Covid-19 pandemic: a (un)predictable event?

Barbara Illi

Institute of Molecular Biology and Pathology – National Research Council (IBPM-CNR), c/o Department of Biology and Biotechnology "Charles Darwin", Sapienza University, Rome, Italy.

Note: highlighted in italic are information for non biologist.

Introduction

For all zoonoses, three conditions have to be satisfied for a pandemic to emerge: 1) a virus has to be compatible with humans; 2) contacts between humans and animals have to be established; 3) human-to-human transmission has to be possible. Unfortunately, Covid-19 pandemic fulfills all these three conditions. However, the effective animal origin of SARS-CoV-2 (CoV-2), the etiological agent of Covid-19, is still not completely clear.

SARS-CoV-2: artificial or natural?

In the conspiracy era, the idea that CoV-2 was an artificially, lab-created virus has destabilized the public opinion. A video circulating on the social media, related to an italian press report about a synthetic SARS coronavirus (SARS-CoV), has immediately raised doubts about CoV-2 origin and about China responsibility in its spread. Nevertheless, that video, which was within a program called "Leonardo", dates back to 2015 and, specifically, reports the construction and characterization of a chimeric virus in which the spike SHC014 protein of a bat coronavirus was inserted into the backbone of a mouse SARS-CoV[1] (figure 1). *The procedure, which is called "cloning", basically consists in a sort of "cut and paste" of a gene into another DNA molecule (which may be circular DNA, named plasmid, or larger viral DNA) by proteins named restriction enzymes, which cut DNA at specific stretch of nucleotides, and other proteins termed "ligases" which, actually, paste the sequences. In this case, the SHC014 gene was cut from an available plasmid and pasted into a mouse SARS-CoV, previously cut by the same restriction enzyme, to permit a correct SHC014 insertion (figure 1).* In that paper, "in vitro" and "in vivo" studies demonstrated the inefficacy of both anti-SARS-CoV neutralizing antibodies and vaccines, respectively, to slow virus replication within host cells and to protect mice from infection. Therefore, a great alarm raised when Leonardo press report was re-published in last March. However, that work was simply a proof of principle that a SARS-CoV-derived virus and related pandemic could

Figure 1

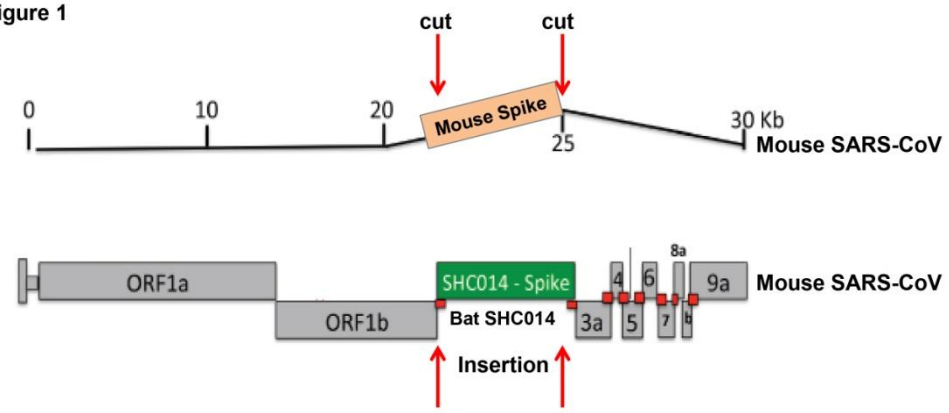

Figure 1. Chimeric mouse/bat SARS-CoV. Scheme of the chimeric construct having a mouse SARS-CoV as the backbone and the SHC014 bat spike protein as the "foreign" gene. Sites of "cut and paste" are highlighted by red arrowheads. *(Adapted from Menachery et al., Nat Med, 2015)*

