Emergence and Continuous Evolution of Genotype 1E Rubella Viruses in China

Zhen Zhu,1 Aili Cui,1 Huanhuan Wang,1 Yan Zhang,1 Chunyu Liu2, Changyin Wang,3 Shujie Zhou,4 Xia Chen,4 Zhenying Zhang,5 Daxin Feng,5 Yan Wang,6 Haiyun Chen,7 Zhengfan Pan,7 Xiangjie Zeng,7 Jianhui Zhou,8 Shuang Wang,8 Xin Chang,8 Yue Lei,9 Hong Tian,9 Yang Liu,9 Shunde Zhou,10 Jun Zhan,11 Hui Chen,11 Suyi Gu,12 Xiaoling Tian,12 Jianfeng Liu,13 Ying Chen,13 Hong Fu,13 Xiuhui Yang,14 Huanying Zheng,15 Leng Liu,15 Lei Zheng,16 Hui Gao,16 Jilan He,17 Li Sun,17 and Wenbo Xu1*

WHO WPRO Regional Reference Measles/Rubella Laboratory and State Key Laboratory for Molecular Virology & Genetic Engineering, National Institute for Viral Disease Control and Prevention, Beijing, People’s Republic of China1; Institute of Pathogen Biology, Chinese Academy of Medical Science & Peking Union Medical College2; Shandong Provincial Centers for Disease Control and Prevention, People’s Republic of China3; Anhui Provincial Centers for Disease Control and Prevention, People’s Republic of China4; Henan Provincial Centers for Disease Control and Prevention, People’s Republic of China5; Liaoning Provincial Centers for Disease Control and Prevention, People’s Republic of China6; Hainan Provincial Centers for Disease Control and Prevention, People’s Republic of China7; Jilin Provincial Centers for Disease Control and Prevention, People’s Republic of China8; Jilin Provincial Centers for Disease Control and Prevention, People’s Republic of China9; Tianjin Provincial Centers for Disease Control and Prevention, People’s Republic of China10; Jiangxi Provincial Centers for Disease Control and Prevention, People’s Republic of China11; Ningxia Provincial Centers for Disease Control and Prevention, People’s Republic of China12; Neimeng Provincial Centers for Disease Control and Prevention, People’s Republic of China13.
People’s Republic of China; Gansu Provincial Centers for Disease Control and Prevention, People’s Republic of China; Fujian Provincial Centers for Disease Control and Prevention, People’s Republic of China; Guangdong Provincial Centers for Disease Control and Prevention, People’s Republic of China; Shanxi Provincial Centers for Disease Control and Prevention, People’s Republic of China; and Sichuan Provincial Centers for Disease Control and Prevention, People’s Republic of China.

Running title: Evolution of genotype 1E rubella viruses in China

*Corresponding author: Dr. Wenbo Xu, No. 155, Changbai Road, Changping District, Beijing 102206, People’s Republic of China. Tel and Fax: 0086-10-58900187. E-mail: wenbo_xu1@yahoo.com.cn
ABSTRACT

In China, rubella vaccination was introduced into the national immunization program in 2008, and a rubella epidemic occurred in the same year. In order to know whether changes in the genotypic distribution of rubella viruses have occurred in the post-vaccination era, we investigate in detail the epidemiological profile of rubella in China, and estimate the evolutionary rate, molecular clock phylogeny, and demographic history of the predominant rubella virus genotypes circulating in China using Bayesian Markov chain Monte Carlo phylodynamic analyses. 1E was found to be the predominant rubella virus genotype since its initial isolation in China in 2001, and no genotypic shift occurred since then. The results suggest that the global 1E genotype may diverge in 1995 and that it has evolved at a mutation rate of $1.65 \times 10^{-3}$ per site per year. The Chinese 1E rubella isolates were grouped into either cluster 1 or cluster 2, which likely originated in 1997 and 2006, respectively. Cluster 1 viruses were found in all provinces examined in this study with the mutation rate of $1.90 \times 10^{-3}$ per site per year. The effective number of infections remained constant until 2007, and along with the introduction of rubella vaccine into the national immunization program, although the circulation of cluster 1 viruses has not been interrupted, some viral lineages have disappeared, and the epidemic started a decline that led to a decrease in the effective population size. Cluster 2 viruses were only found in Hainan province, which is likely caused by the importation.

Keywords: Genotype 1E rubella virus; continuous circulation; evolution rate; demographic history
INTRODUCTION

Rubella had been considered a relatively benign infection in children and adults. However, the rubella virus is a potent, highly infectious, and teratogenic agent. Infection in the first trimester of pregnancy can lead to fetal death or various birth defects, including deafness, cataracts, and heart disease, known as congenital rubella syndrome (CRS) (11, 16). At least 100,000 cases of CRS occur each year globally (24). An effective rubella and CRS surveillance program has not yet been established nationwide in China. Thus, there is a lack of information on the prevalence of congenital rubella-associated defects and disabilities. The estimated number of CRS cases in China in 2005 was at least 20,000 (17).

