

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.

Portuguese

Rastreado o aparecimento e a disseminação do COVID-19 usando sequência de dados

Um resumo da apresentação de Sam Lycett no workshop de SARS-CoV-2 / COVID-19.

As análises filogenéticas das sequências de SARS-CoV-2 mostram que o vírus é muito semelhante ao SARS, mas também a uma variedade de outros beta-coronavírus previamente identificados em morcegos. Podemos inferir essas relações com a ajuda dos dados de sequenciamento viral que estão sendo compartilhados globalmente no banco de dados acessível ao público, GISAID (<https://www.gisaid.org>). Durante janeiro e fevereiro, a maioria das sequências depositadas foram isoladas na Ásia, mas nas últimas semanas vimos um número crescente proveniente da Europa e da América do Norte.

As sequências de SARS-CoV-2 são muito semelhantes entre si, mas vemos algumas mutações entre os isolados. Embora não esperemos que essas mutações afetem a virulência, podemos usá-las para rastrear a propagação da epidemia.

Construindo uma árvore filogenética em escala temporal com todas as sequências disponíveis, podemos estimar a origem do vírus em novembro de 2019. A mesma árvore indica que o vírus

foi introduzido no Reino Unido várias vezes com “clusters” espalhados a cada introdução. Nós também podemos ver “clusters” claros nas sequências provenientes de diferentes continentes, como por exemplo, as sequências da Europa que são muito mais semelhantes entre si do que as da Oceania.

Esses dados podem ser sobrepostos em um mapa do mundo que mostra que em janeiro e fevereiro o epicentro da disseminação viral foi a Ásia, dentro da Europa, Austrália e América do Norte. No entanto, com o passar do tempo, o epicentro mudou-se para a Europa, espalhando-se pela América do Sul e outros lugares.

À medida que coletamos mais dados de sequência, podemos inferir mais detalhes sobre a propagação do vírus, bem como estimar diferenças no R_0 entre diferentes países e continentes. Por sua vez, isso pode ajudar a mostrar quais estratégias de intervenção tem sido mais eficaz.