Measles virus and rinderpest virus divergence dated to the rise of large cities

Authors: Ariane Düx[1,2]†, Sebastian Lequime[3]†, Livia Victoria Patrono[1,2], Bram Vrancken[3], Sengül Boral[4], Jan F. Gogarten[1,2], Antonia Hilbig[1], David Horst[4], Kevin Merkel[1,2], Baptiste Prepoint[2], Sabine Santibanez[5], Jasmin Schlotterbeck[2], Marc A. Suchard[6], Markus Ulrich[1], Navena Widulin[7], Annette Mankertz[5], Fabian H. Leendertz[1], Kyle Harper[8], Thomas Schnalke[7], Philippe Lemey[3]‡, Sébastien Calvignac-Spencer[1,2]‡*

Affiliations:

[1] Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany

[2] Viral Evolution, Robert Koch Institute, Berlin, Germany

[3] KU Leuven Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory of Clinical and Evolutionary Virology, Leuven, Belgium

[4] Institute for Pathology, Charité, Berlin, Germany

[5] National Reference Centre Measles, Mumps, Rubella, Robert Koch Institute, Berlin, Germany

[6] Department of Biostatistics, Fielding School of Public Health and Departments of Biomathematics and of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

[7] Berlin Museum of Medical History at the Charité, Berlin, Germany

[8] Department of Classics and Letters, The University of Oklahoma, Norman, OK, USA

*Correspondence to: CalvignacS@rki.de

† Equal contributions.

‡ Equal contributions.

Abstract:

Many infectious diseases are thought to have emerged in humans after the Neolithic revolution. While it is broadly accepted that this also applies to measles, the exact date of emergence for this disease is controversial. Here, we sequenced the genome of a 1912 measles virus and used selection-aware molecular clock modeling to determine the divergence date of measles virus and rinderpest virus. This divergence date represents the earliest possible date for the establishment of measles in human populations. Our analyses show that the measles virus potentially arose as early as the 6th century BCE, possibly coinciding with the rise of large cities.

One Sentence Summary:

Measles virus diverged from rinderpest virus in the 6th century BCE, which is compatible with an ancient emergence of measles.

Main Text:

Measles is a highly contagious viral disease that presents with rash, fever and respiratory symptoms. Before a live-attenuated vaccine was developed in the 1960s, the disease affected the vast majority of children (*1, 2*). Global vaccination campaigns resulted in a marked reduction of measles transmission and fatal cases and WHO has proclaimed an elimination goal. However, the disease still caused an estimated 110,000 deaths in 2017 (*3*) and incidence has recently been on the rise (*4*). Measles is caused by *Measles morbillivirus* (MeV), a negative sense single-stranded RNA virus from the family *Paramyxoviridae* (Order: *Mononegavirales*). MeV is an exclusively human pathogen whose closest relative was the now eradicated *Rinderpest morbillivirus* (RPV), a devastating cattle pathogen (*5*). It is generally accepted that measles emergence resulted from a

spill-over from cattle to humans, although the directionality of this cross-species transmission event has never been formally established (supplementary text S1; *6*).

It is unclear when measles first became endemic in human populations, but assuming an origin in cattle, the earliest possible date of MeV emergence is defined by the MeV-RPV divergence time. Several studies have provided estimates for this date using molecular clock analyses (*7-10*), with the most reliable (and oldest) estimate falling at the end of the 9th century CE (mean: 899 CE [95% highest posterior density (HPD) interval: 597 – 1144 CE]; (*8*). Here, we reassess the MeV-RPV divergence time using advanced, selection-aware Bayesian molecular clock modelling (*11*) on a dataset of heterochronous MeV genomes including the oldest human RNA virus genome sequenced to date, and show that a considerably earlier emergence can no longer be excluded.

Our re-examination was prompted by the broadly accepted view that molecular dating based on tip date calibration, i.e. the method used in previous efforts to estimate the timing of MeV-RPV divergence, underestimates deep divergence times (*8*). Rapid short-term substitution rates captured by tip calibration can often not be applied over long evolutionary timescales, because of the effects of long-term purifying selection and substitution saturation. This causes a discrepancy between short- and long-term substitution rates, which is referred to as the time-dependent rate phenomenon (*12, 13*). Since measurement timescales matter, a first step to arrive at accurate estimates is to maximize the time depth of tip calibration, for example through the use of ancient viral sequences (*14, 15*).

RNA tends to be much less stable in the environment than DNA, making the recovery of MeV genetic material from archeological remains unlikely (*16*). Pathology collections represent a more realistic source of MeV sequences that predate the oldest MeV genome – the genome of the Edmonston strain that was isolated in 1954 and attenuated to become the first measles vaccine. We examined a collection of lung specimens gathered by Rudolf Virchow and his successors

between the 1870s and 1930s and preserved by the Berlin Museum of Medical History at the Charité (Berlin, Germany), and identified a 1912 case diagnosed with fatal measles-related bronchopneumonia (Fig. 1, fig. S1, supplementary texts S2 and S3). To retrieve MeV genetic material from this specimen, we first heat-treated 200mg of the formalin-fixed lung tissue to reverse macromolecule cross-links induced by formalin and subsequently performed nucleic acid extraction (*17*). Following DNase treatment and ribosomal RNA depletion, we built high-throughput sequencing libraries and shotgun sequenced them on Illumina® platforms. We generated 27,328,219 high quality reads, of which 0.46% were mapped to a MeV genome. Median insert size varied between 95 and 136 nucleotides and little damage was observed, suggesting good preservation of RNA molecules (fig. S2, table S1, supplementary text S4). The resulting 10,960 unique MeV reads allowed us to reconstruct an almost complete 1912 MeV genome: 15,257 of the 15,894 nucleotides in the MeV strain Edmonston (AF266288) were covered by at least 3 unique reads (11,988 nucleotides by at least 20 reads; mean coverage 54x).

In addition to the 1912 genome and the 1954 Edmonston genome, only 2 genomes have been determined from MeV isolated prior to 1990 (Mvi/Lyon.FRA/77: HM562899; T11wild: AB481087). We therefore searched the strain collection of the German National Reference Laboratory (Robert Koch Institute, Berlin, Germany) for pre-1990 isolates. We found two strains from the pre-vaccine era isolated in 1960 by the National Reference Laboratory of former Czechoslovakia in Prague (MVi/Prague.CZE/60/1 and MVi/Prague.CZE/60/2; *18*). We performed serial passages of these strains and determined their genome sequences at a mean coverage of 109x and 70x, respectively. The two genomes were nearly identical, differing at only four sites.

We performed Bayesian and maximum likelihood (ML) phylogenetic analyses to investigate the phylogenetic placement of the 1912 and 1960 genomes with respect to 127 available MeV

genomes. Tip-dated Bayesian phylogenetic trees placed the 1912 genome as a sister lineage to all modern genomes while the two genomes from 1960 clustered together with the Edmonston strain (genotype A; fig. S3). The placement of the 1912 genome in the dated-tip tree was consistent with its placement in a non-clock ML tree reconstruction and with the rooting of a dated-tip tree excluding the 1912 genome (fig. S4 A and B). The relatedness of the 1912 and 1960 genomes to now extinct MeV lineages is in line with a marked reduction of MeV genetic diversity during the 20th century as a product of massive vaccination efforts.

Having extended the time depth of MeV tip calibration, we subsequently focused our attention on estimating the timing of MeV-RPV divergence. We assembled a dataset of 51 genomes comprising MeV (including one of the 1960 genomes and the 1912 genome), RPV and Peste des petits ruminants virus (PPRV, the closest relative to MeV-RPV) sequences, ensuring they represented the known genetic diversity of these viruses (table S2). Prior to inferring a time-scaled evolutionary history for this dataset using a Bayesian phylogenetic framework, we assessed its temporal signal and tested it for substitution saturation. We confirmed a strong temporal signal (fig. S5, table S3) and did not identify strong substitution saturation (table S4). We constructed a series of increasingly complex evolutionary models to accommodate various sources of rate heterogeneity. Models ranged from a standard codon substitution model with a strict molecular clock assumption to a codon substitution model with time-varying selection combined with a clade-specific rate for PPRV and additional branch-specific random effects on the substitution rate. Adequately accommodating different sources of rate heterogeneity is known to provide a better correction for multiple hits in genetic distance estimation and the potential of codon substitution modelling in recovering deep viral divergence has specifically been demonstrated (*8*). This was reflected in the increasingly older estimates of MeV-RPV and PPRV-MeV-RPV divergence times and wider credible intervals for increasingly complex models (Fig. 2

and table S5). Parameter estimates of the substitution and clock models also provided evidence for a significant contribution of these different sources of rate heterogeneity to model fit improvement (table S5). We found a significantly negative coefficient for the time-dependent nonsynonymous/synonymous substitution rate ratio (ω) (*11*), indicating strong long-term purifying selection, a significantly positive coefficient for the fixed effect on the PPRV rate, indicating a faster evolutionary rate in this clade (as suggested by temporal signal analyses, fig. S5), and significant additional unexplained variation as modelled by the random effects (table S5). Our most complex model therefore provided the best description of the evolutionary process and significantly pushed back the divergence date of MeV and RPV, with a mean estimate at 528 BCE [95% HPD interval: 1174 BCE - 165 CE] (Fig. 3). These estimates were robust to (i) including or excluding the 1912 genome in the analyses (table S6), (ii) using a more conservative consensus genome for the 1912 sample (table S7), (iii) the prior specification on the age of the RPV genome or the inclusion of an additional RPV genome (table S7), and (iv) the coalescent prior specification (table. S7). A comparison of the four models in their ability to recover the age of the 1912 genome indicated that the most complex model yielded the best estimate (1929 CE [95% HPD interval: 1889 - 1961 CE]; fig. S6).

The MeV/RPV divergence time provides the earliest possible date for measles emergence in humans, which is now compatible with the emergence of this disease more than 2,500 years ago. It seems plausible that the divergence of these lineages was closely followed by the cattle-to-human host jump and subsequent evolution into two distinct pathogens. However, the spill-over could have occurred at any time between the MeV/RPV divergence and the time to the most recent common ancestor of all MeV known to infect humans (1880 CE [95% HPD interval: 1865 – 1893 CE]). This raises the question of whether other sources of information can narrow down this timeframe and agree with an earlier timing of measles emergence.

The earliest clear clinical description of measles is often attributed to the Persian physician Rhazes, writing in the 10th century CE (*19*). But Rhazes was extremely familiar with all available medical literature at his time, and made use of earlier sources. Indian medical texts possibly describe measles several centuries prior to Rhazes (*20*). While clear descriptions of measles are missing in the Hippocratic corpus and the Greek medical tradition (at least through the prolific second-century writer Galen), such absence alone cannot be decisive. Retrospective diagnosis from pre-modern medical texts is notoriously fraught, especially for diseases like measles whose symptoms were easily confused with a variety of other conditions. Measles differential diagnosis remained a challenge well into more recent times (*21*). Therefore, any number of the large-scale "pestilences" described in ancient sources from Europe or China could reflect MeV outbreaks.

An ancient origin of measles seems all the more plausible in the light of demographic changes that are compatible with our understanding of (contemporary) MeV epidemiology. Populations large enough to support continuous MeV transmission, i.e. larger than the MeV critical community size (CCS) of 250,000-500,000 individuals (*22-24*), could not exist in Neolithic, Bronze Age, and early Iron Age settlements, which lacked both economic and political means to allow such numbers. Even if connectivity between such settlements may have created a larger pool of susceptible individuals, given the speed with which measles epidemics occur, and the efficacy of acquired immunity, epidemiologists have held that MeV could not have become endemic in urban populations below the CCS (*25*). In the late first millennium BCE, technologies (both economic and political) crossed a threshold promoting an upsurge in population sizes in Eurasia and South and East Asia. Although considerable uncertainty exists around population size estimates derived from ancient documents (e.g. literary observations, travelers' reports, censuses, or references to the amount of food distributed in a city) or archaeological proxies (e.g. size of city walls, built-up area of settlement), there is broad agreement that a number of

settlements in North Africa, India, China, Europe, and the Near East began to surpass the CCS for MeV by around 300 BCE, presumably for the first time in human history (Fig. 3; *26*). From this period onward, there were consistently urban populations above the CCS for MeV.

