Researchers develop tool to quickly and accurately identify mobile genetic elements like plasmids and virus

September 22 2023

geNomad is a tool that identifies and classifies MGEs based upon their gene content and their genetic sequences. Credit: Kent Leech for Berkeley Lab

Mobile genetic elements (MGEs) are genetic entities that seek to

replicate themselves and spread from cell to cell. Two of the most common forms of MGEs are viruses and plasmids. They can be found in virtually all of Earth's ecosystems.

A software tool described in *Nature Biotechnology* called geNomad identifies and classifies MGEs based upon their gene content and their genetic sequences.

The software was created by researchers under the direction of Microbiome Data Science Group Lead Nikos Kyrpides at the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility at Lawrence Berkeley National Laboratory. geNomad is now accessible through the National Microbiome Data Collaborative's EDGE platform.

In order to gain more insights into Earth's ecosystems, researchers must examine the interactions that happen between tiny organisms (microbes) located in soils and water.

MGEs like viruses and plasmids drive microbial processes and evolution. This is because MGEs can affect a microbe's ability to cycle nutrients or produce new chemicals. They can also affect ecosystems by killing other cells.

Over time, MGEs can help microbes gain a competitive edge within an ecosystem. They influence the composition of the biogeochemical cycles around them. By understanding more about the genomics of MGEs and their evolution at the cellular level, scientists can better understand ecosystem-wide processes.

geNomad is an annotation and classification framework that combines and builds on two standard techniques for identifying viruses and plasmids. Until now, most other tools have focused on identifying only

specific plasmids or viruses. geNomad combines a broad scope, targeting all known groups of viruses and plasmids. It is also optimized for speed, i.e., it can identify millions of new viruses and plasmids quickly, even in massive datasets.

geNomad employs two distinct approaches to identify both viruses and plasmids; one is marker gene-based, and the second is a neural network approach.

The tool was employed to build version 4 of the Joint Genome Institute's [IMG Virus Resource (IMG/VR)](#), now with more than 15 million viral genomes. It was also key to developing the first version of the [IMG Plasmid Resource (IMG/PR)](#), which currently has more than 700,000 plasmids from genomes, metagenomes and metatranscriptomes.

To lower barriers to accessing and using geNomad, the JGI partnered with the NMDC to integrate this tool into the NMDC EDGE platform. NMDC EDGE's easy to use interface allows beginners and experienced users to import their data and access the geNomad tool to conduct their research.

Designed to reduce the effects of taxonomic representation biases during marker selection, geNomad identifies plasmids and viruses from underrepresented groups more accurately.

Additionally, because it can process large datasets, geNomad is poised to become an essential tool for researching global viral diversity. geNomad has already been downloaded thousands of times, receiving excellent feedback from the general research community. It can be [downloaded through NERSC](#).

More information: Antonio Pedro Camargo et al, Identification of mobile genetic elements with geNomad, *Nature Biotechnology* (2023).

Provided by DOE/Joint Genome Institute

Citation: Researchers develop tool to quickly and accurately identify mobile genetic elements like plasmids and virus (2023, September 22) retrieved 27 April 2024 from https://phys.org/news/2023-09-tool-quickly-accurately-mobile-genetic.html