Figure 2

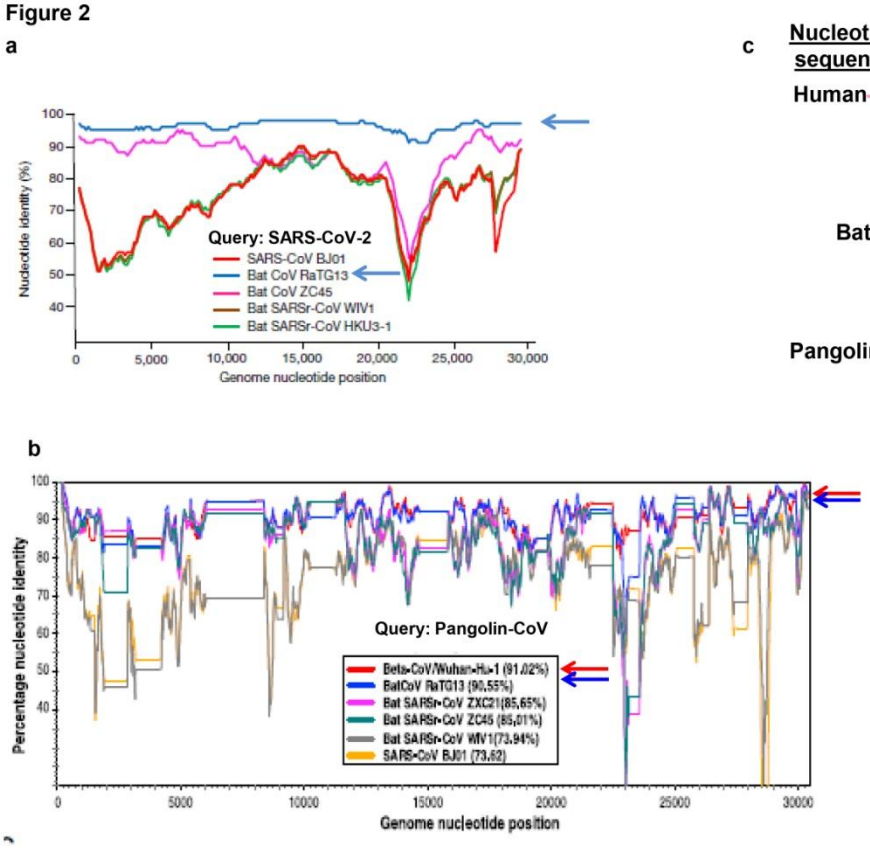

Figure 2. a) Comparison and identity among SARS-CoV-2 genome and other bat SARSr-CoVs. BatCoVRaTG13 (highlighted by a blue arrowhead) is the CoV-2 most closely related CoV. b) Comparison and identity among Pangolin CoV, CoV-2 (red arrowhead) and BatCoVRaTG13 (blue arrowhead) and other bat and SARS-CoVs. CoV-2 and RaTG13 genomes are basically overlapping and almost identical to the genome of Pangolin CoV. c) Phylogenetic analysis based on nucleotide sequence homology (pink lines, left) or aminoacid homology (blue lines, right). *(Adapted from Zhang et al., Current Biol, 2020)*

have raised and been dangerous for the human being due to the potential lack of therapies and vaccines. Indeed, SARS-CoV-2 has no traces, in its genome, of any mouse SARS-CoV sequence. Another concern was about the presence of 4 small aminoacidic sequences in Cov-2 spike protein, the tool the virus uses to enter human host cells, common to the Human Immunodeficiency Virus (HIV). Such similarity was recently

stressed by Prof. Luc Montagnier, one of the “fathers” of HIV, strengthening the hypothesis of Cov-2 as the product of a genetic manipulation. However, those sequences are common to about other 300 unrelated proteins. Furthermore, since last February, GenBank searches for similarities between HIV and CoV-2 genomes excluded the possibility that CoV-2 gained those specific sequences from HIV[2]. More recently, a Correspondence on Nature Medicine Journal, has well clarified that CoV-2 was the product of a natural selection of zoonotic viruses[3], as described in the next paragraph.

SARS-CoV-2: bat or pangolin?

The first information regarding Cov-2 origin came from the analysis of the broncheoalveolar fluid lavage (BALF) of 1 among 7 patients with severe pneumonia, admitted to the Intensive Care Unit of Wuhan Jin Yin-Tan hospital[4]. Database searches for similarities among the viral genome isolated from that patient with known viral genomes addressed that the virus was closely related to SARS coronaviruses (SARSr-CoV). Further analyses demonstrated that the virus likely descends from a bat coronavirus and, specifically, from Rinolophus affinis TG13 virus (BatCoVRaTG13) (figure 2a). Thereafter, the search for CoV-2 intermediate hosts has focused scientists’ attention on other SARSr-CoV viruses reservoirs, such as Malayan Pangolins, which are illegally imported in the Guandong province and sold in chinese wet markets. Indeed, the analysis of the genome of a virus extracted from the lungs of 2 Pangolins, whose death occurred in a time frame compatible with the emergence of Covid-19 outbreak, revealed that those animals were infected by a virus (named Pangolin-CoV) closely related to SARS-CoV-2 and other bat SARS-CoV viruses[5] (figure 2b). Pangolin-CoV and human CoV-2 share five key aminoacids in the receptor binding domain of the spike protein, a homology which makes CoV-2 closer to Pangolin-CoV than to BatCoVRaTG13 suggesting the occurrence of mutations in Pangolins of BATCoVRaTG13 and Pangolins as intermediate hosts (figure 2c, blue lines). However, the phylogenetic analysis of Pangolin-CoV, at the whole genome level revealed Pangolin-CoV as the putative common ancestor of both BatCoVRaTG13 and CoV-2 (figure 2c, pink lines) and altogether clustered in a new group of betacoronaviruses named the “SARS-CoV-2 group”. Nevertheless, the Pangolin origin of SARS-CoV-2 is still under debate.

