

The 2013 Novitski Prize Jonathan Pritchard



THE Edward Novitski prize has recognized an “extraordinary level of creativity and intellectual ingenuity in the solution of significant problems in genetics research” each year since 2008. It is my great pleasure to report that the 2013 Novitski prize goes to Professor Jonathan Pritchard of the University of Chicago, primarily for developing a model-based clustering approach to analyzing multilocus genotype data for inferring population structure. This approach, implemented in the software package *structure*, is now widely used in a range of fields, including human genetics, forensics, molecular ecology, and conservation genetics. As of the time this was written, the original presentation (published in 2000 in *Genetics*) was cited over 8500 times according to Google Scholar.

Pritchard lived in England until age 14. However, like Novitski (Crow *et al.* 2006), Pritchard also lived part of his early life in Pennsylvania, where Pritchard’s father was a professor of mathematics at Pennsylvania State University. Also like Novitski, who discovered a *Drosophila* mutation while in high school (Novitski and Rifenburgh 1938), Pritchard fostered a love of insects in particular and nature in general at a young age, spending his summer catching beetles and trying to identify them. Pritchard enrolled at Penn State for college and began conducting research with Stephen Schaeffer after his freshman year. He quickly developed a passion for population genetics theory and was particularly influenced by Begun and Aquadro’s (1992) report of a strong association between recombination rate and nucleotide diversity. After finishing degrees in biology and math, he pursued a Ph.D. at Stanford University with Marcus Feldman in theoretical modeling using human genetic data. This was an exciting time in the Feldman lab, as Pritchard had the opportunity to work closely with Carl Bergstrom, Michael Lachmann, Noah Rosenberg, and Lauren Ancel Meyers.

Near the end of his time at Stanford, Pritchard took a statistical genetics class with Neil Risch, who was conducting association mapping studies of complex diseases and developing methods for analysis. Association mapping studies (also called “case-control” studies) of diseases test whether particu-

lar alleles are more common in affected than unaffected individuals. Such studies were growing in popularity alongside new methods for genotyping, but as several researchers had noted, spurious associations could easily result from population subdivision/ admixture (Lander and Schork 1994 present an amusing illustration regarding mapping human ability to eat with chopsticks in San Francisco). Based on that class and emerging interest in the field, Pritchard began to examine methods for determining whether population subdivision exists in a sample that would cause problems in association mapping (Pritchard and Rosenberg 1999). The clear next step was to control for population subdivision in such studies.

Pritchard took this interest with him to his postdoctoral appointment at Oxford with Peter Donnelly, also working with Matthew Stephens. A unified approach was needed to detect the existence of population subdivision in a sample and then use a model-based approach to factor out the subdivision by assigning individuals to one or more populations probabilistically. To this end, the three of them developed the original framework for the software package *structure* (Pritchard *et al.* 2000). The framework was broadly flexible to accepting different kinds of data (*e.g.*, SNP genotypes, AFLP, microsatellites) but also varying numbers of contributing subpopulations/clusters (*K*) and prior information about the geographic sampling location of individuals. Pritchard has repeatedly improved the package, collaborating with Daniel Falush to also provide a more user-friendly interface (Falush *et al.* 2003) and address issues associated with dominant markers and null alleles (Falush *et al.* 2007).

The impact of this new approach and software was not immediate—there were few data sets at the time with sufficient quality and quantity on which they could be implemented. Nonetheless, the work climbed steadily in citations each year, with both applications to real data and further simulations confirming the ability of Pritchard *et al.*’s approach to correctly estimating the number of subpopulations or assign individuals under varying conditions (*e.g.*, Evanno *et al.* 2005). An early application of this approach to studying genetic structure in modern humans from 52 populations showed that within-population variation among individuals accounts

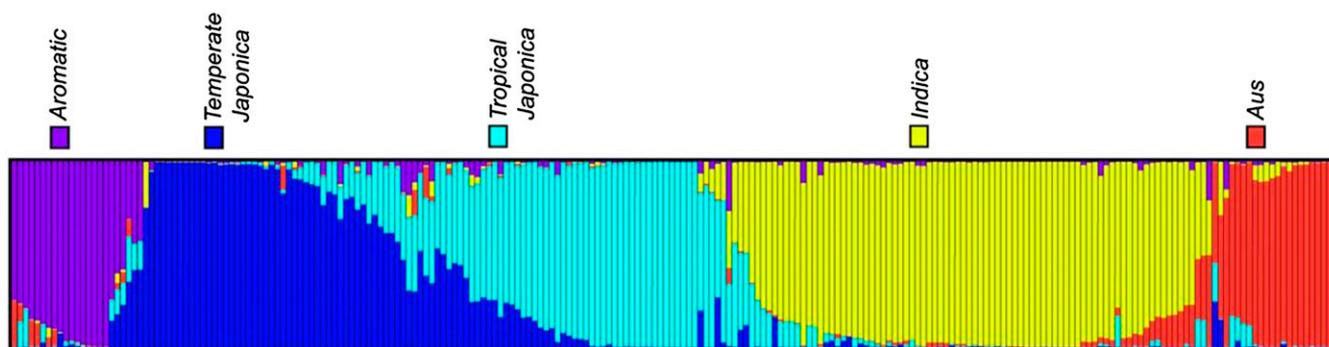


Figure 1 Model-based ancestry for each accession. Color codes are as follows: aromatic, purple; aus, orange; indica, yellow; temperate japonica, dark blue; and tropical japonica, light blue. Reprinted from Garris *et al.* (2005) with permission from the Genetics Society of America.