Based on these considerations, our substantially older MeV/RPV divergence estimate provides grounds for sketching a new model of MeV's evolutionary history. Under this scenario, a bovine virus, the common ancestor of modern strains of RPV and MeV, circulated in large populations of cattle (and possibly wild ungulates) since its divergence from PPRV around the 4th millennium BCE (3199 BCE [95% HPD interval: 4632 - 1900 BCE]; Fig. 3). As a fast-evolving RNA virus, it may have produced variants that were able to cross the species barrier on several occasions, but small human populations could only serve as dead-end hosts. Then, almost as soon as contiguous settlements reached sufficient sizes to maintain the virus' continuous transmission (Fig. 3), it emerged as a human pathogen, the progenitor of modern-day MeV. It has been suggested that numerous concurrent human-bovine epidemics in the early medieval period (here 6th-10th centuries CE) were caused by an immediate ancestor of MeV and RPV that was pathogenic to both cattle and humans (*27*). The new RPV-MeV divergence date allows for the same inference to be made for earlier concurrent human-bovine mortality events well attested in e.g. Roman sources from the 5th century BCE on (*28*). During the following centuries, introduction of MeV into naive human populations and/or flare-ups of the disease might have caused some ancient epidemics whose etiology remains uncertain.

While our findings shed new light on the origin of measles, formally proving that the virus emerged soon after its divergence from RPV would require archeological genomic evidence. Most studies on ancient viruses have thus far focused on viruses with a double-stranded DNA genome (*14*, *29-32*). However, genetic material of parvovirus B19 was also detected in early Neolithic skeletal remains, despite the relatively unstable nature of its single-stranded DNA

genome (*15*). It remains to be determined if viral RNA recovery from such ancient specimens is feasible. Recently, RNA was extracted from the remains of a 14,300-year-old Pleistocene canid preserved in permafrost (*33*). While the majority of RNA fragments were extremely short (<30 nt), the authenticity of the sequences could be validated (*33*). Such advances highlight that it may not be completely impossible for ancient remains to still contain MeV RNA, especially if preserved under favorable circumstances, including natural mummification or preservation in cold environments (*16*). While awaiting such direct evidence, we believe that the proposed model of MeV evolution constitutes a compelling working hypothesis.

References and Notes:

1. A. D. Langmuir, Medical importance of measles. *American journal of diseases of children (1960)* **103**, 224-226 (1962).
2. W. J. Moss, Measles. *Lancet* **390**, 2490-2502 (2017).
3. A. Dabbagh, R. L. Laws, C. Steulet, L. Dumolard, M. N. Mulders, K. Kretsinger, J. P. Alexander, P. A. Rota, J. L. Goodson, Progress toward regional measles elimination—worldwide, 2000–2017. *Morbidity and Mortality Weekly Report* **67**, 1323 (2018).
4. WHO, Provisional data based on monthly data reported to WHO (Geneva) as of January 2020. *https://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/*.
5. P. Roeder, J. Mariner, R. Kock, Rinderpest: the veterinary perspective on eradication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120139 (2013).
6. N. D. Wolfe, C. P. Dunavan, J. Diamond, Origins of major human infectious diseases. *Nature* **447**, 279-283 (2007).
7. Y. Furuse, A. Suzuki, H. Oshitani, Origin of measles virus: divergence from rinderpest virus between the 11th and 12th centuries. *Virology journal* **7**, 52 (2010).
8. J. O. Wertheim, S. L. Kosakovsky Pond, Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol* **28**, 3355-3365 (2011).
9. M. Muniraju, M. Munir, A. R. Parthiban, A. C. Banyard, J. Bao, Z. Wang, C. Ayebazibwe, G. Ayelet, M. El Harrak, M. Mahapatra, G. Libeau, C. Batten, S. Parida, Molecular evolution of peste des petits ruminants virus. *Emerging infectious diseases* **20**, 2023-2033 (2014).
10. H. Kimura, M. Saitoh, M. Kobayashi, H. Ishii, T. Saraya, D. Kurai, H. Tsukagoshi, K. Shirabe, A. Nishina, K. Kozawa, M. Kuroda, F. Takeuchi, T. Sekizuka, H. Minakami, A. Ryo, M. Takeda, Molecular evolution of haemagglutinin (H) gene in measles virus. *Scientific reports* **5**, 11648 (2015).
11. J. V. Membrebe, M. A. Suchard, A. Rambaut, G. Baele, P. Lemey, Bayesian Inference of Evolutionary Histories under Time-Dependent Substitution Rates. *Mol Biol Evol* **36**, 1793-1803 (2019).
12. S. Y. Ho, S. Duchene, M. Molak, B. Shapiro, Time-dependent estimates of molecular evolutionary rates: evidence and causes. *Molecular ecology* **24**, 6007-6012 (2015).

13. P. Aiewsakun, A. Katzourakis, Time-Dependent Rate Phenomenon in Viruses. *Journal of virology* **90**, 7184-7195 (2016).

14. B. Mühlemann, T. C. Jones, P. B. Damgaard, M. E. Allentoft, I. Shevnina, A. Logvin, E. Usmanova, I. P. Panyushkina, B. Boldgiv, T. Bazartseren, K. Tashbaeva, V. Merz, N. Lau, V. Smrcka, D. Voyakin, E. Kitov, A. Epimakhov, D. Pokutta, M. Vicze, T. D. Price, V. Moiseyev, A. J. Hansen, L. Orlando, S. Rasmussen, M. Sikora, L. Vinner, A. Osterhaus, D. J. Smith, D. Glebe, R. A. M. Fouchier, C. Drosten, K. G. Sjogren, K. Kristiansen, E. Willerslev, Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418-423 (2018).

15. B. Mühlemann, A. Margaryan, P. B. Damgaard, M. E. Allentoft, L. Vinner, A. J. Hansen, A. Weber, V. I. Bazaliiskii, M. Molak, J. Arneborg, W. Bogdanowicz, C. Falys, M. Sablin, V. Smrcka, S. Sten, K. Tashbaeva, N. Lynnerup, M. Sikora, D. J. Smith, R. A. M. Fouchier, C. Drosten, K. G. Sjogren, K. Kristiansen, E. Willerslev, T. C. Jones, Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 7557-7562 (2018).

16. O. Smith, M.T. Gilbert, "Ancient RNA" in *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, C. Lindqvist, O. P. Rajora, Eds. (Springer, Cham, 2019).

17. M. T. Gilbert, T. Haselkorn, M. Bunce, J. J. Sanchez, S. B. Lucas, L. D. Jewell, E. Van Marck, M. Worobey, The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? *PloS one* **2**, e537 (2007).

18. S. Santibanez, A. Heider, E. Gerike, A. Agafonov, E. Schreier, Genotyping of measles virus isolates from central Europe and Russia. *Journal of medical virology* **58**, 313-320 (1999).

19. R. Kim-Farley, "Measles" in *The Cambridge World History of Human Disease*, K.F. Kiple, Ed. (Cambridge University Press, Cambridge, 1993).

20. K. R. L. Gupta (translator), *Madhava nidana: ayurvedic system of pathology* (Sri Satguru Publications, Delhi, 1987).

21. B. A. Cunha, Smallpox and measles: historical aspects and clinical differentiation. *Infectious disease clinics of North America* **18**, 79-100 (2004).

22. M. S. Bartlett, Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)* **120**, 48-70 (1957).

23. F. L. Black, Measles endemicity in insular populations: critical community size and its evolutionary implication. *Journal of theoretical biology* **11**, 207-211 (1966).

24. M. J. Keeling, B. T. Grenfell, Disease extinction and community size: modeling the persistence of measles. *Science* **275**, 65-67 (1997).

25. A. D. Cliff, P. Haggett, M. Smallman-Raynor, *Measles: an historical geography of a major human viral disease: from global expansion to local retreat, 1840-1990* (Blackwell, Oxford, 1993).

26. H. Inoue, A. Álvarez, E. N. Anderson, A. Owen, R. Álvarez, K. Lawrence, C. Chase-Dunn, Urban scale shifts since the bronze age: upsweeps, collapses, and semiperipheral development. *Social Science History* **39**, 175-200 (2015).

27. T. P. Newfield, Human–bovine plagues in the early middle ages. *Journal of Interdisciplinary History* **46**, 1-38 (2015).

28. C. A. Spinage, *Cattle Plague: A History* (Springer, Boston, 2003).

29. P. Biagini, C. Theves, P. Balaresque, A. Geraut, C. Cannet, C. Keyser, D. Nikolaeva, P. Gerard, S. Duchesne, L. Orlando, E. Willerslev, A. N. Alekseev, P. de Micco, B. Ludes, E. Crubezy, Variola virus in a 300-year-old Siberian mummy. *The New England journal of medicine* **367**, 2057-2059 (2012).

30. A. T. Duggan, M. F. Perdomo, D. Piombino-Mascali, S. Marciniak, D. Poinar, M. V. Emery, J. P. Buchmann, S. Duchene, R. Jankauskas, M. Humphreys, G. B. Golding, J. Southon, A. Devault, J. M. Rouillard, J. W. Sahl, O. Dutour, K. Hedman, A. Sajantila, G. L. Smith, E. C. Holmes, H. N. Poinar, 17th Century Variola Virus Reveals the Recent History of Smallpox. *Current biology : CB* **26**, 3407-3412 (2016).

31. Z. Patterson Ross, J. Klunk, G. Fornaciari, V. Giuffra, S. Duchene, A. T. Duggan, D. Poinar, M. W. Douglas, J.-S. Eden, E. C. Holmes, The paradox of HBV evolution as revealed from a 16 th century mummy. *PLoS pathogens* **14**, (2018).

32. B. Krause-Kyora, J. Susat, F. M. Key, D. Kuhnert, E. Bosse, A. Immel, C. Rinne, S. C. Kornell, D. Yepes, S. Franzenburg, H. O. Heyne, T. Meier, S. Losch, H. Meller, S. Friederich, N. Nicklisch, K. W. Alt, S. Schreiber, A. Tholey, A. Herbig, A. Nebel, J. Krause, Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife* **7**, (2018).

33. O. Smith, G. Dunshea, M. S. Sinding, S. Fedorov, M. Germonpre, H. Bocherens, M. T. P. Gilbert, Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. *PLoS biology* **17**, e3000166 (2019).

34. I. Morris, *The measure of civilization: how social development decides the fate of nations* (Princeton University Press, Princeton, 2013).

35. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

36. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27**, 824-834 (2017).

37. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

38. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).

39. G. Jäger, ClipAndMerge. *https://github.com/apeltzer/ClipAndMerge*.

40. Broad Institute, Picard. *http://broadinstitute.github.io/picard*.

41. A. Peltzer, DeDup. *https://github.com/apeltzer/DeDup*.

42. H. Jonsson, A. Ginolhac, M. Schubert, P. L. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684 (2013).

43. Geneious 11.1.5. *https://www.geneious.com*.

44. C. H. Woelk, O. G. Pybus, L. Jin, D. W. Brown, E. C. Holmes, Increased positive selection pressure in persistent (SSPE) versus acute measles virus infections. *Journal of General Virology* **83**, 1419-1430 (2002).

45. K. Fukai, K. Morioka, K. Sakamoto, K. Yoshida, Characterization of the complete genomic sequence of the rinderpest virus Fusan strain cattle type, which is the most classical isolate in Asia and comparison with its lapinized strain. *Virus genes* **43**, 249-253 (2011).

46. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).

47. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **1**, vev003 (2015).

48. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).

49. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**, vew007 (2016).

50. S. Duchene, T. Stadler, S. Y. Ho, D. A. Duchene, V. Dhanasekaran, G. Baele, Bayesian Evaluation of Temporal Signal in Measurably Evolving Populations. *bioRxiv*, 810697 (2019).

51. X. Xia, Z. Xie, M. Salemi, L. Chen, Y. Wang, An index of substitution saturation and its application. *Molecular phylogenetics and evolution* **26**, 1-7 (2003).

52. X. Xia, P. Lemey, "Assessing substitution saturation with DAMBE" in *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*, P. Lemey, M. Salemi, A.M. Vandamme, Eds. (Cambridge University Press, Cambridge, 2009).

53. M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016 (2018).

54. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic biology* **67**, 901-904 (2018).

55. O. Chernomor, B. Q. Minh, F. Forest, S. Klaere, T. Ingram, M. Henzinger, A. von Haeseler, Split diversity in constrained conservation prioritization using integer linear programming. *Methods in ecology and evolution* **6**, 83-91 (2015).

56. D. L. Ayres, M. P. Cummings, G. Baele, A. E. Darling, P. O. Lewis, D. L. Swofford, J. P. Huelsenbeck, P. Lemey, A. Rambaut, M. A. Suchard, BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics. *Systematic biology* **68**, 1052-1061 (2019).

57. N. Goldman, Z. Yang, A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular biology and evolution* **11**, 725-736 (1994).

