Indian J. Plant Genet. Resour. 35(3): 194–199 (2022)
DOI 10.5958/0976-1926.2022.00068.7

Genomic Tools in Plant Genetic Resource Management

Dhammaprakash P Wankhede, Rakesh Singh and S Rajkumar*
ICAR-National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi-110012, India

Significant progress has been made in conservation of plant germplasm in last few decades. However, the utilization of genetic resources has remained poor. The reason for the under utilization of plant genetic resources (PGR) are non-availability of structured information and resources from gene banks. This was due to lack of characterization and evaluation data, and under usage of genomic tools in characterization. With the evolution of genomic tools, which get cheaper with discovery of newer technologies, it is pertinent to use these tools for management and utilization of PGR from vast germplasm conserved in gene banks throughout world. In this context the present article describes the tools that can be used in gene bank management and the way to utilize the germplasm with the examples from the Indian National Gene Bank.

Key Words: Core sets, DNA bar coding, Genetic erosion, Genotyping, Phenotyping, Reference sets

Introduction

Plant genetic resources (PGR) represent natural variation in crops and their wild relatives, which support food security, especially under climate change and nutritional challenges. Significant progress has been made in conservation of plant germplasm in last few decades. However, the utilization of genetic resources has remained poor. The reason for the underutilization of PGR are non-availability of structured information and resources from gene banks. This was due to lack of characterization and evaluation data, and under usage of genomic tools in characterization. With the evolution of genomic tools, which get cheaper with discovery of newer technologies, it is pertinent to use these tools for management and utilization of PGR from vast germplasm conserved in gene banks throughout world. In this context the present article describes the tools that can be used in gene bank management and the way to utilize the germplasm with the examples from the Indian National Gene Bank.

Species Integrity of Seed Gene Bank

Plant germplasm conserved in gene bank is important source of material for studies on crop domestication, evolution and for crop improvement. The identity of each accession is important as it affects all the downstream process in crop improvement and related studies. This process is an important aspect in gene bank management. Most of the germplasm in gene bank are cultivated and readily identifiable, but occasional problem exists in crop wild relatives, especially vegetable crops. Most of the vegetable crops come under the family Cucurbitaceae where morpho-taxonomy is quite difficult. DNA barcoding is set of coding regions of nuclear and chloroplast DNA, and is useful in systematics study in plants. These loci are universally accepted and used in identification of plant species. In a study by the authors, barcoding loci useful in delineation of species complex in wild relatives of vegetable crops were identified and effectively used in identification of germplasm with unknown species status and problematic species complex. Among the universally accepted barcoding loci, rbcL and trnH-psbA were more useful in delimiting *Luffa* spp. including *L. acutangula* complex and status of *L. tuberosa*. In vegetable *Amaranthus* spp. nuclear ITS sequence was more appropriate in differentiating vegetable *Amarnathus* species along with delineation of species complexes (blitum and tricolor). In *Trichosanthes cucumerina*, ITS combined with rbcL were effective in delineating *T. cucumerina* species complex. The delineation of taxonomic complex existing within species of *Cucumis melo* were carried out with combine ITS and rbcL and they are able to distinguish different taxonomic varieties under the species.

Estimation of Genetic Erosion

In the agricultural sphere, there is ongoing concern and attention to genetic erosion at all levels, including

