

# Tracking the emergence and spread of COVID-19 using sequence data

A summary of Sam Lycett's presentation at the SARS-CoV-2/COVID-19 workshop.

Phylogenetic analyses of SARS-CoV-2 sequences show the virus is very similar to SARS, but also to a variety of other beta-coronaviruses previously identified in bats. We are able to infer these relationships thanks to viral sequencing data, which are being shared globally on the publicly accessible GISAID database (<https://www.gisaid.org>). Throughout January and February, most sequences deposited were isolated in Asia, but in recent weeks we have seen an increasing number from Europe and North America.

SARS-CoV-2 sequences are very similar to each other, but we do see a few mutations between isolates. Whilst we do not expect these mutations to be affecting virulence, we can use them to trace the spread of the epidemic.

Building a time-scaled phylogenetic tree from all the available sequences, we can estimate the origin of the virus to be November 2019. The same tree indicates that the virus was introduced to the UK multiple times, with clusters spreading from each introduction. We can also see clear clusters in the sequences coming from different continents; for example, the sequences from Europe are all much more similar to each other than the sequences from Oceania.

This data can be overlaid onto a map of the world, which shows that in January and February, the epicentre of viral spread was Asia, into Europe, Australia and North America. However, as time has progressed, the epicentre has moved to Europe instead, spreading to South America and elsewhere.

As we collect more sequence data, we may be able to infer more details about the spread of the virus, such as estimating differences in  $R_0$  between different countries and continents. In turn, this may help to show which intervention strategies have been most effective.

## సీక్వెన్స్ డేటాను ఉపయోగించి COVID-19 యొక్క

### ఆవిర్భావం మరియు వ్యాప్తి ట్రాకింగ్ చేయబడినది

SARS-CoV-2 / COVID-19 సమావేశంలో సామ్ లైసెట్ ప్రదర్శన యొక్క సారాంశం.

SARS-CoV-2 సీక్వెన్సుల యొక్క ఫైలోజెనెటిక్ విశ్లేషణలు వైరస్ SARS కు చాల దగ్గర పోలిక తో ఉంటుంది అని చూపిస్తుంది, ఇది మాత్రమే కాదు గబ్బిలాలలో గతంలో గుర్తించిన అనేక ఇతర బీటా-కరోనావైరస్ తో కూడా దీనికి దగ్గర పోలికలు ఉన్నాయి. వైరల్ సీక్వెన్సింగ్ డేటాకు ధన్యవాదాలు, మేము డేటా తో అర్థం చేసుకున్న విషయాలను బహిరంగంగా ప్రాప్యత చేయగల GISAID డేటాబేస్ (<https://www.gisaid.org>) లో ప్రపంచవ్యాప్తంగా అందుబాటులోకి తెచ్చాము. జనవరి మరియు ఫిబ్రవరి అంతటా, జమ చేసిన చాలా సీక్వెన్సులు ఆసియాలోవి, అయితే ఇటీవలి వారాల్లో యూరప్ మరియు ఉత్తర అమెరికా నుండి పెరుగుతున్న సంఖ్యను చూశాము.

SARS-CoV-2 సీక్వెన్సులు ఒకదానికొకటి చాలా పోలి ఉంటాయి, కాని మేము ఐసోలేట్ల మధ్య కొన్ని ఉత్పరివర్తనాలను చూస్తాము. ఈ ఉత్పరివర్తనలు వైరల్స్ ను ప్రభావితం చేస్తాయని మేము ఊహించనప్పటికీ, అంటువ్యాధి యొక్క వ్యాప్తిని తెలుసుకోవడానికి వాటిని ఉపయోగించవచ్చు.

అందుబాటులో ఉన్న అన్ని సీక్వెన్సుల నుండి సమయ-స్థాయి ఫైలోజెనెటిక్ వృక్షాన్ని నిర్మించడం, వైరస్ యొక్క మూలాన్ని

నవంబర్ 2019 అని మేము అంచనా వేసాము. అదే ఫైలోజెనెటిక్ వృక్షం సూచిస్తుంది - వైరస్ UK కి అనేకసార్లు పరిచయం చేయబడిందని, ప్రతి పరిచయం నుండి వివిధ సమూహాలు వ్యాప్తి చెందుతాయి. వివిధ ఖండాల నుండి వచ్చే సీక్వెన్సులలో స్పష్టమైన జీన్ సమూహాలను కూడా మనం చూడవచ్చు; ఉదాహరణకు, ఐరోపా నుండి వచ్చిన సీక్వెన్సులను పోల్చడగినవి కానీ, ఓషియానియా నుండి వచ్చిన సీక్వెన్సులు వేరుగా ఉంటాయి. ఈ డేటాను ప్రపంచ పటంలో ఉంచినచో, వైరస్ పురోగతి తెలుసుకోవచ్చు. ఇది జనవరి మరియు ఫిబ్రవరిలో, వైరల్ వ్యాప్తికి కేంద్రంగా ఆసియా, యూరప్, ఆస్ట్రేలియా మరియు ఉత్తర అమెరికాలో ఉన్నట్లు చూపిస్తుంది. ఏదేమైనా, సమయం పెరుగుతున్న కొద్దీ, కేంద్రం బదులుగా ఐరోపాకు వెళ్లి, దక్షిణ అమెరికా మరియు ఇతర ప్రాంతాలకు వ్యాపించింది.

మేము మరిన్ని సీక్వెన్స్ డేటాను సేకరించేకొద్దీ, వివిధ దేశాలు మరియు ఖండాల మధ్య R0 లో తేడాలను అంచనా వేయడం వంటి వైరస్ వ్యాప్తి గురించి మరిన్ని వివరాలను అర్థం చేసుకోగలం. క్రమంగా, ఏ ప్రణాళికలు అత్యంత ప్రభావవంతంగా ఉన్నాయో చూపించడానికి ఇది సహాయపడవచ్చు.