58. A. D. Yoder, Z. Yang, Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol* **17**, 1081-1090 (2000).

59. B. Vrancken, A. Rambaut, M. A. Suchard, A. Drummond, G. Baele, I. Derdelinckx, E. Van Wijngaerden, A. M. Vandamme, K. Van Laethem, P. Lemey, The genealogical population dynamics of HIV-1 in a large

transmission chain: bridging within and among host evolutionary rates. *PLoS computational biology* **10**, e1003505 (2014).

60. M. D. Baron, T. Barrett, Rescue of rinderpest virus from cloned cDNA. *Journal of virology* **71**, 1265-1271 (1997).

61. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

62. H. Li, seqtk. *https://github.com/lh3/seqtk*.

63. T. Chandler, *Four thousand years of urban growth: An historical census* (Mellen, Lewiston, 1987).

64. G. Modelski, *World cities: -3000 to 2000*. (Faros, Washington D.C., 2000, 2003).

65. P. Bairoch, *Cities and economic development: from the dawn of history to the present* (University of Chicago Press, Chicago, 1988).

66. D. Pasciuti, C. Chase-Dunn, Estimating the population sizes of cities. *https://irows.ucr.edu/research/citemp/estcit/estcit.htm* (Accessed 10/25/2019)

67. D. Pitt, N. Sevane, E. L. Nicolazzi, D. E. MacHugh, S. D. E. Park, L. Colli, R. Martinez, M. W. Bruford, P. Orozco-terWengel, Domestication of cattle: Two or three events? *Evolutionary applications* **12**, 123-136 (2019).

68. S. Morand, K. M. McIntyre, M. Baylis, Domesticated animals and human infectious diseases of zoonotic origins: domestication time matters. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **24**, 76-81 (2014).

69. J. C. Mariner, J. McDermott, J. A. Heesterbeek, A. Catley, P. Roeder, A model of lineage-1 and lineage-2 rinderpest virus transmission in pastoral areas of East Africa. *Preventive veterinary medicine* **69**, 245-263 (2005).

70. T. Barrett, P. B. Rossiter, Rinderpest: the disease and its impact on humans and animals. *Advances in virus research* **53**, 89-110 (1999).

71. W. H. McNeill, *Plagues and Peoples*. (Anchor Press, Norwell, 1976).

72. P.-P. Patoret, K. Yamanouchi, U. Mueller-Doblies, M. M. Rweyemamu, M. Horzinek, T. Barrett, "Rinderpest – an old and worldwide story: history to c.1902" in *Rinderpest and Peste des Petits Ruminants: Virus Plagues of Large and Small Ruminants*, T. Barrett, P.-P. Pastoret, W. Taylor, Eds. (Elsevier, Amsterdam, 2005).

73. J. Blancou, “Old prophylactic methods“ in *Rinderpest and Peste des Petits Ruminants: Virus Plagues of Large and Small Ruminants*, T. Barrett, P.-P. Pastoret, W. Taylor, Eds. (Elsevier, Amsterdam, 2005).

74. OIE, Peste des Petits Ruminants. *https://www.oie.int/fileadmin/Home/eng/Animal_Health_in_the_World/docs/pdf/Disease_cards/PESTE_DES_PETITS_RUMINANTS.pdf*.

75. D. G. Streicker, A. S. Turmelle, M. J. Vonhof, I. V. Kuzmin, G. F. McCracken, C. E. Rupprecht, Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* **329**, 676-679 (2010).

76. T. Lembo, C. Oura, S. Parida, R. Hoare, L. Frost, R. Fyumagwa, F. Kivaria, C. Chubwa, R. Kock, S. Cleaveland, C. Batten, Peste des petits ruminants infection among cattle and wildlife in northern Tanzania. *Emerging infectious diseases* **19**, 2037-2040 (2013).

77. E. Couacy-Hymann, M. Y. Koffi, V. K. Kouadio, A. Mossoum, L. Kouadio, A. Kouassi, K. Assemian, P. H. Godji, P. Nana, Experimental infection of cattle with wild type peste-des-petits-ruminants virus - Their role in its maintenance and spread. *Research in veterinary science* **124**, 118-122 (2019).

78. J. F. Drexler, V. M. Corman, M. A. Muller, G. D. Maganga, P. Vallo, T. Binger, F. Gloza-Rausch, V. M. Cottontail, A. Rasche, S. Yordanov, A. Seebens, M. Knornschild, S. Oppong, Y. Adu Sarkodie, C. Pongombo, A. N. Lukashev, J. Schmidt-Chanasit, A. Stocker, A. J. Carneiro, S. Erbar, A. Maisner, F. Fronhoffs, R. Buettner, E. K. Kalko, T. Kruppa, C. R. Franke, R. Kallies, E. R. Yandoko, G. Herrler, C. Reusken, A. Hassanin, D. H. Kruger, S. Matthee, R. G. Ulrich, E. M. Leroy, C. Drosten, Bats host major mammalian paramyxoviruses. *Nat Commun* **3**, 796 (2012).

79. W. Ghawar, H. Pascalis, J. Bettaieb, J. Melade, A. Gharbi, M. A. Snoussi, D. Laouini, S. M. Goodman, A. Ben Salah, K. Dellagi, Insight into the global evolution of Rodentia associated Morbilli-related paramyxoviruses. *Scientific reports* **7**, 1974 (2017).

80. J. R. Pulliam, J. H. Epstein, J. Dushoff, S. A. Rahman, M. Bunning, A. A. Jamaluddin, A. D. Hyatt, H. E. Field, A. P. Dobson, P. Daszak, Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *Journal of the Royal Society, Interface* **9**, 89-101 (2012).

81. K. Halpin, P. L. Young, H. E. Field, J. S. Mackenzie, Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. *The Journal of general virology* **81**, 1927-1932 (2000).

82. M. J. Mina, T. Kula, Y. Leng, M. Li, R. D. De Vries, M. Knip, H. Siljander, M. Rewers, D. F. Choy, M. S.

Wilson, Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* **366**, 599-606 (2019).

Acknowledgments:

The 1912 MeV genome was generated from a formalin-fixed lung specimen (museum object ID: BMM 655/1912) from the collection of the Berlin Museum of Medical History at the Charité (Berlin, Germany). Ethics approval was obtained from the ethics committee of the Charité (Berlin, Germany) under the reference number EA4/212/19. We thank Oliver Smith and Joel Wertheim for helpful suggestions. The Titan V GPU used for this research was donated by the NVIDIA Corporation.

Funding:

The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no.~725422-ReservoirDOCS). PL acknowledges support by the Research Foundation -- Flanders (`Fonds voor Wetenschappelijk Onderzoek -- Vlaanderen', FWO: G066215N, G0D5117N and G0B9317N). S.L. and B.V. are postdoctoral research fellows funded by the FWO. M.A.S. was partially supported through National Institutes of Health grant U19 AI135995.

Author contributions:

Conceptualization: PL, SCS; data curation: AD, SL, BV, MAS, PL, SCS; formal analysis: AD, SL, BV, JFG, MU, PL, SCS; funding acquisition: FHL, SCS; investigation: AD, LVP, SB, AH, DH, KM, BP and JS; methodology: SL, BV, MAS, PL; project administration: FHL, PL, SCS; resources: SS, NW, AM, FHL, TS, PL, SCS; software: SL, BV, MAS, PL; supervision: PL, SCS; validation: AD, MU, JFG, SCS; visualization: SL; writing – original draft: AD, SL, KH, PL, SCS; writing – review and editing: all authors.

Competing interests: All authors declare no competing interests.

Data and materials availability: The sequencing data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB36265 (https://www.ebi.ac.uk/ena/data/view/PRJEB36265). Human reads have been removed from 1912 sequencing files prior to uploading (supplementary materials). Alignments, trees, and BEAST xml files are available at https://github.com/slequime/measles-history.

Supplementary Materials:

Materials and Methods

Supplementary text S1-S4

Figures S1-S6

Tables S1-S7

References *35-82*

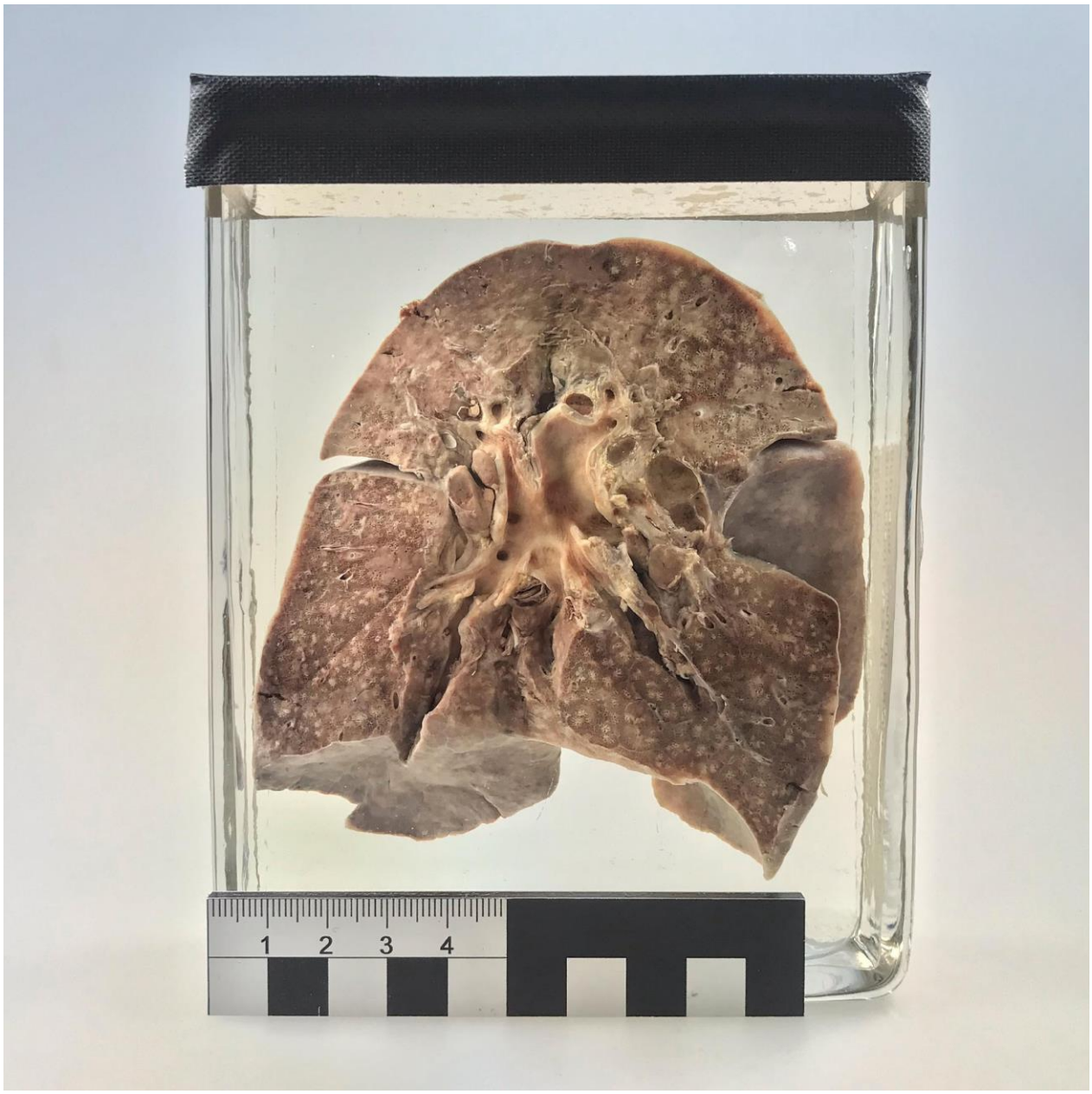

Fig. 1. Formalin-fixed lung specimen collected in 1912 in Berlin from a 2-year old girl diagnosed with measles-related bronchopneumonia (museum object ID: BMM 655/1912).

