kong SAR, Japan, Taiwan, the United Kingdom, and Vietnam (Figure 4). Although Cluster Cα sequences are closely related to the rest of Cluster C sequences (1.4% nucleotide divergence), Cluster Cα contains all the sequences of viruses collected after 2010, suggesting that this cluster is expanding. The geographic distribution of 2B viruses in mainland China since 2010 is shown in Supplementary figure 3.

The TMRCA and evolutionary rate of 2B viruses were estimated using datasets 5 and 6. Dataset 5 contains all the dataset 4 739-nt sequences as well as 14 non-2B viruses sequences collected before 1968 when 2B was first identified (Supplementary Table 1). Dataset 6 consists of full-length E1 sequences: 19 new Chinese sequences, 6 non-Chinese sequences and 14 non-2B sequences. Analysis of the 739-nt sequence dataset showed that the genotype 2B emerged

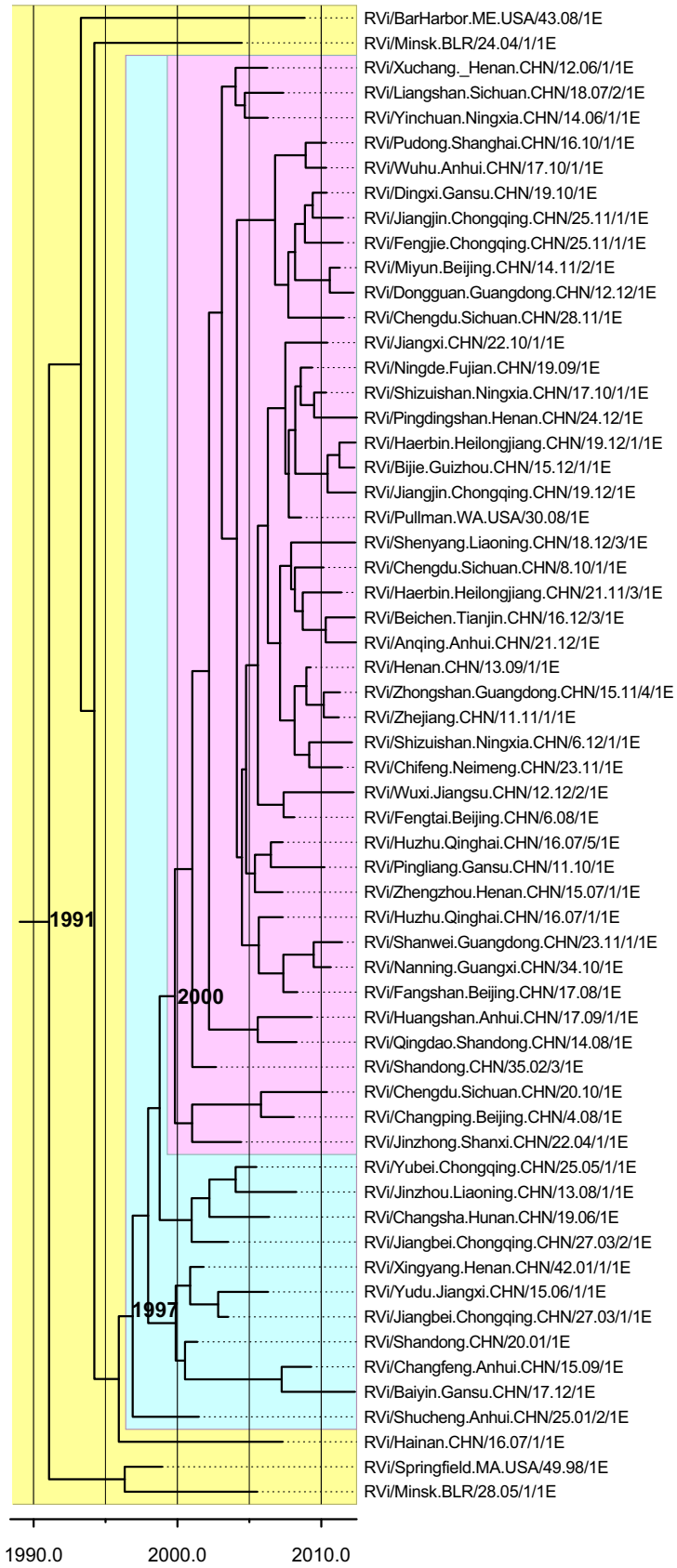


Figure 3 | Maximum clade credibility tree based on the complete E1 sequences of genotype 1E rubella viruses. Sequences corresponding to Cluster A are highlighted in pink. Other sequences assigned to the Chinese lineage are highlighted in blue; the remaining 1E sequences are highlighted in yellow. Mean values of TMRCAs at the node of interest are indicated (see Table1 for 95% HPD values at these nodes). A complete tree is shown in supplementary Figure 2A.

