Supplementary Documentation

Sharing of Results and Management of Intellectual Property

Data Types

This proposal will generate sequence data, genotype, phenotype data, analytical software, teaching resources, germplasm, and publications.

Data Access, Sharing

All sequence data (RNA-seq, whole genome sequencing, and fastq files from genotyping by sequencing) will be submitted immediately upon completion of data quality control to the NCBI sequence read archive (SRA), along with passport information on each parent. A "hold until publication" embargo will be requested at the SRA. Before publication, data will also be made publicly available via the Figshare website (`www.figshare.com`), a free public website allowing dissemination and archiving of large datasets. Data will be released in accordance with the Toronto agreement (2009. Nature 461:168-170. `www.nature.com/nature/journal/v461/n7261/full/461168a.html`) under the stipulation that no whole-genome analyses be performed until we have published our initial analyses. RNA-seq data will include metadata as stipulated by MIAME (`http://www.ncbi.nlm.nih.gov/geo/info/MIAME.html`) and will also be deposited in the NCBI GEO database.

Phenotypic data and genotypes from sequencing and GBS will be uploaded to Figshare, along with appropriate metadata associated with other publications, links to germplasm, SRA experiments, Github code, etc. Phenotypic data will be recorded digitally in the field using the high-throughput techniques developed by Dr. Flint-Garcia. Data will be uploaded at the end of each day into the FieldBook database developed by Dr. Flint-Garcias USDA-ARS group and immediately backed up at a remote location. Data will be grouped into projects, and each project is associated with a unique digital object identifier (DOI). Drs. Ross-Ibarra and Coop have already used Figshare extensively to share and archive data, preprints, and code (see `http://figshare.com/authors/Jeffrey_Ross-Ibarra/98899` and `http://figshare.com/authors/Graham_Coop/101524`). Data on Figshare is publicly available and searchable. We will submit data as soon as we complete quality control, but again with explicit stipulations as to the analyses that the data can be used for prior to our initial publication. All appropriate metadata including plant ID, data collector, sequence run, field location, etc. will be associated with genotype and phenotype data deposited to Figshare.

Analytical software and code from this project will be hosted on Github, a version-controlled public git repository. Upon submission of papers all code will be made publicly available. Drs. Ross-Ibarra and Coop have already done this extensively (see `https://github.com/rossibarra`, `https://github.com/rilab`, and `https://github.com/cooplab`). Publication of all code will ensure reproducibility of all analyses conducted.

Presentations and teaching resources from our field workshop will be made publicly available via Figshare as well.

All data, code, and presentations will be made publicly available via a creative commons CC by 2.0 license (http://creativecommons.org/licenses/by/2.0/) allowing free access to reuse, redistribute, and modify, requiring only citation of the license and the original source.

All publications resulting from this project will be submitted to one or more preprint servers (e.g. arXiv, bioRxiv, PeerJ) such that they will be publicly available immediately upon submission of the paper for publication.

Data Archiving

All data, code, presentations, and publications will be made publicly available online (see above). Prior to public release, all data will be hosted locally. Dr. Ross-Ibarra will maintain a backup of all raw genotyping, sequence, and phenotyping data. His lab maintains a DROBO distributed backup server (currently > 8Tb of free space) which is robust to single disk failure. All analytical code will be hosted on Github, which maintains version-controlled backups, as private repositories until release.

Both our F2:3 families and our near isogenic lines will require multiple generations of development until they are mature resources for mapping traits related to highland adaptation. We will archive a sample from each generation of population development in temperature- and humidity-controlled facilities at Iowa State University and Langebio. Sample accession data will be securely stored in a MySQL server hosted at the University of California, Davis and backed up on a weekly basis offsite. International agreements prohibit some of the maize and teosinte germplasm collected in Mexico from being stored and distributed by USDA. We will, however, deposit small quantities of seed from all our collections with the CIMMYT germplasm bank in Mexico, and deposit samples of our mapping populations (F2:3 seed) in the USDA-ARS Maize Stock Center at the University of Illinois. Both centers provide public access to seed.