Figure 3

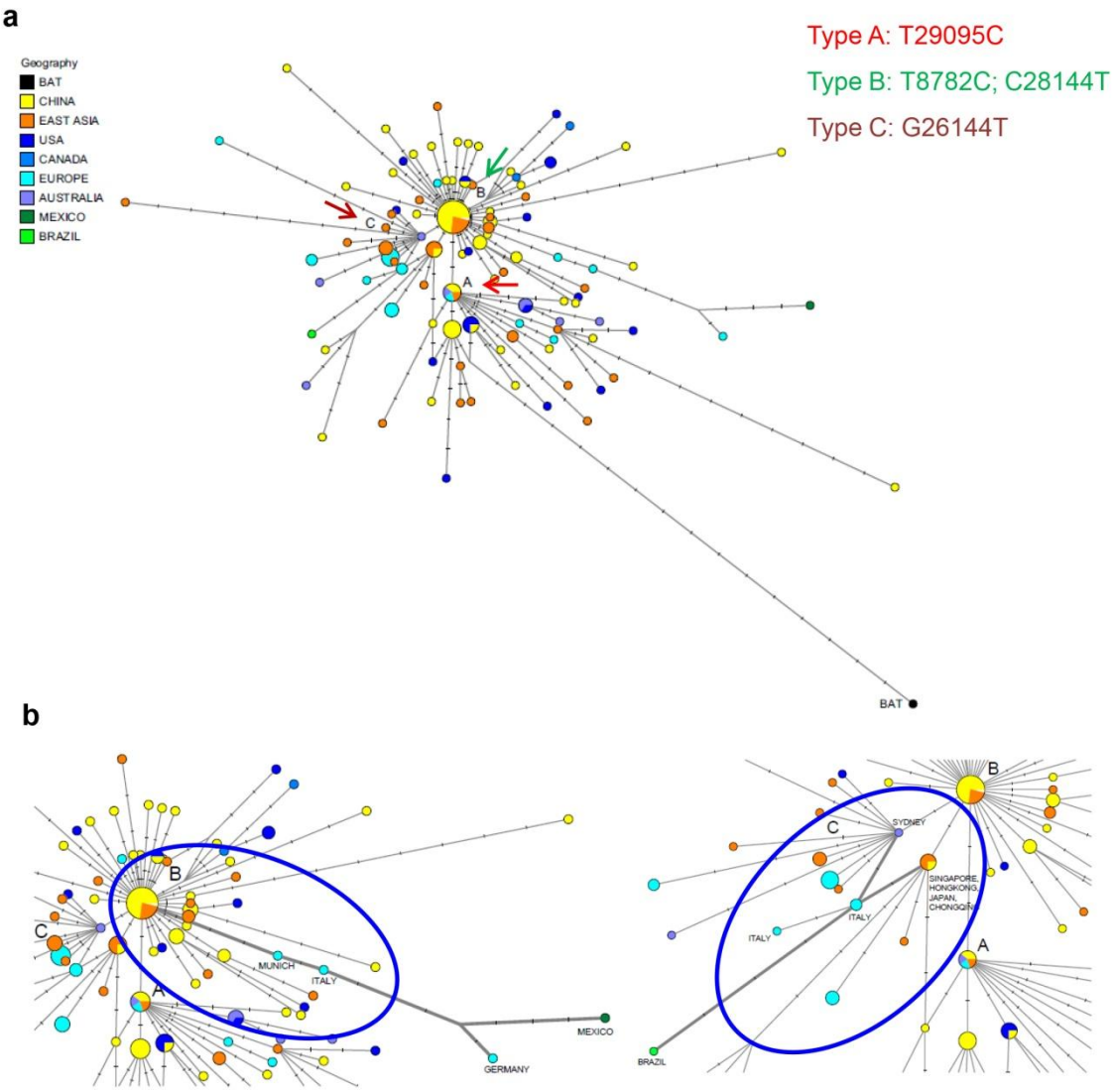

Figure 3. a). A network of 160 CoV-2 genomes. Type A is the root type as defined by the comparison with the BatCoVRaTG13. Each circle area is proportional to the number of genomes included. Each notch on the links represents a mutational event. At the top, right, mutations typical of each node are identified by colors corresponding to arrowheads indicating nodes in the network. b) Routes of infection in Italy. Apparently, in Italy two virus entries occurred, one from Munich, type-B derived (left) and one from Singapore, type C-derived (right). Both are highlighted by blue circles. *(Adapted from Forster et al., PNAS, 2020)*

SARS-CoV-2 mutability and variants

Survival and fitness maintenance underlie the high mutation rate of RNA viruses, such as Coronaviruses. When bottleneck events occur and viruses slow their replication rate they lose their fitness and deleterious mutations for their survival may appear. Conversely, repeated and large population passages result in rapid fitness gain. However, RNA viruses

can tolerate a low number and few types of mutations[6]. Furthermore, it has to be considered that a large part of mutations are “synonymous” or “silent”, *that is the change in the RNA sequence does not correspond to a change in the corresponding aminoacids encoded by that specific RNA sequence.* A study on 95 CoV-2 genomes revealed that all of them were similar at 99.9% both at the DNA and protein levels. Nevertheless, 13 selective genomic sites have been discovered in CoV-2 with high mutation rates. Importantly, some “hot-spot” mutations occur in the genes encoding for spike and for nucleocapsid proteins and this is particularly relevant for viral replication, transmission and induced immunity related studies[7]. Mutation studies have been important for the detection of CoV-2 variants, which have been found to characterize specific geographic areas. The analysis of 160 CoV-2 genomes worldwide has permitted to identify three principle types of CoV-2, marked by specific aminoacid changes[8], termed A, B and C. The ancestral type A is the “father” of type B, which further evolved in type C. Type A, which is further subdivided in two subgroups by the synonymous mutation T29095C (*T=thymine; C=cytosine; 29095 is the nucleotide position in the genome*), spread mainly in China and less in East Asia, Europe, USA and Australia (figure 3a). From the A “node”, node B derived by two mutations, one synonymous (T8782C) and