Charis Cook

The Other Genome Project

Key words
genome
Arabidopsis
database
Green Revolution

*In this article, **Charis Cook** explains the significance of a cress called Arabidopsis, the plant with the best-understood genome.*

Food security hit the headlines in the 1960s, when leading economists were predicting a catastrophic food crisis. That food-based apocalypse never arrived due to the unforeseen Green Revolution which was largely down to a long 20 years of work by Nobel Peace Prize winner Norman Borlaug. Today, global population growth and climate change means the world's food supply is again facing an uncertain future. Fortunately, plant research has changed beyond recognition since Borlaug started his wheat breeding programme in 1944, and scientific progress moves much faster today. The reason for the step change in plant science is extensive research into an insignificant member of the cabbage family, *Arabidopsis thaliana*, subject of The Other Genome Project.

Arabidopsis rows: Arabidopsis growing in a glasshouse at the University of Warwick. Image courtesy of Ruth Bastow, GARNet

Arabidopsis thaliana is a model species like the fruit fly *Drosophila*, yeast, and the bacterium *Escherichia coli*. Scientists work on these species in order to understand as much as possible about a particular kind of organism. Arabidopsis was first suggested as a plant model species for developmental and genetic research in 1943. A German scientist named Friedrich Laibach noted that *A. thaliana* was easy to grow inside in a small space, showed a lot of natural variation, was amenable to cross-breeding between varieties, and generated large numbers of seeds per plant.

Envel Kerdaffrec, Gregor Mendel Institute

Arabidopsis ecotypes: A selection of Arabidopsis ecotypes growing at the 1001 Genomes Project at the Gregor Mendel Institute, Austria.

William M. Gray PLoS Biol via Wikimedia Commons

Auxin mutant: An example of an extreme phenotype in Arabidopsis thaliana*. Wild type* Arabidopsis thaliana *is on the left. An auxin signal-transduction mutant,* axr2*, is on the right. Auxin is a plant hormone involved in many processes, including growth.*

Laibach's research group and a number of other German researchers worked on Arabidopsis well into the 1960s, but it was not until the 1980s that Arabidopsis became widely used worldwide. By this time, it was clear that artificially adding genes to Arabidopsis by gene transfer using *Agrobacterium tumefaciens* was easier than in other plant species, and genetically modified seed were shared between the growing number of Arabidopsis research groups.

Box 1 Transferring DNA into plants

Bacteria called *Agrobacterium tumefaciens* naturally insert their own DNA into plants. Scientists use *A. tumefaciens* to put specific pieces of DNA into plants. First, the DNA fragment that needs to be inserted into the plant genome is put into a circular piece of DNA called a plasmid. The plasmid also contains an antibiotic resistance gene which will act as a 'marker' for bacteria and plants containing the plasmid DNA.

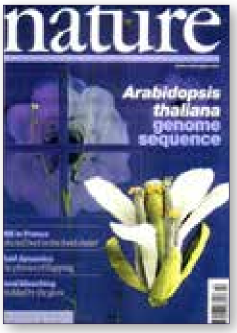
The plasmids containing the insert are put into *A. tumefaciens* cells which are grown on agar plates containing the antibiotic. The colonies that survive contain the plasmid, and are then grown in liquid culture.

