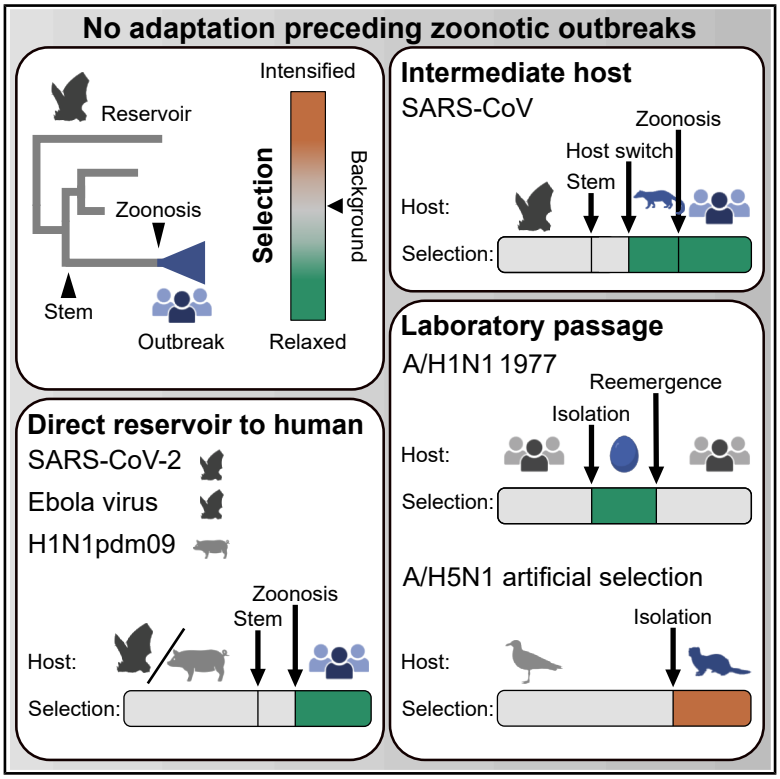
Cell

Dynamics of natural selection preceding human viral epidemics and pandemics

Graphical abstract



Highlights

- Viral adaptation is not a necessary precursor to outbreaks of novel zoonotic viruses
- Selection signatures on SARS-CoV-2 were unchanged until its emergence in humans
- Laboratory and gain-of-function passage produce distinct evolutionary signatures
- 1977 influenza virus reemergence preceded by evolution consistent with laboratory passage

Authors

Jennifer L. Havens, Sergei L. Kosakovsky Pond, Jordan D. Zehr, ..., Michael Worobey, Kristian G. Andersen, Joel O. Wertheim

Correspondence

j.lee.havens@gmail.com (J.L.H.), spond@temple.edu (S.L.K.P.), jwertheim@health.ucsd.edu (J.O.W.)

In brief

Phylogenetic-based selection analysis of viral evolution before and after spillover reveals that viral adaptation in the host reservoir is not a necessary precursor of novel zoonotic virus epidemics and pandemics, including SARS-CoV-2. In contrast, the 1977 reemergence of H1N1 influenza A virus was preceded by selection consistent with laboratory passage.

Havens et al., 2026, Cell *189*, 1–14
April 30, 2026
https://doi.org/10.1016/j.cell.2026.02.006

CellPress

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

Cell

CellPress
OPEN ACCESS

Article

Dynamics of natural selection preceding human viral epidemics and pandemics

Jennifer L. Havens,[1,2,9,*] Sergei L. Kosakovsky Pond,[3,*] Jordan D. Zehr,[3,4] Jonathan E. Pekar,[1,5,6] Edyth Parker,[7] Michael Worobey,[8] Kristian G. Andersen,[7] and Joel O. Wertheim[6,*]

[1]Bioinformatics and Systems Biology Graduate Program, University of California - San Diego, La Jolla, CA 92093, USA
[2]Department of Biostatistics, Fielding School of Public Health, University of California - Los Angeles, Los Angeles, CA 90095, USA
[3]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA 19122, USA
[4]Department of Public and Ecosystem Health, Cornell University, Ithaca, NY 14850, USA
[5]Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK
[6]Department of Medicine, University of California - San Diego, La Jolla, CA 92093, USA
[7]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA
[8]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
[9]Lead contact
*Correspondence: j.lee.havens@gmail.com (J.L.H.), spond@temple.edu (S.L.K.P.), jwertheim@health.ucsd.edu (J.O.W.)
https://doi.org/10.1016/j.cell.2026.02.006

SUMMARY

Using a phylogenetic framework to characterize natural selection, we investigate the hypothesis that zoonotic viruses require adaptation prior to zoonosis to sustain human-to-human transmission. Examining the zoonotic emergence of Ebola virus, Marburg virus, mpox virus, influenza A virus, and SARS-CoV-2, we find no evidence of a change in selection intensity immediately prior to outbreaks in humans compared with typical selection within reservoir hosts. We found a change in selection on SARS-CoV in an intermediate host. We conclude that extensive pre-zoonotic adaptation is not necessary for human-to-human transmission of zoonotic viruses. In contrast, the reemergence of H1N1 influenza A virus in 1977 was preceded by a shift in selection intensity, consistent with the hypothesis of passage in a laboratory setting. Holistic phylogenetic analysis of selection regimes can be used to detect evolutionary signals of host switching or laboratory passage, providing insight into the circumstances of past and future viral emergence.

INTRODUCTION

Human history has been marked by viral epidemics of zoonotic origin, viruses transmitted from a vertebrate reservoir to humans. Recent such epidemics include Ebola virus, Marburg virus, HIV-1, influenza A virus, SARS-CoV, MERS-CoV, SARS-CoV-2, and mpox virus.[1–8] The traditional framework of zoonosis posits that animal viruses must first undergo adaptation in their original host reservoir, an intermediate host, or during stuttering human transmission. Adaptation allows these viruses to evolve the ability to infect human cells and sustain human-to-human transmission.[9,10] An alternative to this framework proposes that some viruses circulating in the natural host reservoir already have the ability to infect and transmit between humans and do not require additional adaptation to cause zoonotic epidemics.[11–13]

Direct evidence to adjudicate these two alternatives, such as sequenced and characterized isolates from infected proximal and intermediate hosts, densely sampled viral genomes from early human outbreaks, and suitably established and measured correlates of transmission and infectivity, is very rarely, if ever, available. Indirect evidence, such as informative evolutionary patterns in viral genomes, is far more common. Guided by this perspective, we define adaptation as the process by which human infectiousness and transmissibility evolve via directional natural selection, which allows us to delineate distinct expectations for these two frameworks for understanding zoonosis. In the first scenario, zoonotic viruses that cause epidemics in humans are the product of adaptive evolution in their host reservoirs, and this selection should be detectable prior to zoonosis. In the latter scenario, the selective sieve for zoonotic viruses occurs at the species barrier, and so-called adaptive mutations necessary for infection and human-to-human transmission are not the product of directional natural selection.

Both of these frameworks presume that a zoonotic virus will continue to adapt to improve human-to-human transmission post zoonosis. This pattern has been well documented: improved ACE2 (angiotensin-converting enzyme 2) receptor binding affinity during the 2002–2004 SARS epidemic,[14–16] increased human cell tropism during the 2013–2016 West African Ebola epidemic,[17] and increased transmissibility and ACE2 receptor binding affinity in early lineages and subsequent variants during the COVID-19 pandemic.[18–20] We do not consider selection on endemic pathogens, because by that stage, adaptive imperatives become complex and varied (e.g., population

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

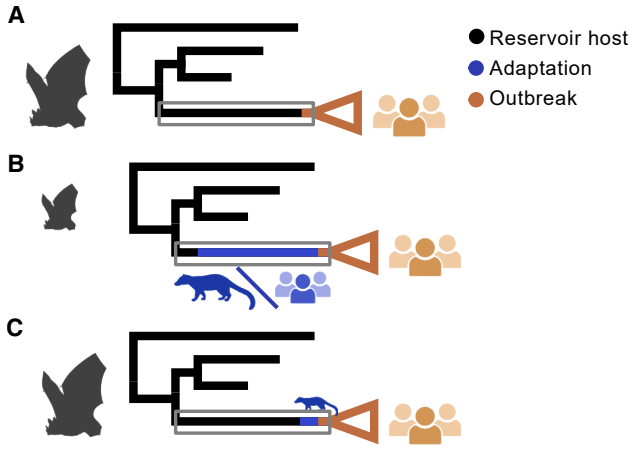


Figure 1. Cartoon viral phylogeny during the evolutionary history of zoonosis

(A) Virus jumps directly from the host reservoir to humans without adaptation and evolution on the stem leading to the MRCA of the human outbreak, which is primarily associated with the natural host reservoir.

(B) Zoonotic emergence occurs after a period of adaptation—either during stuttering transmission between humans, in an intermediate host, and/or in the host reservoir—and selection on the stem leading to the MRCA of the human outbreak is distinct from the host reservoir.

(C) Zoonotic emergence is facilitated, briefly, via an intermediate host, but selection on the stem leading the MRCA of the human outbreak is primarily associated with the natural host reservoir. The size of the animal silhouettes reflects the proportion of the evolutionary change on the stem branch that occurred in that host. Icons created with BioRender.

immunity, drug escape, and lineage competition). To distinguish between these hypotheses concerning the emergence of zoonotic viruses, we need to characterize the evolutionary dynamics prior to the most recent common ancestor (MRCA) in humans, along the stem branch in the phylogeny preceding zoonotic emergence, as compared with evolution in the natural host reservoir.

Comparative phylogenetic methods, which estimate the relative rates of non-synonymous (dN) and synonymous (dS) substitution, have found broad use in understanding how viruses evolve and adapt in changing environments.[21–25] The ratio of these rates, dN/dS or ω, is an informative statistic describing the nature of selective forces, whereby $\omega < 1$ indicates purifying selection, $\omega > 1$ indicates positive diversifying selection, and $\omega \sim 1$ indicates neutral evolution. Modern methods account for the spatial and temporal heterogeneity of selective pressures and correct for many confounding processes such as recombination and variation in synonymous substitution rates.[26]

We can describe a holistic, selection regime over a genome and phylogeny by splitting a single ω into the relative contribution of strongly purifying, weakly purifying, and positive selection.[27–29] Qualitative changes in the selection regime can be inferred by determining whether the strength of natural selection has been relaxed (all ω components shift closer to 1, indicating weaker positive and purifying selection) or intensified (all ω components shift away from 1, indicating stronger positive and purifying selection), as implemented in the RELAX framework.[27] Relaxation of natural selection can be observed when viruses are passaged in cell culture, released from both the purifying and positive selection produced by forces such as host immune responses; viruses have been passaged in cell culture to create live-attenuated vaccine viral strains.[30,31] Importantly, there is sufficient signal along a single evolutionary branch in a phylogeny to detect such changes.[27,30,31]

Here, we characterized genome-wide selection regimes of zoonotic viruses (1) in their natural animal host reservoirs, (2) on the stem branch of the phylogenetic tree immediately preceding a zoonotic emergence, (3) and in the early stages of sustained outbreaks in humans, in four viral families: *Orthomyxoviridae*, *Filoviridae*, *Poxviridae*, and *Coronaviridae*. We use the RELAX[27] modeling framework to conduct a formal statistical comparison with a test for a change in selection between evolution in the natural animal host and stem preceding zoonotic epidemics at the level of the entire coding viral genome. We expanded the RELAX framework to account for recombination and reassortment that can confound selection inference. By using a phylogenetic approach to reconstruct the evolutionary history of viruses, we can assess selection during time periods prior to the recognition of an outbreak and before the first viral genomes were sampled. In several "control" scenarios in which viruses were artificially passaged, we found changes in selection consistent with prior expectations. For natural evolution scenarios, a substantial change in evolutionary dynamics prior to zoonosis is not detectable; thus, adaptation is not a necessary condition for the outbreak of novel zoonotic viruses.

RESULTS

Framework for detecting changes in selection using RELAX

We characterized natural selection prior to successful zoonotic outbreaks to investigate whether adaptation preceded outbreaks in humans. We considered selection regimes (referred to as "selection" henceforth) for three phases of each outbreak, using representative phylogenetic trees and partitioning branches into sets: (1) branches representing evolution in the animal host reservoir, (2) branches encompassing early evolution during the outbreak in humans, and (3) the single stem branch separating the MRCA of the human outbreak and animal host reservoir viruses which captures evolution prior to and immediately after the time of zoonosis. The key question addressed here is whether there is a change on this branch that represents a change in the dynamics of natural selection prior to establishing an outbreak in humans. If there is no change in selection, the evolution on the stem branch will be indistinguishable from the virus in the animal host until the outbreak in humans (Figure 1A). Alternatively, if detectable adaptation occurs on this stem branch—either during stuttering transmission between humans, in an intermediate host, and/or in the host reservoir—then a change in selection will be apparent (Figure 1B). If the virus spends a short period adapting in a novel host (Figure 1C), this evolution is unlikely to result in a detectable change in selection. Intuitively, for the method to be powered, it is necessary to have a selection change that is sufficiently large and an evolutionary time that is sufficiently long to accumulate detectable substitution patterns. A single or a few substitutions, even if they are adaptive, will likely be missed.

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

We estimated and compared selection regimes using an extension of the RELAX framework that accounts for recombination and reassortment.[27] This random-effects framework captures temporal (branch to branch) and spatial (site to site in the genome) variation in the selection regime using a discrete distribution of 2 or 3 ω categories with purifying/negative ($\omega \leq 1$), or diversifying/positive selection ($\omega \geq 1$), to represent the full selection regime. The reference distribution is estimated using maximum likelihood (ML) from the "background" branches and describes evolution in the animal host reservoir. Selection on another set of branches (test), which can be human outbreak only, stem only, or combined (stem + human), is modeled as a power transformation of the reference distribution, ω^K. K is a selection relaxation/intensification parameter (estimated by ML). When $K < 1$, all the ω values in the test branches are closer to 1 (neutrality), implying that selection is overall relaxed, relative to the reference (animal host reservoir) branches. When $K > 1$, ω values are further away from one, implying that selection is further away from neutrality, or is intensified relative to the background. The scalar parameter K offers a simple genome-wide measure of how selection prior to or following a zoonosis compares in intensity with selection in natural reservoir hosts. Branches that are neither test nor background are treated as a nuisance set, with their own inferred ω distribution, which is not of interest. A hypothesis test in which the null is $K = 1$ (selection is identical between the environments) provides a measure of statistical significance for a change in selection. We complement it with a cruder single-value genomic estimate of ω, representing an "average" selection regime.

