

Solving DNA puzzles is overwhelming computer systems, researchers warn

July 15 2013

(Phys.org) —Imagine millions of jigsaw puzzle pieces scattered across a football field, with too few people and too little time available to assemble the picture.

Scientists in the new but fast-growing field of computational genomics are facing a similar dilemma. In recent decades, these researchers have begun to assemble the chemical blueprints of the DNA found in humans, animals, plants and microbes, unlocking a door that will likely lead to better healthcare and greatly expanded life-[science knowledge](#). But a major obstacle now threatens the speedy movement of DNA's secrets into research labs, two scholars in the field are warning.

This logjam has occurred, the researchers say, because the flood of unassembled [genetic data](#) is being produced much faster than current computers can turn it into useful information. That's the premise of a new article, co-written by a Johns Hopkins bioinformatics expert and published in the July 2013 issue of *IEEE Spectrum*. The piece, titled "DNA and the Data Deluge," was co-authored by Michael C. Schatz, an assistant professor of quantitative biology at Cold Spring Harbor Laboratory, in New York state; and Ben Langmead, an assistant professor of computer science in Johns Hopkins' Whiting School of Engineering.

In their article, the authors trace the rapidly increasing speed and declining cost of machines called DNA sequencers, which chop extremely long strands of biochemical components into more

manageable small segments. But, the authors point out, these sequencers do not yield important biological information that researchers "can read like a book."

Instead, the article says, the sequencing machines "generate something like an enormous stack of shredded newspapers, without any organization of the fragments. The stack is far too large to deal with manually, so the problem of sifting through all the fragments is delegated to computer programs."

In other words, the sequencers produce the genetic jigsaw pieces, and a computer is needed to assemble the picture. Therein lies the problem, Schatz and Langmead say: Improvements in computer programs have not kept pace with the enhancements and widespread use of the sequencers that are cranking out huge amounts of data. The result is, the puzzle cannot be pieced together in a timely manner.

"It's a problem that threatens to hold back this revolutionary technology," the authors say in their article. "Computing, not sequencing, is now the slower and more costly aspect of genomics research."

The authors then detail possible computing solutions that could help erase this digital bottleneck. In his own research at Johns Hopkins, co-author Langmead is working on some of these remedies.

"The battle is really taking place on two fronts," he said. "We need algorithms that are more clever at solving these data issues, and we need to harness more computing power."

An algorithm is a recipe or a series of steps—such as searching through data or doing math calculations—that a computer must complete to accomplish a task.

"With cleverer algorithms," Langmead said, "you can do more steps with a fixed amount of computing power and time—and get more work done."

The Johns Hopkins researcher has also had extensive experience in the second digital battle zone: assembling more computing power. This can be accomplished by putting multiple computers to work on assembling the DNA jigsaw puzzle. The linked machines can be at a single location or at multiple sites connected over the Internet through cloud computing. For the latter option, Langmead said, scientists may be able to do their work more quickly by tapping into the huge computing centers run by companies such as Amazon and "renting" time on these systems.

Langmead said he and Schatz wrote the IEEE Spectrum article to call attention to a significant computing problem and to jumpstart efforts to address it. The magazine describes itself as the flagship publication of the IEEE, the world's largest professional technology association.

"We hope the people who read our article can contribute to some solutions and make the work of genomic scientist much easier," he said.

More information: spectrum.ieee.org/biomedical/dna-sequencing/the-dna-data-deluge

Provided by Johns Hopkins University

Citation: Solving DNA puzzles is overwhelming computer systems, researchers warn (2013, July 15) retrieved 26 April 2024 from <https://phys.org/news/2013-07-dna-puzzles-overwhelming.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.