

# Evolutionary genetics of genotype H1 measles viruses in China from 1993 to 2012

Songtao Xu,<sup>1†</sup> Yan Zhang,<sup>1†</sup> Pierre Rivaller,<sup>2‡</sup> Huiling Wang,<sup>1</sup> Yixin Ji,<sup>1</sup> Zhu Zhen,<sup>1</sup> Naiying Mao,<sup>1</sup> Chongshan Li,<sup>3</sup> William J. Bellini,<sup>2‡</sup> Wenbo Xu<sup>1</sup> and Paul A. Rota<sup>2‡</sup>

## Correspondence

Paul A. Rota  
prota@cdc.gov  
Wenbo Xu  
wenbo\_xu1@aliyun.com

<sup>1</sup>WHO Regional Reference Laboratory for Measles for the Western Pacific Region, National Institute for Viral Disease Control and Prevention, China Center for Disease Control and Prevention, Beijing 102206, PR China

<sup>2</sup>Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

<sup>3</sup>Shanghai Center for Disease Control and Prevention, Shanghai City, PR China

Virologic surveillance is a critical component of measles management. One of the criteria for verification of elimination of endemic measles is genetic analysis of wild-type viruses to demonstrate lack of an indigenous genotype. Measles is yet to be eliminated in China, and genotype H1 has been detected continuously since virologic surveillance was initiated in 1993. Virologic surveillance has been very active in China, providing a unique opportunity to conduct a detailed study of the evolution of a single, endemic genotype over a timespan of nearly two decades. Phylogenetic analysis performed on the 450 nt coding sequence for the C-terminal 150 amino acids of the nucleoprotein (N-450), fusion (F) gene and haemagglutinin (H) gene confirmed the continued circulation of genotype H1 viruses for 19 years. No evidence of selective pressure for the H protein was found. The substitution rates ranged from  $0.75 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> for H to  $1.65 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> for N-450. The time of most recent common ancestor (TMRCA) for genotype H1 was estimated as approximately 1985 (95% highest probability density, 1979–1989). Finally, the overall diversity of measles sequences from China decreased from 2005 to 2012, coincident with a substantial decrease in measles cases. The results suggest that detailed evolutionary analyses should facilitate the documentation of eventual measles elimination in China. Moreover, the molecular approaches used in this study can be applied in other countries approaching measles elimination.

Received 8 April 2014

Accepted 4 June 2014

## INTRODUCTION

Measles is a highly contagious disease characterized by high fever, cough and a maculopapular rash. The causative agent, measles virus (MeV), has a single-stranded, negative-sense RNA genome. MeV belongs to the genus *Morbillivirus* within the family *Paramyxoviridae*. The MeV genome is 15 894 nt in length, and contains six genes encoding nucleoprotein (N), phosphoprotein (P), matrix (M), fusion (F), haemagglutinin (H) and polymerase (L). The H and F

proteins play key roles in viral entry and are the main targets of virus-neutralizing antibodies (Griffin & Bellini, 1996).

A safe and effective vaccine is available to control measles. Countries and regions with high vaccination coverage have successfully eliminated the transmission of endemic MeV. The World Health Organization (WHO) Region of Americas achieved measles elimination in 2002. The remaining WHO regions have established measles elimination goals for 2015–2020 (Anonymous, 2013; WHO, 2012a). Global vaccination programmes have achieved major reductions in measles-related mortality. Between 2000 and 2012, estimated global deaths due to measles decreased by 78%, from 562 400 to 122 000, both historically low levels (Perry *et al.*, 2014). In 2005, China set a goal of measles elimination by 2012, and is currently working to achieve this aim by 2015 according to the regional plan of the Western Pacific Region.

Documentation of the interruption of MeV transmission is an essential criterion to confirm elimination, and laboratories play a key role in this process by providing virologic

<sup>†</sup>These authors contributed equally to this work.

<sup>‡</sup>The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

GenBank accession numbers for viruses included in this study are listed in Tables S1, S2 and S3.

Three supplementary tables and three supplementary figures are available with the online version of this paper.

surveillance data (WHO, 2012a). The Chinese Measles Laboratory Network (CMLN), introduced in 2001 in China, is composed of one national, 31 provincial and 331 prefectural laboratories. The National Measles Laboratory performs genetic characterization of MeVs isolated in all the provinces.

Although MeV is monotypic, sequence differences between wild-type viruses have led to the designation of eight clades and 24 genotypes (Rota *et al.*, 2011; WHO, 2012b). Genotype H1 was identified as the endemic genotype in China with the initiation of viral surveillance in 1993 (Ji *et al.*, 2009, 2010; Xu *et al.*, 1998; Zhang *et al.*, 2007, 2008). While genotype H1 viruses have spread to other countries, this genotype is endemic only in China. The introduction of virologic surveillance in China has facilitated the genotyping of several thousand MeVs to date. Sequences from the genotype H1 viruses in China form two distinct clusters, previously designated cluster 1 and cluster 2 (Zhang *et al.*, 2007, 2008, 2009). Here, we took advantage of the unique opportunity to study the evolutionary genetics of a single, endemic and extensively characterized genotype in China over a nearly 20-year timespan. During this time, the incidence of measles in China declined steadily due to expanding vaccination programmes.