The RELAX test, as described here, is powered by evolutionary changes on the branches of interest, not the number of viruses sampled at the tips of the phylogeny.[27] Therefore, all the evolutionary dynamics that precede a zoonotic outbreak, including any genomic changes that influence infectiousness, replication, or transmissibility, will be captured on the stem branch. Similarly, the evolutionary dynamics of human viruses can be inferred by analyzing viruses that capture the genomic diversity within an outbreak. We assembled datasets to characterize evolution around zoonosis, including the stem branch preceding the epidemic, by considering closely related viruses in the host reservoir and in the human outbreak.

We first examined selection dynamics preceding the zoonotic emergences of the 2009 H1N1 influenza pandemic, the 2013–2016 West African Ebola virus disease (Ebola) epidemic, and the 2004–2005 Angolan Marburg epidemic (Figures 2A–2C).

Selection regimes preceding and during the 2009 H1N1 influenza pandemic

The 2009 H1N1 influenza (H1N1pdm09) pandemic arose from a novel reassortment of influenza genomic segments from swine viruses.[32] The stem branch preceding this pandemic encompasses the estimated 9–12 years of evolution separating the MRCA of H1N1pdm09 influenza A virus in humans and their closest swine virus relatives.[32] We found no evidence for an increase in ω (Figure 2D) or for an intensification or relaxation of selection ($K = 1.1$; $p = 0.21$; Figure 2E) on the stem branch leading to the human pandemic viruses (stem branch; Figure 2A) compared with the branches associated with swine viruses (background branches; Figure 2A). Selection on the branch preceding the emergence of H1N1pdm09 is not statistically distinguishable from that of the typical swine influenza virus.

In contrast, we inferred relaxation of selection for H1N1pdm09 in humans within a year following emergence (outbreak branches; Figure 2A), compared with selection in swine hosts ($K = 0.82$; $p < 0.01$; Figure 2E). During the first year of the H1N1pdm09 pandemic, the single inferred ω of the virus in humans was greater than the single ω in the swine host (Figure 2D). This increase in ω was not driven by an intensification of positive selection but by a relaxation of purifying selection. This relaxation of selection is evidence that H1N1 during the human outbreak is evolving under a different selection regime than in swine hosts. When we combine the stem with the human virus branches, conflating the signal pre- and post-zoonosis, the relaxation of selection at the outset of the pandemic compared with swine virus selection is still detectable, though less pronounced ($K = 0.91$, $p = 0.01$; Figure 2E).

Selection regime preceding and during filovirus epidemics

The 2013–2016 West African Ebola epidemic was the largest recorded outbreak of Ebola virus disease,[1] occurring thousands of kilometers from previously documented human cases of Ebola virus infection. We tested for a change in the selection regime along the stem leading to the West African Ebola outbreak relative to viruses in the reservoir along phylogenetic branches between human outbreaks (Figure 2B). We found no support for intensification or relaxation of selection preceding the West African epidemic ($K = 1.09$, $p = 0.34$; Figure 2E), indicating that selection on the stem branch is consistent with selection in the bat reservoir up until the MRCA of the human epidemic clade.

There was, however, evidence for a significant change in the selection regime during the first 18 months of human-to-human transmission during the West African Ebola epidemic compared with selection on viruses in the host reservoir ($K = 0.41$, $p < 0.01$; Figure 2E). This signal was still detectable when we conflated the stem and human-outbreak branches ($K < 0.01$, $p < 0.01$; Figure 2E). The ω distribution inferred for the reservoir/host branches included only $\omega \leq 1$, representing purifying and neutral selection. This ω distribution did not include a non-trivial (weight > 0) component for positive selection ($\omega > 1$), indicating that, for these Ebola virus branches, selection in the animal hosts is adequately described by purifying and neutral selection alone. This situation makes it impossible to attribute the inferred change of selection in the test to a specific direction of change in the selection regime (i.e., intensification or relaxation of selection). In the situation of no $\omega > 1$ in the background partition, an increase in positive selection ($\omega > 1$ intensifies beyond $\omega = 1$) or relaxation of purifying selection ($\omega \leq 1$ relaxes to $\omega = 1$) in the test partition can only be described as a relaxation of the $\omega \leq 1$. We refer to this behavior, in which we observe a significant change in the selection regime but cannot distinguish intensification from relaxation, as an "unbalanced model." When we compared the mean ω from the Ebola virus in humans to the bat reservoir, we found the epidemic clade had a higher estimate of ω than the reservoir (Figure 2D); this behavior is the same as we observed in H1N1pdm09.

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

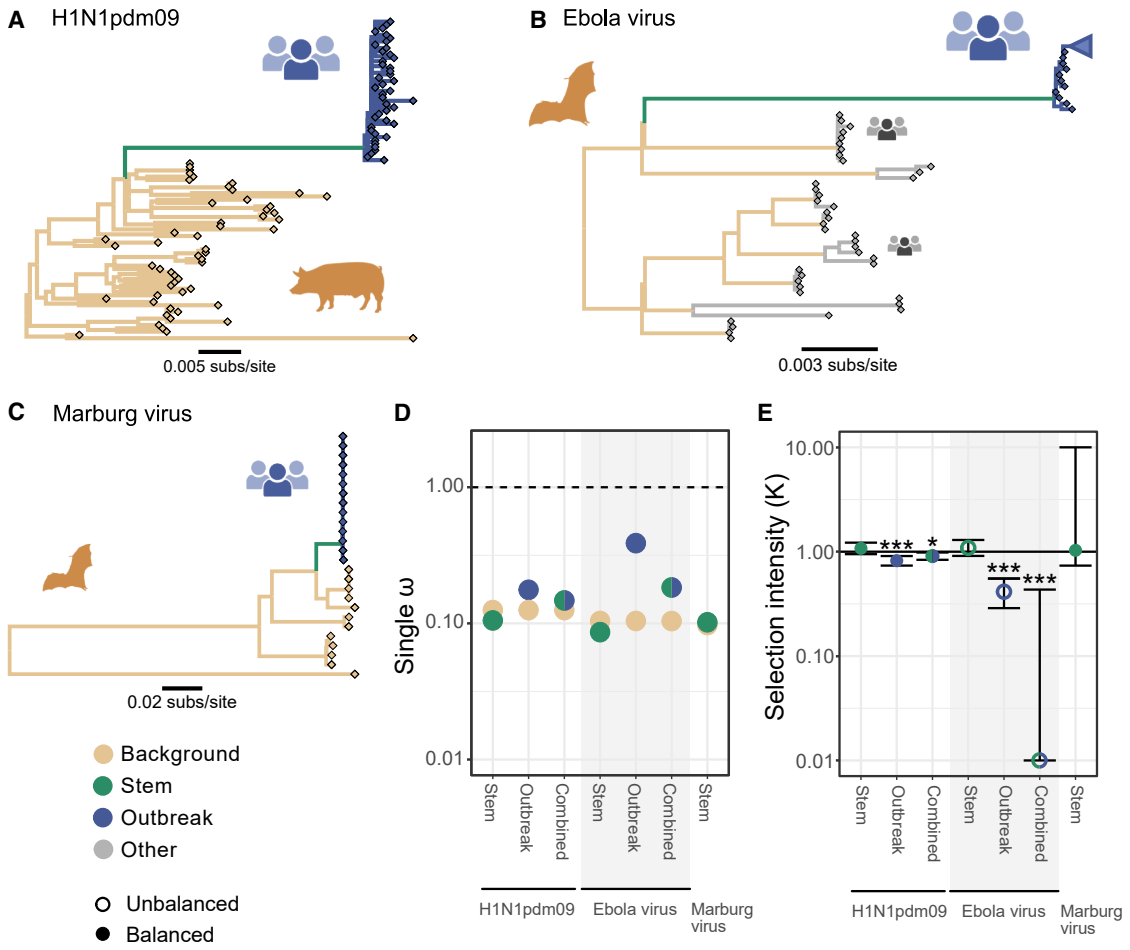


Figure 2. Quantifying selection regimes of epidemics with zoonotic origin

(A–C) Phylogenetic trees of (A) H1N1 2009 influenza pandemic (H1N1pdm09), (B) 2013–2016 West African Ebola outbreak, and (C) Angolan 2004–2005 Marburg virus outbreak.

(D) Single ω for each branch partition of background and the test sets, including "combined," which combines the stem and the epidemic clade in test branches. Points are color-coded according to the branch set: background (brown), stem (green), and outbreak (blue).

(E) Change in selection intensity inferred with RELAX. For the Marburg virus, there was insufficient sequence diversity within human isolates to perform the corresponding tests. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Numerical values reported in Table S1. Icons created with BioRender.

Marburg virus is another filovirus associated with hemorrhagic fever, and fruit bats are known to serve as the natural host reservoir.[33] One of the largest known Marburg outbreaks in humans occurred in Angola in 2004–2005. We characterized selection along the stem preceding the epidemic MRCA and compared it with selection on branches associated with Marburg virus in bats (Figure 2C). There was no evidence of intensification or relaxation on the stem (K = 1.03, $p = 0.90$; Figure 2E). We were unable to estimate selection regimes within the human Marburg outbreak due to insufficient viral genomic diversity among the human samples.

Selection regime preceding and during the early mpox epidemic

In 2022, a clade IIb variant of the zoonotic mpox virus spread across multiple continents.[34] This mpox virus lineage has been circulating in humans in Nigeria since at least 2014, though the identity of its zoonotic host reservoir remains unknown.[35] We characterized selection of the initial detected outbreak in Nigeria, on the stem branches that represent evolution prior to emergence in humans and early undetected human transmission, and in the host reservoir represented by branches between human outbreaks (Figure 3A). We found no difference in the single ω inference between the stem and background reservoir, but we did detect an increase in the single ω within the human outbreak clade (Figure 3B). Additionally, there was no evidence of intensification or relaxation on the stem compared with selection in the host reservoir (Figure 3C; K = 0.80, $p = 0.65$). We found evidence for an intensification of selection during the mpox epidemic compared with selection on viruses in the host reservoir (K = 1.33, $p < 0.01$). Intensification of selection was still detectable when we

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

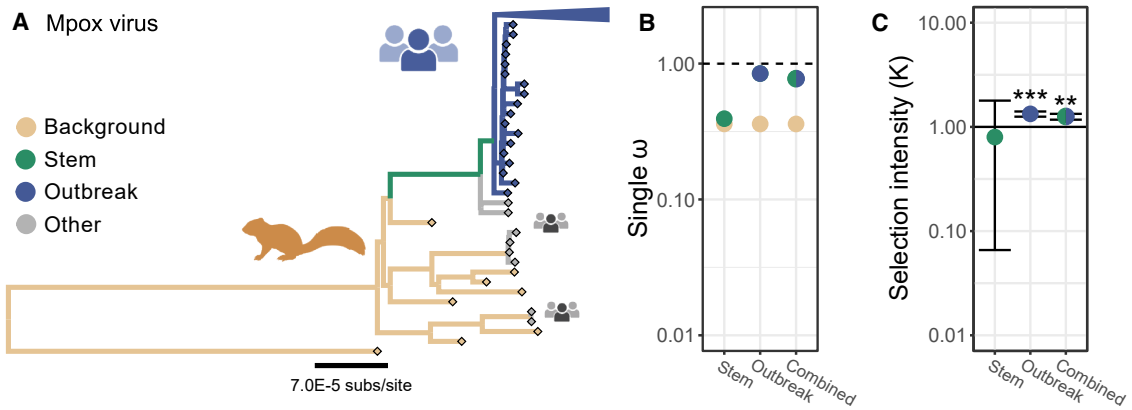


Figure 3. Quantifying the selection regime in the mpox virus

(A) Phylogenetic tree of the mpox virus.

(B) Single ω for each branch partition of background and the test sets, including "combined," which combines the stem and the epidemic clade in test branches. Points are color-coded according to the branch set: background (brown), stem (green), or outbreak (blue).

(C) Change in selection intensity inferred with RELAX. Numerical values in Table S1. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Icons created with BioRender.

conflated the stem and branches within the human outbreak ($K = 1.25$, $p < 0.01$).

Selection regimes in Betacoronaviruses prior to spillovers in humans

Betacoronaviruses are a genus of viruses circulating in bats that have repeatedly jumped into humans. SARS-CoV and SARS-CoV-2 are Betacoronaviruses of the subgenus Sarbecovirus. The 2002–2004 SARS epidemic was traced back to palm civets (*Paguma larvata*) and other animals sold for human consumption in live-animal markets in China.[36] The close contact between humans and these CoV-infected intermediate host animals likely facilitated repeated spillover.[16]

We analyzed SARS-CoV-like sarbecovirus genomes sampled from bats, human SARS-CoV genomes representing two distinct zoonotic introductions, and a SARS-CoV genome sampled from a palm civet intermediate host (Figure 4A). After partitioning the genome into 31 putatively non-recombinant regions, we compared selection along the stem leading from bat SARS-CoV-like viruses with the MRCA of human and palm civet SARS-CoVs. We found no significant difference in selection between the stem and animal hosts in the single ω (Figure 4B) or RELAX analysis ($K = 0.97$, $p = 0.91$; Figure 4C). The lack of change suggests that the SARS-CoV-like viruses had not experienced extensive evolution in an evolutionary environment different from that of the bat host prior to introduction into an intermediate host species.

However, when we combine the stem with the clade comprising palm civet and early human cases—primarily representing evolution in the intermediate palm civet host preceding introduction into humans—we find that the SARS-CoV selection regime was significantly relaxed compared with related bat viruses ($K = 0.65$, $p < 0.01$; Figure 4C). This relaxation is reflected in an increase in the single ω after host-switching (Figure 4B).

The circumstances surrounding the emergence of SARS-CoV-2 are similar to those of SARS-CoV. In late 2019, SARS-CoV-2 was epidemiologically linked to a market selling wild animals in central China,[4,37–39] thousands of kilometers away from closely related bat viruses in southern China and Laos.[40] Genetic and epidemiological evidence support the hypothesis that the virus preceding SARS-CoV-2 briefly circulated in an intermediate host sold at the market.[41–44] However, it has been suggested that SARS-CoV-2 was passaged or modified in a laboratory context, prior to emergence.[45]

If there was extensive evolution in an intermediate host or passage in a laboratory context prior to emergence, we would expect detectable change in selection on the stem preceding SARS-CoV-2. However, our analysis of selection on the stem preceding SARS-CoV-2 emergence across 15 putatively non-recombinant regions found no evidence of intensification or relaxation of selection compared with selection of the bat host reservoir ($K = 1.1$, $p = 0.23$; Figure 5). Hence, we find no evidence to suggest SARS-CoV-2 experienced prolonged selective pressure in an environment different from related bat viruses prior to its emergence in humans. This result does not change if we use a different approach to identifying non-recombinant regions ($K = 1.02$, $p = 0.82$; Figure S1C).