**Author for Correspondence: Email-S.Rajkumar@icar.gov.in*

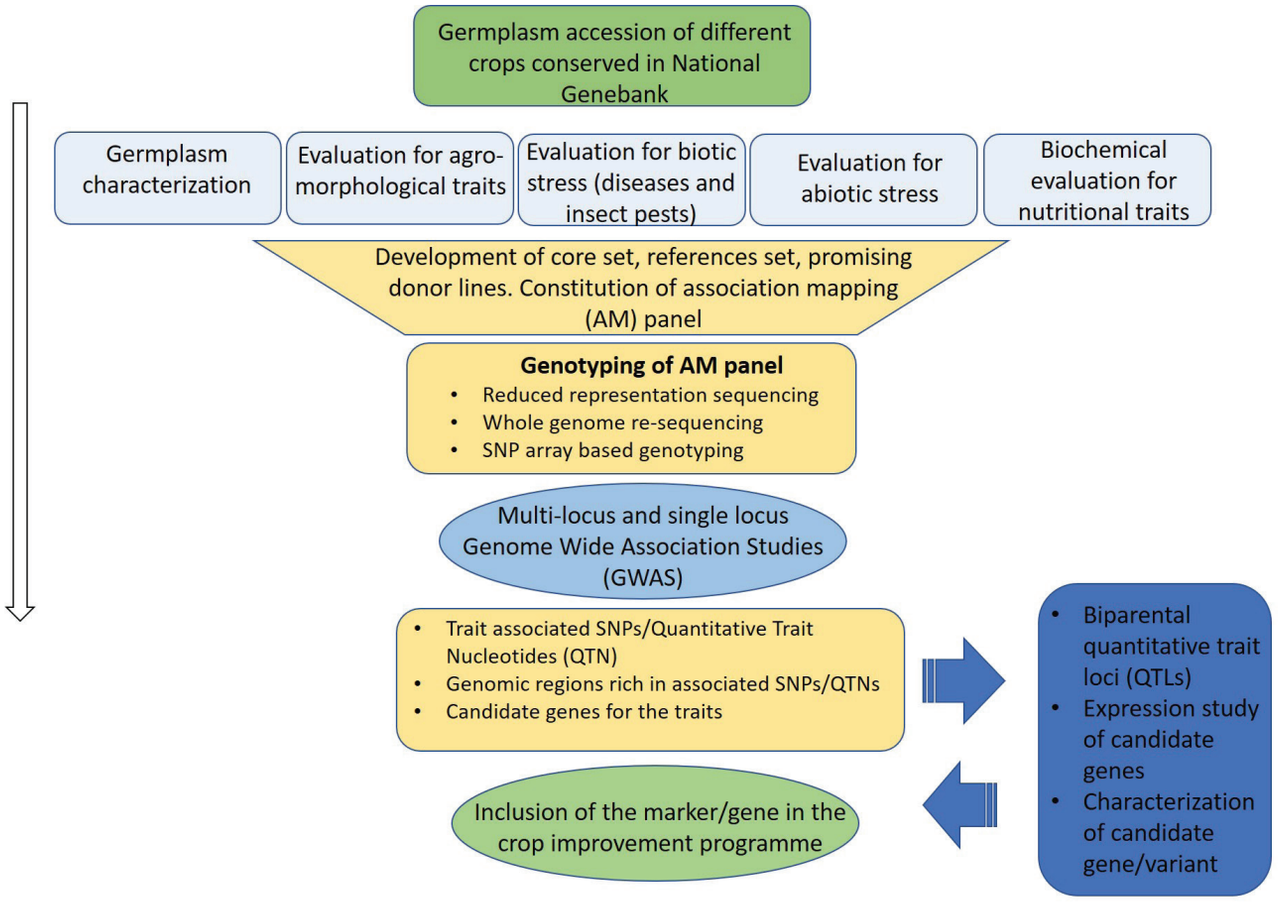

Fig. 1. Overview of approach for utilization of Genebank accessions using Genome Wide Association Studies (GWAS) towards identification marker/gene-trait association for crop improvement