## RESULTS

### Phylogeny of N, F, and H genes of genotype H1 MeVs from China

Among the 63 new viruses analysed, 33 had been collected since 2010 (Fig. 1). Phylogenetic analysis of the 450 nt sequence coding for the C-terminal 150 amino acids of N (N-450) showed that all new sequences are clustered within genotype H1, cluster 1. Phylogenetic analyses of the H and F genes revealed a similar tree topology, with cluster 1 containing the most recent sequences (Fig. 1). Interestingly, the H sequences of cluster 1 could be further divided into two subclusters, each supported by significant bootstrap values (>80%). Viruses collected from the six different Chinese regions were classified into both subgroups, indicating no geographical restriction of subgroups (Fig. 1). Identification of positively selected sites within predicted protein sequences for H and F with fixed-effects likelihood (FEL) and internal FEL (IFEL) did not support the presence of positive selection for either (data not shown), consistent with previous observations (Rima *et al.*, 1997).

Two hundred and sixty-three sequences from 178 viruses of genotype H1 MeVs collected in all six regions of mainland China (Fig. S1, available in the online Supplementary Material) and 30 of 31 provinces were analysed. In addition, sequence data for all three genes were available from most provinces (Table S1). As expected, N-450 sequences were widely available, since N sequences are used for routine genotype identification. The oldest genotype H1 sequence was from a virus isolated in 1993 (MVi/Shandong.CHN/93/

1-AF045207). Since then, viruses have been collected each year (Table S1). Although the dataset contained sequences collected over two decades, only a few sequences were available from 1993 to 1998. The number of available sequences increased in 2005 as a result of increased emphasis on virologic surveillance in China. To accurately estimate the time of most recent common ancestor (TMRCA) for genotype H1, the dataset needed to incorporate sequences of viruses collected before the initial detection of genotype H1. Therefore, sequences of the WHO reference strains collected between 1954 and 1993 were included in the final dataset (Table S2).

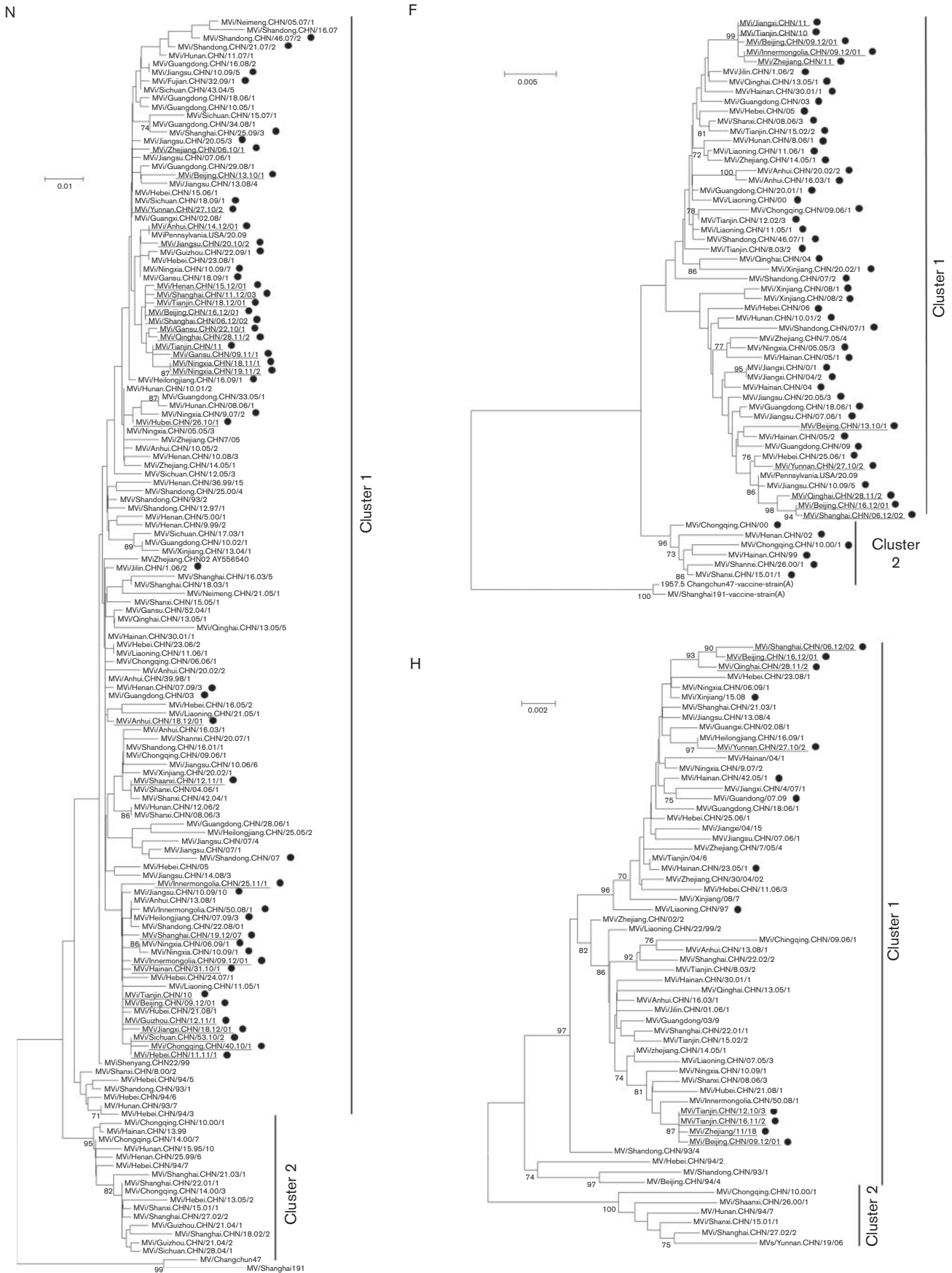
### Estimation of TMRCA for genotype H1 MeVs