We then examined evolution along the SARS-CoV-2 stem in combination with viral evolution during the first 3 months of the outbreak in China, to understand the selection environment of SARS-CoV-2 in humans compared with the bat host reservoir. We find evidence for a significant change in the selection regime, consistent with a host switch causing a change in the evolutionary environment ($K = 0.69$, $p < 0.01$). The change in selection regime is also detectable in human viruses during the first pandemic wave through September 2020, when conflated with the stem ($K = 0.56$, $p < 0.01$; Figure S1). However, we cannot confidently infer the directionality of this change because the model is unbalanced, as in the Ebola virus selection regime.

The reemergence of H1N1 in 1977

In 1977, the H1N1 influenza A virus reemerged in humans after going extinct 20 years prior. The reemergent virus was closely

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

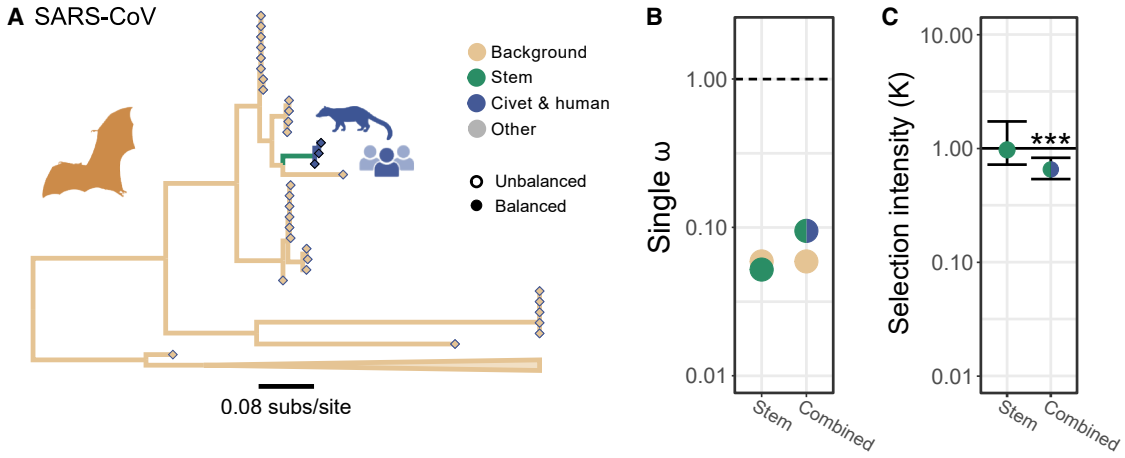


Figure 4. Quantifying the selection regime in SARS-CoV

(A) Phylogenetic tree of SARS-CoV-like sarbecoviruses across a single non-recombinant region. The SARS outbreak includes 2 human virus genomes and 1 palm civet virus genome.

(B) Single ω for each branch partition of branch sets. Points are color-coded according to the branch set: background (brown), stem (green), or outbreak (blue).

(C) Change in selection intensity inferred with RELAX. Numerical values in Table S1. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Icons created with BioRender.

related to a strain that had been circulating in the 1950s.[46,47] It has been proposed that H1N1 reemerged in 1977 without expected evolution because it was frozen in a laboratory before it was accidentally allowed to re-establish "wild" human-to-human transmission, perhaps involving (1) a live-attenuated viral vaccine or (2) a laboratory-adapted virus used as a challenge virus during an influenza vaccine trial.[48]

We inferred 48 substitutions across the entire coding genome along the stem preceding the 1977 reemergence. Genomic segments such as M and PA segments experienced only 2 substitutions each; whereas PB1 experienced 13 substitutions, 12 of which were synonymous (Table S2). With only 48 substitutions over 27 years, the 1977 strain is likely too genetically similar to the 1950s H1N1 virus to support natural reemergence. However, the provenance of this virus prior to its reemergence remains poorly understood.

We compared the selection regime of the coding regions across all H1N1 influenza A virus segments on the stem preceding the 1977 influenza pandemic with evolution post-1977 (Figure 6A); viruses sampled prior to 1950 were excluded from the comparison due to concerns about sequence quality or laboratory passaging. We found evidence for an increase in ω (Figure 6B) and a significant relaxation of selection along the stem ($K = 0.71$, $p = 0.043$) (Figure 6C). This change in selection suggests that, in addition to being frozen in the 1950s, the precursor to the 1977 H1N1 influenza pandemic experienced limited evolution in an environment distinct from human-to-human transmission (e.g., laboratory passage for attenuation or under artificial selection). This result remains consistent even if we exclude the phylogenetic tips associated with influenza A virus isolates with a known history of laboratory passaging (Figure 6C; Figure S2; $K = 0.70$, $p = 0.042$).

Laboratory passage of H1N1 influenza A viruses

To better contextualize the atypical evolutionary dynamics preceding the 1977 H1N1 influenza pandemic, we explored whether the signature of laboratory passage could be detected as a change in the selection regime by RELAX. The Wilson Smith 1933 H1N1 (WSN33) isolate is the ancestor of many laboratory strains used in research.[49] We compared the selection regimes of 3 early laboratory isolates derived from WSN33 with those associated with human-to-human H1N1 transmission across all 8 genomic segments (Figure 6). There was evidence of an increase in overall ω and a significant relaxation of selection in the WSN33 clade ($K = 0.24$, $p < 0.01$; Figures 6B and 6C). This relaxation of selection is consistent with the removal of purifying selection due to the lack of host-specific constraints.

Zhang et al.[50] claimed that the influenza A virus persisted in lake ice based on HA gene sequences. However, Worobey[51] noted that these putative lake ice viruses were monophyletic and descended from the WSN33-derived positive control (Figure S3). These observations indicated that these samples were contaminants derived from this lab strain,[51] which had in fact been used as a positive control by Zhang et al.[50] Accordingly, in our current study, we found that selection on putative lake ice viruses led to an increase in overall ω (Figure 7A) and was significantly relaxed compared with human H1N1 virus ($K = 0.1$, $p < 0.01$; Figure 7B), which is consistent with laboratory passaging in the absence of typical selective constraints.

Selection on attenuated vaccine viruses

We explored whether RELAX could characterize the change in selection during the passage of the virus for the creation of attenuated measles and mumps vaccines. The selection regimes of the measles and mumps viruses passaged for vaccine attenuation were compared with branches representing transmissions between humans (Figures S4A and S4B). We confirmed that the attenuated measles and mumps vaccine virus both had an increased ω relative to the background branches representing wild circulating viruses (Figure 7A). We also detected significant relaxation of selection in the vaccine-attenuated viruses relative

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

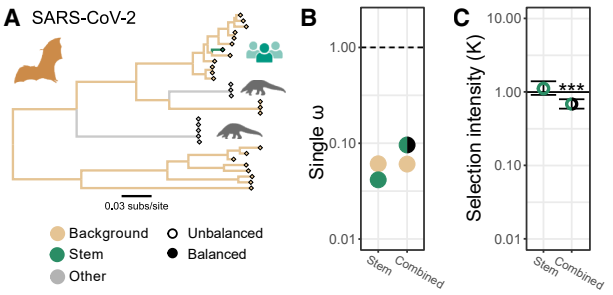


Figure 5. Quantifying selection regimes in SARS-CoV-2

(A) Phylogenetic tree of SARS-CoV-2-like sarbecoviruses in a single non-recombinant region.

(B) Single ω for each branch partition of background and the test sets; "combined" includes both the stem and outbreak partitions in the test set (Figure S1). Points are color-coded according to the branch set: background (brown), stem (green), or outbreak (black).

(C) Change in selection intensity inferred with RELAX. Numerical values in Table S1. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Icons created with BioRender.

See also Figure S1.

to selection in the human host for both measles virus (K = 0.03, $p < 0.01$) and mumps virus (K = 0.84, $p < 0.01$) (Figure 7B).

Serially passaged coronavirus

Murine hepatitis virus (MHV) is a betacoronavirus naturally found in mice. We compared the selection regime of MHV propagated in cell culture with the selection regime of naturally transmitting MHV, using only five MHV genomes for the background based on our confidence that evolution between these genomes is associated with circulating MHV (Figure S5), across 11 putatively non-recombinant regions. We detected an increased ω in the passaged virus relative to the circulating wild MHV (Figure 7A). Further, we found evidence of a change in selection in the passaged viruses (K = 0, $p < 0.01$; Figure 7B); however, we are unable to identify the direction of change due to an unbalanced model. The ability to detect a significant change in lab-passaged MHV demonstrates the power of this framework, even in the presence of sparsely sampled circulating MHV genomes.

Artificial selection on influenza A viruses

Artificial selection, e.g., passaging of viruses under selective constraints to produce novel phenotypes, is expected to intensify the selection regime. We compared selection associated with the H2N2 cold-adapted virus with selection associated with the H2N2 virus transmitted between humans (Figure S6). We found evidence of a change in selection of non-surface proteins resulting from artificial selection (K = 0, $p < 0.01$; Figure 7B); however, the model is unbalanced, and we cannot identify the direction of selection intensity change. Notably, we also inferred an order-of-magnitude increase in ω in passaged viruses relative to the virus in humans (Figure 7A), consistent with adaptation due to laboratory passage.

In an artificial selection experiment, the H5N1 influenza A virus was passaged between ferrets to gain airborne transmissibility.[52] We inferred an increase in the overall ω (Figure 7A) and a significant change in the intensity of selection due to this artificial selection in the new host compared with the virus in the host avian reservoir (Figure 7B; Figure S7; $p < 0.01$). However, the inferred direction of change in selection intensity was unstable; relaxation (K = 0.01) and intensification (K = 2.12) did not have significantly different likelihoods in our model. This instability is partly due to the relatively low level of genomic divergence produced during the gain-of-function experiment. Across our analyses of viruses passaged in a laboratory setting, we consistently found evidence for a change in the selection regime of passaged viruses, with and without artificial selection, compared with evolution in the host environment.

Sensitivity for detecting a change in selection post zoonosis

We explored the sensitivity of our approach to characterizing changes in selection within human outbreaks after zoonotic transmission by evaluating how many months of viral evolution in humans were needed to detect a change in selection intensity. Our primary analyses of H1N1pdm09 influenza A virus and the West African Ebola virus demonstrate that a change in selection is detectable when considering a full year of viral genomic diversity in humans (Figures 2A and 2B). By progressively shortening this window of human viral genome sampling within the outbreak clades, we determined that a change in selection can be detected about 6 months into the H1N1pdm2009 pandemic and 12 months into the West African Ebola virus epidemic (Table S3).

For SARS-CoV-2, our primary analysis (Figure 5C) demonstrated that a change in selection was detectable after 3 months of viral evolution in humans. To assess the sensitivity of our approach, we replaced the first sequenced genome and outbreak clade with single viral genomes sampled progressively later in the pandemic. Using this approach, we detected a change in selection with a single SARS-CoV-2 genome in 2021, when variants emerged (Table S3). These sensitivity analyses indicate that changes in selection intensity due to human-to-human transmission within a single viral genome lineage are relatively weak compared with those induced by laboratory passage, and a combination of multiple sampled genomes over many months is needed to reliably detect this change.

SARS-CoV was the sole zoonotic outbreak in which we detected a change in selection prior to sustained transmission in humans, presumably because of prolonged transmission in the palm civet intermediate host. We explored the sensitivity of our ability to detect this change in selection by downsampling viral genomes from humans, palm civets, and bats. The signal for relaxation could be detected when only two or three viral genomes from palm civets and humans were analyzed, but the test failed to detect a significant difference when only a single palm civet or human virus was included (Table S4). This sensitivity indicates the difficulty in detecting relaxation associated with host-switching with limited genetic diversity, such as in the case of a single genome. When downsampling the bat virus background genomes, we still detected a significant change in selection when excluding 75%, 80%, 85% and 90% of the bat virus genomes (Table S5). Only when we downsampled the background reservoir by 95% (to include just 6 background genomes) were we unable to detect a significant change in selection. Hence, our SARS-CoV inference is strongly robust to viral sampling in the host reservoir.

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

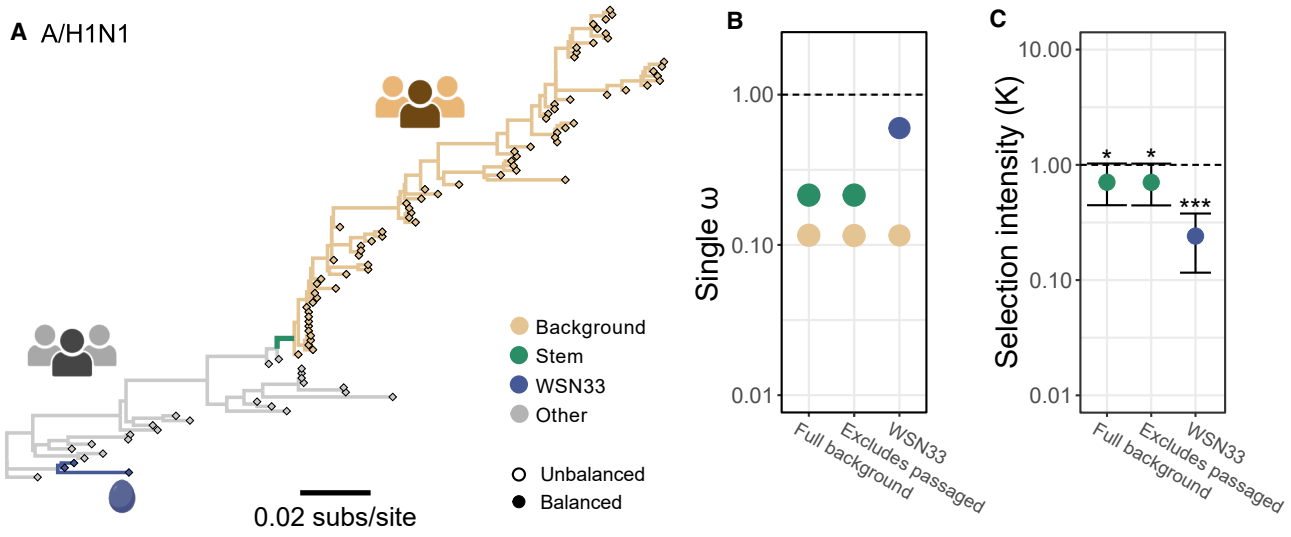


Figure 6. Quantifying the selection regime of H1N1

(A) Phylogenetic tree of H1N1 influenza A virus HA segment. The stem branch connects the 1977 clade and basal ancestors and is foreground for the 1977 flu test. WSN33 (blue) branches are clade-associated with laboratory passage. Other branches are excluded from analysis (treated as a nuisance group).