Chinese immunization program has a long history and a national immunization program was firstly established in 1978, which was a routine immunization schedule that included 4 basic vaccines against 6 contagious diseases (tuberculosis, diphtheria, neonatal tetanus, whooping cough, poliomyelitis and measles). And by 1988, 1990 and 1995, China achieved its universal childhood immunization goals of >85% coverage by province, county and village, respectively. Hepatitis B vaccine was added into the routine immunization schedule in 2002, and in 2008, the national immunization schedule was expanded from five vaccines against seven contagious diseases to fourteen vaccines against fifteen infectious diseases, and rubella vaccine was introduced into the national immunization schedule in this opportunity.

Chinese immunization program is widely regarded as one of the country's most successful public health endeavors under challenging conditions: 1) Very large population (estimated population, 1.336 billion in 2009; proportion of population aged <15 years, 18.5%; birth rate, 12.09‰; birth cohort, estimated at 16.46 million); (2) high population density with unbalanced distribution (average, 134 persons per square
kilometer; eastern coastal area, >400 persons per square kilometer; central area, about 200 persons per square kilometer; western plateau area, <10 persons per square kilometer; (3) hard-to-reach populations in the mountainous area (32). Despite the challenging conditions, all Chinese children have the opportunity to vaccination against these infectious diseases free of charge, and vaccine preventable diseases targeted by the expanded program on immunization are at low levels of incidence overall.

Rubella is one of the preventable contagious diseases, and with the first introduction of the rubella vaccine in 1969, a significant reduction in morbidity and mortality has occurred. The cycle of rubella epidemics has been broken; pandemics have been prevented, and the number of cases of CRS has been reduced (22). Rubella vaccination was instituted in the Chinese National immunization program in 2008. Prior to this, 2 types of rubella vaccine, including the imported RA27/3 and domestic BRDII vaccines, had been used in several large cities in Eastern China since the 1990s (35).

The rubella virus is the sole member of the Rubivirus genus in the Togaviridae family and is a positive-sense RNA virus with a genome of 9762 nucleotides, which encodes 2 nonstructural polypeptides (P150 and P90) and 3 structural polypeptides (C, E2, and E1). A 739-nt region (nt 8731–9469, aa 159-404 within the E1 protein) within the E1 glycoprotein contains important functional domains including a hemagglutination inhibiting and neutralizing epitope, and antigenic sites (5, 11) and has been designated as the minimum acceptable sequence window for assigning genotypes by comparison with reference virus sequences (29). Thus far, 2 clades, including 13 genotypes, have been described: 9 (1B, 1C, 1D, 1E, 1F, 1G, 2A, 2B, and 2C) are recognized and 4 (1a, 1h, 1i, and 1j) are provisional.
The results of molecular epidemiology studies of rubella viruses in China indicate that 5 (1a, 1E, 1F, 2A, and 2B) out of the 13 genotypes have been present in China since 1979, and cocirculation of these different genotypes has been observed (34). Before the introduction of the rubella vaccine into the national immunization program in 2008, rubella epidemics were common. According to the rubella surveillance data, rubella epidemics occurred approximately every 6 to 8 years in China. A shift in the predominant genotype from 1F and 2B to 1E coincided with the 2001 rubella epidemic, and 1E subsequently became the most common genotype (34).

In this study, we examined whether changes in the genotypic distribution of rubella viruses have occurred in the post-vaccination era, particularly following the recent rubella epidemic peak. The epidemiological profile of rubella and the origin and evolution of the predominant rubella viruses circulating in China are described in detail. We sought to provide data that would contribute to our genetic knowledge base on these viruses and help to form a rational scientific basis for the prevention and control of rubella and CRS in China. We also used a Bayesian approach to estimate the evolutionary rates and the demographic history of rubella viruses in a series of isolates collected between 1995 and 2009.

MATERIALS AND METHODS

Rubella incidence data sources. The number of all clinically diagnosed and laboratory confirmed rubella and measles cases and the annual rubella incidence rates were taken directly from reports from the National Notifiable Disease Reporting System (NNDRS), which was in place since the 1950s and covered all hospitals in mainland China. Hospitals reported cases by posting a card to the county Center for Disease Control (CDC), the data were collected monthly in county CDCs and then
submitted through prefecture and provincial CDCs to reach China CDC finally. Based on the official publication of China Ministry of Health “Country norm of information report and management on public health emergency events,” a rubella outbreak was defined as 10 or more laboratory-confirmed cases occurring in the same school, kindergarten, village, or community within 1 week. Nationwide rubella information could be obtained starting from 2004 due to rubella cases reporting was introduced into the NNDRS, and it was the official source of data on number of reported rubella cases since then.

