

II.4 AraCyc: Overview of an Arabidopsis Metabolism Database and its Applications for Plant Research

S.Y. RHEE, P. ZHANG, H. FOERSTER¹, and C. TISSIER

1 Introduction

Metabolism is one of the most fundamental processes of life. Each organism possesses an intricate network of metabolic pathways, whose elaborate regulatory circuitry may be developmentally programmed and hard-wired to respond to changes in the environment. With the release of the fully sequenced plant genomes of *Arabidopsis* and rice (AGI 2000; Goff et al. 2002; Yu et al. 2002), and the initiation of many sequencing projects of other plant species, there is a growing desire to place the sequenced and annotated genomes in a metabolic context. AraCyc (<http://arabidopsis.org/tools/aracyc/>) was the first plant organism-specific metabolism database to be computationally predicted by the PathoLogic component of the Pathway Tools software using MetaCyc as the reference database (Mueller et al. 2003). With continued manual curation, the goal of AraCyc is to describe the complete set of metabolic pathways for *Arabidopsis thaliana* whilst placing genes and enzymes within their metabolic context. Though many enzymes in AraCyc have yet to be manually curated, most of the pathways have been manually validated and it is so far the only genome-wide, comprehensive metabolic database for a single plant species (Zhang et al. 2005).

The benefits of a species-specific metabolic pathway database are substantial: (1) it depicts the biochemical components of an organism; (2) it aids in comparative studies of pathways across species to facilitate metabolic engineering to improve crop metabolic traits; (3) it can be used as a platform to integrate and analyze data from large-scale experiments such as gene expression, protein expression, or metabolite profiling; finally (4) by presenting pathway steps lacking assigned genes, or having genes assigned but solely based on computational prediction, it allows the identification of the biochemical steps that remain to be identified and experimentally characterized. The manual, de novo creation of a pathway database can be labor intensive and time consuming. SoyBase (<http://www.soybase.org/>) is the only other plant pathway database, specific to soybean, which was manually created and made publicly available. Alternatively, there are metabolic pathway databases that cover a wide range of organisms. Examples of comprehensive pathway databases include Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>) (Ogata

¹ Carnegie Institution, Department of Plant Biology, 260 Panama Street, Stanford, CA 94305, USA, e-mail: rhee@acoma.stanford.edu

et al. 1999; Kanehisa 2002; Kanehisa et al. 2004), Enzymes and Metabolic Pathways (EMP, <http://www.empproject.com/>) (Selkov et al. 1996), and MetaCyc (www.metacyc.org) (Krieger et al. 2004). Each has its strengths and weaknesses, some of which have been reviewed (Maranas and Burgard 2001; Kanehisa 2002).

In this review, we describe the content and functionalities of AraCyc database as well as examples of applications that use the information contained in the database, in conjunction with functional genomics data to address systems-wide questions about metabolism. In addition, we discuss the current limitations and future directions of the database.

2 Database Content

AraCyc (version 2.5) currently features 197 pathways, comprising 979 unique reactions and 1071 compounds. Over 63% of the reactions have Arabidopsis genes/enzymes assigned and 1759 unique genes are assigned to the pathways. A metabolic pathway is a set of one or more enzymatic transformations, involved in processes such as biosynthesis, degradation, conversion,

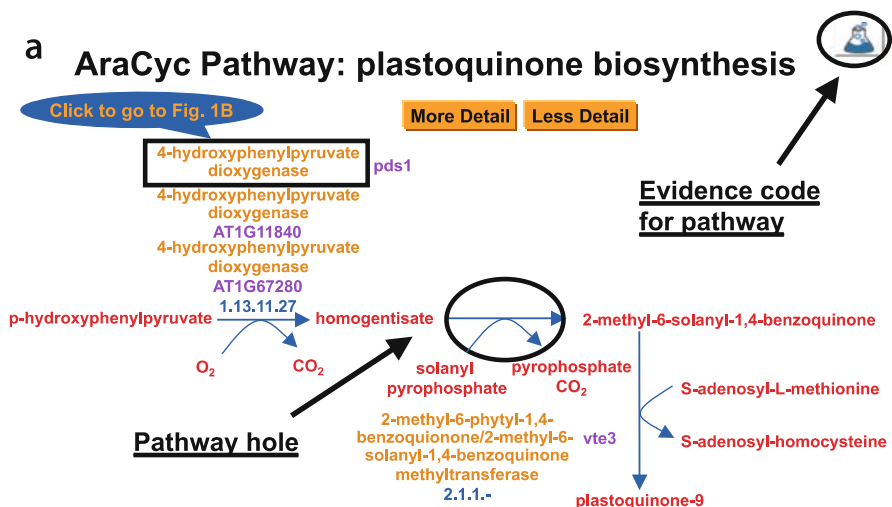


Fig. 1. An example of an AraCyc pathway: a pathway evidence, which could be either computational (indicated by a computer icon) or experimental (indicated by a flask icon), provides assertion of the existence of the pathway in Arabidopsis; b (see next page) similarly, evidence attached to an enzyme provides assertion of its catalytic activity involved in a specific reaction. Each piece of evidence is associated with a citation where the source of the evidence can be found (inset). Pathway can be zoomed to show various levels of details. Compounds, reactions, enzymes and genes on a pathway detail page are clickable to retrieve more information. (Reprinted with permission from Plant Physiology)