one non-synonymous (C28144T), this latter changing the aminoacid leucine to serine. It is interesting to note that type B has basically not left East Asia. Type-B genomes derived from the ancestor B-type and spread outside East Asia present other mutations. Type C differs from its parent B by a non-synonymous mutation (G26144T; *G=guanine*) which changes a glycine into a valine. This is the major CoV-2 type spread in Europe (France, Germany, Italy, Sweden) and which reached California and Brazil. Interestingly, these analyses showed that, apparently, here in Italy we had one entry of type B-derived CoV-2 from Munich (figure 3b left) and one earlier, type C-derived, from Singapore (figure 3b, right). This kind of analyses are particularly important to reconstruct infection paths. However, this work is a picture of the pandemic at early stages.

Potential harmful mutations: the RdRp variant

Very recently other 8 mutations have been detected, 5 predominantly present in Europe and 3 in North America. All are non-synonymous and produce changes in the aminoacid compositions of the proteins encoded by the related mutated genes. Among the “european mutations”, one deserves particular attention because affects CoV-2 RNA-dependent RNA polymerase (RdRp)[8]. SARS-CoV and CoV-2 RdRp are highly conserved, indicating that

Coronaviruses tend to conserved RdRp structure and function. RdRp mutation, specifically at nucleotide 14408 of the viral genome, occurred in February 9th in England, when a dramatic increase in Covid-19 cases was observed in Europe. More importantly, starting from that date, an increase in the mutation rate of other CoV-2 proteins were observed (figure 4) in viral genomes carrying RdRp with respect to genomes with non-mutated RdRp. The 14408 mutation is not in the region of RdRp involved in replication, but may affect the interaction with other non-structural proteins, such as nsp14, which may have, according to the homology with SARS-CoV nsp14, a "proofreading" activity, *that is nsp14 corrects errors which RdRp may make when replicates viral RNA, eliminating improper nucleotide and inserting the correct ones.* This may explain the increase in the mutation frequency of other CoV-2 proteins as a result of replication errors. Further, the site of mutation falls close to the site of interaction with RdRp-interfering drugs, such as Filibuvir and Teogobuvir. Therefore, it will be important to verify whether this mutation may impact RdRp fidelity and/or drug resistance.

