Is there evidence for genetic change in SARS-CoV-2 and if so, do mutations affect virus phenotype?

Wendy Barclay, Imperial College London.

Cariad Evans, Sheffield Teaching Hospitals NHS Foundation Trust.

Alistair Darby, University of Liverpool.

Andrew Rambaut, University of Edinburgh.

Oliver Pybus, University of Oxford.

David Robertson, University of Glasgow.

Ewan Harrison, Wellcome Sanger Institute.

Sharon Peacock, Director, COG-UK.

Meera Chand, Public Health England, Guy's and St Thomas' NHS Foundation Trust.

Julian A. Hiscox, University of Liverpool.

Summary

As the SARS-CoV-2 virus passes from person to person, it is mutating. This is normal, and most mutations will not matter. But there is chance that some mutations will affect how easily the virus spreads between people, how likely it is to make someone ill and how severely ill that person becomes. At the moment, researchers believe that none of the genetic changes found in the virus increase or decrease severity of disease, but they may affect transmission. This might change in the future, as vaccines and treatments have the potential to select different changes in the virus. In the UK, researchers are currently monitoring whether mutations are occurring, but are not systematically checking whether these mutations 'matter.' This is an important gap in our knowledge.

Background context: we can expect mutations to accumulate in the SARS-CoV-2 genome

SARS-CoV-2 is a coronavirus, with a large 30kb positive strand RNA genome. An integral part of the replication mechanism in coronaviruses involves a discontinuous step in the synthesis of viral RNA, with the natural consequence of a high degree of a recombination resulting in the insertion of viral and non-viral sequences into or deletion of viral sequence from the genome. This is one of the major processes by which coronaviruses switch host range or change their pathogenesis/virulence. Recombination with either an unsampled SARS-like virus or human, bat or pangolin host sequence probably gave rise to the furin cleavage site that is found within the SARS-CoV-2 spike glycoprotein and may contribute to infectivity and transmissibility. SARS-CoV and MERS-CoV are thought to have had recombination in their evolutionary history (1, 2) and in new outbreaks (3). Deletions in the genome of the porcine coronavirus transmissible gastroenteritis virus (TGEV) have given rise to a new virus called porcine respiratory coronavirus (PRCV) (4). Human coronavirus OC43 is thought to have acquired the hemagglutinin esterase (HE) gene from recombination between a progenitor coronavirus and influenza C virus.

SARS-CoV-2 will also continually accrue point mutations, measured as single nucleotide changes in the virus genome sequence, as it replicates in host cells and is target by the host immune system (both innate and adaptative) and then transmits among the human population. Most mutations will be of no consequence to the virus biology and the rate of change via point mutation is relatively slow

compared to other RNA viruses such as HIV or influenza because coronaviruses have an error correction enzyme.

Examples from other coronaviruses show that point mutations and recombination events could lead to changes in virus biology, transmissibility (5), antigenicity, pathogenesis (6), virulence and new viruses (7), and thus could have the potential to increase disease severity or affect vaccine efficacy. Mutations in the furin cleavage site of the spike glycoprotein protein during infection of cats with feline enteric coronavirus (FECV), which causes a subclinical infection, can lead to infectious peritonitis virus (FIPV), which causes a systemic and fatal disease. These changes occur in the same animal during infection (8). The avian coronavirus, infectious bronchitis virus, illustrates the need to continuously monitor for genotype changes as there is recombination between distant circulating strains leading to antibody escape mutants (9, 10). There are now hundreds of different isolates of this virus, all with slight changes in the spike glycoprotein, that may allow antibody and vaccine escape mutants similar to other humans coronaviruses such as MERS-CoV (11). Evolutionary theory doesn't predict that there will inevitably be selection to increase or reduce virulence. The most likely scenario is that virulence will remain unchanged.

Is there evidence for genetic change in SARS-CoV-2?