Table 1 | TMRCA and evolutionary rate of genotype 1E and 2B inferred from the Bayesian MCMC method

	739nt-Dataset			E1-Dataset		
Sequences	Number of analyzed sequences	TMRCA[a]	Rate of evolution[b]	Number of analyzed sequences	TMRCA[a]	Rate of evolution[b]
All 1E	158	1990(1984, 1994)	1.56(1.19, 1.94)	60	1990(1984, 1996)	1.34(0.94, 1.74)
Chinese 1E	129	1997(1995, 1999)	1.55(1.12, 2.03)	55	1997(1993, 1999)	1.41(0.98, 1.86)
Chinese 1E-Cluster A	107	1999(1998, 2001)	1.52(1.06, 2.04)	43	2000(1997, 2001)	1.56(0.94, 2.18)
All 2B	101	1962(1953, 1967)	1.98(1.46, 2.48)	25	1969(1943, 1966)	1.45(0.87, 2.02)
2B-Cluster C	65	2001(1999, 2004)	3.03(1.90, 4.12)	15	2004(1999, 2008)	1.80(0.47, 3.25)
2B-Cluster Cα	41	2007(2006, 2008)	1.83(0.72, 3.07)	14	2006(2002, 2009)	2.64(0.01, 5.89)

a, Mean values are indicated, 95% HPD values are in parenthesis.
b, Evolutionary rate is expressed as 10^{-3} substitutions/site/year.

around 1962 (95% HPD; 1953–1967), Cluster C appeared around 2001 (95% HPD; 1999–2004) and Cluster Cα around 2007 (95% HPD; 2006–2008) (Figure 5, Table 1). The 2B sequences are characterized by an evolutionary rate of around 2×10^{-3} substitutions/site/year (95% HPD; 1.46–2.48) (Table 1). The E1 dataset yielded similar estimates although with larger 95% HPD intervals likely due to the limited number of sequences available for analysis (39 sequences).

Genetic diversity of RVs in mainland China. Because 1E is the predominant genotype in China, we used dataset 3, comprising E1 sequences, to analyze the genetic diversity of RVs in mainland China over 10 years (2001–2012) (Figure 6). Prior to 2009, the genetic diversity observed was minimal likely due to an artifact of undersampling, from 2009 to 2010 a rapid increase of genetic diversity occurred, and, lastly, the genetic diversity has remained stable since 2010.

Discussion

This study extends our knowledge of genetic diversity of rubella viruses and the geographic distribution of RV in mainland China. From 2009 to 2010, a sharp increase of genetic diversity was observed likely due to increased surveillance: viruses had been collected from more locations in the country (19 provinces from 2009 to 2012 versus 13 before 2009), thus increasing the likelihood of collecting genetically diverse viruses. Since 2010, the genetic diversity has remained stable; this likely represents the baseline of genetic diversity of RV in mainland China. The incidence of rubella has been gradually decreasing in most provinces since 2010. It is well known, however, that rubella occurs in periodic large epidemics at irregular intervals in endemic countries[20]; thus continued surveillance is necessary to determine if the current low incidence is maintained. Since 2010, an examination of the geographic distribution of genotype 2B in mainland China showed the outbreaks in 10 provinces in the eastern half of the country, whereas genotype 1E viruses were detected in 24 provinces during the same period.

In this study, the sequence data was obtained through RT-PCR after the virus isolation process. Although due to GC rich genome structure, the PCR for rubella virus may not go well in accordance with its sequence, some research indicated that, the nucleotide sequence of the rubella viruses genomes have been shown to be conserved during multiple passages in vivo and in vitro[21]. Although all of the rubella viruses isolates were amplified in cell culture prior to sequencing in this study, the PCR products for rubella viruses may represent their real sequences.