Sequences from the N-450, H, and F genes were analysed. The H gene dataset appeared the most informative, since the entire 1854 nt coding region was included. The F sequences also included the entire coding region, but temporal distribution of the F dataset was limited compared to that of H. Accordingly, we performed the analysis on H sequences and used data from N and F genes to validate these results.

A Bayesian coalescent method was used for H sequence analysis to estimate TMRCA of the H1 genotype, as well as clusters 1 and 2 within the H1 genotype. The TMRCA of the H1 genotype (Table 1, Fig. 2) was estimated as approximately 1985 [95% HPD (Highest Probability Density), 1979–1989]. The oldest genotype H1 virus analysed was collected in 1993. TMRCA suggests that genotype H1 viruses had been circulating in China at least 10 years before identification with virologic surveillance. Cluster 1 diverged at around 1988 (95% HPD, 1984–1991) whereas cluster 2 appeared a few years later (95% HPD, 1987–1994). Based on currently available data, cluster 2 appears to be inactive and cluster 1 viruses are widely distributed throughout China. Comparable date estimates were found with the N-450 and F datasets (Table 1).

BEAST software was used to calculate nucleotide substitution rates for the N-450, F and H genes. The mean substitution rate for the H gene was estimated at  $0.75 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (95% HPD,  $0.51 \times 10^{-3}$ – $0.92 \times 10^{-3}$ ). A similar range was estimated for the F gene with 95% HPD of  $0.54 \times 10^{-3}$ – $1.75 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup>. Notably, the substitution rate for N-450 was approximately twice as high with 95% HPD of  $1.19 \times 10^{-3}$ – $2.11 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (Table 1).

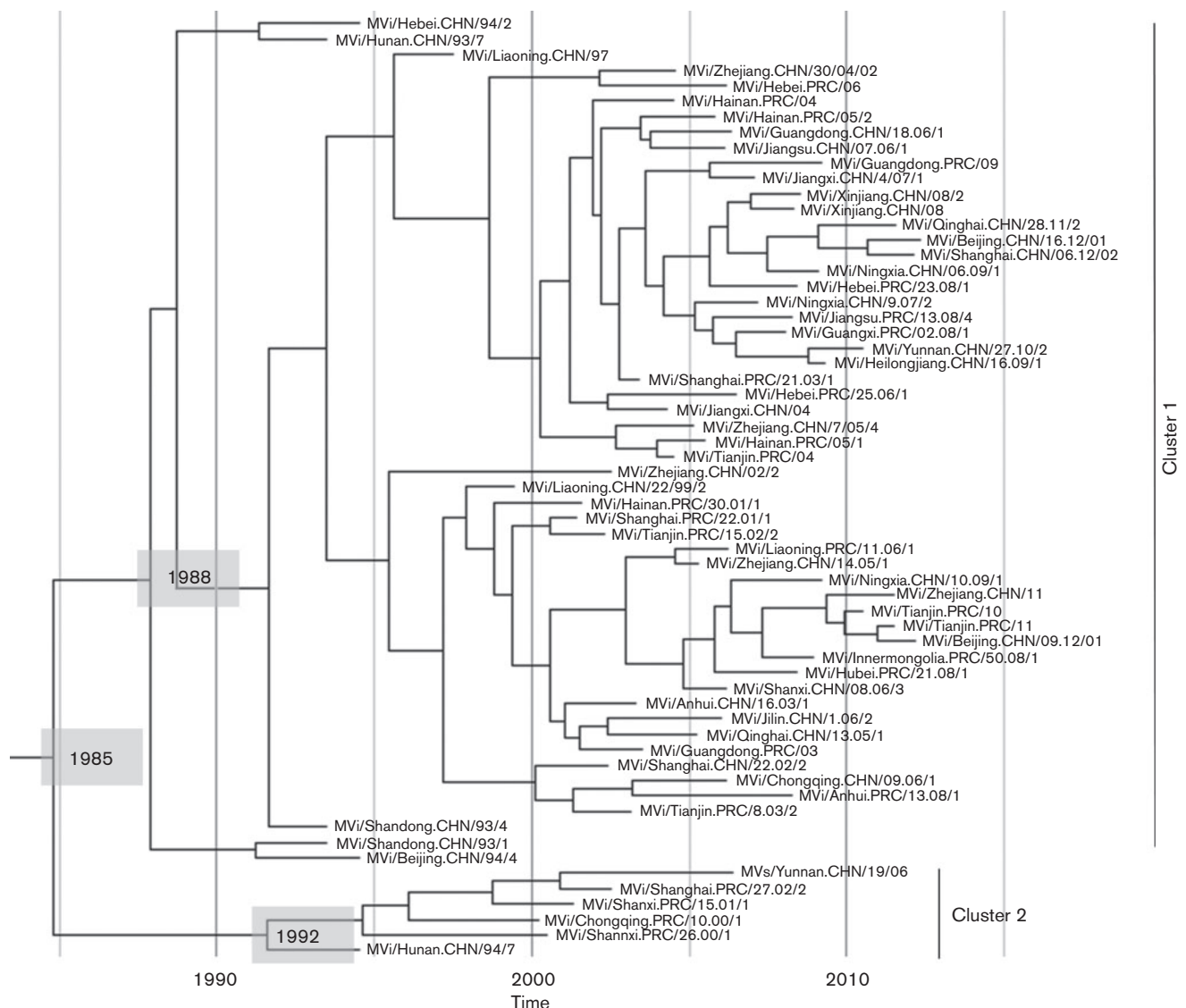
A Bayesian skyline plot was generated for the H gene dataset in order to analyse the genetic diversity of Chinese sequences over time (Fig. 3a). The graph has three distinct phases. From 1993 to 2000, minimal change in genetic diversity was evident, probably due to the limited number of viral sequences available. From 2000 to 2005, genetic diversity increased, consistent with the expansion of virologic surveillance in China during this time. Finally, from 2005 to 2012, genetic diversity steadily decreased, even though the number of sequences analysed was the



