

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.

अनुक्रम डेटा का उपयोग करके COVID-19 के उद्भव और प्रसार पर नज़र रखना

SARS-CoV-2 / COVID-19 कार्यशाला में सैम लाइकेट की प्रस्तुति का सारांश।

SARS-CoV-2 सीक्वेंस के Phylogenetic विश्लेषण से पता चलता है कि यह वायरस SARS से काफी मिलता-जुलता है। इसके आलावा पहले से चमगादड़ में पहचाने जाने वाले अन्य बीटा-कोरोनाविरस की भी एक किस्म है। हम वायरल अनुक्रमण डेटा के कारण इन रिश्तों का अनुमान लगाने में सक्षम हैं, जिन्हें सार्वजनिक रूप से सुलभ GISAID डेटाबेस (<https://www.gisaid.org>) पर वैश्विक रूप से साझा किया जा रहा है। जनवरी और फरवरी के दौरान अधिकांश अनुक्रमों को एशिया में अलग जमा किया गए था, लेकिन हाल के हफ्तों में हमने यूरोप और उत्तरी अमेरिका अनुक्रमों की संख्या बढ़ती देखी गई है।

SARS-CoV-2 सीक्वेंस एक-दूसरे के समान हैं, लेकिन हम आइसोलेट्स के बीच कुछ परिवर्तन देखते हैं। अभी हम इन उत्परिवर्तन को विरुलेन्स होने की उम्मीद नहीं करते हैं, पर हम महामारी के प्रसार का पता लगाने के लिए उनका उपयोग कर सकते हैं।

सभी उपलब्ध अनुक्रमों से एक टाइम-स्केलेड फ़ाइलोजेनेटिक ट्री का निर्माण करके, हम वायरस की उत्पत्ति का अनुमान नवंबर 2019 लगा सकते हैं। वही पेड़ इंगित करता है कि वायरस यूके में कई बार प्रवेश हुआ था, और प्रत्येक प्रवेश से विभिन्न क्लस्टर बने। हम विभिन्न महाद्वीपों से आने वाले अनुक्रमों में स्पष्ट क्लस्टर भी देख सकते हैं; उदाहरण के लिए, यूरोप से अनुक्रम एक दूसरे के समान हैं, लेकिन ओशिनिया के दृश्यों की तुलना में अलग हैं।

यह डेटा दुनिया के एक नक्शे पर मढ़ा जा सकता है, जिससे पता चलता है कि जनवरी और फरवरी में, वायरल प्रसार के उपरिकेंद्र एशिया से यूरोप, ऑस्ट्रेलिया और उत्तरी अमेरिका में फैलाव हुआ। हालाँकि, जैसे-जैसे समय आगे बढ़ रहा है, का वायरल प्रसार का केंद्र यूरोप हुआ, और वहां से दक्षिण अमेरिका और अन्य जगहों तक फैल गया है।

जैसे जैसे कि हम अधिक अनुक्रम डेटा एकत्र करते हैं, हम वायरस के प्रसार के बारे में अधिक जानकारी का अनुमान लगाने में सक्षम हो सकते हैं, जैसे कि विभिन्न देशों और महाद्वीपों के बीच R0 में अंतर का अनुमान लगाना। इससे यह भी पता चलता है कि कौन से हस्तक्षेप की रणनीति सबसे प्रभावी रही है।