There are tens to hundreds of genetically distinct lineages of SARS-CoV-2 currently circulating in the UK. Some will be descended from first wave introductions; others will be recently imported. Each lineage accrues new nucleotide changes at a rate of ~2 nucleotides per month. Between 11th March and 25th of September 2020, the COVID-19 Genomics UK (COG-UK) Consortium has sequenced more than 70,000 SARS-CoV-2 genomes, representing >50% of the genomes available in the global repository. Sequencing currently comprises of ~1400 positive samples each day, routed from pillar 1 and pillar 2 testing labs. So far, mutations observed have been used to provide insights into transmission by providing the ability to link cases during outbreaks. These nucleotide changes may also affect the efficiency of diagnostics for both nucleic acid based systems such as RT-PCR but particularly on serology given the variability of the spike gene/protein in other coronaviruses.

Sequence data from COG-UK shows that that the SARS-CoV-2 genome is certainly changing as the pandemic ensues, and this is reflected by other sequencing programs globally. The viruses that are circulating in the UK show considerable diversity; there were introductions of at least 1200 transmission lineages between March and April 2020 with representation of all of the global lineages, distinguishable by shared patterns of single nucleotide polymorphisms (SNPs) in the viral genome and potentially insertions and deletions. The overall rate of evolution of SARS-CoV-2 has been estimated to be approximately 1.0 x 10^{-3} substitution/site/year across the entire genome but this includes both amino acid replacements and 'silent' nucleotide changes (12). This is similar in terms of rate to Ebola virus during the West African outbreak (13).

Does the genetic change observed in SARS CoV2 lead to any phenotypic change of consequence?

There are two ways to attempt to answer this question. The first approach is already funded in UK and underway: genomic data from COG-UK is considered in context of epidemiological and clinical data: A pipeline within COG-UK (CoV-GLUE) scans the available data for non-synonymous changes (http://cov-glue.cvr.gla.ac.uk/#/home). Changes of concern for the pandemic response would be those emerging at a rapid rate or becoming highly prevalent. This approach has been taken to consider the effect of the D614G spike protein mutations as described below. In addition, in the near future a new module in the daily COG-UK analysis pipeline will pull out such mutations in a way that can be easily analysed. These genomes will be linked to severity and outcome data.

The second way to assess the consequence of genetic change is to perform laboratory based virological assays that compare the characteristics of different strains of virus, and to understand the mechanisms that underlie the changes observed. However, there is not currently a UK virology pipeline set up to analyse the phenotypic consequences of SARS-CoV-2 mutations. Thus, although mutations are being documented in the UK, we do not have the ability to clearly and definitively answer whether they 'matter'. For example, whether particular mutations have begun to predominate due to them conferring a fitness advantage that can be measured or are merely the result of founder effects, or whether reinfections are due to antigenic drift of the virus rather than waning immunity (14).

What do we know about the consequence of genetic changes that have been observed in UK and globally thus far?

Several genetic changes are reported in the global dataset, some of which are present in viruses in the UK. These are reviewed here although investigations of most are incomplete:

D614G change in the spike glycoprotein gene. The D614G mutation in spike has been of particular interest since the change from 614D to 614G is predicted to enhance ACE2 receptor binding by retaining the spike protein in the 'up' conformation. Indeed, this substitution has been associated with increased infectivity *in vitro* using pseudotyped lentiviruses (15) (https://www.biorxiv.org/content/10.1101/2020.07.04.187757v2?). The frequency of the 614G variant has grown to dominate in many locations that reported 614D earlier in the pandemic, but this can be hard to disentangle from epidemiological factors. A COG-UK analysis using more than 25,000 SARS-CoV-2 genomes demonstrated that this increase occurred in a manner consistent with a selective advantage for transmission (~20%), although no evidence of higher COVID-19 mortality or clinical severity was found (https://www.medrxiv.org/content/10.1101/2020.07.31.20166082v2). In particular the COG-UK preprint that described this work suggests any decrease in case fatality seen since the D614G mutation began to predominate can be attributed to either the increased numbers of cases in younger age group, or to improvements in treatment such as introduction of dexamethasone for severe cases (Figure 1).

