July 2017

Big Data: Computational Biology opens a new window on the world's challenges for Colby scientists

Kate Carlisle

BIG DATA:

By Kate Carlisle

COMPUTATIONAL BIOLOGY OPENS A NEW WINDOW ON THE WORLD'S CHALLENGES FOR COLBY SCIENTISTS

Pick a number, any number.

Then multiply that number by, say, a trillion.

You're getting close; you're in terabyte territory.

Imagine that the resulting product, that huge number, represents tiny bits of data: information about what makes you human, what gives you brown eyes, and why your hair is curly.

Now, take these data, these trillions of facts, this micro-universe of information, and use them to change the world.

Far-fetched? Not anymore, and not at Colby, where students and faculty are studying clues to the world's biggest challenges through the use of computational biology, an imposing blend of data and life science and the College's newest major.

"What makes us 'us' and not a plant? Not a bacteria, or a virus," asks Andrea Tilden, The J. Warren Merrill Associate Professor of Biology and a genomics expert. "Any one genome has six thousand novels worth of information. Computational biology is the tool we use to read them."

Simply put, comp bio (a very short moniker for a very big field) is the study of biological questions through the use of massive data sets, integrating biological, statistical, and computational understanding. Many scientists argue that computation, or the quantitative method, is now absolutely central to biology, imposing order and providing testable concepts on a large scale. They say that someday the "computation" label will disappear, subsumed into the larger label of "biology" as mathematical and statistical tools become as much a part of the science as the agar, Bunsen flame, and microscope were on a 20th-century laboratory bench.

Indeed, the twinning of computing capacity and scientific research is having an effect approaching science fiction, not just on big issues like climate change, but on very personal levels, said Professor and Chair of Computer Science Bruce A. Maxwell. A generation hence, individual health monitors could be as common as the ubiquitous iPhone today. "Imagine if I were standing in my house and a monitor went off and said 'Bruce, your blood sugar is low. Better eat something,'" he says. "That's not far-fetched. That is a thing that could happen, and the technology is based in the type of work that is being done right now."

Maxwell points out, however, that scientists are grappling with big questions right now using computational models that exceed resources available even 10 years ago. He ticks off a long list of applications, including agriculture, social studies, including population and migration, psychology, robotics, and advanced medicine.

SCIENTIST MIKE SASNER: DATA KEY TO ALZHEIMER'S BREAKTHROUGH

For years, researchers have been banging their heads against the wall as they search for a treatment for Alzheimer's disease. Researchers like Mike Sasner '83 hope that's about to change.

Sasner—a research scientist at The Jackson Laboratory (JAX) in Bar Harbor, Maine—is at the forefront of a new wave of Alzheimer's research, which is using fast-evolving scientific technology and data management tools to accelerate the search for the causes of the devastating disease.

JAX is a global leader in development of genetic mouse models for scientific research, and one of the research institutions with which Colby has a growing affiliation. Now, boosted with a $25-million grant and a partnership with Indiana University School of Medicine, JAX is launching a new effort to rapidly develop more—and more precise—models to be used in sorely needed clinical trials.

"We need to understand the brain before we can know how to fix the brain," Sasner said.

He's been mulling this problem at JAX since he finished his post-doctoral fellowship at the National Institutes of Health. With a Ph.D in neuroscience from the University of Connecticut, the Colby biology and biochemistry major (and varsity soccer player) is co-directing the disease-modeling core of the new Alzheimer's Disease Precision Models Center. The center will team with the new Center for Biometric Analysis, a state-of-the-art imaging lab that will analyze and assess the new mouse models. The Bar Harbor behavioral lab will also begin testing potential compounds for treatment of the disease.

"We really have a nucleus here," Sasner said.

It couldn't come soon enough, as Alzheimer's research has repeatedly stalled at the clinical level. Scientists have identified conditions in the brain that coincide with Alzheimer's: tangles of beta-amyloid and tau, two types of invasive protein. Modeling has focused on genetic mutations that are associated with these proteins, but recent research has shown that these genetic variants are relatively rare.

