The complete genome sequence of the first higher plant is revealing unexpected similarities to—and differences from—the genomes of other organisms, including humans

Plants Join the Genome Sequencing Bandwagon

ARABIDOPSIS

These three stories describe the sequencing of the first plant genome and the research that achievement has made possible. See also Editorial on p. 2071, Policy Forum on p. 2077, and Research Articles beginning on p. 2105.

▶SEQUENCE
ECOLOGY
CENTROMERES

COLD SPRING HARBOR, NEW YORK—In simultaneous press conferences across the globe, an international consortium announced this week that it has finished the first genome sequence of a higher plant. For plant biologists, the eagerly awaited genome of this small weed, *Arabidopsis thaliana*, offers a window into the genetic makeup of all plants, including key crops. And it's a clear window indeed, as the six international sequencing teams on three continents have pulled off a coup: They produced a genome sequence that is more accurate than that of any multicellular organism which has been published to date. Through this window, they are seeing for the first time that plants may be much more complex than many biologists have imagined. And that realization turns out to be a bit humbling.

Toasting a genome. (Left to right) Daphne Preuss, Machi Dilworth, Robert Martienssen, Elliot Meyerowitz, James Watson, Hans-Werner Mewes, W. Richard McCombie, and Michael Bevan were on hand to celebrate the completion of the *Arabidopsis* sequence.

CREDIT: MIRIAM CHUA

With its estimated 125-million-base sequence, *Arabidopsis* has a surprisingly high number of predicted genes—some 25,500, compared with 13,600 in the fruit fly. That total—which exceeds by several thousand that of any other genome yet analyzed—comes impressively close to low-end estimates of the number of human genes (*Science*, 19 May, p. 1146). "It says that you have to have more respect for plants," concedes Nobel laureate James Watson, co-discoverer of the double helical structure of DNA. "Plants may be sessile, but they are not stupid," adds geneticist Michael Sussman of the University of Wisconsin, Madison. Both were on hand this week at Cold Spring Harbor Laboratory (CSHL) along with more than 150 *Arabidopsis* researchers for a gathering that was part science and part celebration.

The assembled researchers were giddy with pride and anticipation. For them, the sequence, which is described in the 14 December issue of *Nature*, signals the dawn of a new era. "This is a very special time in the whole history of plants," says Daniel Cosgrove, president of the American Society of Plant Physiologists. He expects that the sequence data will speed the identification of genes to improve agriculture as well as expand knowledge of basic plant biology.

Not bad for a plant that 20 years ago was considered decidedly offbeat (*Science*, 6 October, p. 32). Even a decade ago, when *Arabidopsis* enthusiasts in the United States and Europe proposed understanding the genome of this mustardlike weed, many of their colleagues had to be convinced. Little happened until 1994, when, with encouragement from the European Commission, several European labs joined forces and began a pilot sequencing project. Coordinated by Michael Bevan, a geneticist at the John Innes Centre in Norwich, United Kingdom, the project quickly swelled to 17 European groups. It went transatlantic in 1996, when The Institute for Genomic Research (TIGR) in Rockville, Maryland, signed on; CSHL teamed up with Washington University; and Stanford, the University of Pennsylvania, and the U.S. Department of Agriculture lab at the University of California, Berkeley, formed a third group. Satoshi Tabata and his team at the Kazusa DNA Research Institute in Kisarazu, Japan, also joined the consortium.

Their efforts have yielded 115 million bases of *Arabidopsis* DNA that cover the five chromosomes. These sequencers also ventured into the sequence and structure of the centromere, an important but rather inscrutable region of the chromosome that assists in chromosome pairing (see p. 2057). To date, chromosomes 1 and 5 have about three gaps each left to fill, far fewer, for instance, than the 1200 gaps in the *Drosophila* genome sequence. "It's by far the most high-fidelity multicellular eukaryotic genome to date," says Jeffrey Dangl, a molecular geneticist at the University of North Carolina, Chapel Hill.

Although each sequencing group analyzed the sequence as they produced it, two groups in particular scrutinized the data intensely, using computers and the human eye to pick out all the genes and their coding regions. TIGR and the Munich Information Center for Protein Sequences in Germany found that it has lots of duplicated genes. They estimate that, overall, duplicated regions make up 58% of the genome, likely because of whole-genome duplications that occurred 100 million years ago. They also assigned tentative functions to roughly 70% of the genes, in part by comparing the genes and their proteins to those of other organisms already in the genome databases.

From these comparisons other surprises have emerged. For one, some 100 *Arabidopsis* genes have counterparts in humans that cause disease, including the genes involved in cystic fibrosis and breast cancer. This plant also has a relatively large number of genes for water-channel proteins, as well as many more protein kinase genes than expected. Because protein kinases are involved in cell signaling pathways, "there are an awful lot of ways *Arabidopsis* cells are talking to each other and getting information from outside the cells," says Elliot Meyerowitz, a geneticist at the California Institute of Technology in Pasadena.

