MDPI

Article

Dating the Emergence of Human Endemic Coronaviruses

Diego Forni [1],*[iD]**, Rachele Cagliani** [1][iD]**, Uberto Pozzoli** [1]**, Alessandra Mozzi** [1][iD]**, Federica Arrigoni** [2]**, Luca De Gioia** [2]**, Mario Clerici** [3,4][iD] **and Manuela Sironi** [1][iD]

[1] Bioinformatics, Scientific Institute IRCCS E. MEDEA, 23842 Bosisio Parini, Italy; rachele.cagliani@lanostrafamiglia.it (R.C.); uberto.pozzoli@lanostrafamiglia.it (U.P.); alessandra.mozzi@lanostrafamiglia.it (A.M.); manuela.sironi@lanostrafamiglia.it (M.S.)

[2] Department of Biotechnology and Biosciences, University of Milan-Bicocca, 20126 Milan, Italy; federica.arrigoni@unimib.it (F.A.); luca.degioia@unimib.it (L.D.G.)

[3] Department of Physiopathology and Transplantation, University of Milan, 20122 Milan, Italy; mario.clerici@unimi.it

[4] Don Carlo Gnocchi Foundation ONLUS, IRCCS, 20148 Milan, Italy

* Correspondence: diego.forni@lanostrafamiglia.it

Abstract: Four endemic coronaviruses infect humans and cause mild symptoms. Because previous analyses were based on a limited number of sequences and did not control for effects that affect molecular dating, we re-assessed the timing of endemic coronavirus emergence. After controlling for recombination, selective pressure, and molecular clock model, we obtained similar tMRCA (time to the most recent common ancestor) estimates for the four coronaviruses, ranging from 72 (HCoV-229E) to 54 (HCoV-NL63) years ago. The split times of HCoV-229E and HCoV-OC43 from camel alphacoronavirus and bovine coronavirus were dated ~268 and ~99 years ago. The split times of HCoV-HKU1 and HCoV-NL63 could not be calculated, as their zoonoticic sources are unknown. To compare the timing of coronavirus emergence to that of another respiratory virus, we recorded the occurrence of influenza pandemics since 1500. Although there is no clear relationship between pandemic occurrence and human population size, the frequency of influenza pandemics seems to intensify starting around 1700, which corresponds with the initial phase of exponential increase of human population and to the emergence of HCoV-229E. The frequency of flu pandemics in the 19th century also suggests that the concurrence of HCoV-OC43 emergence and the Russian flu pandemic may be due to chance.

Keywords: coronavirus; molecular dating; emerging infectious diseases; viral disease; virus evolution

1. Introduction

Coronaviruses (order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*) are a diverse group of positive-sense, single-stranded RNA enveloped viruses with high zoonotic potential [1–3]. In 2002, a highly pathogenic coronavirus, severe acute respiratory syndrome coronavirus (SARS-CoV), spilled over from palm civets to humans and caused ~8000 cases in several countries [4]. These events were followed by the appearance, in 2012, of Middle East respiratory syndrome coronavirus (MERS-CoV), a camel-derived pathogen that caused multiple outbreaks of respiratory disease mainly in the Arabic Peninsula [5]. Containment and surveillance strategies allowed for the control of these viruses, which have never (SARS-CoV) or only occasionally (MERS-CoV) reappeared in human populations [6,7]. However, at the end of 2019, SARS-CoV-2 emerged in China and is now recognized as the cause of COVID-19 [8]. The virus rapidly spread worldwide, and the World Health Organization declared the SARS-CoV-2 pandemic in early March 2020.

The epidemic behavior of SARS-CoV, MERS-CoV, and SARS-CoV-2, as well as their clinical severity, have clearly raised awareness of the potential danger posed by coronaviruses, which were considered relatively harmless to humans before 2002. In fact, four

other human coronaviruses (HCoV) (HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E), sometimes referred to as "common cold coronaviruses", have been circulating in human populations for decades, usually causing mild symptoms [2,9,10].

Like the highly pathogenic coronaviruses, the endemic coronaviruses have a zoonotic origin [2,3,11]. Although with some controversy [12], most previous estimates indicated that the endemic coronaviruses entered human populations in the last 1000 years [2,13–18]. However, these early analyses were often based on a small number of sequences and did not control for effects that are now recognized to affect molecular dating (e.g., recombination). For instance, using the first complete genome of HCoV-OC43 and the sequences of 15 BCoV spike proteins, Vijgen and coworkers estimated that the time to the most recent common ancestor (tMRCA) of HCoV-OC43 and BCoV dated to the end of the 19th century [14]. A follow-up analysis with seven additional HCoV-OC43 sequences and based on both the S and N sequences confirmed this estimate, placing the tMRCA at the end of the 19th century or at the beginning of the 20th [17]. A study that sequenced and analyzed several BCoV spike sequences also reached the same conclusion [18]. With respect to HCoV-229E, previous dating analysis did not include the camelid viruses, as they were unavailable at that time [13,15]. In the wake of the COVID-19 pandemic, a better understanding of the evolutionary dynamics of endemic coronaviruses, as well as of the tendency of viral disease emergence, might provide valuable insight into the possible trajectories of SARS-CoV-2 evolution.

2. Materials and Methods

2.1. Sequences and Alignments

Complete or almost complete genome sequences for all four endemic coronaviruses were downloaded from the NCBI database (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/, accessed on 15 April 2022). Only sequences with known sampling dates were included in the analyses (Table S1). The HCoV-OC43 Paris strain was excluded, as its sampling date is uncertain [19]. For HCoV-229E and HCoV-OC43, the closest phylogenetically related animal viruses were also retrieved, namely camel alphacoronavirus and bovine coronavirus (BCoV) (Table S1).

