

## Stephen Smith: The Botanist Hacker

**S**tephen Smith just wanted to study honeysuckle. As a new PhD student at Yale University in 2003, “I was going to do a strict monograph,” says Smith. But when he sat down to research the evolution of the genus, he learned that the honeysuckle fossil record is poor, so Smith decided to temporarily expand his research to include the fossils of related plants.

Smith decided to build an evolutionary tree, or phylogeny, of all the Dipsacales, a major group of flowering plants that includes honeysuckles. His advisor, Yale botanist Michael Donoghue, was skeptical. It was a project that he, a world expert on Dipsacales, had attempted but eventually abandoned: There are over 1,000 species of Dipsacales, and at the time, most plant phylogenies maxed out at about 200 organisms due to the complexity of comparing genetic sequences.

But within a few weeks, Smith walked into Donoghue’s office with a sprawling Dipsacales phylogeny in hand. Smith, a skilled banjo player and “retired punk rocker,” had also been studying computer programming since age 12. Confronted with the fact that the computational tools to build such a phylogeny didn’t exist, Smith simply created them. Looking at the 300+ species tree for the first time, Donoghue was speechless. “I thought, ‘Oh my God, I didn’t know you could do that,’” he recalls.

The tree was remarkable for its size, but also for the surprising pattern that emerged from its branches. Lines on the graph representing herbaceous (nonwoody) plants sprouted from the tree like long whiskers, their length representing the quantity of molecular mutations accumulated over time. But woody plant species appeared as short, stubby lines; they acquired significantly

fewer mutations over the same period of time. Smith and Donoghue inferred that woody plants have a slower rate of molecular evolution due to their long generation time—often 10+ years—compared to herbaceous plants, which typically reproduce annually. It was compelling evidence that molecular evolution rates are linked to generation time, a theory that has long been debated among botanists.

Smith decided to look for the same pattern in other plant groups. Within 2 weeks, using more than 30 computer programs (25 built from scratch), Smith constructed phylogenies for four other major branches of flowering plants, the largest including 4,657 species. “He’s a hacker in the best sense of the word—able to pull things together and use them in ways they weren’t intended to be used,” says Brown University biologist Casey Dunn, Smith’s former classmate and current collaborator. Each of the four new trees depicted the same branching pattern: short branches for woody species, long branches for herbaceous species.<sup>1</sup>

Recently, Smith designed analytical tools for the Tree of Life,<sup>2</sup> an online collection of plant and animal phylogenies, and after many email requests from researchers, published a methodology for building mega-phylogenies.<sup>3</sup> “He has streamlined the making of phylogenies,” says Donoghue. “It’s revolutionary.”

Smith joined the National Evolutionary Synthesis Center in Durham, North Carolina, in 2008 after completing his PhD. Fascinated with biogeography, Smith plans to incorporate geographic and climate data into mega-phylogenies, a project he will pursue at Brown during a second postdoc beginning in May. He’s long since abandoned the idea of being a single-genus man. “I’m trying to build the largest trees I can,” he says with a smile. —Megan Scudellari

**Title:** Postdoctoral Fellow,  
National Evolutionary  
Synthesis Center

**Age:** 28

**Representative publications:**

1. S.A. Smith, M.J. Donoghue, “Rates of molecular evolution are linked to life history in flowering plants,” *Science*, 322:86–89, 2008. (Cited in 28 papers)
2. C.W. Dunn et al., “Broad taxon sampling improves resolution of the Animal Tree of Life in phylogenomic analyses,” *Nature*, 452:745–49, 2008. (Cited in 211 papers)
3. S.A. Smith et al., “Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches,” *BMC Evol Bio*, 9:37, 2009. (Cited in 6 papers)