Open Peer Review on Qeios

Sequence evidence that the D614G clade of SARS-CoV-2 was already circulating in northern Italy in the fall of 2019

Xuhua Xia[1]

1 University of Ottawa

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

The D614G clade is characterized by TTTG at four nucleotide sites (sites 241, 3037, 14408 and 23403 following the reference genome NC_045512), in contrast to CCCA shared among early SARS-CoV-2 genomes sampled in China and those that can be traced to China. It was believed that the TTTG lineage descended from the early viral CCCA lineages. A set of SARS-CoV-2 sequences collected from Sept. 12 to Dec. 18, 2019, in Lombardy, Milan and Turin in Italy provided, for the first time, strong evidence that the D614G/TTTG lineage has already been circulating in Italy in 2019. I discussed extensively the controversies arising from this set of early SARS-CoV-2 sequences.

Xuhua Xia[1,2]

[1] *Department of Biology, University of Ottawa, Ottawa, Canada K1N 6N5,xxia@uottawa.ca*

[2] *Ottawa Institute of Systems Biology, Ottawa, Canada K1H 8M5*

Keywords: SARS-CoV-2; COVID-19; viral evolution; viral origin; viral dating; D614G; Lombardy; Italy.

Two key questions are asked following a viral outbreak, i.e., when and where the zoonotic or lab-leak event occurred. The "When" question is typically addressed by estimating the viral evolutionary rate and dating the most recent common ancestor (MRCA) of representative viral strains [1][2][3][4][5][6]. The “Where” question is approximated by the location where the earliest viral lineage was sampled [5][6][7]. Obviously, such large-scale analyses are better performed with full-length high-quality viral genomes than short sequence fragments. For SARS-CoV-2, there are currently more than 13 million genomic sequences. Thus, there seems no shortage of data for any large-scale data analysis. Indeed, the MRCA of SARS-CoV-2 genomes have been dated to the summer of 2019 with a gigantic tree of 455,251 leaves [6].

One major shortcoming of using full-length high-quality SARS-CoV-2 genomes is that these genomes are sampled after the viral outbreak in Wuhan in late December 2019, when SARS-CoV-2 has already spread globally. SARS-CoV-2 samples collected before the viral outbreak are present in GenBank only as short sequence fragments. However, these sequence fragments provide critical insights into the origin and early evolution of the D614G lineage.

The D614G clade differs from the early Wuhan strains at four nucleotide sites (sites 241, 3037, 14408 and 23403 following the reference genome NC_045512) [8]. The D614G strain has TTTG at these four sites, in contrast to CCCA shared among early SARS-CoV-2 genomes sampled in Wuhan. It was believed that the D614G/TTTG lineage descends from the CCCA lineage. However, sequence data from early SARS-CoV-2 samples collected in northern Italy (Lombardy, Milan and Turin regions) from Sept. 12 to Dec. 18, 2019, provided strong evidence that the D614G/TTTG lineage was already circulating in northern Italy in the fall of 2019.

Thirteen of these Italy-derived samples of SARS-CoV-2 were partially sequenced and deposited in GenBank [9][10] (Table 1). These sequence fragments vary in length from 209 nt to 778 nt and do not cover the entire SARS-CoV-2 genome. However, they do cover three of the four critical sites distinguishing the D614G/TTTG lineage from the early CCCA lineage (Table 1). Site 3037 is represented by two sequences with a T, site 14408 by four sequences with a T, and site 23403 by one sequence with a G. This “?TTG” configuration is compatible with that of D614G/TTTG lineage but not with the early CCCA lineage represented by the reference sequence (NC_045512). The inescapable conclusion is that the D614G/TTTG lineage has been circulating in northern Italy in 2019, concurrently with or even earlier than the CCCA lineages represented in early SARS-CoV-2 samples in Wuhan.

Three longest fragments (427, 409 and 778 nt, respectively) match perfectly a D614G/TTTG genome (ACCN: OP777407) sampled in Scotland in March 2020 (as well as other D614G/TTTG genomes sampled in Scotland in April and May 2020). There should be little doubt that the sequence fragments listed in Table 1 belong to D614/TTTG strains, although one may still raise the issue of contamination which I will address later.