(B) Single ω for each branch partition of the background and the test sets. "Full background" compares the selection on the stem to the displayed background. "Excluded passaged" compares selection on stem to background, excluding tips that have a history of passaging (49/74; Figure S2). "WSN33" compares selection of WSN33 branches to the full background. Points are color-coded according to the branch set: background (brown), stem of 1977 pandemic (green), or WSN33 clade (blue).

(C) Change in selection intensity inferred with RELAX. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Numerical values in Table S1. Icons created with BioRender.

See also Figure S2.

Robustness of change in selection inference in laboratory-passaged viruses

Lastly, we explored the impact of genome sampling on our ability to detect a change in selection in some "control" datasets: passaged H1N1 lake ice, measles virus vaccine passaging, and H5N1 artificial selection. In all three cases, our results are robust to (1) downsampling genomes from the background reservoir viruses, (2) downsampling the number of passaged viral genomes, and (3) intentionally misspecifying branches representing evolution in the background reservoir (i.e., the stem branch leading to the passaged viruses) as a passaged virus.

We still found evidence for relaxation in both H1N1 lake ice and measles vaccine genomes when the number of background viruses was reduced by 25%, 50%, and 75% (Table S5). For lab-passaged H5N1, we continued to detect a significant change in selection when downsampling the background viruses by 25%, 50%, 75%, 85%, 90%, and 95%.

Similarly, reducing the number of passaged viral genomes did not influence our findings. When including only a single-passaged viral genome in the phylogeny, we detected a significant change in selection for H1N1 lake ice (K = 0.00, $p = 0.002$) and H5N1 influenza A virus (K = 2.6, $p = 0.002$), results consistent with the original analysis. When including only a single measles vaccine strain in the phylogeny at a time, we were able to detect a significant change in selection for 12 of 13 vaccine strains (mean K = 0.03, $p \leq 0.05$); we then tested 20 combinations of 2 randomly chosen strains, 19 found a change in selection (mean K = 0.03; $p \leq 0.001$); when 3 vaccine strains were randomly chosen, we detected relaxation in all 100 combinations tested (mean K = 0.022, $p \leq 0.001$). Hence, our ability to detect changes in selection is robust to sampling depth for both background and passaged viral genomes.

Furthermore, our ability to detect a change in selection was robust to phylogenetic misspecification, which we explored by removing the virus used at the initial passaging, a modeling error mimicking the imprecision of inferring the virus that jumps into a new host context during analysis of zoonosis. Phylogenetic misspecification that considered the stem branch, as part of laboratory passaging for H5N1 influenza A virus and measles vaccine strains, still produced evidence for significant change in selection: H5N1 influenza A virus (K = 1.51, $p = 0.04$) and measles virus vaccine (K = 0.24, $p < 0.001$). H1N1 lake ice was not considered in this analysis because the ancestral virus used for passaging is unknown. We note that, as expected, the strength of the signal for the change in selection was weaker than in the properly specified analysis.

Discussion

How much adaptation must a zoonotic virus undergo before it becomes adept at sustained human-to-human transmission, and where does this adaptation take place? Observed empirical data from early in outbreaks can be sparse and inconclusive; however, we can employ a phylogenetic framework that captures the evolutionary dynamics prior to the recognition of a new outbreak. We reason that sufficiently extensive adaptation prior to an outbreak in humans will leave detectable evolutionary

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

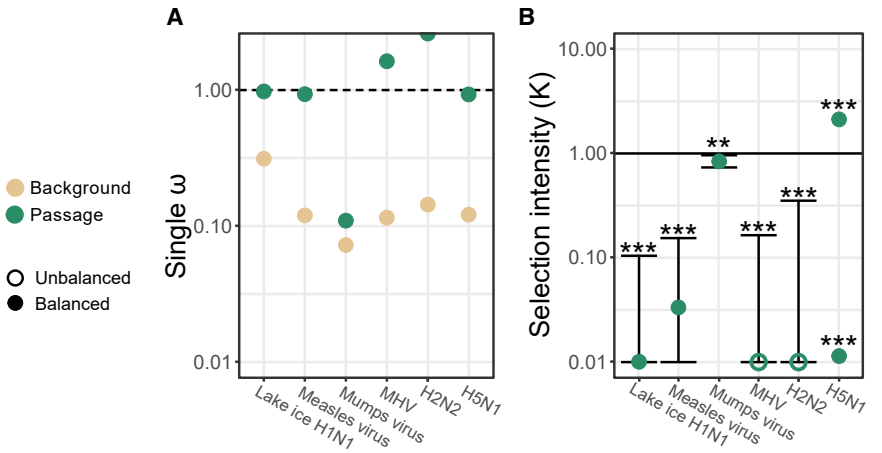


Figure 7. Quantifying selection regimes of laboratory-passaged viruses

(A) Single ω for each branch partition of comparison between wild and passaged viruses. H2N2 ω is shown at the top edge, indicating it is beyond the drawn scale at ω = 4.5. Points are color-coded according to the branch set: background (brown) and passage (green).

(B) Change in selection intensity of laboratory-passaged viruses compared with wild circulating viruses inferred with RELAX. H5N1 depicts points above and below K = 1, indicating the instability in the model estimate. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Numerical values in Table S1. Trees for analysis in Figures S3–S7.

signals in the viral genome, specifically on the phylogenetic branch immediately preceding the observed outbreak. Our study reveals that the intensity of natural selection did not change on evolutionary branches preceding zoonotic outbreaks. This finding stands in stark contrast to the shift in evolutionary dynamics we observed once these viruses began transmitting between humans. Additionally, we identified several instances of a change in the intensity of natural selection during viral passaging in culture and in intermediate host animals, which serve as "controls" for our statistical framework.

We find that for recent zoonotic epidemics—H1N1pdm09 influenza A virus, Ebola virus, Marburg virus, mpox virus, and SARS-CoV-2—these viruses continued to evolve in a manner indistinguishable from evolution in their host reservoir up until the time of the MRCA of the epidemic in humans (as outlined in Figure 1A). These findings challenge the model in which zoonotic viruses must progressively evolve the ability to sustain human-to-human transmission.[9,10,53] Instead, this evolutionary pattern is consistent with a process whereby humans are repeatedly exposed to viruses originating from host reservoirs, but only viruses that happen to be sufficiently transmissible in humans to sustain transmission tend to be the seed of zoonotic epidemics.[11,12]

This conclusion is supported by previous analysis of SARS-CoV-2 and its closest relatives, which found no evidence of significant adaptation during the evolutionary period immediately preceding the human outbreak.[22] Our results expand on this finding by establishing a lack of signal for a change in the overall selection regime on the pre-outbreak stem compared with the selection in bats. Such a change would be expected if the virus had been extensively passaged in a laboratory setting or had been circulating in an intermediate host for a prolonged period. If, as a large body of evidence indicates, SARS-CoV-2 did circulate in an intermediate mammalian host before jumping into humans,[41–44] the duration of this circulation must have been brief,[43] and was not sufficient to leave a signal of a change in the selection regime (Figure 1C).

In contrast, we detected a change in selection for SARS-CoV, associated with the virus circulating for months or years in palm civets, prior to its emergence in humans (as shown in Figure 1B). As with the other zoonotic outbreaks we investigated, however, there was no evidence of a change in selection in the bat reservoir prior to the MRCA of human and palm civet viruses. This lack of change suggests that if a SARS-like virus were to jump directly from bats to humans, we would not expect it to be preceded by a change in selection intensity. The detection of bat sarbecoviruses capable of binding to human cell receptors (such as ACE2) suggests that a subset of bat sarbecoviruses may be capable of infecting and transmitting between humans directly, without an intermediate host.[54] Notably, this observed shift in selection in the palm civet intermediate host did not preclude SARS-CoV from continuing to evolve mutations associated with improved human-to-human transmission during the SARS epidemic.[14–16]

Bats are thought to serve as the host reservoir of not just sarbecoviruses, but also the Ebola virus. The appearance of a Central African strain of Ebola virus in humans in West Africa in late 2013 was unexpected,[55–57] raising questions about how the virus traveled so far without being previously detected. However, serological evidence for the presence of Ebola virus prior to 2013 in people living in West Africa[58] is consistent with repeated introductions of Ebola virus to humans in the region without sustained transmission in the years prior to the epidemic. These dead-end introductions cannot exert selective pressure on the virus in the bat reservoir. Correspondingly, we found no evidence of change in selection prior to the Ebola virus emergence of the 2013–2016 outbreak, suggesting that this Ebola virus lineage remained in its natural host until the West African epidemic.

Unlike the previous examples of zoonotic epidemics, we report evidence of a change in selection preceding the reemergence of H1N1 influenza A virus in 1977, providing a new line of evidence supporting the theory that its reemergence was associated with an influenza vaccine challenge trial.[48] The stem branch preceding the reemergence of H1N1 in 1977 has much less evolution than expected given the temporal length of the branch, suggesting that the virus had been frozen.[46,47] If the precursor to the 1977 H1N1 influenza pandemic had been only frozen for over two decades prior to its reemergence, and not passaged, the selection regime along the phylogenetic branches leading to the re-emergent virus would be similar to that of other circulating H1N1 viruses. Instead, we found that

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

the virus preceding the 1977 influenza pandemic evolved under a selection regime markedly different from typical human-to-human transmission; this change in selection is consistent with the change in selection of the influenza A virus during laboratory passage.

H1N1 viral isolates sampled shortly after the 1977 reemergence were temperature sensitive, replicating more efficiently at temperatures below those of the human body. Remarkably, in the years following the reemergence, its descendant viruses reverted away from temperature sensitivity, indicating that this trait was not optimal for human-to-human transmission.[48,59] Cold adaptation via laboratory passage was a common technique for producing live-attenuated influenza vaccines.[60] We note that modern techniques for producing live-attenuated vaccines via genomic reassortment are not subject to the same ease of reversion to wild-type transmissibility; additionally, subunit and mRNA vaccines are incapable of reverting.[61–64] Based on our analysis of cold-adapted H2N2, cold adaptation via laboratory passage could produce the change in selection dynamics we observed for the 1977 influenza H1N1 reemergence. The missing decades of evolution, temperature sensitivity, and the evidence of a change in selection intensity indicative of viral passaging are all consistent with the reemergence of H1N1 in 1977 being due to an escape during an influenza vaccine trial. Accordingly, the 1977 H1N1 virus and its descendants could be designated a putative circulating vaccine-derived influenza A virus (“cVDIAV”), following the convention used when a similar process occurs with live-attenuated oral poliovirus vaccines (i.e., circulating vaccine-derived polioviruses, “cVDPVs”).

Our approach to characterizing selection regimes overcomes a common shortcoming of similar investigations: the need to independently analyze non-recombinant or non-reassortant genomic segments.[27,65,66] We extended the RELAX phylogenetic framework to account for different evolutionary histories by applying a single selection-regime model to multiple genomic segments with distinct phylogenetic histories. We applied this multi-region analysis to combine regions for viruses in which recombination or reassortment were likely (e.g., positive-sense RNA viruses and influenza A virus). Using multiple genomic regions in the same model increased the power to distinguish changes in the selection regime.

Other analytic methods are also useful for identifying instances in which a pathogen spent time in an intermediate host or was passaged in a laboratory setting. A shift in the rate of the molecular clock can point to transmission in a new host,[25] and a slowdown—or cessation—in the rate is indicative of laboratory isolation,[51] or a persistent—possibly latent—infection.[67,68] Further, changes in nucleotide substitution bias,[69] including APOBEC hypermutation,[70] are useful in characterizing host-switching.

In this study, we have distinguished epidemics that are characterized by viruses that evolved primarily under a selection regime in the natural host reservoir prior to emergence from those that did not. Humans are constantly exposed to animal viruses.[58,71] However, most of these exposures do not result in ongoing outbreaks with human-to-human transmission, due to low fitness of the virus or lack of sufficient transmission opportunities (such as in rural communities).[72] In recent zoonotic epidemics with sustained human-to-human transmission, we found no detectable change in selection preceding the epidemics. Applying multi-region RELAX to the stem of novel viral outbreaks is a tool that we can apply to future outbreaks to rapidly assess the possibility of evolution in an intermediate host or a laboratory setting, compared with zoonosis directly from the natural host reservoir. Increased sampling of viruses in reservoir species will improve the power of approaches for investigating future zoonotic outbreaks.

Limitations of the study

Despite these advances, our multi-region RELAX approach still has limitations. When evolutionary history is complex due to factors such as extensive recombination, repeated host switching, or biased sampling, it becomes difficult to assign correct host species to viral lineages and thus difficult to correctly specify branches for the RELAX test. For example, assigning host species across the reconstructed MERS-CoV phylogeny is complicated by recombination and low statistical confidence.[73] Therefore, we are unable to apply the host-specific RELAX framework to MERS-CoV. Another limitation in our approach is when the stem branch preceding an outbreak becomes longer, the majority of evolution represented on this branch occurs in the reservoir. In these instances, any signal for change in selection preceding the outbreak will be diluted by evolution in the reservoir host species. For example, the stem preceding the variola virus (the *Orthopoxvirus* causing smallpox) leads back to the MRCA with camelpox virus and taterapox virus thousands of years ago.[74–76] Further, this effect is compounded by saturating effects deeper in the evolutionary history of RNA viruses, for example, saturation since the evolutionary divergence of measles virus and rinderpest virus in their ungulate hosts would preclude robust inference.[77] Conversely, the power of our approach can also be limited by exceptionally short evolutionary branches, as in the 1977 H1N1 virus. Changes in selection that result in minimal changes to a viral genome (i.e., one or two nonsynonymous substitutions) can be exceedingly difficult to detect in any circumstance. In our study, the accuracy of our inference was limited when the lack of positive selection in the background lineages resulted in an unbalanced model, which precluded our ability to distinguish intensified selection from relaxed selection in the test branches (e.g., Figures 2E and 7E). Lastly, our approach requires homologous nucleotides to characterize selection, which prevents the characterization of adaptation involving insertions or deletions.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Jennifer L. Havens (j.lee.havens@gmail.com).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- Accession numbers and sequence names for all data are available in Table S6. Alignments of publicly available sequences and annotated trees are available in Data S1.
- The version of RELAX used here is included in HyPhy (v2.5.62 or later), available at github.com/veg/hyphy/.

ACKNOWLEDGMENTS

We gratefully acknowledge all the authors and contributors who generated and shared the viral genomic sequences and metadata, including the originating and submitting laboratories that shared data on GISAID (available at EPI_SET_240827gw), on which this research is based. We thank Stephen A. Goldstein, Mark Zeller, Tetyana I. Vasylyeva, and Marc Suchard for their insightful discussions.