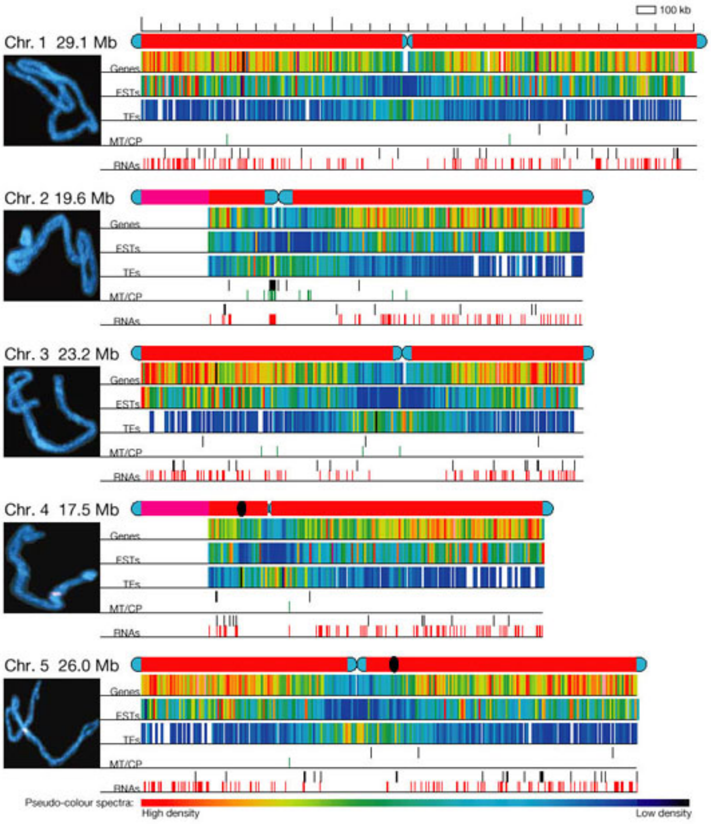
Arabidopsis thaliana is dipped into the *A. tumefaciens* culture flower-first, so that the bacteria can transfer the plasmid to the pollen and ovules. This is called floral dipping. The seeds produced by these plants are grown on media containing the antibiotic, and the seedlings that grow contain the plasmid DNA. DNA analysis can confirm the presence of the required DNA fragment in the seedlings.

Participants at the First International Symposium on Arabidopsis Research in 1965, held in honour of the 80th birthday of Friedrich Laibach (standing at centre with glasses).

In 1989, James Watson, one of the team that discovered the structure of DNA and by then the Director of Cold Spring Harbour Research Laboratory in the USA, called a meeting to discuss the use of Arabidopsis as a model for genetic research. This was when the concept of an Arabidopsis genome project was first discussed. A year later, a group of leading plant scientists published a report outlining plans to sequence the whole Arabidopsis genome. At the time, that sounded overly ambitious and probably impossible – *A. thaliana* has a relatively small genome, but it is around 120 million base pairs long.

In 2000, after ten years of work from scientists in more than 50 research groups from all over the world, the genome was sequenced. The author of the research paper was 'The Arabidopsis Genome Initiative', recognition that this huge project was only possible due to the research community working together. The genome is published and publicly available, but its analysis is an on-going project.





A representation of the five genes of Arabidopsis, from the paper in the journal Nature where the details of the genome were first published (Nature 408:796-815, 14 December 2000).

The next Green Revolution

During the 1940s and 50s, scientists physically tested the wheat varieties that would spearhead the Green Revolution for pathogen resistance and higher yield. They grew them to maturity, observed the phenotype, interbred them, and repeated the cycle with the hybrid offspring. Thanks to the Arabidopsis genome project, today scientists get the information they need by analysing the genome of seedlings, a far faster process than breeding over many life cycles. This is because the genes conferring resistance to disease, or other beneficial characteristics, have been identified. Most were first found in Arabidopsis, and though crop genetics is advancing very fast and more genomes have now been sequenced, many useful genes are still discovered by researchers doing fundamental plant molecular biology on *Arabidopsis thaliana*.

Box 2 Some definitions

Model species: An organism that is widely used in genetic studies because it is convenient to work on.