The last six sequence fragments in Table 1 cannot be individually allocated to either the D614G/TTTG lineage or the CCCA lineage. They are from highly conserved sections in SARS-CoV-2 genomes and matches perfectly to both early D614G/TTTG and CCCA lineages. Similarly, the four SARS-CoV-2 sequence fragments sampled from human sewage in Brazil (MT925972, MT925973, MT925976, MT925977) [11] also match perfectly to early D614G/TTTG and CCCA lineages.

The sequence evidence that the early SARS-CoV-2 sequences from northern Italy belong to the D614G/TTTG lineage is consistent with the dating of the MRCA to the summer of 2019 [5][6]. It is also consistent with the twin-beginnings hypothesis on the origin, spread and evolution of SARS-CoV-2 [12].

How would one know if these early Italian D614G/TTTG sequences did not result from contamination? After all, the first seven viral sequences (Table 1) were submitted to GenBank on May 18, 2021 when D614G/TTTG strains have been around for more than a year. The hypothesis of contamination was advocated in particular by those who think it unlikely to have D614G/TTTG sequences in Italy in late 2019. I will not repeat the cautions against contamination taken by researchers who sequenced those early samples [9]. Instead, I will provide various lines of evidence suggesting that D614G/TTTG strains may well be circulating in Italy much earlier than most of us have thought.

The first two cases of COVID-19 in Italy were confirmed on Jan. 30, 2020. Among 16 fully sequenced SARS-CoV-2 genomes collected in Italy in February 2020, eight are CCCA strains (OP288481, OP288482, OP288483, OP288484,

MT509652, MT509660, MT509661, MT509653) which appeared to cause more serious symptoms and could almost all be traced back to China. The other eight are D614G/TTTG strains (MT509664, MT509667, MW530510, MW530511, MT479216, MT483875, MT483881, MT527178), which appeared to cause only mild symptoms. One D614G/TTTG strain (MT527178) was also circulating in Lazio (including Rome) in February 2020. Thus, D614G/TTTG strains were already frequent in Italy in February 2020.

Around Feb. 18, a 51 year old man from Scotland came to visit Rome via Lombardy, by his private rental car. He subsequently became the first COVID-19 patient in Scotland, confirmed on Mar. 1, 2020 [13][14]. Given the average incubation period of about 12 days, he likely was infected in Lombardy on his way to Rome. The patient exhibited only mild symptoms throughout admission with no supportive care provided [14]. The SARS-CoV-2 genome from the patient, belonging to the D614G/TTTG lineage, was named Scotland/CVR01/2020 (GISAID: EPI_ISL_413221). Thus, while it remains unclear how D614G/TTTG strains moved into Italy, these viral strains might have been circulating cryptically in Italy for some time, only to be revealed after the import of the more virulent CCCA strains into Italy in late January, 2020. Simultaneous import of both the CCCA strains and the D614G/TTTG strains, although not impossible, should be a low-probability event. Even if such an event did happen, it would still imply that that D614G/TTTG strains have been circulating elsewhere before they were transmitted into Italy. Fig. 1 included a number of D614G/TTTG and TTCG strains reported elsewhere in January and February 2020. Putting all these together, there is really nothing extraordinary to find D614G/TTTG strains in Italy (or anywhere else) in late 2019. The sequence fragments in Table 1 provided more direct confirmation of early existence of D614G/TTTG strains than inferences based on collection dates of genomic records.

These sequence fragments have been overshadowed by the huge number of full-length high-quality SARS-CoV-2 genomes and forgotten for too long [9]. In spite of their high relevance, these sequence fragments (Table 1) have not been used in any previous studies addressing the “When” and “Where” questions, including the recent study by Pekar et al.[15]. Not only did Pekar et al.[15] not analyze these early SARS-CoV-2 sequence fragments, but they also omitted full-length high-quality SARS-CoV-2 genomes from the D614G clade sampled in January 2020 in both GenBank and GISAID (Table 2). While there is no guarantee that collection dates found in GenBank and GISAID records are accurate, it is inappropriate to omit crucially relevant data because such omission will almost certainly bias the estimation of the root of the tree. If one includes early samples in Location A but omits early samples from all other locations, then Location A will necessarily become associated with the earliest SARS-CoV-2 lineage. Similarly, if one included early samples of lineage X but excluded early samples of lineage Y, then lineage X would be more likely associated with a statistical ancestor than lineage Y.