Sequence alignments were generated using MAFFT (v7.427) (multiple alignment using fast Fourier transform) [20], with default parameters.

Recombination Analysis

Recombination can affect phylogenetic tree branch length estimates and, consequently, molecular evolution analyses [21]. Thus, each coronavirus alignment was tested for evidence of recombination signals using the 3SEQ software (v.1.7, software for identifying recombination in sequence data) [22]. This method scans a given alignment searching for mosaic recombination signals in all possible sequence triplets. The result is the identification of genomic regions in which one of the three sequences is the recombinant (child) of the other two (parental). 3SEQ full scans were run with a recombination significance threshold of 0.01. All significant recombination events were mapped onto coronavirus alignments, and the longest non-recombinant genomic regions, defined as the genomic region between two recombination breakpoints, were selected for subsequent analyses (Figure 1). This generated four non-recombinant alignments with the following lengths: HCoV-OC43: 12,691 nucleotides, HCoV-229E: 17,271 nucleotides, HCoV-HKU1: 10,820 nucleotides, HCoV-NL62: 2320 nucleotides. The unique recombination events identified in genome alignments are shown in Figure 1.
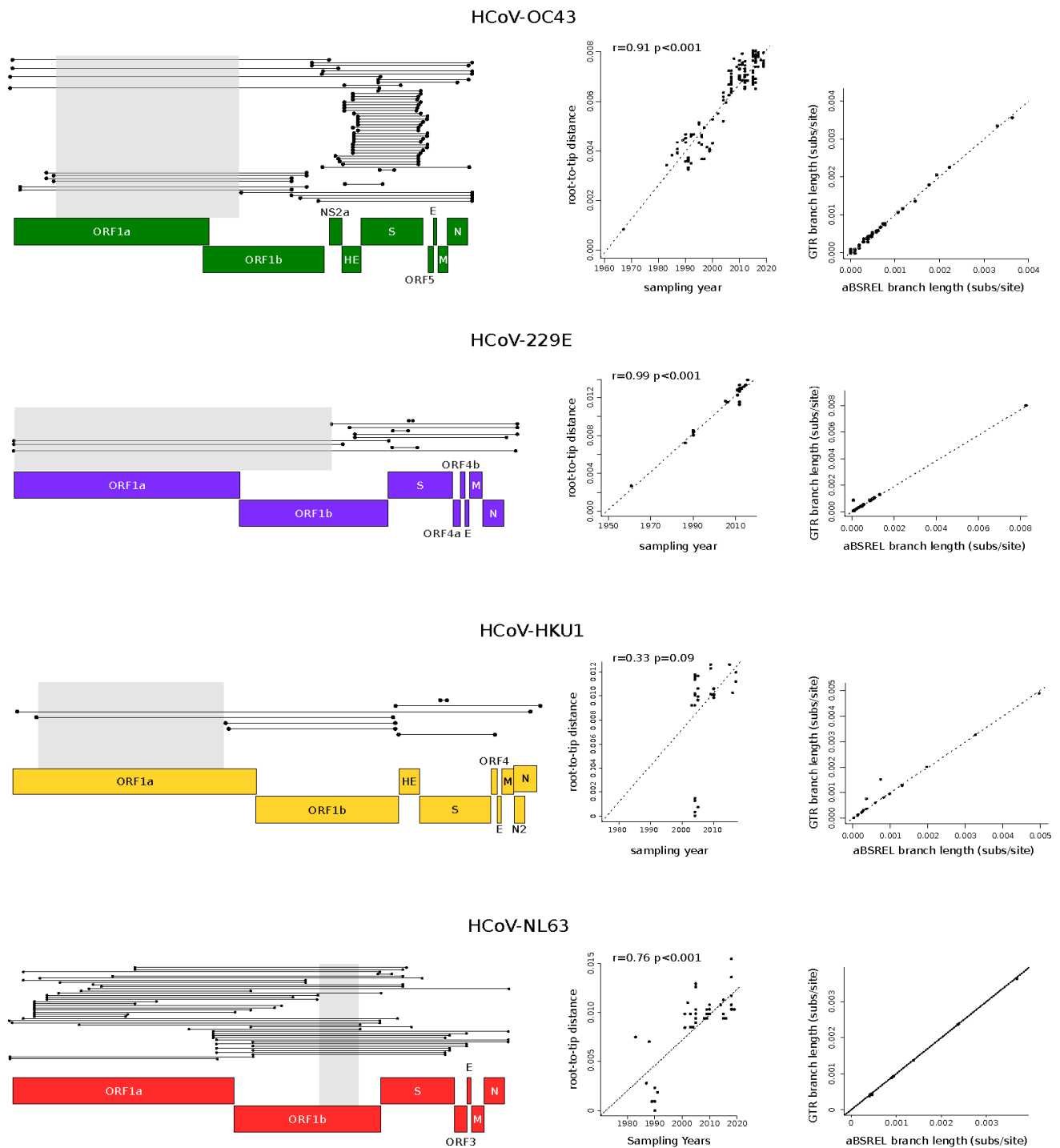
Figure 1. Recombination events and temporal signal. Unique recombination events in endemic coronaviruses (**left panels**). Each event is shown as a line with dots representing the start and the end. The non-recombinant regions used in the analyses are indicated with gray shadows. Schematic representations of coronavirus genomes are also reported. Plots of the root-to-tip distance as a function of sampling years are shown in the central panels. Each point corresponds to a viral sequence and the dotted line is the linear regression calculated using a method that minimizes the residual mean squares. The r coefficient and the corresponding *p* value are also shown. (**right panels**) report comparisons of branch lengths obtained using the aBSREL and the GTR models. Each dot represents an internal branch of a phylogenetic tree calculated using the longest non-recombinant regions of each endemic coronavirus.