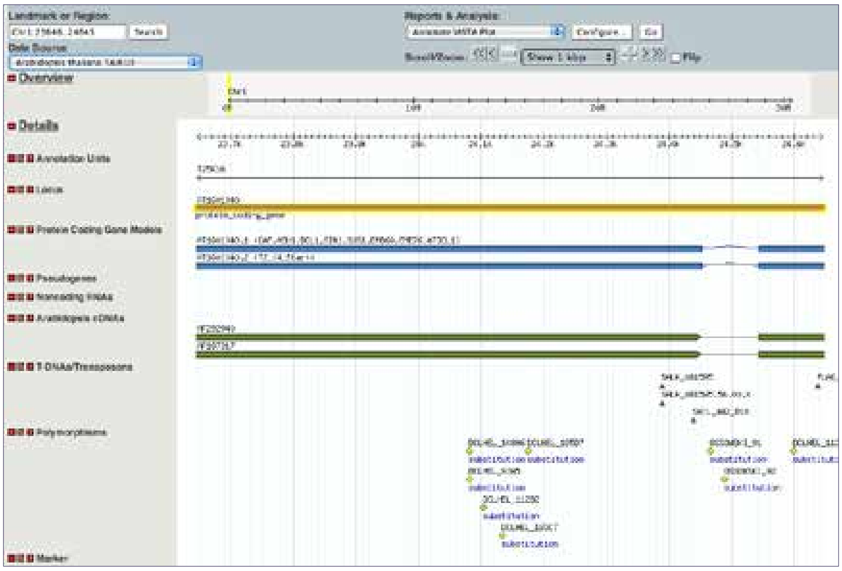
Phenotype: An organism's physical characteristics, defined by its genetic make-up and its interaction with the environment. An example is resistance to draught.

Homologue: Many genes are similar in multiple species. These genes are homologues of each other.

Genome-wide association studies: Analysing entire genomes to identify genetic features associated with a specific phenotype.

Ecotype: A genetically distinct variety of a species, which is usually adapted to a particular habitat.

With the sequence for 'Arabidopsis Protein A', a wheat researcher is likely to be able to find the homologue 'Wheat Protein A' quite easily – even if it turns out that there are actually two wheat proteins that do the job of one Arabidopsis protein. Searching through genomes and comparing gene sequences to find gene homologues in different species is done through free online databases. It is possible thanks to the open nature of modern science, which was born at the start of the Arabidopsis genome project. The world-wide effort to sequence the genome brought together a global research community in a way that had never been done before. Importantly, the resulting data were made public so the community continued to grow as new researchers used and built on the data.



A screenshot from TAIR, The Arabidopsis Information Resource, an online database of genetic information on Arabidopsis.

Working together

Community projects on the Arabidopsis genome are still on going and use new sequencing technologies. New sequencers are able to sequence multiple copies of the Arabidopsis genome in a single run. A notable recent large-scale research effort, the 1001 Genomes Project, was a change in direction for the plant science community. Genomics has traditionally been used to look within the genome to understand molecular processes, but the 1001 Genomes Project is a genome-wide association study that aims to link genetic variants to physical variation in *Arabidopsis thaliana* ecotypes from all over the world. As the name of the project suggests, it involves sequencing and analysing hundreds of genomes. Among the genes identified by the project so far are those responsible for flowering variation, and the single protein that is the primary determinant of levels of the metallic element cadmium in leaves. These could not have been found using traditional genetics experiments, which investigate one gene at a time.

The Arabidopsis Genome project changed plant science in the 1990s by giving researchers a common cause and community resources. Over the last 20 years, scientists working on other plant species have mimicked this, forming multi-group collaborations for large-scale projects. Arabidopsis researchers are still the pioneers of plant science. Although the public see the impact of plant science in crop plants, fundamental research is essential for progress in applied biotechnology. Now, as it was 30 years ago, *Arabidopsis thaliana* is the first choice for researching basic plant science and making new discoveries about how plants work.

Charis Cook is Liaison and Communication Officer for the UK Arabidopsis research network, GARNet (the Genomic Arabidopsis Resource Network).

Look here!

The Arabidopsis Information Resource (TAIR) is at *www.arabidopsis.org*. It provides tools for analysing the genome as well as pages of background information about Arabidopsis.

Nearly all gene sequences from any species (including Arabidopsis) are deposited at the National Centre for Biotechnology Information (*www.ncbi.nlm.nih.gov*). All the data and tools are free to access.