**Viruses.** All rubella virus strains were isolated from throat swabs and urine samples from patients with clinically suspected rubella within 7 d after rash onset, between 2008 and 2009 in 15 provinces (Anhui, Hainan, Henan, Jilin, Jiangxi, Liaoning, Inner Mongolia, Ningxia, Shandong, Sichuan, Fujian, Gansu, Guangdong, Shanxi, and Tianjin), which covered the east, central, and western regions of China. Clinical specimens were inoculated onto monolayers of African green monkey kidney (Vero) cells, Vero/SLAM cells, or rabbit kidney (RK13) cells, according to standard methods (35). Cells inoculated with clinical specimens were incubated at 35°C for 7 days. The culture supernatant was harvested and used to inoculate fresh cells for up to 2 additional passages. The presence of viral RNA was detected via reverse transcription-polymerase chain reaction (RT-PCR), which amplified a 185-nt fragment of the E1 coding region as previously described (35).

**RT-PCR amplification and sequence determination.** RT-PCR was performed using the Titanium One-step RT-PCR kit (BD Bioscience, Palo Alto, CA) to amplify a 1,107-nt (nt 8656–9762) product containing the 739-nt WHO-recommended sequence window (nt 8731–9469) as previously described (3). The Titanium Taq DNA polymerase used in this kit contains a thermo stable DNA polymerase, and TaqStart™
Antibody, which provides an integrated hot start for increased specificity and yield, and this powerful PCR feature made amplification of GC-rich templates possible by the addition of GC-Melt reagent, which destabilizes base-pairing in GC-rich regions. The PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA), and the amplicons were bi-directionally sequenced using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Hitachi, Japan).

**Phylogenetic analysis.** The 739-nt E1 sequences of the rubella virus strains were aligned and phylogenetic analysis using Neighbor-Joining (NJ) and Maximum Likelihood (ML) was performed in order to gain further insights about molecular epidemiology using the MEGA 5.03 program (Sudhir Kumar, Arizona State University, Tempe, AZ) (25). Tamura-Nei (which takes into account unequal base frequencies, variable transition frequencies, equal transversion frequencies and unequal nucleotide substitution rates among sites) and Kimura two-parameter (which takes into account that transitions should occur more often than transversions, equal base frequencies) evolutionary models were selected for rubella virus phylogenetic analysis, respectively. The branch lengths of the NJ tree were determined from the topology of the trees and were obtained by majority rule consensus among 1000 bootstrap replicates, and bootstrap values greater than 80% were considered statistically significant for grouping. For ML trees, a heuristic search was performed with a subtree pruning- re-grafting (SPR) branch swapping algorithm.

**Evolutionary analysis based on the Bayesian Markov chain Monte Carlo method.** The evolutionary rate, molecular clock phylogeny, and the demographic history of the 1E rubella virus global genotypes were co-estimated using the Bayesian Markov chain Monte Carlo (MCMC) method in BEAST 1.6.1 program (9), and the time of the most recent common ancestor (tMRCA) with 95% highest posterior
density (HPD) was estimated. A Bayesian skyline plot (BSP—a non-parametric
piecewise-constant model) under both strict and relaxed (uncorrelated log-normal
distributed, UCLD) clock conditions was used estimate the demographic history. BSP
uses an MCMC method that allows estimates of effective population size over time
with credibility intervals at every time depending on errors due to the phylogeny
reconstruction and the stochastic nature of the coalescent process (10).

In order to reduce the computation load, sequences with high homogeneity and
identical isolation years were deleted. Data were analyzed under both the
Hasegawa-Kishino-Yano (HKY) and the general time reversible (GTR) nucleotide
substitution models with gamma distribution of among-site rate variation. Two
different models of rate variation among branches were implemented in our analysis:
the strict clock and the UCLD relaxed molecular clock. Both constant and exponential
growth (EG) population size coalescent were used as tree priors. For each model, the
MCMC chain was run for 30,000,000 steps and sampled every 1,000 steps.
Uncertainty in the estimates was indicated by 95% highest posterior density (95%
HPD) intervals. The parameter outputs generated by Bayesian MCMC runs and
convergence on the basis of the effective sampling size (ESS) after a 10% burn-in
were analyzed by a TRACER 1.5 program and the trees were summarized in a target
tree by the Tree Annotator program included in the Beast package by choosing the
tree with the maximum product of posterior probabilities after a 10% burn-in.

Nucleotide sequence accession numbers. The nucleotide sequences of 53
representative viruses of the 118 Chinese rubella strains that were isolated in this
study have been deposited in the GenBank database under the accession numbers
JF702819–JF702871. An additional 35 sequences of rubella viruses from China
during 2000–2007 were retrieved from the GenBank database with accession numbers
RESULTS

The epidemiological profile of rubella in China. Since 2004, the average reported incidence rate of rubella cases increased from 1.85/100,000 (24,051 cases) to 9.13/100,000 (120,614 cases) in 2008 and then decreased to 5.26/100,000 (69,821 cases) in 2009 (Figure 1a). Compared to measles, the incidence of rubella was lower during 2004–2008, while it was slightly higher in 2009 (measles: 3.98/100,000). During 2004–2009, rubella cases were reported in all 12 months, and most of them were concentrated between March and June, but a small peak also occurred in December (Figure 1b).