Fig. 1 is a phylogenetic tree including early SARS-CoV-2 genomes from Europe and USA, together with representatives of lineage A and lineage B samples from China. There were two early but now obsolete reasons for placing the root within the shaded clade (Fig. 1). First, the shaded clade represents the earliest SARS-CoV-2 samples from Wuhan and those that can be traced to China. Second, Huanan Seafood Market is likely where the zoonotic transmission occurred. Associated with these two reasons is the assumption that SARS-CoV-2 was sampled and sequenced right after its presumed zoonotic transmission [16]. The specific placement of the root at the red dot or within the red clade was tenuously based on two nucleotides at sites 8782 and 28144[16][17]. Lineage A (represented by the red

clade in Fig. 1) has T and C at these two sites, and the close relatives of bat-derived viruses such as RaTG13 also have T and C at these two sites. This TC configuration is therefore assumed to be ancestral. In contrast, the CT configuration at these two sites observed in lineage B, which was sampled earlier than lineage A and originally represented by the non-red lineages within the shaded clade (Fig. 1), was deemed derived. Pekar et al.[15] did briefly mention an alternative hypothesis that the TC configuration may not be ancestral because of the existence of SARS-CoV-2 with TT and CC configurations at these two sites indicating an evolutionary trajectory of CT (lineage B) (TT or CC) TC (lineage A).

All these interpretations fall apart given the evidence of early circulation of the D614G strain in Table 1. The root of SARS-CoV-2 is more likely at the black dot than at the red dot (Fig. 1). This new root immediately answers two key questions raised by Pekar et al.[15] at beginning of their paper: (i) why were lineage B viruses detected earlier than lineage A viruses and (ii) why did lineage B predominate early in the pandemic? The answers to these two questions are obvious given the root at the black dot, but become difficult if the root is at the red dot as assumed by Pekar et al.[15].

There are a number of unresolved problems concerning the origin of the CCCA lineage and the D614G/TTTG lineage. SARS-CoV-2-like coronaviruses isolated from bats and pangolins in Southeast Asia also feature the CCCA signature. What remains controversial is what happened between an ancestral CCCA lineage in the natural coronavirus reservoir and the viral strain causing the Wuhan outbreak. The lab-leak hypothesis states that the ancestral CCCA lineage first entered a lab and then infected human. Two major sub-hypotheses exist, one claiming that the lab-leak occurred in a Chinese laboratory and the other claiming that the lab-leak occurred in a US laboratory. The natural origin hypothesis states that the ancestral CCCA lineage infected human through a zoonotic transmission. The origin of the D614G/TTTG lineage is even less clear. If it is derived from an CCCA lineage, then that CCCA lineage would have to be circulating much earlier than the Wuhan outbreak. It is even possible, although not highly probable, that the coronavirus reservoir in nature has two separate lineages (i.e., CCCA and D614G/TTTG lineages) that were transmitted to human independently. At present, the gap is wide between the possible and the actual.

Sequence data currently available is inadequate for one to infer when and where the D614G/TTTG strain originated. Although D614G with the TTTG configuration was not found in China except for imported cases, viral genomes with TTCG configuration were later found in China through retrospective sequencing[18]. What is clear is that viral lineages with the TTTG/TTCG configuration coexisted with the CCCA strain at the time when the latter caused the viral outbreak in Wuhan. Both the TTTG clade and the CCCA clade are likely descendants of a much earlier common ancestor. If scientists around the world would respond to WHO's call to sequence archived samples, then there is a chance to find when and where SARS-CoV-2 originated, either through a natural zoonotic event or a lab-leak event. This task is obviously equivalent to searching a needle in a haystack, given the necessarily low rate of SARS-CoV-2 circulation in the pre-pandemic period. However, there is no alternative for one to get a more precise answer to the “When” and “Where” question. Limiting the search for a common ancestor within the early Wuhan clade is reminiscent of a mink government insisting that the common ancestor of SARS-CoV-2 is within the mink farm in the Netherlands where the first SARS-CoV-2 outbreak was recorded [19].