"So the issue is we're testing drugs in models that represent about one percent of the patient population," Sasner said. "Ninety-nine percent of the people [with Alzheimer's] don't have these mutations."

It turns out that most patients have a number of more common genetic variants that singly present little risk, he said, and some combination of these presumably cause Alzheimer's. "This is what our goal in the center is," Sasner said, "to make models to express these variants. And therefore test drugs in those which … mimic the ninety-nine percent of the patient population." That way, he said, "we're much more likely to get a therapy that's going to work in the clinic."

The mission is to get the models and related information out to the research community as soon as possible. Sasner said JAX and partner Indiana University will be working with a bioinformatics company, Sage Bionetworks in Seattle, to do human-to-mouse matching. If a biomarker is increased in a mouse model, does that biomarker show a correlating increase in the patient population? "Do they track the same? Untreated? When we test with compounds?" Sasner said. "The power of this is correlating the human to the mouse. … This has never been done on this scale before."

Sasner says computational biology is key to future scientific breakthroughs. "Science is no longer about generating data," he said. "The data are easy. It's analyzing the data. Understanding the tools to manage that data is really where the world is going."

—Gerry Boyle '78

> " Science is no longer about generating data. It's analyzing the data. Understanding the tools to manage that data is really where the world is going."
>
> —Mike Sasner '83, research scientist, The Jackson Laboratory

> **"Imagine if I were standing in my house and a monitor went off and said 'Bruce, your blood sugar is low. Better eat something.' That is not far-fetched. That is a thing that could happen."**
>
> —Bruce Maxwell,
> professor and chair of computer science

The field of computational biology is not brand-new at Colby, Maxwell points out. After arriving in 2007, he guided the first interdisciplinary computation programs here. The official designation of the major (believed to be rare if not unique among small liberal arts colleges in the United States) was a

WELCOME TO COLBY—AND BIOINFORMATICS

Jay Moore '18 is one of Colby's new comp bio majors, and he's already using technology that was the stuff of researchers' dreams when he was born.

Moore, of Kennebunk, Maine, is one of the early adopters of nanopore gene-sequencing technology brought to campus by The J. Warren Merrill Associate Professor of Biology Andrea Tilden, who has played an instrumental role in providing research opportunities to Colby students, from Colby's partnership with The Jackson Laboratory to helping land the National Science Foundation grant to support a data network among Colby, JAX, and the University of Maine.
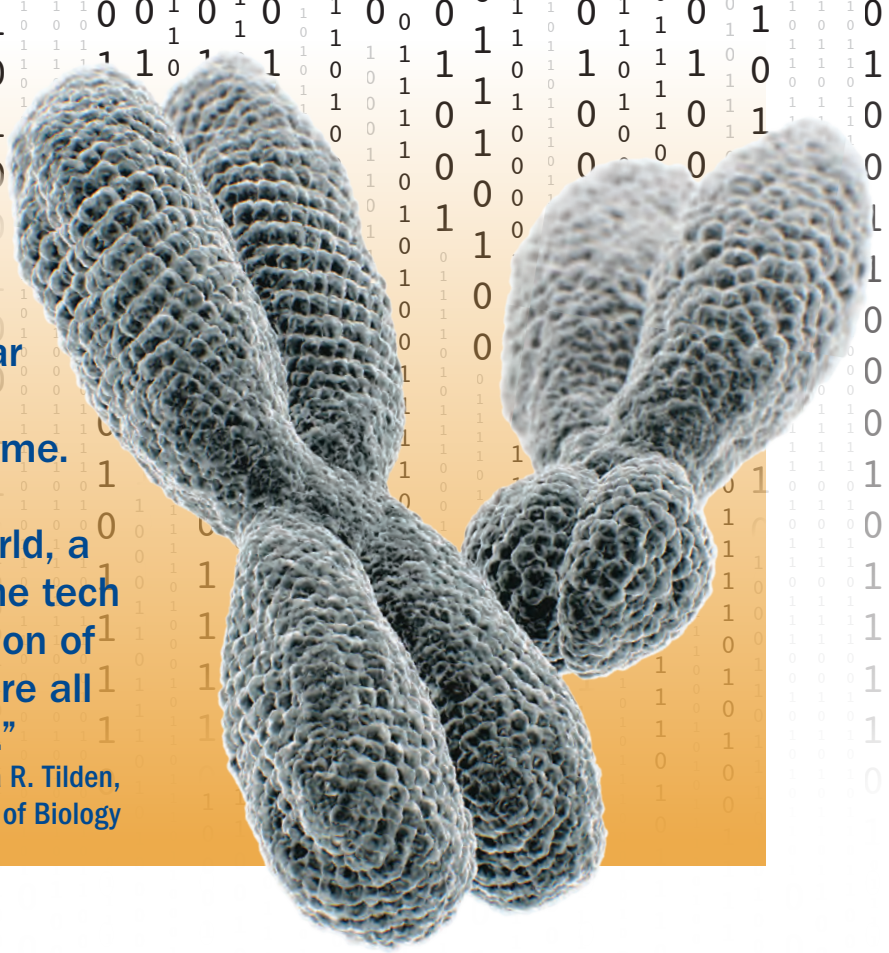
Now Tilden has introduced the MinION—a portable, real-time sequencer for genomic material—to Colby. About the size of a cell phone and linked to cloud computing using a laptop computer, the technology is highly accessible and affordable, allowing Colby to bring both the laboratory and computational aspects of genomics directly to students, Tilden said. Said Moore, "It's pretty amazing what you can do (with bioinformatics). ... Lab work is really about creating data."

