**Response to IWGPG request for input re NPGI Five Year Plan**

**Maize Genetics Executive Committee**

**March 26, 2013**

**Executive Summary**

The grand challenge to produce as much food in the next 40 years as in the previous 10,000 years requires rapid deployment of research discoveries from the laboratory to the field. As the leading U.S. commodity with a rich history of foundational discoveries from genetics research, maize must be a top target. Investments made under the NPGI in the last two decades have yielded the core genomic resources to make maize the pre-eminent monocot model for translational research. When genome sequencing was expensive, C3 rice had a role in basic plant research; however, it is a poor model for conducting translational research under U.S. environments and field growth conditions. Therefore, C4 maize dominates and offers direct translation of scientific discoveries to increased field productivity critical to sustain our production of food, fiber and fuel. This unique system that spans basic to applied research is underpinned by investments by federal funding agencies that have resulted in a completed genome sequence, a detailed understanding of natural variation, and a growing set of resources for reverse-genetic analysis of gene function. The maize community through surveys and community forums has identified key areas in which resources are needed to capitalize on these past investments and continue to increase the productivity of this important crop plant. Anchoring these needs are a strong, global infrastructure for data management that includes quality metrics and public access; foundational tools/resources (facile maize transformation, gene targeting, functional annotation, rapid phenotyping); and grants/fellowships to support the cross-training of students in interdisciplinary research. This overarching infrastructure is critical as we now move toward translating genomic data to an improved understanding of plant processes and predictive models of how genetic variation will influence field performance. The NPGI has proven to be an outstanding mechanism for guiding plant genomics from basic discoveries to deployment of science-based production agriculture. The Maize Genetics Executive Committee appreciates the opportunity to offer suggestions for the 2014-2018 plan.

**Foundational Tools**

Past investments by federal funding agencies in the U.S. have resulted in the development of foundational tools for research in maize. However, surveys of the maize research community make clear the community’s broad agreement that additional such tools are necessary to enable basic, translational, and applied research. In a 2012 survey of the maize community there was very strong support for (1) increased functional annotation of maize genes, (2) a more robust system for transformation and production of transgenic maize, and (3) enhanced reverse-genetics resources. Support for a multi-national, well-coordinated ZeaENCODE-like project would provide valuable functional annotation data for the maize genome that would be widely used in the maize community. An improved ability to transform maize would increase our ability to probe the phenotypic consequences of altering genes, enabling the targeted creation of precise changes (including both transgenic and non-transgenic alterations to the genome - e.g., TALENs) to identify and provide useful agronomic characteristics. A more complete reverse-genetics resource that provides sequence-indexed mutations in nearly all maize genes would enable the understanding of gene function in maize in a manner that is currently unavailable. There is also a rapidly growing demand for high throughput phenotyping tools to enable large-scale field trials for agronomically meaningful traits. However, traits screened by biochemical or physiological assays typically require destructive sampling of a specific tissue that is transported to and assayed in the lab. These more traditional methods are simply not feasible for large sample numbers. Some measurement systems have recently become available to improve field measurements for complex physiological traits, but additional new tools are required for rapid, non-destructive sampling in the field. Together, these tools would revolutionize the pace by which the products of maize research can be deployed into farmers’ fields.

**Integrating Plant Processes**

A major goal for the post-genome sequencing era is to gain a full and holistic understanding of how plants grow, develop, and reproduce. This information will integrate genome datasets of the last decade to help translate the knowledge of sequenced-based plant biology to crop improvements in the field. Such a goal can be accomplished only if we have a comprehensive understanding of plant processes that span from the cellular to the organismal level throughout the plant life cycle. This effort will require a detailed developmental and physiological understanding of how cells function to make tissues, how the tissues are organized into structures and organs, how changes in metabolic flux influence cell functions, and how these processes coordinate to produce an integrated whole. Starting with the farmer planting a seed, we need to determine the regulatory mechanisms of seed germination and seedling emergence at the cellular and physiological level. What are the different tissues in the seed? How are they coordinated? What prevents or delays germination? How are activities of the root, shoot, and endosperm coordinated? As the plant grows and produces seeds, we need a mechanistic understanding of how leaves and branches initiate, how cells remember their developmental fate, how physiology is coordinated at the tissue and whole plant level, how metabolic resources are allocated, and how biotic and abiotic factors combine to trigger growth, development, and reproduction. The information needed to completely describe these plant processes will couple careful observations and experiments starting at the microscopic level with new information about the genome, regulome, transcriptome, proteome, and metabolome. Cross-disciplinary approaches will be needed to generate, analyze, and synthesize the large datasets that result from deep sequencing at the genetic and epigenetic levels. When integrated over developmental time, in different tissues, and under different environments, this information will provide a basis for selecting traits of interest by targeted phenotyping and for validating predictive models. The resulting knowledge will lead to new decision tools for improving maize and other economically important plants.

**Genomes in the field**

Genomic sequencing technology is providing tremendous insight into how the natural variation of maize performs on average across environments. However, as the great drought of 2012 illustrated: *there are no average environments*. While a couple of advanced breeding efforts showed promise in this arena, we do not have varieties of maize that withstand the full range of climatic variation that we will likely encounter over the next decades. This lack of varieties stems from a lack knowledge in four areas: (1) how genetic variation performs in specific environments, (2) how physiology varies across genetic and environmental variation, (3) how to develop crop models that integrate genetic variation, physiology, and environments together, and (4) how to develop genome wide predictions that include biological and physiological mechanisms. A better understanding of the basic science of how genomes perform in their environment will facilitate tremendous applications to agriculture, and with side benefits in medicine and conservation.