Acknowledgments

I thank B. Foley, A. Rambaut, and J. Wertheim for comments, and A. Kovarik and F. Broecker for reviews of a pre-print posted to qeios.com. This study is supported by Natural Science and Research Council (NSERC) of Canada Discovery Grant (RGPIN/2018-03878).

Figures and Tables

Sequence ID	Start(1)	End(1)	3037	14408	23403
MZ223388_2019-10-22_Lombardy	2977	3185	T		
MZ223391_2019-11-22_Lombardy	2977	3185	T		
MZ223389_2019-10-17_Lombardy	14366	14655		T	
MZ223387_2019-10-19_Lombardy	14366	14655		T	
MZ223386_2019-10-23_Lombardy	14366	14655		T	
MZ223392_2019-12-15_Lombardy	14366	14655		T	
MZ223393_2019-12-15_Lombardy	22904	23681			G
MW303957_2019-12-05_Milan	22935	23343			
MZ223385_2019-09-12_Lombardy	22904	23330			
MZ223390_2019-10-12_Lombardy	22904	23330			
MT843234_2019-12-18_Milan	18484	18770			
MT843235_2019-12-18_Turin	18484	18770			
MT843236_2019-12-18_Turin	18484	18770			

(1) *Start and ending site of the sequence with site numbering according to the reference genome (NC_045512)*

Table 2. Early SARS-CoV-2 samples in the D614G lineage with collection date (Coll. Date), country, and Type (nucleotide configuration at sites 241, 3037, 14408 and 23403) that are not used in Pekar et al.[15].

Accession	Coll. Date	Country	Type
ON085102	2020-01-01	USA	TTTG
MZ047270	2020-01-20	Poland	TTTG
MZ500923	2020-01-07	USA	TTTG
MZ500486	2020-01-20	USA	TTTG
MZ500330	2020-01-21	USA	TTTG
MZ500751	2020-01-25	USA	TTTG
EPI_ISL_8311708	2020-01-10	Lebanon	TTTG

