# Annex B[1]

# Mini case study: International Arabidopsis Genome Research Project

Prepared by the Delegation of the United States

## History

In the late 1980s, the Human Genome Project was being established in the United States and Europe. For the model organism genome project, the US National Institutes of Health (NIH) selected *E. coli*, yeast, and *C. elegans* (a worm), but decided not to include a plant. However, the National Science Foundation (NSF) recognised the importance of a plant genome project, and held a workshop attended by NIH, the Department of Energy (DOE), the Department of Agriculture (USDA) and NSF programme officials as well as plant science researchers from universities and private industry. The workshop participants recommended the establishment of an international plant genome project using Arabidopsis as the model system. A series of workshops were held to develop an international Arabidopsis genome project.

In the 1990s, nine Arabidopsis researchers from the United States, Europe, Australia and Japan formed an *ad hoc* committee and drafted a long-range plan for the Multinational Co-ordinated Arabidopsis Thaliana Genome Project. The draft plan was discussed at the International Conference on Arabidopsis Research held in Vienna and approved by the approximately 400 attendees.

The Multinational Arabidopsis Steering Committee was established in the 1990s to: *i)* co-ordinate programmatic aspects of the Arabidopsis genome project; *ii)* communicate with the informatics and biological resource centres that were to be established; *iii)* monitor and summarise progress of scientific activities in participating laboratories; *iv)* serve as liaison to the broader plant biology community; *v)* identify needs and opportunities of the Arabidopsis research community and communicate these to the funding agencies of participating nations; and *vi)* periodically update the long-range plan.

In 1996, an international consortium consisting of the United States and the United Kingdom supported by the European Commission, France and Japan began sequencing the Arabidopsis genome. In 2000, the sequencing was completed. In 2001, the Multinational Co-ordinated Arabidopsis Functional Genome Project was established as a follow-up to the Multinational Co-ordinated Arabidopsis Genome Project. In 2002, The Multinational Arabidopsis Steering Committee published a long-range plan for the follow-up project.

From 2001 to 2010, various funding agencies around the world established programmes to support this international effort: the NSF's Arabidopsis 2010 Project; the UK Biotechnology and Biological Sciences Research Council's (BBSRC) "Genomic

1. As noted in Chapter 1 and Box 1.1, two smaller institutional case studies were undertaken to complement the in-depth core case studies, and one is presented in this annex.

Arabidopsis Resource Network (GARNeT); and the German Research Foundation's (DFG) Arabidopsis Functional Genomics Network (AFGN), to name a few.

In 2010, The Multinational Arabidopsis Steering Committee discussed developing the next long-range plan at its annual meeting held in Yokohama, Japan, during the 21st International Conference on Arabidopsis Research. There were 1 400 attendees at the Yokohama conference (compared to 400 in 1990 in Vienna.)

## Why Arabidopsis?

The plant scientific community identified Arabidopsis as the ideal model for several reasons. Arabidopsis is a plant that is easily manipulated, genetically tractable, and about which much was already known. Arabidopsis also has a relatively small genome consisting of approximately 25 500 genes. It therefore allowed for testing hypotheses quickly in the laboratory. In 2002, the Multinational Arabidopsis Steering Committee acknowledged the importance of collaboration: "We have put forth a goal of no less than complete understanding of the biology of an organism: the only way to achieve success is to work together with realization that we are all wedded to the same goal."

The Arabidopsis Project has had two phases. The first resulted in a complete annotated genetic sequence and the adoption of ontological standards. Once this phase was completed, the scientific community began work "to determine the role of every gene in the control of the metabolic and developmental processes of the plant".

The stated mission of the project was to increase the knowledge and understanding of a flowering plant, using Arabidopsis as an experimental model system. The long-term goal is to increase crop yields and nutritional value, enhance resistance to drought, disease and other stresses, and decrease dependence on chemical pesticides. In addition to producing more and higher-quality food, the research also advances understanding of plants for use in sustainable energy and pharmaceutical products.

## Project management

The project was driven by science and managed by scientists. The Multinational Arabidopsis Steering Committee met once a year at the annual International Conference on Arabidopsis Research and issued annual progress reports. Funding agency programme staff (especially NSF, EC, BBSRC and DFG) worked closely together and with the Multinational Arabidopsis Steering Committee. For joint activities, such as the whole genome sequencing project, funding agencies co-ordinated funding although funds were never co-mingled.

Since 2001, NSF, DFG and BBSRC have provided funding for a co-ordinator of the Multinational Arabidopsis Steering Committee. All co-ordinating activities have been carried out over 20 years without any memoranda of understanding or other officially signed agreements.

## Factors of efficiency and effectiveness in modes of governance

- Open communication and immediate sharing of information, data, as well as biological and research resources.
- Sense of ownership by the research community.
- Commitment to the project by international leaders in the field.
- Common goals shared by all participating researchers and funding agencies.
- Stable leadership at funding agencies.
- No initial issues with respect to data release and industry because Arabidopsis is a model system and has no commercial value as a crop plant.
- Establishment of a co-ordinator to assist the Steering Committee in the co-ordination of international research efforts.
- Establishment of repositories for long-term storage and dissemination of data and biological resources in the United States and the United Kingdom.

## Challenges to effective collaboration

During the functional phase, some of the researchers who received industry funding encountered obstacles for sharing data and research resources. Over time, the scientific community became more educated about management of IP and related issues and developed solutions to enable them to share the data as they were generated.

Communicating the value of research on Arabidopsis to decision makers and the public has been difficult. Participating scientists must be constantly reminded of their obligations to communicate the value of their work to the taxpayers.

## Benefits and outcomes

Arabidopsis has become the model for sequencing the genomes of plants. The techniques and standards developed in the Arabidopsis project became a model for studies of all plants, including rice and maize. The Arabidopsis Project was not inexpensive; it cost approximately USD 75 million. The knowledge gained by the Arabidopsis project enabled subsequent sequencing efforts to be carried out at a much lower cost.

The Arabidopsis Project has had broader impacts, including the development of enabling technologies for sequencing. It resulted in an increase in US utility patents. The first sequence was published in 1996. The Multinational Coordinated Function Genomic Project Annual Report for 2008 maps (p. 22, Figure 3) US utility patents referencing Arabidopsis from 1987-2007. The data demonstrate that dissemination of the sequence information correlates with a significant increase in utility patents; sharing of the sequence information enabled scientific studies that led to patent inventions.

The Arabidopsis Project has advanced knowledge of plant biology, moved the field of plant biology to the age of systems biology, and led to the development of more advanced tools to sequence and elucidate the function of plant genes. It has served as a model for sequencing the genomes of plants with greater commercial potential. This research is crucial to increasing food feed, and fibre production and addressing the global challenge of food security.

## Lessons learned

- It is very important for the international scientific community to agree on a data and research materials release policy at the beginning of the project.
- The Arabidopsis Project was successful because it was science-driven and had broad support from the international scientific community. The science community identified the need to sequence a flowering plant to advance plant genomic research and identified Arabidopsis as the ideal candidate. Both phases of the project were successfully implemented without formal international arrangements.
- The government role in the project, however, was also critical to its success. Federal funding agencies funded critical research under the project and facilitated sharing of the resulting sequencing and functional data by requiring that they be shared with the project partners in a timely manner.
- In addition to funding research, national funding agencies can play a crucial role by establishing policies that encourage timely data sharing, requiring grantees to develop a data management plan acceptable to the scientific community as part of their research proposal, and assisting researchers in the identification of appropriate repositories for the results of the research.