## Charge to Working Group

**The primary role of the Working Group is to help the MaizeGDB Team with strategic planning.** This is especially important now, as we hope to hire or re-train personnel, and as we start preparing for the next USDA-ARS 5-year plan next year. We would greatly appreciate both your comments on what we have presented today, and your guidance on the following four charges. We would appreciate your written report by March 31, 2016. Thank you for your ongoing guidance and support!

**Charge 1: Genome Assembly Stewardship.** Below is a list of question. If you agree that these items are important, could you please provide a priority or ranking to help us set priorities?

1. Should we devote curation effort into patch and new assembly releases for B73? We now have a tool to collect assembly errors. We plan to continue to collect them, and show them, unedited, on a browser track. Should we spend effort to vet these and create patch assembly releases?
2. Should we continue to collect new assemblies and related metadata?
3. If so, to what extent should we integrate the data into MaizeGDB; eg, make gene models pages for every set of genes? Provide a genome browser? Provide a Cyc view?
4. What should MaizeGDB's role be in encouraging researchers to submit genome assemblies to GenBank? We plan to require that genomes served at MaizeGDB be submitted to GenBank and are developing a metadata collection template to ease the submission process. The template will be made freely available to researchers at MaizeGDB.
5. Should we map any data from B73 RefGen_v2/v3 to the v4 assembly? B73 RefGen_v2 had 58 tracks of data. Twenty-six of those were computationally mapped onto v3, while only 11 tracks were mapped directly to v3. Tracks are listed here http://www.maizegdb.org/gbrowse under the "select track" tab.

**Charge 2: Big Data set identification, evaluation and incorporation**. Below is a list of question. Can you offer advice and suggestions?

1. How do we triage Big Data sets reported in the literature? It is becoming difficult to decide which large data sets will be useful to the maize community. This includes data sets that could be accessible by a genome browser, incorporated into MaizeMine, or by special genotype-phenotype queries or other tools. Can you suggest new ways to evaluate large datasets? Ways to get community help to recommend data?

2. Our Project Plan Objective 2 is to incorporate experimentally confirmed functional genomic annotation including: GO annotation, phenotype/trait annotation, quantitative trait values, and Metabolic Pathway data. This requires extensive manual literature curation. Can you suggest better, faster, less labor intensive ways to accomplish this objective? Possibilities include using the Editorial board, automated literature annotation, work with journals to get authors to submit pre-publication, develop and send templates to authors post-publication, and more.
3. How valuable is comprehensive integration of public QTL and GWAS data at MaizeGDB? Currently we integrate into MaizeGDB a subset of available trait scores with metadata, but do not add researcher-defined QTL loci, either defined by a SNP or more loosely by a genetic region.

**Charge 3: Tool Development**. Below is a list of question. If you think these items are important, could you please provide a priority or ranking to help us make priorities:

1. Should we improve existing tools to visualize and access large-scale diversity data (genotype, SNP and GBS data) such as haplotype viewers?
2. Should we transition to the JBrowse genome browser to handle larger datasets?
3. Should we improve query tools that use the hierarchical nature of ontologies with regard to phenotypes, mutants and genes? For example, this tool would allow searches for mutant phenotype with parent term “leaf” to return phenotypes from all child terms (ligule, sheath, blade, margins, etc) as well as whole leaf phenotypes.
4. Should we update curation tools in collaboration with the Maize Genetics Stock Center? This would allow easier, faster manual literature curation.
5. Should we find ways to better integrate information about Mu and AC tagged sequences into all areas of MaizeGDB? For example, add available tagged alleles to gene model pages?

**Charge 4: Future needs and expectations.** Below is a list of expected needs in the next 5 years. Can you identify more, and can you comment on the importance of each of these?

1. Identifying and storing Genomes to Fields (G2F) type data (linking genotype and environment to phenotypes)
2. Finding a path to a pan genome infrastructure
3. Using large-scale semi-automated literature curation, such as annotation by publisher and authors check as part of proof-reading, etc.
4. Multiple maize genome comparison and display