

Betacoronaviruses

Subjects: Virology | Pathology

Contributor: Ricardo Leonart

In the 21st century, three highly pathogenic betacoronaviruses have emerged, with an alarming rate of human morbidity and case fatality. Genomic information has been widely used to understand the pathogenesis, animal origin and mode of transmission of coronaviruses in the aftermath of the 2002–2003 severe acute respiratory syndrome (SARS) and 2012 Middle East respiratory syndrome (MERS) outbreaks. Furthermore, genome sequencing and bioinformatic analysis have had an unprecedented relevance in the battle against the 2019–2020 coronavirus disease 2019 (COVID-19) pandemic, the newest and most devastating outbreak caused by a coronavirus in the history of mankind. Here, we briefly review the application of genomics and bioinformatics in the molecular epidemiology of pathogenic betacoronaviruses.

Keywords: betacoronavirus ; genomics ; molecular epidemiology ; SARS-CoV ; MERS-CoV ; SARS-CoV-2

1. Introduction

Molecular epidemiology focuses on the contribution of genomic, genetic and other molecular factors to etiology, distribution and prevention of diseases. Central to molecular epidemiology of coronaviruses is their circulation among different animal hosts, as well as the evolutionary forces that facilitate these cross-species jumps. Here, we discuss how genomic information has been used to better understand the rate of evolution of betacoronaviruses and their transmission in human populations, as well as the evolutionary changes associated with host and tissue tropism.

Coronaviruses (CoV) are important pathogens of vertebrates with the ability to cause respiratory, enteric and systemic diseases in humans and animals. They are enveloped, single-stranded, positive-sense RNA viruses belonging to subfamily *Orthocoronavirinae* of family *Coronaviridae*, order *Nidovirales*. The subfamily is further divided into four genera, namely, *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*. The majority of clinically relevant coronaviruses belong to the *Alphacoronavirus* and *Betacoronavirus* genera. Genus *Alphacoronavirus* comprises species infecting a diverse group of mammals, including two (229E and NL63) of the seven known species of human coronaviruses. In the case of genus *Betacoronavirus*, the International Committee on Taxonomy of Viruses (ICTV) currently divides it into five subgenera, *Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* and *Hibecovirus*, established based on phylogenetic analysis of conserved protein domains. The first four of these subgenera were formerly known as lineages or subgroups A, B, C and D, respectively.

Subgenus *Embecovirus* includes two human coronaviruses (HKU1 and OC43), as well as several animal coronaviruses of veterinary relevance such as bovine, canine, equine, porcine and murine coronaviruses. *Sarbecovirus* comprises the severe acute respiratory syndrome (SARS)-related coronaviruses such as SARS-CoV and SARS-CoV-2 (2019-nCoV), respectively responsible for the 2002–2003 SARS outbreak and the 2019–2020 coronavirus disease 2019 (COVID-19) pandemic. Several SARS-related bat coronaviruses, mainly isolated from Chinese horseshoe bats (*Rhinolophus* spp.), also belong to this subgenus. Subgenus *Merbecovirus* comprises the Middle East respiratory syndrome (MERS)-related coronaviruses, including the MERS-CoV responsible for the 2012 MERS outbreak, as well as two additional species of bat coronaviruses isolated from *Tylonycteris* and *Pipistrellus* bats. Subgenera *Nobecovirus* and *Hibecovirus* comprise only bat coronaviruses, mainly isolated from *Rousettus* and *Hipposideros* bats, respectively.

Since the 2002–2003 SARS outbreak, genomic information has become ever-increasingly significant to address outbreaks caused by pathogenic coronaviruses. Before the 2019–2020 COVID-19 pandemic, there were ~1200 complete genomes of betacoronaviruses deposited in the GenBank database. The number of available genomes has increased dramatically during the pandemic, with more than 6000 complete genomes available in Genbank as of June 2019, and almost 50,000 genomic sequences in other public repositories. A variety of information including phylogenetic relationships, mode of transmission, evolutionary rates and the role of mutations in infection and disease severity can be deduced from comparing multiple genomes.