for ~95% of genetic variation (Rosenberg *et al.* 2002), and this article won *Lancet's* Paper of the Year award in 2003 (Horton *et al.* 2003). Also recognizing its emerging impact, the original article won the 2002 Toby Mitchell Prize from the International Society for Bayesian Applications.

Around this time, the fields of molecular ecology and conservation genetics began to adopt higher-throughput genotyping approaches, and they quickly adopted *structure*. I recall what I perceived to be a turning point at the 2006 conference of the Society for the Study of Evolution, when it seemed every other talk presented a *structure* plot similar to Figure 1, depicting ancestry fractions among rice strain accessions (Garris *et al.* 2005). Indeed, the popularity of the approach has grown so much that the journal *Molecular Ecology* recently published an editorial specifically dedicated to recommendations on using the software and how to report results and archive data related to published works (Gilbert *et al.* 2012).

I should stress that the method employed in *structure* is, by far, not Pritchard's only contribution to the field of genetics. Nearly all of his research focuses on understanding genetic variation, but a large part of his current research is about understanding how and why phenotypes change over time. He is currently a Howard Hughes Medical Institute investigator funded to explore how genetic variation affects gene regulation and specifically how to use eQTLs as way of linking genetic variation to function. From this effort, he has recently been able to link about half of all eQTLs to mechanisms involving changes in chromatin accessibility (Degner *et al.* 2012) and some to changes in mRNA decay rates (Pai *et al.* 2012).

It is an exciting time to be working in the field of genetics. We have more genome-wide sequence and genotype data than we can easily parse and statistical methods like *structure* are exceptionally useful, both for examining population history and for interpreting mapping data as are now leveraged by both researchers and "personal genomics" consumers. We are also finally getting a handle on how the many genes that contribute to complex traits actually work together. The need for creativity in bioinformatic analysis, like that of this year's recipient, has never been more acute, and we congratulate him.

I close with a small personal anecdote of a debt I owe to Professor Pritchard. In 1994, Pritchard interviewed for a graduate

position at the University of Chicago, where I was a second-year Ph.D. student. Being as much a freeloader then as I am now, I immediately volunteered to take him to dinner. I asked several current Ph.D. students to join me, but I literally got to my seventh request before one fellow student, Julie Furneaux (with whom I had never had dinner previously), agreed to come along. After dinner, I asked Julie if she wanted to chat for a while rather than going straight home. One thing led to another, and now, 17 years of marriage and 2 children later, I could not be happier. I cannot help but feel that Pritchard's scintillating dinner conversation about population genetics may have played a small role in getting that wonderful process started.

Literature Cited

- Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356: 519–520.
- Crow, J. F., D. Lindsley, and J. Lucchesi, 2006 Edward Novitski: *Drosophila* virtuoso. *Genetics* 174: 549–553.
- Degner, J. F., A. A. Pai, R. Pique-Regi, J. B. Veyrieras, D. J. Gaffney *et al.*, 2012 DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482: 390–394.
- Evanno, G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611–2620.
- Falush, D., M. Stephens, and J. K. Pritchard, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Falush, D., M. Stephens, and J. K. Pritchard, 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7: 574–578.
- Garris, A. J., T. H. Tai, J. Coburn, S. Kresovich, and S. McCouch, 2005 Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169: 1631–1638.
- Gilbert, K. J., R. L. Andrew, D. G. Bock, M. T. Franklin, N. C. Kane *et al.*, 2012 Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Mol. Ecol.* 21: 4925–4930.
- Horton, R., N. A. Rosenberg, J. K. Pritchard, J. L. Weber, H. M. Cann *et al.*, 2003 Read all about it: the *Lancet's* Paper of the Year, 2003. *Lancet* 362: 2101–2103.
- Lander, E. S., and N. J. Schork, 1994 Genetic dissection of complex traits. *Science* 265: 2037–2048.
- Novitski, E., and S. A. Rifenburgh, 1938 Heldout, a recessive wing mutation in *Drosophila melanogaster*. *Proc. Indiana Acad. Sci.* 47: 256–260.

- Pai, A. A., C. E. Cain, O. Mizrahi-Man, S. De Leon, N. Lewellen *et al.*, 2012 The contribution of RNA decay quantitative trait Loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet.* 8: e1003000.
- Pritchard, J. K., and N. A. Rosenberg, 1999 Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* 65: 220–228.
- Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd *et al.*, 2002 Genetic structure of human populations. *Science* 298: 2381–2385.

Mohamed A. F. Noor