Dated: January 7, 2010

Ref: Letter of Collaboration with MaizeCyc metabolic pathway database

Dear Andrew,

We are writing this letter as PIs and co-PIs of the MaizeGDB (www.maizegdb.org) and Gramene database projects (www.gramene.org) to outline our desire collaborate on your proposal entitled "PGRP: Comparative genomics-driven discovery of B vitamin pathways in maize". We appreciate your having contacted us in a timely way to plan the coordination of your proposal with our ongoing efforts in maize genome annotation.

As you are aware, our plans at Gramene and MaizeGDB for functional annotation of maize metabolic genes are currently centered on MaizeCyc and reported metabolic pathways. Similar to RiceCyc and SorghumCyc (available from http://gramene.org/pathway/), MaizeCyc (to be available soon) will catalog known and predicted biochemical pathways from maize. Pathways and genes in this catalog are based mainly on the annotations carried out by the Gramene database project, which are computationally generated via sequence-inferred homology and HMM predictions. MaizeCyc pathways consequently will be incomplete and may contain errors that require expert manual curation. When fully implemented, MaizeCyc will provide a framework for researchers to search for known and/or predicted pathways and find current annotations. Together with known annotations gaps and anomalies in the computationally derived datasets will become evident. These will require identification and further experimental work.

There is a powerful complementarity between your comparative genomics-based approach and the MaizeCyc effort. By combining prokaryote comparative genomic data and advanced metabolic reconstructions, your work will dovetail with ours by (i) locating and correcting errors in current pathway annotations, (ii) identifying novel candidates to fill pathway gaps that cannot be found by homology methods, and (iii) predicting functions for new metabolic genes, and verifying the function of others. Your initial focus on B vitamin-related metabolism is very welcome because this area is poorly known in plants compared to, say, carbon cycle, carbohydrate, and amino acid metabolism and hence stands to benefit greatly from application of comparative genomics.

The deliverables and timeline you anticipate are compatible with our plans. The datasets described below will be integrated by doing an incremental updates to the MaizeCyc to be made available via the Gramene and the MaizeGDB project sites. We also anticipate that gene based annotations will be added to the primary genome annotations provided by forthcoming maize genome sequencing and annotation projects including the community genome annotation underway as a joint project by the Lawrence and Brendel groups. We will work collaboratively with your group to setup data exchange protocols, formats and timelines coherent with the data release cycles of the respective projects. Specifically we will work with you to make available:

1. A dataset comprising all maize genes of unknown or uncertain function that have strong homologs in prokaryotes, with a link for each gene to the SEED database and, in cases where strong association evidence is found, a short summary of this evidence for inclusion in MaizeCyc. This dataset is expected to contain several thousand genes, and will be delivered in desired format in years 1 and 2.

2. Metabolic reconstructions and validation of primary metabolic pathways in maize, with special focus on B vitamin pathways (estimated to involve ~400 genes).
   - An independent set of annotations will be produced for the maize genome using RAST
     - The MaizeCyc and RAST annotations will be compared and reconciled collaboratively
   - A flux-balance-analysis-ready genome-scale metabolic model will be constructed from the reconciled annotation of maize

- A biomass reaction will be generated including essential cofactors and small molecule building blocks (e.g. lipids, cell wall components, amino acids, nucleotides)
- An optimization-based gap-filling process will be performed, which will identify the reactions/functions that must be added to the model to enable the synthesis of all compounds included in the biomass reaction from inorganic nutrients
- Flux balance analysis (FBA) will be utilized to identify active and inactive pathways in the metabolic model (and in the annotations by association), which will facilitate the identification of additional annotation gaps

About 80% of metabolic reconstruction data will be done in year 1, the remainder in year 2.

3. In years 3 and 4, experimentally verified functional annotations for 10 novel B vitamin-related genes, together with validated cDNA sequences if these are not already correct in the genome database.

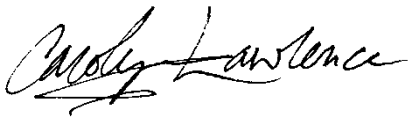
We understand that your project will have the necessary bioinformatics support and will provide us the datasets in formats we specify as they are ready for integration into the MaizeCyc. Together, our groups will establish a format for data exchange and update project information on a 6-month cycle with links back to your project database. We are pleased to know that unless otherwise specified, you will be the primary contact person to work with us on data exchange. We should note that the development and curation of the the MazieCyc is a deliverable specified in Gramene's NSF award (DBI # 0703908) and Gramene's collaboration with your project can only be guaranteed through the end of our current round of NSF funding (09/30/2011). MaizeGDB has long-term funding through the USDA-ARS and is committed to supporting this collaboration throughout your project's funded period.

In conclusion, we believe that your project is timely, that it has major potential for function discovery in an under-researched area of plant based B vitamins, and that it will enrich maize genome annotation. We wish the proposal success!

Sincerely,

Doreen Ware
USDA-ARS and Cold Spring Harbor Laboratories
Gramene Database PI

Pankaj Jaiswal
Assistant Professor
Dept. of Botany and Plant Pathology
Oregon State University
Co-PI Gramene database



Carolyn J. Lawrence
USDA-ARS Research Geneticist
Lead Scientist, MaizeGDB