THE REANNOTATION OF THE *ARABIDOPSIS THALIANA* GENOME

Roger K. Smith, Jr., Brian J. Haas, Rama Maiti, Agnes P. Chan, Linda I. Hannick, Catherine M. Ronning, Yong-Li Xiao, Christopher D. Town, Owen R. White
The Institute of Genomics Research (TIGR), 9712 Medical Center Drive, Rockville, MD20850, USA (email: owhite@tigr.org)

Keywords Arabidopsis, genome

1. INTRODUCTION

The small crucifer *Arabidopsis thaliana* emerged as a “model” plant species due to the many experimental advantages the plant has over more economically important plants (reviewed in Meinke, et al., 1998). These economically important plants, mainly crop species, are generally difficult to manipulate in the lab, are large, and have large and complex genomes. Arabidopsis, on the other hand, has a rapid life cycle (6 weeks), a small size, and a small diploid genome (125 Mb, compared to maize, 2,500 Mb, and wheat, 16,000 Mb)(Meyerowitz and Somerville, 1994). These advantages led to an explosion in the number of members of the scientific community using Arabidopsis to investigate biological questions. To date, numerous tools and resource materials exist for those wishing to pose questions using this model plant.

To support the ongoing research efforts in Arabidopsis, the Arabidopsis Genome Initiative (AGI) was formed (Bevan, 1997). This international collaboration facilitated the coordinated sequencing of the genome, and the release of the first completed plant genome was “a scientific event of some importance” (Dennis and Surridge, 2000). Plants evolved along similar paths, and diversification among species is a relatively recent event. Thus, most developmental and physiological processes, and the genes controlling them, are thought to be conserved among various plant species. The elucidation of the genomic sequence of Arabidopsis provides a means for analyzing gene function relevant to a range of plant species.

The multinational effort of AGI produced large amounts of sequence data, but without annotation of genes and other biologically important features, these sequences are of little utility for the research community. AGI groups realized this, and the annotation of the Arabidopsis genome was completed