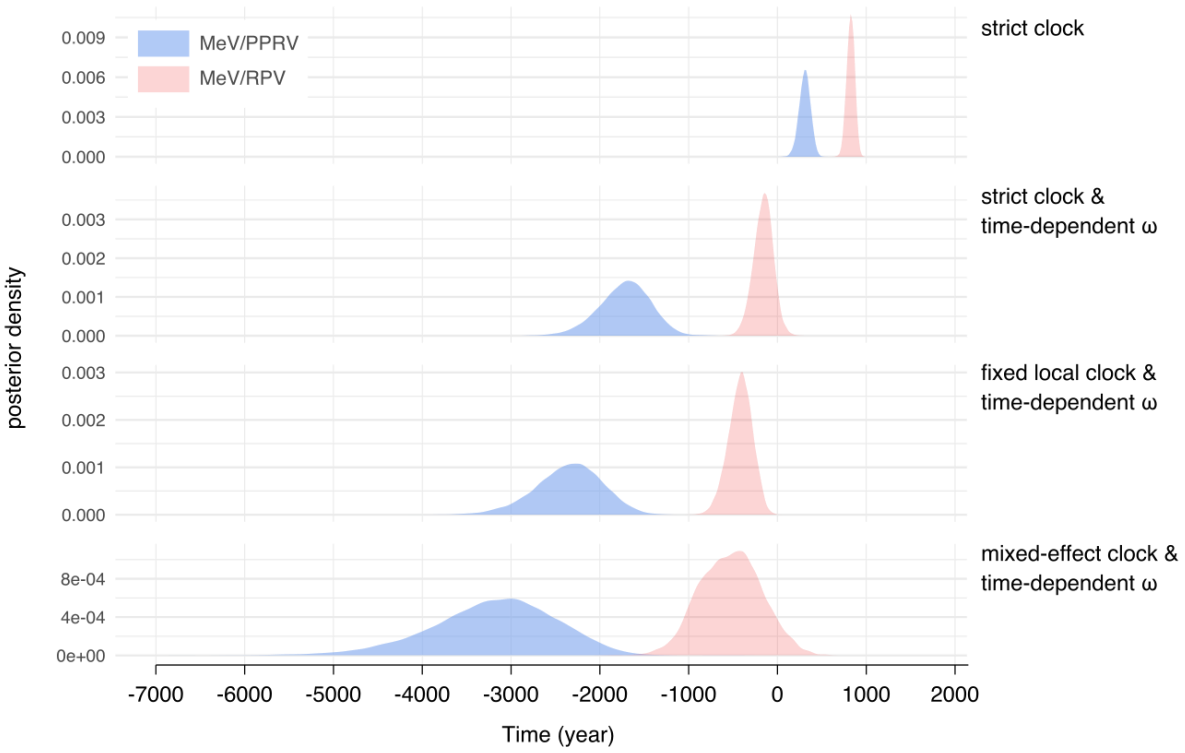

Fig. 2. Divergence time estimates for MeV and RPV (red) and for MeV/RPV and PPRV (blue) under increasingly complex evolutionary models. Estimates for parameters of interest (posterior mean and 95% highest posterior density interval) under each model are provided in table S5.

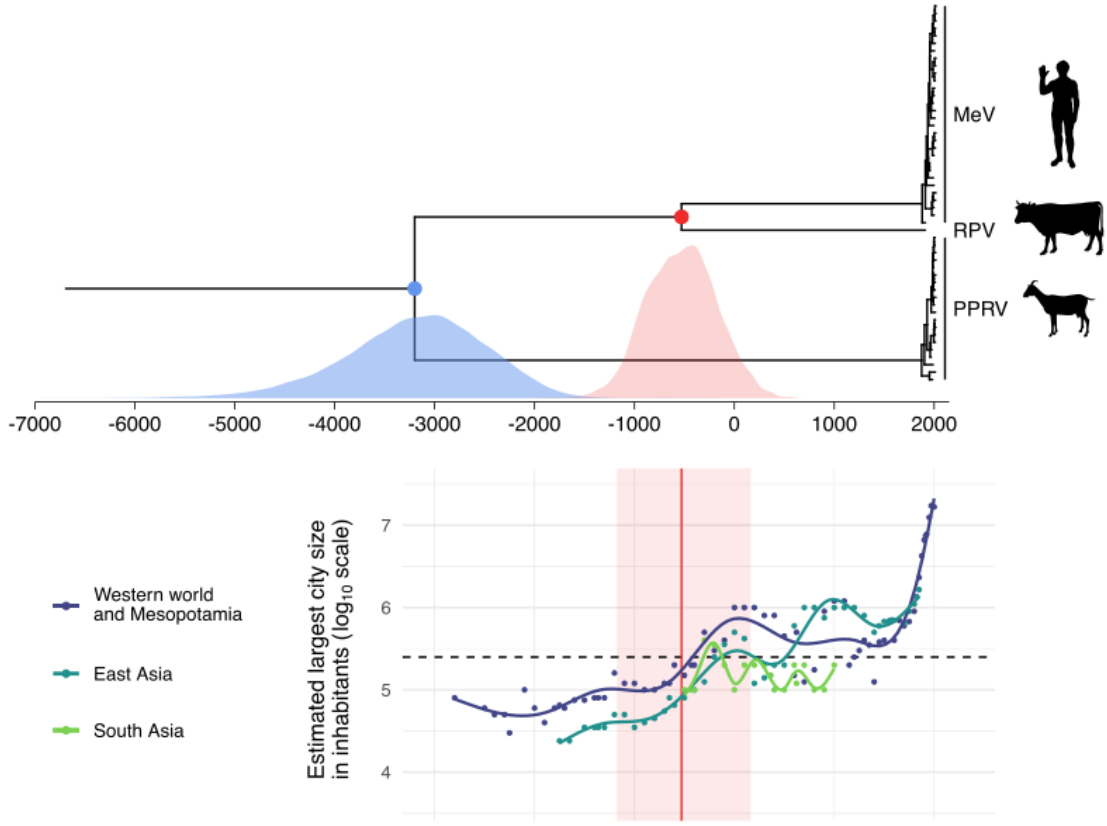

Fig. 3. Time-measured evolutionary history for MeV, RPV and PPRV, and largest city size over time in three well-studied regions of the world. Upper figure displays maximum clade credibility (MCC) tree summarized from a Bayesian time-measured inference using tip-dating and accounting for long-term purifying selection. The red and blue points represent the mean estimates for the divergence times between MeV and RPV and MeV/RPV and PPRV, respectively; the corresponding divergence date estimates are depicted below as marginal posterior distributions. The lower figure represents the estimated size (log10 scale) of the largest city in the western world including Mesopotamia (dark blue), East Asia (teal) and South Asia (green) over time. The red vertical line represents the mean divergence time estimate between MeV and RPV and the red area its 95% highest posterior density interval. The dashed horizontal line represents the classical threshold for MeV maintenance in a population (i.e. 250,000

individuals). Dots show data points according to Morris (*34*) and Inoue et al. (*26*). Each line represents the fit of a generalized additive model with a cubic spline smoothing function.

Supplementary Materials for

Measles virus and rinderpest virus divergence dated to the rise of large cities

Authors: Ariane Düx[1,2]†, Sebastian Lequime[3]†, Livia Victoria Patrono[1,2], Bram Vrancken[3], Sengül Boral[4], Jan F. Gogarten[1,2], Antonia Hilbig[1], David Horst[4], Kevin Merkel[1,2], Baptiste Prepoint[2], Sabine Santibanez[5], Jasmin Schlotterbeck[2], Marc A. Suchard[6], Markus Ulrich[1], Navena Widulin[7], Annette Mankertz[5], Fabian H. Leendertz[1], Kyle Harper[8], Thomas Schnalke[7], Philippe Lemey[3]‡, Sébastien Calvignac-Spencer[1,2]‡*

Correspondence to: calvignacs@rki.de

This PDF file includes:

Materials and Methods

Supplementary Text S1 to S4

Figures S1 to S6

Tables S1 to S7

References *35* to *82*

Materials and Methods

Precautions for ancient DNA

To avoid contamination with modern DNA, all laboratory analyses preceding amplification for dual indexing were performed in the clean room of our ancient DNA laboratory. Here, specific precautions are applied, including restricted entry to trained personnel wearing full body suits, masks, overshoes and double gloves, and daily UV light decontamination of the whole room. Prior to this study, no measles positive samples were handled in these facilities.

Sample material

We obtained a sample of a formalin-fixed lung specimen (museum object ID: BMM 655/1912) from the collections of the Berlin Museum of Medical History at the Charité (Berlin, Germany). The specimen is a lung that was collected during an autopsy of a 1912 measles patient from Berlin. According to the case file, the lung belonged to a 2-year-old child that died on June 3rd, 1912, of measles-related bronchopneumonia. Since the exact composition and concentration of the formalin fixative was unknown, we stored the collected lung sample in PBS to avoid further damage to nucleic acids by adding fresh formalin.

In addition to the 1912 measles case, we obtained two *Measles morbillivirus* (MeV) strains (MVi/Prague.CZE/60/1 and MVi/Prague.CZE/60/2) which were isolated by the National Reference Laboratory of former Czechoslovakia in Prague in 1960. These strains were kept by the National Reference Laboratory of the former German Democratic Republic until they became part of the strain collection of the German National Reference Laboratory (Robert Koch Institute, Berlin, Germany) after the German reunification. The strains were isolated in the intraamniotic cavity of chick embryos and passaged inVero cells for 5 and 6 passages, respectively, and for one passage in human SLAM-expressing Vero cells.

Extraction

To maximize chances of viral RNA recovery, we performed eight separate nucleic acid extractions from different areas of the lung using the DNeasy® Blood & Tissue Kit (Qiagen) with modifications for formalin-fixed samples. For each separate extraction, a pea-sized piece (ca. 25mg) of lung was washed in 1 ml PBS to remove residual fixative. The washed tissue was cut into smaller pieces using sterile scissors and added to bead tubes containing tissue lysis buffer (ATL). To reverse formalin-induced crosslinking, the tissue was heated to 98°C for 15 minutes. To facilitate lysis, the tissue was homogenized by bead beating for three times 20s at 4m/s with a Fast Prep® (MP Biomedicals). We added 20 µl Proteinase K and kept the homogenate at 56°C until the tissue was completely lysed (ca. 1 hour). Subsequent steps were performed according to the manufacturer's protocol and nucleic acids were eluted in 35 µl elution buffer (AE).

Library preparation

To maximize viral RNA content in the final sequencing libraries, we removed DNA and ribosomal RNA from the nucleic acid extracts before conversion to double-stranded cDNA. For DNase treatment we used the TURBO DNA-*free*™ Kit (Ambion). We performed ribosomal RNA depletion and clean-up separately on the first four DNase-treated extracts, using NEBNext® rRNA Depletion Kit (Human/Mouse/Rat) with RNA Sample Purification Beads (New England Biolabs) according to protocol. To reduce costs, we pooled and concentrated the next four DNase-treated extracts using RNA Clean & Concentrator-5 Kit (Zymo Research) and eluted in 13 µl nuclease free water as input for one ribosomal RNA depletion reaction. In both cases, following bead clean-up, RNA was eluted in 12 µl nuclease free water. We performed cDNA synthesis, using the SuperScript™ IV First-Strand Synthesis System (Invitrogen) and

converted cDNA into dsDNA with the NEBNEXT® mRNA Second Strand Synthesis Module (New England Biolabs). Double-stranded DNA was purified using MagSi-NGS^{prep} Plus Beads (Steinbrenner Laborsysteme) and eluted in 50 µl Tris-HCl (10mM) EDTA (1mM) Tween20 (0.05%) buffer. We prepared five separate libraries with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs) without prior fragmentation of double-stranded cDNA and without size-selection upon adapter ligation. All clean-up steps during the library preparation were conducted with MagSi-NGS^{prep} Plus Beads (Steinbrenner Laborsysteme). Libraries were dual indexed with NEBNext® Multiplex Oligos for Illumina® (New England Biolabs), quantified using the KAPA Library Quantification Illumina Universal Kit (Roche), amplified with the KAPA HiFi HotStart ReadyMix (Roche) and Illumina adapter-specific primers, and diluted to a concentration of 4nM for sequencing.

Sequencing

Libraries were sequenced on an Illumina® MiSeq platform using the v3 chemistry (2x300-cycle) and on an Illumina® NextSeq platform using v2 chemistry (2x150-cycle) for a total of 48,326,978 unfiltered paired reads.

Genome assembly

Sequencing reads were filtered (adapter removal and quality filtering) using Trimmomatic (*35*) with the following settings: LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:30. For the 1912 sequences, we attempted *de novo* assembly on a subsample of 2,083,813 reads (up to 500,000 reads per library) using metaSPAdes (*36*). However, since the largest MeV contig generated with this approach covered only 1657 nt, we chose to use a two-step reference based mapping approach. We first used BWA-MEM (*37*) to map trimmed reads to

the MeV RefSeq genome (NC_001498). We then called the consensus of the two largest contigs and determined the closest MeV genome via BLASTn search against the NCBI non-redundant nucleotide collection (*38*). The best hit for both contigs was the measles virus strain Edmonston (AF266288). Subsequently, we merged paired reads with ClipAndMerge (*39*) and used BWA-MEM to map all reads (merged, unmerged, unpaired) to the Edmonston genome. We sorted the mapping files, removed duplicates and generated insert size distributions with the tools SortSam, MarkDuplicates and CollectInsertSizeMetrics from the Picard suite (*40*). We compared MarkDuplicates to Dedup (*41*) to select the more stringent tool for duplicate removal, finding that for all libraries MarkDuplicates removed more reads. We proceeded with these more conservative mapping results. We also generated damage profiles for all libraries using mapDamage 2.0 (*42*). For the final genome assembly, we merged the mapped reads (bam files) of the five libraries using MergeSamFiles from the Picard suite (*40*). Using Geneious 11.1.5 (*43*), we assembled two consensus genomes, 1) with a minimum coverage of 20x and a 95% majority for base calling (modern DNA settings), and 2) with a minimum 3x coverage and 50% majority consensus (ancient DNA settings). The consensus sequences contained 11,988 and 15,257 unambiguous positions, respectively. The 1960s MeV strains were treated in the same way resulting in consensus of 15,870 nts and 15,868 nts (minimum 3x coverage; 50% majority consensus), and 15,780 nts and 15,713 nts (minimum 20x coverage; 95% majority consensus), respectively.