2. Evolutionary Rates and Divergence

Estimation of evolutionary rates is an important step to characterize the genetic diversity among viral lineages and to place a timescale in phylogenetic hypotheses explaining their origin and divergence. The rate of evolution of viruses is often assessed through the number of errors occurring during replication of the viral genome (the mutation rate) and the frequency at which such mutations become fixed in the population (the substitution rate) ^[1]. The substitution rate depends on several factors, including the underlying mutation rate and the presence of selective forces that influence fixation of mutations in association with their fitness. Mutation rates of RNA viruses are generally higher than those of DNA viruses, due to the lack of a proofreading activity and consequent low fidelity of their RdRp ^[2]. However, due to the proofreading activity of Nsp14, members of the order *Nidovirales* have relatively lower mutation rates ^[3].

The substitution rate is often expressed in substitutions per nucleotide site per year (s/n/y) and can be estimated from phylogenetic reconstructions when divergence time is known for particular lineages. Although several methods have been traditionally used to estimate substitution rates, including linear regression and maximum likelihood (ML), the most popular method nowadays is the Bayesian Markov chain Monte Carlo (MCMC) approach, such as that implemented in the BEAST package ^[4]. Globally, substitution rates of coronaviruses have been estimated to be in the order of 10^{-3} – 10^{-4} s/n/y ^{[5][6]}. Studies conducted in SARS-CoV and MERS-CoV have estimated whole genome substitution rates to be between 0.80 – 2.38×10^{-3} and 0.88 – 1.37×10^{-3} s/n/y, respectively ^{[7][8]}. However, variation in the estimates for particular genes have been observed for both SARS-CoV ^{[9][10][11][12][13]} and SARS-CoV-2 ^[14,15], suggesting that the S gene and the region between ORF7b and ORF8 may be subjected to positive selective pressure in some lineages. In contrast, more conserved regions of the genome such as ORF1a and ORF1b appear to be under strong negative or purifying selection.

An important application of this type of analysis is the estimation of the time to most recent common ancestor (TMRCA) between two lineages, as an approximate measure of their time since divergence. Several studies have estimated that the SARS-CoV lineage within the SARS-related coronaviruses most probably emerged between 1961–1985, while the civet SARS-CoV strains may have originated around 1986–1995 ^{[11][12][13]}. TMRCA estimates for the SARS-CoV and MERS-CoV strains involved in previous outbreaks have been roughly consistent with the dates when the first cases were reported ^{[8][13]}. A preliminary study has estimated that the group containing SARS-CoV-2 and its closest bat coronavirus, RaTG13, may have diverged between 40–70 years ago ^[16].

3. Recombination, RBD Mutations and Host/Tissue Tropism

Recombination events are often inferred by comparing phylogenetic trees built from different genes or genomic regions, since occurrence of recombination often leads to inconsistent topologies in such trees. One of the most popular methods for detecting recombination in viral genomes is bootscan analysis ^[17]. To use this method, sequences of target genomes are aligned against reference sequences thought to be involved in the recombination events. The alignments are divided into short sequential segments and phylogenetic trees are then built from these segments. Recombination is suspected in segments for which the trees exhibit an alternative topology, involving different reference sequences. Bootscan and other complementary methods, such as sequence similarity plots for the putative recombinant regions, are implemented in packages such as SimPlot ^[18] or the Recombination Detection Program (RDP) ^[19]. Studies relying on these methods have documented the occurrence of recombination in several coronavirus genera and have also provided ample evidence supporting the important role of this process in coronavirus cross-species transmission ^{[20][21]}.

The first reported example of natural recombination in human coronaviruses was that occurring between two different HKU1 genotypes ^[22]. Putative recombination events between genotypes have also been documented for human coronaviruses NL63 ^[23] and OC43 ^[24]. In the case of SARS-related and MERS-related coronaviruses, recombination appears to occur among strains infecting several animal hosts, including bats, intermediary hosts and humans (reviewed by Su et al. ^[24]). Studies considering several bat SARS-related coronaviruses have suggested the occurrence of recombination in lineages leading to human and/or civet strains of SARS-CoV, with breakpoints often located close or within the S and ORF8 genes ^{[12][25][26]}. These findings suggest that recombination between existing strains can result in new strains or species, with possible differences in host and tissue tropism. For instance, bat coronavirus strains Rs3367 (WIV1) and WIV16 have been reported to have high sequence similarity to human/civet SARS-CoV at the S gene, allowing them to use ACE2 as a receptor for cell entry ^{[27][28]}. Recombination analysis suggested that at least one civet SARS-CoV strain (SZ3) may have originated by recombination between WIV16 and another bat SARS-CoV strain (Rf4092) ^[29].

