

The web meets genomics: a DNA search engine for microbes

February 4 2019



Credit: CC0 Public Domain

Researchers at EMBL's European Bioinformatics Institute (EMBL-EBI) have combined their knowledge of bacterial genetics and web search algorithms to build a DNA search engine for microbial data. The search engine, described in a paper published in *Nature Biotechnology*, could enable researchers and public health agencies to use genome sequencing data to monitor the spread of antibiotic resistance genes. By making this vast amount of data discoverable, the search engine could also allow

researchers to learn more about bacteria and viruses.

A search engine for microbes

The [search engine](#), called Bitsliced Genomic Signature Index (BIGSI), fulfils a similar purpose to internet [search](#) engines, such as Google.

The amount of sequenced microbial DNA is doubling every two years. Until now, there was no practical way to search this data.

This type of search could prove extremely useful for understanding disease. Take, for example, an outbreak of food poisoning, where the cause is a Salmonella strain containing a drug-resistance plasmid (a 'hitchhiking' DNA element that can spread drug resistance across different bacterial species). For the first time, BIGSI allows researchers to easily spot if and when the plasmid has been seen before.

Google and other search engines use natural language processing to search through billions of websites. They are able to take advantage of the fact that human language is relatively unchanging. By contrast, microbial DNA shows the imprint of billions of years of evolution, so each new microbial genome can contain new 'language' that has never been seen before. The key to making BIGSI work was finding a way to build a search index that could cope with the diversity of microbial DNA.

Monitoring infectious diseases

"We were motivated by the problem of managing infectious diseases and antibiotic resistance," explains Zamin Iqbal, Research Group Leader at EMBL-EBI. "We know that bacteria can become resistant to antibiotics either through mutations or with the help of plasmids. We also know that

we can use mutations in bacterial DNA as a historical record of bacterial ancestry. This allows us to infer, to some extent, how bacteria might spread across a hospital ward, a country or the world. BIGSI helps us study all of these things at massive scale. For the first time, it allows scientists to ask questions such as 'has this outbreak strain been seen before?' or 'has this drug resistance gene spread to a new species?'. "

Quick and easy search

"This search [engine](#) complements other existing tools and offers a solution that can scale to the vast amounts of data we're now generating," explains Phelim Bradley, Bioinformatician at EMBL-EBI. "This means that the search will continue to work as the amount of data keeps growing. In fact, this was one of the biggest challenges we had to overcome. We were able to develop a search engine that can be used by anybody with an internet connection."

"As DNA sequencing becomes cheaper, we will see a whole new host of users outside basic research, and a rapid increase in the volume of data generated," continues Iqbal.

"We will very likely see DNA sequencing used in clinics, or in the field, to diagnose patients and prescribe treatment, but we could also see it used for a range of other things, such as checking what type of meat is in a burger. Making genomics data searchable at this point is essential and it will allow us to learn a huge amount about biology, evolution, the spread of disease, and much more."

Why do we care about microbes?

A microbe is a living thing that is too small to be seen with the naked eye and requires a microscope. 'Microbe' is a general term used to describe

different types of life forms, including bacteria, viruses, fungi, and more.

A small but important fraction of microbes, primarily some specific types of bacteria and virus, are responsible for infectious diseases. When bacteria are able to "survive" antibiotic treatment, they become extremely dangerous to patients. This is happening increasingly around the world and is known as antibiotic resistance.

By comparing the DNA of multiple bacterial species, we can start to understand how they are related and study the dynamics of antibiotic resistance as it spreads—both geographically and across species. For example, DNA analysis can help us predict how dangerous a certain strain of tuberculosis is, and what kinds of drugs that particular strain might respond to.

More information: Ultrafast search of all deposited bacterial and viral genomic data, *Nature Biotechnology* (2019). [DOI: 10.1038/s41587-018-0010-1](https://doi.org/10.1038/s41587-018-0010-1) , www.nature.com/articles/s41587-018-0010-1

Provided by European Molecular Biology Laboratory

Citation: The web meets genomics: a DNA search engine for microbes (2019, February 4) retrieved 20 April 2024 from <https://phys.org/news/2019-02-web-genomics-dna-microbes.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.