2.2. Phylogenetic Trees and Temporal Signal

Phylogenetic trees for the non-recombinant regions of all endemic coronaviruses were reconstructed using the phyML (phylogenetic tree estimation under the maximum likelihood (ML) principle) software under a general time-reversible (GTR) model plus gamma-distributed rates and 4 substitution rate categories [23]. The substitution models were estimated using JmodelTest 2 [24,25].

Internal-GTR-estimated branch lengths were compared to branch lengths calculated using a model that accounts for different selective pressures among lineages. This model is implemented in the aBSREL (adaptive branch site random effects likelihood [26]) tool from the HYPHY (hypothesis testing using phylogenesis) suite (version 2.5) [27].

To evaluate whether the non-recombinant genomic regions selected for the analyses carried sufficient temporal signal, we calculated the correlation coefficients (r) of regressions of root-to-tip genetic distances against sequence sampling years. We applied a previously proposed method, described by Murray and co-workers [28], that minimizes the residual mean squares of the models rather than one that maximizes r^2 [28]. We calculated p values by performing clustered permutations (1000) of the sampling dates, as previously suggested [28,29]. We considered significant a regression with $p < 0.05$ (Figure 1).

2.3. Molecular Dating

A time estimate phylogenetic reconstruction was performed using a Bayesian approach implemented in the Bayesian Evolutionary Analysis by Sampling Trees (BEAST, v.1.10.4) software [30].

To select the best-fit molecular clock and tree prior, we ran the path sampling tool implemented in BEAST to choose between a constant size, an exponential growth, or a coalescent Bayesian skyline tree prior, and between a strict and an uncorrelated relaxed log-normal clock (100 steps, 1,000,000 iterations each). For all parameters, default priors were chosen only if they had proper distributions; otherwise, they were changed accordingly (i.e., for the population size parameter, an uninformative lognormal prior distribution was used, instead of the 1/x default prior).

A Bayes factor test was applied to compare the different likelihoods (Table S2). Since none of the models were favored when compared to all the others, we selected the simplest among the favored ones (Table S2). Thus, a constant population size tree prior with a strict clock model was used for HCoV-229E and HCoV-NL63, whereas a constant population and a relaxed clock with a log-normal distribution were used for HCoV-OC43.

For the HCoV-HKU1 phylogeny, we used the mean rates estimated for the other betacoronavirus HCoV-OC43 (1.78×10^{-4} substitutions per site yr^{-1}) as an informative rate prior following a normal distribution.

We performed two different Markov chain Monte Carlo runs for all four endemic coronaviruses, one hundred million iterations each, and sampled every 10,000 steps after a 10% burn-in. The runs were combined after checking for convergence and for heaving effective sampling sizes > 100.

We generated a maximum clade credibility tree using TreeAnnotator [31], which was visualized with FigTree (http://tree.bio.ed.ac.uk/, accessed on 15 April 2022).

2.4. Data on Influenza Pandemics and Human Population Size

The timing of influenza pandemics was obtained from a previous work [32], as well as from references therein [33–42]. Estimates of human population size were obtained from the "Our World in Data" website (https://ourworldindata.org/, accessed on 15 April 2022).

3. Results

3.1. Time-Frame of Human Endemic Coronavirus Emergence

As mentioned above, all endemic coronaviruses were estimated to have recently emerged as human pathogens [2,13–18]. However, besides being generally based on a limited number of sequences, most previous analyses did not include some of the viruses

that are now recognized to be closely related to endemic human coronaviruses (e.g., the dromedary camel alphacoronaviruses related to HCoV-229E). Moreover, it is now recognized that the presence of recombination, the lack of a temporal signal in the sequence data, and the pervasive effect of purifying selection can affect molecular dating [21,43,44]. Accounting for these effects has become common practice only in recent years. Thus, because their zoonotic source is known, we decided to reassess the timing of HCoV-OC43 and HCoV-229E emergence. As their animal origin is unknown, we instead estimated the time when circulating strains of HCoV-NL63 and HCoV-HKU1 last shared a common ancestor. For this purpose, we retrieved all available sequences with known sampling date for the four coronaviruses (HCoV-OC43, n = 167; HCoV-229E, n = 31; HCoV-NL63, n = 68; HCoV-HKU1, n = 34), together with the sequences of the reference genomes of BCoV and camel alphacoronavirus (Table S1).

Because recombination is known to be frequent in all coronavirus genera [45–47], we used 3SEQ to identify recombination events, which were detected in all datasets (Figure 1) [22]. Based on the location of breakpoint positions, we then selected the longest non-recombining region for each alignment. Specifically, we obtained relatively long regions for HCoV-OC43, HCoV-229E, and HCoV-HKU1, whereas the non-recombining region was shorter for HCoV-NL63 (Figure 1). For all the selected non-recombining regions, maximum likelihood phylogenetic trees were constructed, and we checked for the presence of a temporal signal by performing regression of root-to-tip genetic distances against sampling dates. These analyses indicated a strong temporal signal for all regions, with the exclusion of the HCoV-HKU1 region (Figure 1). In this latter case, the lack of a temporal signal is most likely due to the short time span among virus sampling dates, with the earliest sequences collected in 2003 (Table S1).

