

1st National Conference on Biotechnology  
Lima, Peru, 12 May 2009

## **Genomics in the Gramineae and Application to Productivity and Nutritional Quality for Sorghum, Corn, Sugarcane, and Rice**

Joachim Messing, Waksman Institute, Rutgers University, New Jersey, USA

### **Summary**

Ultimately, important agricultural traits are written into the genes of the genomes of our crops. Therefore, the recent efforts to sequence the entire genomes of these crops hold the great potential to uncover the complicated interactions of genes and their regulation that make up the desirable properties, often optimized only after long periods of breeding if at all. Rationale approaches to accelerate breeding by targeting specific genes by genetic engineering requires an understanding of the hierarchical organization of gene function and specificity. While sequencing entire genomes cannot bring us the answers to these questions directly, the first step of such rational approaches is to take inventories of all the genes present in any of the genomes of our most important crops and learn by comparison how characteristic features of individual crops might relate to differences in their organization and differential amplification of genes. Such comparisons then should permit us to focus on the functional characterization of a subset of genes that could be ectopically expressed in their host plants for achieving new desirable traits in crop plants.

Even before sequencing entire genomes, the use of conserved DNA probes has shown that genetic maps of species within the same taxonomic family exhibit large contiguous collinear regions as a mosaic of synteny blocks. These findings implicated that speciation involved polyploidization, chromosome breakage and fusion events, neocentromere formation, loss of centromeres, as well as formation of telomeres. As a result the chromosome numbers and ploidy

of closely related species have changed, but progenitor chromosomes could be reconstructed via the junction of these chromosome breaks using collinearity.

One of the most important agronomical taxonomic families is the *Poaceae* or the grasses, which is also one of the largest with over 10,000 species. This family includes crops like wheat, rice, maize, sorghum, sugarcane, barley, millets, and the biofuel prospects of *Miscanthus* and switchgrass. Because species within the same family have collinear gene order by descent as discussed above, sequencing genomes within the same family can also be used to determine gene amplification and deletion. Such an analysis can reveal gene-copying events in specific crops that potentially acquired new properties by subfunctionalization of gene copies. However, sequencing entire genomes has been hampered because of their variable sizes, which is largely due to the differential amplification of transposable elements. Consequently, initial focus has been on sequencing a crop with one of the smaller genomes first. Rice has 12 chromosomes and a total length of 390 million base pairs (Mb), representing one of the smallest genomes among crops, and has been sequenced as the first monocot species. Sorghum with 730 Mb has also been sequenced and maize with 2,300 Mb is also nearly complete. Another grass genome, even smaller than rice with 300 Mb, that has recently been sequenced is *Brachypodium* belonging the *Pooideae* subfamily as a reference for the large wheat and barley genomes. Critical to sequencing any of these genomes has been the construction of a physical map so that sequences can be assembled in a contiguous and ordered fashion, which is critical for the comparative analysis. This presentation seeks to review the status of genome sequencing, highlights of the comparative analysis, and the differential amplification of one gene family, the prolamins, which composition and regulation determine the nutritional value of our major crops.

## References

- Bolot, S., Abrouk, M., Masood-Quraishi, U., et al. (2009). The 'inner circle' of the cereal genomes. *Curr Opin Plant Biol* 12, 119-125.
- Gardner, R.C., Howarth, A.J., Hahn, P., et al. (1981). The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucleic acids research* 9, 2871-2888.
- Goff, S.A., Ricke, D., Lan, T.H., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296, 92-100.
- Huo, N., Vogel, J.P., Lazo, G.R., You, F.M., Ma, Y., McMahon, S., Dvorak, J., Anderson, O.D., Luo, M.C., and Gu, Y.Q. (2009). Structural characterization of *Brachypodium* genome and its syntenic relationship with rice and wheat. *Plant Mol Biol* 70, 47-61.
- Kellogg, E.A. (2001). Evolutionary history of the grasses. *Plant Physiol* 125, 1198-1205.
- Maizel, J.V., Jr., and Lenk, R.P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc Natl Acad Sci U S A* 78, 7665-7669.
- Matsumoto, T., Wu, J.Z., Kanamori, H., et al. (2005). The map-based sequence of the rice genome. *Nature* 436, 793-800.
- Messing, J., and Dooner, H.K. (2006). Organization and variability of the maize genome. *Current opinion in plant biology* 9, 157-163.
- Messing, J., Crea, R., and Seeburg, P.H. (1981). A system for shotgun DNA sequencing. *Nucleic acids research* 9, 309-321.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551-556.
- Sanger, F., Coulson, A.R., Friedmann, T., et al (1978). The nucleotide sequence of bacteriophage phiX174. *J Mol Biol* 125, 225-246.
- Segal, G., Song, R., and Messing, J. (2003). A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. *Genetics* 165, 387-397.
- Smith, L.M. (1991). High-speed DNA sequencing by capillary gel electrophoresis. *Nature* 349, 812-813.
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674-679.
- Song, R., and Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. *Proceedings of the National Academy of Sciences of the United States of America* 100, 9055-9060.
- Song, R., Llaca, V., Linton, E., and Messing, J. (2001). Sequence, regulation, and evolution of the maize 22-kD alpha zein gene family. *Genome research* 11, 1817-1825.
- Tanaka, T., Antonio, B.A., Kikuchi, S., et al. (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res* 36, D1028-1033.
- Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259-268.
- Wei, F., Coe, E., Nelson, W., Bharti, A.K., et al. (2007). Physical and Genetic Structure of the Maize Genome Reflects Its Complex Evolutionary History. *PLoS Genet* 3, e123.
- Xu, J.-H., and Messing, J. (2008a). Diverged Copies of the Seed Regulatory Opaque-2 Gene by a Segmental Duplication in the Progenitor Genome of Rice, Sorghum, and Maize. *Mol Plant* 10.1093/mp/ssn038 1, 760-769.
- Xu, J.H., and Messing, J. (2008b). Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc Natl Acad Sci U S A* 105, 14330-14335.
- Yu, J., Hu, S., Wang, J., Wong, et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296, 79-92.