In 2004 and 2005, an incidence of rubella of >10/100,000 occurred only in Xinjiang province (western China). The incidence of rubella in all 31 provinces was <10/100,000 in 2006. However, the number of rubella cases increased in 2007, with incidence rates of >10/100,000 in 5 of 31 provinces. Tianjin and Liaoning provinces (eastern China), and Chongqing and Qinghai provinces (western China) had an incidence rate of >15/100,000. Among these, Liaoning had the highest incidence, at 48.90/100,000. The nationwide rubella epidemic continued and reached a peak in 2008. In 2008, 10/31 provinces had a rubella incidence of >15/100,000 [Tianjin, Zhejiang, Hainan (eastern China), Jilin (central China), Chongqing, Yunnan, Tibet, Gansu, Ningxia, and Xinjiang (western China)]. The epidemic was widespread throughout China, with the highest incidence in Tianjin (49.98/100,000). In 2009, the epidemic subsided, with an incidence of >15/100,000 limited to 4 provinces,
including Zhejiang (eastern China), Tibet, Gansu, and Ningxia province (western China), with Tibet having the highest incidence (58.15/100,000) (Figure 2).

Between 2004 and 2009, the total number of reported rubella outbreaks was approximately 972, of which, 931 (95.78%) occurred in schools. The reported rubella cases were concentrated in those under 15 years of age (Figure 3a), with the proportion being 78.70%, 75.34%, 66.66%, 58.69%, and 64.05%, for each year between 2004 and 2009, respectively (Figure 3b). The proportion of reported rubella cases within the 15 to 39 age group increased each year from 2004 (19.77%) to 2007 (42%), maintained a similar level in 2008 (40.51%), and then decreased to 34.71% in 2009 (Figure 3b). The proportion of rubella cases in individuals aged <15 years among the eastern, central, and western regions of China was quite varied. Between 2004 and 2009, the proportion of rubella cases in individuals aged <15 years in eastern China was 62.84%, 58.07%, 42.04%, 37.02%, 46.40%, and 42.45%, respectively, for each of these years. These were relatively lower than those in central and western China, whereas the proportion of reported rubella cases within the 15 to 39 age group in eastern China was higher than that in central and western China (35.80%, 40.27%, 56.63%, 62.07%, 52.51%, and 55.57%, respectively, for each of these years) (Figure 3c).

**Phylogenetic analysis of rubella viruses.** A total of 118 rubella isolates were collected from throat swabs or urine specimens in 15 of 31 provinces in China during the rubella epidemic between 2008 and 2009. The sequences of the 739-nt region within the E1 gene from virus isolates from the outbreaks were very similar (<0.54% difference) or identical, so 53 rubella viruses were selected (randomly selected on the basis of their genetic relationships) as representative viruses for phylogenetic analysis. All viruses were named according to the WHO systematic nomenclature for rubella
viruses. The nomenclature system of strain naming includes epidemiological information that is essential for interpretation of the molecular data. And strains or sequences will be designated as either RVi, rubella virus isolate in cell culture; or RVs, rubella virus sequence derived from RNA extracted from clinical material. Other information to be included in the strain/sequence name include town/city of isolation, country, date of specimen collection by epidemic week and year, isolate number, genotype, and special designation for sequences if it derived from CRS Cases.

Both Neighbor-Joining (NJ) and Maximum Likelihood (ML) analysis with Tamura-Nei and kimura two-parameter models gave the similar results. The 53 sequences of Chinese rubella viruses during 2008 and 2009 were divided into 2 genotypes: genotype 1E (51 strains, 96.2%) and genotype 2B (2 strains, 3.8%), indicating that 1E was predominant. These genotype assignments were supported by high bootstrap scores (Figure 4).

Genotype 1E rubella viruses were found in all 15 provinces, and a 96.7–100% nucleotide identity (99.1%–100% amino acid identity) was found among these viruses. Identical and similar sequences of genotype 1E rubella viruses during 2008–2009 were found in different provinces, including Guangdong, Liaoning (eastern China), Anhui, Jilin (central China), and Ningxia (western China) (Figure 4); this indicates that highly similar or identical 1E rubella virus sequences circulated in the various provinces with no apparent geographic restriction.