**Fig. 1.** Phylogeny of N, F and H genes of genotype H1. Neighbour-joining phylogenetic trees were reconstructed using MEGA5. Sequences generated in this study are indicated with black circles. The names of viruses collected since 2010 are underlined. Groups corresponding to clusters 1 and 2 and bootstrap values greater than 70 % are indicated. Phylogenetic trees for N and F genes were rooted to genotype A Chinese vaccine strains Changchun47 and Shanghai191. GenBank ID numbers for the sequences used in the N, F and H datasets are presented in Table S1.

same or greater than that for previous years (Figs S2a, 3a). The decrease in genetic diversity coincided with a significant reduction in the number of reported measles cases in China. A similar analysis was performed on H sequences of genotype B3 viruses of African origin (Fig. 3b, Table S3). The B3 skyline plot disclosed an increase in

diversity from 1980 to 1992, possibly due to increased surveillance. From 1992 to 2012, the graph showed no change in diversity over time. Thus, it appears that the decrease in genetic diversity of genotype H1 reflects the decline in reported cases due to effective vaccination campaigns.



**Fig. 2.** Maximum Clade Credibility (MCC) tree of H gene sequences of genotype H1 viruses from China. TMRCA of interest are shown (see Table 1). Cluster 1 and cluster 2 are indicated. The complete MCC tree, featuring 95% HPDs node bars and posterior probabilities for the entire dataset is shown in Figure S3.

**Table 1.** Nucleotide substitution rates and TMRCA for N-450, F and H genes of genotype H1 viruses from China

|                    |           | MeV gene          |                  |                  |
|--------------------|-----------|-------------------|------------------|------------------|
|                    |           | H                 | F                | N                |
| Substitution rate* |           | 0.75 (0.51–0.92)† | 1.07 (0.54–1.75) | 1.65 (1.19–2.11) |
| TMRCA              | H1        | 1985 (1979–1989)  | 1989 (1978–1996) | 1987 (1982–1994) |
|                    | Cluster 1 | 1988 (1984–1991)  | 1994 (1987–1998) | 1989 (1985–1992) |
|                    | Cluster 2 | 1992 (1987–1994)  | 1994 (1987–1998) | 1992 (1983–1994) |

\*Expressed as  $10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup>.

†Mean value is indicated first, followed by 95% HPD in parentheses.

## DISCUSSION

In this investigation, we had the unique opportunity to study the evolutionary genetics of a single, endemic and well characterized MeV genotype (H1) in China over a nearly 20-year timespan. Virologic surveillance began in China in 1993, and expanded significantly after 2005 to meet national programme goals. Currently, virologic surveillance is conducted by the National Measles Laboratory at the China Centers for Disease Control and Prevention, which oversees a network of more than 300 laboratories. From 2005 to 2012, the incidence of measles in China has decreased steadily due to expanding vaccination programmes.

Analysis of the H genes for genotype H1 viruses led to estimation of TMRCA for genotype H1 as around 1985, with divergence of two clusters (cluster 1 in 1987 and cluster 2 in 1991). Viral strains in cluster 2 have not been detected since 2005, and this lineage is presumably inactive, which is not an unusual occurrence for measles genotypes. Among the 24 measles genotypes recognized by WHO, only eight have been detected through global surveillance in the past five years (WHO, 2012b). Viruses in cluster 1 have been prevalent in mainland China since at least 1987. Our TMRCA estimates for the N and F genes were in close agreement with that for the H gene, even though the N and F datasets were limited by the short sequence length for N and incomplete representation for F.