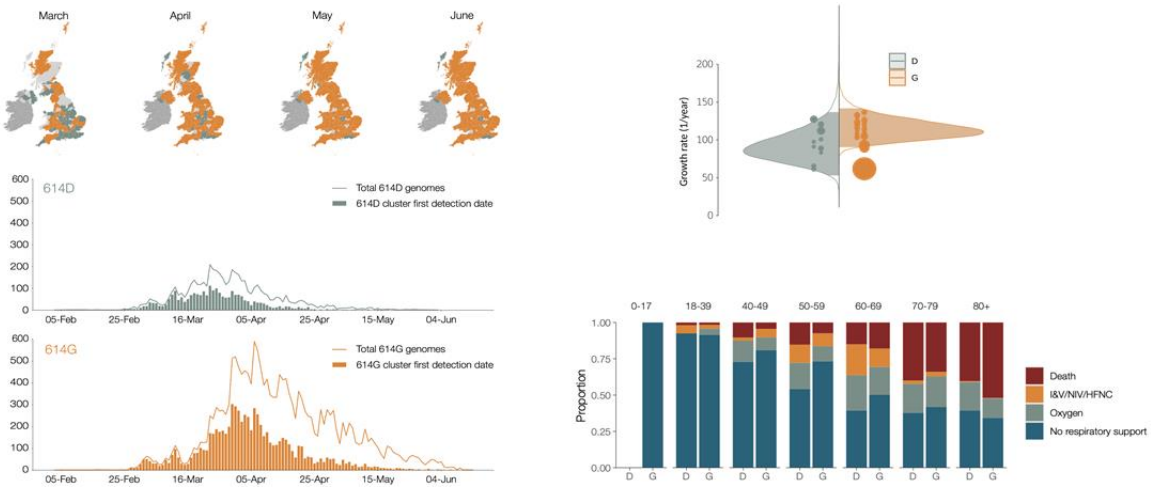

Figure 1. COG-UK data showing the effects of the SARS-CoV-2 D614G mutation on transmission and clinical management.

This illustrates the power of the combined epidemiological and genomics approach. Very recently a preprint has been posted describing experiments that validate these hypotheses using traditional biochemical and virological approaches and provide mechanistic confirmation for the increased transmission. This combination of genomics and molecular virology and the utility of reverse genetics

to engineer precise mutations into the virus has shown that the D614G increased ACE2 binding and fusion and thus onwards transmission measured in animal models (https://www.biorxiv.org/content/10.1101/2020.07.04.187757v1). Importantly, this mutation does not affect the ability of antibodies to neutralize the virus thus no requirement for vaccine update. The work performed by scientists in US illustrates a pipeline of work that is currently lacking in UK.

N439K in the receptor binding motif (RBM) in the spike glycoprotein gene. This mutation was present in a lineage, until recently almost unique to Scotland, that had infected more than 500 people. The mutation increases the binding affinity of spike protein to ACE2 which is one mechanism by which the virus might escape neutralization by monoclonal antibodies. The Scottish lineage defined by N439K has not been observed since the 20th June 2020. Detailed investigation of the available clinical data showed no increased severity, although any effects for this mutation alone are difficult to determine as it occurs on the D614G background. N439K has also been identified in another significant lineage which has been sampled in Romania, Norway, Switzerland and Ireland and now all parts of the UK. It has also been detected in four linked infections in the US and sporadically in the data. Multiple other mutations in the RBM specifically and spike (some of known antigenic significance) have been observed (some in linked infections) demonstrating that although (relatively) slow evolving, SARS-CoV-2 can readily tolerate mutations which have the potential to be antigenic/vaccine escape mutants or alter the interaction between virus and ACE2 receptor.

Deletions within and around the furin cleavage site of spike protein. During propagation of SARS-CoV-2 in cell culture such as Vero cells, a deletion in or near to the furin cleavage site between S1 and S2 spike subunits is often reported (16, 17). Similar deletions have been detected in clinical samples at low frequency, around 1-2%, and as a minor variant in many samples (18). There is no report of viruses circulating that lack the furin cleavage site. Molecular virological assessment show that such viruses are attenuated for infection of upper respiratory tract cells and animal studies illustrate these mutants are less pathogenic and less transmissible in vivo (https://www.biorxiv.org/content/10.1101/2020.09.28.317685v1). Whether they are associated with any impact clinically is not clear.

P323L mutation in the viral polymerase. This mutation always co-occurs with the D614G mutation in the spike glycoprotein, although the relevance of this association for the properties of the virus are at present unclear. There is not currently a system available to measure the consequence of polymerase mutations in vitro.

