Breeding and Modern Genomic Tool Integration

Warren Kruger, Monsanto, 3302 SE Convenience Boulevard, Ankeny, IA 50021
warren.m.kruger@monsanto.com

Overview

The new challenge for breeders today is to explore and find ways to apply and/or leverage (gen)omic datatypes in a breeding context to augment and improve agronomic performance to meet global demand for food, feed, and fuel (Edgerton, 2009). The rationale for the substantial growth in "omic" data (and financial support) accumulated over the last decade has centered on developing a fundamental understanding and characterization of the function of each gene and the molecular processes underlying phenotypes of interest (e.g. NSF funded Arabidopsis 2010 project beginning in early 2000 to determine the function of all genes in the model plant *Arabidopsis thaliana* by the year 2010). This fundamental level of understanding molecular processes has not historically been the domain of plant breeders, although the competitive need to "drive" complex selection models within breeding has brought functional genomics into focus. When analyzed collectively across the (sub)disciplines of genomics there is a rich repository that a breeder can query to gain insights into linkage, gene level expression, gene expression and metabolite networks. Genomic information can inform correlations between different traits, and provide insight into linkage effects. Interaction networks assembled from transcript or metabolite profiling by way of systems biology can illuminate pleiotropy and epistasis, and these phenomena are squarely within the domain of breeders. There is an ever expanding universe of (gen)omic data about fundamental characteristics of plant genes which has been and continues to be realized with the advancements made in sequencing, transcript, protein and metabolite profiling. While a lot of information is in proprietary databases, the amount of data collected (and in process) within the public domain for model and crop plants is substantial and should not be ignored (for review see Childs, 2009).

For genomic technologies to be widely adopted in a breeding context it is not enough to show "it works", it must be scalable, i.e. cheap, robust, and integrates (or can be integrated) within breeding schemes and IT systems. A similar scenario played out during the development and deployment of marker assisted breeding applications. This has been nicely summarized by Dick Johnson (2007), "It was apparent, however, that to achieve success with marker-aided breeding, half-way measures were inadequate. Marker- aided breeding needed to be integrated with all phases of hybrid and inbred line development .... Achieving the integration, however, required massive investment in genotyping capabilities, including tissue sampling, DNA extraction, robotics, automated marker scoring, and information management systems for collection, integration, analysis, and distribution of data from performance trials, nurseries, and laboratories ...". The utility of marker assisted breeding has gained wide acceptance with private and public breeders as markers can improve the rate of gain for yield and associated traits such as grain moisture and stalk lodging (Eathington *et al.*, 2007). The challenges confronting integration of genomics and breeding are the same and occur in both the input or data acquisition and the execution or output streams.

On the input side the challenge is to (i) demonstrate opportunity and utility for breeding, and (ii) improve “scalability” in acquiring data, i.e. drive down the cost and make it robust. Rapid evolution in the domain of high throughput (HT) sequencing and the challenge to get to a $1000 genome are leading to dramatic gains in efficiency of this technology, and it is likely that further technological improvements will follow in HT sequencing and other “omic” technologies.

Equally important on the output side, the challenge is (i) a cost effective and robust way for deployment in breeding, (ii) IT tools/Information systems to run a large scale (proprietary) breeding effort with this information, (iii) technical skills and training to support and help people understand and work with this information, and (iv) proven applications that deliver cost effective gain.

An emerging GAB (Genomics Assisted Breeding) paradigm would need to integrate “omic” data with the framework for genetic and phenotypic data presently deployed within breeding. A fully integrated (genetic and physical) map can serve as the reference medium for connecting and integrating the domains of breeding and genomics. It is routine to use integrated maps as the basis for positioning data types as diverse as DNA sequence and gene structure, to genetic marker position and trait associations (see public domain examples - Gramene, Maize Genome Browser). This essentially provides the first order connectivity between the genomic (expression maps of transcript and/or metabolite and their molecular interactions) and the genetic (linkage maps of markers and traits) contexts. As higher orders of information are added (e.g. high throughput sequencing of gene level allele variation) the “read through” access to this vast *a priori* information can provide breeders with insight on linkage, pleiotropy, and interaction effects. In return breeders are able to inform and execute a selection and recombination strategy by way of a competent and well established marker assisted selection infrastructure within breeding.

References

Childs, K.L. 2009. Genomic and Genetic Database Resources for the Grasses. Plant Physiology 149:132-136.

Eathington, S., Crosbie, T.M, Edwards M.D., Reiter R.S., and Bull J.K. 2007. Molecular markers in a commercial breeding program. Crop Science 47(Suppl3): S154–S163.

Edgerton, M. 2009. Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel. Plant Physiology 149:7-13.

Johnson, G.R. 2007. The Industrialization of Corn Breeding. In M.S. Kang and P.M. Priyadarshan (eds.) Breeding Major Food Staples , Blackwell Publishing, Oxford, UK.