Phylogenetic analysis of genotype 1E sequences showed that almost all viruses from mainland China grouped into a single lineage. This pattern has been reported previously[3,5] and analysis of all mainland China 1E viruses collected between 2010 and 2012 showed that this lineage remained endemic until 2012. The analysis also showed that two closely related clusters (Cluster A and Cluster B) could be identified within the Chinese lineage. Cluster A has been greatly expanding since 2010. The fact that most of the recent sequences are found in Cluster A is unlikely to be due to a surveillance gap due to extensive surveillance since 2010. In fact, Cluster B seems to have been replaced by Cluster A in recent years. Both clusters of Chinese 1E sequences were interspersed with a small number of viruses from close or neighboring countries, such as Japan, Taiwan, and Russia. A study on rubella molecular surveillance using standard epidemiological data in Taiwan found that 15 out of 56 1E viruses collected in Taiwan from 2006 to 2011 were imported from mainland China[22]. The sequences of the exported viruses segregated into one of three 1E lineages in the Taiwan study; that lineage corresponds to the Chinese lineage identified in this paper. In addition, epidemiological data was available to verify that a 1E virus isolated in the United States in 2008 (RVi/Pullman.WA.USA/30.08, JN635288) was imported from mainland China[23]. Thus, this 1E lineage appears to be mainly geographically restricted to China. Previously, RV virological surveillance in the Americas found that genotype 1C was geographically restricted to parts of Central and South America[24]. This restriction was useful in that monitoring the disappearance of genotype 1C viruses provided evidence of the elimination of endemic rubella in these areas due to increased vaccine coverage. Likewise, continued virological surveillance in China could document the disappearance of the Chinese 1E lineage thus providing evidence of the effectiveness of the vaccination program. The current study extends previous work by finding that a specific cluster (Cluster A) within this lineage is now predominant in mainland China.

Phylogenetic analysis showed that the majority of 2B viruses from mainland China group to a lineage (called Cluster C in this report) with the earliest known detection of this lineage being a CRS case in France in 2004. This lineage has been found to have a wide geographic distribution from 2006 to the present. Viruses from this lineage were responsible for rubella epidemics in parts of South America and establishment of genotype 2B endemicity in Brazil and Argentina, regions with no previous documentation of 2B viruses[24,25]. Viruses of the 2B Cluster Cα likely spread to eastern Asia around 2008 when they were first detected in Taiwan and Hong Kong. Subsequently, rubella epidemics caused by Cluster Cα were reported in Vietnam in 2009–2010[26] and Japan in 2011–2013[27]. Sequences from all but one 2B virus obtained between 2011 and 2012 in mainland China share a high degree of similarity with viruses from the Vietnamese and Japanese outbreaks, suggesting this lineage was introduced into mainland China, perhaps via multiple introductions. It appears that a similar pattern occurred in Taiwan, as 28 of the 38 traceable 2B viruses imported into Taiwan from 2006 to 2011 originated in Vietnam[22]. In fact, the geographic proximity and similar timeframes suggest that outbreaks in these three countries could be considered as a multi-country epidemic linked to travel between these countries. It appears that 2B Cluster Cα has become endemic

