

In a world awash with data, projects proliferate for informaticist

July 16 2014

In a small, nondescript office on the Northern Arizona University campus, those keystrokes Greg Caporaso is fervently tapping into his laptop spell possibility.

They could be a line of code for his interactive informatics textbook, or advice to a graduate student analyzing complex communities of DNA. Maybe he's posting his own microbial ecology research results to share with other scientists, or putting down a few thoughts about a potential link between the human gut microbiome and the severity of autism.

As DNA sequencing churns out one rushing tributary of the exponentially growing river known as "big data," and as high performance computing delivers greater potential for discerning meaning from the torrent, those who combine computer science and biology occupy a world where there is no shortage of territory to explore.

Informatics is no place for hunt and peck.

The field broadly defines the infusion of computation into more familiar disciplines: biology was a forerunner, now joined by everything from geology and astronomy to business, virtual reality and even the arts.

So much information is being generated in all these areas that dealing with it effectively has become a pursuit in its own right. The field is young enough that widely agreed-upon standards are still catching up with developments.

"For more straightforward questions, we've got pretty good protocols in place," said Caporaso, assistant professor of biological sciences at NAU. But especially in the realm of DNA sequencing, even the groundbreaking work that emerged less than a decade ago has already receded into history.

"It's one problem to take a sequence and figure out what organism it belongs to and what its function is," Caporaso said. "It's a very different problem to do that with a hundred million sequences."

Most of Caporaso's published work combines his knowledge of the human microbiome—the trillions of microorganisms living in your gut—with expertise in developing software to interpret DNA sequencing results. The research requires ever more powerful computing capacity, so the recent addition of "monsoon," the [high-performance computing](#) cluster at NAU, opens new research and educational opportunities to Caporaso and his students.

One of Caporaso's graduate students has focused on building models of known DNA sequences to identify bacterial species through unknown sequences. "Her project wouldn't have been possible without monsoon," Caporaso said.

A post-doctoral student is preparing an endeavor so big "it could tie the whole system up for a month or two," Caporaso said, although it won't—multiple researchers on campus share the resource. "Essentially, what she's doing is benchmarking some of the standard approaches for analyzing complex communities of DNA."

Not only are such benchmarks desperately needed, Caporaso said, but the work extends ongoing microbiome research and sets up a major project just getting under way with collaborators at Arizona State University.

"A lot of the studies we do involve going out into the environment somewhere, which might be the soil or a swab of the human mouth, and looking at all the microbes that are present," Caporaso said. "One of the things we're really interested in comparing is the functional potential of the organisms that live in and on your body versus my body."

Some of that potential may be disease. Funded with a tri-university grant from the Arizona Board of Regents, Caporaso's team, along with Rosa Krajmalnik-Brown and James Adams at ASU and Matthew Sullivan at the University of Arizona, will examine associations between the microbiome and the severity of autism.

"There's been some recent evidence, some of it published out of ASU, that microbes living in the gut may be producing chemicals that can affect the severity of the symptoms of autism," Caporaso said. "And there also has been quite a bit of anecdotal evidence that altering that community of microbiomes with certain probiotics can reduce the severity of those symptoms."

The project is still in the planning and discussion stage, leaving Caporaso time to give some additional attention to another regents-funded pursuit, this one directly related to his teaching. He has built an online, interactive textbook and used it in his undergraduate bioinformatics course—cross-listed, of course, in biology and computer science.

Caporaso recreated his slides and notes as an online, interactive tool that includes executable code. Students can learn about bioinformatics methods in the context of their implementation.

"If we're going through the lecture and somebody has a question about how changing a parameter might change the result of an algorithm, I don't have to give a theoretical description because we've all got the code right in front of us," Caporaso said. "So we end up working on it

together."

Caporaso hopes to find the funding to develop the project into a "fully stand-alone, open source, completely free online bioinformatics textbook. It's not necessarily a research project like my others, but it's one of the projects that I'm most excited about right now."

Provided by Northern Arizona University

Citation: In a world awash with data, projects proliferate for informaticist (2014, July 16) retrieved 18 July 2024 from <https://phys.org/news/2014-07-world-awash-proliferate-informaticist.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.