In the case of SARS-CoV-2, comparison of its genome to those of other SARS-related coronaviruses did not provide enough evidence supporting recent recombination as a possible explanation for its origin ^[30]. However, two putative breakpoints, possibly derived from a past recombination event, were identified within the S gene and flanking its receptor-

binding domain (RBD) [30]. Globally, the SARS-CoV-2 genome is more similar to those of bat coronaviruses SL-CoVZXC21 and SL-CoVZC45 however, the region between the two breakpoints was found to be more similar to human/civet SARS-CoV and WIV1. Putative recombination signals have also been reported between SARS-CoV-2, RaTG13 and the pangolin-related coronaviruses [31]. Although SARS-CoV-2 is more similar to RaTG13 than to the pangolin coronaviruses, some of the latter have higher sequence similarity to SARS-CoV-2 in the RBD. It has also been suggested that these similarities at the amino acid level may be due to convergent evolution, arising from positive selection instead of recombination [32].

The fact that different evolutionary events often involve the RBD is likely to be associated with the role of this domain and its RBM in receptor recognition and adaptation to different animal hosts. In SARS-CoV-like viruses, six RBD amino acids have been found to be essential for binding to ACE2, five of which differ between SARS-CoV and SARS-CoV-2 [33]. Particular sets of RBD mutations appear to be associated with a specific host range for each coronavirus species, as is the case of humans and civets in SARS-CoV or humans and camels in MERS-CoV. Although it has been suggested that SARS-CoV-2 is optimized for binding to human ACE2, it may also infect other animals with highly similar ACE2 homologs such as pigs, ferrets, cats and primates [33]. In fact, the S gene of a SARS-CoV-2 strain recently isolated from a tiger (GenBank accession number MT365033) is identical to those of human isolates, both clustering into the same branch in phylogenetic trees.

Comparative sequence analysis has suggested that positive selection may have a role in shaping the evolution of the S protein and the RBD of SARS-CoV and MERS-CoV [34][35][36]. The role of natural selection in the evolution of particular genes is typically inferred by computing the ratio of the rates of nonsynonymous to synonymous changes (K_a/K_s) between groups or lineages, with a value greater than 1 indicating an overall positive selective pressure. Although evolution of the SARS-CoV genome during the 2002–2003 SARS outbreak was found to be largely neutral or nearly neutral, at least six mutations occurred in the S protein during the early, middle and late phases of the outbreak, all of which were present in the epidemic strain (Urbani) [34][37]. However, the average K_a/K_s values for the early phase were found to be significantly higher than those for the middle and late phases, suggestive of initial positive selection in the S gene, followed by purifying selection and stabilization. Likewise, a recent study has suggested limited episodes of positive selection during divergence of SARS-CoV-2 from RaTG13, although there is still insufficient evidence to associate these changes with its adaptation to humans [38].

4. Genetic Variation and Transmission in Human Populations

As pathogenic viruses replicate and spread during outbreaks, their genomes accumulate random mutations that can be used to track the spread of the disease, reconstruct their transmission routes and detect lineages with different levels of virulence and transmissibility. There are several methods for tracking mutations and inferring the mode of transmission from genomic data, most of which require the alignment of sequences from new isolates to reference genomes and the subsequent identification of genetic variants such as single nucleotide variants (SNV) and insertion/deletions (indels). The simplest and fastest methods are based on pairwise distances among samples computed from these variants. However, these methods do not consider evolutionary models and can be highly inaccurate when there is substantial divergence between donor and recipients in transmission chains [39]. More advanced methods are based on ML or MCMC approaches, often within a Bayesian framework, applying an explicit model of evolution to phylogenetic estimation. When these methods are combined with sampling dates, estimations of the presence of significant molecular evolution over a sampling period is possible [39]. Such analyses are implemented in packages like TransPhylo [40][41], PhyloScanner [42], Outbreaker2 [43] or Phybreak [41].