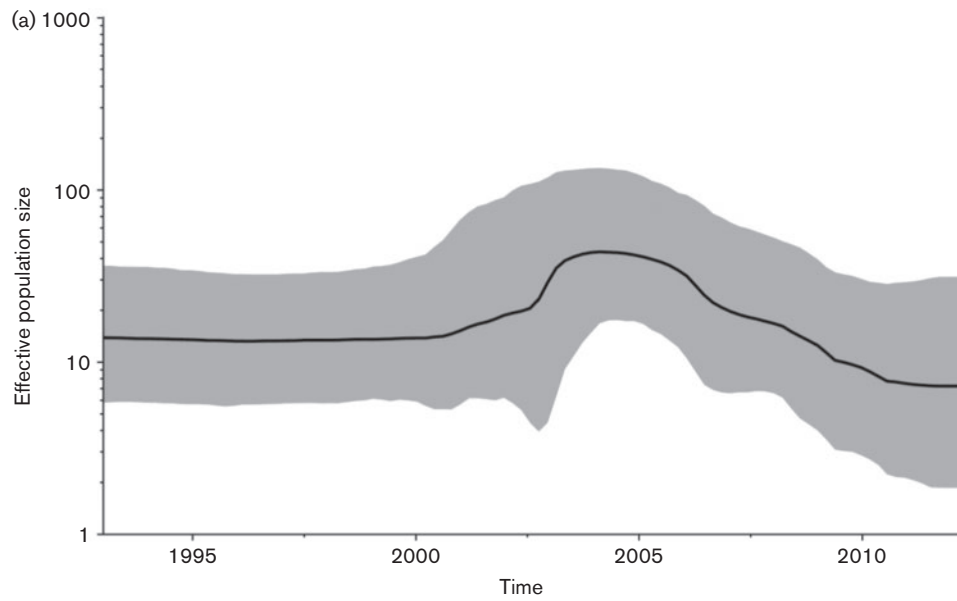
The main focus of the current study was to analyse the evolution of a single, endemic genotype in a country with a strong surveillance system, with a view to limiting environmental effects and minimizing surveillance bias. Our analysis estimated appearance of the H1 genotype at around the mid-1980s, consistent with previous data on Asian genotypes D3, D5, D9 and H1 (Saitoh *et al.*, 2012) showing that H1 diverged in 1977 (95% confidence interval, 1968 and 1984). Analysis of a limited number of Chinese N gene sequences led to TMRCA estimation for genotype H1 of around 1975 (Wei *et al.*, 2012). Previous studies also confirmed that the most reliable estimates of TMRCA are provided by analysis of the H gene, with N sequences giving comparable estimates (Pomeroy *et al.*, 2008; Saitoh *et al.*, 2012).

Studies on the molecular evolution of MeV are limited, since relatively few MeVs from the pre-vaccine era are

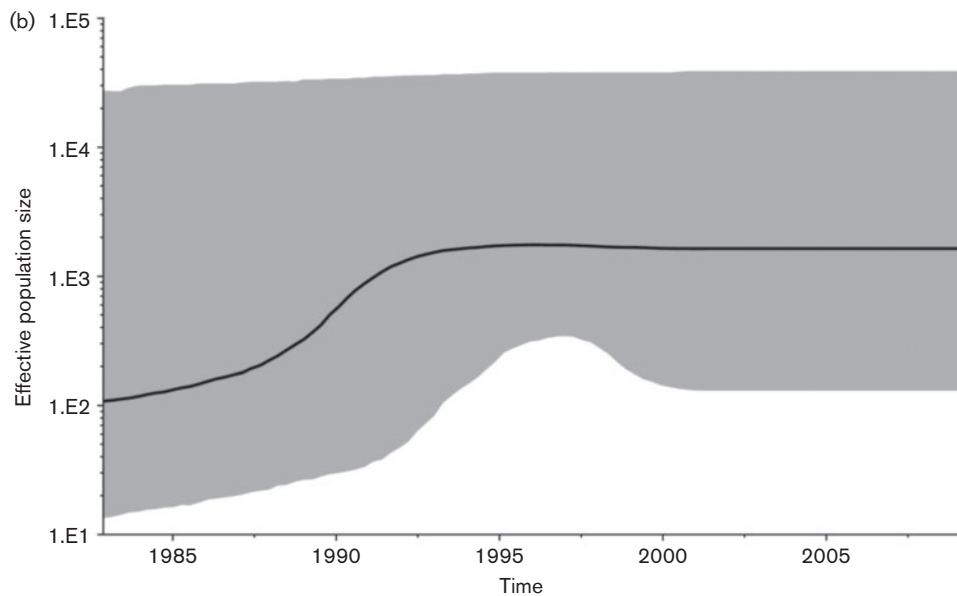
available for genotyping. The majority of these older sequences are genotype A, which may have been distributed globally in the 1950s, 1960s and 1970s. The Chinese vaccine strain, S191, a member of genotype A, was isolated from a child in 1960 in Shanghai. Routine isolation and genotyping of wild-type MeVs did not begin on a global scale until the mid-to-late 1990s following the global measles resurgence in the late 1980s, and very few wild-type isolates are available from any part of the world before 1990.

