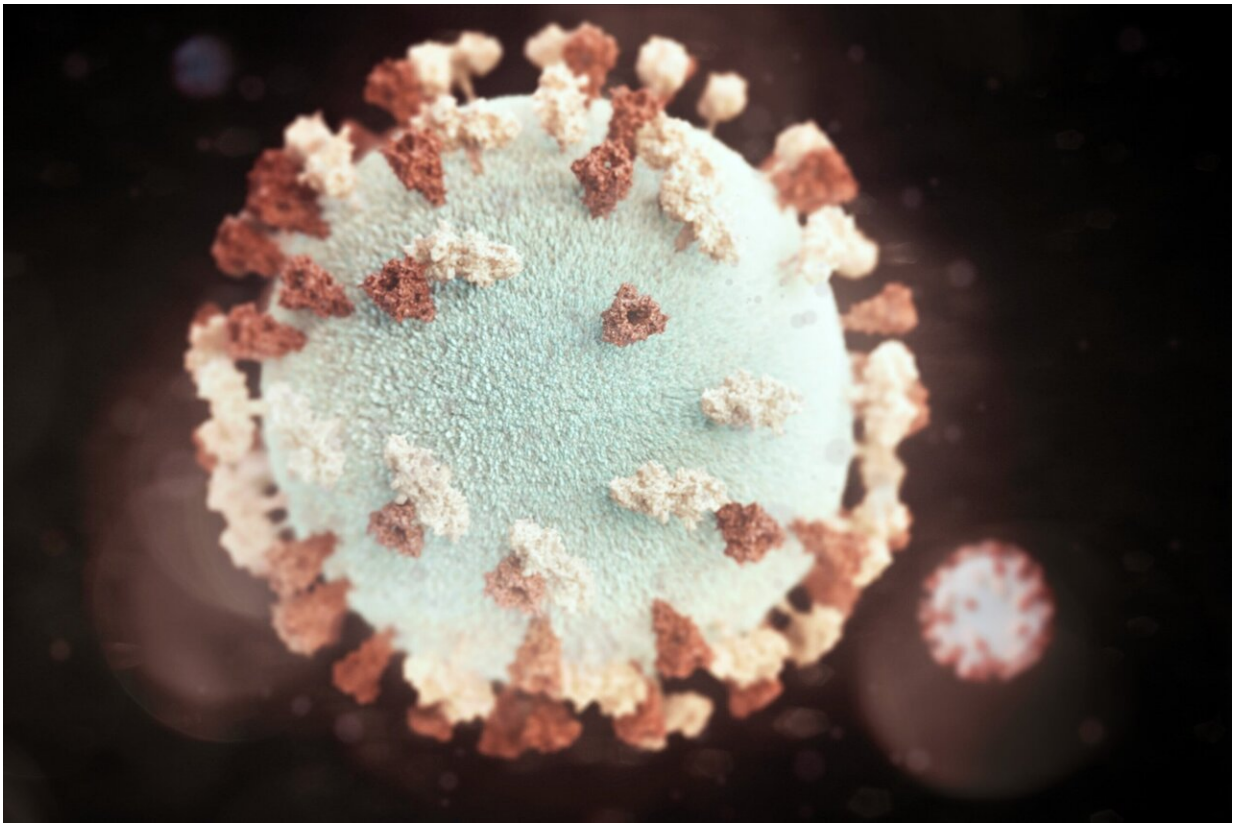


An automated system for generating a million new names for bacteria

December 7 2020



Credit: Unsplash/CC0 Public Domain

Researchers are discovering thousands of new bacterial species that live with us, on us and even inside us. Our relationship with them affects our health and that of the planet—but to define that relationship, new species

need new names.

Scientists at the Quadram Institute have developed an [automated system](#) for generating well-formed names based on the traditional Linnaean system, providing a pool of over a million names— both for newly discovered bacteria and those we've yet to find.

Bacteria have colonized almost every environment on Earth, from the upper atmosphere to the deep subsurface. Estimates vary for the number of [bacterial species](#) on our planet—from millions to billions. We know that we humans alone carry thousand of bacterial species in our guts.

In the 18th century, the Swedish naturalist Linnaeus came up with a system for naming living things that included a two-word Latin name for each species. So, we have the name *Homo sapiens* for humans, along with around a million other such names for animals and plants.

There are also some familiar two-word Latin names for bacterial species, such as *Escherichia coli* (often abbreviated to *E. coli*), which is named after its discoverer, the German microbiologist Theodor Escherich, and after the colon where the organism lives. But, despite there being millions of bacterial species out there, so far only around 20,000 have been given Latin names.

One reason for a lack of new names is that huge numbers of new species are being discovered through DNA analysis rather than grown in the lab, which sits uneasily with the current rules for naming bacterial species, while also creating a huge volume of work. Additional problems are that most bacteriologists don't have enough knowledge of classical languages like Latin and Greek to come up with well-formed names and even the experts struggle to come up with new names quickly enough to cope with the deluge of discoveries.

Professor Mark Pallen, working at the Quadram Institute on the Norwich Research Park, has come up with a solution to the problem, using an automated approach to combine a small set of Latin and Greek roots to create over a million new names. The input files for this process have been carefully checked over by an expert in bacterial nomenclature, Professor Aharon Oren, at the Hebrew University of Jerusalem in Israel, while a Quadram Institute bioinformatician, Dr. Andrea Telatin has written a computer program to automate the process.

Professor Mark Pallen says, "We have shown how combinatorial use of Greek and Latin roots can be used to create more than a million well-formed taxonomic names for bacteria. This probably represents the largest creation of names in the history of science. We expect that our approach could be broadened to cover the need for well-formed names across the whole tree of life."

Pallen and his colleagues are keen to point out that they have not actually named any bacterial [species](#). Instead, they have created a digital warehouse of grammatically correct names that can be used off the shelf as needed by their fellow bacteriologists. Included within their approach are names for bacteria associated with particular animals or particular systems—for example, a name like *Equimonas* could be applied to a new bacterium isolated from horses while *Intestinimicrobium* could be applied to a new bacterium isolated from the intestine. They also included a set of names honoring fellow scientists, for example *Darwiniimicrobium* or *Darwiniimonas* after the British naturalist Charles Darwin.

Bioinformatician Andrea Telatin says, "This is just the start of a process. We predict that, one day, naming bacteria might be as easy as performing a search on Google. All of our computer code is available for anyone to use."

Nomenclature expert Aharon Oren says: "We chose the name Great Automatic Nomenclator for our program after one of my favorite short stories by Roald Dahl. This abbreviates to Gan, which pleasingly means "garden" in Hebrew, which speaks to the rich fertility of our approach. It is thrilling to see bacterial nomenclature joining the world of big data."

More information: Mark J. Pallen et al. The Next Million Names for Archaea and Bacteria, *Trends in Microbiology* (2020). [DOI: 10.1016/j.tim.2020.10.009](https://doi.org/10.1016/j.tim.2020.10.009)

Provided by Quadram Institute

Citation: An automated system for generating a million new names for bacteria (2020, December 7) retrieved 25 April 2024 from <https://phys.org/news/2020-12-automated-million-bacteria.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.