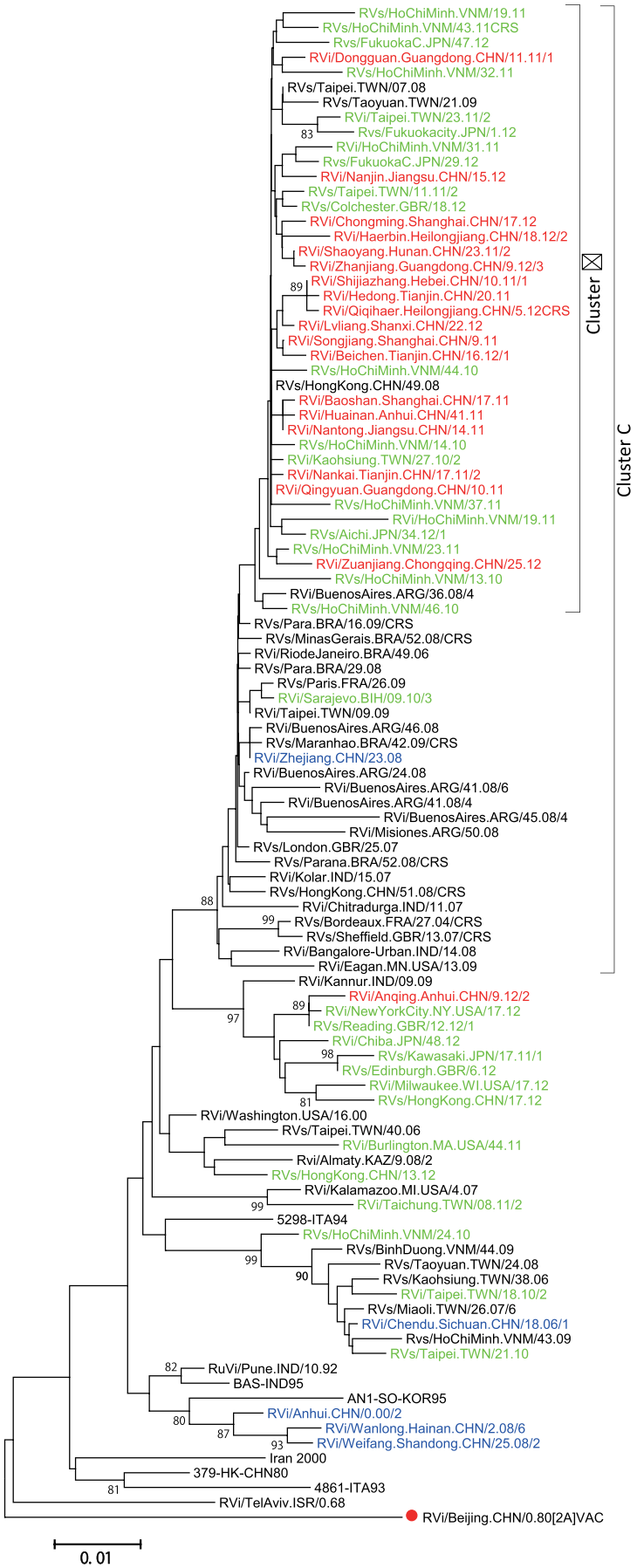


Figure 4 | Neighbor joining phylogenetic tree of 739-nt sequences of genotype 2B rubella viruses. Taxa names are colour-coded according to Figure 2. Bootstrap values greater than 70% are shown. The Chinese vaccine strain BRD II (red dot) was used as an out-group.

