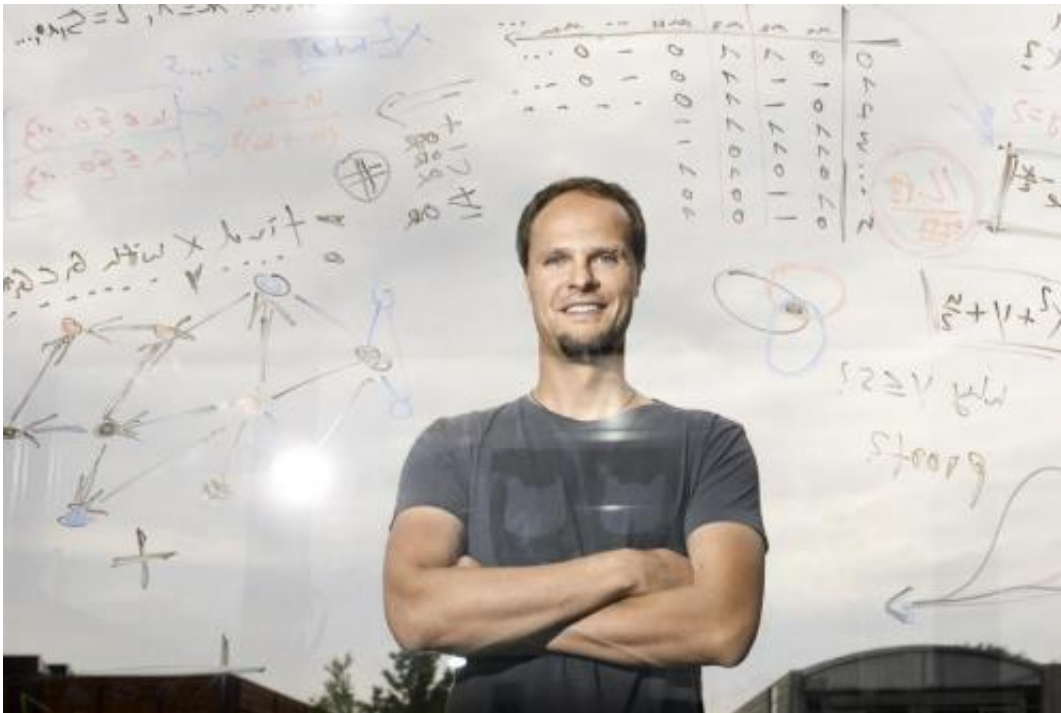


Scientists about sequencing data: We drown in data but thirst for knowledge

June 18 2014



Associate Professor Jan Baumbach, University of Southern Denmark. Credit: University of Southern Denmark

The availability of genome data has revolutionized modern biology and molecular medicine. However, with the costs for genome sequencing dropping by several orders of magnitude down to 200 EUR for a bacterial genome, the amount of species with available whole-genome sequences has exploded over the last years. On the other hand,

information does not equal knowledge, say researchers from University of Southern Denmark, who have analyzed bacteria genome sequences.

While more and more genomic information is becoming available at a drastically increasing pace, the knowledge we can gain about how microorganisms interact with their surrounding, infect hosts and alter their molecular programs in accordance to changing environmental conditions remains widely not deducible from genomic data alone, the researchers from University of Southern Denmark claim. This raises questions regarding the value of newly sequenced species.

The researchers have analyzed the genomes that are available from the past 20 years of sequencing bacterial DNA. They tried to use this data pile to answer a simple questions: Can one distinguish between pathogenic and non-pathogenic bacteria based on their DNA content only?

No valuable knowledge about dangerous bacteria

When they found out that this is not possible in several cases, i.e. you cannot use these data to make such a simple but extremely important distinction, why should we bother collecting even more of this kind of data, the University of Southern Denmark scientists, Associate Professor Jan Baumbach and his doctoral student Eudes Barbosa from the Department of Mathematics and Computer Science at the University of Southern Denmark, now ask in a new study.

Almost 3,000 bacterial species have been sequenced so far. Another 24,000 sequencing projects are presently under way, and there are numerous additional projects on sequencing many more organisms from all kingdoms of life.

"One may ask for the value of all this", the researchers say.

Their research results now show that when it comes to bacteria science cannot count on getting useful information on their pathogenicity from DNA sequencing.

"Should we continue to sequence the DNA of bacteria on such a large scale? Maybe some of the effort and resources could be spent better", say Baumbach and Barbosa.

Proteins provide more valuable knowledge than DNA

Together with colleagues from the Max Planck Institute for Informatics in Germany and the Bioinformatics Department at the Federal University of Minas Gerais in Brazil, the two researchers performed in-depth investigations of 240 whole-genome DNA sequences from actinobacteria, one of the oldest clades on earth. It covers species of high medical relevance, such as *Corynebacterium diphtheriae* (causing diphtheria), *Mycobacterium tuberculosis* (tuberculosis) and *Mycobacterium leprae* (leprae). In average, their genomes have around three million base pairs and five thousand genes.

Since the first sequenced genome of the influenza virus in 1995, researchers have deciphered several thousand of species and ca. 50 million genes. In total, we know about ten thousand [bacterial species](#) and bacteria-like archaea, but it is estimated that there are many more. Conservative bids suggests well above 100 million.

The University of Southern Denmark researchers emphasize that they are not generally opposed to DNA sequencing as a scientific tool at all. One should just be aware of its limited value regarding important follow-

up questions, such as pathogenicity, virulence and infectiousness.

"We drown in data but starve for knowledge", Jan Baumbach says and continues:

"Modern sequencing technologies, so-called next-generation sequencers, are also used to study gene expression – by sequencing so-called mRNA."

This allows for measuring the activity of the genes under a specific condition (after infection, for instance) rather than their mere occurrence, which turns out to be uninformative, at least for bacterial infectivity.

"Such data can be expected to carry more information than the DNA sequence alone, and it can be used to illuminate the interplay of genes, as they do not act in isolation but in an orchestra," the bioinformatics group leader explains.

The important aspects of disease-causing bacteria are found in the genes activity, not in their DNA sequence.

"It's like a plane crash. The color of the plane does not matter. What matters is unraveling the parallel sequence of activities that lead to the accident." says Eudes Barbosa.

Provided by University of Southern Denmark

Citation: Scientists about sequencing data: We drown in data but thirst for knowledge (2014, June 18) retrieved 25 April 2024 from <https://phys.org/news/2014-06-scientists-sequencing-thirst-knowledge.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.