RNA viruses, such as MeV, have the ability to undergo rapid genetic change (Domingo & Holland, 1997). These viruses evolve at an approximate rate of  $10^{-3}$  nucleotide substitutions site<sup>-1</sup> year<sup>-1</sup> (Duffy *et al.*, 2008). For example, the capsid gene of GII.3 norovirus is reported to evolve at  $4.2 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (Mahar *et al.*, 2013), the HN gene of human parainfluenza virus type 1 at  $0.77 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (Mizuta *et al.*, 2011), and the G glycoprotein gene of human respiratory syncytial virus at  $4.68 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (Pretorius *et al.*, 2013). In this report, we calculated the nucleotide substitution rates for genotype H1 MeVs in China. The substitution rate calculated for the H gene of genotype H1 ( $0.75 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup>) was slightly higher than that of  $0.65 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> reported by Pomeroy *et al.* (2008). A more recent study determined the substitution rate for the H gene of genotypes D3, D8, D9 and H1 as  $0.56 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> (Saitoh *et al.*, 2012), which was similar to  $\sim 0.4 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> calculated for genotype D3 viruses circulating in Europe in the early 1990s and  $0.34 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> for sequences from cases of subacute sclerosing panencephalitis (Rima *et al.*, 1997; Woelk *et al.*, 2002). Here, we estimated a rate of  $1.65 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> for N-450, consistent with  $1.12 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup> calculated based on a dataset containing multiple genotypes, including H1 (Wei *et al.*, 2012). The substitution rate for N-450 was slightly higher than that of  $0.86 \times 10^{-3}$  substitutions site<sup>-1</sup> year<sup>-1</sup>, which was calculated based on sequences for the entire coding region of the N gene (Pomeroy *et al.*, 2008).

Skyline plots showed that genetic diversity of genotype H1 MeVs in China decreased after 2005, while that of H genes of genotype B3 viruses from Africa remained constant. The



| Year            | 1995  | 1999  | 2000  | 2001  | 2002  | 2003  | 2004  | 2005   | 2006   | 2007   | 2008   | 2009  | 2010  | 2011  | 2012 |
|-----------------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|-------|-------|-------|------|
| Number of cases | 57281 | 61840 | 73567 | 88962 | 58341 | 74813 | 71013 | 124219 | 100267 | 109827 | 132602 | 52854 | 38448 | 10204 | 5993 |



**Fig. 3.** Bayesian skyline plots showing changes in the genetic diversity of H genes from genotypes H1 and B3 over time. Panel (a) shows sequence diversity for the H gene of genotype H1 viruses from China isolated between 1993 and 2012. The inset table shows the total number of measles cases reported in China over selected years. Panel (b) shows the diversity of H gene sequences for genotype B3 viruses detected in Africa. For both graphs, highlighted areas correspond to 95% HPD intervals.

decrease in genetic diversity in China coincided with a significant reduction in the number of reported measles cases. Thus, the decreased genetic diversity of genotype H1 possibly reflects the decline in measles due to effective vaccination campaigns. Of course, vaccination campaigns

are also ongoing in Africa, leading to a dramatic drop in measles-associated mortality. However, vaccination coverage in most African countries is still considerably lower than that in China, and genotype B3 viruses continue to circulate widely as the endemic genotype in most

sub-Saharan African countries (WHO, 2012b). Coalescent analyses were recently used to document the effectiveness of hepatitis B vaccine in the Netherlands (Hahné *et al.*, 2013) and antiviral therapy for hepatitis C virus in Egypt (Stadler *et al.*, 2013). The analytic approach described in the current report may provide a means for countries to use data from virologic surveillance for documenting progress towards measles elimination.

## METHODS

**Virus isolation and sequencing.** Previously uncharacterized MeVs were sequenced to generate 52 N, 13 H, and 55 F sequences. B95a cells before June 2004 and after July 2004 (Kobune *et al.*, 1990) and Vero/hSLAM cells (Ono *et al.*, 2001) were used to isolate MeV. Viruses were passaged no more than twice. Both cell lines express a receptor for wild-type measles virus, and therefore passage in these cells would not require amino acid substitution associated with adaptation. RNA extraction, amplification and sequencing of N and H genes were performed as described previously (Pond & Frost, 2005; Zhang *et al.*, 2007). The complete coding region (1662 nt) of the F gene was amplified with the sequence-specific primers, MeV-F-for: GAA TCC CAG AAT CAA GAC TCA TC and MeV-F-rev: CAT TTG TGT TTC AAG AGT TG, using the SuperScript<sup>TM</sup>III One-Step reverse transcription (RT)-PCR kit (Invitrogen). The sequencing primers used were MeV-F-seq1: CCT CTG CTC CAA GAA TGC CTC CGG, MeV-F-seq2: TCT GTA TAG TAT CTG AGC AAT TTG AG, and MeV-F-seq3: TGC ACA ATT GGC TAT TAG G. Metadata, including GenBank ID numbers, are available in Table S1.

**Sequence datasets used for analysis.** Three sequence datasets, one for each gene, were established. The N and H gene datasets were derived mostly from previously reported sequences and complemented with those of viruses collected after 2008 (for N) and 2009 (for H). In order to avoid oversampling due to large outbreaks, similar sequences (less than 1 % nucleotide divergence) from viruses collected at the same times and locations were deleted. To obtain a meaningful estimation of the TMRCA for genotype H1, sequences of viruses collected before the estimated appearance of genotype H1 (1993) were added to the dataset. The F gene sequence dataset was more straightforward, as all available F gene sequences were included. Metadata for all three datasets are shown in Tables S1 and S2. The final datasets contained 170 N (N-450) sequences (144 genotype H1 and 26 from non-H1 WHO reference strains), 86 H sequences (60 genotype H1 and 26 from non-H1 WHO reference strains), and 69 F sequences (57 genotype H1 and 12 from non-H1 WHO reference strains).

