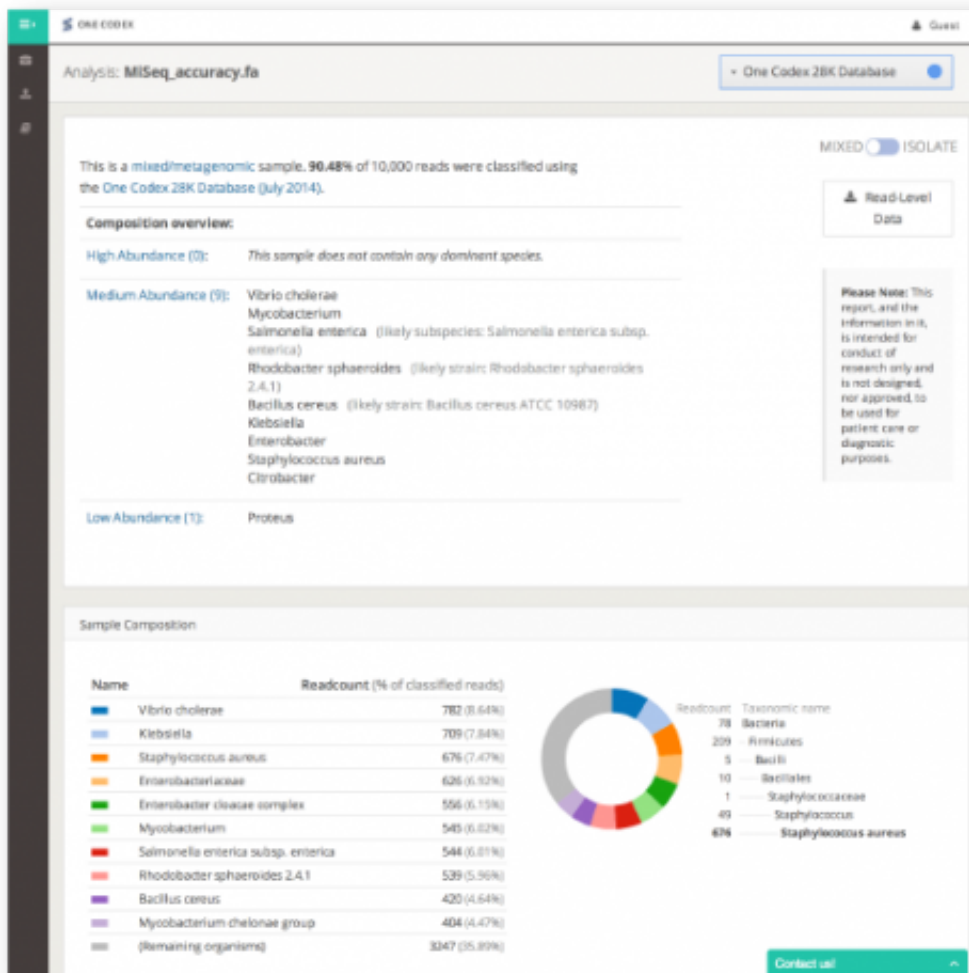# One Codex in open beta for genomic data search

August 17 2014, by Nancy Owano

Data, data everywhere and now as ever researchers need the best tools to make the data useful. In medicine, searching through genomic data can take some time. A startup called One Codex hopes to make difference with their genetic search platform that can process data sets quickly. A report on their work on Friday in TechCrunch noted the advantage of One Codex speed. "Currently," wrote Julian Chokkattu, "the most commonly used tool for genome searching is by using an algorithm called BLAST, Basic Local Alignment Search Tool, which compares primary biological sequence information." For Nick Greenfield, cofounder of One Codex, uploading a file to BLAST took two minutes and 30 seconds to process, compared with the One Codex system where the number was less than 1/20th of a second. The company defines One Codex as a search engine for genomic data. The TechCrunch piece describes what they offer as a service platform for genomics. Apart from using search technology," said Chokkattu, the platform also acts as an indexed, curated reference.

The company said that it can search the world's largest index of bacterial, viral, and fungal genomes. A key advantage is speed. The product can, said the company, "process next-generation datasets in minutes, not days (millions of DNA base pairs per second)."

The two founders are Nick Greenfield, former data scientist, and Nik Krumm, who has a PhD in genome sciences from the University of Washington.

Sample applications would be in clinical diagnostics, food safety and biosecurity. Right now, said TechCrunch, the company is focusing on testing their platform with hospitals and agencies. One Codex is in open beta.

Scientific interest in being able to search genomic data faster has been in evidence for some years. In 2012, MIT's news office reported on a study

in *Nature Biotechnology*, where MIT and Harvard researchers described an algorithm "that drastically reduces the time it takes to find a particular gene sequence in a database of genomes. Moreover, the more genomes it's searching, the greater the [speedup](#) it affords, so its advantages will only compound as more data is generated."

The authors of that paper, titled "Compressive genomics," said, "In the past two decades, genomic sequencing [capabilities](#) have increased exponentially, outstripping advances in computing power. Extracting new insights from the data sets currently being generated will require not only faster computers, but also smarter algorithms." They stated that although compression schemes for BLAST and BLAT that they presented yield an increase in computational speed and in scaling, "they are only a first step."

More information: One Codex: onecodex.com/

© 2014 Phys.org

Citation: One Codex in open beta for genomic data search (2014, August 17) retrieved 25 April 2024 from https://phys.org/news/2014-08-codex-beta-genomic.html