Phylogenetic and molecular clock analyses

Dataset preparation

We collected all available full length MeV, rinderpest virus (RPV) and "peste des petits ruminants" virus (PPRV) genomes from GenBank, and excluded all vaccine strains, as well as

subacute sclerosing panencephalitis MeV strains (*44*), and RPV strain Fusan (AB547189) because of its extensive and not completely resolved passage history (*45*). This resulted in a dataset of 133 MeV (including the 3 new sequences from this study), 1 RPV, and 73 PPRV full-length genomes. Sequences from MeV and PPRV were aligned independently using MAFFT v7 (*46*) and checked for recombination using RDP4 (*47*). Sequences exhibiting a significant result for more than three of the selected recombination detection methods (RDP, GENECONV, Chimaera, MaxChi, Bootscan, SiScan, and 3Seq) were excluded from subsequent analyses. The final dataset consisted of complete genomes for 130 MeV strains, 65 PPRV strains and 1 RPV strain.

Temporal signal and saturation analyses

Phylogenetic trees for the final full-length MeV and PPRV genomes were reconstructed using IQ-TREE version 1.6.10 (*48*) under a general time-reversible (GTR) substitution model with discrete Γ-distributed rate variation among sites. We used TempEst v1.5.1 (*49*) to visually explore the temporal signal in the resulting maximum likelihood trees by plotting root-to-tip divergence as a function of sampling time with the root that optimized their correlation for both MeV and PPRV. These plots indicated a clear accumulation of divergence as a function of sampling time with a faster rate for PPRV as compared to MeV (fig. S5). We confirmed the presence of temporal signal using a formal Bayesian approach recently developed by Duchene et al. 2019 (table S3; *50*). This test involves comparisons of marginal likelihood estimates under a dated-tip model with an estimable evolutionary rate and an ultrametric model with a fixed rate. Bayes factor support values of 165 and 43 for MeV and PPRV respectively provide strong evidence for temporal signal (table S3).

To investigate saturation in the data sets, we performed the substitution saturation test developed by Xia et al. 2003 (*51)*, and Xia and Lemey 2009 (*52*). The test employs an information entropy-based index of substitution saturation and compares the value of this index for subsets of the sequence data to the corresponding critical values obtained by computer simulations on symmetrical and asymmetrical tree topologies. Table S4 summarizes these indices and their critical values for all the data sets, showing that the tests fail to identify strong substitution saturation.

Phylogenetic analysis of the 1912 sample

A time-measured phylogenetic tree for the 130 full-length MeV genomes was constructed using BEAST v1.10.4 (*53*) under a GTR + Γ nucleotide substitution model, an uncorrelated relaxed lognormal clock and a constant population size. Multiple independent MCMC chains were run sufficiently long to ensure proper mixing after convergence to the posterior of the model parameters, as measured by sufficiently high effective sample sizes (ESS > 200) using Tracer v1.7 (*54*). Trees and log files were combined after removing burn-in and posterior distribution of trees was summarized as a maximum clade credibility (MCC) trees using TreeAnnotator 1.10.4 (fig. S3; *53*). The Bayesian phylogenetic tree placed the 1912 genome as a sister lineage to all modern genomes (fig. S3).

To confirm the phylogenetic placement of the 1912 genome and examine to which extent its position was driven by its age in the dated-tip tree, we explored how this sample branched in a full-genome MeV maximum likelihood (ML) phylogeny without date information. We inferred a ML tree for the 130 full-length MeV genomes in IQ-TREE version 1.6.10 (*48*) under a general time-reversible (GTR) and discrete Γ model. In the non-clock ML phylogeny, the 1912 sample branched very deeply in the tree, on the branch linking the H1 genotype to the remaining MeV

diversity (fig. S4A), which is consistent with the clustering in the dated-tip phylogeny (fig. S3). In addition, we also performed the BEAST time-measured phylogenetic inference without the 1912 genome. This confirmed that the root is positioned on the branch ancestral to genotype H1 (fig. S4B).

Phylogenetic estimation of the RPV-MeV divergence time

We inferred maximum likelihood (ML) trees for the remaining 130 MeV and 65 PPRV full-length genomes in IQ-TREE version 1.6.10 (*48*) under a general time-reversible (GTR) and discrete Γ model. These were used to subsample the dataset while keeping an optimal representation of the genetic diversity of each virus using Phylogenetic Diversity Analyzer v0.5 (*55*). This step was necessary to ensure manageable run-time during subsequent analysis with computationally expensive codon substitution models in a Bayesian framework. The final set (table S2) consisted of 30 MeV genomes, including the 1912 genome and one of the two nearly identical 1960 genomes generated in this study, 20 PPRV genomes and 1 RPV genome. To represent the 1912 genome we use the consensus genome called by ancient DNA settings (a minimum 3x coverage and 50% majority consensus), but we also explore the use of the consensus genome called by modern DNA settings (cfr. below).

All 51 full-length genomes were aligned using MAFFT and manually edited to keep only conserved coding regions among the three viral species and to exclude the P/C/V overlapping gene region. The alignment was manually checked at the codon level to ensure all sequences retained a proper translation frame and did not contain stop codons; one MeV sequence (GQ376026) had an improper coding frame that could be corrected by inserting an N at position 7,025, resulting in a correct amino-acid sequence, similar to other isolates of the same species. The final dataset is composed of 13,131 nucleotide sites (4,377 codons). Using the same

approaches as applied to the separate MeV and PPRV data sets, we also confirmed the presence of temporal signal and the absence of strong saturation in this data set (table S3 and S4).

The time-measured evolutionary history of MeV, RPV, and PPRV was reconstructed using BEAST v1.10.4 (*53*) and the BEAGLE 3 v3.1.2 library to improve computational performance (*56*). Four different evolutionary models, representing an increasing level of complexity, were fitted to the data. All of them used a Goldman-Yang 1994 (GY94) codon substitution model (*57*) and a constant population size tree prior. The first model assumed a strict clock. The second model considered evolution under a strict clock while allowing time-dependency of the nonsynonymous to synonymous substitution rate ratio (dN/dS or ω) under an epoch structure to take long-term purifying selection into account during the phylogenetic reconstruction (*11*). The third model combined the same time-dependent ω substitution model with a fixed local clock (*58*), allowing for an estimable difference in substitution rate between the PPRV and MeV-RPV clades. The fourth model further extends the time-dependent ω substitution process and fixed local clock specification with branch-specific random effects to model additional branch-to-branch rate variation (*59*). Within the fixed local clock specification of the third and fourth model, there is no prior information to determine whether the branch ancestral to the PPRV clade (its 'stem' branch) should share the fixed rate effect assigned to that clade or not. Therefore, we modelled the inclusion of that effect on the stem branch as an estimable parameter using a Bayesian variable selection procedure.

For each model, multiple independent MCMC chains were run until convergence and proper mixing of the model parameters were achieved, as measured by sufficiently high effective sample sizes (ESS > 200) using Tracer v1.7 (*54*). Trees and log files were combined after removing

burn-in and posterior distributions of trees were summarized as maximum clade credibility (MCC) trees using TreeAnnotator 1.10.4 (*53*). The time to most recent common ancestor of the whole tree and the MeV-RPV clade were summarized using mean estimates and 95% highest posterior density (HPD) intervals.

By fitting the four models with increasing complexity to the morbillivirus data set, we estimate increasingly older divergence times (with wider 95% HPD intervals) indicating that additional model complexity allows for a better correction of multiple substitutions, in particular multiple synonymous substitutions as indicated by the codon substitution model estimates (table S5, Fig. 2). The estimates for the parameters that represent additional complexity with respect to simpler models (in a nested fashion) also indicate that there is significant evidence for the complexity in the data, e.g. the credible interval for a negative time-dependent effect on ω does not include zero, the credible intervals of the fixed effect rates do not overlap, and the credible interval for the standard deviation of the random effects excludes zero (table S5). The most complex model therefore provides the best fit to the data (table S5). The branch ancestral to the PPRV clade did not receive any support for sharing the same fixed effect modelled within the PPRV clade.

We explored the robustness of our estimates with respect to several particular differences in the data set and model specifications. Estimating the age of the 1912 genome or excluding the 1912 genome yielded highly similar RPV/MeV divergence time estimates (table S6), indicating that the data point represented by the 1912 genome is in line with our model-based extrapolation from modern genomes to ancient divergence times. Using a more conservative 1912 consensus genome (called by a minimum coverage of 20x and a 95% majority rule) also did not affect the RPV/MeV divergence time estimates (table S7).

The same holds true when using different priors on the age of the RPV (Kabete 'O') strain (table S7). The above models used an exponential prior centered on 1960 with a mean of 10 for the age of this RPV isolate in order to model its uncertainty in the presence of a rich passage history (*60*). Changing this prior to a uniform between 1950 and the year of GenBank deposition did not have a strong impact on the RPV/MeV divergence time estimates (table S7). Similarly, adding an additional RPV genome (Fusan: AB547189) with an age that is difficult to assess given its unclear passage history did not influence the RPV/MeV divergence time estimates (table S7). The Fusan strain sequence was originally collected in the early 1920's and maintained for about 1,000-1,500 passages in animals until April 1946 (*45*). It was subsequently passaged another 219 times in animals until an undisclosed date or decade (*45*). To accommodate this uncertainty, we specified a uniform prior over its age bounded by the final year of the first series of passages and the year of GenBank deposition.

Finally, we also tested the non-parametric Bayesian Skygrid coalescent prior as an alternative to the constant size population model. While this produced a somewhat older mean RPV/MeV divergence time estimate, it is covered by the credible interval of the estimate under a constant size population model and 80% of this credible interval overlaps with the credible interval of the estimate under a skygrid model (table S7).

<u>Removal of human reads prior to sequence uploading</u>

In agreement with the ethics committee of the Charité, Berlin, human reads have been removed from the 1912 sequencing files prior to uploading them to the European Nucleotide Archive. Filtered reads were mapped to the human RefSeq genome assembly GRCh38.p13 (GCF_000001405.39) using BWA-MEM (*37*). Forward and reverse reads were mapped

separately. Unmapped reads were extracted using Samtools (*61*). These reads were then extracted by name from the original fastq files using seqtk (*62*).

Estimating historical urban populations

To bring the MeV/RPV divergence estimate in a historical context, we connected it to demographic data. We obtained estimates of the past population sizes of cities from standard historical and archaeological literature, collating data from Inoue et al. 2015 (*26*) and Morris 2013 (*34*), who in turn draw from Chandler 1987 (*63*), Modelski 2003 (*64*), Bairoch 1988 (*65*), and other more localized studies. It should be noted that estimating historical urban populations entails considerable methodological and empirical difficulties. The very definition of a settlement is a complex question. Inoue et al. (*26*) define a settlement as a "spatially contiguous built-up area," and this definition reasonably allows us to capture what is most epidemiologically important, the number of people living in contact within a single area. An overview of the methodological issues is provided by Pasciuti and Chase-Dunn 2002 (*66*).

Supplementary Text

S1. Directionality of the host switch: from cattle to humans or vice versa?

Cattle domestication, which started about 10,000 years ago (*67*), increased contact between humans and bovines and provided ample opportunity for spill-over of infectious diseases in either direction (*68*). In the case of measles, the predominant view is that the common ancestor of MeV and RPV was a cattle-infecting virus that eventually was transmitted to humans (*6*). MeV and RPV both require large populations to sustain endemicity (i.e. greater than the critical community size (CCS) of 250,000 – 500,000 individuals for measles (*22-24*) and ca. 200,000 for rinderpest (*69*)). The assumption that herds of cattle and wild artiodactyls were sufficiently large to support an RPV-like pathogen long before human populations met the MeV CCS is a central argument to support the hypothesis that measles originated from a cattle-infecting ancestor (*70, 71*). Indeed, it seems plausible that growing herds of domesticated cattle in combination with wild aurochs and other susceptible artiodactyls had the numbers to support circulation of a multi-host virus like RPV. However, we lack reliable estimates of population sizes of cattle and wild ungulates throughout ancient history which would be required to validate this theory.

