



Green Genomes

With about 30 plant genomes in hand, researchers trace the evolution of our flora and discover that plant DNA is unusually dynamic

ELEVEN YEARS AGO, SCIENTISTS PUBLICLY unveiled the near-complete DNA sequence of a small weed known as *Arabidopsis thaliana*. The debut of this first plant genome marked “a very special time in the whole history of plants,” said Daniel Cosgrove, then president of what is now the American Society of Plant Biologists.

It’s still a special time. True, the sequencing of plants has lagged behind that of animal genomes (*Science*, 20 July 2007, p. 317). But there are now nearly 30 plant genomes available for analysis and comparison, including most recently one belonging to spike-moss, which reproduces via spores rather than seeds and isn’t truly a moss (*Science*, 20 May, p. 960). Buried within all this DNA sequence has been no shortage of surprises and insights into how the planet’s flora and their DNA evolve. “These are really exciting days,” says Patrick Schnable, a geneticist at Iowa State University in Ames. “There is a lot of power in comparing genomes.”

To date, angiosperms, flowering plants that produce seeds and make up the majority of flora seen on Earth today, dominate the roster of the sequenced genomes. Besides

A. thaliana, scientists have deciphered DNA from poplar, maize, rice, and about 20 other angiosperms. But a few other genomes come from species—algae, a moss, the spikemoss—representing key stages in the evolution of land plants. These DNA sequences are beginning to reveal what it took for plants to move onto land, grow tall, and produce seeds and flowers. For example, many genes key to the colorful blossoms that brighten the landscape turn out to have unexpectedly deep evolutionary roots, some stretching back to algae.

“Very few genes seem to be without precedents,” says Jeffrey Bennetzen, a geneticist at the University of Georgia, Athens.

The genomes are also showing that plant evolution is surprisingly dynamic. Plants seem to frequently undergo whole genome duplications, their DNA repertoire swelling in size, then shrinking again through time, gaining and losing genes much more quickly than do animal genomes. “You can see what events have led to the evolution of the genome,” says

Elliot Meyerowitz, founding director of the Sainsbury Laboratory, a major new plant science facility in Cambridge, U.K.

Genes that mattered

The sequenced plant genomes cover a billion years of evolution. The green alga *Chlamydomonas reinhardtii* represents early photosynthesizers, confined to water and never expanding beyond a simple single cell (*Science*, 12 October 2007, p. 245). Its genome showed what it took to be “green”: It had everything necessary to utilize photosynthesis. The alga’s DNA offered a few surprises as well. *Chlamydomonas* has the core of a flowering gene discovered first in snapdragons, but in the alga it’s involved in sensing copper—another example of how a gene gets co-opted during evolution to serve a new purpose. When researchers compared *Chlamydomonas*’s genome with that of the more complex green colonial alga, *Volvox carteri*, they were surprised to find that the transition to multicellularity in this group was a small step, relatively speaking, as the two algae have roughly the same number of genes, 14,500 (*Science*, 9 July 2010, p. 128).

In 2008, bryophytes, which evolved 450 million years ago and were among the first plants to colonize land, entered the genomic age. This group includes mosses, liverworts, and hornworts; the moss *Physcomitrella patens* was the first to fully bare its DNA (*Science*, 4 January 2008, p. 64). Scientists studying its genome have found

Online
sciencemag.org
 Podcast interview
 with author
 Elizabeth Pennisi.

that the climb to shore required new genes for surviving dry spells and temperature swings, and that resulted in a more complex genome, with expanded families of genes.

As spore-forming plants, mosses don't produce seeds, yet *P. patens* unexpectedly revealed genes that in other plants encode proteins that help seeds survive desiccation. The team that decoded this moss's genome, led by Stefan Rensing of the University of Freiberg in Germany, concluded that these progenitors of seed-plant genes must have originated in the earliest land plants to protect the entire plant during dry spells. In a new comparative analysis, Tomoaki Nishiyama and Mitsuyasu Hasebe of the Japan Science and Technology Agency in Okazaki found that this moss also has 80% of the several hundred genes responsible for development in *Arabidopsis*, indicating that the tool kit for more complex plant architectures was already in place long before angiosperms evolved.