**Sequence analysis and phylogeny inference.** Sequence alignments were generated with MUSCLE (Edgar, 2004) and edited with BioEdit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Phylogenies were inferred with MEGA software (version 5) using the neighbour-joining method and a maximum composite likelihood model (Tamura *et al.*, 2011; Saitou & Nei, 1987). The reliability of phylogenetic inference at each branch node was estimated by the bootstrap method with 1000 replications using MEGA (Efron & Tibshirani, 1994). TMRCA was estimated using a coalescent-based Bayesian method implemented in BEAST (version 1.7.2) (Drummond *et al.*, 2012). The nucleotide substitution model was tested in jModelTest0.1 using Akaike information criterion (AIC) (Guindon & Gascuel, 2003; Posada, 2008). The best-fit model of nucleotide substitution for the three datasets was determined as GTR+I+G [general time reversible model (GTR); proportion of invariable sites (I); gamma distribution shape parameter (G)]. Initially, three molecular clock models were tested on each dataset: strict, uncorrelated exponential relaxed and uncorrelated lognormal

relaxed clock. The Bayes factor test determined that the uncorrelated exponential relaxed clock (UCER) with eight rate categories was the best fit for each dataset (Drummond *et al.*, 2012). Therefore, only results using this model are shown. Three independent analyses were run for each dataset and combined using LogCombiner (Drummond *et al.*, 2012). Sixty to ninety million generations were run, and log and tree files saved to generate 10 000 trees, ensuring sufficient sample size ( $n > 200$ ) for each parameter of interest. Maximum clade credibility (MCC)-dated phylogenetic trees were summarized after the removal of 10 % sample burn-in with TreeAnnotator, and edited in FigTree version 1.3.1 (Drummond *et al.*, 2012; Rambaut & Drummond, 2010). Bayesian skyline plots depicting relative viral genetic diversity through time were generated for the H gene in TRACER (Drummond *et al.*, 2005). As a control, 63 H sequences of genotype B3 genotype were analysed via skyline plot (Table S3).

**Selective pressure analysis.** To evaluate selective pressure on the H and F coding regions among the genotype H1 MeV strains, the rates of synonymous (dS) and non-synonymous (dN) substitutions were estimated with both FEL and IFEL methods using maximum-likelihood approaches available on the Datamonkey webserver (<http://www.datamonkey.org/>) (Pond & Frost, 2005). Positive (dN > dS) selections were predicted at a significance level (*P*-value) of <0.05.

## ACKNOWLEDGEMENTS

We thank all the provincial and prefecture measles laboratory staff and epidemiologists in mainland China for providing clinical specimens, isolates and epidemiological data. We are grateful to WHO headquarters, WPRO and US CDC for technical and financial support. We additionally acknowledge anonymous reviewers for comments that improved the manuscript. This work is supported by a grant from the National Natural Science Foundation of China (grant no. 81371791), the Key Technologies R&D Program of National Ministry of Science (2013ZX10004-202, 2012ZX10004215, 2013ZX10004-101, 2012YQ03026104, 2012ZX10004201-003, 2012ZX10004401), WHO Measles Regional Reference Laboratory funding and WHO funding (no. WP CHN AAA 011 XZ 08).

## REFERENCES

- Anonymous (2013).** Global Vaccine Action Plan: decade of vaccines collaboration. *Vaccine* **31** (Suppl 2), B5–B31.
- Domingo, E. & Holland, J. J. (1997).** RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151–178.
- Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. (2005).** Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185–1192.
- Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012).** Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973.
- Duffy, S., Shackleton, L. A. & Holmes, E. C. (2008).** Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**, 267–276.
- Edgar, R. C. (2004).** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- Efron, B. & Tibshirani, R. J. (1994).** *An Introduction to the Bootstrap*. Boca Raton: CRC Press.
- Griffin, D. E. B. & Bellini, W. J. (1996).** Measles viruses. In *Fields Virology*, 3rd edn, pp. 1267–1296. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott-Raven.

