

Researchers weigh methods to more accurately measure genome sequencing

February 3 2012

Lost in the euphoria of the 2003 announcement that the human genome had been sequenced was a fundamental question: how can we be sure that an individual's genome has been read correctly?

While the first full, individual genome was sequenced a decade ago, given the vast [genetic variation](#) across the world's seven billion people, not to mention the differences in makeup even among close relatives, the question of accurate sequencing for individuals has continued to vex researchers.

With companies now projecting they can sequence a genome for a \$1,000, down from \$25,000 just a few years ago, and efforts to develop "personalized" medicines, this matter is taking on increased significance in today's marketplace. These cheaper endeavors rely on newer technologies, which assume that scientists can continue to use the standard shotgun approach of randomly chopping down the genome into smaller pieces and then reassembling them algorithmically. Specifically, today's lower cost is achieved by breaking the DNA into even tinier pieces and rapidly and cheaply reading a massive amount of them. But it is not clear how to assess the accuracy of the newer assembly algorithms and the basic shotgun approach, especially if the accuracy of the earlier [genomic data](#) is questionable.

Among the particular challenges in confirming the accuracy of the sequencing of an individual's genome is matching a person's phenotype, or physical trait, with his or her [genotype](#), or [genetic makeup](#). This has

served, in particular, as a barrier to successful development of personalized medicines, which were predicted shortly after the first sequencing of the [human genome](#), but have yet to truly materialize.

In an article in the journal *PLoS One*, researchers at New York University's Courant Institute of [Mathematical Sciences](#) evaluate some current methods to sequence individual genomes—a study that serves as a "stress test" of the efficacy of these practices.

The researchers employed testing procedures that aim to identify key, or representative, features of the genome as well as how each of these features is related to others.

"Most current technologies, when assembling a genome, make several kinds of mistakes when they encounter a repeated region—where a substring of the letters that make up [DNA](#) strands re-occur in many locations in the genome," explained Bud Mishra, a professor of computer science and mathematics and the study's senior author. "The input random reads tend to collect in one such location, and also show much higher discrepancies among themselves."

To test the viability of these procedures, the NYU researchers relied on a collection of features from an open-source software, AMOS, developed by a public consortium of genomicists and bioinformaticists. If a method has accurately sequenced an individual's entire genome, the researchers hypothesized, then the components of that method's creation should "fit together," and will be consistent with other auxiliary data like "mate pairs," "optical maps," or "strobed sequences," all of which constitute long-range information from the genome. Currently, the use of mate pairs is quite common in sequence assembly and validation algorithms, but not the other two.

While they found shortcomings in all examined methods for sequencing

an individual's genome, some assemblers showed promise. The NYU researchers' conclusions were derived from a procedure called Feature-Response Curve (FRCurve), which effectively shows a global picture of how different assemblers are able to deal with different regions and different structures in a large complex genome. In this way, it also points out how an assembler might have traded off one kind of quality measure at the expense of another kind. For instance, it shows how aggressively a genome assembler might have tried to pull together a group of genes into a contiguous piece of the genome, while incorrectly rearranging their correct order and copy-numbers.

"Such errors have important consequences, especially if the technology is being used to study the genome of a tumor, which often can be highly heterogeneous, making each tumor cell's [genome](#) rearranged and mutated very differently from its neighbors'," explained Mishra.

Provided by New York University

Citation: Researchers weigh methods to more accurately measure genome sequencing (2012, February 3) retrieved 12 May 2024 from <https://medicalxpress.com/news/2012-02-methods-accurately-genome-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.
