



This issue's interview is with Laura Kubatko, Professor of Statistics and Evolution, Ecology and Organismal Biology at Ohio State University. Laura's research interests lie in statistical genetics, particularly relating to the inference of phylogenetic trees. She is the chair for the upcoming SMB 2023 meeting in July.

Who or what inspired you to pursue a career in statistical genetics and mathematical biology?

I actually began my undergraduate studies as a biology major. As a child, I enjoyed spending time outdoors, and I wanted to find a career that would allow me to study the natural world. I was encouraged to take Calc III during my first semester of college after getting AP credit for the first two semesters of calculus, and I subsequently decided to continue taking math classes because I enjoyed them.

After my sophomore year in college, I was selected for an REU program in a biophysics lab. I found the lab work challenging, but I wandered into the math-bio section of the library at the university where I was working and began browsing books on the shelves (yes, I am old enough to love physical books!). Until then, I hadn't realized that mathematical biology was a field that one could study. After my junior year of college, I participated in a research project that involved the collection and analysis of data for an ecological study of a forest. When I realized how much I enjoyed the data analysis aspects of this project, I decided to finish a math major in addition to my biology major and to pursue graduate study in biostatistics.

What is something exciting you are currently working on?

So many things! I love research. But I'd say that the statistical approach I'm most excited about right now is the use of composite likelihood (also called pseudolikelihood) for estimation in problems where it's difficult or impossible to compute the full likelihood. There's a lot of interesting theory being developed in this setting for a range of problems, many of which feature the availability of large-scale data. The approaches seem especially well-suited for problems in phylogenetics for which genome-scale data are available for numerous species.

What is an aspect of your work that you are most proud of?

The work that James Degnan did for his dissertation stands out for sure. James derived the probability distribution for gene tree topologies conditional on the species tree under the coalescent model. James was only my second Ph.D. student, so as a young professor I had some anxiety about what would come out of his project. It turns out that being able to compute the

gene tree distribution has led to numerous interesting research questions, as well as to the development of methodology that is used to analyze empirical data.

Phylogenetics and population genetics is naturally interdisciplinary, existing at the intersection of molecular biology and mathematical modeling. What are some challenges of interdisciplinary work, and what are some of the best aspects of it?

The biggest challenge is without a doubt the terminology. Students that begin working with me almost always comment that they can't even get through a paragraph of a paper in the primary literature without needing to stop and google several terms (and yes, I do remind them that some of us did this before google existed!). But biologists feel the same way ... probabilities, likelihoods, conditioning ... none of these terms are meaningful until you see how they are used.

The best aspect to me is a deeper involvement in science than can be achieved by working exclusively in either field. I have the opportunity to learn exactly how data are generated and what the significance of the biological questions being asked is, as well as to understand how the analysis proceeds and what conclusions are being made (and why). This "start-to-finish" understanding is very important to me.

Sequencing capacity has increased wildly in the past decade, and particularly in a COVID-era. How has this data availability shaped your work?

The change in data availability from the start of my career to the present is truly remarkable. It has allowed for new techniques that were not possible when I was working on my Ph.D. It has also required researchers to have substantial skill in computation, including data processing and data manipulation. I like coding, so this change has suited me well. The current approaches that focus on extracting as much "information" as possible from data are likely to be the ones that lead to rapid advances.

Do you have any advice for those starting out their academic careers?

Be open to changes of direction. I started as a biologist who became interested in statistics. I went to graduate school planning to work on ecological statistics, and ended up studying evolutionary biology instead. I'm glad that I was curious and willing to spend time exploring things that interested me. It's also important to ask questions of colleagues and collaborators, and this part was hard for me since I'm naturally pretty shy. But it's also the thing that has helped me the most, and something that I've gotten better at over time.

What does your perfect weekend look like?

I still love being outdoors and spend my free time outside whenever I can. I've biked from Cincinnati to Cleveland. I spent 10 days hiking the Appalachian Trail last summer. So I'd say any weekend where I'm outside most or all of the time is perfect!