Historic sources that may refer to rinderpest exist much earlier than for measles. Descriptions consistent with rinderpest date back as far as the 2nd millennium BCE to an Egyptian veterinary papyrus and can be found on an Indian palm leaf text from ca. 1000 BCE as well as in several antique texts (*28, 72, 73*). However, as in the case of measles (and in fact most diseases) an unambiguous identification of rinderpest from such historical documents is impossible.

Perhaps the most compelling and supported argument in favor of cattle to human spill-over derives from the phylogeny of the morbillivirus genus. The closest relative to MeV/RPV is PPRV, a virus that mainly infects sheep and goats, but is also known to infect other artiodactyls (*74*). Phylogenetic proximity facilitates host switches of pathogens (*75*), making a transmission

from small ruminants to cattle and then to humans more likely than a switch from small ungulates to humans and then back to cattle (a host switch from humans via small ruminants to cattle seems unlikely given the virus tree topology). Modern day PPRV causes occasional subclinical infections in cattle (*76, 77*), showing that barriers to spill-over between small ruminants and cattle are low. It is also conceivable that the common ancestor of PPRV and MeV/RPV infected a broad range of ungulates including cattle and diverged into two more specialized pathogens.

At a deeper evolutionary scale, small mammals appear to be the ancestral hosts of paramyxoviruses (rodents in the case of morbilliviruses; (*78, 79*). Most human-infecting paramyxoviruses are thought to have a relatively recent zoonotic origin, including a number of instances where it is known that domesticates have acted as intermediate hosts (e.g. pigs and horses for Nipah and Hendra viruses; (*80, 81*). Thus far, the opposite directionality (human to domestic animal) has never been documented. This imbalance might either reflect insufficient discovery effort in domestic animals or represent a complex biological trait shared by all paramyxoviruses – the latter possibility rather supporting cattle to human transmission of the ancestor of MeV and RPV.

All these considerations remain speculative and a scenario of human to cattle transmission, though there is little support for this, cannot be excluded. However, irrespective of the direction of the host switch, the MeV/RPV divergence had enormous consequences on human health (*82*), either directly as a major childhood disease or indirectly via its devastating effects on cattle and thereby on human food supplies.

S2. Pathological report

The pathological report pertaining to the 1912 museum specimen of a 2-year-old child states that the patient was hospitalized with a diagnosis of measles, bronchitis, bronchiolitis and

bronchopneumonia. The child died after 3 days in the hospital and a post mortem examination was performed one day after death. The gross pathological findings were translated from German and are listed below.

> "Pathological Diagnosis: Measles-related bronchopneumonia of both lungs with multiple foci of inflammation and mild interstitial pneumonia, bronchitis, tracheitis, bronchiectasis of the left lung, pharyngitis, sub-pleural hemorrhage, alveolar emphysema, swelling of mesenterial and isolated ileal lymph nodes.
>
> The body of the child was well muscled, had a strong built and was well nourished. The skin of the upper chest showed red spots that did not vanish upon pressure. The peritoneum was smooth and glossy. The liver protruded two finger's breadth under the lower end of the costal arch. Position of the diaphragm on both sides: 5th rib. The cartilage-bone junctions of the ribs are distended. After removal of the sternum, the lungs hardly retract into their cavities. The pleural cavities are free of foreign content. The pericardium is filled with a few ccm of clear, serous liquid. Upon opening the heart, uncoagulated blood flows from the heart ventricles and atriums.
>
> The heart itself, about 1.5 times the size of the subject's fist, is taut, myocardium left: 1.1 – 0.2 cm, myocardium right: 0.2 cm, of brownish-red color, the endocardium is smooth and glossy, the cardiac valves are thin and delicate. The interior of the heart is filled with blood and postmortem clots.
>
> Left lung: pleura is smooth and glossy, the color varies between a light greyish red and a dark blueish red. The light areas protrude over the surrounding tissue, smaller air bubbles are visible and big sub-pleural bubbles present like a string of pearls. The dark areas feel dense and rubbery; the light areas are air-filled and elastic. There are several red spots under the pleura that do not vanish upon pressure. The cut surface is very uneven; the

color is light mottled with a dark reddish blue. The liquid that escapes the tissue contains only very little air and is bloody. In case of the smallest bronchi, there is a small amount of yellowish green liquid. Some bronchi with more pronounced reddening of the mucosa are also filled with greenish yellow mucus. Isolated bronchi show cylindrical dilations. Pulmonary lymph nodes are swollen to the size of hazelnuts, are very firm and extremely red.

Right lung: findings correspond to left lung.

Throat: the pharyngeal mucosa and the area around the larynx opening are reddened. Both tonsils are slightly swollen and filled with a small amount of greenish yellow liquid that spills out when tonsils are cut. Esophagus no particular findings. The tracheal mucosa is extremely red, the lymph nodes of the neck are swollen and very red. Thyroid gland, thoracic aorta without particular findings.

Spleen: 7.5 x 4.5 x 2.5 cm. The color is blue-red, the consistency firm-elastic. The surface is smooth and glossy. The cut surface clearly shows the structure of the lymph follicles, the color is blue-red, the pulpa is intact.

Kidneys: (2.5 x 3 x 3 cm) capsule can be removed easily, color grey-red, surface smooth and glossy. Consistency: firm-elastic. On the cut surface cortex and medulla can easily be distinguished. The parenchyma is slightly cloudy, renal pelvis no particular findings. Right kidney (7.5 x 3 x 2.5 cm): findings correspond to left kidney. Adrenal glands are very pale, otherwise without particular findings.

Urinary bladder, genital tract, rectum without particular findings.

Liver: (16.5 x 11 x 5.5 cm) surface is smooth and glossy. Color: brown-red. Consistency: firm-elastic. On the cut surface the hepatic lobules are completely clear, the parenchyma is clear, blood content is moderate. Gall bladder no particular findings.

Stomach, pancreas, duodenum without particular findings. Some ileal lymph nodes are slightly swollen. Abdominal aorta no particular findings."

S3. Histopathological report

We performed histopathological analyses of the 1912 formalin-fixed lung specimen. For conventional histopathology, slides of lung tissue were stained with hematoxylin and eosin (fig. S1). We observed an active, focally purulent bronchopneumonia with widespread interstitial pneumonia, associated with multinucleated giant cells of the Warthin-Finkeldey cell type. These findings, while not specific for MeV infection, are consistent with measles-related bronchopneumonia. Immunohistochemical approaches were unsuccessful due to the age and low quality of the specimen.

S4. Ancient RNA characterization

While ancient DNA (aDNA) characteristics are now well described, much less is known about ancient RNA (aRNA). We therefore investigated characteristics of aRNA sequences generated from formalin-fixed tissue. We analyzed insert size distributions and damage profiles of reads that mapped to the MeV genome (using our consensus genome as reference) or to the human mitochondrial genome (using NC_012920 as a reference). We found that aRNA molecules were generally much longer than previously described (*33*). For MeV, median lengths fell between 93-136 nt and maximum lengths between 217-373 nt; for human mitochondrial DNA (mtDNA), median lengths comprised 66-143 nt and maximum lengths 171-373 nt[1] (table S1, fig. S2A). We did not find any evidence of RNA damage, whether considering MeV or human mtDNA reads

[1] Our protocol included a ribosomal RNA (rRNA) depletion step so we could not use the expected excess of rRNA as a criterion to authenticate these molecules as being RNA ones (as opposed to being their DNA genomic template). However, our maps clearly showed that very few reads mapped to the non-coding D-loop region of the mtDNA genome (data not shown), suggesting few DNA molecules were present in our libraries.

(fig. S2B). These observations suggest a surprisingly good preservation of RNA molecules. Analyses of additional formalin-fixed specimens will reveal whether this is an idiosyncratic finding or a general characteristic of this type of archival specimen.

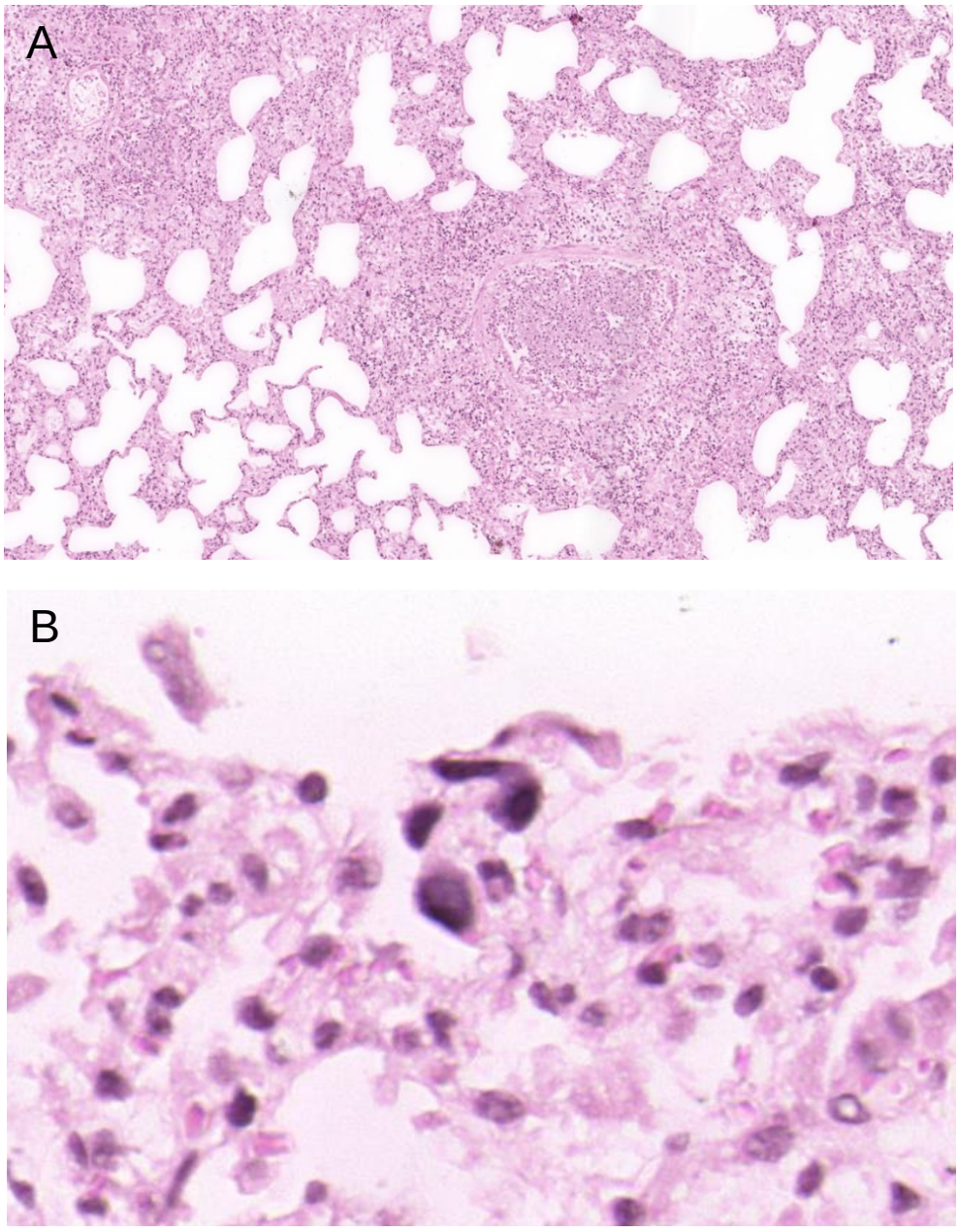

Fig. S1. Histopathology of formalin fixed lung. Hematoxylin and eosin stain. Magnified 50x (A) and 400x (B). Active, focally purulent bronchopneumonia with widespread interstitial pneumonia, associated with multinucleated giant cells of the Warthin-Finkeldey cell type.