natural evolution of this curriculum development, according to Tilden. "It started out in a more raw form, where we were just looking at courses we could put together that could combine this growing interest and need to use computational tools to analyze big, biological data," she said. Tilden, whose work had moved into bioinformatics (using computer tools to look at biological questions), created a ground-breaking Jan Plan genomics course with the education team at Colby partner The Jackson Laboratory (JAX), a biomedical research institution in Bar Harbor, Maine, where more than 1,900 employees search for genomic solutions to human health problems.

Through Tilden's program, Colby students working at the lab were able to wrangle big data in their search for answers to basic genetic questions. (Additionally, Colby has a bioinformatics relationship set up with Mount Desert Island Biological Laboratory, where Tilden will spend the coming year as a visiting scientist in bioinformatics.)

The JAX Jan Plan researchers were in good company. Life scientists and medical researchers around the world have begun to use bioinformatics in their approach to every problem, from population ecology to agriculture to cancer.

The field has exploded at a rapid clip, and when current Colby students were born, computational biology was in its infancy, too. "Very recently, around the year 2000, we had just finished sequencing the human genome," Tilden said. "It had taken 15 years at 20 different labs around the world, a vast project. We now have the tech to accomplish this in a fraction of the time, and while there is still a network of labs and researchers around the world, we can share all this data in virtually no time."

For example, medical researchers studying a child born with a metabolic condition would have once believed it was a unique genetic problem, but they had no easy means of comparison to help them address or prevent the condition. Through

Opposite page: from left, Jackson Laboratory scientists Stacey Rizzo, Gareth Howell, Mike Sasner '83, and Greg Carter.

bioinformatics, "you can sequence that child's genome and compare it to an [international] database to understand what is the one difference," Tilden said. "Then you begin to have a base to find a solution." What does it take in 2017 to get this done? "It's a thousand bucks and one day—we are really close to what we call the thousand-dollar genome."

And recent Colby graduates are already helping move the science in new directions.

Adam Lavertu '16, one of Colby's first comp bio majors, is currently pursuing a Ph.D. in bioinformatics at Stanford University. Lavertu was initially attracted to both biology and computer science at Colby, but envisioned a fairly standard life science path as an academic. "My interests collided," he said, while analyzing the genetic sequence of an algae species. "Our goal was to discover how algae and coral synchronize their cell cycles, a vital part of their symbiotic relationship. If we could increase our understanding of how their partnership works, we might be able to help slow or even reverse the damage climate change has inflicted on the world's coral reefs," Lavertu explained.

