

Title: Archival influenza virus genomes from Europe reveal genomic and phenotypic variability during the 1918 pandemic

Authors: Livia Victoria Patrono^{1, 2, †}, Bram Vrancken^{3, †}, Matthias Budt^{4, †}, Ariane D ux^{1, 2}, Sebastian Lequime⁵, Seng l Boral⁶, M. Thomas P. Gilbert^{7, 8}, Jan F. Gogarten^{1, 2}, Luisa Hoffmann⁴, David Horst⁶, Kevin Merkel^{1, 2}, David Morens⁹, Baptiste Prepoint^{2, 10}, Jasmin Schlotterbeck², Verena Schuenemann¹¹, Marc A. Suchard^{12, 13, 14}, Jeffery K. Taubenberger¹⁵, Luisa Tenkhoff⁴, Christian Urban¹¹, Navena Widulin¹⁶, Eduard Winter¹⁷, Michael Worobey¹⁸, Fabian H. Leendertz¹, Thomas Schnalke^{16, †}, Thorsten Wolff^{4, †}, Philippe Lemey^{3, †}, S bastien Calvignac-Spencer^{1, 2, †*}

Affiliations:

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

²Viral Evolution, Robert Koch Institute, Berlin, Germany.

³Laboratory of Clinical and Evolutionary Virology, Department of Microbiology, Immunology and Transplantation, Rega Institute, Katholieke Universiteit Leuven, Leuven, Belgium.

⁴Unit 17 Influenza and other Respiratory Viruses, Robert Koch Institute, Berlin, Germany.

⁵Cluster of Microbial Ecology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands.

⁶Institute for Pathology, Charit , Berlin, Germany.

⁷Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark.

⁸University Museum, NTNU, Trondheim, Norway.

⁹ Office of the Director, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland.

¹⁰D partement de Biologie, Ecole Normale Sup rieure, PSL Universit  Paris, Paris, France.

¹¹Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland.

¹²Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA.

¹³Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.

¹⁴Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.

¹⁵Viral Pathogenesis and Evolution Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland.

¹⁶Berlin Museum of Medical History, Charit , Berlin, Germany.

¹⁷Pathological-anatomical collection in the Narrenturm, Natural History Museum of Vienna, Vienna, Austria.

¹⁸Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona.

*Correspondence to: calvignacs@rki.de

†These authors contributed equally to this work

‡ These authors contributed equally to this work

5

Abstract:

The 1918 influenza pandemic was the deadliest respiratory pandemic of the 20th century and determined the genomic make-up of subsequent human influenza A viruses (IAV). Here, we analyze the first 1918 IAV genomes from Europe and from the first, milder wave of the pandemic. 1918 IAV genomic diversity is consistent with local transmission and frequent long-distance dispersal events and *in vitro* polymerase characterization suggests potential phenotypic variability. Comparison of first and second wave genomes shows variation at two sites in the nucleoprotein gene associated with resistance to host antiviral response, pointing at a possible adaptation of 1918 IAV to humans. Finally, phylogenetic estimates based on extended molecular clock modelling suggests a pure pandemic descent of seasonal H1N1 IAV as an alternative to the hypothesis of an intrasubtype reassortment origin.

15

One Sentence Summary:

Much can be learned about past pandemics by uncovering their footprints in medical archives, which we here demonstrate for the 1918 flu pandemic.

20

Main Text:

As the COVID-19 pandemic unfolds (1), we are facing many questions regarding the emergence and spread of a novel pathogen in human populations. Although every pandemic is unique and making meaningful comparisons is challenging, the current situation naturally rekindles interest in past pandemics (2).

25

The 1918 influenza A (H1N1) pandemic (hereafter 1918 pandemic) was the largest global catastrophe of infectious origin to affect humankind in the last century. The disease proceeded in three waves: a mild one in the spring of 1918, followed by deadlier waves in the 1918 fall and 1919 winter (3). Beyond mortality records from that period that allow for comparisons of the scale of this pandemic with subsequent ones, very little is known about even basic biological

30

consequences of the fast and global spread of the 1918 virus. For example, it remains essentially unknown how much genomic diversity arose during the 1918 pandemic and how this diversity impacted viral phenotypes.

In the late 1990s, molecular analyses of formalin-fixed, paraffin-embedded tissue blocks and permafrost-preserved bodies from victims of the 1918 pandemic (4) ascertained that the causative agent was an influenza A virus (IAV) of the H1N1 subtype (5). The reconstruction of two complete IAV genomes from Brevig Mission, Alaska (hereafter BM) (6) and Camp Upton, New York (hereafter CU) (7), as well as the identification of a few other partial sequences (8, 9) informed evolutionary and functional investigations. These revealed that the 1918 virus harbored gene segments drawn from the diversity of IAV strains circulating in an avian reservoir (10), and that the hemagglutinin (HA) and polymerase complex genes were major determinants of its pathogenicity (11, 12). However, our knowledge about an agent that caused an estimated 50-100 million deaths (13) remains extremely limited, notably because it relies on a remarkably sparse corpus of genomic/genetic data, the majority stemming from North America.