This work was funded in part with federal funds from the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIH-NIAID), and the National Science Foundation (NSF). J.L.H. acknowledges support from the NIH (grant R01AI153044). S.L.K.P. and J.D.Z. acknowledge support from the NIH (AI183870, GM151683, and GM144468) and the NSF (grant DBI/2419522). J.E.P. acknowledges support from the NIH-NIAID (T15LM011271) and the UC San Diego Merkin Fellowship. M.W. acknowledges support from the NIH-NIAID (contract no. 75N93021C00015). E.P. and K.G.A. acknowledge support from the NIH (grant U01AI151812). K.G.A. also acknowledges support from the NIH (grant U19AI135995). J.O.W. acknowledges support from the NIH-NIAID (R01AI135992).

AUTHOR CONTRIBUTIONS

Conceptualization, J.L.H, and J.O.W.; methodology, J.L.H., S.L.K.P., J.D.Z., and J.O.W.; software, J.L.H., S.L.K.P., and J.D.Z.; validation, J.L.H., S.L.K.P., and J.D.Z.; formal analysis, J.L.H., S.L.K.P., J.D.Z., and J.O.W.; investigation, J.L.H., S.L.K.P., J.E.P., E.P., M.W., K.G.A., and J.O.W.; resources, S.L.K.P., and J.O.W.; data curation, J.L.H., S.L.K.P., J.E.P., E.P., M.W., K.G.A., and J.O.W.; writing—original draft preparation, J.L.H., S.L.K.P., and J.O.W.; writing—review and editing, all authors; visualization, J.L.H., S.L.K.P., J.D.Z., and J.O.W.; supervision, S.L.K.P., K.G.A., M.W., and J.O.W.; project administration, J.L.H., and J.O.W.; funding acquisition, S.L.K.P., K.G.A., M.W., and J.O.W.

DECLARATION OF INTERESTS

J.E.P., M.W., K.G.A., and J.O.W. have received consulting fees and/or provided compensated expert testimony on SARS-CoV-2 and the COVID-19 pandemic.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Phylogenetic selection analysis
  - Method sensitivity to sampling
  - Data sets
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.02.006.

Received: January 8, 2025
Revised: July 9, 2025
Accepted: February 5, 2026

REFERENCES

1. Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M.S., Keïta, S., De Clerck, H., et al. (2014). Emergence of Zaire Ebola Virus Disease in Guinea. N. Engl. J. Med. *371*, 1418–1425. https://doi.org/10.1056/NEJMoa1404505.
2. Cauchemez, S., Nouvellet, P., Cori, A., Jombart, T., Garske, T., Clapham, H., Moore, S., Mills, H.L., Salje, H., Collins, C., et al. (2016). Unraveling the drivers of MERS-CoV transmission. Proc. Natl. Acad. Sci. USA *113*, 9081–9086. https://doi.org/10.1073/pnas.1519235113.
3. Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., et al. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature *397*, 436–441. https://doi.org/10.1038/17130.
4. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.
5. Isidro, J., Borges, V., Pinto, M., Sobral, D., Santos, J.D., Nunes, A., Mixão, V., Ferreira, R., Santos, D., Duarte, S., et al. (2022). Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat. Med. *28*, 1569–1572. https://doi.org/10.1038/s41591-022-01907-y.
6. Peiris, J.S.M., Lai, S.T., Poon, L.L.M., Guan, Y., Yam, L.Y.C., Lim, W., Nicholls, J., Yee, W.K.S., Yan, W.W., Cheung, M.T., et al. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet *361*, 1319–1325. https://doi.org/10.1016/S0140-6736(03)13077-2.
7. Schrauwen, E.J., and Fouchier, R.A. (2014). Host adaptation and transmission of influenza A viruses in mammals. Emerg. Microbes Infect. *3*, e9. https://doi.org/10.1038/emi.2014.9.
8. Towner, J.S., Khristova, M.L., Sealy, T.K., Vincent, M.J., Erickson, B.R., Bawiec, D.A., Hartman, A.L., Comer, J.A., Zaki, S.R., Ströher, U., et al. (2006). Marburgvirus Genomics and Association with a Large Hemorrhagic Fever Outbreak in Angola. J. Virol. *80*, 6497–6516. https://doi.org/10.1128/JVI.00069-06.
9. Lloyd-Smith, J.O., George, D., Pepin, K.M., Pitzer, V.E., Pulliam, J.R.C., Dobson, A.P., Hudson, P.J., and Grenfell, B.T. (2009). Epidemic Dynamics at the Human-Animal Interface. Science *326*, 1362–1367. https://doi.org/10.1126/science.1177345.
10. Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. Nature *447*, 279–283. https://doi.org/10.1038/nature05775.
11. Holmes, E.C., and Drummond, A.J. (2007). The Evolutionary Genetics of Viral Emergence. In Wildlife and Emerging Zoonotic Diseases: the Biology, Circumstances and Consequences of Cross-Species Transmission Current Topics in Microbiology and Immunology, J.E. Childs, J.S. Mackenzie, and J.A. Richt, eds. (Springer), pp. 51–66. https://doi.org/10.1007/978-3-540-70962-6_3.
12. Magouras, I., Brookes, V.J., Jori, F., Martin, A., Pfeiffer, D.U., and Dürr, S. (2020). Emerging Zoonotic Diseases: Should We Rethink the Animal–Human Interface? Front. Vet. Sci. *7*, 582743. https://doi.org/10.3389/fvets.2020.582743.
13. Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., and Lloyd-Smith, J.O. (2017). Pathways to zoonotic spillover. Nat. Rev. Microbiol. *15*, 502–510. https://doi.org/10.1038/nrmicro.2017.45.
14. Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N., et al. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. *24*, 1634–1643. https://doi.org/10.1038/sj.emboj.7600640.
15. Qu, X.-X., Hao, P., Song, X.-J., Jiang, S.-M., Liu, Y.-X., Wang, P.-G., Rao, X., Song, H.-D., Wang, S.-Y., Zuo, Y., et al. (2005). Identification of Two Critical Amino Acid Residues of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Its Variation in Zoonotic Tropism Transition via a Double Substitution Strategy. J. Biol. Chem. *280*, 29588–29595. https://doi.org/10.1074/jbc.M500662200.
16. Chinese; SARS; Molecular; Epidemiology Consortium (2004). Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China. Science *303*, 1666–1669. https://doi.org/10.1126/science.1092002.

17. Urbanowicz, R.A., McClure, C.P., Sakuntabhai, A., Sall, A.A., Kobinger, G., Müller, M.A., Holmes, E.C., Rey, F.A., Simon-Loriere, E., and Ball, J.K. (2016). Human Adaptation of Ebola Virus during the West African Outbreak. Cell *167*, 1079–1087.e5. https://doi.org/10.1016/j.cell.2016.10.013.

18. Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell *182*, 812–827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

19. Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., et al. (2021). Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell *184*, 64–75.e11. https://doi.org/10.1016/j.cell.2020.11.020.

20. Martin, D.P., Weaver, S., Tegally, H., San, J.E., Shank, S.D., Wilkinson, E., Lucaci, A.G., Giandhari, J., Naidoo, S., Pillay, Y., et al. (2021). The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. Cell *184*, 5189–5200.e7. https://doi.org/10.1016/j.cell.2021.09.003.

21. Frost, S.D.W., Magalis, B.R., and Kosakovsky Pond, S.L. (2018). Neutral Theory and Rapidly Evolving Viral Pathogens. Mol. Biol. Evol. *35*, 1348–1354. https://doi.org/10.1093/molbev/msy088.

22. MacLean, O.A., Lytras, S., Weaver, S., Singer, J.B., Boni, M.F., Lemey, P., Kosakovsky Pond, S.L.K., and Robertson, D.L. (2021). Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol. *19*, e3001115. https://doi.org/10.1371/journal.pbio.3001115.

23. Pybus, O.G., Rambaut, A., Belshaw, R., Freckleton, R.P., Drummond, A.J., and Holmes, E.C. (2007). Phylogenetic Evidence for Deleterious Mutation Load in RNA Viruses and Its Contribution to Viral Evolution. Mol. Biol. Evol. *24*, 845–852. https://doi.org/10.1093/molbev/msm001.

24. Tan, C.C.S., Van Dorp, L., and Balloux, F. (2024). The evolutionary drivers and correlates of viral host jumps. Nat. Ecol. Evol. *8*, 960–971. https://doi.org/10.1038/s41559-024-02353-4.

25. Worobey, M., Han, G.-Z., and Rambaut, A. (2014). A synchronized global sweep of the internal genes of modern avian influenza virus. Nature *508*, 254–257. https://doi.org/10.1038/nature13016.

26. Kosakovsky Pond, S.L., Poon, A.F.Y., Zárate, S., Smith, D.M., Little, S.J., Pillai, S.K., Ellis, R.J., Wong, J.K., Leigh Brown, A.J., Richman, D.D., et al. (2008). Estimating selection pressures on HIV-1 using phylogenetic likelihood models. Stat. Med. *27*, 4779–4789. https://doi.org/10.1002/sim.3192.

27. Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. Mol. Biol. Evol. *32*, 820–832. https://doi.org/10.1093/molbev/msu400.

28. Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. Mol. Biol. Evol. *32*, 1342–1353. https://doi.org/10.1093/molbev/msv022.

29. Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., and Scheffler, K. (2011). A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. Mol. Biol. Evol. *28*, 3033–3043. https://doi.org/10.1093/molbev/msr125.

30. Hughes, A.L. (2009). Relaxation of purifying selection on the SAD lineage of live attenuated oral vaccines for rabies virus. Infect. Genet. Evol. *9*, 827–831. https://doi.org/10.1016/j.meegid.2009.04.016.

31. Hughes, A.L. (2009). Relaxation of Purifying Selection on Live Attenuated Vaccine Strains of the Family Paramyxoviridae. Vaccine *27*, 1685–1690. https://doi.org/10.1016/j.vaccine.2009.01.036.

32. Smith, G.J.D., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G., Ma, S.K., Cheung, C.L., Raghwani, J., Bhatt, S., et al. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature *459*, 1122–1125. https://doi.org/10.1038/nature08182.

33. Amman, B.R., Bird, B.H., Bakarr, I.A., Bangura, J., Schuh, A.J., Johnny, J., Sealy, T.K., Conteh, I., Koroma, A.H., Foday, I., et al. (2020). Isolation of Angola-like Marburg virus from Egyptian rousette bats from West Africa. Nat. Commun. *11*, 510. https://doi.org/10.1038/s41467-020-14327-8.

34. Paredes, M.I., Ahmed, N., Figgins, M., Colizza, V., Lemey, P., McCrone, J.T., Müller, N., Tran-Kiem, C., and Bedford, T. (2024). Underdetected dispersal and extensive local transmission drove the 2022 mpox epidemic. Cell *187*, 1374–1386.e13. https://doi.org/10.1016/j.cell.2024.02.003.

35. Parker, E., Omah, I.F., Djuicy, D.D., Magee, A., Tomkins-Tinch, C.H., Otieno, J.R., Varilly, P., Ayinla, A.O., Sijuwola, A.E., Ahmed, M.I., et al. (2025.). Genomics reveals zoonotic and sustained human mpox spread in West Africa. Nature *643*, 1343–1351. https://doi.org/10.1038/s41586-025-09128-2.

36. Shi, Z., and Hu, Z. (2008). A review of studies on animal reservoirs of the SARS coronavirus. Virus Res. *133*, 74–87. https://doi.org/10.1016/j.virusres.2007.03.012.

37. Xiao, X., Newman, C., Buesching, C.D., Macdonald, D.W., and Zhou, Z.-M. (2021). Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. Sci. Rep. *11*, 11898. https://doi.org/10.1038/s41598-021-91470-2.

38. Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

39. Liu, W.J., Liu, P., Lei, W., Jia, Z., He, X., Shi, W., Tan, Y., Zou, S., Wong, G., Wang, J., et al. (2024). Surveillance of SARS-CoV-2 at the Huanan Seafood Market. Nature *631*, 402–408. https://doi.org/10.1038/s41586-023-06043-2.

40. Temmam, S., Vongphayloth, K., Baquero, E., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Chrétien, D., Sanamxay, D., et al. (2022). Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature *604*, 330–336. https://doi.org/10.1038/s41586-022-04532-4.

41. Crits-Christoph, A., Levy, J.I., Pekar, J.E., Goldstein, S.A., Singh, R., Hensel, Z., Gangavarapu, K., Rogers, M.B., Moshiri, N., Garry, R.F., et al. (2024). Genetic tracing of market wildlife and viruses at the epicenter of the COVID-19 pandemic. Cell *187*, 5468–5482.e11. https://doi.org/10.1016/j.cell.2024.08.010.

42. Holmes, E.C. (2024). The Emergence and Evolution of SARS-CoV-2. Annu. Rev. Virol. *11*, 21–42. https://doi.org/10.1146/annurev-virology-093022-013037.

43. Pekar, J.E., Magee, A., Parker, E., Moshiri, N., Izhikevich, K., Havens, J.L., Gangavarapu, K., Malpica Serrano, L.M., Crits-Christoph, A., Matteson, N.L., et al. (2022). The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science *377*, 960–966. https://doi.org/10.1126/science.abp8337.

44. Worobey, M., Levy, J.I., Malpica Serrano, L., Crits-Christoph, A., Pekar, J.E., Goldstein, S.A., Rasmussen, A.L., Kraemer, M.U.G., Newman, C., Koopmans, M.P.G., et al. (2022). The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science *377*, 951–959. https://doi.org/10.1126/science.abp8715.

45. Office of the Director of National Intelligence. Updated assessment on COVID-19 origins. (2021). https://www.dni.gov/files/ODNI/documents/assessments/Declassified-Assessment-on-COVID-19-Origins.pdf.

46. Nakajima, K., Desselberger, U., and Palese, P. (1978). Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. Nature *274*, 334–339. https://doi.org/10.1038/274334a0.

47. Wertheim, J.O. (2010). The Re-Emergence of H1N1 Influenza Virus in 1977: A Cautionary Tale for Estimating Divergence Times Using Biologically Unrealistic Sampling Dates. PLOS One *5*, e11184. https://doi.org/10.1371/journal.pone.0011184.

48. Rozo, M., and Gronvall, G.K. (2015). The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate. mBio *6*, e01013-15. https://doi.org/10.1128/mBio.01013-15.