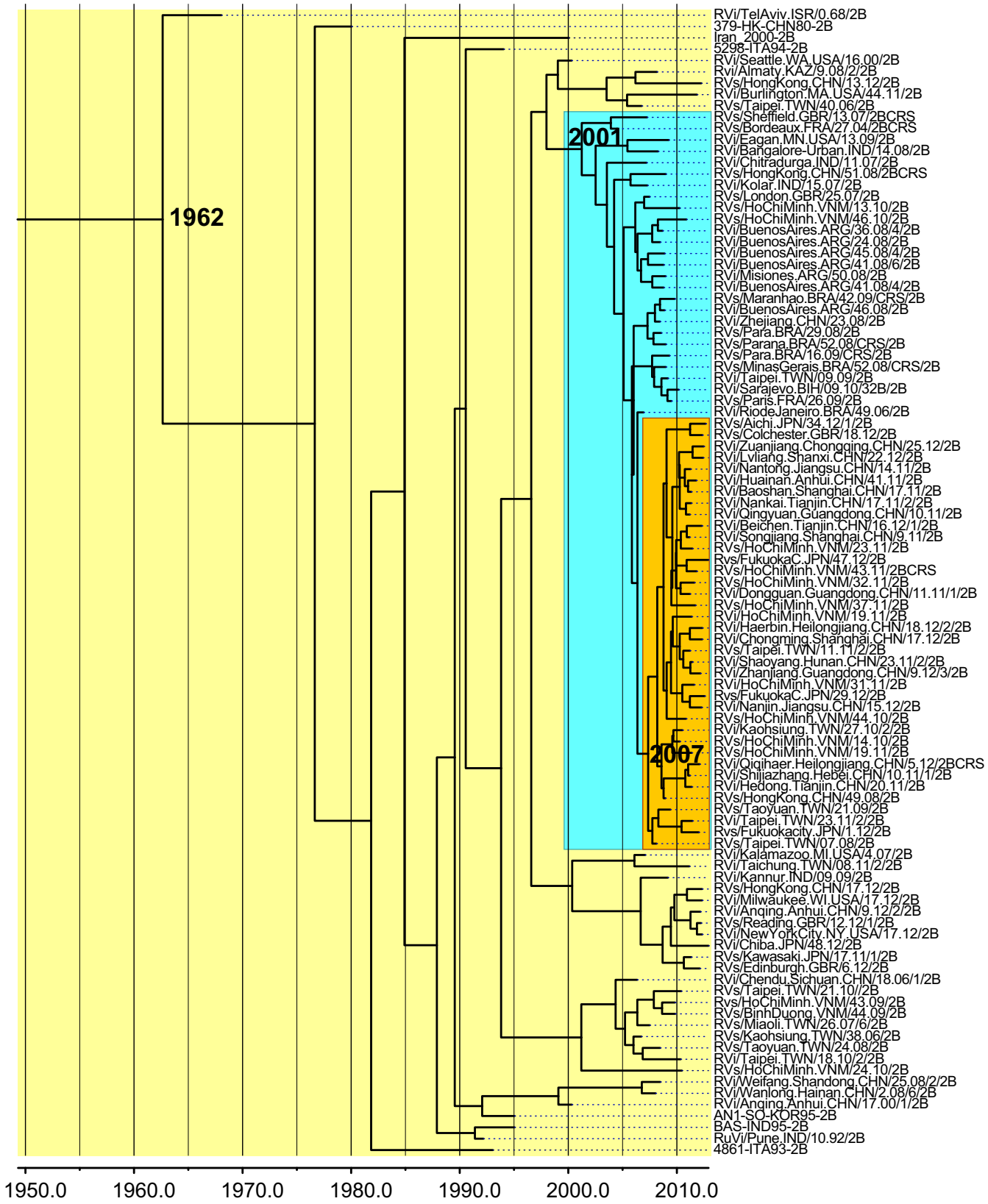
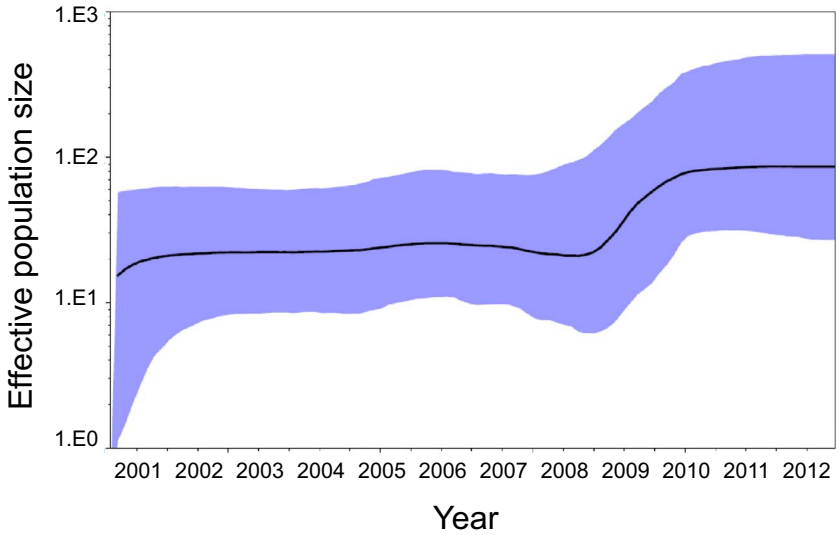


Figure 5 | Maximum clade credibility tree based on the 739-nt sequences of genotype 2B rubella viruses. Sequences corresponding to Cluster Cα are highlighted in orange. Other sequences belonging to the broader global lineage are highlighted in blue; the remaining 2B sequences are highlighted in yellow. Mean values of TMRCAs at the node of interest are indicated (see Table1 for 95% HPD values at these nodes). A complete tree is shown in supplementary Figure 2B.



Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Number of rubella cases reported	ND	ND	ND	24,051	25,545	37,137	74,868	120,614	69,821	43,246	65,663	40,362

Figure 6 | Bayesian skyline plot constructed with 55 Chinese genotype 1E rubella virus sequences sampled from 24 provinces during 2001-2012. The blue area represents the 95% HPD of the effective population size at time estimates. The number of rubella cases according to NNDRS is shown for each year below the skyline plot. ND: Not Determined.

in mainland China (defined as continuous circulation of a lineage for more than 1 year)[28]. Although the number of viruses isolated is low (24 in 2 years), this 2B virus has now been detected in 10 provinces representing 5 of 6 regions of mainland China, with most activity concentrated in the Huadong and Huabei regions. Viruses from this cluster have also been exported to Taiwan: strains RVi/Hsinchu.TWN/30.10/2 and RVi/Tainan.TWN/30/11/1 were confirmed to be imported from China[22]. Interestingly, one 2B virus from 2010–2012 (RVi/Anqing.Anhui.CHN/9.12/2) grouped with a separate cluster. This virus was isolated from an outbreak in a foreign language college with a total of 60 cases among students and teachers and was likely due to a separate importation event.

In addition to being endemic in China, genotypes 1E and 2B are the most common genotypes globally[8,23]. However, these two genotypes do not present the same epidemic pattern. The TMRCA of genotype 1E was estimated at around 1990 in this study. The first genotype 1E virus was detected in 1995 (RVs/Caen.FRA/23.95), suggesting that 1E circulated undetected for at least five years. However, because the number of RV sequences was limited before the mid-1990s, little is known about rubella genotypes from much of the world during the early 1990s[29]. The Chinese lineage appears to have become established a few years after genotype 1E emerged worldwide. Interestingly, this Chinese lineage has apparently not become established in other countries despite international travel. In contrast, the primary lineage of 2B viruses now circulating in China has a wider geographic distribution, including at least three different countries (China, Japan, and Vietnam). Virological surveillance is poor in many rubella-endemic regions, however, so the complete geographic distribution of this lineage is not yet known. Currently, 2B viruses are much less prevalent in China than 1E; however, viruses from the Cluster Cα continue to circulate. It will be interesting to track this 2B lineage to determine whether it follows the 1E pattern of dissemination throughout mainland China, especially as the 1E and Cluster Cα lineages were introduced to China before and after the start of the national vaccination program, respectively. The BEAST analysis identified a slightly higher evolutionary rate for 2B viruses (2×10^{-3} substitutions/site/year; 95% HPD (1.46×10^{-3}, 2.48×10^{-3} substitutions/site/year) compared to 1E viruses (1.56×10^{-3} substitutions/site/year 95% HPD (1.19×10^{-3}, 1.94×10^{-3} substitutions/site/year). These evolutionary rates are comparable to other RNA virus reports[30–32]. The HPD values of the evolutionary rates for 1E and 2B showed significant overlap; we therefore consider these evolutionary rates are not significantly different.

RV remains endemic in many countries and its elimination from China would be an important step toward global the elimination of the virus. Discovery of source cases poses challenges in mainland China. For example, the lack of epidemiologic investigations makes it difficult to distinguish between possible multiple introductions of 2B viruses and continuous transmission. Therefore, prompt and comprehensive case-based epidemiology investigations should be developed to trace the sources of cases, as has been done in other countries. Monitoring the reduction of both Chinese genotype 1E lineage and the currently circulating 2B lineage will aid in assessing the effectiveness of the rubella vaccination program in mainland China.

1. Frey, T. K. Molecular biology of rubella virus. *Adv Virus Res* **44**, 69–160 (1994).
2. Xu, H. *et al.* A rubella outbreak investigation and BRD-II strain rubella vaccine effectiveness study, Harbin city, Heilongjiang province, China, 2010-2011. *Vaccine* **32**, 85–9 (2013).
3. Chen, M. *et al.* Rubella epidemic caused by genotype 1E rubella viruses in Beijing, China, in 2007–2011. *Virol J* **10**, 122 (2013).
4. WHO. Rubella vaccines: WHO position paper. *Wkly Epidemiol Rec* **86**, 301–16 (2011).
5. Zhu, Z. *et al.* Emergence and Continuous Evolution of Genotype 1E Rubella Viruses in China. *J Clin Microbiol* **50**, 353–63 (2012).
6. WHO. Update of standard nomenclature for wild-type rubella viruses, 2007. *Wkly Epidemiol Rec* **82**, 216–22 (2007).
7. WHO. Standardization of the nomenclature for genetic characteristics of wild-type rubella viruses. *Wkly Epidemiol Rec* **80**, 126–32 (2005).
8. WHO. Rubella virus nomenclature update: 2013. *Wkly Epidemiol Rec* **88**, 337–43 (2013).
9. Zhu, Z. *et al.* Rubella virus genotypes in the People's Republic of China between 1979 and 2007: a shift in endemic viruses during the 2001 Rubella Epidemic. *J Clin Microbiol* **48**, 1775–81 (2010).
10. Abernathy, E. *et al.* Analysis of whole genome sequences of 16 strains of rubella virus from the United States, 1961–2009. *Virol J* **10**, 32 (2013).
11. Zhu, Z. *et al.* Genomic analysis of the Chinese genotype 1F rubella virus that disappeared after 2002 in China. *J Med Virol* **86**, 2114–21 (2014).

