Data Release, Sharing, and Management

Management of Data Produced by this Project: This project will generate several large, complex datasets including both short and long-read whole-genome sequences. I am well-versed in working with these types of data at these scales thanks to my dissertation work.

The University of Minnesota also houses institutional resources that allow the secure storage of this data with off-site back-up of the primary data and secondary analytical products.

Release of Data and Materials Produced by this Project:

Raw data/code will be released through public repositories, such as GitHub or NCBI. Processed datasets will be released through CyVerse Data Commons, Dryad, or DRUM (www.lib.umn.edu/datamanagement/drum). All processed data and any files that may be of interest to the broader community will be made publicly available through CyVerse or GitHub and will also be shared with MaizeGDB. Below, I provide specific details for each type of data/materials.

Sequence datasets:

The first aim of this project is to characterize specific changes in repetitive DNA content within maize x teosinte hybrids using long-read, whole-genome sequencing. The second aim of the project will generate short-read sequencing data using genotype-by-sequencing approaches. Raw datasets will be released through NCBI following quality control to ensure correct samples and properly associated metadata.

BAM, VCF, and associated metadata files (genotype, tissue, growth conditions, library preparation and sequencing conditions for all samples) generated will be released through NCBI's Sequence Read Archive (SRA). All code used to process raw data, call structural variants and identify differences between samples will be made available through GitHub and will be continuously updated to reflect the most recent version as we adapt our approach. Versions of all software used will be documented and shared for reproducibility purposes.

Manisha Munasinghe will be responsible for the release of this data.

Biological materials: The second aim of this project relies on natural sampling of additional teosinte plants from Valle de Bravo, Mexico. Seeds collected will be stored at the University of Minnesota within controlled environmental conditions at a secure site. Materials will be freely shared with other researchers as long as there is sufficient seed availability.

Seed stocks from the initial collections may be somewhat limited but we anticipate larger amounts of seed for advanced backcross generations. Stocks for the advanced backcross generations will be deposited at the Maize Genetics Cooperative Stock Center (***www.maizecoop.cropsci.uiuc.edu/***).

Evolutionary Models: Several evolutionary models will be developed throughout the course of Aim 3. All associated code used to simulate and evaluate models will be made publicly available through GitHub.

Manisha Munasinghe will be responsible for the release of this code.

Upon publication, all scripts and data summaries will be made available on dryad (datadryad.com) so that work is reproducible and accessible.