- Guindon, S. & Gascuel, O. (2003).** A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Hahné, S., van Houdt, R., Koedijk, F., van Ballegooijen, M., Cremer, J., Bruisten, S., Coutinho, R. & Boot, H. (2013).** Selective hepatitis B virus vaccination has reduced hepatitis B virus transmission in the Netherlands. *PLoS ONE* **8**, e67866.
- Ji, Y., Zhang, Y., Xu, S., Zhu, Z., Zuo, S., Jiang, X., Lu, P., Wang, C., Liang, Y. & other authors (2009).** Measles resurgence associated with continued circulation of genotype H1 viruses in China, 2005. *Virology* **6**, 135.
- Ji, Y., Xu, S., Zhang, Y., Zhu, Z., Mao, N., Jiang, X., Ma, C., Lu, P., Wang, C. & other authors (2010).** Genetic characterization of wild-type measles viruses isolated in China, 2006–2007. *Virology* **7**, 105.
- Kobune, F., Sakata, H. & Sugiura, A. (1990).** Marmoset lymphoblastoid cells as a sensitive host for isolation of measles virus. *J Virol* **64**, 700–705.
- Mahar, J. E., Bok, K., Green, K. Y. & Kirkwood, C. D. (2013).** The importance of intergenic recombination in norovirus GI.3 evolution. *J Virol* **87**, 3687–3698.
- Mizuta, K., Saitoh, M., Kobayashi, M., Tsukagoshi, H., Aoki, Y., Ikeda, T., Abiko, C., Katsushima, N., Itagaki, T. & other authors (2011).** Detailed genetic analysis of hemagglutinin-neuraminidase glycoprotein gene in human parainfluenza virus type 1 isolates from patients with acute respiratory infection between 2002 and 2009 in Yamagata prefecture, Japan. *Virology* **8**, 533.
- Ono, N., Tatsuo, H., Hidaka, Y., Aoki, T., Minagawa, H. & Yanagi, Y. (2001).** Measles viruses on throat swabs from measles patients use signaling lymphocytic activation molecule (CDw150) but not CD46 as a cellular receptor. *J Virol* **75**, 4399–4401.
- Perry, R. T., Gacic-Dobo, M., Dabbagh, A., Mulders, M. N., Strebel, P. M., Okwo-Bele, J. M., Rota, P. A., Goodson, J. L. & Centers for Disease Control and Prevention (CDC) (2014).** Global control and regional elimination of measles, 2014. *MMWR Morb Mortal Wkly Rep* **63**, 103–107.
- Pomeroy, L. W., Bjørnstad, O. N. & Holmes, E. C. (2008).** The evolutionary and epidemiological dynamics of the paramyxoviridae. *J Mol Evol* **66**, 98–106.
- Pond, S. L. & Frost, S. D. (2005).** Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533.
- Posada, D. (2008).** jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–1256.
- Pretorius, M. A., van Niekerk, S., Tempia, S., Moyes, J., Cohen, C., Madhi, S. A., Venter, M. & SARI Surveillance Group (2013).** Replacement and positive evolution of subtype A and B respiratory syncytial virus G-protein genotypes from 1997–2012 in South Africa. *J Infect Dis* **208** (Suppl 3), S227–S237.
- Rambaut, A. & Drummond, A. (2010).** FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Rima, B. K., Earle, J. A., Bacsko, K., ter Meulen, V., Liebert, U. G., Carstens, C., Carabaña, J., Caballero, M., Celma, M. L. & Fernandez-Muñoz, R. (1997).** Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. *J Gen Virol* **78**, 97–106.
- Rota, P. A., Brown, K., Mankertz, A., Santibanez, S., Shulga, S., Muller, C. P., Hübschen, J. M., Siqueira, M., Beirnes, J. & other authors (2011).** Global distribution of measles genotypes and measles molecular epidemiology. *J Infect Dis* **204** (Suppl 1), S514–S523.
- Saitoh, M., Takeda, M., Gotoh, K., Takeuchi, F., Sekizuka, T., Kuroda, M., Mizuta, K., Ryo, A., Tanaka, R. & other authors (2012).** Molecular evolution of hemagglutinin (H) gene in measles virus genotypes D3, D5, D9, and H1. *PLoS ONE* **7**, e50660.
- Saitou, N. & Nei, M. (1987).** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. (2013).** Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A* **110**, 228–233.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011).** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739.
- Wei, C., Shi, J., Liu, B., Shi, Y., Zheng, J., Xu, G., Ma, J., Wang, G. & Li, F. (2012).** Molecular characterization of the measles virus genotypes in JiLin Province, China. *PLoS ONE* **7**, e46011.
- WHO (2012a).** The Measles & Rubella Initiative Welcomes World Health Assembly Commitment to Measles and Rubella Elimination Goals.
- WHO (2012b).** Measles virus nomenclature update: 2012. *Wkly Epidemiol Rec* **87**, 73–81.
- Woelk, C. H., Pybus, O. G., Jin, L., Brown, D. W. & Holmes, E. C. (2002).** Increased positive selection pressure in persistent (SSPE) versus acute measles virus infections. *J Gen Virol* **83**, 1419–1430.
- Xu, W., Tamin, A., Rota, J. S., Zhang, L., Bellini, W. J. & Rota, P. A. (1998).** New genetic group of measles virus isolated in the People's Republic of China. *Virus Res* **54**, 147–156.
- Zhang, Y., Zhu, Z., Rota, P. A., Jiang, X., Hu, J., Wang, J., Tang, W., Zhang, Z., Li, C. & other authors (2007).** Molecular epidemiology of measles viruses in China, 1995–2003. *Virology* **4**, 14.
- Zhang, Y., Ji, Y., Jiang, X., Xu, S., Zhu, Z., Zheng, L., He, J., Ling, H., Wang, Y. & other authors (2008).** Genetic characterization of measles viruses in China, 2004. *Virology* **5**, 120.
- Zhang, Y., Zhou, J., Bellini, W. J., Xu, W. & Rota, P. A. (2009).** Genetic characterization of Chinese measles vaccines by analysis of complete genomic sequences. *J Med Virol* **81**, 1477–1483.