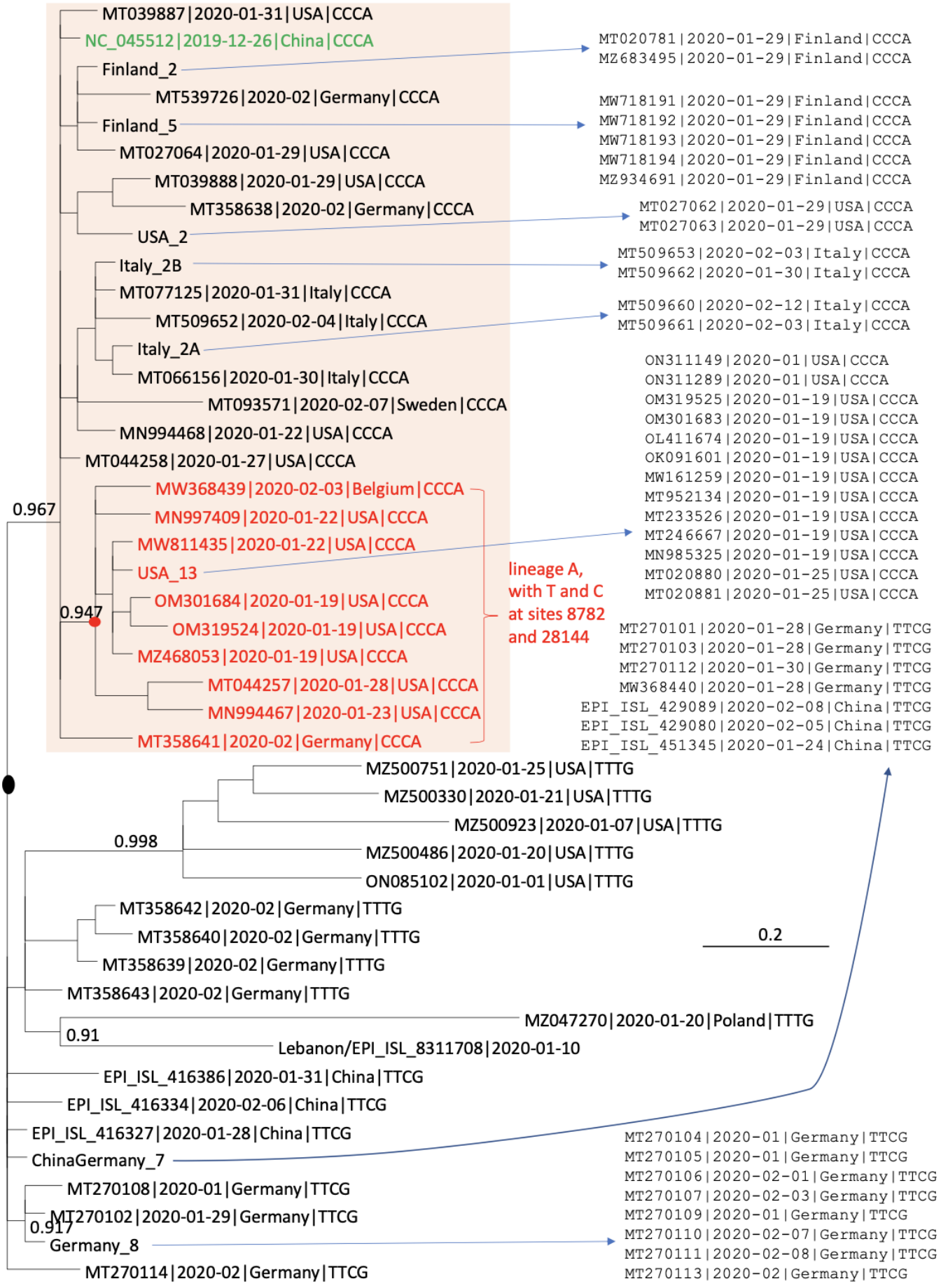
MT039887|2020-01-31|USA|CCCA
NC_045512|2019-12-26|China|CCCA
Finland_2
MT539726|2020-02|Germany|CCCA
Finland_5
MT027064|2020-01-29|USA|CCCA
MT039888|2020-01-29|USA|CCCA
MT358638|2020-02|Germany|CCCA
USA_2
Italy_2B
MT077125|2020-01-31|Italy|CCCA
MT509652|2020-02-04|Italy|CCCA
Italy_2A
MT066156|2020-01-30|Italy|CCCA
MT093571|2020-02-07|Sweden|CCCA
MN994468|2020-01-22|USA|CCCA
MT044258|2020-01-27|USA|CCCA
MW368439|2020-02-03|Belgium|CCCA
MN997409|2020-01-22|USA|CCCA
MW811435|2020-01-22|USA|CCCA
USA_13
OM301684|2020-01-19|USA|CCCA
OM319524|2020-01-19|USA|CCCA
MZ468053|2020-01-19|USA|CCCA
MT044257|2020-01-28|USA|CCCA
MN994467|2020-01-23|USA|CCCA
MT358641|2020-02|Germany|CCCA
0.967
0.947
lineage A, with T and C at sites 8782 and 28144
MT020781|2020-01-29|Finland|CCCA
MZ683495|2020-01-29|Finland|CCCA
MW718191|2020-01-29|Finland|CCCA
MW718192|2020-01-29|Finland|CCCA
MW718193|2020-01-29|Finland|CCCA
MW718194|2020-01-29|Finland|CCCA
MZ934691|2020-01-29|Finland|CCCA
MT027062|2020-01-29|USA|CCCA
MT027063|2020-01-29|USA|CCCA
MT509653|2020-02-03|Italy|CCCA
MT509662|2020-01-30|Italy|CCCA
MT509660|2020-02-12|Italy|CCCA
MT509661|2020-02-03|Italy|CCCA
ON311149|2020-01|USA|CCCA
ON311289|2020-01|USA|CCCA
OM319525|2020-01-19|USA|CCCA
OM301683|2020-01-19|USA|CCCA
OL411674|2020-01-19|USA|CCCA