the Food and Agriculture Organization of the United Nations (FAO). It is important to note that along with the loss of diversity at the species, varietal and allelic levels, genetic erosion can also occur at the level of germplasm collections and gene banks due to improper management and inadequate regeneration procedures. A large amount of genetic diversity has and is being collected and stored in gene banks, in which every sample (usually of seed) is kept in airtight containers at -10 to -200 ºC and 5 to 7% humidity for 50 to 100 years (Damania, 2008). From time-to-time, a given amount of material is taken out of the gene bank, planted in the field and 'rejuvenated': the fresh seed is then stored again. Worldwide, 1308 gene banks are registered and conserve a total of 6.1 million accessions, including major crops, minor or neglected crop species as well as trees and wild plants. Of the 30 main crops, more than 3.6 million accessions are conserved *ex situ*. On one hand these collections serve a very important purpose - avoiding the loss of individuals and species, and of the genes they carry, which may be unique. On the other hand, by 'freezing' seeds they also 'freeze' evolution at the time of the collection. Therefore, many advocate that together with the *ex situ* conservation in gene banks, diversity should also be conserved in its original locations (*in situ*), where the plant populations can continue to evolve.

Genetic erosion as a reduction in evenness originates from the diversity indices used in vegetation ecology and population genetics, such as Shannon's index or Nei's gene diversity index. Diversity is measured using the frequencies of alleles within a group of genotypes or using the production areas of landraces, cultivars or crop species in a region. Diversity levels are lowered due to increasing dominance of a single or small number of crop species, genotypes or alleles, even though alleles or varieties are not necessarily lost. Using evenness, rare varieties or rare alleles contribute little to the diversity.

The risks of losing alleles or varieties are higher when distributions are very skewed. Using evenness as a measure for genetic erosion offers the opportunity to take action before a reduced diversity results in an absolute loss and reduced richness. Furthermore, it is not as sensitive to the sampling procedure as compared with the previous measures. Considerable overlap between these three views on genetic erosion exists, and most studies use a combination of the different approaches.

A case study was conducted to identify and characterize changes in long-term conserved samples of safflower genetic resources in National Genebank (NGB) and to estimate the risk of genetic erosion during the conservation process. Total four accessions (50 individuals each) were selected based on the availability of historic material in the gene bank and deposition of regenerated material after constant cycle of regeneration. For estimating variation in allelic frequency, genomic SSR markers were used. The markers with high polymorphic information content (PIC) content from the earlier studies were used to generate allelic diversity in both regimes of the accessions. The results showed changes mainly due to some highly significant differences in allele frequencies, whereas the majority of alleles occur in similar frequencies. This implies that either regeneration protocols should be improved or the composition of the collection should be changed in the gene bank.

Molecular Core Development

A large number of genetic materials have been conserved in gene banks, but their use is limited due to an unmanageable number of accessions and the continuous expansion of accession numbers. Core germplasm development has been proposed for better management and use of collections available in gene banks. This requires the development of a core set of accessions to more precisely characterize, explore, and conserve gene bank resources, monitor the genetic drift during preservation, and identify gaps in genetic diversity. Genetic diversity and population structure knowledge form the backbone in building core sets adequately representing variations found in the whole collection, and thereby making the collection small and condensed (Yan *et al*., 2007; Agrama *et al.*, 2009; El Bakkali *et al.*, 2013).

India's east coast rice collections were characterized using SNP markers. The genetic diversity and population structure were studied, and core and mini core collections with maximum diversity and minimum redundancy were developed. A total of 2,242 east coast rice accessions from three different states of India, i.e., Andhra Pradesh, Orissa, and Tamil Nadu, have been characterized, and a wide range of gene diversity and PIC was observed. A phylogenetic analysis of the total east coast rice collection revealed three groups, and a population structure analysis revealed four populations. The 36-SNP assay used in this study was validated by comparing the genetic diversity parameters (gene diversity, PIC, major allele frequency, and heterozygosity) across two different rice collections, i.e., east coastal rice and northeast rice collection, and it was observed these markers were sufficient to decipher all genetic parameters very efficiently; hence, they can be effectively utilized for core development and diversity study of different rice genotypes (Choudhury *et al* 2021).