Before performing molecular dating, we evaluated whether natural selection strongly affected branch length estimates in the viral phylogenies. In fact, it is now recognized that purifying selection and saturation effects contribute to the time-dependent substitution rate variation in viruses, which, in turn, affects molecular dating [43,48]. We thus estimated branch lengths using the aBSREL (adaptive branch-site random effects likelihood) model, which accounts for different selective pressures among lineages and is relatively robust to substitution saturation [49]. For all phylogenies, branch lengths estimated with aBSREL were comparable to those obtained with a GTR (general time reversible) model (Figure 1), suggesting that molecular dating can be performed with minor effects related to the time dependency of substitution rates.

Thus, for the three phylogenies (HCoV-OC43, HCoV-229E, and HCoV-NL63) showing a temporal signal, we used a Bayesian approach to estimate substitution rates and time-measured evolutionary histories. Substitution rates in the range of 1.78×10^{-4} to 2.03×10^{-4} substitutions per site yr^{-1} were obtained, in line with previous analyses [46]. For the HCoV-HKU1 phylogeny, date estimates were obtained by using the substitution rate of HCoV-OC43 (another betacoronavirus) as a prior. For the circulating strains of all coronaviruses, we obtained similar tMRCA estimates, which ranged from 72 (HCoV-229E) to 54 (HCoV-NL63) years ago (Figure 2 and Figure S1).
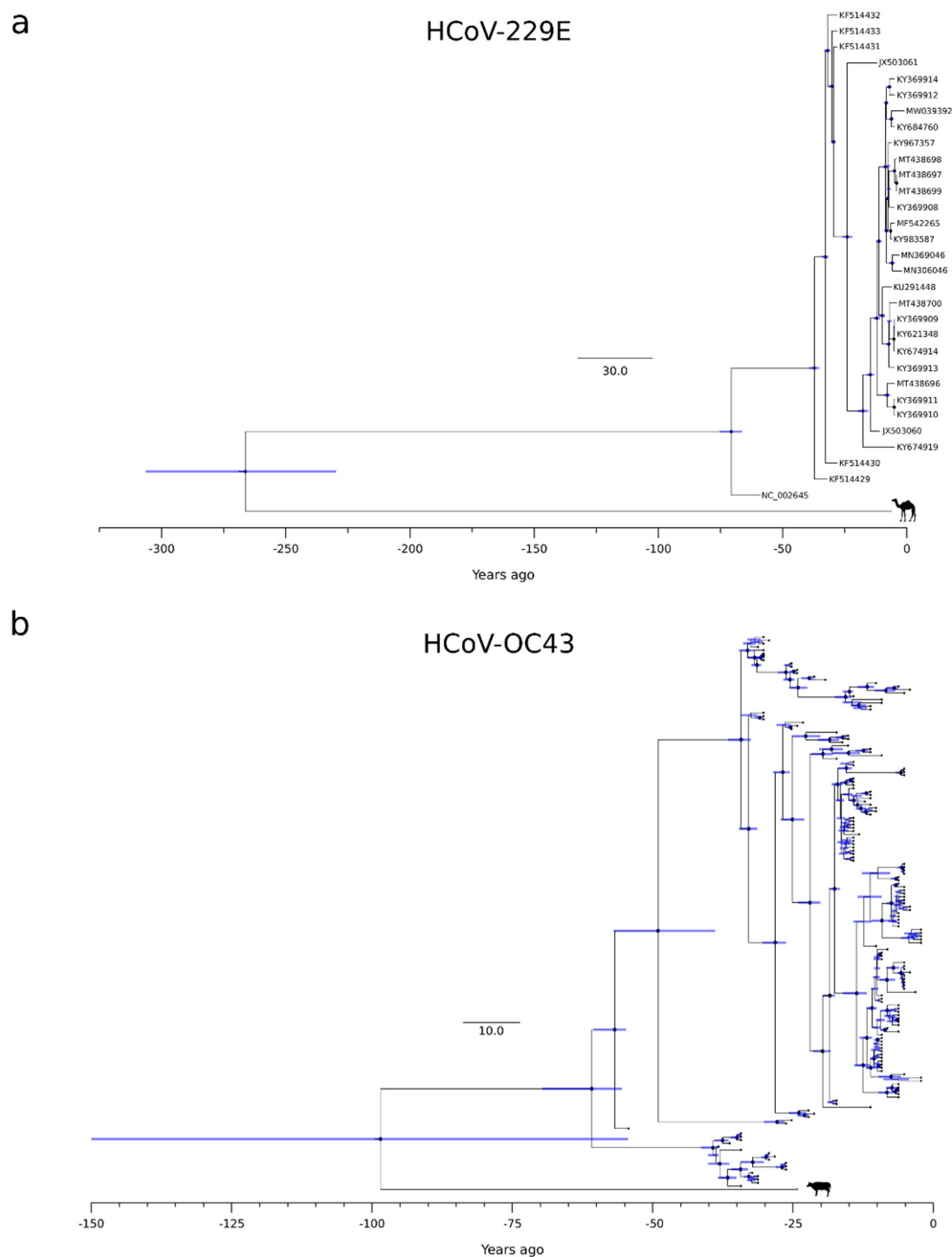
Figure 2. HCoV-229E and HCoV-OC43 timescaled phylogenetic trees. Maximum likelihood trees estimated for the non-recombinant region of HCoV-229E (**a**) and HCoV-OC43 (**b**). Branch lengths represent the evolutionary time measured by the grids corresponding to the timescale shown at the tree base (in years). For internal nodes, 95% HPD bars are shown, and black dots indicate a posterior probability > 0.80 for that node.

The splits of HCoV-OC43 and HCoV-229E from their most closely related animal viruses were more variable. Specifically, we estimated that HCoV-OC43 split from the bovine coronavirus (BCoV) lineage around 1923 (HPD: 1872–1967), whereas HCoV-229E separated from the camel alphacoronavirus in the 18th century (1754, HPD: 1714–1791) (Figure 2 and Figure S1). It should, however, be noted that the 95%HPD intervals for the split of HCoV-OC43 from the animal virus were very large, and the inference should, therefore, be taken with caution.