OK091601|2020-01-19|USA|CCCA
MW161259|2020-01-19|USA|CCCA
MT952134|2020-01-19|USA|CCCA
MT233526|2020-01-19|USA|CCCA
MT246667|2020-01-19|USA|CCCA
MN985325|2020-01-19|USA|CCCA
MT020880|2020-01-25|USA|CCCA
MT020881|2020-01-25|USA|CCCA
MT270101|2020-01-28|Germany|TTCG
MT270103|2020-01-28|Germany|TTCG
MT270112|2020-01-30|Germany|TTCG
MW368440|2020-01-28|Germany|TTCG
EPI_ISL_429089|2020-02-08|China|TTCG
EPI_ISL_429080|2020-02-05|China|TTCG
EPI_ISL_451345|2020-01-24|China|TTCG
MZ500751|2020-01-25|USA|TTTG
MZ500330|2020-01-21|USA|TTTG
MZ500923|2020-01-07|USA|TTTG
0.998
MZ500486|2020-01-20|USA|TTTG
ON085102|2020-01-01|USA|TTTG
MT358642|2020-02|Germany|TTTG
MT358640|2020-02|Germany|TTTG
MT358639|2020-02|Germany|TTTG
0.2
MT358643|2020-02|Germany|TTTG
MZ047270|2020-01-20|Poland|TTTG
0.91
Lebanon/EPI_ISL_8311708|2020-01-10
EPI_ISL_416386|2020-01-31|China|TTCG
EPI_ISL_416334|2020-02-06|China|TTCG
EPI_ISL_416327|2020-01-28|China|TTCG
ChinaGermany_7
MT270104|2020-01|Germany|TTCG
MT270105|2020-01|Germany|TTCG
MT270106|2020-02-01|Germany|TTCG
MT270107|2020-02-03|Germany|TTCG
MT270109|2020-01|Germany|TTCG
MT270110|2020-02-07|Germany|TTCG
MT270111|2020-02-08|Germany|TTCG
MT270113|2020-02|Germany|TTCG
MT270108|2020-01|Germany|TTCG
MT270102|2020-01-29|Germany|TTCG
0.917
Germany_8
MT270114|2020-02|Germany|TTCG