Origin, evolutionary rate, molecular clock phylogeny, and the demographic history. The sequences of the 739-nt window within the E1 gene of the representative Chinese rubella isolates from 2008 to 2009 (n = 51), 2001 to 2007 (n = 35, from 13 provinces), and 13 other countries (n = 19), including the Bahamas (33), Canada (33), Germany (33), Italy (33), USA (33), Suriname (33), France (26), Tunisia (26), Belarus
(12), UK (14), Russia, Ukraine, Sudan, and Kazakhstan (Table 1), representing the global genotype 1E rubella virus population, were selected for divergence time and substitution rate estimation using the Bayesian MCMC method. Different models were used for data analysis, and it was found that UCLD with EG fit our data best, while the HKY and GTR nucleotide substitution models had no significant impact on the analysis (Table 2). The coefficient of variation of the estimated evolutionary rates among branches was 0.19 (95% HPD, 0.01 to 0.44) and 0.20 (95% HPD, 0.01 to 0.46) estimated by the HKY model and GTR model, respectively, indicating that rate heterogeneity exists among the different branches. Analysis of the partial E1 fragment sequences showed that the substitution rate of the 1st + 2nd codon positions was significantly lower than that calculated for the 3rd codon position (mean relative substitution rates of 0.17 with a 95% HPD of 0.08–0.28, and 2.69 with a 95% HPD of 2.56–2.82, respectively).

Chinese rubella viruses were distributed in multiple different clusters within genotype 1E, and the sequences of 1E rubella viruses circulating in 2008 and 2009 interdigitated with the sequences from 2001 to 2007. In the molecular clock phylogenetic tree (Figure 5), the Chinese strains belonged to either cluster 1 or cluster 2, composed of isolates from 2001 to 2009 and 2007 to 2008, respectively. Cluster 1 was unique to China. Most of the viruses (81/86) belonged to cluster 1 and were present in all 19 provinces. Some viruses in cluster 1 (such as isolates from Shandong, Anhui, and Henan provinces) were isolated in 2001, and 9 years have passed since the last isolation. In contrast, recent isolates from these provinces are located in different lineages, indicating that the older strains may have been eliminated. Five viruses (5/86) from Hainan province belonged to cluster 2. Although most Chinese genotype 1E rubella viruses had a distinct cluster (cluster 1), some (cluster 2) still group with
viruses isolated from the United Kingdom, Belarus, Suriname, Italy, Germany, the United States, Bahamas, Russia, Canada, France, Tunisia, Ukraine, and Kazakhstan between 1997 and 2008 (Figure 5).

Different estimated evolutionary rates for different branches were observed, and the mean rate was $1.65 \times 10^{-3}$ substitutions per site per year (HKY) or $1.66 \times 10^{-3}$ substitutions per site per year (GTR). Cluster1 viruses within the 1E genotype had a faster estimated evolutionary rate of $1.90 \times 10^{-3}$ per site per year (HKY) or $1.95 \times 10^{-3}$ per site per year (GTR). The most recent common ancestor of global genotype 1E rubella virus may be traced back to February 1995 (range, January 1995 to April 1995). The $t_{MRCA}$ estimates for Chinese clusters were dated to May 1997 (cluster 1) and March 2006 (cluster 2).

Both the strict and relaxed clock models were implemented using the Bayesian skyline model for population growth, with the second being preferred. The BSP (Fig. 6) showed that the effective number of infections remained constant from the time of the root until about 4 years ago (2001–2007), when the epidemic started a decline that led to a decrease in the effective population size, although a rubella epidemic occurred in the epidemiology in 2008.

DISCUSSION

The rubella epidemiological data collected in this study indicate that the high prevalence of rubella in China alternated between different regions and that the endemic continuously occurred in areas covering eastern, central, and western portions of the country. Overall, the incidence of rubella in western China was significantly higher than that in other regions.
Reported rubella cases were concentrated in those under 15 years of age, most of whom were primary and middle school students. However, in eastern China, the proportion of rubella cases in those less than 15 years of age decreased significantly, and the proportion within the 15 to 39 age group increased. This may be due to the introduction of the rubella vaccine into some provinces in eastern China, such as Shandong, since the 1990s. The immunization strategy involved 2 doses of vaccine, including the first dose for infants between 8–18 months and the second dose for 7- or 12-year-olds (30). If routine vaccine coverage in children is not maintained, immunization of children could alter transmission dynamics and potentially lead to an increase in susceptibility in older age groups (27). A shift in risk to older age groups has already occurred in Brazil and Costa Rica (4). This is of concern because older age groups include women of childbearing age, thus raising the potential for an increase in the incidence of CRS.