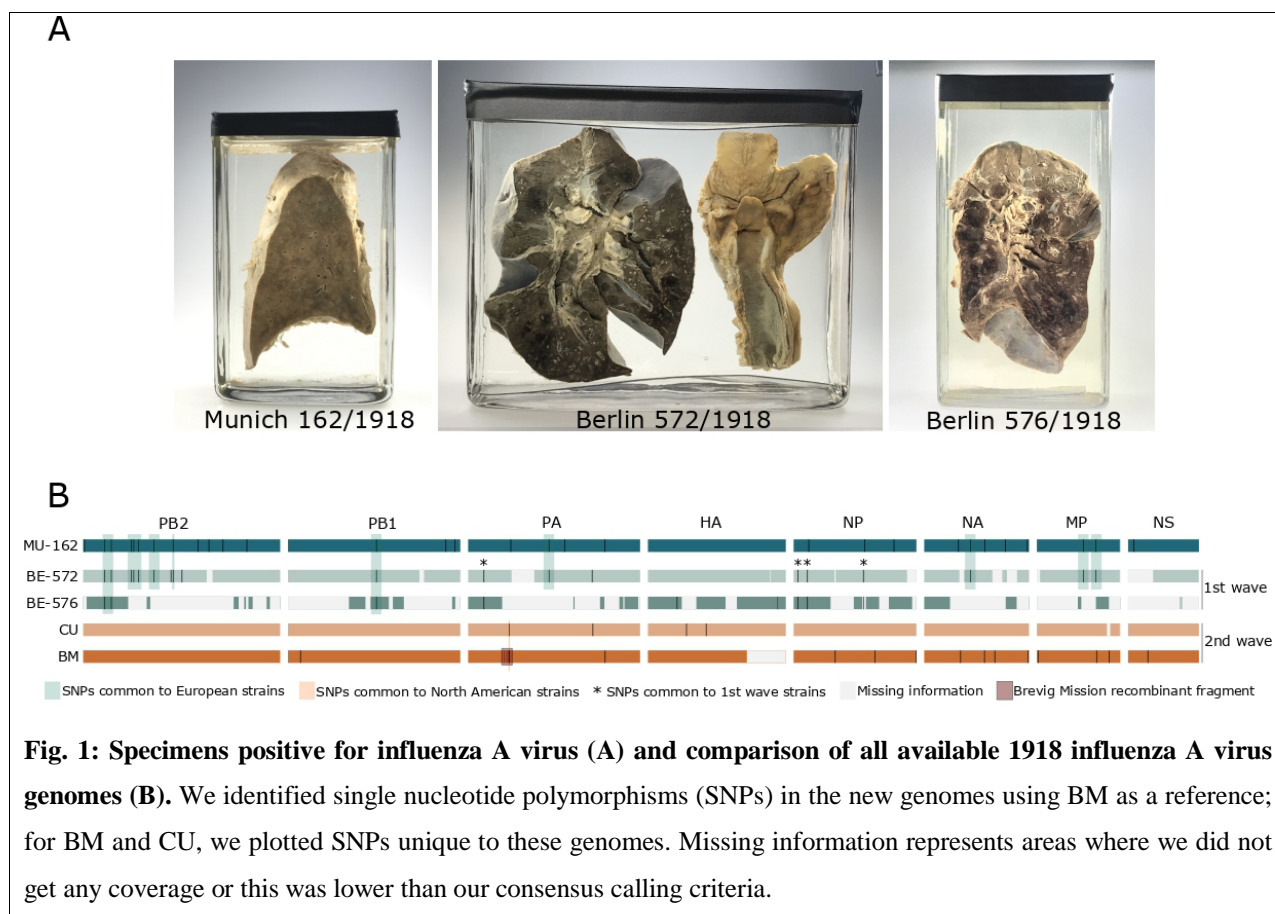
We have started to meet this gap by applying recent advances in nucleic acid recovery from archival samples (7, 14) to as-yet unexplored European pathology collection material, including the medical archive started by Rudolf Virchow in mid 1800s Germany (15). This enabled us to sequence complete and partial 1918 influenza virus genomes from specimens sampled in two German cities, then use these data to provide new insights into the genomic and phenotypic diversity of the pandemic strains.

1918 influenza virus genomes from Munich and Berlin

To explore the emergence, pandemic and post-pandemic periods, we selected 13 formalin-fixed lung specimens dated between 1900 and 1931, preserved within the Berlin Museum of Medical History at the Charité (Berlin, Germany, n=11) and the pathology collection (Narrenturm) of the Natural History Museum in Vienna, Austria (n=2). This set included six specimens collected during pandemic years in Europe (n=4 in 1918, n=2 in 1919). Details about all specimens, including initial diagnosis, are available in the supplementary information. For each specimen, we heat treated 200 mg of formalin-fixed lung tissue to reverse macromolecule cross-links induced by formalin, then performed nucleic acid extraction. Following DNase treatment and

ribosomal RNA depletion, we built high-throughput sequencing libraries and shotgun sequenced them on Illumina® platforms.

We identified IAV reads in libraries from 3 of 13 specimens, all of which date to 1918 and were associated with histopathological findings of bronchopneumonia (**Fig. 1A; supplementary text S1 and Fig. S1**). Viral RNA preservation was generally good, as determined by median fragment lengths well above 100 nucleotides and the lack of any obvious damage pattern (**Fig. S9-S11**). Human RNA fragments were on average shorter, possibly indicating better protection of encapsidated viral RNA (**supplementary text S2**). We did not detect any other viral agent likely to be involved in severe respiratory disease, but identified a number of bacteria which have been previously associated with respiratory diseases (e.g. *Pseudomonas aeruginosa* and *Mycobacterium kansasii* in specimens that did not produce influenza reads) and with 1918 influenza (*Staphylococcus aureus*, *Streptococcus pneumoniae*, *Klebsiella pneumoniae*, *Pasteurella multocida*); in 2 specimens bacterial colonization was also evident from histopathology (**supplementary text S1 and Fig. S1**).



Increased sequencing efforts for the three influenza-positive samples MU-162 (Munich), BE-572 and BE-576 (Berlin) yielded 40,133,161, 31,989,479 and 14,965,377 high quality reads, respectively. Of these, 55.5%, 0.33% and 0.1% mapped to the influenza virus strain A/Brevig Mission/1/1918, which allowed us to reconstruct a complete 1918 influenza virus genome at a 1944x average coverage depth for MU-162, as well as significant portions of the genomes for BE-572 and BE-576 (89.3% and 57.2%, respectively).

Genomic comparison and phylodynamics of 1918 influenza viruses

Together with the available BM and CU sequences, we used these influenza genomes from Germany to assess the genomic diversity of: (i) strains simultaneously involved in local transmission (BE-572 and BE-576), (ii) strains circulating in Europe (MU-162, BE-572 and BE-576) and North America (BM and CU), (iii) strains circulating during the first (BE-572 and BE-576) and second wave (BM and CU) of the pandemic.

The two Berlin genomes sampled on June 28th 1918, for which only a portion of the genome could be compared, differed from each other at only two (non-synonymous) nucleotide positions in the HA gene (**Fig. 1B**). These positions were however polymorphic in BE-576, with the minor variant being identical to BE-572. At a country/continent scale, BE-572 and MU-162 on one hand and BM and CU on the other differed both by 22 and 15 single nucleotide polymorphisms (SNPs; 11 non-synonymous, all genes but HA, NS and MP and 7 non-synonymous, all genes but NP and PB2, respectively, **Table 1 and S4; Fig. 1B**). The four pairwise comparisons of European and North American strains identified 22-43 SNPs, with the two best genomes (MU-162 and BM) differing at 43 positions, 16 of which coding for amino acid changes (all genes but HA). When comparing the first wave European strain BE-572 with second wave strains BM and CU, we identified 29 and 22 SNPs (11 non-synonymous, all genes but HA and NS genes, and 8 non-synonymous, all genes but NS genes, respectively).

