

Recent advances in crop genomics for ensuring food security

Mehanathan Muthamilarasan, P. Theriappan and Manoj Prasad

Prevalence of hunger and undernutrition among the world population at alarming rates challenges the food security. According to FAO Hunger Report 2012, about 12.5% of the global population (one in eight people) is starving, excluding 100 million children under the age of five. About 2.5 million children die every year due to starvation (<http://www.fao.org/docrep/016/i3027e/i3027e00.htm>). Hence ensuring food security is an immediate requirement for saving the global human potential, where the role of crop genomics is substantial. The remarkable innovations in high-throughput technologies had significantly accelerated the crop genomic research, of which, we highlight the recent advances reported towards improving the crop plants for ensuring global food security (Box 1).

Primarily, whole genome sequence serves as the blueprint for crop improvement by facilitating the scientific community towards understanding the functional genomics mostly related to yield, stress-tolerance, engineering the crops with important agronomic traits, etc. The relation between the genome sequence information and the respective genomics is illustrated in Figure 1. The release of reference genome sequences for many important crops and model plants holds the potential to aid in the understanding of the enduring promise of plant genomics to significantly accelerate crop improvement¹ (Table 1). Though the model plant *Arabidopsis thaliana* was the first eukaryote to have its whole genome sequenced², presently sequenced animal genomes (> 100) outnumber the plant genomes sequenced (< 50). In 2012, the whole-genome sequence for three distinct crops was released – (i) foxtail millet (*Setaria italica*)^{3,4}, a tractable model crop for studying the functional genomics of bioenergy grasses⁵; (ii) bread wheat (*Triticum aestivum*)⁶, the world's largest cultivated cereal and staple food, produced in broad-spectrum climatic environments and geographic regions⁷, and (iii) barley (*Hordeum vulgare*)⁸, a predominantly used animal feed which also contributes about 30% of the calories consumed by global human population⁹. In 2013, the draft genome sequence of chickpea (*Cicer arietinum*)

was released by two independent groups^{10,11}. The whole-genome sequence data serve as the foundation for biotechnological research aimed at crop improvement by promoting large-scale development of molecular markers for linkage mapping, association mapping, wide crosses and alien introgression, epigenetic modifications, transcript profiling, population genetics and *de novo* genome/organellar genome assembly¹². Noteworthy, the whole-genome sequences of two wheat progenitors, namely *Triticum urartu* (ancestor of the A genome) and *Aegilops tauschii* (the D donor) were also released recently^{13,14}, which would hence serve as a powerful tool for the study of complex, polyploid wheat genomes and a valuable resource for genetic improvement of wheat.

One well evidenced application on the utility of reference genome sequence was the genome-wide mining of molecular markers. Recently, Pandey *et al.*¹⁵ performed an *in silico* genome-wide scanning for the presence of microsatellite markers using the draft genome sequence of foxtail millet and reported about 28,342 microsatellite repeat-motifs spanning 405.3 Mb of the foxtail millet genome. A comprehensive cross-genera transferability test across millet, non-millet and bioenergy grass species showed a higher transferability percentage in bioenergy grasses (~ 79%, switchgrass and ~ 93%, pearl millet)¹⁵, which indicates the importance of the developed microsatellite markers in studying germplasm characterization, phylogenetics, construction of genetic linkage maps for gene/quantitative trait loci discovery and comparative mapping in bioenergy grass species. Of the 28,342 microsatellite markers, about 15,573 microsatellites were physically mapped on nine chromosomes of the foxtail millet genome to generate a high-density physical map which would serve as a potential reference for rapid selection of microsatellite markers either across genome-wide/chromosomal level or at the targeted chromosomal location¹⁵. Thus it would accelerate the construction of high-density genetic linkage map and fine mapping of genes/Quantitative Trait Loci (QTLs) for important agronomic traits in foxtail millet.

