

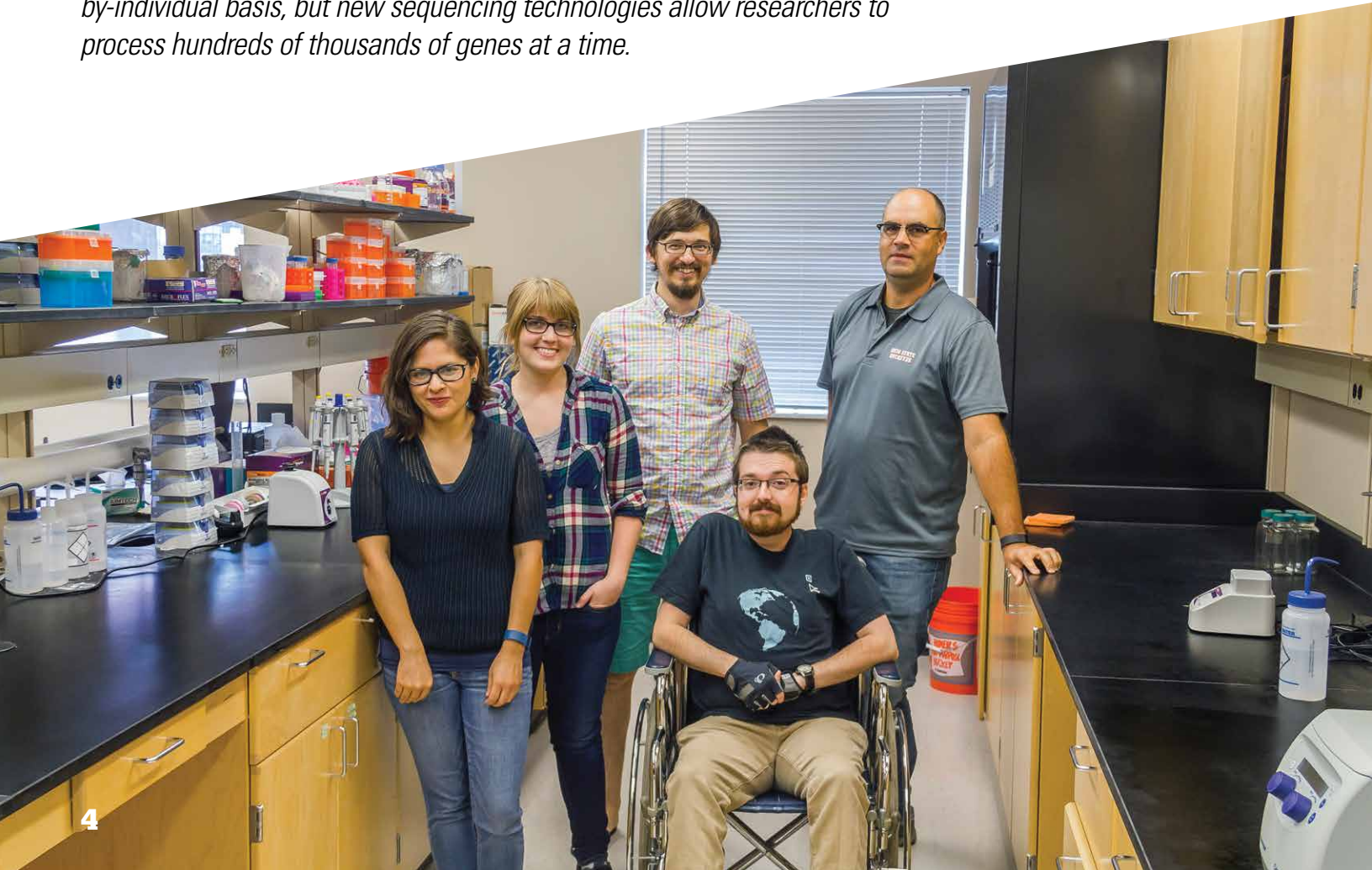
Biological Sciences

On a daily basis, each one of us asks ourselves: Where do we come from? Where are we going? And how can we thrive? The study of modern biology seeks to answer those questions on a broader scale to improve each one of our lives in some significant way. The biosciences represent a vast, eclectic and fascinating field. Research being done at the Ohio Supercomputer Center illustrates the breadth of critical work being done to understand life, including everything from evolutionary questions to focused issues of fighting cancer, improving human hearing and developing plants to boost economies.