12. Zhu, Z. *et al*. Comparison of four methods using throat swabs to confirm rubella virus infection. *J Clin Microbiol* **45**, 2847–52 (2007).
13. Abernathy, E. *et al*. Confirmation of rubella within 4 days of rash onset: comparison of rubella virus RNA detection in oral fluid with immunoglobulin M detection in serum or oral fluid. *J Clin Microbiol* **47**, 182–8 (2009).
14. Tamura, K. *et al*. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–9 (2011).
15. Saitou, N. & Nei, M. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. *J Mol Evol* **24**, 189–204 (1986).
16. Kumar, S., Tamura, K. & Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–63 (2004).
17. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–25 (1987).
18. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214 (2007).
19. Posada, D. jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–6 (2008).
20. Dowell, S. F. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg Infect Dis* **7**, 369–74 (2001).
21. Frey, T. K. & Abernathy, E. S. Identification of strain-specific nucleotide sequences in the RA 27/3 rubella virus vaccine. *J Infect Dis* **168**, 854–64 (1993).
22. Cheng, W. Y., Wang, H. C., Liu, M. T. & Wu, H. S. Molecular surveillance of rubella viruses in Taiwan from 2005 to 2011. *J Med Virol* **85**, 745–53 (2013).
23. Abernathy, E. S. *et al*. Status of global virologic surveillance for rubella viruses. *J Infect Dis* **204 Suppl 1**, S524–32 (2011).
24. Icenogle, J. P. *et al*. Virologic surveillance for wild-type rubella viruses in the Americas. *J Infect Dis* **204 Suppl 2**, S647–51 (2011).
25. Figueiredo, C. A. *et al*. Epidemiological and molecular characterization of rubella virus isolated in Sao Paulo, Brazil during 1997–2004. *J Med Virol* **84**, 1831–8 (2012).
26. Tran, D. N. *et al*. Phylogenetic analysis of rubella viruses in Vietnam during 2009–2010. *J Med Virol* **84**, 705–10 (2012).
27. Centers for Disease, C. & Prevention. Nationwide rubella epidemic--Japan, 2013. *MMWR Morb Mortal Wkly Rep* **62**, 457–62 (2013).
28. Reef, S. E., Redd, S. B., Abernathy, E., Kutty, P. & Icenogle, J. P. Evidence used to support the achievement and maintenance of elimination of rubella and congenital rubella syndrome in the United States. *J Infect Dis* **204 Suppl 2**, S593–7 (2011).
29. Frey, T. K. *et al*. Molecular analysis of rubella virus epidemiology across three continents, North America, Europe, and Asia, 1961-1997. *J Infect Dis* **178**, 642–50 (1998).
30. Kuhne, M., Brown, D. W. & Jin, L. Genetic variability of measles virus in acute and persistent infections. *Infect Genet Evol* **6**, 269–76 (2006).
31. Cui, A., Myers, R., Xu, W. & Jin, L. Analysis of the genetic variability of the mumps SH gene in viruses circulating in the UK between 1996 and 2005. *Infect Genet Evol* **9**, 71–80 (2009).
32. Zhang, Y. *et al*. Molecular evidence of persistent epidemic and evolution of subgenotype B1 coxsackievirus A16-associated hand, foot, and mouth disease in China. *J Clin Microbiol* **48**, 619–22 (2010).