3.2. Human Coronavirus Emergence in the Context of Viral Outbreaks

The molecular dating analyses reported above indicate that, most likely, all endemic coronaviruses emerged as human pathogens earlier than 50 years ago, and possibly in a more distant past (Figure 3). This implies that, at least between ~1970 and 2002 (when SARS-CoV appeared), no coronavirus gained the ability to spread widely in our species. Thus, the pattern of coronavirus emergence seems to be highly irregular and to have intensified in recent years. To compare the timing of coronavirus emergence to that of another respiratory virus, we recorded the occurrence of known influenza pandemics since 1500. As previously noted [50], this pattern is also irregular. Although there is no clear relationship between pandemic occurrence and human population size, the frequency of influenza pandemics seems to intensify starting around 1700, which corresponds with the initial phase of the exponential increase of human population (Figure 3). This time also roughly corresponds to the emergence of HCoV-229E.
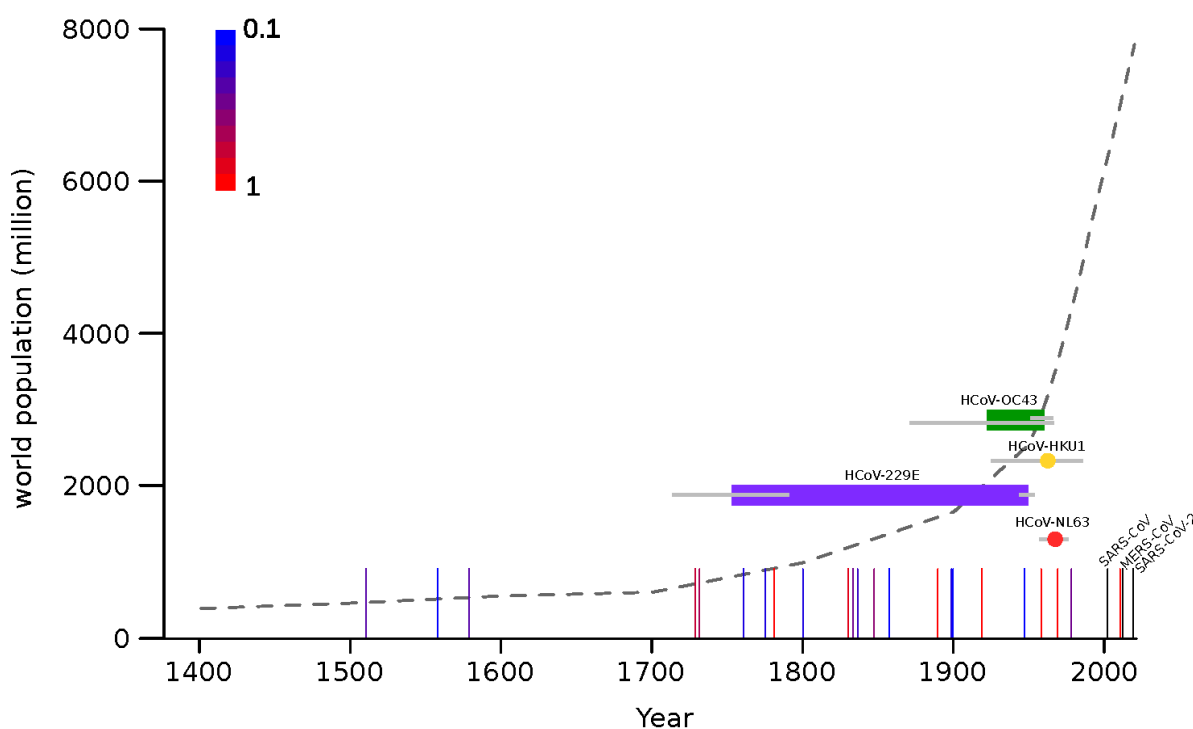

Figure 3. Timeline of endemic coronavirus emergence. Colored horizontal bars represent the time span between the divergence of each coronavirus from the closest known animal virus and the tMRCA (time to the most recent common ancestor) of circulating strains. In the case of HCoV-HKU1 and HCoV-NL63 tMRCA are shown as colored dots. Gray bars indicate 95% HPD. Vertical lines represent influenza pandemic events and the scaled colors (from blue to red, see legend on the plot) indicate the proportion of articles reporting that event as pandemic [33–42]. The three recent coronavirus zoonoses are shown as black vertical lines. The world human population count from 1400 C.E. is reported as a dotted gray line.

4. Discussion

Many uncertainties surround the origin of coronaviruses as human pathogens. Recent molecular dating analyses have estimated that the tMRCA of sarbecoviruses dates back to about 21,000 years ago [51], which roughly corresponds to the time when a set of human proteins that interact with coronaviruses started to experience positive selection in Asian populations [52]. Whether the selective pressure was accounted for by a coronavirus or another pathogen remains to be clarified. In the case that the agent that infected humans back then was indeed a betacoronavirus, it must have gone extinct, at least in human populations. Indeed, it seems difficult to imagine that a newly emerged coronavirus might

be able to spread and persist in early human communities, which were small and poorly connected. Whatever the nature of that early infectious agent, it is clear that much information on the possible trajectories of SARS-CoV-2 evolution can be gained by the analysis of previous human epidemics, especially those caused by coronaviruses. We thus leveraged the increasing availability of sequence data and advances in molecular dating approaches to re-estimate the time when endemic coronaviruses entered human populations.