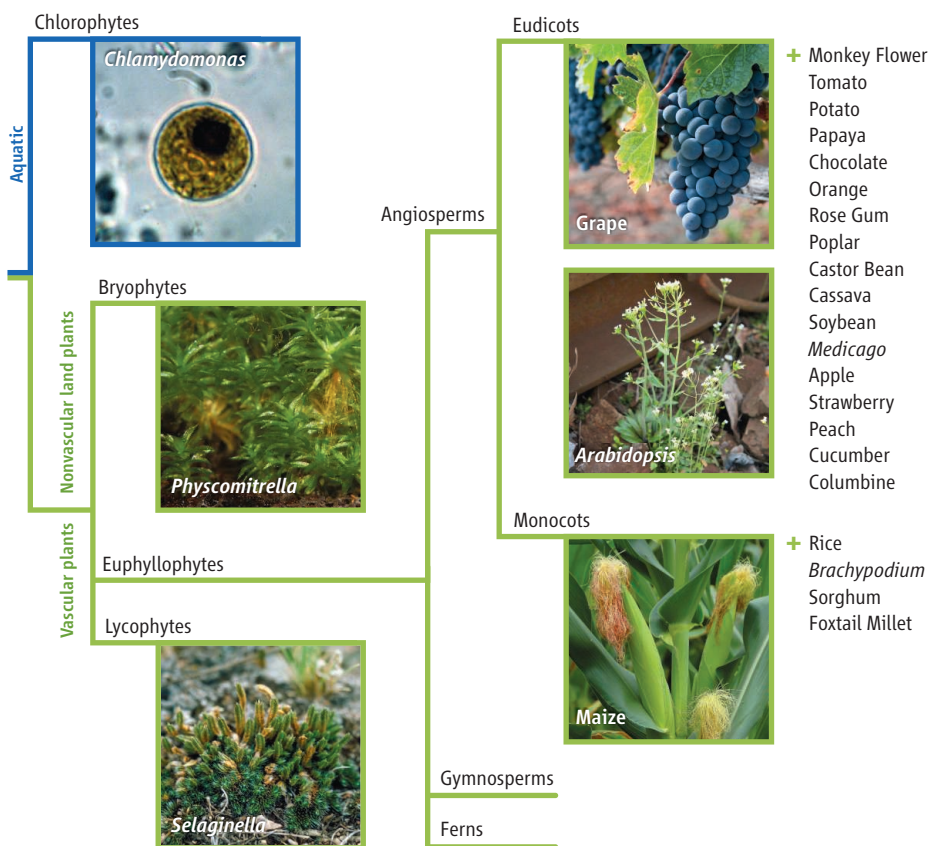
After plants made the leap to land, a key innovation enabled them to grow tall: a system of tissue that conducts water and nutrients. One group that evolved this vascular tissue early on was the lycophytes, which represent an interim step in plant evolution because they have a vascular system but do not produce seeds. Some 300 million years ago, lycophytes were the first giants of the plant world, dominating the carboniferous forests whose remains became coal beds. Today's lycophytes, which include spike-mosses and clubmosses, are small plants with scalelike leaves on branching stems found in forests and gardens.

Like the moss genome, the spikemoss DNA revealed genes expected in more complex plants, suggesting that those genes have deep roots. "We see the evolution and origin of genes and gene families that we know are very important in wood development" even though lycophytes are not woody plants, says Claude dePamphilis, a plant genomicist at Pennsylvania State University, University Park.

The emergence of lycophytes also marks a transition in the evolution of plant reproductive strategies. Plants undergo sexual and asexual life stages. In moss and algae, the sexual stage is the dominant phase; in lycophytes, ferns, and seed plants, the asexual stage is the form we are most familiar with. Researchers had thought this transition would have required a lot of genetic innovation, but a comparison of the spikemoss genome with other plant genomes shows that other transitions required much more heavy lifting.

Purdue University geneticist Jo Ann Banks and her colleagues looked at all the

PLANT GENOMES ON THE TREE OF LIFE



Lopsided coverage. Flowering plants (angiosperms) have dominated the floral genome scene, with moss, algae, and spikemoss representing the few exceptions.

available plant genomes, grouping genes into families. They found 3814 gene families common to every plant from algae to grape, from which they concluded that these were in the common ancestor to all green plants. A comparison of algae and mosses to other plants indicated that the shift in reproductive strategies and the transition to vascular plants required only about 516 new gene families. The evolution of angiosperms required an additional 1350 families. The move to land demanded the most evolutionary innovation, with 3006 new gene families appearing, they reported. "In number and types of genes, the most dramatic change was what was acquired to move onto land," dePamphilis says.

Genomic anarchy

Comparisons of plant genomes have done more than highlight genes that came on the scene through time. They have also revealed insights into the broader evolution of the genomes themselves. Plant genomes change more quickly than do animal genomes, leading to more variation among closely related species and even within a species. Chimeric DNA—bits of two dif-

ferent genes that have merged—arises relatively frequently, and transposable elements, mobile DNA sequences in a cell, are actively remodeling the genome, creating new regulatory elements. The overall conclusion: Plants practice "genomic anarchy," says Detlef Weigel, a geneticist at the Max Planck Institute for Developmental Biology in Tübingen, Germany.

Why such anarchy? Plants have such a different lifestyle from that of animals that they apparently need more flexible genomes. "The different features of their genomes reflect their different evolutionary histories, and these can be correlated with their different survival strategies," says Andrew Leitch, a plant evolutionary biologist at Queen Mary, University of London.