We could not fully disentangle these comparisons (the two Berlin genomes are not complete, and the two accurately dated genomes from the first wave are from European specimens, while the two accurately dated genomes from the second wave are from North America) and our sample size was very small. However, keeping these limitations in mind, it appears that within-outbreak variability in Berlin was negligible, while genomes sampled on the

same continents (0.11-0.16%) and during the same wave (0.11%) exhibited lower overall divergence than genomes sampled on different continents (0.16-0.32%) and during different waves (0.16-0.21%). This pattern is compatible with spatio-temporal effects of local transmission.

5

	BM	CU	MU-162	BE-572
BM	-			
CU	15 (11)	-		
MU-162	43 (36)	33 (30)	-	
BE-572	29	22	22	-

Table 1. Number of differences between the highest quality genomes. Numbers in () represent the differences counted in the positions covered by BE-572 (12023 nt). BE-576 was not included in these comparisons because of its lower quality. A 69 nt fragment in PA with a signal of recombination (8 SNPs, 1 coding) was excluded from this comparison.

10

To further investigate the question of 1918 virus spread, we took advantage of the availability of a total of 21 HA sequences sampled in Europe and North America during all pandemic waves to infer a time-scaled evolutionary history in a Bayesian phylogenetic framework (16) This reconstruction demonstrated an interspersed clustering of the three German sequences with North American strains, suggesting no geographic segregation between continents (**Fig. 2**). Similarly, two sequences sampled three months apart in London during the 1918-1919 winter waves of the pandemic (November 1918 and February 1919, respectively) clustered separately, albeit with low support (posterior probability .75 and .73, **Fig. 2 and S7**). Models in which monophyletic clustering was enforced on the European lineages showed a decisively lower model fit (ln Bayes factor [BF] ≥ 8.7 , **supplementary text S6**), supporting the hypothesis of extensive, and most likely bi-directional, transatlantic movement of pandemic strains, in line with the historical context of human movement near the end of World War I. Altogether, genomic data support a scenario of predominantly local transmission with frequent long-distance dispersal events.

15

20

25

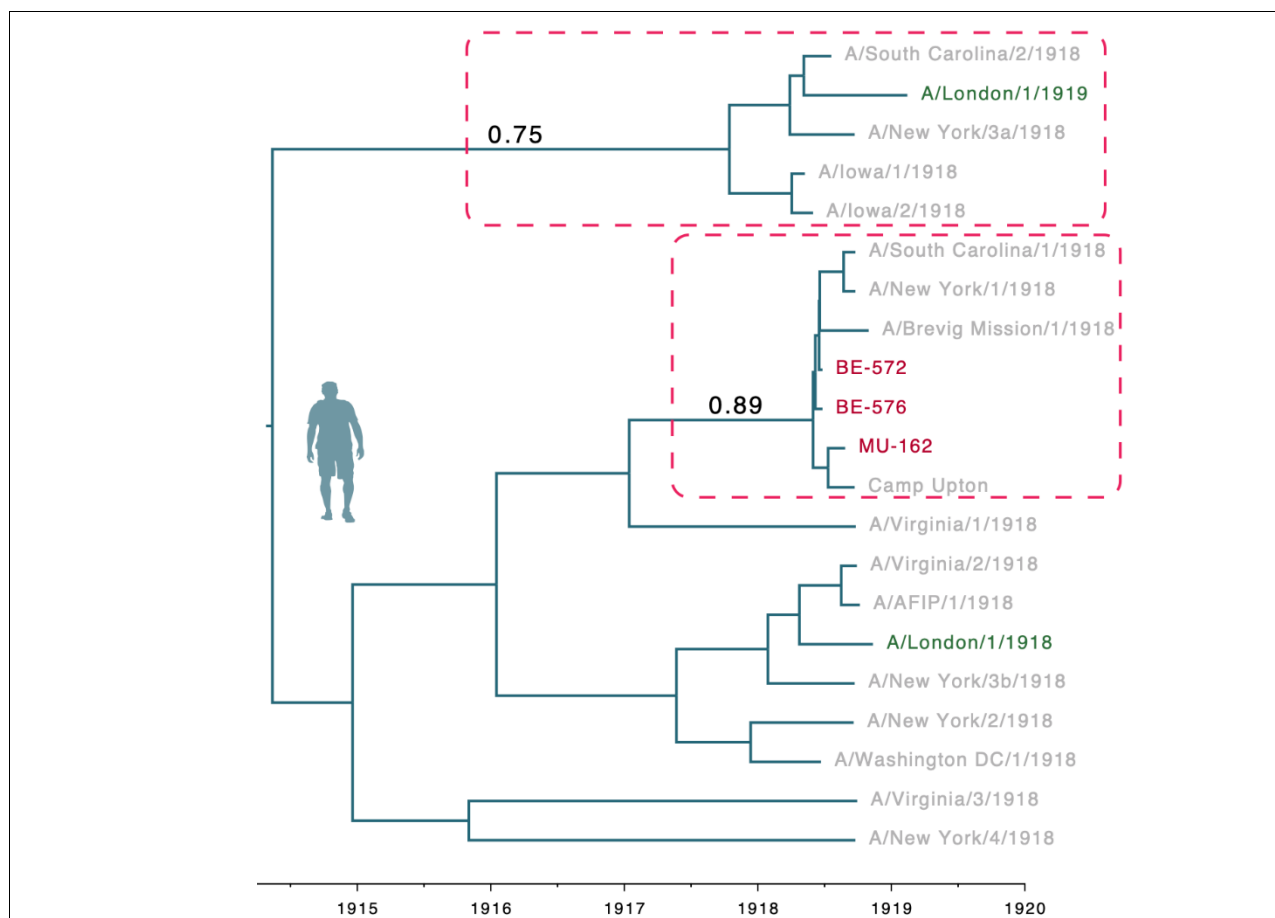


Fig. 2: Maximum clade credibility time tree estimated for HA sequences from 1918 flu strains. US strains are in light grey pink, European strains are in dark red (Germany) and dark green (UK). Dashed rectangles highlight clades comprising strains from different continents; for these clades posterior probabilities are reported above stem branches. The BE-576 consensus genome with ambiguities was used in this analysis; an equivalent reconstruction using majority rule consensus base calling can be found in Fig. S7.

