

# Bioinformatics: At the forefront, behind the scenes

September 4 2020, by Sebastian Maurer-Stroh

---



Credit: Freepik

Bioinformatics is an inherently interdisciplinary effort, combining the molecular biology revolution heralded by pioneers like Watson and Crick with advances in computer science that have placed previously unimaginable abilities at our fingertips. Reflecting the collaborative

nature of bioinformatics, my colleagues and I at the A\*STAR Bioinformatics Institute (BII) study everything from sequence analysis to image processing, working with academic, clinical and industry partners alike.

Nowhere has the synergistic power of bioinformatics been made more clear, however, than during the ongoing COVID-19 pandemic, where over 15 million people have been infected and over half a million have sadly lost their lives in the first six months of the pandemic. Amid the uncertainty and upheaval, bioinformaticians have been at the forefront of supporting the development of urgently needed diagnostics and drug repurposing, as well as behind the scenes carefully monitoring the [virus genome](#) for potentially dangerous mutations and tracing the [virus evolution](#) to study and help curb transmission.

## **From sequence to test kit**

We first heard reports of an unusual viral pneumonia spreading in Wuhan in late December 2019. In the field of infectious diseases, this is usually not a cause for concern; experts are always on the lookout for 'the next big one' but most of the time, it turns out to be nothing. This one, however, did progress further and by the second week of January, the world was informed by authorities in China that it was caused by a coronavirus, the same family of viruses that caused severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). We immediately sat up and took notice, and were soon called up by our colleagues from the Global Initiative on Sharing All Influenza Data (GISAID) to take action.

Originally designed to rapidly disseminate information about [influenza viruses](#), GISAID was called upon by affected countries to make available its platform known for its unique sharing mechanism, so countries could share their virus sequences with unprecedented speed. It all started with

five genomes from three Chinese labs.

The realization that we were dealing with something bigger than we had ever faced before didn't happen the very first moment the sequences of what later became known as SARS-CoV-2 were made publicly available. Instead, the realization grew gradually as the cases climbed to the hundreds and then thousands, spreading from the city of Wuhan to other Chinese cities.

From an average of three or four new sequences per day, the number of submissions quickly escalated to hundreds per day and we soon had to build a more robust database to store all 30 kilobase pairs of each sequence, along with important metadata like where the virus was isolated and when the sample was taken. GISAID could rely on programmers and scientists in different parts of the world working literally day and night for weeks and months. To deal with the influx, we also had to adapt our procedures for handling incoming genomes, and quickly develop [computational tools](#) to screen for mistakes and quality so my colleagues at A\*STAR's BII and Genome Institute of Singapore (GIS) developed software tools that allow us to automatically flag genomes for issues such as illegal characters, as well as sort them into categories based on the quality of coverage. Colleagues in France from Institut Pasteur improved checking of metadata and colleagues in Brazil and Argentina also joined our team of up to 50 people in this huge curation effort to be able to respond at every hour of the day by covering all time zones. As of July 2020, the GISAID database has over 60,000 sequences and that number is growing by the day.

While tracking the difference in SARS-CoV-2 sequences can give us important insight into where the virus could possibly have come from and how it is evolving over time, the first point of action everyone had to do to prepare for the virus was to develop accurate and reliable diagnostic kits. My A\*STAR colleague Dr. Masafumi Inoue at the

Experimental Drug Development Center (EDDC) and Dr. Timothy Barkham at Tan Tock Seng Hospital immediately got down to work. Aided by Dr. Sidney Yee, CEO of the Diagnostics Development (DxD) Hub, and the entire A\*STAR ecosystem, we were able to launch the Fortitude quantitative reverse transcription polymerase chain reaction (qRT-PCR) diagnostic kits by the first week of February.

To develop Fortitude, we needed access not just to one genome but reference multiple genomes—from this outbreak relative to previous viruses—to identify a region that is not only unique to the new virus but also relatively stable, so that it is common among all the current outbreak strains. This is where bioinformatics had a profound impact: not only did it enable us to make comparisons rapidly across the thousands of bases in the genome, but it also continues to help us make sense of the constant flow of sequences so that we can ensure that the diagnostic kits will continue to work even as the virus mutates.

Beyond diagnostics, once you have the genome sequenced, you can also use it to predict drug targets and start screening existing drugs in silico, greatly speeding up the search for much-needed therapies.

## **The meaning of mutations**

Apart from addressing the urgent need for diagnostics and drugs and triggering vaccine development on the frontlines, bioinformatics also plays a crucial role in helping us understand how the virus is mutating. First of all, I would like to stress that mutations are normal, particularly for RNA viruses like SARS-CoV-2 which naturally make mistakes when replicating, making imperfect copies of themselves. But just because the virus is mutating does not necessarily mean that it is more dangerous.

Secondly, most mutations are small and are either bad for the virus or have no impact at all. To give you an analogy, if the entire [virus genome](#)

is like a car, the mutated form of the virus would be the same car in the same color with only a tiny difference such as a single letter difference in the license plate. Just as this change in the license plate doesn't affect the performance of the car or make it more fuel-efficient, these mutations do not mean that the virus has become more or less virulent.

However, this "changed license plate" can tell us where the car came from and when it was registered. Similarly, mutations can give us a sense of how the different viral "cars" are related to each other, a piece of information that we can then use in contact tracing.

On very rare occasions, it is possible that there are mutations that do actually change the performance and fitness of the virus and it often requires multiple steps. For example, one such set of mutations in the evolution of SARS-CoV-2 is thought to have given it the ability to jump from animals like bats or pangolins into human hosts.

With real-time genomic surveillance enabled by platforms such as GISAID in combination with modern tools of bioinformatics, we can quickly detect these rare changes when they occur and judge if they affect diagnostics, treatments or increase virulence. For example, the virus that caused the outbreak in Europe had evolved to be slightly different from the original strain, such that it was not so well detected by the RT-PCR kit initially developed by colleagues in China based on the first outbreak genomes. As soon as we saw that change, we notified our colleagues and, armed with that information, they were able to quickly change their protocols and subsequently were able to fully detect the new European strains as well.

As the battle against SARS-CoV-2 continues to be waged across the globe, bioinformaticians around the world are racing against the continually evolving virus. By following its genomic evolution, we hope to catch up with it or sometimes even be one step ahead.

Provided by Agency for Science, Technology and Research (A\*STAR),  
Singapore

Citation: Bioinformatics: At the forefront, behind the scenes (2020, September 4) retrieved 4  
May 2024 from <https://medicalxpress.com/news/2020-09-bioinformatics-forefront-scenes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.