Exploring the genomic information for genetic and breeding applications will be rapid in plants whose genome sequences are available, but the progress would be hindered in plants whose genome sequences have yet to be established. This impediment could be evaded by comparative genomics, where targeted application of phylogenetically related genome sequence resources offers a major opportunity for unsequenced plant species in efficiently investigating the genomic information for genetic and breeding applications. Recently, Pfeifer *et al.*¹⁶ presented a model of the perennial ryegrass (*Lolium perenne*) genome on the basis of conserved synteny to barley (*Hordeum vulgare*), and the model grass genome Brachypodium (*Brachypodium distachyon*) with rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*). A transcriptome-based genetic linkage map of *L. perenne* served as a scaffold to establish the chromosomal arrangement of syntenic genes from model grass species¹⁶. A high degree of synteny and macrocollinearity was divulged by this scaffold, which was then used to anchor *L. perenne* genes to their expected genomic locations. Thus the resultant *L. perenne* GenomeZipper included 4035 conserved grass gene loci that were utilized for a genome-wide sequence variance analysis between *L. perenne*, *H. vulgare*, *B. distachyon*, *O. sativa* and *S. bicolor*, which showed a well-organized, highly informative genome scaffold, expediting map-based cloning and genome assembly in *L. perenne* and closely related Poaceae species¹⁶. This study is deemed a landmark in unfolding the synteny between unsequenced *L. perenne* and completely sequenced model grass genomes, thus expanding the knowledge on genome organization and evolution in the most significant temperate forage and turf grass species¹⁶.

Next-generation sequencing assisted with high-throughput gene characterization and function analysis technologies has revolutionized the genome-decoding projects, thus resulting in the generation of large data on genomics and genetics. The application of the knowledge thus generated towards crop improvement is termed 'plant translational genomics'¹⁷.

Box 1. Glossary

Genomics is the study of whole genome of any particular organism. Deciphering the genomic content of crop species using high-throughput and next-generation approaches will allow the research community to generate elite cultivars with agronomically important traits, thus ultimately addressing the food insecurity. Genomics is a multi-disciplinary subject which has many modules such as the following:

Structural genomics: This is the study of DNA sequencing, sequence assembly, sequence organization and management. Being the initial step of genome analysis, it includes construction of high resolution genetic, physical or sequence maps of the organism for conducting genome-wide studies. Availability of high-throughput technologies such as 3D visualization of biomolecules, easy identification of molecular interactions, etc. has accelerated the progress of structural genomics.

Functional genomics: By utilizing the sequence information, functional genomics examines the functions and interactions of genes. It examines the dynamic aspects at genome level, such as transcription, translation and biomolecular interactions. Functional genomics deals with answering the questions on the function of DNA at the levels of genes, RNA transcripts and translational products.

Epigenomics: This is the study of complete set of epigenetic modifications in the genetic material of a cell. It includes DNA methylation and histone modification, which play a significant role in gene expression and regulation during both development and stress response.

Comparative genomics: It is the study of comparing two genomes to investigate the genomic differences, similarities, adaptive traits, evolution and gene functions. It would address the key questions such as what makes closely related species different.

Translational genomics: It represents the adaptation of data derived from genome technologies for crop improvement. Use of molecular markers in breeding for agronomic traits in crop plants appears promising in circumventing the food insecurity.

Next-gen population genomics: Producing a new generation of genetic-mapping populations using genomics is called next-gen population genomics. The next-gen populations promisingly overcome many of the limitations of biparental QTL mapping and association mapping.