He was deeply impressed by how computational methods could guide traditional biological research—and have a tremendous impact on some of the planet's most pressing problems.

This interdisciplinary approach is fast becoming the rule, not the exception.

In fact, computational biology is a great exemplar of Colby's success with interdisciplinary learning, according to Associate Professor of Computer Science Stephanie Taylor, who teaches systems biology, which draws both computer science students and biology scholars. "It's clearly more interdisciplinary than whatever we've been able to do before," she said. "I find when I'm teaching a class like this, I can go further with computer science students who have a bio background. And there's a greater depth of learning for students with background in each, because they can have greater interaction with the material. Even those who don't know certain things yet can learn from their peers."

The problems and projects Taylor assigns are based on her own research, in many cases her work on circadian rhythms (which has real-world impact on problems ranging from jet lag to cancer drug dosing). "The students work on mathematical models of circadian clocks. We ask, how do we write the code that solves those problems numerically?" By comparing models published in scientific literature, Taylor's students learn to evaluate the tools that will someday be used on tough problems. It's not easy, but the challenge of using data in this way sets the students up for future success in this field. "If they were grad students, this might be the first project I would assign them," she said.

For Lavertu, who grew up in Maine's Aroostook County interested in life sciences, it was a class with Taylor that first sparked interest in combining computer science with biology. "After a couple of late nights spent on the projects, I was hooked on coding and determined to build my computational skill set," he said.

But it wasn't his first brush with comp bio as a discipline: as a high school student he took a course offered at JAX that

used computational techniques to examine genetic data. And he points out that the field is not new: it has simply exploded in capacity thanks to technological advances. "In the late 70s and early 80s an entire Ph.D. thesis could be based on a graduate student sequencing a single gene; now I work with data sets that contain genetic information for hundreds of thousands of people," Lavertu said.

Given the importance of big data in biology research, is it time to abandon the "comp bio" designation and just call it "biology?" Some scientists think so. "In all the sciences, advances in technology have made computational work so necessary that yes, it's an integral part of the discipline," Maxwell said.

But Lavertu agrees only up to a point. "It is becoming increasingly difficult for a biologist with no computational background to remain effective in the modern laboratory," he says. "Computational research now has a presence in almost all current day research." That said, he believes in a "need for and value in computational biology remaining distinct within biology. Computational biologists receive a level of specialized training that is comparable to someone who has selected a particular area within biology such as cancer or development."

And the road to that specialized training can now begin at Colby. The program is growing—a distinction for a liberal arts college. "While (the program) is very, very new and undergraduate programs in the field are rare, we are ahead of the game," Tilden said. "We are aware of the magnitude of this."

And the program doesn't stop outside the classroom or Mayflower Hill. Colby's partnerships with JAX, the Bigelow Laboratory for Ocean Sciences, and other research centers offer opportunities for students to put their training to use on challenging problems. Charles Wray, director of courses and conferences at JAX, also worked closely with Colby faculty to launch genomics study at the College. "These partnerships are extraordinary and distinctive," said Tilden, whose class last year worked on a JAX "mouse model" investigating Down syndrome.

A National Science Foundation grant awarded last year to a partnership among Colby, the University of Maine, and JAX will support a dedicated research network connecting Colby faculty and students to resources at UMaine and JAX's large genomic databases. And this is where the "big" in big data really comes in. Colby researchers can quickly gain access to numerous gigabyte- and terabyte-scale data sets. For comparison, the Hubble Space Telescope generates about 10 terabytes of data a year. "It's huge," Maxwell said.

Working with this kind of data, and in concert with top-drawer research partners, affords Colby's comp bio students the type of experience Lavertu (who worked with JAX as an undergraduate) found so valuable. He's now putting that experience to work in his doctoral studies, aiming to guide drug research toward more safe and effective medicines. And that is truly big. Ⓒ

gigabyte
1 billion bytes

kilobyte
1,000 bytes

megabyte
1 million bytes

terabyte
1 trillion bytes

petabyte
1 quadrillion bytes

■ one byte = one letter/character