

Building A Network of Plant Metabolic Pathway Databases and Communities

Project Summary

This project aims to create a network of plant metabolism databases. The proposers will draw on their expertise in annotating genomes, generating metabolic pathway databases, curating biochemical information from the literature, and forming extensive network of collaborations with biological databases and biochemistry researchers. At the center of the network will be PlantCyc containing pathways from many plants, supported by experimental evidence for pathways, reactions or enzymes. PlantCyc will be initialized from currently available plant metabolism databases such as AraCyc, TomatoCyc, RiceCyc, MedicagoCyc and Soybase. It will be used as a reference database (in conjunction with MetaCyc) to create plant pathway genome databases (PGDBs) for 23 organisms with substantial sequence data. To build a PGDB, putative enzyme sequences will be identified for each organism using several sequence analysis methods and Pathway Tools software will be used to generate the initial PGDBs from the annotated sequences. As each PGDB is built, all of the pathways and enzymes in the new PGDB will be validated and added to PlantCyc, and subsequently curated. Therefore, with each round of PGDB prediction, the quantity and quality of PlantCyc will be increased. The project will leverage the curation teams at other databases interested in different species as well as biochemistry experts who are interested in specific domains of metabolism in improving the content of PlantCyc and the PGDBs. All of the data will be made freely available and updates will be released on a regular basis.

Intellectual Merit: As the worldwide demand for production of biofuels, food, animal feed and new medicines continues to grow, there is an increasingly urgent need to develop new technologies using plants. The long-term goal of developing these technologies has prompted the sequencing of plant genomes and gene complements. To date, genome and mRNA sequences are available for a number of representative plant species and many more are underway. There is a growing need to place the sequenced and annotated genomes in a biochemical context in order to facilitate discovery of enzymes and engineering of metabolism. However, only a handful of these organisms have a large body of literature and corresponding community-level databases. This proposal will generate an infrastructure for comprehensive plant metabolism information. Metabolic databases built from this project will address the need of the scientific community to store, analyze and display the growing body of data that is emerging from both conventional biochemistry and high-throughput/large-scale data experiments. The proposed network of databases will facilitate the discovery of new enzymes and pathways, and the engineering of metabolic pathways. In addition, researchers studying individual genes and mutants will be able to put their findings in the context of overall metabolic scheme of an organism, thereby deepening their understanding of the roles played by their genes of interest.

Broader Impacts: Developing a comprehensive infrastructure for a large domain of knowledge is difficult for a single group to achieve. This project aims to foster cooperation and sharing of resources and expertise ranging from genome annotation databases and metabolism databases to individual biochemists by creating a set of plant metabolic databases and engaging experts in specific domains to participate in their improvement. This project will create a network of database curators and biochemistry researchers via organizing annotation jamborees, establishing data exchange mechanisms between databases and forming a board of editors and reviewers. The resulting network of databases will not only facilitate research, but also provide an up-to-date, easily accessible set of teaching and educational resources for students. In addition to the metabolic pathway diagrams and detailed descriptions of the enzymes and the pathways, this project will generate a set of concise overview diagrams of the main plant compound classes, comprehensive list of glossaries and tutorials as teaching tools. All of the resources will be made freely available and presented in user-friendly ways via the web using a standard web browser.