The main output of these methods is a transmission tree indicating which individuals infected others. Although a transmission tree cannot be directly inferred from a phylogenetic tree, it must be consistent with the underlying phylogeny. Together, phylogenetic and transmission trees help us trace back the origin of an outbreak, detect multiple introductions of a pathogen into a given territory, identify mutations that define specific lineages and predict the potential existence of unsampled individuals that may have acted as missing transmission links. Due to their relatively recent development, transmission trees based on genomic data were not widely used to study the transmission routes of previous SARS and MERS outbreaks; however, a recent study analyzing data from the 2003 SARS outbreak has provided new insights into its early stages [44].

Progress in genome sequencing technologies has resulted in an exceptionally high number of SARS-CoV-2 genomes sequenced during the 2019–2020 COVID-19 pandemic. In fact, the COVID-19 pandemic is the second one in history, after the 2009 H1N1 influenza pandemic [45][46], for which genomic data have been generated almost in a real-time fashion, allowing a detailed reconstruction of transmission trees. Genome sequences have been made publicly accessible through several repositories, including a data sharing service hosted by the Global Initiative on Sharing All Influenza Data

(GISAID) (<https://www.gisaid.org/>) [47][48]. In addition to public genome repositories, open-source platforms for real-time data visualization and analysis of genomic data are also available, including NextStrain (<https://nextstrain.org>) and CoV-GLUE (<http://cov-glue.cvr.gla.ac.uk>). NextStrain is fed with sequences from the GISAID repository and uses the Augur bioinformatics toolkit (<https://github.com/nextstrain/augur>) for tracking molecular evolution and the Auspice software (<https://nextstrain.github.io/auspice/>) for interactive visualization of phylogenomic data. Conversely, CoV-GLUE is a web application based on an integrated software environment called GLUE (Genes Linked by Underlying Evolution), designed to create bioinformatic resources based on viral genome sequences [49]. CoV-GLUE is also based on GISAID data and contains a database of replacements and indels that have been found in previously sampled SARS-CoV-2 sequences. New SARS-CoV-2 genome sequences can be loaded into the platform to identify novel or known mutations, assign them to potential lineages and visualize them in a phylogenetic context.

Several attempts have been made to classify circulating strains of SARS-CoV-2 into lineages or genotypes with potential differences in transmissibility and disease severity. Among these, a study comparing 103 SARS-CoV-2 genomes suggested the existence of two lineages, termed L and S, with potential differences in prevalence [32]. Another recent study compared 160 genomes from different countries and suggested the existence of three subtypes, A, B and C, with differences in geographic distribution and prevalence [50]. However, such studies have been criticized for possible sampling biases and misinterpretation of results [51][52][53][54], stressing that caution should be taken when drawing conclusions from genomic analyses. Limited or inappropriate sampling can bias the inference of transmission networks, potentially hiding introduction events and intermediate states and resulting in inaccurate mutation rate estimates [39]. When describing new lineages based on SNVs and other genetic variants, fixation of the corresponding mutations should be first demonstrated in local populations.

5. Concluding Remarks

Since the 2002–2003 SARS outbreak, genomic information has been crucial to tackle epidemics caused by betacoronaviruses. During the 2019–2020 COVID-19 pandemic, quick availability of genomic data has allowed a very rapid, detailed and accurate follow-up of disease progression worldwide and has tremendously supported the development of diagnostic systems, drug candidates and vaccines. Full viral genome analysis has swiftly changed the way scientists deal with epidemic viruses in two main ways: first, the speed that allows the description and classification of the responsible pathogen in a record timeframe, and second, the ability to generate massive amounts of viral genome data, contributing to establishing sound hypotheses on evolution and transmission. It is remarkable that genome analyses at such a scale are now increasingly feasible, without having to culture the viruses, many of which are classified as Biosafety Level 3 agents. However, it is important to stress that genomic information must be used carefully when drawing conclusions related to human and animal health. Sampling bias, selection of inadequate bioinformatic tools and misinterpretation of results can all lead to unreliable conclusions. Furthermore, the quality of genomic sequences used in comparative analyses is also crucial to establish sound conclusions. Currently, there are several platforms used for genome sequencing, each of them with their own patterns of systematic sequencing errors. Data curation and normalization is extremely important before conducting further analyses, particularly when comparing data from different sequencing platforms. If analyzed properly, genomic data will indisputably serve as strong basis for addressing future outbreaks caused by highly pathogenic emerging viruses such as SARS-CoV-2.