Identification of Germplasms with Superior Trait Value

The National Gene Bank at ICAR-NBPGR facilitates the conservation of genetic diversity by harboring the diverse plant genetic resources important for food and agriculture. These genetic resources constitute vast variability for several agronomically and economically important traits in respective crop plants. The conserved diversity and variability can cater the present requirements and also provide insurance for the future adverse conditions including biotic and abiotic stress (Paroda and Arora, 1991, Yadav *et al.,* 2018). In this perspective, it is crucial to pinpoint the desired trait specific accessions by the exercise of thorough characterization and comprehensive evaluation for specific traits. Identification of accessions with desired specific traits can be used as promising donors in breeding program and thereby facilitate enhanced utilization of the conserved germplasm. The NGB accessions of several crops have been evaluated for identification of accessions with desired traits. In wheat, 498 and 868 accessions have been identified potentially resistant for different wheat rusts and spot blotch, respectively from initial 19460 wheat accession (Kumar *et al.,* 2016). The characterization of entire NGB collection of barley (6,778 accessions) has helped identification of trait-specific accessions for agro-morphologically and economically important traits including days to spike emergence, days to maturity, plant height, spike length, number of grains per spike and hundred-grain weight (Kaur *et al.,* 2022). Among pulses, in chickpea, novel promising donors with high disease resistance to

Ascochyta blight (*Ascochyta rabiei*) have been identified by evaluation of chickpea germplasm accessions of NGB (Gayacharan *et al.,* 2020). The characterization of complete lentil germplasm collection of NGB over two years led to identification superior accessions for crucial traits such as early maturity, number of secondary branches, pods per plant and seed weight and suitability for mechanical harvesting (Tripathi *et al.,* 2022). In linseed germplasm accessions with superior trait values have been identified for flowering, maturity, test seed weight, seed area, capsules numbers per plant and plant height (Kaur *et al.,* 2018; Saroha *et al.,* 2022a). With increase in awareness about healthy diet among general public, a special emphasis should be given to evaluation of germplasm of NGB for various nutritional aspects to identify germplasm for high nutritional values like protein, fibre, important vitamins, mineral, lignans and with healthy fatty acid ratio and deficient in anti-nutritional factors. Recently, nutri-dense accessions have been identified in cowpea with high protein, total soluble sugar, amylose, and total dietary fibers from by nutritional profiling of 120 biochemically diverse cowpea accessions (Padhi *et al*., 2022).

In recent years, under the mission projects funded by Department of Biotechnology at institutes of special expertise of ICAR, DBT and SAUs, the comprehensive phenotyping, genotyping and whole genome sequencing have been undertaken in important cereals, pulses and oilseeds. In these projects either complete germplasm accessions of respective crops or substantial number of accessions of crops like including rice, wheat, chickpea, minor pulses (cowpea, greengram, blackgram, mothbean and horse gram), sesame, safflower, linseed and niger are being characterized. These characterizations and evaluation under different biotic and abiotic stress condition and nutritional profiling are expected to unravel the diversity and variability for different traits in these crops and help development of core sets, reference sets for agro-morphological and nutritionally important traits. Further, such exercise would lead to identification of several trait specific accessions which can be utilized in breeding programme for crop improvement.

Employing Genome Wide Association Studies for Accelerated Utilization of Genebank Germplasm

Germplasm collection of Genebanks constitute genetic resources from diverse geographical areas and distinct populations which often show extensive phenotypic variation for several traits. Such a collection of germplasm accessions is the treasure trove to understand the underlying genetic architecture for the traits in question. Plant scientists have traditionally used the biparental mapping for identification of quantitative (QTL)/ genomic regions associated with the desired complex traits. However, this approach assays genetic variation limited to two parents and fewer recombination events in a biparental population. Genome-wide association study (GWAS) on the other hand assays a wide swathe of existing natural variation through population-scale samples and takes into account the historic recombination events across lineages, enabling a finer resolution of QTL (Burghardt *et al.,* 2017). Exploitation of such natural variation is more pertinent in the context of changing climatic conditions and to offer environmentally sustainable crop production.