Unfortunately, we could not estimate the emergence time for HCoV-HKU1 and HCoV-NL63, as the hosts from which they spilled over are unknown. Both viruses have their closest relatives in wild rodents and bats, but there is no information concerning the unsampled diversity in these mammals or in other domestic ones. Thus, for these two viruses, we can only estimate the time when circulating strains last shared a common ancestor, which was in the 1960s. Thus, we can place an upper-bound limit and state that they emerged earlier than ~50 years ago. This time roughly corresponds to the tMRCAs of HCoV-229E and HCoV-OC43. However, for these viruses, we estimated that the splits from the viruses hosted by domestic animals occurred in the 18th and 20th centuries, respectively. Because bovines and camels are plausible zoonotic sources for human infections, these split dates may be considered as good proxies for the time when HCoV-OC43 and HCoV-229E entered human populations. The long time spans that, especially for HCoV-229E, separate the split times and the tMRCAs are most likely accounted for either by extinct ancestral lineages or by unsampled viral diversity. In the case of HCoV-OC43, our estimate of the split from BCoV in 1923 (HPD: 1872–1967) is based on a large non-recombining region of ORF1a. This result is in good agreement with previous works that analyzed the *S* or *N* gene regions and obtained split dates ranging from 1873 to 1910 (plus credible intervals) [14,17,18]. Previous studies on HCoV-229E, as well as on HCoV-NL63, mainly analyzed the split times from bat viruses and are thus not comparable with the data we present herein [13,15]. In general, our results and previous studies agree that the four endemic coronaviruses entered human populations earlier than 50 years ago [14–18]. This indicates that, for several decades, no coronavirus spilled over to humans (or at least caused a registered outbreak) until three highly pathogenic coronaviruses emerged in tight temporal succession. The factors responsible for the timing of epidemics or pandemics have remained unknown for decades in the case of influenza, as virological and non-virological elements have been associated with the occurrence of pandemics with poor explanatory power [32,34,53–57]. Thus, it is presently impossible to predict the timing of such events and their severity. As noted elsewhere [50], there is no clear role of human population size in the frequency of pandemics either. Thus, the recent exponential growth has not determined a comparable increase in pandemic occurrence. However, some intensification seems to be detectable starting around the beginning of the 18th century, which also corresponds to the time when HCoV-229E emerged. This period was characterized by the industrial revolution and the colonial expansion, which resulted in larger cities and long-distance travels. Whether these changes in human behavior contributed to the intensification of viral disease emergence remains to be evaluated. Alternatively, it is possible that the increase in pandemic occurrence since 1700 simply reflects the increasing accuracy and reliability of medical or historical records.

In the same way as we are unable to predict the timing, we have very little ability to anticipate which viruses will emerge and how pathogenic they will be [58,59]. In this respect, it is worth mentioning that, because they have now circulated in (and adapted to) human populations for decades, if not centuries, we cannot exclude that the endemic coronaviruses were once more pathogenic than they are now. Indeed, it was previously suggested that the 1889–1890 flu pandemic (known as the Russian flu), which was characterized by pronounced central nervous system symptoms, was actually caused by HCoV-OC43 [14,60]. If we allow for credible intervals, our estimate of the timing of HCoV-OC43 emergence is still compatible with the hypothesis that the virus, which displays some neurotropism, was the causative agent of Russian flu. However, the frequency of flu pandemics in the 19th century suggests the concurrence of HCoV-OC43 emergence and the Russian flu pandemic may be

due to chance. Only the retrieval of historical samples from the pandemic will prove or refute this hypothesis.

References

1. Ye, Z.W.; Yuan, S.; Yuen, K.S.; Fung, S.Y.; Chan, C.P.; Jin, D.Y. Zoonotic Origins of Human Coronaviruses. *Int. J. Biol. Sci.* **2020**, *16*, 1686–1697. [CrossRef] [PubMed]
2. Forni, D.; Cagliani, R.; Clerici, M.; Sironi, M. Molecular Evolution of Human Coronavirus Genomes. *Trends Microbiol.* **2017**, *25*, 35–48. [CrossRef] [PubMed]
3. Cui, J.; Li, F.; Shi, Z.L. Origin and Evolution of Pathogenic Coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181–192. [CrossRef] [PubMed]
4. Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H.R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R.A.; et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976. [CrossRef]
5. Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.; Fouchier, R.A. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814–1820. [CrossRef]
6. Lipsitch, M.; Cohen, T.; Cooper, B.; Robins, J.M.; Ma, S.; James, L.; Gopalakrishna, G.; Chew, S.K.; Tan, C.C.; Samore, M.H.; et al. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* **2003**, *300*, 1966–1970. [CrossRef]
7. Peiris, M.; Perlman, S. Unresolved Questions in the Zoonotic Transmission of MERS. *Curr. Opin. Virol.* **2022**, *52*, 258–264. [CrossRef] [PubMed]
8. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [CrossRef]
9. Bucknall, R.A.; King, L.M.; Kapikian, A.Z.; Chanock, R.M. Studies with Human Coronaviruses. II. some Properties of Strains 229E and OC43. *Proc. Soc. Exp. Biol. Med.* **1972**, *139*, 722–727. [CrossRef]
10. Woo, P.C.; Lau, S.K.; Tsoi, H.W.; Huang, Y.; Poon, R.W.; Chu, C.M.; Lee, R.A.; Luk, W.K.; Wong, G.K.; Wong, B.H.; et al. Clinical and Molecular Epidemiological Features of Coronavirus HKU1-Associated Community-Acquired Pneumonia. *J. Infect. Dis.* **2005**, *192*, 1898–1907. [CrossRef]
11. Corman, V.M.; Muth, D.; Niemeyer, D.; Drosten, C. Hosts and Sources of Endemic Human Coronaviruses. *Adv. Virus Res.* **2018**, *100*, 163–188. [PubMed]
12. Brandão, P.E. Could Human Coronavirus OC43 have Co-Evolved with Early Humans? *Genet. Mol. Biol.* **2018**, *41*, 692–698. [CrossRef] [PubMed]
13. Huynh, J.; Li, S.; Yount, B.; Smith, A.; Sturges, L.; Olsen, J.C.; Nagel, J.; Johnson, J.B.; Agnihothram, S.; Gates, J.E.; et al. Evidence Supporting a Zoonotic Origin of Human Coronavirus Strain NL63. *J. Virol.* **2012**, *86*, 12816–12825. [CrossRef] [PubMed]
14. Vijgen, L.; Keyaerts, E.; Moes, E.; Thoelen, I.; Wollants, E.; Lemey, P.; Vandamme, A.M.; Van Ranst, M. Complete Genomic Sequence of Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event. *J. Virol.* **2005**, *79*, 1595–1604. [CrossRef] [PubMed]
15. Pfefferle, S.; Oppong, S.; Drexler, J.F.; Gloza-Rausch, F.; Ipsen, A.; Seebens, A.; Muller, M.A.; Annan, A.; Vallo, P.; Adu-Sarkodie, Y.; et al. Distant Relatives of Severe Acute Respiratory Syndrome Coronavirus and Close Relatives of Human Coronavirus 229E in Bats, Ghana. *Emerg. Infect. Dis.* **2009**, *15*, 1377–1384. [CrossRef]