Finally, we identified a 69 nt stretch of the PA of BM comprising 8 SNPs that shows a high degree of identity to IAV strains circulating in 1933 or later (**Table S4**). Given the size of this fragment and that of the fragments initially targeted to characterize the BM PA sequence (17), this apparent recombination may reflect the contamination of one of the PCR reactions that allowed for the reconstruction of the BM genome with a PCR product derived from a later influenza A virus strain. However, we cannot formally exclude that the BM PA fragment is a genuine but rare case of natural recombination (18) or that it represents a sequencing artefact due to intrahost variation. Resequencing of the BM genome using PCR-independent methods should clarify this issue.

Potential signatures of avian-to-human adaptation in 1918 influenza viruses

In light of the hypothesis of an avian origin of the pandemic virus, we investigated the presence of amino acid (aa) signatures known or suspected to be associated with avian-to-human adaptation, either because they have been characterized functionally or found to be distributed differentially across bird- and human-infecting influenza viruses. We first examined the high quality (but imprecisely dated) genome derived from MU-162 and identified two such positions. In PB2, we found a M631L aa change within the PB2 627 host range domain. Although this mutation has recently been described as a main mediator of adaptation and lethality of an avian influenza virus in mice (19), all but one human H1N1 strains (including all other 1918 influenza viruses) present a methionine at this position, suggesting that this change did not have profound evolutionary implications. On the contrary, we found that MU-162 has a leucine (L) at position 61 in NP, which is the residue most commonly found in human H1N1 strains prior to the 2009pdm (**Fig. S12**), whereas an overwhelming majority of avian influenza A viruses present an isoleucine (I) at this position (20). Interestingly, MU-162 was the only 1918 influenza virus presenting the I61L change, indicating this mutation only reached fixation after the first two waves of the pandemic.

We then considered potential coding variation between first and second wave strains. The second wave of the 1918 pandemic was indeed much more severe than the first, and potential adaptations of the virus have been proposed as one possible explanation, among others (3). The presence of the aa residue G222 in the receptor binding domain of the H1 subtype HA protein has already been discussed in this context (8). This residue confers a binding affinity for both avian and human glycans, while the human-like D222 only efficiently binds human glycans (21, 22). In a recent study, three of four first wave strain sequences presented G222 while only two of nine second wave strains did (all other strains presented D222) (8). The three German HAs sequenced here carried the human-like D222, which reduces the likelihood that the frequency of the two variants varied significantly between the first two pandemic waves.

The availability of genome-wide information about two first wave strains also allowed us to identify two other aa changes that potentially mark a difference between first and second wave strains. We detected these variations at sites of the NP known to influence host range and susceptibility/resistance to the interferon-induced MxA antiviral protein (20, 23, 24): first wave

strains BE-572 and BE-576 carried avian-like residues at position 16 (G) and 283 (L), whereas the fall strains and MU-162 carried D16 and P283 found in most human H1N1 strains with the exception of the 2009 pandemic H1N1 virus (see references (23–26) and **Fig. S12**). Position 283 is located in the body of the NP and presence of a proline at this site confers resistance to the human MxA protein, to which D16 also contributes, albeit to a lower extent (24, 26). In addition to supporting the hypothesis of an avian origin of this gene, these mutations may represent hallmarks of early adaptation to humans: during the first months of the pandemic, 1918 influenza viruses may have evolved a better capacity to evade the innate interferon response, which is an important aspect of influenza virus pathogenicity.

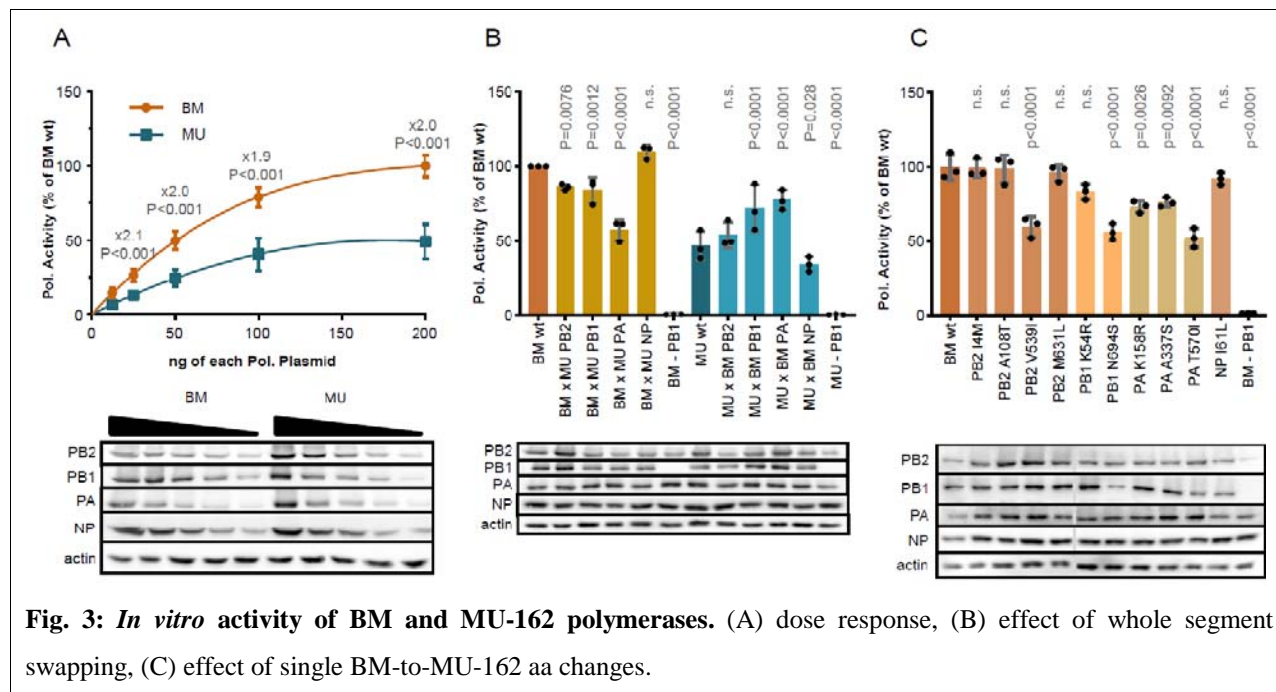
Comparison of *in vitro* polymerase activity of 1918 influenza viruses

To start exploring the phenotypic impact of the genomic differences observed *in vitro*, we focused on the MU-162 strain for which we obtained a high-quality complete genome sequence. Since the polymerase complex contained most of the coding mutations found in MU-162 (4 in PB2, 2 in PB1, 3 in PA and 1 in NP proteins), and is a major pathogenicity factor for BM, we resynthesized viral genes and performed a functional comparison of the activity of the BM and MU-162 polymerases after reconstitution in transfected cells. Using reporter assays, dose response curves revealed a two-fold higher activity of BM compared to MU-162 polymerase (**Fig. 3A**).