In order to undertake a GWAS study, a precise phenotyping of large number of accessions for the desired traits is crucial. In case of field phenotyping, multi-location-season phenotyping would be beneficial to reduce the environmental error. Next, to have the phenotype-genotype association, it is important to perform genotyping of the same set of accessions at substantial genome coverage. Different genotyping approaches can be employed as per the crop species, available genomic resources and economics. More the available variants, higher are the chances to identify causative variants, therefore the best approach would be to have the complete genome sequencing of the accessions in the association panel (Gua *et al*., 2019). Besides the whole genome sequencing, array-based SNP genotyping and reduced representation sequencing approaches have also been employed to get significant markers associated with the trait (Saroha *et al*., 2022b, Vikas *et al.,* 2022).

Many researchers have preferred to constitute a subset of total germplasm collection (such as core set) as association mapping panel and identified significant QTLs for the complex traits (Soto-Cerda *et al.,* 2021). Once phenotype and genotyping of an AM panel is achieved, the next crucial step is application of statistical methods to identify marker trait association. There are several single locus models for example MLM, EMMAX, GEMMA and ECMLM are available for genetic dissection of traits. However, using these models it is essential to control the false positive rates by applying stringent corrections (Bonferroni corrections) which often result in

exclusion of important loci (Zhang *et al.,* 2019). Multi-locus methods have advantage over the single locus methods as it simultaneously tests multiple markers and thereby increases the statistical power while reducing type 1 error (Zhang *et al*. 2019). Once the statistically significant marker traits association is established, candidate genes for the traits can be selected either by direct identification of causative variant in the candidate gene or on the basis of the functional relevance of genes in the strong linkage disequilibrium (Burghardt *et al.,* 2017). Gene annotations of candidate genes, KEGG analysis and biological relevance have to be established to take the call for candidate genes. Further, validation of the candidate genes can be undertaken such as gene expression studies, experimental validation using RNAi, and genome editing methods including CRISPR/Cas9. Having established the association, the marker/gene can be used in breeding programmes. The illustration of application of GWAS in utilization of genetic resources of NGB is shown in Fig. 1. The similar approaches have been used by several researchers to identify QTLs/QTNs in different crops for agro-morphological traits as well as for biotic and abiotic stress (Pradhan *et al*., 2020; *Kumar et al.,* 2020; Chaurasia *et al*., 2021; Vikas *et al.,* 2022; Saroha *et al.,* 2022a). The GWAS approach can accelerate the pace of identification of gene/marker-trait association, genetic dissection of important traits and thereby utilization of germplasm collection of NGB in breeding programmes for varietal development.

References

Agrama HA, Yan W, Lee F, Fjellstrom R, Chen M-H, Jia M, et al. (2009). Genetic Assessment of a Mini-Core Subset Developed from the USDA Rice Genebank. *Crop Sci.* **49:** 1336–1346. 10.2135/cropsci2008.06.0551

Burghardt, L.T., Young, N.D., and Tiffin, P. 2017. A guide to genome-wide association mapping in plants. Curr. Protoc. Plant Biol. 2:22-38. doi: 10.1002/cppb.20041

Chaurasia, S., Singh, A. K., Kumar, A., Songachan, L. S., Yadav, M. C., Kumar, S., et al. (2021). Genome-wide Association Mapping Reveals Key Genomic Regions for Physiological and Yield-Related Traits under Salinity Stress in Wheat (*Triticum aestivum* L.). Genomics 113 (5), 3198–3215. doi:10.1016/j. ygeno.2021.07.014

Choudhury DR, Kumar R, S VD, Singh K, Singh NK, Singh R. Identification of a Diverse Core Set Panel of Rice From the East Coast Region of India Using SNP Markers. Front Genet. 2021 Nov 25;12:726152.