16. Al-Khannaq, M.N.; Ng, K.T.; Oong, X.Y.; Pang, Y.K.; Takebe, Y.; Chook, J.B.; Hanafi, N.S.; Kamarulzaman, A.; Tee, K.K. Molecular Epidemiology and Evolutionary Histories of Human Coronavirus OC43 and HKU1 among Patients with Upper Respiratory Tract Infections in Kuala Lumpur, Malaysia. *Virol. J.* **2016**, *13*, 33. [CrossRef]

17. Vijgen, L.; Keyaerts, E.; Lemey, P.; Maes, P.; Van Reeth, K.; Nauwynck, H.; Pensaert, M.; Van Ranst, M. Evolutionary History of the Closely Related Group 2 Coronaviruses: Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human Coronavirus OC43. *J. Virol.* **2006**, *80*, 7270–7274. [CrossRef]

18. Bidokhti, M.R.M.; Tråvén, M.; Krishna, N.K.; Munir, M.; Belák, S.; Alenius, S.; Cortey, M. Evolutionary Dynamics of Bovine Coronaviruses: Natural Selection Pattern of the Spike Gene Implies Adaptive Evolution of the Strains. *J. Gen. Virol.* **2013**, *94*, 2036–2049. [CrossRef]

19. Vijgen, L.; Lemey, P.; Keyaerts, E.; Van Ranst, M. Genetic Variability of Human Respiratory Coronavirus OC43. *J. Virol.* **2005**, *79*, 3223–3224, author reply 3224-5. [CrossRef]

20. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

21. Schierup, M.H.; Hein, J. Recombination and the Molecular Clock. *Mol. Biol. Evol.* **2000**, *17*, 1578–1579. [CrossRef] [PubMed]

22. Lam, H.M.; Ratmann, O.; Boni, M.F. Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Mol. Biol. Evol.* **2018**, *35*, 247–251. [CrossRef] [PubMed]

23. Guindon, S.; Delsuc, F.; Dufayard, J.F.; Gascuel, O. Estimating Maximum Likelihood Phylogenies with PhyML. *Methods Mol. Biol.* **2009**, *537*, 113–137. [PubMed]

24. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. JModelTest 2: More Models, New Heuristics and Parallel Computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef]

25. Guindon, S.; Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [CrossRef]

26. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S.L. Less is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. [CrossRef]

27. Pond, S.L.; Frost, S.D.; Muse, S.V. HyPhy: Hypothesis Testing using Phylogenies. *Bioinformatics* **2005**, *21*, 676–679. [CrossRef]

28. Murray, G.G.; Wang, F.; Harrison, E.M.; Paterson, G.K.; Mather, A.E.; Harris, S.R.; Holmes, M.A.; Rambaut, A.; Welch, J.J. The Effect of Genetic Structure on Molecular Dating and Tests for Temporal Signal. *Methods Ecol. Evol.* **2016**, *7*, 80–89. [CrossRef]

29. Duchene, S.; Duchene, D.; Holmes, E.C.; Ho, S.Y. The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data. *Mol. Biol. Evol.* **2015**, *32*, 1895–1906. [CrossRef]

30. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian Phylogenetic and Phylodynamic Data Integration using BEAST 1.10. *Virus Evol.* **2018**, *4*, vey016. [CrossRef]

31. Bouckaert, R.; Heled, J.; Kuhnert, D.; Vaughan, T.; Wu, C.H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003537. [CrossRef] [PubMed]

32. Towers, S. Sunspot Activity and Influenza Pandemics: A Statistical Assessment of the Purported Association. *Epidemiol. Infect.* **2017**, *145*, 2640–2655. [CrossRef] [PubMed]

33. Hampson, A.W.; Mackenzie, J.S. The Influenza Viruses. *Med. J. Aust.* **2006**, *185*, S39–S43. [CrossRef] [PubMed]

34. Morens, D.M.; Taubenberger, J.K. Pandemic Influenza: Certain Uncertainties. *Rev. Med. Virol.* **2011**, *21*, 262–284. [CrossRef]

35. Mamelund, S. Influenza, Historical. *Medicine* **2008**, *54*, 361–371.

36. Lattanzi, M. Non-recent history of influenza pandemics, vaccines, and adjuvants. In *Influenza Vaccines for the Future*; Springer: Berlin, Germany, 2008; pp. 245–259.