References

1. Duffy, S.; Shackelton, L.A.; Holmes, E.C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* 2008, 9, 267–276.
2. Sanjuán, R.; Nebot, M.R.; Chirico, N.; Mansky, L.M.; Belshaw, R. Viral Mutation Rates. *J. Virol.* 2010, 84, 9733–9748.
3. Gorbalenya, A.E.; Enjuanes, L.; Ziebuhr, J.; Snijder, E.J. Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* 2006, 117, 17–37.
4. 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, 1–5.
5. Sánchez, C.M.; Gebauer, F.; Suñé, C.; Mendez, A.; Dopazo, J.; Enjuanes, L. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* 1992, 190, 92–105.
6. Vigen, L.; Keyaerts, E.; Moës, 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.

7. Zhao, Z.; Li, H.; Wu, X.; Zhong, Y.; Zhang, K.; Zhang, Y.P.; Boerwinkle, E.; Fu, Y.X. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* 2004, 4, 21.
8. Cotten, M.; Watson, S.J.; Zumla, A.I.; Makhdoom, H.Q.; Palser, A.L.; Ong, S.H.; Al Rabeeah, A.A.; Alhakeem, R.F.; Assiri, A.; Al-Tawfiq, J.A.; et al. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *MBio* 2014, 5, e01062-13.
9. Song, H.D.; Zhang, G.W.; Cheng, F.; Pan, C.M.; Chen, S.J.; Tu, C.C.; Lei, L.C.; Gao, Y.W.; Xiang, H.; Xuan, H.; et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. USA* 2005, 102, 2430–2435.
10. Salemi, M.; Fitch, W.M.; Ciccozzi, M.; Ruiz-Alvarez, M.J.; Rezza, G.; Lewis, M.J. Severe Acute Respiratory Syndrome Coronavirus Sequence Characteristics and Evolutionary Rate Estimate from Maximum Likelihood Analysis. *J. Virol.* 2004, 78, 1602–1603.
11. Vijaykrishna, D.; Smith, G.J.D.; Zhang, J.X.; Peiris, J.S.M.; Chen, H.; Guan, Y. Evolutionary Insights into the Ecology of Coronaviruses. *J. Virol.* 2007, 81, 4012–4020.
12. Hon, C.-C.; Lam, T.-Y.; Shi, Z.-L.; Drummond, A.J.; Yip, C.-W.; Zeng, F.; Lam, P.-Y.; Leung, F.C.-C. Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus. *J. Virol.* 2008, 82, 1819–1826.
13. Lau, S.K.P.; Li, K.S.M.; Huang, Y.; Shek, C.-T.; Tse, H.; Wang, M.; Choi, G.K.Y.; Xu, H.; Lam, C.S.F.; Guo, R.; et al. Ecoepidemiology and Complete Genome Comparison of Different Strains of Severe Acute Respiratory Syndrome-Related Rhinolophus Bat Coronavirus in China Reveal Bats as a Reservoir for Acute, Self-Limiting Infection That Allows Recombination Events. *J. Virol.* 2010, 84, 2808–2819.
14. Wang, C.; Liu, Z.; Chen, Z.; Huang, X.; Xu, M.; He, T.; Zhang, Z. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* 2020, 92, 667–674.
15. Pachetti, M.; Marini, B.; Benedetti, F.; Giudici, F.; Mauro, E.; Storici, P.; Masciovecchio, C.; Angeletti, S.; Ciccozzi, M.; Gallo, R.C.; et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* 2020, 18, 179.
16. Boni, M.F.; Lemey, P.; Jiang, X.; Lam, T.T.-Y.; Perry, B.; Castoe, T.; Rambaut, A.; Robertson, D.L. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *bioRxiv* 2020.
17. Salminen, M.O.; Carr, J.K.; Burke, D.S.; McCutchan, F.E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retrovir.* 1995, 11, 1423–1425.
18. Lole, K.S.; Bollinger, R.C.; Paranjape, R.S.; Gadkari, D.; Kulkarni, S.S.; Novak, N.G.; Ingersoll, R.; Sheppard, H.W.; Ray, S.C. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 1999, 73, 152–160.
19. Martin, D.P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefevre, P. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* 2010, 26, 2462–2463.
20. Graham, R.L.; Baric, R.S. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *J. Virol.* 2010, 84, 3134–3146.
21. Su, S.; Wong, G.; Shi, W.; Liu, J.; Lai, A.C.K.; Zhou, J.; Liu, W.; Bi, Y.; Gao, G.F. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016, 24, 490–502.
22. Woo, P.C.Y.; Lau, S.K.P.; Yip, C.C.Y.; Huang, Y.; Tsoi, H.-W.; Chan, K.-H.; Yuen, K.-Y. Comparative Analysis of 22 Coronavirus HKU1 Genomes Reveals a Novel Genotype and Evidence of Natural Recombination in Coronavirus HKU1. *J. Virol.* 2006, 80, 7136–7145.
23. Pyrc, K.; Dijkman, R.; Deng, L.; Jebbink, M.F.; Ross, H.A.; Berkhout, B.; van der Hoek, L. Mosaic structure of human coronavirus NL63, one thousand years of evolution. *J. Mol. Biol.* 2006, 364, 964–973.
24. Lau, S.K.P.; Lee, P.; Tsang, A.K.L.; Yip, C.C.Y.; Tse, H.; Lee, R.A.; So, L.-Y.; Lau, Y.-L.; Chan, K.-H.; Woo, P.C.Y.; et al. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J. Virol.* 2011, 85, 11325–11337.
25. Lau, S.K.P.; Feng, Y.; Chen, H.; Luk, H.K.H.; Yang, W.-H.; Li, K.S.M.; Zhang, Y.-Z.; Huang, Y.; Song, Z.-Z.; Chow, W.-N.; et al. Severe Acute Respiratory Syndrome (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from Greater Horseshoe Bats through Recombination. *J. Virol.* 2015, 89, 10532–10547.
26. Wu, Z.; Yang, L.; Ren, X.; Zhang, J.; Yang, F.; Zhang, S.; Jin, Q. ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* 2016, 213, 579–583.