We advocate developing a research ecosystem that promotes extensive research in how genetic variation performs across environmental variation. Key aspects of the proposed research ecosystem would be (1) extensive field trialing with ~100 environments per year for 10 years (substantial portion of the locations would be through private sector collaboration), (2) integrated genomic and physiological analysis, (3) development of high throughput and innovative phenomics tools to measure physiology and the environment, (4) core national effort to make germplasm available, run central environments, and provide bioinformatics to enable numerous group projects focused on physiology, agronomy, genetics, and modeling to coordinate their efforts.

This is the time to make the shift to understanding how genomes perform in the field. Maize is already a superior model species for population and quantitative genetics and serves as the preeminent model for outcrossing plant species. A national effort initiated at this time would leverage new public personnel in breeding and newly emerging, amazing technologies to achieve both basic and translational research in how genomes interact with the environment.

**Coordination and Informatics**

In order to select target genes for functional characterization and to effect crop improvement, significant coordination is required, not only by researchers involved in generating data through diverse, independent projects, but among informatics groups that manage and analyze genome-scale information. This translates into coordination and broad data management needs that involve documentation of ongoing project goals and timelines as well as data including, e.g., genomics, functional genomics, and diverse annotations of phenotype. These needs can be met by developing, documenting, and requiring the use of gold standard datasets and best practices. The availability and pervasive use of such guidelines enables broad access to materials that serve as benchmarks to check computational pipelines and supports effective data collection that enables the simultaneous use of diverse datasets generated by multiple project teams. To benefit from the work of others in this area, engagement with outside groups, notably the NSF-funded Genomic Standards Consortium (GSC; <http://gensc.org/gc_wiki/index.php/Main_Page>), already deeply involved in these sorts of endeavors is key. It is noteworthy that the GSC has created a peer-reviewed venue called Standards in Genomic Sciences (SiGS;<http://standardsingenomics.org/>) to report work in this area, thus ensuring that such foundational endeavors are encouraged and valued by academic standards (e.g., publications to develop and deploy standards count toward tenure review). The GSC has made requests for plant researchers’ involvement in their work already. By participating to enable the expansion of existing gold standards and best practices to accommodate plant-specific requirements, engagement by maize researchers will benefit both plant biology and the GSC. The maize community database MaizeGDB (<http://www.maizegdb.org>) is committed to serving as a centralized clearinghouse for access to maize-specific standard practices by deploying a “Standards” data center with linkages to outside groups and datasets that are not specific to maize, where appropriate and available.

**Cross-training**

Technological advances in genomic tools have fundamentally and rapidly changed how students must be educated and trained for careers in agriculture and life sciences. Team-building, communication and analytical skills essential for success in today’s and tomorrow’s complex research endeavors must overlay foundational disciplinary knowledge in core areas of biology, genetics, genomics, statistics, physiology, etc. The maize community has long embraced and established a strong track record of success stories for cooperative research, sharing of resources and mentoring of students, postdocs and junior faculty. Meeting today’s global challenges requires a workforce able to move seamlessly across traditional disciplinary boundaries to use large datasets, develop and improve analytical tools and expand our understanding of fundamental biological, physiological, biochemical and developmental processes. Support through research assistantships and training grants for students, postdoctoral grant programs and re-tooling opportunities for established faculty are critical for building interdisciplinary bridges and preparing a new generation of researchers who can leverage existing resources and build new ones to solve complex crop production problems.

**International** **Challenges and Opportunities**

The maize community is fortunate to have such a tremendous community of scientists who support the long-standing tradition of cooperation that is made possible by intentional collaboration and resource-sharing. While in the past this has been mostly focused among U.S. scientists, research in China, Europe, and the developing world plays an increasingly important role in the advancement of maize genetics. Although this global community of scientists has expanded the horizon for opportunities that promise to accelerate scientific advancement, lack of international agreement on germplasm access and data sharing impede effective collaboration and prevent worldwide resources from being leveraged most effectively. Specifically, agreements need to be in place to share genetics stocks that currently are encumbered by MTAs. Although advanced breeding stocks are anticipated to maintain IP limitations, uniform methods to enable sharing older (>20 years) breeding germplasm are needed. Additionally, genomic-phenomic resources (whole genome sequence, RNA profiles, and field phenotypes) are not currently shared through publically recognized venues (e.g., <http://www.insdc.org/>; the International Nucleotide Sequence Database Collaboration, INSDC, that coordinates the activities of DDBJ, ENA, and GenBank) by many groups outside the U.S. Alignment of U.S. open data approaches with European and Chinese research would help even the playing field in this arena to accelerate science (as recognized by the G-8 through their upcoming International Conference on Open Data for Agriculture; see <https://sites.google.com/site/g8opendataconference/home>). Many of the issues with data and germplasm sharing are not caused by individual scientists, but rather by high level policy decisions with consequences that inhibit science. Policies that support and even require international data and germplasm sharing are needed to support the success of future international agreements.

**Diversity of Grant Portfolios**

Global access to public genomic resources and analysis tools continues to generate tremendous returns on investment as scientists mine datasets and develop hypotheses about important biological processes. Creative researchers working at small colleges or as individual principal investigators provide the spokes extending from this hub of genomic resources to test hypotheses, provide excellent training for students and postdocs, and make seminal advances in plant sciences. Such leveraging of public data, however, requires that granting agencies support single investigator awards for smaller scale projects that draw upon rather than generate genomic resources. We strongly support individual research grant programs that fund established as well as new investigators.