Figure 4

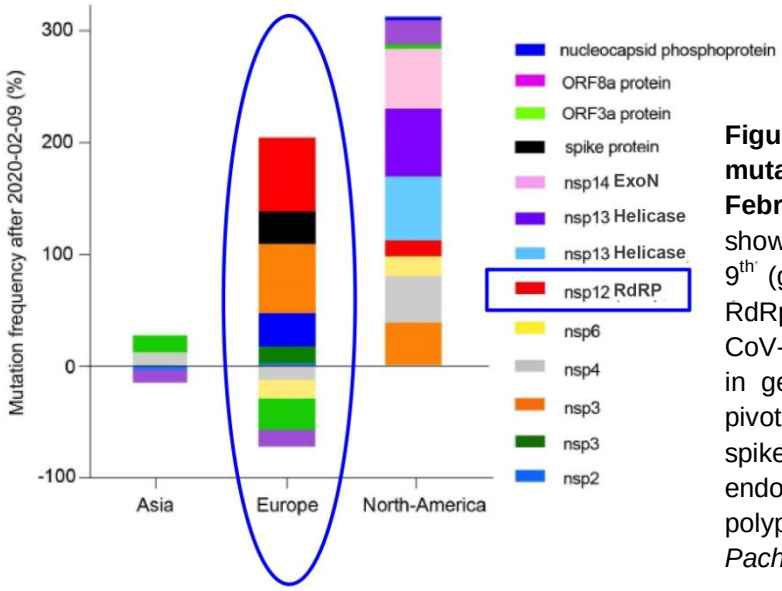

Figure 4. Increase of CoV-2 mutation frequency starting from February 9th 2020. The histogram shows that, starting from February 9th (grey line), when a mutation in RdRp was detected, an increase in CoV-2 mutation frequency occurred in genes encoding proteins playing pivotal roles in virus biology, such as spike, nucleocapsid, nsp3 (an endopeptidase which cleaves CoV-2 polyprotein pp1ab). *(Adapted from Pachetti et al., J Transl Med, 2020)*

Conclusions

Since its appearance CoV-2 has mutated, generating a huge number of variants. However, (viral) genotype-(clinical) phenotype correlations are still lacking. Also, mortality

rate in specific geographic areas has not been yet correlated with the presence of specific viral variants. However, it has been already demonstrated that mutations in RNA viruses may transform apathogenic phenotypes into pathogenic ones[6]. Importantly, a RdRp mutation occurred in England is strictly related to the increase of mutations in European viral genomes as well as to the increase of Covid-19 clinical cases. Therefore, it will be important to verify whether this mutation affects RdRp replication fidelity and whether it may have produced the appearance of drug-resistant viral phenotypes. Furthermore, the zoonotic origin of Covid-19, as well as of SARS and MERS epidemic, suggests this pandemic a predictable event. In particular, the contamination between urban and forest areas will repeatedly foster the jump of viruses from animals to humans and makes preventive strategies extremely compelling.

References

1. Menachery VD, Yount BL Jr, Debbink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH, Lanzavecchia A, Marasco WA, Shi ZL, Baric RS. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med. 2015; 21: 1508-13. doi: 10.1038/nm.3985.
2. Xiao C, Li X, Liu S, Sang Y, Gao SJ, Gao F. HIV-1 did not contribute to the 2019-nCoV genome. Emerg Microbes Infect. 2020; 9:378-381. doi: 10.1080/22221751.2020.1727299.
3. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020; 26: 450-452. doi: 10.1038/s41591-020-0820-9.
4. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579: 270-273. doi: 10.1038/s41586-020-2012-7.
5. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV 2 Associated with the COVID-19 Outbreak. Curr Biol. 2020; 30: 1346-1351.e2. doi: 10.1016/j.cub.2020.03.022.
6. Domingo E, Holland JJ.RNA virus mutations and fitness for survival. Annu Rev Microbiol. 1997;51:151-78.

7. 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. doi: 10.1002/jmv.25762.
8. Forster P, Forster L, Renfrew C, Forster M.Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci U S A. 2020; 117: 9241-9243. doi: 10.1073/pnas.2004999117.
9. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020; 18: 179. doi: 10.1186/s12967-020-02344-6.