In 2008, the rubella vaccine was introduced into the national immunization program in China. Measles-rubella vaccine (MR) was introduced for 8-month-old infants, and measles-rubella-mumps vaccine (MMR) was given to infants 18–24 months of age. However, in some circumstances, when rubella vaccines were not available, they were replaced by measles vaccines in order to achieve the measles elimination goal initiated by WHO (19). Furthermore, seroepidemiological surveys of Chinese females in the cities of Shenzhen (20), Beijing, and Chongqing (18) revealed that only about 80% were immunized against rubella, which is too low to provide herd immunity to the population. Considering the shift in the ages at risk and the seroepidemiological data, the development of a routine rubella vaccination program should be a priority, and both children and women of childbearing age should be included.
In 2005, the WHO Regional Committee of the Western Pacific Region (WPR) formally declared regional measles elimination a goal with a target date set for 2012 (1). Measles and rubella are similar diseases, both characterized by a rash that may be difficult to differentiate clinically (35). In addition, rubella epidemics usually occur in the spring or early summer, similar to measles, peaking between April and May (13). Therefore, during the measles elimination campaign in China, particularly in 2009, when the rubella incidence was higher than that of measles, measles control might have been hampered because large numbers of suspected measles cases, which were in reality rubella cases, may have appeared. Therefore, it is crucial to eliminate rubella during measles eradication campaigns. However, countries wishing to control and eliminate rubella must not only maintain high vaccine coverage but also be supported by high-quality surveillance, including molecular epidemiological studies, in order to obtain information about the circulation of indigenous viruses and the importation of new strains from other parts of the world (29).

The data presented in this study demonstrate that 1E rubella has been the predominant virus genotype since 2001 and that only 2 provinces had incidences of the genotype 2B virus in 2008. Therefore, the genotype 1E rubella virus is most likely to cause a nationwide epidemic. During 2001–2007, multiple transmission chains of genotype 1E rubella viruses were found in different parts of China (34). Some transmission chains of 1E faded, while other similar sequences continued to circulate in various provinces between 2008 and 2009. This might be related to the large-scale use of rubella vaccine during the nationwide immunization program in 2008, which may have interrupted some transmission chains. Some genotype 1E strains still survive and circulate.
Chinese genotype 1E rubella virus isolates were grouped into 2 clusters. Cluster 1 isolates were collected from 2001 to 2009, revealing the long-term circulation of this cluster in China since its emergence in 1997 (based on Bayesian MCMC analysis). Cluster 2 consisted of isolates from Hainan province between 2007 and 2008 (that may emerge in 2006, based on Bayesian MCMC analysis), suggesting that it may have been introduced into Hainan province recently.

Cluster 1 strains within the 1E genotype predominate in China and have not been found elsewhere, although sequences of many other 1E strains from other countries are available for comparison. This indicates that cluster 1 is unique to China. These viruses may have arisen from mutations and random genetic drifts that conferred a selective advantage to this lineage following its emergence in 1997.

Various methods based on Bayesian statistics have recently been developed that allow evolutionary rates, molecular phylogeny, and population dynamics to be co-estimated in a single analysis starting from a nucleotide sequence alignment. The phylogeny and times of divergence of the rubella virus lineages using partial E1 fragment sequences sampled at different times was inferred using a relaxed molecular clock model, which incorporate the time-dependent nature of the evolutionary process by assuming independent rates in different branches rather than a strict clock.

Recently developed methods based on coalescent theory for inferring the demographic history of a population on the basis of gene sequences of a sample have allowed the reconstruction of the past history of epidemics due to highly variable RNA viruses (6, 21). BSP, a non-parametric piecewise-constant model of population size, can fit several models, which solved the problem that coalescent methods usually require the assumption of a demographic model that is a mathematical description of the changes in effective population size, and seldom information was obtained about
the demographic behavior of a study in advance. The BSP approach makes it possible to reconstruct novel and complex demographic scenarios (10). The greater sensitivity of the non-parametric BSP method showed that the rubella virus population size remained constant until the 2007, when a decline in the effective number of infections occurred with the introduction of rubella vaccine into the national immunization program in 2008.

In vitro, RNA virus nucleotide misincorporation rates per site ranged from $10^{-3}$ to $10^{-6}$, due to the intrinsic error rates of the RNA polymerase and the lack of proofreading (2, 8). Genetic mutations are the basis of RNA virus evolution, as they allow the virus population to rapidly adapt to new environments and escape host anti-viral responses (7). As an RNA virus, rubella has the potential to continually mutate, so close monitoring of the genetic variations of wild-type rubella strains is necessary. In this study, we estimated that the mean mutation rate of genotype 1E rubella viruses was $1.65 \times 10^{-3}$ substitutions per site per year based for the 739-nt window of the E1 gene, which is lower than that of the measles virus ($0.78 \times 10^{-2}$) (15), mumps viruses ($1.86 \times 10^{-2}$) (7), and Coxsackie virus A16 ($0.91 \times 10^{-2}$) (31). The effects of strong negative selection can be seen in the sixteen-fold lower evolutionary rate in the 1st + 2nd codon positions of the partial E1 genome sequence, frequently causing amino acid changes, than in the 3rd codon position, rarely resulting in amino acid substitution. The evolution rate of rubella is relatively slow may due to highly conserve in both nucleotide and amino acid sequences of rubella virus (11).