Another intriguing find is that the set of genes for cell-to-cell communication varies dramatically between *Arabidopsis* and humans. By contrast, many genes involved in basic cell functions, such as cell division, tend to be conserved among species—in other words, a yeast gene might work just as well in *Arabidopsis*—suggesting that the genes existed in a common ancestor of all organisms. But not so the cell communication genes. "[That] fits with the idea that multicellularity evolved separately [in plants and animals]," says CSHL's Robert Martienssen, a plant geneticist.

These are the first of many insights expected to pour out of the genome, not just for learning about *Arabidopsis* but also about other plants. With this genome in hand, plant biologists should be able to find key genes in other plants far more easily. As Bevan points out, such insights may lead to the development of crops better suited to developing countries, plants designed to soak up more carbon dioxide, or other applications that cannot yet be imagined. For that reason, he predicts, the genome sequence from this tiny plant "will have as much impact as the human genome."

–ELIZABETH PENNISI

ARABIDOPSIS

▶ **ECOLOGY**

Stalking the Wild Mustard

Arabidopsis thaliana **is not just for molecular biologists anymore. A few intrepid ecologists and evolutionary biologists are now touting its merits**

Several years into his Ph.D. studies, Massimo Pigliucci realized he had a problem. He wanted to study how the environment interacts with a species' genetic makeup over time, and he had been focusing on *Lobelia*, a tall, native perennial much studied by ecologists. But *Lobelia* "was hard to use," he recalls, because little was known about its genetics. So Pigliucci shifted to a small weed typically studied by molecular and developmental biologists rather than ecologists. His committee at the University of Connecticut was appalled. The weed, *Arabidopsis thaliana*, "wasn't a real plant" as far as his advisers were concerned. And, they asserted, it had no intriguing features worthy of ecological or evolutionary investigation.

A decade later, Pigliucci, now at the University of Tennessee, Knoxville, and a small but growing number of other plant ecologists and evolutionary biologists are proving those advisers wrong. They have searched the globe for natural variants of this weed and of its relatives, which will enable them to probe the evolutionary history of this plant (see sidebar). Bucking the trend among field biologists to focus on wild flora in pristine places, these researchers are capitalizing on decades of intensive research on *Arabidopsis*. Thousands of research papers have already examined the physiology, development, biochemistry, and genetics of this plant. Indeed, this week the full genome sequence of *Arabidopsis* is being published—the first complete sequence of any plant (see *Nature*, 14 December, p. 796, and this issue of *Science*, p. 2105). Now it's time for more ecologists and evolutionary biologists to make use of these data, says Thomas Mitchell-Olds, an ecologist-turned–quantitative geneticist at the Max Planck Institute of Chemical Ecology in Jena, Germany. For instance, he and others want to tease out how genetic variation helps *Arabidopsis* thrive under different conditions.

Even mainstream ecologists are beginning to see the potential of this lowly weed. Stanford plant ecologist Harold Mooney, for one, is convinced that *Arabidopsis* will help evolutionary biologists and ecologists answer numerous other questions, and he is encouraging his colleagues to take a chance on this unlikely plant. Specifically, says Mooney, *Arabidopsis* offers "an incredible opportunity" for them to understand the genetic basis of the traits they study. Until now, researchers have had trouble connecting the genotype—an organism's particular mix of genes—with its phenotype—how that organism develops and acts. But with *Arabidopsis*, "you can precisely answer this question, because it's got great genetics," says Michael Purugganan, an evolutionary geneticist at North Carolina State University in Raleigh.

A very real plant. Much used in the lab, *Arabidopsis thaliana* also thrives in the wild, where it is providing insights into how plants adapt to different environments.

CREDIT: MICHAEL PURUGGANAN AND JOHANNA SCHMITT

Those great genetics include not only the sequence but also mutants that *Arabidopsis* biologists have developed for almost every gene. By growing these mutants—or natural variants—under different conditions, Pigliucci and other plant ecologists are homing in on the genes responsible for specific traits, such as the timing of flowering or the production of protective compounds. The sequence information "allows us to plan our experiments more carefully and analyze our data more thoroughly," says Naohiko Miyashita, a population geneticist at Kyoto University, who switched from *Drosophila* to *Arabidopsis* 6 years ago.

Rodney Mauricio, an ecological geneticist at the University of Georgia, Athens, is using *Arabidopsis* to look at how insect pests affect the evolution of plants. Although he complains about how delicate this weed is—for his next experiment he recently planted 4800 seedlings, each no bigger than a thumbtack—he's one of the pioneers in using *Arabidopsis* in the field. Over the past decade he and his colleagues have established that insect pests prompt plants to develop tiny hairs called trichomes and that these have a "cost": The hairier plants produce fewer seeds. "It's a fairly important adaptive story," Mauricio notes. Now he's tracking down the genes responsible for trichome growth and density. Says Mauricio, "The idea is to move toward a complete understanding of evolution from the ecological side to the molecular side."