A mustardlike weed was the first to alert researchers to this anarchy. In the late 1990s, the expense and difficulty of sequencing prompted researchers to pick small, simple genomes; hence they tackled *A. thaliana*, with its 125 million bases, ahead of much more economically and politically important maize, an estimated 2.5 billion bases. Also, *Arabidopsis* was supposedly diploid,



Filling out the tree. Now the gymnosperms (left to right) Douglas fir, loblolly pine, sugar pine, and Norway spruce are being sequenced.

with just two copies of each chromosome, whereas maize seemed to come from a polyploid; its extra chromosomes might have made piecing together sequenced DNA an even more daunting task.

But the *Arabidopsis* DNA revealed that at least twice, the plant had somehow duplicated its entire genome and survived the unusual event. Subsequent genomes have revealed this to be a recurrent theme in plant evolution, with different species experiencing one or more instances of whole genome duplications at different times in their history. Moreover, it's becoming clear that many plant genomes go through cycles in which their genomes double in size; most extra genes disintegrate over the next few million years, returning genomes to near preduplication gene numbers; and then doubling occurs again. Poplar's duplication was more recent than *A. thaliana*'s, for example, and thus its genome has many more protein-coding genes.

Indeed, the latter seems to be in the winning stage. Tina Hu of the University of Southern California in Los Angeles and her colleagues sequenced a close relative, *A. lyrata*, and reported online 10 April in *Nature Genetics* that it has more than 500 additional genes. Moreover, lining up the two *Arabidopsis* genomes shows that *A. thaliana* has undergone hundreds of thousands of small deletions in the 10 million years since the two plants diverged.

These cycles significantly alter a plant's genetic landscape. In *A. thaliana*, many of the documented deletions occurred in the DNA regions that don't encode proteins and are likely sites of gene regulation. "We are beginning to get a sense for how dramatically the regulatory programs of these plants have been reset with each new round of polyploidization and subsequent diploidization," says Todd Vision, a plant evolutionary biologist at the

University of North Carolina, Chapel Hill.

Plant genomes are generating novelty in other ways as well. For one, their transposable elements are much more active than are those in animal genomes, hopping in and out of chromosomes, dragging bits of DNA with them and in doing so positioning that DNA where it can help regulate genes in new ways. At the same time, researchers examining plant genomes have discovered thousands of examples of gene shuffling, in which fragments of two or more genes have been brought together to form an apparently functional new gene. What most of them do remains unknown, but in maize, 8% of these chimeras are under some form of selection, indicating that they are important to the species' survival, Bennetzen says.

"What's also been surprising is how rapidly plant genomes change" compared with those of animals, Bennetzen notes. One can match a fish genome up with a primate's and find noncoding DNA sequences that are conserved between these organisms even though they diverged 400 million years ago. (Noncoding DNA tends to change faster than that which encodes proteins.) But in plants, such conservation disappears fast. Researchers have compared *Arabidopsis* with rice, maize with sorghum, and rice with sorghum—finding little in common in the noncoding DNA between genes. Even between two corn varieties, this DNA can differ by 20%. "That's much more different than between chimps and humans," Weigel says.

Because DNA bases between genes are often sites that control gene activity, plants likely undergo much faster evolution in gene regulation than do animals. "If we don't have changes in gene regulation 100-fold faster than in mammals, I would be shocked," Bennetzen says.

Such insights have made plant biolo-

gists hunger for more genome sequences. In some cases, researchers are going after just the full set of expressed genes. For example, since 2009, the 1000 Plant Genomes Project has looked at about 500 plant species with the goal of examining a species from every angiosperm family, many medicinal plants, and hundreds of green algae by the end of this year. The dozen subprojects include one that will survey the plants' light-sensitive proteins and one that is examining the more advanced C4 photosynthetic system, which evolved at least 62 times in angiosperms. Moreover, a collaboration of U.S. universities is pulling out all the expressed genes of so-called parasitic plants, which invade the tissues of other plants.

Some researchers are also looking at varieties of already sequenced species. The 1001 Genomes Project is sequencing that many strains of *A. thaliana* to get a better sense of how variation affects trait differences among strains, with about 250 already completed. Multiple maize genomes are being deciphered as well.

Still other researchers are picking specific plant genomes to broaden coverage across the tree of life. DePamphilis is leading an effort to sequence *Amborella*, which sits at the base of the angiosperms. Four conifer sequencing projects began this year: the loblolly pine, sugar pine, Douglas fir, and Norway spruce. They will be the first gymnosperms, the other major group of seed plants, to be sequenced. Banks and her colleagues are working to get support to study a fern's genome. Despite advances in DNA sequencing efficiency, that would still be daunting; ferns can have up to 1000 chromosomes. Still, a fern genome may help answer why they're so genetically complex. All in all, Vision says, it's "heady days to study plant genome evolution."

—ELIZABETH PENNISI