27. 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. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 2013, 503, 535–538.
28. Yang, X.-L.; Hu, B.; Wang, B.; Wang, M.-N.; Zhang, Q.; Zhang, W.; Wu, L.-J.; Ge, X.-Y.; Zhang, Y.-Z.; Daszak, P.; et al. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* 2015, 90, 3253–3256.
29. Hu, B.; Zeng, L.P.; Yang, X.L.; Ge, X.Y.; Zhang, W.; Li, B.; Xie, J.Z.; Shen, X.R.; Zhang, Y.Z.; Wang, N.; et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* 2017, 13, 1–27.
30. 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. A new coronavirus associated with human respiratory disease in China. *Nature* 2020, 579, 265–269.
31. Lam, T.T.Y.; Shum, M.H.H.; Zhu, H.C.; Tong, Y.G.; Ni, X.B.; Liao, Y.S.; Wei, W.; Cheung, W.Y.M.; Li, W.J.; Li, L.F.; et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* 2020, in press.
32. Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; et al. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* 2020, 2020, nwaa036.
33. Wan, Y.; Shang, J.; Graham, R.; Baric, R.S.; Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* 2020, 94, 1–9.
34. He, J.F.; Peng, G.W.; Min, J.; Yu, D.W.; Liang, W.J.; Zhang, S.Y.; Xu, R.H.; Zheng, H.Y.; Wu, X.W.; Xu, J.; et al. Molecular Evolution of the SARS Coronavirus, during the Course of the SARS Epidemic in China. *Science* (80) 2004, 303, 1666–1669.
35. Zhang, C.Y.; Wei, J.F.; He, S.N. Adaptive evolution of the spike gene of SARS coronavirus: Changes in positively selected sites in different epidemic groups. *BMC Microbiol.* 2006, 6, 88.
36. Zhang, Z.; Shen, L.; Gu, X. Evolutionary Dynamics of MERS-CoV: Potential Recombination, Positive Selection and Transmission. *Sci. Rep.* 2016, 6, 25049.
37. Bolles, M.; Donaldson, E.; Baric, R. SARS-CoV and emergent coronaviruses: Viral determinants of interspecies transmission. *Curr. Opin. Virol.* 2011, 1, 624–634.
38. Cagliani, R.; Forni, D.; Clerici, M.; Sironi, M. Computational inference of selection underlying the evolution of the novel coronavirus, SARS-CoV-2. *J. Virol.* 2020, in press.
39. Lam, T.T.Y.; Hon, C.C.; Tang, J.W. Use of Phylogenetics in the Molecular Epidemiology and Evolutionary Studies of Viral Infections. *Crit. Rev. Clin. Lab. Sci.* 2010, 47, 5–49.
40. Didelot, X.; Gardy, J.; Colijn, C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol. Biol. Evol.* 2014, 31, 1869–1879.
41. Klinkenberg, D.; Backer, J.A.; Didelot, X.; Colijn, C.; Wallinga, J. Simultaneous Inference of Phylogenetic and Transmission Trees in Infectious Disease Outbreaks. *PLoS Comput. Biol.* 2017, 13, e1005495.
42. Wymant, C.; Hall, M.; Ratmann, O.; Bonsall, D.; Golubchik, T.; De Cesare, M.; Gall, A.; Cornelissen, M.; Fraser, C. PhyloScanner: Inferring transmission from within- and between-host pathogen genetic diversity. *Mol. Biol. Evol.* 2018, 35, 719–733.
43. Campbell, F.; Didelot, X.; Fitzjohn, R.; Ferguson, N.; Cori, A.; Jombart, T. Outbreaker2: A modular platform for outbreak reconstruction. *BMC Bioinform.* 2018, 19, 363.
44. Jombart, T.; Cori, A.; Didelot, X.; Cauchemez, S.; Fraser, C.; Ferguson, N. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput. Biol.* 2014, 10, e1003457.
45. Fraser, C.; Donnelly, C.A.; Cauchemez, S.; Hanage, W.P.; Van Kerkhove, M.D.; Hollingsworth, T.D.; Griffin, J.; Baggaley, R.F.; Jenkins, H.E.; Lyons, E.J.; et al. Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* (80) 2009, 324, 1557–1561.
46. Smith, G.J.D.; Vijaykrishna, D.; Bahl, J.; Lycett, S.J.; Worobey, M.; Pybus, O.G.; Ma, S.K.; Cheung, C.L.; Raghvani, J.; Bhatt, S.; et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza epidemic. *Nature* 2009, 459, 1122–1125.
47. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* 2017, 1, 33–46.
48. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* 2017, 22, 2–4.
49. Singer, J.B.; Thomson, E.C.; McLauchlan, J.; Hughes, J.; Gifford, R.J. GLUE: A flexible software system for virus sequence data. *BMC Bioinform.* 2018, 19, 1–18.

50. Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. USA* 2020, 117, 9241–9243.
51. Ruffell, D. Coronavirus SARS-CoV-2: Filtering fact from fiction in the infodemic: Q&A with virologist Professor Urs Greber. *FEBS Lett.* 2020, 594, 1127–1131.
52. Sánchez-Pacheco, S.J.; Kong, S.; Pulido-Santacruz, P.; Murphy, R.W.; Kubatko, L. Median-joining network analysis of SARS-CoV-2 genomes is neither phylogenetic nor evolutionary. *Proc. Natl. Acad. Sci. USA* 2020, 117, 12518–12519.
53. Mavian, C.; Pond, S.K.; Marini, S.; Magalis, B.R.; Vandamme, A.-M.; Dellicour, S.; Scarpino, S.V.; Houldcroft, C.; Villabona-Arenas, J.; Paisie, T.K.; et al. Sampling bias and incorrect rooting make phylogenetic network tracing of SARS-COV-2 infections unreliable. *Proc. Natl. Acad. Sci. USA* 2020, 117, 12522–12523.
54. Chookajorn, T. Evolving COVID-19 conundrum and its impact. *Proc. Natl. Acad. Sci. USA* 2020, 117, 12520.

Retrieved from <https://encyclopedia.pub/entry/history/show/6974>