Table 1. List of sequenced crops species which serve as food and feed

Year of release	Crop	Botanical name	Reference
2002	Rice	<i>Oryza sativa</i> ssp. <i>indica</i>	27
2002	Rice	<i>Oryza sativa</i> ssp. <i>japonica</i>	28
2009	Maize	<i>Zea mays</i>	29
2009	Sorghum	<i>Sorghum bicolor</i>	30
2010	Castor bean	<i>Ricinus communis</i>	31
2010	Soybean (cultivated)	<i>Glycine max</i>	32
2011	Potato	<i>Solanum tuberosum</i>	33
2011	Pigeon pea	<i>Cajanus cajan</i> genotype ICPL87119	34
2011	Pigeon pea	<i>Cajanus cajan</i> genotype ICPL8711	35
2011	Mustard	<i>Brassica rapa</i>	36
2012	Tomato	<i>Solanum lycopersicum</i>	37
2012	Foxtail millet	<i>Setaria italica</i> accession Yugu1	3
2012	Foxtail millet	<i>Setaria italica</i> strain Zhang gu	4
2012	Flax	<i>Linum usitatissimum</i>	38
2012	Cassava	<i>Manihot esculenta</i>	39
2012	Wheat	<i>Triticum aestivum</i>	6
2012	Barley	<i>Hordeum vulgare</i>	8
2013	Chickpea (kabuli type)	<i>Cicer arietinum</i> variety CDC frontier	10
2013	Chickpea (desi type)	<i>Cicer arietinum</i> cultivar ICC4958	11

Heterologous expression of genes related to biotic and abiotic stress tolerance, plant development, yield and quality aspects in agronomically important crops has been recorded since long back.

Hanafy *et al.*¹⁸ demonstrated a better drought and salinity tolerance in faba beans (*Vicia faba* L.) transformed with potato *PR10a* gene. The leaves of the transgenic plants were observed to be

green even after withholding water for three weeks and also displayed a higher tolerance to salinity stress¹⁸.

In addition, numerous databases loaded with genomic and genetic

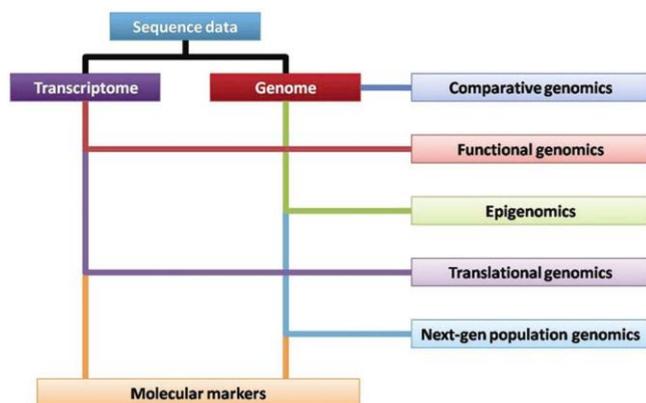


Figure 1. The association of genome sequence data with the respective genomic studies.

information have been developed for researchers and plant breeders aiming for crop improvement. Noteworthy, the crop-, genus- and family-specific genomic resource databases have immense applications. Some examples are CMD, Cotton Marker Database; VegMarks, a marker database for vegetables and Peanut Marker Database, etc. With the release of foxtail millet draft genome^{3,4} and the subsequent generation of microsatellite markers¹⁵, Foxtail millet Marker Database (FmMDB) has been developed by our laboratory for providing the generated information to the researchers and breeders free of cost. Realizing the significance of transcription factors (TFs), which play the crucial role as controls for gene expression, in regulating their temporal and spatial expression particularly during stress, many databases for TFs have also been developed. Plant Transcription Factor Database (PlantTFDB), Database of *Arabidopsis* Transcription Factors (DATF), Database of Rice Transcription Factors (DRTF), Database of Poplar Transcription Factors (DPTF) were the outcome of such attempts, which enable the user to browse for the TF of interest either by name or chromosomal location; some databases also provide the phylogenetic tree-view of the respective TF family. Considering the potential abiotic stress tolerance of foxtail millet, the construction of similar database for foxtail millet TFs is in progress in our laboratory, which would promisingly assist the plant research community in selecting the stress-responsive TFs for studies aimed at imparting abiotic stress tolerance in crop plants.

