

A new assembler for decoding genomes of microbial communities developed

October 8 2020



Credit: CC0 Public Domain

Researchers from the Center for Algorithmic Biotechnology at St Petersburg University, as part of a group of Russian and American scientists, have developed the metaFlye assembler. It is designed to assemble DNA samples from microbial communities. With its help, it is possible to solve a wide range of fundamental and applied problems, among which is the control of the process of treating patients and even

the creation of new drugs.

At present, to study the DNA of any living organism, scientists around the world use complex biotechnological instruments—DNA sequencers. These special machines cannot 'read' the genome from start to finish (like people read books). They do it in separate short fragments—reads. Combining reads into longer fragments, and ideally into a single sequence of the original genome, is an extremely complex computational problem. It is like assembling a million-piece puzzle. The problem is complicated by the fact that genomes often contain a large number of identical repetitive sequences, which often exceed the length of reads. It is possible to cope with this challenging problem using specialized software—genome assemblers.

Several dozen different assemblers are being developed in leading bioinformatics laboratories around the world, and they are available to scientists. This diversity is because the algorithms that assemblers are based on need to be adapted to: different types of input data obtained on different types of DNA sequencers; and different organisms. For example, approaches for assembling bacterial genomes may not be suitable at all for assembling the human genome and vice versa. Additionally, the developers of genomic assemblers are constantly striving to improve their solutions so that: their programs run faster and use less memory; and the resulting assemblies are longer and more accurate than those produced by the competing software.

The new metaFlye assembler is designed for assembling metagenomes. These are DNA samples from microbial communities obtained from various environments, such as the deep sea, soil in a park, or human gut. Having received an assembly of such a sample, it is possible to determine what kind of and how many organisms are presented in it. Using additional assembly analysis, it is often possible to find out: what these organisms can feed on; how they interact; and what substances they

synthesize. All this information can be used in the future, for example: to search for new drugs of natural origin; to determine the reasons underlying the extreme soil fertility; when checking the course of treating patients; and in solving many other fundamental and applied problems.

The metaFlye assembler is designed for data obtained using the current state-of-the-art sequencing technology—long-read sequencing. There are already several metagenomic assemblers working with short-read sequencing, or next-generation sequencing (NGS) data generated on Illumina instruments. Among these assemblers there is the metaSPAdes assembler. It was developed at the Center for Algorithmic Biotechnology at St Petersburg University in 2016. There are also software for assembling isolate genomes from long reads. metaFlye makes it possible to take advantage of the new technology for complex metagenomic data. It is the first metagenome assembler specially designed to work with Oxford Nanopore and PacBio technologies.

"The impetus to develop metaFlye was the absence of a specific metagenomic assembler for long-read technology," says Mikhail Rayko, one of the project's authors, a senior research fellow at the Center for Algorithmic Biotechnology at St Petersburg University. "This technology has already changed dramatically the whole modern genomic science. We have learned to obtain much more complete assemblies. For example, with its help, many missing fragments of the human genome have recently been sequenced and localized. The original Flye tool was used for that, and the members of our laboratory also took part in this project. However, such data have just begun to appear for metagenomes, and, of course, special tools are needed for processing it."

Work on metaFlye started about two years ago. It is four years if we count from the creation of its predecessor, the genomic assembler Flye, on the basis of which the new project was implemented.

"In our study, published in the journal *Nature Methods*, we used metaFlye and other assemblers to analyze several simulated (i.e., computer generated, without real DNA sequencing) and real metagenomic samples from the gastrointestinal tract of a human, a cow and a sheep," says Alexey Gurevich, a co-author of the assembler and a senior research fellow at the Center for Algorithmic Biotechnology at St Petersburg University. "A sample of the sheep microbiome is perhaps of principal interest. It was first obtained and studied in this work, while the initial sequencing data for the other two samples were taken from the works of third-party authors. metaFlye made it possible to assemble an order of magnitude more viral genomes and one and a half times more plasmids in this sample than when using the best existing analog programs."

Another intriguing result was that it was possible to assemble in the sample the genomes of not only bacteria and archaea, but also eukaryotes. At the same time, bioinformatics analysis revealed that almost half of eukaryotic genomic fragments belong to representatives of nematodes, or roundworms. This result fully complies with the autopsy report of the animal, which showed signs of parasitic infection.

"The metaFlye assembler is a tool for solving a wide range of tasks. It will be available to all researchers working with such data. Of the specific projects carried out in our laboratory, we use the assembler to study the soil composition in Chernevaya taiga—a unique biocoenosis of Western Siberia with abnormally high fertility," says Alexey Gurevich.

More information: Mikhail Kolmogorov et al, metaFlye: scalable long-read metagenome assembly using repeat graphs, *Nature Methods* (2020). [DOI: 10.1038/s41592-020-00971-x](https://doi.org/10.1038/s41592-020-00971-x)

Provided by St. Petersburg State University

Citation: A new assembler for decoding genomes of microbial communities developed (2020, October 8) retrieved 17 April 2024 from <https://phys.org/news/2020-10-decoding-genomes-microbial.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.