Bioinformatics Tools

Carstens gauging how statistical models fit various genetic data

In less than 10 years, the way genetic data has been collected has sped up in a major way. Previously, collecting data from a species was done one gene at a time, on an individual-by-individual basis, but new sequencing technologies allow researchers to process hundreds of thousands of genes at a time.



That massive uptick has created a huge need for bioinformatics tools. Bryan Carstens, Ph.D., and researchers in his lab within the Department of Evolution, Ecology and Organismal Biology at The Ohio State University are working to create these bioinformatics tools through the Ohio Supercomputer Center (OSC).

The Carstens lab relies heavily on statistical models to determine parameters for evolutionary processes, such as mutation, selection, genetic drift, migration and phylogenetic diversification. While there are plenty of software packages available for these processes, there hasn't been a lot of attention dedicated to assessing how well one model fits a given empirical data set.

"The work we do is looking at fitting statistical models to genetic data, and we do this to learn about the evolutionary history of different species," Carstens said. "Sequencing now is massively parallel so it's ideal for a parallel computing environment."

Carstens has several projects with OSC, some of which are packaged into a larger one: Developing bioinformatics tools for evolutionary genetics. Those projects include: Developing P2C2M (Posterior Predictive Checks of Coalescent Models), a package that evaluates the fit of coalescent models; developing Phrapl, which evaluates coalescent model selection; genome assembly for the carnivorous plant *Sarracenia alata*; and using bGMYC, a species delimitation software, to evaluate sequence data from more than 20,000 Malagasy ants. In addition, other lab researchers are working on predicting the future geographic distributions of threatened and endangered species under differing models of climate change.

Carstens' lab is seeking to understand demography of specific species because it allows for more precise inferences about changes in natural selection, how strongly the selection has been and other related questions.

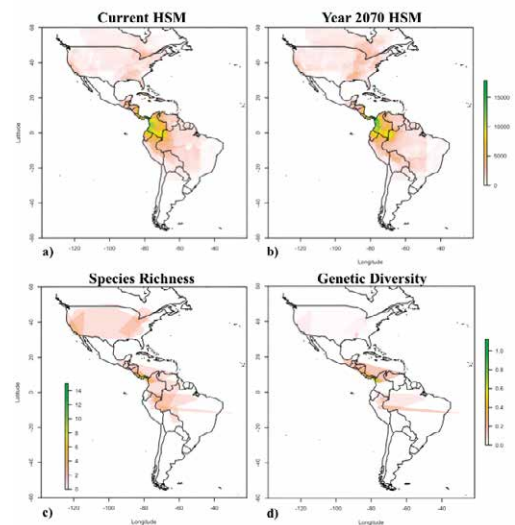
That insight helps gain understanding into many areas such as fighting disease. For example, some stricken by cancer might have a genetic variance that can be pinpointed to develop a treatment that has worked for someone with a similar genotype.

Another example relates to conservation. For instance, why are certain species of birds, like blue jays or cardinals, plentiful while others are rare?

"It's a mistake to assume an abundant species has always been abundant," Carstens said. "If you're a conservation manager and you're trying to make decisions about how to allocate resources, it becomes useful to know some species have undergone a dramatic expansion because it looks like they're able to thrive in human-moderated environments versus another species that has declined since the time of European colonization. This history, you can't really get it any other way."

That makes high performance computing critical to bioinformatics, which may require thousands of computing hours for the statistical analyses necessary to understand a genetic data set.

"What happens at one part of the genome is very likely to impact some other part of the genome," Carstens said. "As geneticists, we need to understand the statistical properties and distributions of our analyses by replicating them over and over. If we do one analysis and don't have replication we often can't interpret it. So what we need to do is thousands of different analyses with slightly different parameters and positions to understand how much confidence we should place in the answers we've gotten. And that's only possible in a parallel computing environment, so it's a huge competitive advantage to house my research program at The Ohio State University because of OSC's resources." •



▶ Left: Bryan Carstens, Ph.D., and researchers in his lab within the Department of Evolution, Ecology and Organismal Biology at The Ohio State University are working to create these bioinformatics tools through the Ohio Supercomputer Center.

▶ Geographic distributions of American amphibians, under current and projected climatic regimes. Carstens' analysis of intraspecific genetic diversity can contribute to the identification of regions that should be targeted for conservation.

Project Lead: Bryan Carstens, Ph.D., The Ohio State University
Research Title: Developing bioinformatics tools for evolutionary genetics
Funding Source: National Science Foundation
Website: carstenslab.osu.edu