

Cost-effective method accurately orders DNA sequencing along entire chromosomes

November 7 2013

A new computational method has been shown to quickly assign, order and orient DNA sequencing information along entire chromosomes. The method may help overcome a major obstacle that has delayed progress in designing rapid, low-cost—but still accurate—ways to assemble genomes from scratch. Data gleaned through this new method can also validate certain types of chromosomal abnormalities in cancer, research findings indicate.

The advance was reported in *Nature Biotechnology* by several University of Washington scientists led by Dr. Jay Shendure, associate professor of [genome](#) sciences.

Existing technologies can quickly produce billions of "short reads" of segments of DNA at very low cost. Various approaches are currently used to put the pieces together to see how DNA segments line up to form larger stretches of the genetic code.

However, current methods produce a highly fragmented genome assembly, lacking long-range information about what sequences are near what other sequences, making further biological analysis difficult.

"Genome science has remained remarkably distant from routinely assembling genomes to the standards set by the Human Genome Project," said the researchers. They noted that the Human Genome Project tapped into many different techniques to achieve its end result. Many of these are too expensive, technically difficult, and impractical

for large-scale initiatives such as the Genome 10K Project, which aims to sequence and assemble the genomes of 10,000 vertebrate species.

Members of the Shendure lab that developed what they hope will be a more scalable strategy were Joshua N. Burton, Andrew Adey, Rupali P. Patwardhan, Ruolan Qiu, and Jacob O. Kitzman.

To more completely assemble genomes, they tapped into a technology called Hi-C, which measures the three-dimensional architecture and physical territories of chromosomes within the nuclei of cells. Hi-C maps the physical interactions between regions of the chromosomes in a genome, including contact within a chromosome and with other chromosomes. The results indicate which regions tend to occur near each other within three-dimensional space in a cell's nucleus.

The researchers speculated that this interaction data, because it offers clues about the position of and distances between various regions of the chromosome, might reveal how DNA sequences are grouped and lined up along entire chromosomes. They wondered if the interaction data could show them which regions of the genome are near each other on each chromosome.

Their investigation of this possibility led them to create what they named LACHESIS (an acronym for "ligating adjacent chromatin enables scaffolding in situ"). The map of physical interactions generated by Hi-C was interpreted by the LACHESIS computational program to assign, order and orient genomic sequences into their correct position along chromosomes, including DNA positioned close to the centromere, the "pinch waist" gap in the chromosome shape.

The researchers combined their new approach with other cheap and widely used sequencing methods to generate chromosome-scale assemblies of the human, mouse and fruit fly genomes. The researchers

were able to cluster nearly all scaffolds—collections of short DNA segments whose position relative to each other is unknown—into groups that corresponded to individual [chromosomes](#).

They then ordered and oriented the scaffolds assigned to each chromosome group, and validated their results by comparing them to the high-quality reference genomes for these species that were generated by the Human Genome Project. In the case of human genomes, they achieved 98 percent accuracy in assigning tens of thousands of sequences of contiguous DNA to chromosome groups and 99 percent accuracy in ordering and orienting these sequences within chromosome groups.

"We think the method may fundamentally change how we approach the assembly of new genomes with next-generation sequencing technologies," noted Shendure.

While he and his team cite many areas in which the computational and experimental methods can be improved, the approach is an important step in his lab's long-term goal to facilitate the assembly, for a variety of species, of low-cost, high-quality genomes that meet the rigorous standards set by the Human Genome Project.

Provided by University of Washington

Citation: Cost-effective method accurately orders DNA sequencing along entire chromosomes (2013, November 7) retrieved 27 April 2024 from <https://phys.org/news/2013-11-cost-effective-method-accurately-dna-sequencing.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.