Fig. 1. Phylogenetic tree of early lineages of SARS-CoV-2 isolated in Europe (excluding UK) and USA, together with one early D614G genome from Lebanon and seven key representative SARS-CoV-2 genomes from China, including the reference genome (NC_045512, colored green). All genomes have at least 29740 nt, and "human" as host. Collection dates were up to February 2020 for European sequences, and by Jan. 31, 2020 for USA sequences. OTU names are in the form of accession|collection date|country|type where "accession" is either GenBank or GISAID accession and "type" is defined by the four nucleotides at sites 241, 3037, 14408 and 23403 that separate the D614G (TTTG) lineage from those early strains isolated in China (CCCA). Viral genomes missing any of these four sites were excluded. Lineage A is characterized by C8782T and T28144C. The shaded area includes the original viral classification of lineage A and lineage B. Genomes were aligned by MAFFT[20] with the FFT-NS-2 option. The unrooted phylogenetic tree was reconstructed with PhyML [21], with a GTR model and optimization of topology, branch lengths and rates.

References

1. ^*Oscar A. MacLean, Spyros Lytras, Steven Weaver, Joshua B. Singer, Maciej F. Boni et al. (2021). Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLOS Biology. 19(3):e3001115. doi:10.1371/journal.pbio.3001115.*
2. ^*Hongru Wang, Lenore Pipes, Rasmus Nielsen. (2021). Synonymous mutations and the molecular evolution of SARS-CoV-2 origins. Virus evolution. 7:veaa098. doi:10.1093/ve/veaa098.*
3. ^*Maciej F. Boni, Philippe Lemey, Xiaowei Jiang, Tommy Tsan-Yuk Lam, Blair Perry et al. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol. 5:1408–1417. doi:10.1101/2020.03.30.015008.*
4. ^*S. Lytras, J. Hughes, D. Martin, Klerk Arné de, Lourens Rentia et al. (2021). Exploring the natural origins of SARS-CoV-2 in the light of recombination. bioRxiv (accessed on Sept 1, 2021). doi:10.1101/2021.01.22.427830 ,.*
5. a, b, c*Xuhua Xia. (2021). Dating the Common Ancestor from an NCBI Tree of 83688 High-Quality and Full-Length SARS-CoV-2 Genomes. Viruses. 13(9):1790. doi:10.3390/v13091790.*
6. a, b, c, d*X. Xia. (2022). Improved method for rooting and tip-dating a viral phylogeny. In: H. H.-S. Lu, B. Scholkopf, M. T. Wells, H. Zhao, editors. Handbook of Computational Statistics, II. Berlin: Springer (In press). p.*
7. ^*M. T. Gilbert, A. Rambaut, G. Wlasiuk, T. J. Spira, A. E. Pitchenik et al. (2007). The emergence of HIV/AIDS in the Americas and beyond. Proceedings of the National Academy of Sciences of the United States of America. 104(47):18566-18570. PubMed PMID: 17978186.*
8. ^*B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 182(4):812-827 e819. doi:10.1016/j.cell.2020.06.043. PubMed PMID: 32697968; PubMed Central PMCID: PMC7332439.*
9. a, b, c*A. Amendola, M. Canuti, S. Bianchi, S. Kumar, C. Fappani et al. (2022). Molecular evidence for SARS-CoV-2 in samples collected from patients with morbilliform eruptions since late 2019 in Lombardy, northern Italy. Environ Res.113979. doi:10.1016/j.envres.2022.113979. PubMed PMID: 36029839; PubMed Central PMCID: PMC9404229.*
10. ^*Giuseppina La Rosa, Pamela Mancini, Giusy Bonanno Ferraro, Carolina Veneri, Marcello Iaconelli et al. (2021).*

SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. Science of the Total Environment. 750:141711. doi:https://doi.org/10.1016/j.scitotenv.2020.141711.

11. ^*G. Fongaro, P. H. Stoco, D. S. M. Souza, E. C. Grisard, M. E. Magri et al. (2021). The presence of SARS-CoV-2 RNA in human sewage in Santa Catarina, Brazil, November 2019. Sci Total Environ. 778:146198. doi:10.1016/j.scitotenv.2021.146198. PubMed PMID: 33714813; PubMed Central PMCID: PMC7938741.*
12. ^*Yongsen Ruan, Haijun Wen, Mei Hou, Ziwen He, Xuemei Lu et al. (2022). The twin-beginnings of COVID-19 in Asia and Europe—one prevails quickly. Natl Sci Rev. 9(4):nwab223. doi:10.1093/nsr/nwab223.*
13. ^*Rebecca Speare-Cole. Scotland confirms first case of coronavirus. Evening Standard. 2020 01 March 2020.*
14. a, b*K. J. Hill, C. D. Russell, S. Clifford, K. Templeton, C. L. Mackintosh et al. (2020). The index case of SARS-CoV-2 in Scotland. J Infect. 81(1):147-178. doi:10.1016/j.jinf.2020.03.022. PubMed PMID: 32205138; PubMed Central PMCID: PMC7118628.*
15. a, b, c, d, e, f*Jonathan E. Pekar, Andrew Magee, Edyth Parker, Niema Moshiri, Katherine Izhikevich et al. (2022). The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science. 377(6609):960-966. doi:doi:10.1126/science.abp8337.*
16. a, b*Andrew Rambaut, Edward C. Holmes, Verity Hill, Áine O'Toole, J. T. McCrone et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. bioRxiv.2020.2004.2017.046086. doi:10.1101/2020.04.17.046086.*
17. ^*Xiaolu Tang, Changcheng Wu, Xiang Li, Yuhe Song, Xinmin Yao et al. (2020). On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 7(6):1012–1023. doi:10.1093/nsr/nwaa036.*
18. ^*E. Volz, V. Hill, J. T. McCrone, A. Price, D. Jorgensen et al. (2020). Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 184:64-75. doi:10.1016/j.cell.2020.11.020. PubMed PMID: 33275900; PubMed Central PMCID: PMC7674007.*
19. ^*Bas B. Oude Munnink, Reina S. Sikkema, David F. Nieuwenhuijse, Robert Jan Molenaar, Emmanuelle Munger et al. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science. 371(6525):172. doi:10.1126/science.abe5901.*
20. ^*K. Katoh, H. Toh. (2008). Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinformatics. 9:212. doi:10.1186/1471-2105-9-212. PubMed PMID: 18439255; PubMed Central PMCID: PMC2387179.*
21. ^*S. Guindon, O. Gascuel. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52(5):696-704.*