49. Smith, W., Andrewes, C.H., and Laidlaw, P.P. (1933). A VIRUS OBTAINED FROM INFLUENZA PATIENTS. Lancet *222*, 66–68. https://doi.org/10.1016/S0140-6736(00)78541-2.

50. Zhang, G., Shoham, D., Gilichinsky, D., Davydov, S., Castello, J.D., and Rogers, S.O. (2006). Evidence of Influenza A Virus RNA in Siberian Lake Ice. J. Virol. *80*, 12229–12235. https://doi.org/10.1128/JVI.00986-06.

51. Worobey, M. (2008). Phylogenetic Evidence against Evolutionary Stasis and Natural Abiotic Reservoirs of Influenza A Virus. J. Virol. *82*, 3769–3774. https://doi.org/10.1128/JVI.02207-07.

52. Herfst, S., Schrauwen, E.J.A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V.J., Sorrell, E.M., Bestebroer, T.M., Burke, D.F., Smith, D.J., et al. (2012). Airborne transmission of influenza A/H5N1 virus between ferrets. Science *336*, 1534–1541. https://doi.org/10.1126/science.1213362.

53. Morse, S.S., Mazet, J.A.K., Woolhouse, M., Parrish, C.R., Carroll, D., Karesh, W.B., Zambrana-Torrelio, C., Lipkin, W.I., and Daszak, P. (2012). Prediction and prevention of the next pandemic zoonosis. Lancet *380*, 1956–1965. https://doi.org/10.1016/S0140-6736(12)61684-5.

54. Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., et al. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature *503*, 535–538. https://doi.org/10.1038/nature12711.

55. Bausch, D.G., and Schwarz, L. (2014). Outbreak of Ebola Virus Disease in Guinea: Where Ecology Meets Economy. PLoS Negl. Trop. Dis. *8*, e3056. https://doi.org/10.1371/journal.pntd.0003056.

56. Dudas, G., and Rambaut, A. (2014). Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. PLoS Curr. *6*, ecurrents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d. https://doi.org/10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

57. Goldstein, T., Anthony, S.J., Gbakima, A., Bird, B.H., Bangura, J., Tremeau-Bravard, A., Belaganahalli, M.N., Wells, H.L., Dhanota, J.K., Liang, E., et al. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. Nat. Microbiol. *3*, 1084–1089. https://doi.org/10.1038/s41564-018-0227-2.

58. Akoi Boré, J., Timothy, J.W.S., Tipton, T., Kekoura, I., Hall, Y., Hood, G., Longet, S., Fornace, K., Lucien, M.S., Fehling, S.K., et al. (2024). Serological evidence of zoonotic filovirus exposure among bushmeat hunters in Guinea. Nat. Commun. *15*, 4171. https://doi.org/10.1038/s41467-024-48587-5.

59. Oxford, J.S., Corcoran, T., and Schild, G.C. (1980). Naturally Occurring Temperature-sensitive Influenza A Viruses of the H1N1 and H3N2 Subtypes. J. Gen. Virol. *48*, 383–389. https://doi.org/10.1099/0022-1317-48-2-383.

60. Klimov, A.I., Cox, N.J., Yotov, W.V., Rocha, E., Alexandrova, G.I., and Kendal, A.P. (1992). Sequence changes in the live attenuated, cold-adapted variants of influenza A/Leningrad/134/57 (H2N2) virus. Virology *186*, 795–797. https://doi.org/10.1016/0042-6822(92)90050-y.

61. Ullah, S., and Ross, T.M. (2022). Next generation live-attenuated influenza vaccine platforms. Expert Rev. Vaccin. *21*, 1097–1110. https://doi.org/10.1080/14760584.2022.2072301.

62. Wareing, M.D., and Tannock, G.A. (2001). Live attenuated vaccines against influenza; an historical review. Vaccine *19*, 3320–3330. https://doi.org/10.1016/S0264-410X(01)00045-7.

63. Van Damme, P., De Coster, I., Bandyopadhyay, A.S., Revets, H., Withanage, K., De Smedt, P., Suykens, L., Oberste, M.S., Weldon, W.C., Costa-Clemens, S.A., et al. (2019). The safety and immunogenicity of two novel live attenuated monovalent (serotype 2) oral poliovirus vaccines in healthy adults: a double-blind, single-centre phase 1 study. Lancet *394*, 148–158. https://doi.org/10.1016/S0140-6736(19)31279-6.

64. Yeh, M.T., Bujaki, E., Dolan, P.T., Smith, M., Wahid, R., Konz, J., Weiner, A.J., Bandyopadhyay, A.S., Van Damme, P., De Coster, I., et al. (2020). Engineering the Live-Attenuated Polio Vaccine to Prevent Reversion to Virulence. Cell Host Microbe *27*, 736–751.e8. https://doi.org/10.1016/j.chom.2020.04.003.

65. Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J., and Fitch, W.M. (1999). Predicting the Evolution of Human Influenza A. Science *286*, 1921–1925. https://doi.org/10.1126/science.286.5446.1921.

66. Furuse, Y., Shimabukuro, K., Odagiri, T., Sawayama, R., Okada, T., Khandaker, I., Suzuki, A., and Oshitani, H. (2010). Comparison of selection pressures on the HA gene of pandemic (2009) and seasonal human and swine influenza A H1 subtype viruses. Virology *405*, 314–321. https://doi.org/10.1016/j.virol.2010.06.018.

67. Whitmer, S.L.M., Ladner, J.T., Wiley, M.R., Patel, K., Dudas, G., Rambaut, A., Sahr, F., Prieto, K., Shepard, S.S., Carmody, E., et al. (2018). Active Ebola Virus Replication and Heterogeneous Evolutionary Rates in EVD Survivors. Cell Rep. *22*, 1159–1168. https://doi.org/10.1016/j.celrep.2018.01.008.

68. Keita, A.K., Koundouno, F.R., Faye, M., Düx, A., Hinzmann, J., Diallo, H., Ayouba, A., Le Marcis, F., Soropogui, B., Ifono, K., et al. (2021). Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks. Nature *597*, 539–543. https://doi.org/10.1038/s41586-021-03901-9.

69. Worobey, M., Han, G.-Z., and Rambaut, A. (2014). Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. Proc. Natl. Acad. Sci. USA *111*, 8107–8112. https://doi.org/10.1073/pnas.1324197111.

70. O'Toole, Á., Neher, R.A., Ndodo, N., Borges, V., Gannon, B., Gomes, J.P., Groves, N., King, D.J., Maloney, D., Lemey, P., et al. (2023). APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. Science *382*, 595–600. https://doi.org/10.1126/science.adg8116.

71. Sánchez, C.A., Li, H., Phelps, K.L., Zambrana-Torrelio, C., Wang, L.-F., Zhou, P., Shi, Z.-L., Olival, K.J., and Daszak, P. (2022). A strategy to assess spillover risk of bat SARS-related coronaviruses in Southeast Asia. Nat. Commun. *13*, 4380. https://doi.org/10.1038/s41467-022-31860-w.

72. Pekar, J., Worobey, M., Moshiri, N., Scheffler, K., and Wertheim, J.O. (2021). Timing the SARS-CoV-2 index case in Hubei province. Science *372*, 412–417. https://doi.org/10.1126/science.abf8003.

73. Dudas, G., Carvalho, L.M., Rambaut, A., and Bedford, T. (2018). MERS-CoV spillover at the camel-human interface. eLife *7*, e31257. https://doi.org/10.7554/eLife.31257.

74. Forni, D., Molteni, C., Cagliani, R., Clerici, M., and Sironi, M. (2023). Analysis of variola virus molecular evolution suggests an old origin of the virus consistent with historical records. Microb. Genomics *9*, mgen000932. https://doi.org/10.1099/mgen.0.000932.

75. Mühlemann, B., Vinner, L., Margaryan, A., Wilhelmson, H., de la Fuente Castro, C., Allentoft, M.E., de Barros Damgaard, P., Hansen, A.J., Holtsmark Nielsen, S., Strand, L.M., et al. (2020). Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science *369*, eaaw8977. https://doi.org/10.1126/science.aaw8977.

76. Li, Y., Carroll, D.S., Gardner, S.N., Walsh, M.C., Vitalis, E.A., and Damon, I.K. (2007). On the origin of smallpox: Correlating variola phylogenics with historical smallpox records. Proc. Natl. Acad. Sci. USA *104*, 15787–15792. https://doi.org/10.1073/pnas.0609268104.

77. Wertheim, J.O., and Kosakovsky Pond, S.L. (2011). Purifying Selection Can Obscure the Ancient Age of Viral Lineages. Mol. Biol. Evol. *28*, 3355–3365. https://doi.org/10.1093/molbev/msr170.

78. Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W. (2006). Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. Mol. Biol. Evol. *23*, 1891–1901. https://doi.org/10.1093/molbev/msl051.

79. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

80. Sagulenko, P., Puller, V., and Neher, R.A. (2018). TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. *4*, vex042. https://doi.org/10.1093/ve/vex042.

81. Delport, W., Scheffler, K., and Seoighe, C. (2008). Models of coding sequence evolution. Brief. Bioinform. *10*, 97–109. https://doi.org/10.1093/bib/bbn049.

82. Pekar, J.E., Lytras, S., Ghafari, M., Magee, A.F., Parker, E., Wang, Y., Ji, X., Havens, J.L., Katzourakis, A., Vasylyeva, T.I., et al. (2025). The recency and geographical origins of the bat viruses ancestral to SARS-CoV and SARS-CoV-2. Cell *188*, 3167–3183.e18. https://doi.org/10.1016/j.cell.2025.03.035.

83. Wisotsky, S.R., Kosakovsky Pond, S.L., Shank, S.D., and Muse, S.V. (2020). Synonymous Site-to-Site Substitution Rate Variation Dramatically Inflates False Positive Rates of Selection Analyses: Ignore at Your Own Peril. Mol. Biol. Evol. *37*, 2430–2439. https://doi.org/10.1093/molbev/msaa037.

84. Anisimova, M., Bielawski, J.P., and Yang, Z. (2002). Accuracy and Power of Bayes Prediction of Amino Acid Sites Under Positive Selection. Mol. Biol. Evol. *19*, 950–958. https://doi.org/10.1093/oxfordjournals.molbev.a004152.

85. Gibbs, A.J., Armstrong, J.S., and Downie, J.C. (2009). From where did the 2009 "swine-origin" influenza A virus. Virol. J. *6*, 207. https://doi.org/10.1186/1743-422X-6-207.

86. Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Délicat, A., Paweska, J.T., Gonzalez, J.-P., and Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. Nature *438*, 575–576. https://doi.org/10.1038/438575a.

87. Seifert, S.N., Fischer, R.J., Kuisma, E., Badzi Nkoua, C.B., Bounga, G., Akongo, M.-J., Schulz, J.E., Escudero-Pérez, B., Akoundzie, B.-J., Ampiri, V.R.B., et al. (2022). Zaire ebolavirus surveillance near the Bikoro region of the Democratic Republic of the Congo during the 2018 outbreak reveals presence of seropositive bats. PLoS Negl. Trop. Dis. *16*, e0010504. https://doi.org/10.1371/journal.pntd.0010504.

88. Lam, H.M., Ratmann, O., and Boni, M.F. (2018). Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. Mol. Biol. Evol. *35*, 247–251. https://doi.org/10.1093/molbev/msx263.

89. Lam, T.T.-Y., Jia, N., Zhang, Y.-W., Shum, M.H.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., Liao, Y.-S., et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature *583*, 282–285. https://doi.org/10.1038/s41586-020-2169-0.

90. Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.-J., Li, N., Guo, Y., Li, X., Shen, X., et al. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature *583*, 286–289. https://doi.org/10.1038/s41586-020-2313-x.

91. Düx, A., Lequime, S., Patrono, L.V., Vrancken, B., Boral, S., Gogarten, J.F., Hilbig, A., Horst, D., Merkel, K., Prepoint, B., et al. (2020). Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science *368*, 1367–1370. https://doi.org/10.1126/science.aba9411.

92. Graepel, K.W., Lu, X., Case, J.B., Sexton, N.R., Smith, E.C., and Denison, M.R. (2017). Proofreading-Deficient Coronaviruses Adapt for Increased Fitness over Long-Term Passage without Reversion of Exoribonuclease-Inactivating Mutations. mBio *8*, e01503-17. https://doi.org/10.1128/mBio.01503-17.

93. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. *9*, 286–298. https://doi.org/10.1093/bib/bbn013.

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Virus genomes	NCBI-GenBank, GISAID	Listed in Table S6
Software and algorithms		
MAFFT (v7.409)	Katoh et al.[93]	https://mafft.cbrc.jp/alignment/software/
HyPhy – GARD (v2.5.42)	Kosakovsky Pond et al.[78]	github.com/veg/hyphy/
IQtree2 (v2.0.4)	Minh et al.[79]	http://www.iqtree.org/
HyPhy – RELAX (v2.5.62)	Wertheim et al.[27]	github.com/veg/hyphy/
TreeTime (v0.8.6)	Sagulenko et al.[80]	https://github.com/neherlab/treetime

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The sources of bioinformatic data analyzed in this study are available in the key resources table.

METHOD DETAILS

Phylogenetic selection analysis

For all influenza A virus data sets, coding regions were extracted from each genomic segment. Then each region was aligned using MAFFT (v7.409)[77]; Alignments were checked manually. For all other data sets, full genomes were aligned to the reference genome and split into non-recombinant regions. For MHV, a positive-sense RNA virus, putative non-recombinant regions were identified using GARD (v2.5.42).[78] For SARS-CoV and SARS-CoV-2, predefined putative non-recombinant regions were used.[40,82] For all data sets, coding regions were extracted based on the reference genome annotation. For each region, a maximum likelihood phylogeny was inferred with IQtree2 (v2.0.4)[79] using the GTR+F+Γ_4 model.

Change in intensity of selection was tested across all genomic regions with the multi-region (joint) RELAX model implemented in HyPhy (v2.5.62).[27] In the RELAX testing framework, we partition branches in a phylogenetic tree into "*test*" and "*reference*", with an optional "*nuisance*" set for the remaining (if any) branches. Evolution along each tree branch is described using a continuous time discrete space Markov model of codon evolution (for a review see Delport et al.[81]); the ratio of non-synonymous to synonymous relative substitution rates (ω) is used for model selection. Along each branch, the evolutionary process is modeled as an independent draw from discrete distributions on ω, with one rate allocated to positive selection ($\omega \geq 1$) and one or two rates allocated to negative selection or neutral evolution ($\omega \leq 1$). Phylogenetic likelihood is computed by integrating over the assignments of ω to branches, and all model parameters, including ω distributions, are estimated by maximum likelihood. The selection regimes (ω distributions) of the test and reference branches are related by a selection intensity parameter K, via $\omega(\text{test}) = \omega(\text{background})^K$. Selection intensity (K) greater than 1 indicates intensification, where positive and purifying increase in strength (move away from $\omega = 1$), and K less than 1 indicates relaxation, where positive and purifying selection decrease in strength (move closer to $\omega = 1$). As described in Wertheim (2014)[27], we use the likelihood ratio test, and the asymptotic approximation to the test statistic distribution under the null (chi squared distribution and 1 degree of freedom), to compare the null model with K = 1 (when the selection regime in test and reference are equal) to the alternative model of an unrestricted K. This comparison can determine whether the null model model (selection regimes are equal) can be rejected in favor of the alternative model (relaxation or intensification of selection).

