

Scientists use new technology to assemble genome of Zika virus mosquito

March 23 2017



Transmission electron micrograph (TEM) of Zika virus. Credit: Cynthia Goldsmith/Centers for Disease Control and Prevention

A team spanning Baylor College of Medicine, Rice University, Texas Children's Hospital and the Broad Institute of MIT and Harvard has developed a new way to sequence genomes, which can assemble the

genome of an organism, entirely from scratch, dramatically cheaper and faster. While there is much excitement about the so-called "\$1000 genome" in medicine, when a doctor orders the DNA sequence of a patient, the test merely compares fragments of DNA from the patient to a reference genome. The task of generating a reference genome from scratch is an entirely different matter; for instance, the original human genome project took 10 years and cost \$4 billion. The ability to quickly and easily generate a reference genome from scratch would open the door to creating reference genomes for everything from patients to tumors to all species on earth. Today in *Science*, the multi-institutional team reports a method - called 3D genome assembly - that can create a human reference genome, entirely from scratch, for less than \$10,000.

To illustrate the power of 3D [genome](#) assembly, the researchers have assembled the 1.2 billion letter genome of the *Aedes aegypti* mosquito, which carries the Zika virus, producing the first end-to-end assembly of each of its three chromosomes. The new genome will enable scientists to better combat the Zika outbreak by identifying vulnerabilities in the mosquito that the virus uses to spread.

The human genome is a sequence of 6 billion chemical letters, called base-pairs, divided up among 23 pairs of chromosomes. Despite the decline in the cost of DNA sequencing, determining the sequence of each chromosome from scratch, a process called de novo [genome assembly](#), remains extremely expensive because chromosomes can be hundreds of millions of base-pairs long. In contrast, today's inexpensive DNA sequencing technologies produce short reads, or hundred-base-pair-long snippets of DNA sequence, which are designed to be compared to an existing reference genome. Actually generating a reference genome and assembling all those long chromosomes involves combining many different technologies at a cost of hundreds of thousands of dollars. Unfortunately, because human genomes differ from one another, the use of a reference genome generated from one person in the process of

diagnosing a different person can mask the true genetic changes responsible for a patient's condition.

"As physicians, we sometimes encounter patients who we know must carry some sort of genetic change, but we can't figure out what it is," said Dr. Aviva Presser Aiden, a physician-scientist in the Pediatric Global Health Program at Texas Children's Hospital, and a co-author of the new study. "To figure out what's going on, we need technologies that can report a patient's entire genome. But, we also can't afford to spend millions of dollars on every patient's genome."

To tackle the challenge, the team developed a new approach, called 3D assembly, which determines the sequence of each chromosome by studying how the chromosomes fold inside the nucleus of a cell.

"Our method is quite different from traditional genome assembly," said Olga Dudchenko, a postdoctoral fellow at the Center for Genome Architecture at Baylor College of Medicine, who led the research.

"Several years ago, our team developed an experimental approach that allows us to determine how the 2-meter-long human genome folds up to fit inside the nucleus of a human cell. In this new study, we show that, just as these folding maps trace the contour of the genome as it folds inside the nucleus, they can also guide us through the sequence itself."

By carefully tracing the genome as it folds, the team found that they could stitch together hundreds of millions of short DNA reads into the sequences of entire chromosomes. Since the method only uses short reads, it dramatically reduces the cost of de novo genome assembly, which is likely to accelerate the use of de novo genomes in the clinic.

"Sequencing a patient's genome from scratch using 3D assembly is so inexpensive that it's comparable in cost to an MRI," said Dudchenko, who also is a fellow at Rice University's Center for Theoretical Biological Physics. "Generating a de novo genome for a sick patient has

become realistic."



A multi-institutional team has sequenced the genome of the mosquito that transmits the Zika virus using a new way to sequence the genome of an organism entirely from scratch, dramatically cheaper and faster. Credit: Robert Lang

Unlike the genetic tests used in the clinic today, de novo assembly of a patient genome does not rely on the reference genome produced by the Human Genome Project. "Our new method doesn't depend on previous knowledge about the individual or the species that is being sequenced," Dudchenko said. "It's like being able to perform a [human genome project](#) on whoever you want, whenever you want."

"Or whatever you want," said Dr. Erez Lieberman Aiden, director of the Center for Genome Architecture at Baylor and corresponding author on the new work. "Because the genome is generated from scratch, 3D assembly can be applied to a wide array of species, from grizzly bears to tomato plants. And it is pretty easy. A motivated high school student with access to a nearby biology lab can assemble a reference-quality

genome of an actual species, like a butterfly, for the cost of a science fair project."

The effort took on added urgency with the outbreak of Zika virus, which is carried by the *Aedes aegypti* mosquito. Researchers hoped to use the mosquito's genome to identify a strategy to combat the disease, but the *Aedes* genome had not been well characterized, and its chromosomes are much longer than those of humans.

"We had been discussing these ideas for years - writing a chunk of code here, doing a proof-of-principle assembly there," said Lieberman Aiden, also assistant professor of molecular and human genetics at Baylor, computer science at Rice and a senior investigator at the Center for Theoretical Biological Physics. "So we had assembly data for *Aedes aegypti* just sitting on our computers. Suddenly, there's an outbreak of Zika virus, and the genomics community was galvanized to get going on *Aedes*. That was a turning point."

"With the Zika outbreak, we knew that we needed to do everything in our power to share the *Aedes* genome assembly, and our methods, as soon as possible," Dudchenko said. "This de novo genome assembly is just a first step in the battle against Zika, but it's one that can help inform the community's broader effort."

The team also assembled the genome of the *Culex quinquefasciatus* mosquito, the principal vector for West Nile virus. "*Culex* is another important genome to have, since it is responsible for transmitting so many diseases," said Lieberman Aiden. "Still, trying to guess what genome is going to be critical ahead of time is not a good plan. Instead, we need to be able to respond quickly to unexpected events. Whether it is a patient with a medical emergency or the outbreak of an epidemic, these methods will allow us to assemble de novo genomes in days, instead of years."

More information: "De novo assembly of *Aedes aegypti* using Hi-C yields chromosome-length scaffolds," *Science*, [science.sciencemag.org/lookup/... 1126/science.aal3327](https://science.sciencemag.org/lookup/.../1126/science.aal3327)

Provided by Baylor College of Medicine

Citation: Scientists use new technology to assemble genome of Zika virus mosquito (2017, March 23) retrieved 19 April 2024 from <https://phys.org/news/2017-03-scientists-technology-genome-zika-virus.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.