37. Potter, C. Chronicle of Influenza Pandemics. In *Textbook of Influenza*; Nicholson, K.G., Webster, R., Hay, A., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 1997.

38. Garrett, L. *The Coming Plague: Newly Emerging Diseases in A World out of Balance*; Macmillan: New York, NY, USA, 1994.

39. Beveridge, W.I. The Chronicle of Influenza Epidemics. *Hist. Philos. Life Sci.* **1991**, *13*, 223–234.

40. Kilbourne, E.D. History of influenza. In *Influenza*; Springer: Berlin, Germany, 1987; pp. 3–22.

41. Pyle, G.F. *The Diffusion of Influenza: Patterns and Paradigms*; Rowman & Littlefield: Lanham, MD, USA, 1986.

42. Patterson, K.D. *Pandemic Influenza, 1700–1900: A Study in Historical Epidemiology*; Rowman & Littlefield: Totowa, NJ, USA, 1986.

43. Duchene, S.; Holmes, E.C.; Ho, S.Y. Analyses of Evolutionary Dynamics in Viruses are Hindered by a Time-Dependent Bias in Rate Estimates. *Proc. Biol. Sci.* **2014**, *281*, 2014.0732. [CrossRef]

44. Rieux, A.; Balloux, F. Inferences from Tip-Calibrated Phylogenies: A Review and a Practical Guide. *Mol. Ecol.* **2016**, *25*, 1911–1924. [CrossRef]

45. Graham, R.L.; Baric, R.S. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *J. Virol.* **2010**, *84*, 3134–3146. [CrossRef]

46. Boni, M.F.; Lemey, P.; Jiang, X.; Lam, T.T.; Perry, B.W.; Castoe, T.A.; Rambaut, A.; Robertson, D.L. Evolutionary Origins of the SARS-CoV-2 Sarbecovirus Lineage Responsible for the COVID-19 Pandemic. *Nat. Microbiol.* **2020**, *5*, 1408–1417. [CrossRef]

47. Forni, D.; Cagliani, R.; Sironi, M. Recombination and Positive Selection Differentially Shaped the Diversity of Betacoronavirus Subgenera. *Viruses* **2020**, *12*, 1313. [CrossRef] [PubMed]

48. Aiewsakun, P.; Katzourakis, A. Time-Dependent Rate Phenomenon in Viruses. *J. Virol.* **2016**, *90*, 7184–7195. [CrossRef] [PubMed]

49. Wertheim, J.O.; Chu, D.K.; Peiris, J.S.; Kosakovsky Pond, S.L.; Poon, L.L. A Case for the Ancient Origin of Coronaviruses. *J. Virol.* **2013**, *87*, 7039–7045. [CrossRef] [PubMed]

50. Morens, D.M.; Taubenberger, J.K. The Mother of all Pandemics is 100 Years Old (and Going Strong)! *Am. J. Public Health* **2018**, *108*, 1449–1454. [CrossRef]

51. Ghafari, M.; Simmonds, P.; Pybus, O.G.; Katzourakis, A. A Mechanistic Evolutionary Model Explains the Time-Dependent Pattern of Substitution Rates in Viruses. *Curr. Biol.* **2021**, *31*, 4689–4696.e5. [CrossRef]

52. Souilmi, Y.; Lauterbur, M.E.; Tobler, R.; Huber, C.D.; Johar, A.S.; Moradi, S.V.; Johnston, W.A.; Krogan, N.J.; Alexandrov, K.; Enard, D. An Ancient Viral Epidemic Involving Host Coronavirus Interacting Genes More than 20,000 Years Ago in East Asia. *Curr. Biol.* **2021**, *31*, 3504–3514.e9. [CrossRef]

53. Dowdle, W.R. Influenza A Virus Recycling Revisited. *Bull. World Health Organ.* **1999**, *77*, 820–828.

54. Hayes, D.P. Influenza Pandemics, Solar Activity Cycles, and Vitamin D. *Med. Hypotheses* **2010**, *74*, 831–834. [CrossRef]

55. Snyder, M.R.; Ravi, S.J. 1818, *1918*, 2018: Two Centuries of Pandemics. *Health. Secur.* **2018**, *16*, 410–415. [CrossRef]

56. Taubenberger, J.K.; Morens, D.M. Pandemic Influenza–Including a Risk Assessment of H5N1. *Rev. Sci. Tech.* **2009**, *28*, 187–202. [CrossRef]

57. Viboud, C.; Lessler, J. The 1918 Influenza Pandemic: Looking Back, Looking Forward. *Am. J. Epidemiol.* **2018**, *187*, 2493–2497. [CrossRef] [PubMed]

58. Holmes, E.C.; Rambaut, A.; Andersen, K.G. Pandemics: Spend on Surveillance, Not Prediction. *Nature* **2018**, *558*, 180–182. [CrossRef] [PubMed]

59. Morse, S.S.; Mazet, J.A.; Woolhouse, M.; Parrish, C.R.; Carroll, D.; Karesh, W.B.; Zambrana-Torrelio, C.; Lipkin, W.I.; Daszak, P. Prediction and Prevention of the Next Pandemic Zoonosis. *Lancet* **2012**, *380*, 1956–1965. [CrossRef]

60. Rozen, T.D. Daily Persistent Headache After a Viral Illness during a Worldwide Pandemic may Not be a New Occurrence: Lessons from the 1890 Russian/Asiatic Flu. *Cephalalgia* **2020**, *40*, 1406–1409. [CrossRef] [PubMed]