The joint RELAX test is an extension of RELAX in which multiple evolutionary histories (i.e., trees) share the inferred selection regime (ω distributions) and selection intensity test parameter (K). We ran joint RELAX for each context with every combination of (i) with and without multiple hits, which allows instantaneous multi-nucleotide substitutions (MNS), (ii) with and without synonymous rate variation (SRV), which accommodates site-to-site variation of synonymous substitution rates, and (iii) with 2 or 3 rate partitions, which has 1 or 2 partitions for purifying or neutral selection respectively. From these 8 options, the best model was chosen using small sample AIC (c-AIC).

Method sensitivity to sampling

Statistical properties of dN/dS estimators in codon models are not analytically tractable, but they have been studied extensively using simulations and using empirical data analyses.[83,84] To assess how sensitive our inference of the relative selective strength is to which sequences are included in the analyses, we performed a series of empirical robustness analyses, by changing the size and composition of reference background (reservoir) and foreground (stem and/or outbreak) sequences. We then applied RELAX, then

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

compared them to the original analyses. Depending on the scenario (explained below), we expected either minimal changes, or specific, predictable changes. We describe sensitivity analyses within individual dataset descriptions. For RELAX, we used the best fitting model (number of rates, presence or absence of SRV, support for MNS) chosen during the primary complete dataset analysis.

Data sets

Selection regime preceding and during the H1N1 2009 pandemic (H1N1pdm09)

To characterize the selection regime preceding H1N1pdm09, we used previously curated virus genomes,[85] with up to 58 stains for each of the 8 influenza genome segments. Additionally, to characterize the selection regime during the initial human outbreak, we considered 39 stains randomly subset from Influenza Virus Resource search with parameters of pandemic H1N1 virus that are full length with all segments sampled from a human host in North America from the FLU project collected from 1 April 2009 through 1 March 2010. We used branches associated with transmission between swine viruses, to represent selection associated with the host reservoir (reference partition). The stem branch leading to the MRCA of H1N1pdm09 was used as the test branch of interest. Additionally, selection on the stem branch preceding H1N1pdm09 and branches within the outbreak clade were tested against the selection regime of the same reference partition.

Sensitivity Analysis. We hypothesized that if we used the stem lineage together with the early isolates as the foreground (keeping swine isolates as the background), the choice of the sampling window for the human lineages would have a predictable effect on the estimates of K. Specifically, the original analysis noted that stem and outbreak sequences combined were evolving under relaxed selection (compared with the reservoir), but stem alone was not significantly different. Hence, as we added human sequences sampled later during the human outbreak to the stem alone, the stem and outbreak selective regime would eventually tip into being relaxed compared to the reservoir.

We further evaluated robustness by selecting a different set of human isolates compared to the original analysis. We queried NCBI for all complete (all 8 segments) H1N1 influenza A virus isolates from a human host, with recorded collection dates between April 2009 and April 2010. We ascertained that these were H1N1 swine origin sequences by computing pairwise genetic distances from the NCBI RefSeq sequence. We then randomly selected up to 10 isolates (not all two month intervals had 10 isolates) for each two-month sampling window (1 April 2009 through 31 May 2009, 1 June 2009through 31 July 2009/, etc.), and conducting a rolling time-series analysis comparing stem and human outbreak clade to background.

Selection regime preceding and during Ebolavirus 2013–2016 West Africa Ebola outbreak

We assembled a dataset of 101 full length Ebola virus genomes: 61 genomes were from the West African Ebola epidemic (collected 1 January 2014 through 1 June 2015) and the remaining 40 were representative of all previous known Ebola Zaire outbreaks in humans during which a viral genome had been sequenced. All coding regions concatenated in a single open-reading frame, excluding the sections of the glycoprotein gene containing overlapping reading frames.

We identified branches between human Ebola outbreaks, to represent selection associated with the presumed bat natural host reservoir (reference partition).[86,87] Branches that were associated within human outbreak clades were excluded. The long branch leading to the MRCA of the 2013–2016 West African outbreak was used as the test branch of interest. Additionally, selection on the stem branch preceding the 2013–2016 outbreak and branches within the outbreak clade, was tested against the selection regime of the same reference partition.

Sensitivity Analysis. Similar to H1N1pdm2009 above, we ran selected subsets of human EBOV genomes from 2014–2015 with fully resolved collection dates, occurring at or before a particular date, and ran RELAX using the stem lineage and the branches between human isolates as the foreground.

Selection regime preceding Marburg virus disease Angolan outbreak

We analyzed all Marburg virus genomes sampled from bats available on NCBI (n = 12) and 14 samples from the 2004–2005 outbreak in Angola,[8] which is the largest known human outbreak of Marburg virus disease to date. Marburg virus is a negative-sense single-stranded RNA virus, and we concatenated all coding regions for a single genome wide analysis. The branches associated with bat samples were used to represent selection in the natural host reservoir, which is understood to be bats. The stem branch leading to the human outbreak was used as the test branch of interest.

Selection regime preceding and during mpox virus outbreak

We analyzed the coding regions of 201 mpox virus (MPXV) genomes, including 185 from Clade IIb representing the initial mpox outbreak in Nigeria. We masked all sites associated with APOBEC3 activity (C in the dinucleotide context TC, and the reverse complement of G in the dinucleotide context GA) to prevent this signature from biasing our results. Samples were chosen based on previous literature.[35] MPXV is a double-stranded DNA genome, and recombination was not assessed. The branches between isolated human cases and outbreaks (excluding outbreak clades) were used as reference branches to characterize selection in the natural host reservoir. We compared background selection to stem branches of one branch that leads to the outbreak, which is suggested to include some amount of host reservoir transmission and some human-to-human transmission based on the signature of APOBEC3 activity,[35] and the parent branch which represents evolution prior to emergence. Sequences from Clade IIb represent the human outbreak starting in 2014, identified in 2017, that lead to the 2022–2023 global mpox epidemic.

Selection regimes in SARS-CoV-like virus

In 2002–2004, SARS-CoV spilled over into humans, likely from palm civets (*Paguma larvata*) sold for human consumption in live-animal markets in China.[36] We used a total of 142 sequences, including 3 SARS-CoV sequences (2 from humans associated with recent

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

introductions and 1 from a palm civet), using 31 predefined putative non-recombinant regions.[16,82] We considered selection preceding the SARS outbreak by comparing the stem branch leading to the MRCA of SARS-CoV civet and human samples to the branches associated with a bat host. We also compared the early outbreak, using all branches in the SARS-CoV clade and the branch preceding it with selection in the host reservoir, using the same reference partition.

Sensitivity Analysis. We considered the power to detect changes in selective pressure as a function of which branches were included in the foreground set with the stem. The complete dataset includes a civet isolate (AY304486) and two human isolates (AY394995, AY394996). We ran RELAX by restricting the foreground to one or two sequences, including all six possible combinations. Additionally, we drastically downsampled the background lineages to determine the impact of background sampling on detection of a change in selection.

Selection regimes in SARS-CoV-2-like virus

We considered the selection of SARS-CoV-2-like virus preceding COVID-19 pandemic and compared it to selection of SARS-CoV-2-like sarbecoviruses sampled from the natural bat host reservoir. We analyzed a total 31 genomes, previously used to characterize zoonotic evolution[82] with 15 putatively non-recombinant regions pre-defined from previous literature.[40] The robustness of our selection analysis to changes in defining non-recombinant regions was tested by repeating the RELAX analyses with non-recombinant regions inferred by 3seq,[88] which identified 21 regions. We considered selection preceding the COVID-19 outbreak by comparing the stem branch leading to SARS-CoV-2, represented by the reference sequence Hu-1 to branches associated with the natural host reservoir of bats.

We also considered selection early in the COVID-19 pandemic using two different contexts. First, we included a subsample of 50 SARS-CoV-2 genomes representing the early outbreak in China, over 3 months, from a previous dataset.[43] Second, we analyzed a random sample of 39 sequences from NCBI Virus with complete genomes sampled from 15 September 2020 to 1 October 2020 to represent early worldwide transmission. For each of these sequence sets representing the early outbreak, we compared selection of the host reference partition to the outbreak clade and the stem leading to SARS-CoV-2.

Of the non-bat SARS-CoV-2-like sarbecoviruses, all 8 were sampled from pangolins, which form 2 clades.[89,90] The stem leading to the pangolin clades, and the pangolin clades were excluded from the reservoir host background.

Sensitivity Analysis. We extracted full-length (>29,000nt) SARS-CoV-2 genomes with known collection dates from GenBank, further requiring fewer than 1% total length and fewer than 10 consecutive ambiguous nucleotides. Next, we randomly selected one sequence every three months, starting in January 2020 and ending in January 2023. For each "monthly" sample, we segmented it into 15 non-recombinant segments used for the original analysis by aligning to the corresponding segment consensus sequence (with the cawlign tool). We then replaced all the human viral genomes in the phylogenetic tree with this sequence, on a per-segment basis, and ran the RELAX analysis using the monthly sample as foreground, and the same background as was used for the original analysis. The goal of these analyses is two-fold: assess the sensitivity of inference to the specific sample used, and the effect of sampling delay on inference.

Selection of H1N1 influenza A viruses

In 1957 H1N1 influenza A was replaced by another strain of influenza A virus: H2N2. In 1977, an H1N1 strain nearly identical to that which circulated in the 1950s reemerged in the human population.[46] We used 100 genomes sampled from before and after 1977 to build a phylogenetic tree and define the branch leading to the 1977 influenza pandemic as the stem. To characterize typical human-to-human transmission, we considered the branches descended from the 1977 influenza pandemic. Genomes were chosen based on previous curated analysis.[47] We excluded branches associated with samples from before 1977 from the background due to concerns of laboratory passaging, erroneous inclusion of lab strains, and sequence quality. As a robustness check, we performed an additional analysis excluding 49 tips associated with post-1977 isolates with a documented history of laboratory passaging.

We inferred substitutions on the above tree using TreeTime (v0.8.6)[80] ancestral state reconstruction with the default parameters. We determined synonymous and nonsynonymous substitutions by comparing the inferred mutations in the context of the alignment.

We then tested if laboratory passage of virus isolates changed selection regime by comparing the background with the clade that includes 3 early laboratory isolates derived from WSN33[49]: WSN/33, and two derivatives A/United Kingdom/1/1933 and A/Wilson-Smith/1933(H1N1), all of which are known to be passaged before sequencing.

Selection of H1N1 influenza A viruses putatively from lake ice

Zhang et al.,[50] purportedly isolated H1N1 influenza A virus from Serbian lake ice and suggested these samples were frozen in the ice from a local outbreak. We considered the 19 HA sequences purportedly isolated from lake ice, 1 positive control,[50] and 18 HA genes from the H1N1 dataset from the 1930s through 1950s, excluding tips associated with sequences that have a history of lab passaging[51] for a total of 38 sequences. Branches associated with transmission between humans were used for the reference partition, and branches in the 'lake ice' clade were used as the test partition.

Sensitivity Analysis. We removed the positive control and sampled (randomly) a single HA gene from "lake ice" samples to form a much smaller foreground, to see how well RELAX is able to infer selection intensity change with limited data. Additionally, we randomly downsampled the reference sequences to 25%, 50%, or 75% of the original size.

Selection on attenuated vaccine measles virus

We analyzed 28 measles virus full genomes that were assembled previously for a molecular clock analysis.[91] This set of genomes is characteristic of measles virus evolution within humans, and we used branches associated with these samples to characterize background selection. We then used 13 genomes from vaccine associated strains of measles virus, previously identified,[31] which were

Please cite this article in press as: Havens et al., Dynamics of natural selection preceding human viral epidemics and pandemics, Cell (2026), https://doi.org/10.1016/j.cell.2026.02.006

passaged in cell culture to produce an attenuated virus intended for use as a vaccine. Branches associated with the vaccine strains were used as the test branches.

Sensitivity Analysis. We reduced the foreground (vaccine strains) to see how well the method is able to infer selection intensity change with limited data. Firstly, we tested the foreground composed of a single vaccine strain; then 20 random combinations of two vaccine strains; and 100 random combinations of three vaccine strains as foreground. Additionally, we randomly downsampled the reference sequences to 25%, 50%, or 75% of the original size. Finally, we deliberately included an internal branch of natural evolution in the foreground set of passaged strains, thereby creating a model misspecification.

Selection on attenuated vaccine mumps virus

We analyzed 59 mumps virus genomes that were selected by subsetting from Nextstrain (https://nextstrain.org/mumps/na), and collecting the associated full genomes from NCBI. Branches associated with these samples were used to characterize background selection. We characterized mumps virus passaged in the lab for attenuation, with 10 mumps virus genomes associated with vaccine strains, which were previously identified.[31]

Serially passaged coronavirus MHV

MHV is a betacoronavirus naturally found in mice. Graepel et al. passaged MHV 250 times in cell culture.[92] We used the passaging control that did not include experimental modifications to characterize selection associated with laboratory passaging, using branches leading to the passaged virus from the input wild type strain. Background selection was characterized with branches between 5 MHV lineage defining strains. Eleven non-recombinant regions were inferred using GARD.

Artificial selection on influenza A virus H2N2

An H2N2 influenza A virus (wt A/Leningrad/134/57) was passaged from 34°C down to 25°C to produce the non-surface proteins for a reassortant live attenuated vaccine that that was cold adapted and had attenuated replication at human body temperature.[60] We characterized the selection regime of passaging with artificial selection as the branches that were descendant from the wildtype leading to isolates sequenced at passage 17, passage 47, and the input wild type sequence. The branch leading to the input wild type was excluded from the RELAX paritions. The background context for comparing to cold adaptation of an H2N2 stain was characterized by a random subsample of 50 full H2N2 genomes from Influenza Virus Resource, with all proteins available, excluding vaccine strains, sampled from a human host.