Although the 739-nt region within the E1 glycoprotein for diagnostic applications contains important functional domains including a hemagglutination inhibiting and neutralizing epitope, and antigenic sites, negative or positive selection pressures may
differ considerably across a viral genome, further studies on other genes or whole genome are required to confirm this finding.

To our best knowledge, genotype 1E viruses were firstly identified in 1995 in France (26) and then in 1997 in the United States, Canada, the Caribbean, and Italy (23, 33). In addition to North America and Europe, 1E viruses have now been isolated in South America, Africa, and Asia (28, 35). Global genotype 1E viruses investigated in this study share the recent common ancestor that originated in 1995, the same year of the first isolation in France, suggesting that genotype 1E viruses first appeared in Europe. However, analysis of additional sequences is needed to determine the origin and evolution of genotype 1E rubella viruses with greater accuracy.

A shift in the predominant genotype from 1F and 2B to 1E was found with the 2001 rubella epidemic in China. Subsequently, 1E rubella viruses continually circulated in China for more than 9 years. Epidemiological data shows that a rubella epidemic occurred in 2008. The transmission dynamics of endemic viruses may change due to vaccine introduction in 2007 and these may be contained or even eliminated. Ongoing molecular epidemiological surveillance of circulating rubella viruses is necessary.

Because human beings are the only reservoir of rubella virus, so the transmission of rubella virus is influenced by various factors relating to the amount and virulence of the viruses, the number of unimmunized hosts, and their interaction. So it can be postulated that the more recent decrease in the effective number of the rubella virus infections (those actually transmitted) might be due to a decrease in the virus population that is related to routine vaccination with rubella vaccine in national immunization program and decreasing in immunity gap related to the susceptibility to the infection.
In conclusion, after the first isolation of 1E genotype rubella virus (cluster 1) in 2001 in China, it continuously circulated throughout the epidemic of 2008, but unlike the last rubella outbreak in 2001, no genotypic shift occurred. In addition, although the circulation of the virus has not been interrupted following the introduction of the rubella vaccine into the national immunization program in 2008, some lineages within the 1E genotype disappeared. Cluster 2 viruses within the 1E genotype were only found in Hainan province, which is most likely the result of importation from other regions due to high rates of migration and tourism. The most recent common ancestor of global genotype 1E rubella virus may be traced back to February 1995 (range, January 1995 to April 1995). The \( t_{\text{MRCA}} \) estimates for Chinese clusters were dated to May 1997 (cluster 1) and March 2006 (cluster 2). Estimated by BSP method, the effective number of infections remained constant until 2007, although a rubella epidemic occurred in the epidemiology in 2008, the epidemic started a decline that led to a decrease in the effective population size along with the introduction of rubella vaccine into the national immunization program in the same year.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (project no. 81101244), National Infectious Diseases Surveillance Program (2008ZX10004-001, 2008ZX10004-008, 2009ZX10004-201, and 2009ZX10004-202), and WHO measles regional reference laboratory funding (WPCHN1002802). We thank all the provincial and prefectural measles and rubella laboratory staffs and the epidemiologists in China for providing clinical specimens, isolates, and epidemiologic data; we thank WHO HQ and WPRO for the technical and financial support.
We would also like to acknowledge all of the laboratories that isolated the viruses used in this study. And we thank anonymous reviewers for comments that improved the manuscript.

All authors report no conflicts of interest.
REFERENCES


Figure 1. Reported rubella cases in China by year of onset (A) and by month of onset (B), 2004–2009

Figure 2. Geographic distribution of rubella cases in China, 2004–2009

Figure 3. Age distribution of rubella cases in China, 2004–2009
   (a) Age-specificity of rubella cases in mainland China from 2004 to 2009, as reported by NDRIS
   (b) The proportion of rubella incidences in different age groups from 2004 to 2009 in mainland China, as reported by NDRIS
   (c) Rubella incidences in different age groups from 2004 to 2009 in eastern (E), central (C), and western regions (W) of China, as reported by NDRIS

Figure 4. Phylogenetic analyses of sequences of Chinese rubella viruses. (a) Phylogenetic analysis of sequences of 53 representative Chinese rubella viruses from 2008 to 2009 compared to the WHO reference sequences based on the WHO standard sequence window within the E1 gene (nt 8731–9469). Numbers in parentheses are the numbers of identical or similar sequences found in the same outbreak. The genotype 1E rubella viruses isolated during 2008–2009 are indicated by solid triangles. (b) Phylogenetic analysis of representative Chinese rubella virus strains of genotype 1E compared to rubella viruses from the other countries, based on the WHO standard sequence window. The 51 representative rubella virus strains isolated during 2008–2009 and 35 virus strains isolated during 2001–2007, which were downloaded from Genbank, are indicated by solid triangles and solid rounded diamonds,
respectively. WHO reference strains are indicated by a solid square. The Chinese
vaccine strain BRD II was used as an out-group.