To identify the subunit(s) responsible for this difference, we determined the effect of swapping single polymerase subunits between BM and MU-162 (**Fig. 3B**). The introduction of PA from MU-162 into the BM polymerase complex caused a 1.7-fold reduction of its activity (**Fig. 3B**, $P < 0.0001$). In line with this finding, BM PA increased the activity of the MU-162 polymerase by 1.7-fold ($P < 0.0001$). Exchange of PB1 (1.2-fold reduction for BM x MU-162 PB1, $P = 0.0012$ and 1.5-fold increase for MU-162 x BM PB1, $P < 0.0001$, respectively) and PB2 (1.2-fold decrease for BM x MU-162 PB1, $P = 0.0076$ and no significant change for MU-162 x BM PB1, respectively) subunits caused a similar, but more modest impact on activity. No effect was recorded upon exchange of NP proteins.

To more precisely pinpoint the variations responsible for the observed differences, we generated BM point mutants, each carrying one of the aa changes identified in the MU-162 strain. All three PA mutants significantly reduced the activity of the polymerase compared to

BMwt (1.4-fold ($P=0.0026$) for K158R, 1.3-fold, ($P=0.0092$) for A337S and 1.9-fold ($P<0.0001$) for T570I, respectively) (Fig. 3C). Of these, K158R was located in the region identified as being of uncertain origin in BM. In PB1, the N694S mutation apparently reduced polymerase activity by 1.8-fold ($P<0.0001$), which might however be explained by a reduced protein expression level. The V539I aa change in PB2 caused a similar 1.7-fold reduction ($P<0.0001$) whereas the other three aa changes in this gene did not (including the aforementioned M631L change). These results indicate that, *in vitro*, amino acid polymorphisms encoded in MU-162 affect the activity of the viral RNA polymerase complex. Although this does not necessarily predict functional differences *in vivo*, it is, at this stage, compatible with the notion of phenotypically polymorphic 1918 influenza viruses.



1918 influenza viruses and the origin of human seasonal H1N1 influenza viruses

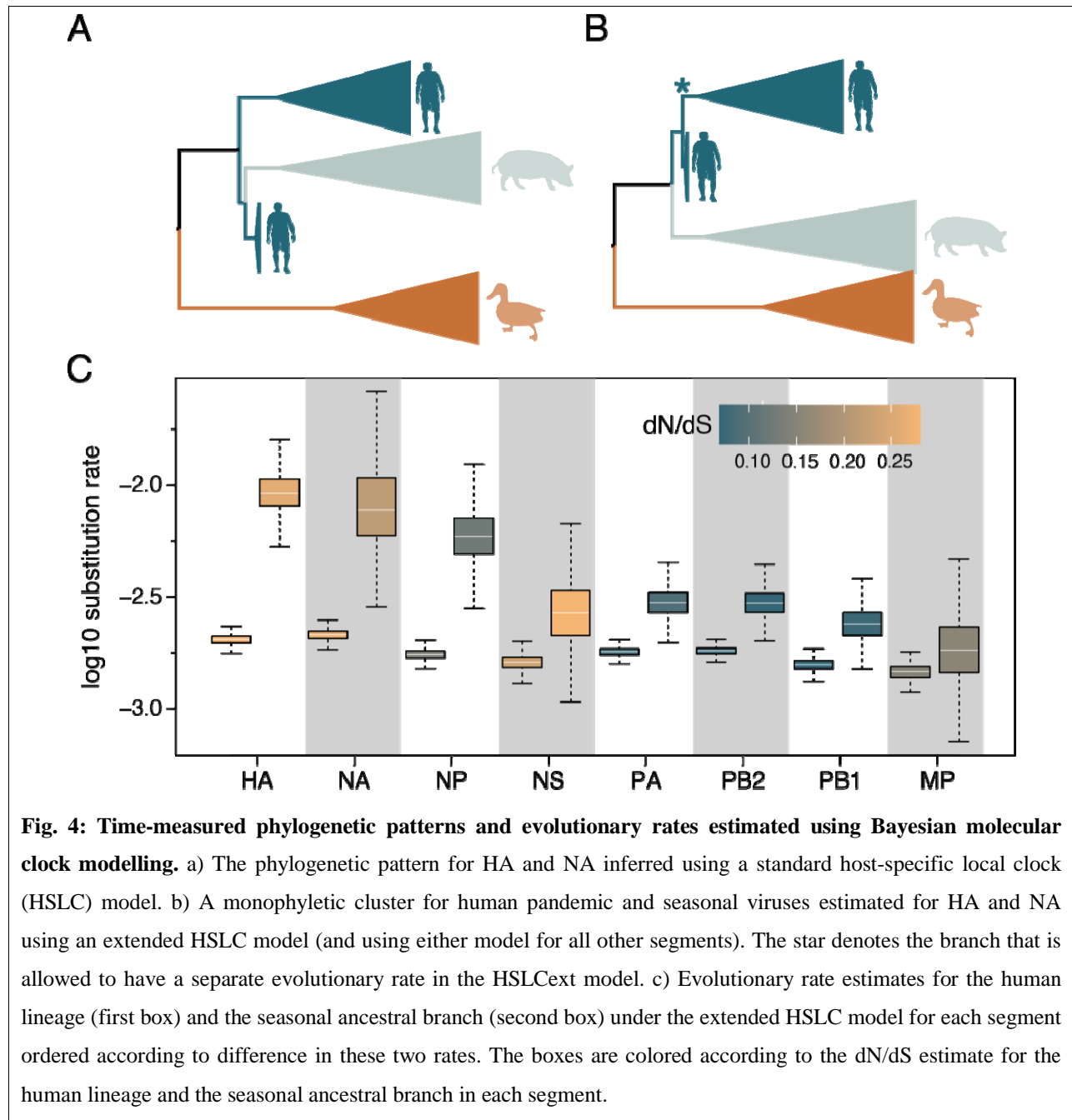
It had long been assumed that seasonal H1N1 IAVs descended (all or in part) from 1918 influenza viruses. The sequencing of the first segments and then genomes of 1918 strains showed their very close phylogenetic relationships with seasonal H1N1 viruses over all segments: non-clock tree reconstructions suggested that the latter directly descended from the former (17). However, the development and application of molecular clock models allowing for host-specific rates of nucleotide evolution (thereafter host-specific local clocks; HSLC) supported an