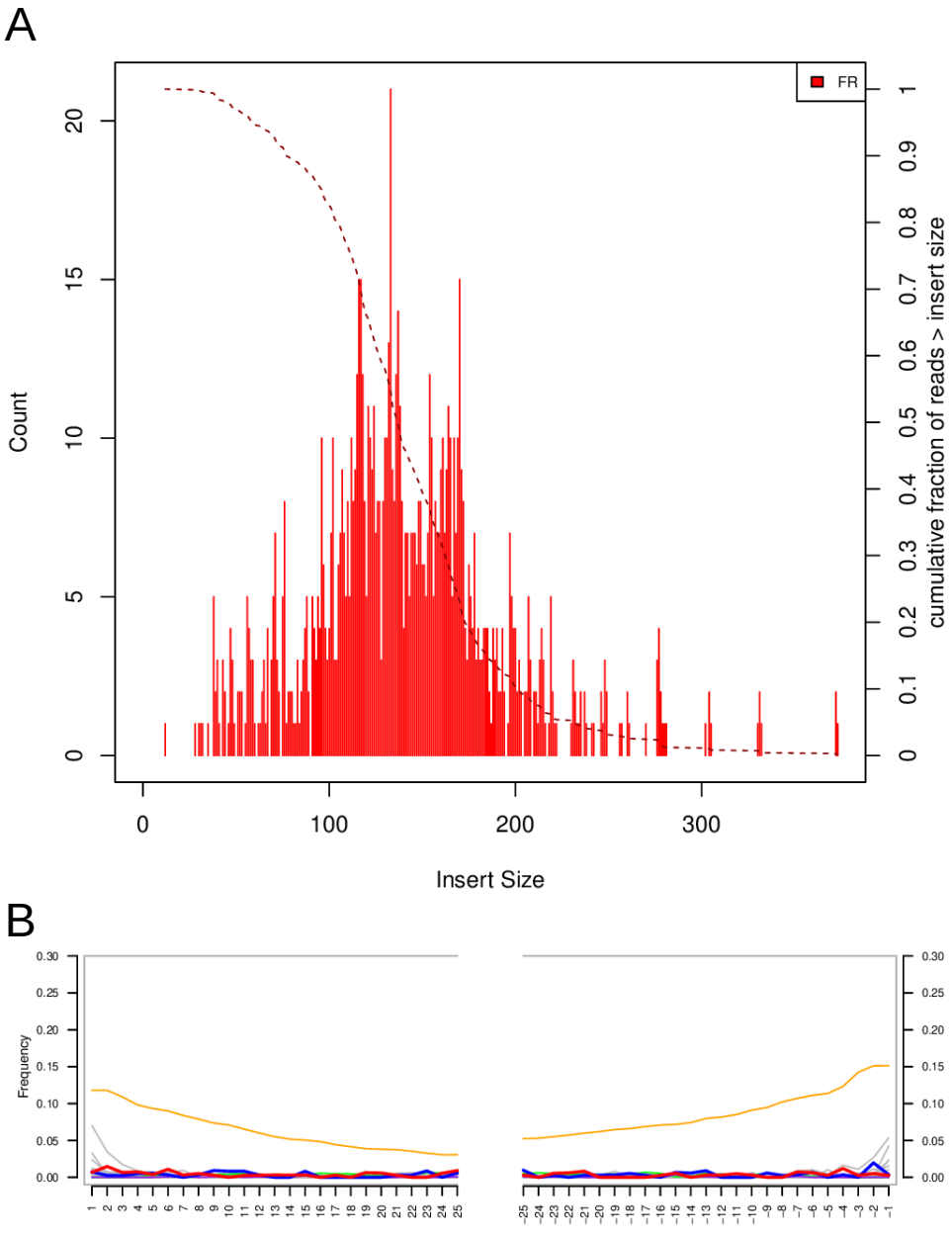

Fig. S2. Insert size distribution (A) and mapDamage profile (B) of reads in library #5. Library #5 comprised longer fragments than libraries #1-4 but showed a very similar damage profile.

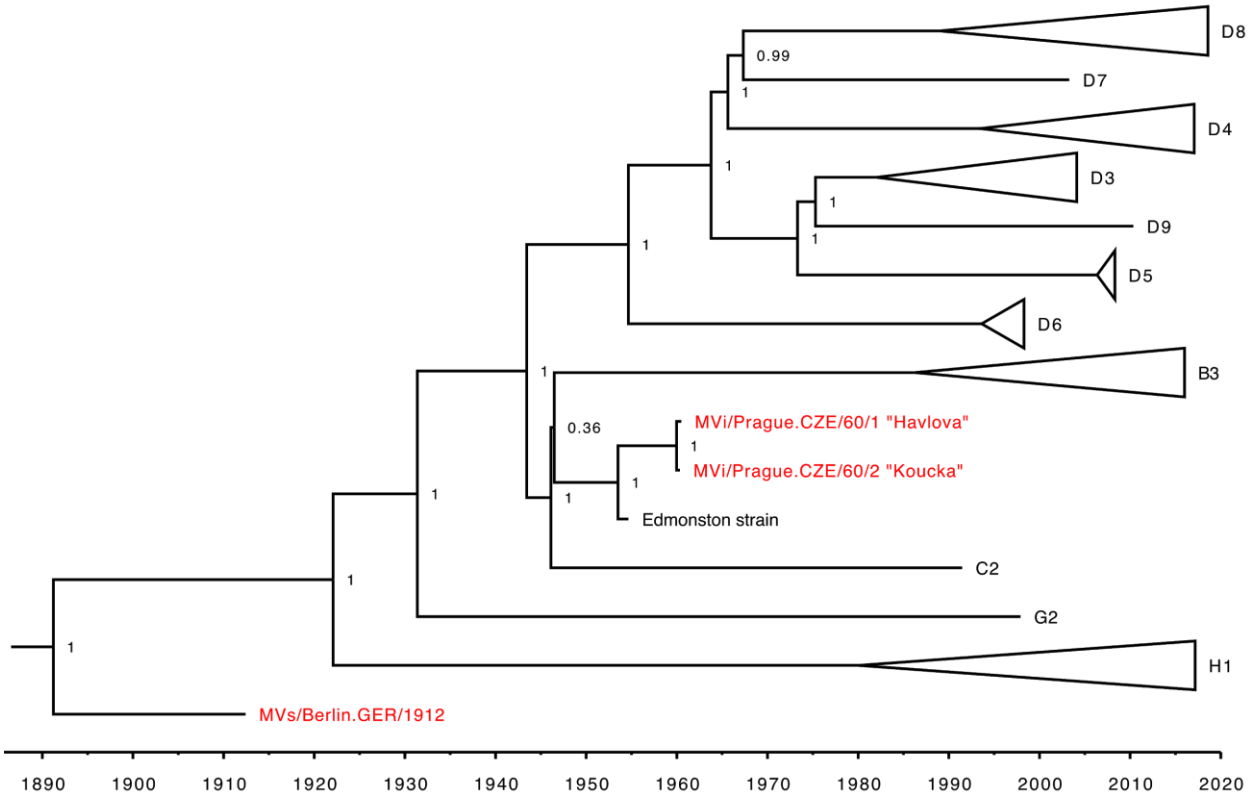

Fig. S3. Timed phylogenetic tree of measles virus (n=130) including the three sequences generated in this study (highlighted in red), showing the main genotypes. This tree is the maximum clade credibility (MCC) tree derived from a Bayesian evolutionary inference for 130 full genomes of MeV. A time scale is shown at the bottom. Numbers represent node posterior probability support.

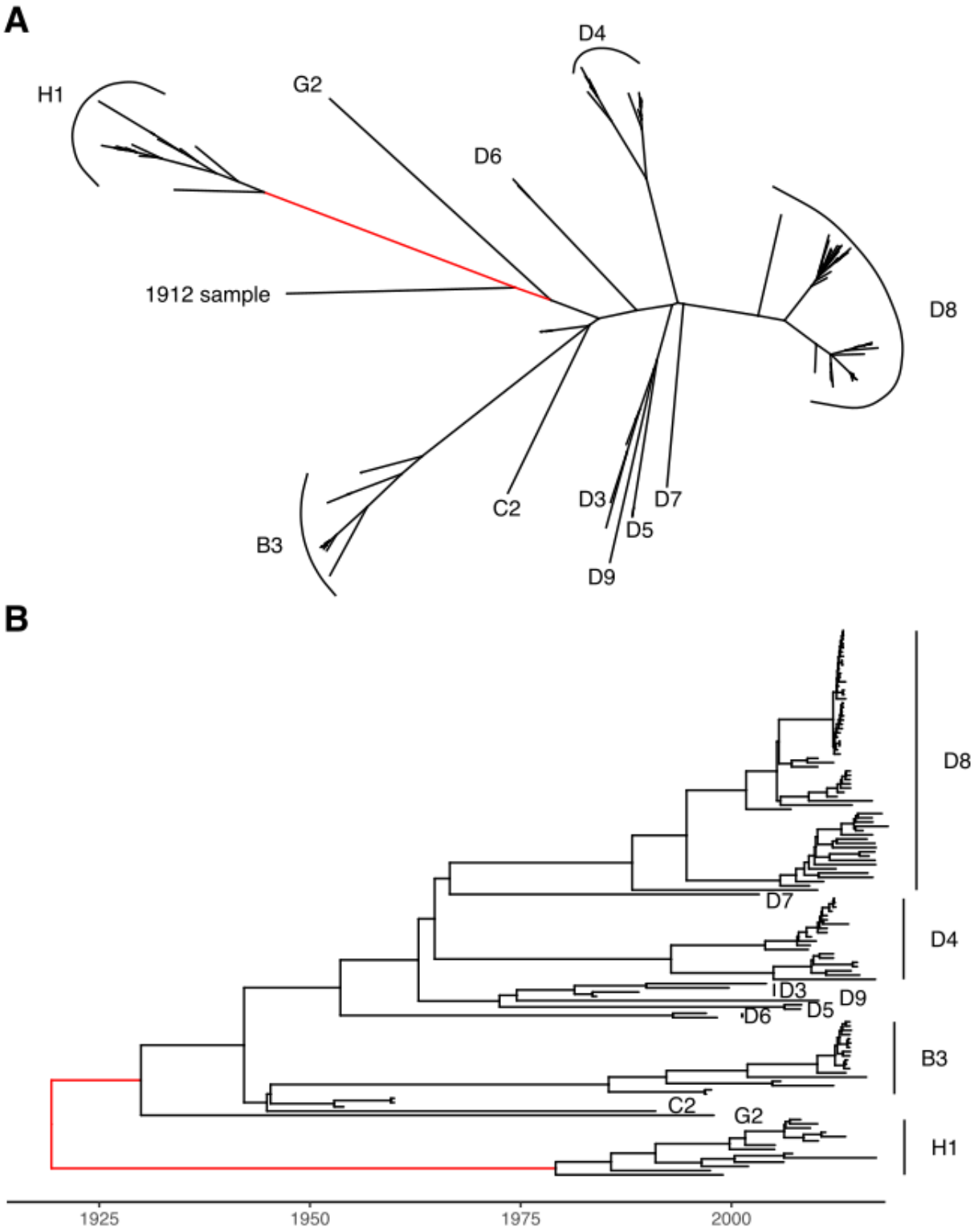

Fig. S4. Phylogenetic exploration of the 1912 measles genome sequence. (A) Unrooted tree for all 130 MeV including the 1912 measles genome. The branch where the 1912 genome roots is

highlighted in red. (B) Maximum clade credibility (MCC) tree summarized from a Bayesian evolutionary inference for 129 full genomes of MeV (excluding the 1912 sample). A time scale is shown at the bottom. The branch where the 1912 genome roots in the unrooted tree is highlighted in red.

