

Assessment shows metagenomics software has much room for improvement

October 17 2017

A recent critical assessment of software tools represents a key step toward taming the "Wild West" nature of the burgeoning field of metagenomics, said an Oregon State University mathematical biologist who took part in the research.

Metagenomics refers to the science of genetically studying whole communities of microorganisms, as opposed to sequencing single species grown in culture.

"Microbes are ridiculously important to life," said David Koslicki, assistant professor of mathematics in the OSU College of Science. "They not only can cause terrible things to happen, like blight and disease, but in general, overwhelmingly, microbes are our friends. Without them doing their jobs, crops couldn't grow as well, it would be hard to digest our food, we might not get sleepy at appropriate times. Microbes are so fundamental to life, to health, we really need to know as much as we can about them."

Koslicki, a leader in a university-wide research and education program known as OMBI - the OSU Microbiome Initiative - described the findings, published recently in *Nature Methods*, as "sobering."

"There are not a lot of well-established, well-characterized computational techniques and tools that biologists can use," he said. "And the assessment showed that a lot of the tools being used do not do nearly as well as had been initially thought, so there's definitely room for

improvement there.

"That said, depending on the situation that a biologist is interested in, there are definitely different tools that have proven to be the best so far."

Metagenomics is a relatively new field that developed quickly once next-generation sequencing grew inexpensive enough that looking at entire microbial communities became economically feasible, said Koslicki.

"The typical view of biology is a wet lab and everything like that, but a whole other facet has to do with these high-throughput ways of accessing genetic material," he said. "You end up with a ton of data, and when you end up with a ton of data, you introduce new problem: How do I get the important information out of it? You have to come up with an algorithm that allows biologists to answer the questions they find important: What critters are there, how many are there, what are they doing, are there any viruses? We need to answer those questions and not just answer them quickly but also have some sort of idea how accurate the answer is."

The dizzying array of tools biologists are using to try to answer those questions is "kind of like the Wild West," Koslicki said. "If you want to learn what bacteria are in a sample, there are no less than three or four dozen different tools people have come up with, and in a rather disjointed manner. You have teams of statisticians, mathematicians, biologists, microbiologists, engineers all looking at this from their own perspectives and coming up with their own tools. Then the end-user [biologist](#) comes along and is faced with 40 different tools, and how do they know how good they are at answering the questions they need answered?"

Koslicki's research, known as the CAMI challenge - critical assessment of metagenome interpretation - was aimed at ranking those tools to provide a road map for biologists.

"The challenge engaged the global developer community to benchmark their programs on highly complex and realistic data sets, generated from roughly 700 newly sequenced microorganisms and about 600 novel viruses and plasmids and representing common experimental setups," he said. "This was an independent initiative. Typically when tools are compared, it's attached to the publication of a new method that's compared to other tools that do worse, so the new method looks good. There hasn't been a lot of independent research into which tools actually work, how well they work, what kind of data do they well on, etc."

More information: Alexander Sczyrba et al, Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software, *Nature Methods* (2017). [DOI: 10.1038/nmeth.4458](https://doi.org/10.1038/nmeth.4458)

Provided by Oregon State University

Citation: Assessment shows metagenomics software has much room for improvement (2017, October 17) retrieved 27 April 2024 from <https://medicalxpress.com/news/2017-10-metagenomics-software-room.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.