Acknowledgments

We thank the provincial and prefectural measles and rubella laboratory staffs and the epidemiologists in mainland China for providing clinical specimens, isolates, and epidemiologic data. This work is supported by the National Natural Science Foundation of China (project no.81101244), Key Technologies R&D Program of National Ministry of Science (2013ZX10004-202, 2012ZX10004201-003, 2012ZX10004215, and 2011ZX10004-001), National Science and Technology Major Project for Creation of Major New Drugs (2013ZX09304101), and WHO Measles Regional Reference Laboratory Funding (WPCHN1002802). The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the U.S. Department of Health and Human Services.

Author contributions

W.B.X. and Z.Z. conceived and designed the experiments. Z.Z., P.R. and E.A. performed data analysis. Z.Z., A.L.C., Y.Z., N.Y.M., S.T.X., S.J.Z., Y.L., Y.W., H.Y.Z., J.L.H., Y.C., C.S.L., F.B., C.F.Z., M.C., P.S.L., F.C.L., S.Y.G., H.G., Y.G., H.C., D.X.F., S.W., X.M.T., Y.K.L., Y.F., L.L.D., T.G., L.X.F. and R.V.S.W.G. performed the experiments. Z.Z., P.R., E.A., J.I. and W.B.X. wrote the main manuscript text, and Z.Z. and P.R. prepared Fig. 1–6. All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhu, Z. *et al*. Evolutionary analysis of rubella viruses in mainland China during 2010–2012: endemic circulation of genotype 1E and introductions of genotype 2B. *Sci. Rep.* **5**, 7999; DOI:10.1038/srep07999 (2015).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

Consortia

The Rubella Virology Surveillance Working Group: Xia Chen[1], Hong Tian[2], Yan Ma[3], Leng Liu[4], Li Liu[5], Jianfeng Liu[6], Hong Fu[6], Yuying Yang[7], Yujie Ma[8], Hua Zhao[9], Fang Huang[10], Ying Hu[11], Hong Zhang[12], Xiaoling Tian[13], Hui Du[14], Xuemin Ma[15], Zhenying Zhang[16], Jin Xu[16], Jianhui Zhou[17], Xufang Ye[18], Jing Li[19], Yiyu Lu[20], Wei Liu[21], Yanni Zhang[22], Shengcang Zhao[23] & Zhuoma Ba[23].

[1]Anhui Provincial Centers for Disease Control and Prevention, People's Republic of China; [2]Tianjin Provincial Centers for Disease Control and Prevention, People's Republic of China; [3]Liaoning Provincial Centers for Disease Control and Prevention, People's Republic of China; [4]Guangdong Provincial Centers for Disease Control and Prevention, People's Republic of China; [5]Sichuan Provincial Centers for Disease Control and Prevention, People's Republic of China; [6]Gansu Provincial Centers for Disease Control and Prevention, People's Republic of China; [7]Shanghai Provincial Centers for Disease Control and Prevention, People's Republic of China; [8]Heilongjiang Provincial Centers for Disease Control and Prevention, People's Republic of China; [9]Chongqing Provincial Centers for Disease Control and Prevention, People's Republic of China; [10]Beijing Provincial Centers for Disease Control and Prevention, People's Republic of China; [11]Jiangsu Provincial Centers for Disease Control and Prevention, People's Republic of China; [12]Hunan Provincial Centers for Disease Control and Prevention, People's Republic of China; [13]Neimeng Provincial Centers for Disease Control and Prevention, People's Republic of China; [14]Hebei Provincial Centers for Disease Control and Prevention, People's Republic of China; [15]Ningxia Provincial Centers for Disease Control and Prevention, People's Republic of China; [16]Henan Provincial Centers for Disease Control and Prevention, People's Republic of China; [17]Jilin Provincial Centers for Disease Control and Prevention, People's Republic of China; [18]Guizhou Provincial Centers for Disease Control and Prevention, People's Republic of China; [19]Hubei Provincial Centers for Disease Control and Prevention, People's Republic of China; [20]Zhejiang Provincial Centers for Disease Control and Prevention, People's Republic of China; [21]Guangxi Provincial Centers for Disease Control and Prevention, People's Republic of China; [22]Jiangxi Provincial Centers for Disease Control and Prevention, People's Republic of China; [23]Qinghai Provincial Centers for Disease Control and Prevention, People's Republic of China