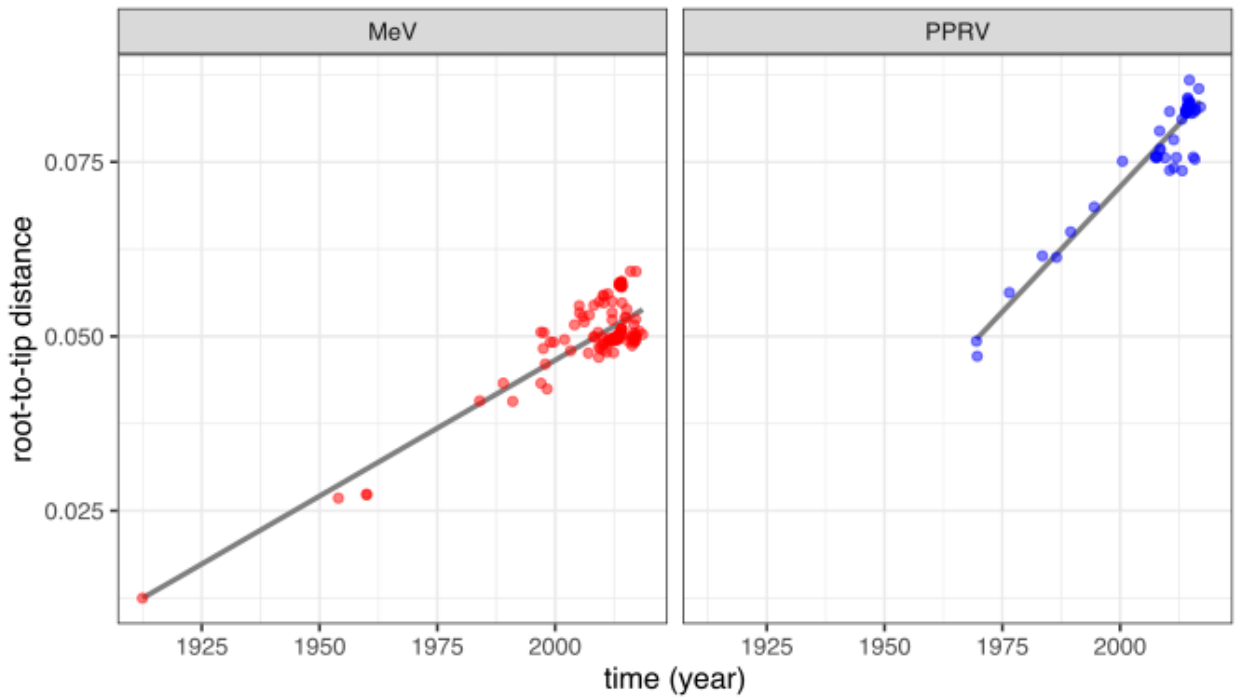

Fig. S5. Correlation between sampling time and root-to-tip divergence for the complete genome dataset of MeV (*n*= 130) in red and PPRV (*n*=65) in blue. Rooting is optimized to minimize residual mean squared distance of the sampling time and root-to-tip divergence correlation in TempEst.

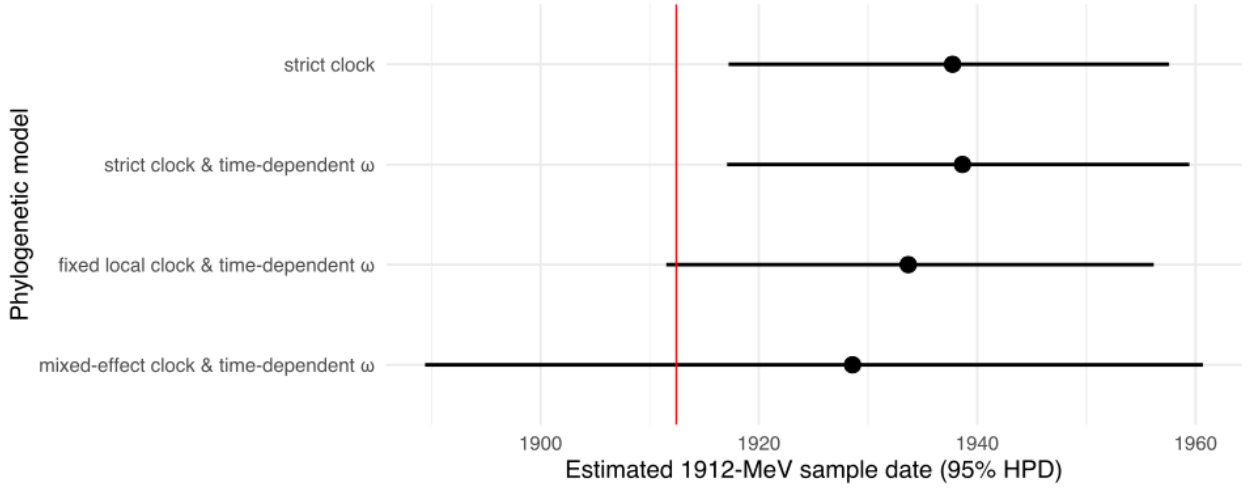

Fig. S6. Estimation of the sampling date of the 1912 genome for the four models. Points represent mean estimates and ranges the estimates' 95% HPD intervals. The red vertical line represents the 1912 genome's real sampling date.

Library	Mapping reference	Median insert size (nt)	Maximum insert size (nt)
#1	Measles virus	95	260
	Human mtDNA	68	178
#2	Measles virus	102	284
	Human mtDNA	71	171
#3	Measles virus	98	227
	Human mtDNA	66	233
#4	Measles virus	114	234
	Human mtDNA	83	311
#5	Measles virus	136	373
	Human mtDNA	143	473

Table S1. Median and maximum insert sizes for reads mapping to the MeV or human mtDNA genome.

GenBank accession number	Date (in decimal year)	Uncertainty (prior)	Virus
PRJEB36265 (1912: ERS4249335)	1912.426	-	MeV
PRJEB36265 (1960: ERS4249363)	1960	1 year (uniform)	MeV
AF266288	1954	1 year (uniform)	MeV
AB481087	1989	1 year (uniform)	MeV
MG912589	1991	1 year (uniform)	MeV
DQ227319	1997	1 year (uniform)	MeV
MG912590	1997.62739726027	-	MeV
MG912594	1997.49315068493	-	MeV
GQ376026	1999.69863013699	-	MeV
KJ755974	1999	1 year (uniform)	MeV
JN635410	2003.28219178082	-	MeV
JN635409	2004.12568306011	-	MeV
JN635406	2008.18852459016	-	MeV
MH356248	2009.10684931507	-	MeV
KY969476	2010.35342465753	-	MeV
KY969479	2013.87945205479	-	MeV
MG912591	1997.91506849315	-	MeV
MH356249	2014.20547945205	-	MeV
MH356237	2016.69672131148	-	MeV
MH356243	2016.77322404372	-	MeV
MH356238	2017.07945205479	-	MeV
MH356240	2017.02191780822	-	MeV
KT588921	2013.53424657534	-	MeV

KC117298	2010.21917808219	-	MeV
MG972194	2017.17534246575	-	MeV
KT732214	2012.1174863388	-	MeV
KT732224	2014.14794520548	-	MeV
KX838946	2016.00819672131	-	MeV
KT732261	2013.55342465753	-	MeV
MH173047	2017.84657534247	-	MeV
X98291	1910	(exponential prior, mean 10 year, offset 57.84658 years)*	RPV
KR781450	1969	1 year (uniform)	PPRV
EU267274	1976	1 year (uniform)	PPRV
KJ466104	2010	1 year (uniform)	PPRV
KU236379	2015.51506849315	-	PPRV
MF741712	2011.95890410959	-	PPRV
KR781449	2011.38356164384	-	PPRV
KR140086	1994	1 year (uniform)	PPRV
JX217850	2008	1 year (uniform)	PPRV
MG581412	2008.3306010929	-	PPRV
KR261605	2014.70684931507	-	PPRV
KY888168	2016.66666666667	-	PPRV
NC_006383	2000	1 year (uniform)	PPRV
KC594074	2008	1 year (uniform)	PPRV
KY885100	2015.8602739726	-	PPRV
KJ867541	2010	1 year (uniform)	PPRV
MF678816	2017	1 year (uniform)	PPRV

KR828813	2013.12328767123	-	PPRV
EU267273	1989	1 year (uniform)	PPRV
KJ867544	1983	1 year (uniform)	PPRV
KM463083	2011.32876712329	-	PPRV

Table S2. Accession numbers of sequences used in the Bayesian phylogenetic analyses performed to date MeV/RPV/PPRV divergence. * This sequence derives from a field isolate from 1910 that was then passaged until the late 1950's (*60*). To model the uncertainty on its age, we used an exponential prior with an offset of 57.84658 years (date of the most recent sequence minus 1960) and a mean of 10.

Data set	MLE(dated tip)	MLE(ultrametric)
MeV	-72,672	-72,837
PPRV	-57,830	-57,872
Morbillivirus	-93,593	-93,636

Table S3. Marginal likelihood estimates (MLE) comparing the fit of a dated-tip model with a free evolutionary rate parameter and an ultrametric model with an arbitrarily fixed rate.

Dataset	NumOTU	Iss	Iss.cSym	Iss.cAsym
MeV	4	0.136	0.860	0.848
	8	0.140	0.845	0.761
	16	0.144	0.854	0.675
	32	0.151	0.819	0.574
PPRV	4	0.130	0.860	0.847
	8	0.130	0.845	0.761
	16	0.138	0.853	0.675
	32	0.152	0.819	0.574
Morbillivirus	4	0.391	0.859	0.847
	8	0.365	0.844	0.761
	16	0.366	0.852	0.675
	32	0.363	0.819	0.572

Table S4: Substitution saturation test (*51, 52*) applied to the three data sets. All differences between the index of substitution saturation (Iss) and the critical values for this index based on computer simulations on a symmetrical (Iss.cSym) and asymmetrical tree topology (Iss.cAsym) were statistically significant. These differences are computed and tested for a subset of 4, 8, 16 and 32 sequences of the data set because 32 represents the maximum number of sequences for

which the simulation-based critical value was determined, while taking care of the proportion of invariable sites, estimated to be 0.57, 0.54 and 0.39 for the MeV, PPRV and Morbillivirus datasets respectively. A statistically significant difference with Iss < Iss.c(Sym/Asym) indicates little saturation.

464

Model	**Parameters**	**Mean**	**95% HPD**
Strict clock	log ω	-2.703	[-2.659;-2.749]
	Clock rate *r*	1.279×10^{-3}	[1.206×10^{-3}; 1.350×10^{-3}]
	Divergence $date_{MeV - RPV}$	825	[738; 909]
	Divergence $date_{PPRV - MeV/RPV}$	308	[185; 430]
Strict clock & time-dependent ω	log ω: Intercept β_0	-1.339	[-1.470, -1.213]
	log ω: Slope β_1	-0.329	[-0.359, -0.301]
	Clock rate *r*	1.246×10^{-3}	[1.176×10^{-3}; 1.323×10^{-3}]
	Divergence $date_{MeV - RPV}$	-159	[-386; 48]
	Divergence $date_{PPRV - MeV/RPV}$	-1721	[-2335; -1197]
Fixed local clock and time-dependent ω	log ω: Intercept β_0	-1.346	[-1.476, -1.217]
	log ω: Slope β_1	-0.329	[-0.360, -0.300]
	Clock rate $r_{MeV/RPV}$	1.135×10^{-3}	[1.048×10^{-3}; 1.218×10^{-3}]
	Clock rate r_{PPRV}	1.666×10^{-3}	[1.526×10^{-3}; 1.805×10^{-3}]
	Divergence $date_{MeV - RPV}$	-420	[-686; -157]
	Divergence $date_{PPRV - MeV/RPV}$	-2353	[-3137; -1651]
Mixed-effect clock	log ω: Intercept β_0	-1.367	[-1.503; -1.227]

and time-dependent ω	log ω: Slope β_1	-0.32	[-0.354; -0.288]
	Clock rate $r_{MeV/RPV}$	1.033×10^{-3}	[0.889×10^{-3}; 1.202×10^{-3}]
	Clock rate r_{PPRV}	1.701×10^{-3}	[1.427×10^{-3}; 2.040×10^{-3}]
	Clock rate *sd*	0.192	[0.135; 0.255]
	Divergence $date_{MeV - RPV}$	-528	[-1174; 165]
	Divergence $date_{PPRV - MeV/RPV}$	-3199	[-4633; -1901]

Table S5. Parameter estimates for the four evolutionary models with increasing complexity. The clock rate (*r*) estimates are in units of codon substitutions per site per year.

Model	1912 age specified		1912 age estimated		1912 genome excluded	
	Mean age	**95% HPD**	**Mean age**	**95% HPD**	**Mean age**	**95% HPD**
Strict clock	825	[738; 909]	843	[754; 929]	845	[763; 938]
Strict clock & time-dependent ω	-159	[-386; 48]	-132	[-345; 71]	-146	[-365; 59]
Fixed local clock & time-dependent ω	-420	[-686; -157]	-387	[-655; -141]	-423	[-709; -158]
Mixed-effects clock & time-dependent ω	-528	[-1174;165]	-518	[-1197;157]	-485	[-1160; 174]

Table S6. Divergence time estimates for the RPV/MeV split with the age of the 1912 genome specified, estimated, and without that genome for the 4 different models

Data Set or Model Specifications	Mean	95% HPD
standard data set	-528	[-1174;165]
different age prior for RPV (Kabete 'O') strain	-759	[-1523; -3]
Additional RPV sample (Fusan)	-735	[-1457; 70]
Conservative 1912 consensus sequence	-490	[-1184; 170]
Skygrid tree prior	-1073	[-2026; -232]

Table S7. Sensitivity of the MeV/RPV divergence date estimate (with 95% HPD) for the mixed-effect clock model with time-dependent ω with respect to: (i) the prior specification on the RPV strain, (ii) an additional RPV strain, the 1912 consensus genome criteria, and a different coalescent tree prior.