The study of complete set of epigenetic modifications in the genome termed ‘epigenomics’, has now become the hot

topic of research. Precisely, epigenetic modifications are reversible changes that occur in the DNA (DNA methylation) or histones (histone modification), thus affecting the gene expression without altering the DNA sequence. Natural epigenetic variation serves as a source for the evolution of phenotypic diversity. However, to understand its contribution towards this phenotypic diversity, its interaction with genetic variation entails further research. One such investigation was conducted by Schmitz *et al.*¹⁹ who studied the level, pattern and origin of epigenomic variation in *A. thaliana* by characterizing the genomes, methylomes and transcriptomes of wild populations of *A. thaliana*. The study proposes that, though single cytosine methylation polymorphisms do not have a genetic basis in this species, genetic variation does affect RNA-directed DNA methylation (RdDM) which occurs at differentially methylated regions, for which thousands of methylation quantitative trait loci were identified. The study provides evidence that RdDM-targeted genes might have chosen the transposon silencing mechanism to preserve their silenced condition in vegetative tissues and transgenerationally, in addition to ensuring appropriate expression vital for pollen, seed and germ-line development¹⁹.

According to World Health Organization (WHO), approximately 250,000 to 500,000 vitamin A-deficient children become blind every year, half of them dying within 12 months of losing their sight (<http://www.who.int/nutrition/topics/vad/en/>). This is a simple example of under-nutrition and malnutrition, while the listing on the effect of nutrition deficiency on human population is infinite. In this scenario, biofortification, the strategy of breeding nutrients into food crops

appears promising in evading malnutrition and under-nutrition. It provides a comparatively economic, sustainable and durable means of delivering more micronutrients. The historical ‘Golden Rice’ is an outcome of one such attempt, where rice crop is engineered genetically with the potential to biosynthesize beta-carotene, a precursor of vitamin A, in the edible parts of rice²⁰. Later, many researchers had demonstrated the applicability of selecting the lines with diverse vitamin and mineral profiles for exploiting them towards genetic improvement^{21,22}. Some recent examples include the following: Lee *et al.*²³ showed an increase in iron content in rice seeds twofold through overexpressing *OsNAS2* gene; Salunke *et al.*²⁴ studied the bioavailability of zinc in a diploid wheat (*T. monococcum*) *lpa* mutant using *Caco2* cell lines and observed that it essentially had higher bioavailability of zinc than wild type. Despite these merits, there exist certain major gaps in knowledge with respect to biofortification, the most important being the requirement of more efficient trials and efficacy studies to confirm and enhance the promising evidence thus far obtained^{25,26}.

The milestone scientific advances such as discovery of structure of the DNA, and improvement of molecular biology assisted with the establishment of sophisticated high-throughput technologies have resulted in the development of two novel tools for crop improvement, namely marker-assisted selection and transgene-based genetic modification. As discussed, both strategies were evidenced to be wisely implemented in many crop improvement programmes and they still appear promising in ensuring food security now and in future.

1. Paterson, A. H., Freeling, M. and Sasaki, T., *Genome Res.*, 2005, **15**, 1643–1650.
2. The Arabidopsis Genome Initiative. *Nature*, 2000, **408**, 796–815.
3. Bennetzen, J. L. *et al.*, *Nature Biotechnol.*, 2012, **30**, 555–561.
4. Zhang, G. *et al.*, *Nature Biotechnol.*, 2012, **30**, 549–554.
5. Lata, C., Gupta, S. and Prasad, M., *Crit. Rev. Biotechnol.*, 2012; doi: 10.3109/07388551.2012.716809.
6. Brechley, R. *et al.*, *Nature*, 2012, **491**, 705–710.
7. Muthamilarasan, M. and Prasad, M., *Curr. Sci.*, 2013, **104**, 286.
8. International Barley Genome Sequencing Consortium, *Nature*, 2012, **491**, 711–716.