Artificial selection on influenza A virus H5N1

Herfst et al., passaged virus derived from A/Indonesia/5/2005 10 times in ferrets, which was then passaged twice more in two parallel sets of transmission.[52] We considered all branches in the clade of all available passaged sequences ($n = 6$) and the branch leading to this clade from the genome put into passaging (A/Indonesia/5/2005) as the test partition. The background context was characterized by a random subsample of 94 whole H5N1 genomes from Influenza Virus Resource, with all proteins available, sampled from avian sources from 1980 to 2005. We compared selection dynamics along with branches associated with the H5N1 samples from the avian population (reference partition). The branch leading to the human A/Indonesia/5/2005 from the avian samples was excluded from main RELAX analysis.

Sensitivity Analysis. We tested the impact of downsampling the foreground by removing all but one serially passaged H5N1 isolate (*CEIRS-CIP047-A_Indonesia_5_2005.F7*) from the foreground. Next, we introduced model misspecification, by removing the unpassaged reference strain (A/Indonesia/5/2005), and designating all passaged sequences and their MRCA stem as foreground. This partitioning conflates natural evolution of Avian H5N1 sequences with passaging. Finally, we randomly downsampled the background (reservoir) sequences (5%, 10%, 15%, 25%, 50%, and 75% of the original size).

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantification and statistical analyses were performed using procedures described in STAR Methods phylogenetic selection analysis with details on data inclusion described in "Data sets."

Supplemental figures

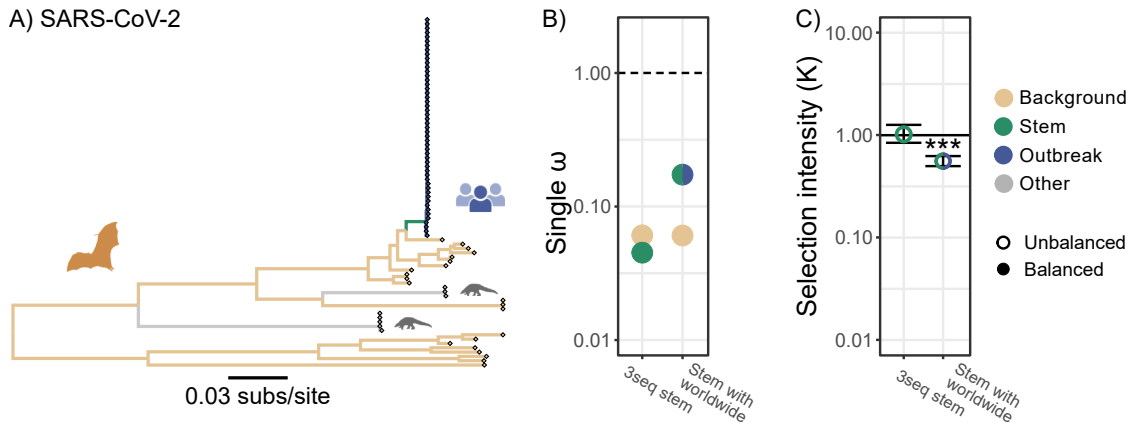


Figure S1. Quantifying selection regimes in SARS-CoV-2, related to Figure 5

(A) Phylogenetic tree of SARS-CoV-2-like sarbecoviruses, with the SARS-CoV-2 early worldwide outbreak characterized by 39 sequences sampled up to October 2020, non-recombinant region 08 branch color indicating partition.

(B) Single ω for each branch partition of background and the test sets, including 3seq stem, which compares the stem branch with background based on alignment of non-recombinant regions inferred with 3seq (no tree displayed), and "combined with worldwide," which compares stem and worldwide outbreak to background, color-coded to match the tree. Points are color-coded according to the branch set: background (brown), stem (green), or outbreak (blue).

(C) Change in selection intensity comparing the selection associated with the background with the test set. Filled circles indicate a balanced model where directionality is identifiable, open circles indicate an unbalanced model, and the direction above or below 1 is not identifiable. Numerical values in Table S1. Significance indicated with * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. Whiskers show 95% confidence interval. Icons created with BioRender.

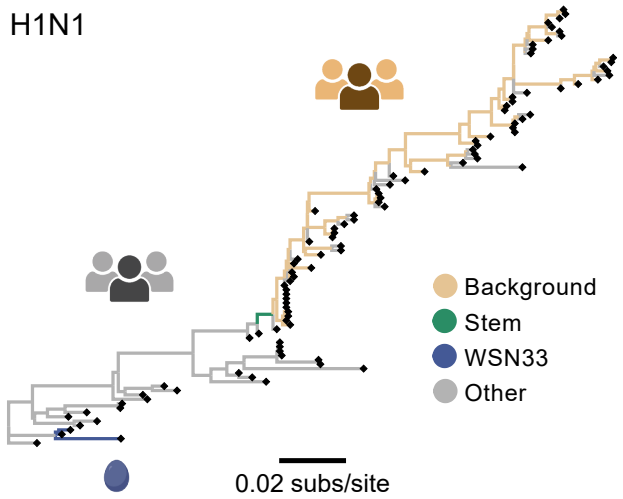


Figure S2. Phylogenetic tree of H1N1 HA segment, related to Figure 6

Branch color indicates a partition. The stem branch connects the 1977 pandemic reemergence and basal ancestors and is foreground for the 1977 flu test. WSN33 (blue) branches are clade-associated with that laboratory-passaged positive control. Post-1977 branches associated with human-to-human transmission (brown) used as background. Other branches that are pre-1977 or tips that have a history of passaging (gray) are excluded from analysis. Icons created with BioRender.

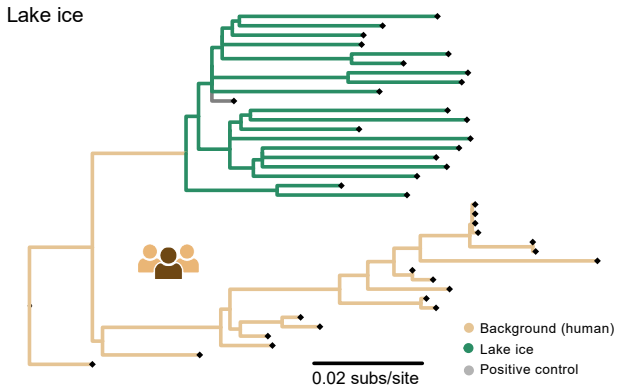


Figure S3. Phylogenetic tree of H1N1 HA segment with lake ice, related to Figure 7

Branch color indicates a partition. Lake ice clade (green) is the foreground, and branches associated with human-to-human transmission (brown) are the background of the lake ice test. The study positive control (gray) is excluded. Icons created with BioRender.

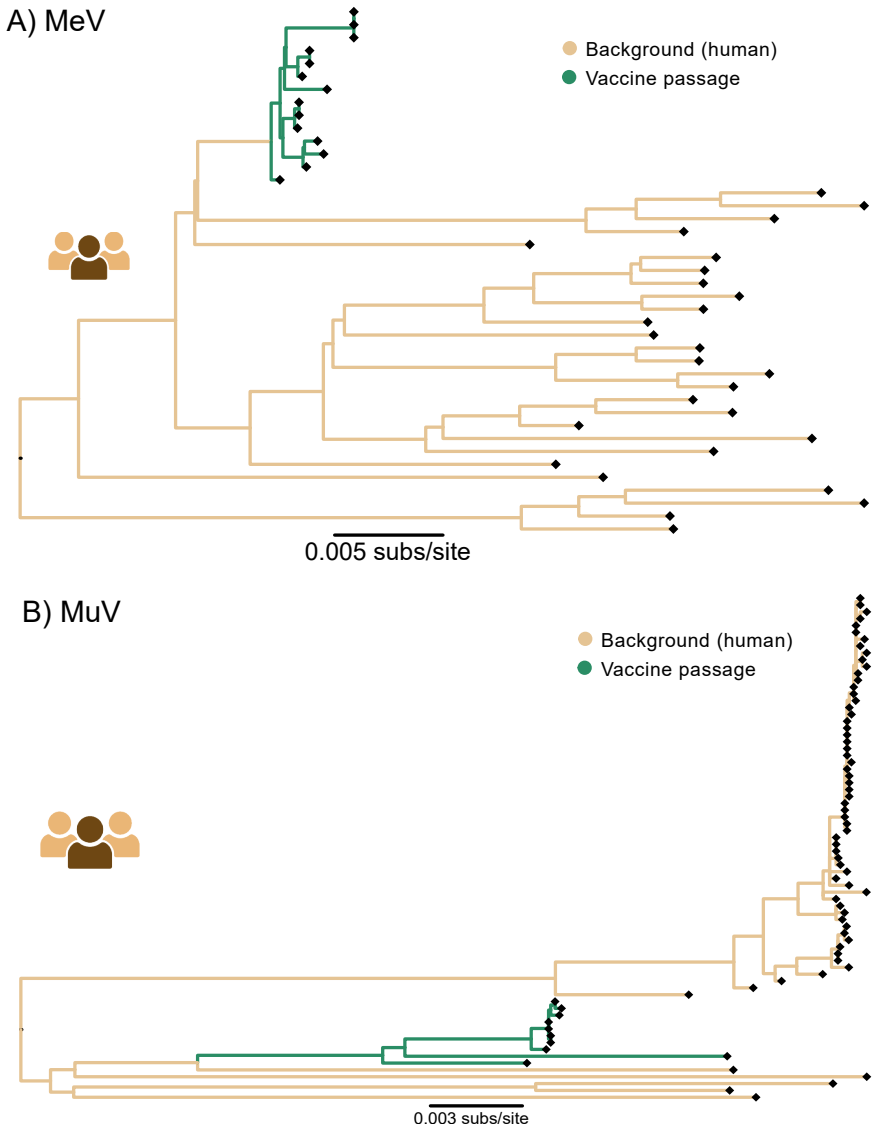


Figure S4. Phylogenetic tree of MeV genome and MuV genome, related to Figure 7

Branch color indicates a partition. Vaccine attenuated by passage (green) is the foreground, and branches associated with human-to-human transmission (brown) are the background. Icons created with BioRender.

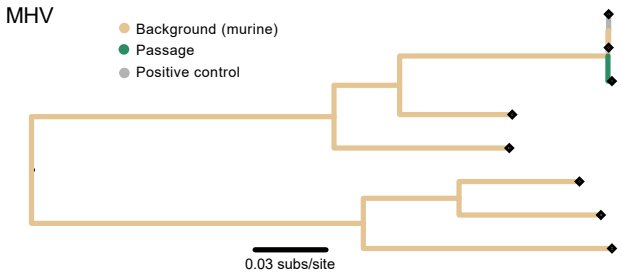


Figure S5. Phylogenetic tree of MHV non-recombinant region 07, defined by GARD, related to Figure 7

Branch color indicates a partition. Laboratory-passaged virus (green) is the foreground, and branches between murine strains (brown) are the background. The isolate used at the start of passaging (gray) is excluded.

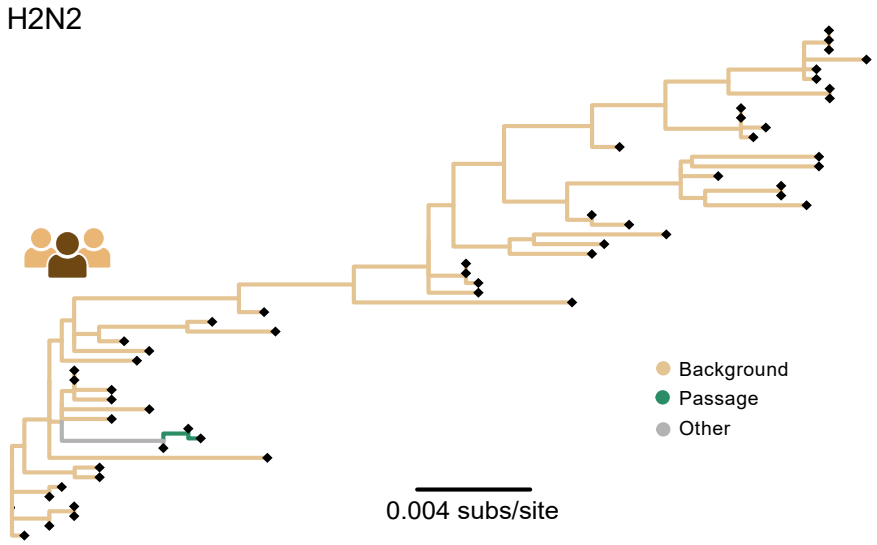


Figure S6. Phylogenetic tree of H2N2 PB1 segment, related to Figure 7

Branch color indicates a partition. Cold adaptation passage (green) is the foreground, and branches associated with human-to-human transmission (brown) are the background. The stem to isolate used at the start of passaging (gray) is excluded. Icons created with BioRender.

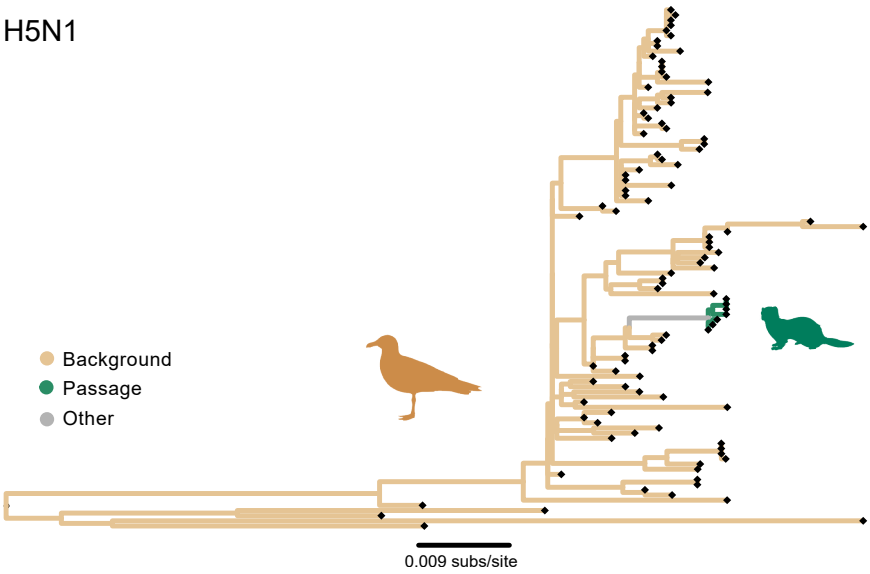


Figure S7. Phylogenetic tree of H5N1 PB1 segment, related to Figure 7

Branch color indicates a partition. Airborne ferret adaptation passage (green) is the foreground, and branches associated with the avian host reservoir (brown) are the background. The stem to isolate used at the start of passaging (gray) is excluded. Icons created with BioRender.