alternative scenario, whereby seasonal H1N1 viruses did not evolve directly from the pandemic virus in HA. These models rather suggested that the HA of seasonal IAVs was acquired from a co-circulating homosubtypic H1 IAV through intrasubtype reassortment (10). Intriguingly, this scenario is also compatible with other non-genetic information. First, serological studies provided strong evidence that an H3 IAV circulated prior to about 1900 but weak evidence for H3 circulation from ~1900-1918 (27–29). Then, the fact that individuals born after 1900 were spared severe outcomes compared to young adults 20-40 years of age was compatible with an initial exposure to H1 viruses better matched to the 1918 pandemic viruses than the putative H3 viruses to which young adults were first exposed to (10). We revisited these conflicting hypotheses about the origin of human seasonal influenza using our extended sequence data set and first used the same inference approaches as previously applied. Non-clock maximum likelihood (ML) reconstructions indicated that human seasonal H1N1 and 1918 pandemic viruses cluster together with reasonably high bootstrap support, with the seasonal lineage nested within the 1918 pandemic variants for all segments but PB2 (**Fig. S2**). Time-measured phylogenetic inference using a host-specific local clock (HSLC) model broadly recovers previously reconstructed phylogenetic patterns for all segments but NP. For all segments but HA and NA the human seasonal lineage nested within pandemic flu diversity, while for HA and NA pandemic viruses form a monophyletic cluster with swine influenza viruses, which is positioned as a sister lineage to humans seasonal lineage (10) (**Fig. 4a vs b, Fig. S3**); this pattern was previously also observed in NP, but not in our reconstructions.

Because of the persisting conflict between the difference inference approaches for HA and NA (**Fig. 4a, S2 and S3A**), we explored an alternative scenario that allows for a higher evolutionary rate on the branch ancestral to the seasonal lineage. Such a scenario would result in considerably higher divergence between pandemic and seasonal viruses for H1 and N1 than would be expected under a strict clock, and could therefore induce a sister lineage pattern with a relatively deep most recent common ancestor (MRCA) under the HSLC model - which assumes a different but constant rate of evolution in each of the host-specific lineages. We ran simulations that showed that if the rate is elevated on the relevant branch, and this is not accounted for in the clock model, the HSLC inference generally fails to recover the seasonal lineage as a direct descendent of the pandemic virus (**Fig. S5**). Fitting an extended HSLC model that allows for a separate rate on the relevant branch inferred the seasonal lineage as a descendant of the

pandemic virus in both HA and NA, consistent with the non-clock trees. Across segments, this model resulted in consistently higher posterior mean rates on the seasonal ancestral branch compared to the seasonal clade rates (**Fig. 4**), but with a variability that largely reflected the variation in selective constraints on the segments. Less stringent purifying selection in HA and NA accompanied by a stronger rate acceleration potentially explains the observed patterns (**Fig. 4c**). To investigate whether variation in selection pressures could explain the acceleration on the branch leading to the seasonal lineage, we performed selection analyses using codon substitution models. These models did not identify diversifying episodic selection or relaxed selective constraints on this branch in any of the segments, implying that only mutation rate/generation time effects may explain the branch-specific elevated rate. Interestingly, a higher mutation rate for the seasonal predecessor (due to a higher replication rate) has indeed been suggested by some experimental evidence comparing the Brevig Mission strain to seasonal strains (30, 31).

Altogether, these new analyses revive the scenario of a pure pandemic origin of seasonal H1N1 viruses. However, the essentially phenomenological nature of our modelling approach does not, for now, allow us to definitely favor it over the alternative scenario of a homosubtypic reassortment.



5

10

Conclusion

We derived considerable insight from the addition of only one complete and two partial 1918 flu genomes. Our analyses show significant genomic variation whose spatial distribution suggests frequent long-distance dispersal events, identify potentially adaptive substitutions in NP between first and second wave viruses, hint at viral polymerase phenotypic variation and rekindle the scenario of a pandemic origin for all segments of subsequent seasonal H1N1 influenza viruses.

15

We acknowledge that these findings remain preliminary. Our sample of genomic diversity is extremely small (four good quality genomes) and *in vitro* phenotypic characterization cannot accurately predict *in vivo* phenotypes. Additional genomes from archival samples from various time points starting in 1918, as well as their phenotypic characterization *in vitro* and *in vivo*, will undoubtedly provide the opportunity for more robust tests of our hypotheses. As 3 out of the 4 archival samples from 1918 analyzed yielded good quality viral RNA, influenza genomic research using medical collections can now probably be considered a low risk-high gain enterprise. Therefore, we anticipate that the main obstacle to a better understanding of the evolution of 1918 flu viruses will be the identification of surviving pathological specimens, which highlights the importance of long-neglected museum activities (32).

References and Notes:

1. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China. *Nature*. **579**, 265–269 (2020).
2. D. M. Morens, P. Daszak, H. Markel, J. K. Taubenberger, Pandemic COVID-19 Joins History's Pandemic Legion. *MBio*. **11**, 1–9 (2020).
3. J. K. Taubenberger, D. M. Morens, 1918 Influenza: The mother of all pandemics. *Emerg. Infect. Dis.* **12**, 15–22 (2006).
4. J. K. Taubenberger, A. H. Reid, A. E. Krafft, K. E. Bijwaard, T. G. Fanning, Initial genetic characterization of the 1918 “Spanish” influenza virus. *Science (80-.)*. **275**, 1793–1796 (1997).
5. A. H. Reid, T. G. Fanning, J. V. Hultin, J. K. Taubenberger, Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1651–1656 (1999).
6. J. K. Taubenberger, J. C. Kash, D. M. Morens, The 1918 influenza pandemic: 100 years of questions answered and unanswered. *Sci. Transl. Med.* **11** (2019), doi:10.1126/scitranslmed.aau5485.
7. Y. L. Xiao, J. C. Kash, S. B. Beres, Z. M. Sheng, J. M. Musser, J. K. Taubenberger, High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *J. Pathol.* **229**, 535–545 (2013).
8. Z. M. Sheng, D. S. Chertow, X. Ambroggio, S. McCall, R. M. Przygodzki, R. E. Cunningham, O. A. Maximova, J. C. Kash, D. M. Morens, J. K. Taubenberger, Autopsy series of 68 cases dying before and during the 1918 influenza pandemic peak. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16416–16421 (2011).
9. A. H. Reid, T. A. Janczewski, R. M. Lourens, A. J. Eillot, R. S. Daniels, C. L. Berry, J. S. Oxford, J. K. Taubenberger, 1918 Influenza pandemic caused by highly conserved viruses with two receptor-binding variants. *Emerg. Infect. Dis.* **9**, 1249–1253 (2003).
10. M. Worobey, G. Z. Han, A. Rambaut, Genesis and pathogenesis of the 1918 pandemic