Figure 5. MCMC tree of the 739-nt region of the E1 sequences of genotype 1E rubella
viruses throughout the world visualized in FigTree. The width of a branch reflects the
evolutionary rate of individual sequences and their reconstructed ancestors. Genotype
1E rubella viruses from China during 2001–2009 segregated into 2 clusters (1 and 2).
Chinese genotype 1E viruses were named according to WHO standard nomenclature

Figure 6. Bayesian skyline plot obtained by analyzing the 73 1E genotype rubella
virus sequences sampled at different times. Ordinate: the number of effective
infections at time; abscissa: time (in years). The thick solid line represents the median
and the grey area the 95% HPD of the number of effective infections at time
estimates.
## Table 1. Representative genotype 1E strains from countries other than China used in this study.

<table>
<thead>
<tr>
<th>Virus isolate</th>
<th>Isolation site and year</th>
<th>GenBank no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rvs/Caen.FRA/23.95/1E</td>
<td>France, 1995</td>
<td>FN546967</td>
<td>[16]</td>
</tr>
<tr>
<td>FRI-BAH97</td>
<td>Bahamas 1997</td>
<td>AY326359</td>
<td>[15]</td>
</tr>
<tr>
<td>DES/MB-CAN97</td>
<td>Manitoba, Canada 1997</td>
<td>AY326358</td>
<td>[15]</td>
</tr>
<tr>
<td>G432-GER99</td>
<td>Germany 1999</td>
<td>AF551761</td>
<td>[15]</td>
</tr>
<tr>
<td>6423/PV-ITALY-1997</td>
<td>Pavia, Italy 1997</td>
<td>AY161374</td>
<td>[15]</td>
</tr>
<tr>
<td>CAS/FL-USA97</td>
<td>Florida, USA 1997</td>
<td>AY326356</td>
<td>[15]</td>
</tr>
<tr>
<td>CAB/NY-USA00</td>
<td>New York, USA 2000</td>
<td>AY326355</td>
<td>[15]</td>
</tr>
<tr>
<td>Rvs/Angers.FRA/36.03[1E]</td>
<td>France, 2003</td>
<td>FN547016</td>
<td>[16]</td>
</tr>
<tr>
<td>Rvs/TUN/7.03[1E]</td>
<td>Tunisia, 2003</td>
<td>FN547014</td>
<td>[16]</td>
</tr>
<tr>
<td>Rvi/Deweim.SDN/24.05[1E]CRS</td>
<td>Sudan, 2005</td>
<td>FJ775000</td>
<td>-</td>
</tr>
<tr>
<td>Rvi/Minsk.BLR/18.05/2[1E]</td>
<td>Minsk, Belarus 2005</td>
<td>AM258955</td>
<td>[17]</td>
</tr>
<tr>
<td>Rvs/London.GBR/08.05[1E]CRS</td>
<td>London, UK 2005</td>
<td>EF210051</td>
<td>[18]</td>
</tr>
<tr>
<td>Rvs/Lille.FRA/25.05[1E]</td>
<td>France, 2005</td>
<td>FN547019</td>
<td>[16]</td>
</tr>
<tr>
<td>Bar4-108.RUS/06</td>
<td>Barnaul city, Russia 2006</td>
<td>EF421978</td>
<td>-</td>
</tr>
<tr>
<td>Rvs/Chernivtsi.UKR/13.07[1E]</td>
<td>Ukraine, 2007</td>
<td>FJ711683</td>
<td>-</td>
</tr>
<tr>
<td>Rvi/Vladimir.RUS/9.08[1E]</td>
<td>Russia, 2008</td>
<td>FJ711681</td>
<td>-</td>
</tr>
<tr>
<td>Rvi/Almaty.KAZ/13.06[1E]</td>
<td>Kazakhstan, 2006</td>
<td>FJ711684</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2. Origin and evolutionary rate inferred from the Bayesian MCMC method on the 739-nt window within the E1 gene

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean value of the parameter (95% HPD) as determined by:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HKY+UCLD+EG</td>
</tr>
<tr>
<td>Mean evolutionary rate (10^{-3} substitutions/site/yr)</td>
<td>1.65 × 10^{-3} (1.32 × 10^{-3}–2.00 × 10^{-3})</td>
</tr>
<tr>
<td>Cluster 1 evolutionary rate (10^{-3} substitutions/site/yr)</td>
<td>1.90× 10^{-3} (1.44 × 10^{-3}–2.40 × 10^{-3})</td>
</tr>
<tr>
<td>Coefficient of variation</td>
<td>0.19 (0.01–0.44)</td>
</tr>
</tbody>
</table>

2 Rate of molecular evolution given as numbers of nucleotide substitutions per site per year.

3