Deletion in the orf8 gene. The ORF8 gene is a viral accessory protein that is proposed to interfere with the innate response of cells to infection. Viruses harbouring a deletion of this gene have been reported in patients from Singapore and were associated with milder infection (19), but the virus lineage carrying this deletion appears to have been eliminated. Deletions in this gene and others involved encoding proteins that antagonise the innate immune response have been identified at a minor variant level within patients but are presumably not selected.

Conclusions

In summary, there is no evidence that the limited number of genetic changes reported and investigated to date have brought about major shifts in viral tropism or virulence of SARS-CoV-2. Whilst limited diversity has emerged thus far, this may change in the next phase of the epidemic as selective pressures exerted by increased levels of natural immunity in the human population as well as using vaccines, treatments and non-pharmaceutical interventions increase. It is particularly

important that surveillance and analysis for antigenic change is established in the lead up to the roll out of a vaccination programme. COG-UK provides a 'genome first' surveillance programme by scanning all sequenced viruses. However, we currently lack the capability in the UK to rapidly and systematically assess the biological significance of detected genetic change. This represents a major gap in our ability to respond and hone mitigations to the virus and requires a coordinated molecular virology-based enterprise. This has to be unpinned by continued sequencing of genomes in patients and we note that work is underway to increase capacity further to enable the sequencing of up to 10,000 whole SARS-CoV-2 genomes per week. The ability to rely on genomic analysis and clinical linked parameters alone is further hampered because viral load measurements (Ct values from testing) are not routinely obtained or if available are acquired using different platforms that does not allow cross comparison. A proposed rapid assessment pipeline of genotype to phenotype changes underpinned by genomic/clinical data is illustrated (Figure 2).