- H1N1 influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8107–8112 (2014).
11. T. Watanabe, J. Tisoncik-Go, N. Tchitchek, S. Watanabe, A. G. Benecke, M. G. Katze, Y. Kawaoka, 1918 Influenza Virus Hemagglutinin (HA) and the Viral RNA Polymerase Complex Enhance Viral Pathogenicity, but Only HA Induces Aberrant Host Responses in Mice. *J. Virol.* **87**, 5239–5254 (2013).
12. T. Watanabe, Y. Kawaoka, Pathogenesis of the 1918 pandemic influenza virus. *PLoS Pathog.* **7**, 5–8 (2011).
13. N. P. A. S. Johnson, J. Mueller, Updating the accounts: global mortality of the 1918-1920 “Spanish” influenza pandemic. *Bull. Hist. Med.* **76**, 105–115 (2002).
14. A. M. Devault, G. B. Golding, N. Waglechner, J. M. Enk, M. Kuch, J. H. Tien, M. Shi, D. N. Fisman, A. N. Dhody, S. Forrest, K. I. Bos, D. J. D. Earn, E. C. Holmes, H. N. Poinar, Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N. Engl. J. Med.* **370**, 334–340 (2014).
15. T. Schnalke, in *Wissenschaft im Museum – Ausstellung im Labor* (2014), pp. 18–38.
16. 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**, 1–5 (2018).
17. J. K. Taubenberger, A. H. Reid, R. M. Lourens, R. Wang, G. Jin, T. G. Fanning, Characterization of the 1918 influenza virus polymerase genes. *Nature.* **437**, 889–893 (2005).
18. G. Z. Han, M. Worobey, Homologous recombination in negative sense RNA viruses. *Viruses.* **3**, 1358–1373 (2011).
19. X. Zhang, G. Xu, C. Wang, M. Jiang, W. Gao, M. Wang, H. Sun, Y. Sun, K. C. Chang, J. Liu, J. Pu, Enhanced pathogenicity and neurotropism of mouse-adapted H10N7 influenza virus are mediated by novel PB2 and NA mutations. *J. Gen. Virol.* **98**, 1185–1195 (2017).
20. G. W. Chen, S. C. Chang, C. K. Mok, Y. L. Lo, Y. N. Kung, J. H. Huang, Y. H. Shih, J. Y. Wang, C. Chiang, C. J. Chen, S. R. Shih, Genomic signatures of human versus avian influenza A viruses. *Emerg. Infect. Dis.* **12**, 1353–1360 (2006).
21. J. Stevens, O. Blixt, L. Glaser, J. K. Taubenberger, P. Palese, J. C. Paulson, I. A. Wilson, Glycan Microarray Analysis of the Hemagglutinins from Modern and Pandemic Influenza Viruses Reveals Different Receptor Specificities. *J. Mol. Biol.* **355**, 1143–1155 (2006).
22. A. Srinivasan, K. Viswanathan, R. Raman, A. Chandrasekaran, S. Raguram, T. M. Tumpey, V. Sasisekharan, R. Sasisekharan, *Proc. Natl. Acad. Sci.*, in press, doi:10.1073/pnas.0711963105.
23. C. Pan, B. Cheung, S. Tan, C. Li, N. Li, S. Liu, S. Jiang, Genomic signature and mutation trend analysis of pandemic (H1N1) 2009 influenza A virus. *PLoS One.* **5**, 1–7 (2010).
24. B. Mänz, D. Dornfeld, V. Götz, R. Zell, P. Zimmermann, O. Haller, G. Kochs, M. Schwemmler, Pandemic Influenza A Viruses Escape from Restriction by Human MxA through Adaptive Mutations in the Nucleoprotein. *PLoS Pathog.* **9** (2013), doi:10.1371/journal.ppat.1003279.
25. J. Tu, H. Zhou, T. Jiang, C. Li, A. Zhang, X. Guo, W. Zou, H. Chen, M. Jin, Isolation and molecular characterization of equine H3N8 influenza viruses from pigs in China. *Arch. Virol.* **154**, 887–90 (2009).
26. O. Ashenberg, J. Padmakumar, M. B. Doud, J. D. Bloom, Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLoS Pathog.* **13**, 1–23 (2017).

27. L. Simonsen, T. A. Reichert, M. A. Miller, The virtues of antigenic sin: consequences of pandemic recycling on influenza-associated mortality. *Int. Congr. Ser.* **1263**, 791–794 (2004).
28. F. M. Davenport, E. Minuse, A. V Hennessy, T. J. Francis, Interpretations of influenza antibody patterns of man. *Bull. World Health Organ.* **41**, 453–460 (1969).
29. N. Masurel, J. Mulder, Studies on the content of antibodies for equine influenza viruses in human sera. *Bull. World Health Organ.* **34**, 885–893 (1966).
30. T. M. Tumpey, C. F. Basler, P. V Aguilar, H. Zeng, A. Solórzano, D. E. Swayne, N. J. Cox, J. M. Katz, J. K. Taubenberger, P. Palese, A. García-Sastre, Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science.* **310**, 77–80 (2005).
31. T. Watanabe, J. Tisoncik-Go, N. Tchitchek, S. Watanabe, A. G. Benecke, M. G. Katze, Y. Kawaoka, 1918 Influenza virus hemagglutinin (HA) and the viral RNA polymerase complex enhance viral pathogenicity, but only HA induces aberrant host responses in mice. *J. Virol.* **87**, 5239–5254 (2013).
32. Deutscher Wissenschaftsrat, “Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastruktur” (2011), (available at <https://www.wissenschaftsrat.de/download/archiv/10464-11.html>).
33. M. T. P. 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** (2007), doi:10.1371/journal.pone.0000537.
34. L. Wu, N. Patten, C. T. Yamashiro, B. Chui, Extraction and amplification of DNA from formalin-fixed, paraffin-embedded tissues. *Appl. Immunohistochemistry Mol. Morphol.* **10**, 269–74 (2002).
35. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* **30**, 2114–2120 (2014).
36. D. E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 1–13 (2019).
37. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754–60 (2009).
38. P. P. Nurk S, Meleshko D, Korobeynikov A, metaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res.* **1**, 30–47 (2017).
39. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Kniep, J. Krause, K. Nieselt, EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
40. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* **29**, 1682–1684 (2013).
41. M. Worobey, G. Z. Han, A. Rambaut, A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature.* **508**, 254–257 (2014).
42. P. Villesen, FaBox: An online toolbox for FASTA sequences. *Mol. Ecol. Notes.* **7**, 965–968 (2007).
43. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, 1–5 (2015).
44. G. O. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. - PubMed - NCBI. *Syst. Biol.* **29**, 307–21 (2010).

45. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
46. A. Larsson, AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* **30**, 3276–3278 (2014).
- 5 47. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, E. Teeling, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
48. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences. *Am. Math. Soc. Lect. Math. Life Sci.* **17** (1986), pp. 57–86.
- 10 49. B. Q. Minh, M. A. T. Nguyen, A. Von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
50. M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
- 15 51. M. Worobey, G. Han, A. Rambaut, A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature.* **508**, 254–257 (2014).
52. A. Düx, S. Lequime, L. V. Patrono, B. Vrancken, S. Boral, J. F. Gogarten, A. Hilbig, D. Horst, K. Merkel, B. Prepoint, S. Santibanez, J. Schlotterbeck, M. A. Suchard, M. Ulrich, N. Widulin, A. Mankertz, F. H. Leendertz, K. Harper, T. Schnalke, P. Lemey, S. Calvignac-Spencer, *Science (80-.)*, in press, doi:10.1126/science.aba9411.
- 20 53. 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. *Syst. Biol.* **68**, 1052–1061 (2019).
- 25 54. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
55. S. L. Kosakovsky Pond, A. F. Y. Poon, R. Velazquez, S. Weaver, N. L. Hepler, B. Murrell, S. D. Shank, B. R. Magalis, D. Bouvier, A. Nekrutenko, S. Wisotsky, S. J. Spielman, S. D. W. Frost, S. V. Muse, HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **37**, 295–299 (2020).
- 30 56. M. D. Smith, J. O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, S. L. Kosakovsky Pond, Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
- 35 57. J. O. Wertheim, B. Murrell, M. D. Smith, S. L. K. Pond, K. Scheffler, RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).
58. F. Bielejec, P. Lemey, L. M. Carvalho, G. Baele, A. Rambaut, M. A. Suchard, π BUSS: A parallel BEAST/BEAGLE utility for sequence simulation under complex evolutionary scenarios. *BMC Bioinformatics.* **15**, 1–8 (2014).
- 40 59. V. Czudai-Matwich, M. Schnare, O. Pinkenburg, A simple and fast system for cloning influenza A virus gene segments into pHW2000- and pCAGGS-based vectors. *Arch. Virol.* **158**, 2049–2058 (2013).
60. G. Gabriel, B. Dauber, T. Wolff, O. Planz, H. D. Klenk, J. Stech, The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18590–18595 (2005).
- 45 61. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: advanced multi-sample

- quality control for high-throughput sequencing data. *Bioinformatics*. **32**, 292–294 (2016).
62. D. O. Cliver, Capsid and Infectivity in Virus Detection. *Food Environ. Virol.* **1**, 123–128 (2009).
63. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and
5 dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).

Acknowledgments:

Funding: BV and SL were supported by postdoctoral fellowship grants of the Research Foundation -- Flanders (Fonds voor Wetenschappelijk Onderzoek, 12U7121N and 12R4718N, respectively). MTPG acknowledges Danish National Research Foundation Award DNRF143 for
10 funding. This work was funded in part by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases of the NIH (DMM and JKT). MAS was supported by US National Institutes of Health grants HG006139 and AI135995. MW was supported by the Bill and Melinda Gates Foundation (INV-004212) and the David and Lucile Packard
15 Foundation. 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) and from the European Union's Horizon 2020 project MOOD (grant agreement no. 874850). The Artic Network receives funding from the Wellcome Trust through project 206298/Z/17/Z. PL acknowledges support by the Research
20 Foundation - Flanders ('Fonds voor Wetenschappelijk Onderzoek - Vlaanderen', G0D5117N and G0B9317N). This project was also supported by a grant to SCS from the National Research Platform for Zoonoses (Federal Ministry of Education and Research, 01KI1714).

Author contributions:

25 Conceptualization: LVP, BV, MB, FHL, TS, TW, PL, SCS

Methodology: LVP, BV, MB, MTPG, MAS, MW, TW, PL, SCS

Software: MAS, PL

Validation: LVP, BV, MB, DM, JKT, TS, TW, PL, SCS

Formal analysis: LVP, BV, MB, SL, TS, TW, PL, SCS

30 Investigation: LVP, BV, MB, AD, JFG, LH, DH, KM, BP, JS, VS, LT, CU, TW, PL, SCS

Resources: FHL, NW, EW, TS, TW, PL, SCS

Data Curation: LVP, BV, MB, AD, JFG, NW, TS, TW, PL, SCS

Writing - Original Draft: LVP, BV, MB, MW, TS, TW, PL, SCS

Writing - Review & Editing: all authors

Visualization: LPV, BV, MB, NW, TW, PL, SCS

Supervision: DH, VS, TS, TW, PL, SCS

Project administration: LVP, SCS

5 Funding acquisition: FHL, TS, TW, PL, SCS

Competing interests: MAS reports contracts from the US Department of Veteran Affairs, the US Food & Drug Administration and Janssen Research & Development, all outside the scope of this work.

10

Data and materials availability: All raw reads generated for this study have been deposited to the European Nucleotide Archive under project number PRJEB41631 (sample numbers ERS5447401-413). Kraken2 results for all specimens can be visualized through Krona plots at <https://zenodo.org/record/4384755>. Nucleotide alignments for the 8 viral segments are available at <https://zenodo.org/record/4384715>.

15

Supplementary Materials:

Materials and Methods

Supplementary text

20

Figures S1-S12

Tables S1-S5

References (33-63)