RESEARCH NEWS

9. Muthamilarasan, M. and Prasad, M., *Curr. Sci.*, 2013, **104**, 998–999.
10. Varshney, R. K. *et al.*, *Nature Biotechnol.*, 2013, **31**, 240–246.
11. Jain, M. *et al.*, *Plant J.*, 2013; doi: 10.1111/tpj.12173.
12. Varshney, R. K., Nayak, S. N., May, G. D. and Jackson, S. A., *Trends Biotechnol.*, 2009, **27**, 522–530.
13. Ling, H. Q. *et al.*, *Nature*, 2013, **496**, 87–90.
14. Jia, J. *et al.*, *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 2013, **496**, 91–95.
15. Pandey, G., Misra, G., Kumari, K., Gupta, S., Parida, S. K., Chattopadhyay, D. and Prasad, M., *DNA Res.*, 2013; doi: 10.1093/dnares/dst002.
16. Pfeifer, M. *et al.*, *Plant Physiol.*, 2013, **161**, 571–582.
17. Salentijn, E. M. J., Pereira, A., Angenent, G. C., van der Linden, C. G., Krens, F., Smulders, M. J. M. and Vosman, B., *Mol. Breed.*, 2007, **20**, 1–13.
18. Hanafy, M. S., El-Banna, A., Schumacher, H. M., Jacobsen, H. J. and Hassan, F. S., *Plant Cell Rep.*, 2013; doi: 10.1007/s00299-013-1401-x.
19. Schmitz, R. J. *et al.*, *Nature*, 2013, **495**, 193–198.
20. Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P. and Potrykus, I., *Science*, 2000, **287**, 303–305.
21. Ashok Kumar, A., Reddy, B. V. S., Ramaiah, B., Sahrawat, K. L. and Pfeiffer, W. H., *Eur. J. Plant Sci. Biotechnol.*, 2012, **6**, 66–70.
22. Velu, G. *et al.*, *Field Crops Res.*, 2012, **137**, 261–267.
23. Lee, S., Kim, Y.-S., Jeon, U. S., Kim, Y.-K., Schjoerring, J. K. and An, G., *Mol. Cell*, 2012, **3**, 33.
24. Salunke, R., Rawat, N., Neelam, K., Tiwari, V. K., Randhawa, G. S., Dhaliwal, H. S. and Roy, P., *J. Food Comp. Anal.*, 2012, **25**, 149–159.
25. Saltzman, A., Birol, E., Bouis, H. E., Boy, E., de Moura, F. F., Islam, Y. and Pfeiffer, W. H., *Global Food Security*, 2013, **2**, 9–17.
26. Rawat, N., Neelam, K., Tiwari, V. K. and Dhaliwal, H. S., *Plant Breed.*, 2013; doi: 10.1111/pbr.12040.
27. Yu, J. *et al.*, *Science*, 2002, **296**, 79–92.
28. Goff, S. A. *et al.*, *Science*, 2002, **296**, 92–100.
29. Schnable, P. S. *et al.*, *Science*, 2009, **326**, 1112–11125.
30. Paterson, A. H. *et al.*, *Nature*, 2009, **457**, 551–556.
31. Chan, A. P. *et al.*, *Nature Biotechnol.*, 2010, **28**, 951–956.
32. Schmutz, J. *et al.*, *Nature*, 2010, **463**, 178–183.
33. The Potato Genome Sequencing Consortium. *Nature*, 2011, **475**, 189–195.
34. Varshney, R. K. *et al.*, *Nature Biotechnol.*, 2011, **30**, 83–89.
35. Singh, N. K. *et al.*, *J. Plant Biochem. Biotechnol.*, 2011, **21**, 98–112.
36. The *Brassica rapa* Genome Sequencing Project Consortium, *Nature Genet.*, 2011, **43**, 1035–1039.
37. The Tomato Genome Consortium, *Nature*, 2012, **485**, 635–641.
38. Wang, Z. *et al.*, *Plant J.*, 2012, **72**, 461–473.
39. Prochnik, S. *et al.*, *Trop Plant Biol.*, 2012, **5**, 88–94.

Mehanathan Muthamilarasan and Manoj Prasad are in the National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110 067, India; P. Theriappan is in Prathyusha Institute of Technology and Management, Poona-mallee–Tiruvallur High Road, Tiruvallur 602 025, India.*

**e-mail: manoj_prasad@nipgr.ac.in*