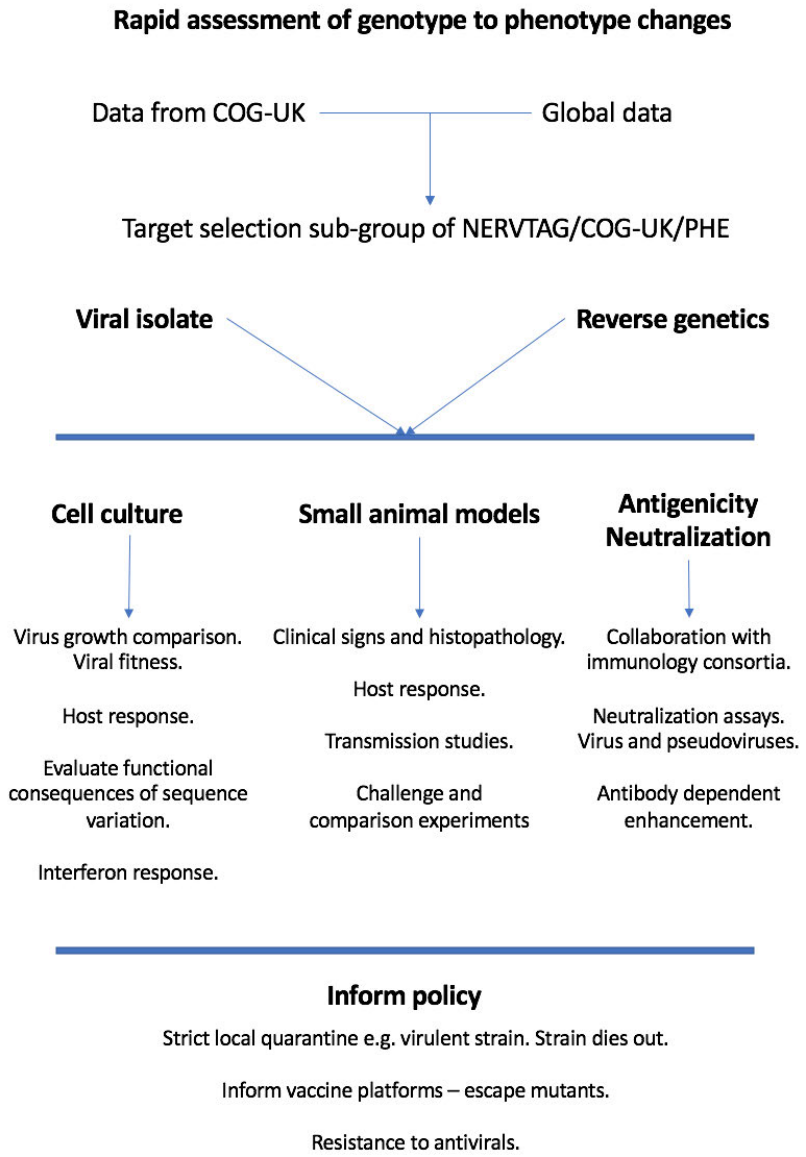

Figure 2. Work flow showing the proposed step in investigating, validating and identifying whether genome changes result in phenotypic changes related to key issues surrounding the infectivity and disease profile of SARS-CoV-2.

1. Lau SK, Feng Y, Chen H, Luk HK, Yang WH, Li KS, 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(20):10532-47.
2. Dudas G, Rambaut A. MERS-CoV recombination: implications about the reservoir and potential for adaptation. Virus Evol. 2016;2(1):vev023.
3. Kim JI, Kim YJ, Lemey P, Lee I, Park S, Bae JY, et al. The recent ancestry of Middle East respiratory syndrome coronavirus in Korea has been shaped by recombination. Sci Rep. 2016;6:18825.
4. Chen F, Knutson TP, Rossow S, Saif LJ, Marthaler DG. Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States. Sci Rep. 2019;9(1):3953.
5. Zhang Y, Li J, Xiao Y, Zhang J, Wang Y, Chen L, et al. Genotype shift in human coronavirus OC43 and emergence of a novel genotype by natural recombination. J Infect. 2015;70(6):641-50.
6. So RTY, Chu DKW, Miguel E, Perera R, Oladipo JO, Fassi-Fihri O, et al. Diversity of Dromedary Camel Coronavirus HKU23 in African Camels Revealed Multiple Recombination Events among Closely Related Betacoronaviruses of the Subgenus Embecovirus. J Virol. 2019;93(23).
7. Jackwood MW, Boynton TO, Hilt DA, McKinley ET, Kissinger JC, Paterson AH, et al. Emergence of a group 3 coronavirus through recombination. Virology. 2010;398(1):98-108.
8. Licitra BN, Millet JK, Regan AD, Hamilton BS, Rinaldi VD, Duhamel GE, et al. Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. Emerg Infect Dis. 2013;19(7):1066-73.
9. Pohuang T, Chansiripornchai N, Tawatsin A, Sasipreeyajan J. Sequence analysis of S1 genes of infectious bronchitis virus isolated in Thailand during 2008-2009: identification of natural recombination in the field isolates. Virus Genes. 2011;43(2):254-60.
10. Mardani K, Noormohammadi AH, Ignjatovic J, Browning GF. Naturally occurring recombination between distant strains of infectious bronchitis virus. Arch Virol. 2010;155(10):1581-6.
11. Kim YS, Aigerim A, Park U, Kim Y, Rhee JY, Choi JP, et al. Sequential Emergence and Wide Spread of Neutralization Escape Middle East Respiratory Syndrome Coronavirus Mutants, South Korea, 2015. Emerg Infect Dis. 2019;25(6):1161-8.
12. Pereson MJ, Mojsiejczuk L, Martinez AP, Flichman DM, Garcia GH, Di Lello FA. Phylogenetic analysis of SARS-CoV-2 IN THE FIRST MONTHS SINCE ITS EMERGENCE. J Med Virol. 2020.
13. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. Nature. 2015;524(7563):97-101.
14. To KK, Hung IF, Ip JD, Chu AW, Chan WM, Tam AR, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis. 2020.
15. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020;182(4):812-27 e19.
16. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ, et al. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med. 2020;12(1):68.
17. Klimstra WB, Tilston-Lunel NL, Nambulli S, Boslett J, McMillen CM, Gilliland T, et al. SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely infected hospitalized COVID-19 patients. J Gen Virol. 2020.
18. Liu Z, Zheng H, Lin H, Li M, Yuan R, Peng J, et al. Identification of Common Deletions in the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus 2. J Virol. 2020;94(17).
19. Young BE, Fong SW, Chan YH, Mak TM, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet. 2020;396(10251):603-11.