Damania AB. 2008. History, Achievements, and Current Status of Genetic Resources Conservation. *Agron. J.* **100:** 9–21

El Bakkali A., Haouane H., Moukhli A., Costes E., van Damme P., Khadari B. (2013). Construction of Core Collections Suitable for Association Mapping to Optimize Use of Mediterranean Olive (Olea Europaea l.) Genetic Resources. PLoS One 8, e61265. 10.1371/journal.pone.0061265

Gayacharan, Rani U, Singh S, Basandrai AK, Rathee VK, Tripathi K, et al. (2020) Identification of novel resistant sources for ascochyta blight (Ascochyta rabiei) in chickpea. *PLoS ONE* **15(10):** e0240589. https://doi.org/10.1371/journal.pone.0240589

Guo, D., Jiang, H., Yan, W., Yang, L., Ye, J., Wang, Y., Yan, Q., Chen, J., Gao, Y., Duan, L., *et al.* (2019). Resequencing 200 flax cultivated accessions identifies candidate genes related to seed size and weight and reveals signatures of artificial selection. Front. Plant Sci., 10, 1682.

Kaur V, Aravind J, Manju, Jacob SR, Kumari J, Panwar BS, Pal N, Rana JC, Pandey A and Kumar A (2022) Phenotypic Characterization, Genetic Diversity Assessment in 6,778 Accessions of Barley (*Hordeum vulgare* L. ssp. vulgare) Germplasm Conserved in National Genebank of India and Development of a Core Set. Front. *Plant Sci*. 13: 771920. doi: 10.3389/fpls.2022.771920

Kaur V, S. Kumar, R. Yadav, D.P. Wankhede, J. Aravind, J. Radhamani, J.C. Rana and A. Kumar (2018) Analysis of genetic diversity in Indian and exotic linseed germplasm and identification of trait specific superior accessions. *J. environ. boil.* **39-5:** 702-709. DOI : http://doi.org/10.22438/jeb/39/5/MRN-849.

Kumar S, Archak S, Tyagi RK, Kumar J, VK V, Jacob SR, *et al.* (2016) Evaluation of 19,460 Wheat Accessions Conserved in the Indian National Genebank to Identify New Sources of Resistance to Rust and Spot Blotch Diseases. *PLoS ONE* **11(12):** e0167702. doi:10.1371/journal.pone.0167702

Padhi SR, Bartwal A, John R,Tripathi K, Gupta K, Wankhede DP, Mishra GP, Kumar S, Archak S and Bhardwaj R (2022) Evaluation and Multivariate Analysis of Cowpea [*Vigna unguiculata* (L.) Walp] Germplasm for Selected Nutrients—Mining for Nutri-Dense Accessions. Front. Sustain. Food Syst. 6:888041. doi: 10.3389/fsufs.2022.888041

Paroda RS, Arora RK (1991) Plant genetic resources: general perspective (English) In: Plant genetic resources: conservation and management. Concepts and approaches/International Board for Plant Genetic Resources, New Delhi (India). Regional Office for South and Southeast Asia, 1991, 1-23

Pradhan AK, Kumar S, Singh AK, Budhlakoti N, Mishra DC, Chauhan D, Mittal S, Grover M, Kumar S, Gangwar OP, Kumar S, Gupta A, Bhardwaj SC, Rai A and Singh K (2020) Identification of QTLs/Defense Genes Effective at Seedling Stage Against Prevailing Races of Wheat Stripe Rust in India. *Front. Genet*. 11:572975. doi: 10.3389/fgene.2020.572975

Saroha A, Pal D, Gomashe SS, Akash, Kaur V, Ujjainwal S, Rajkumar S, Aravind J, Radhamani J, Kumar R, Chand D, Sengupta A and Wankhede DP (2022a) Identification of QTNs Associated with Flowering Time, Maturity, and Plant Height Traits in *Linum usitatissimum* L. Using Genome-

Wide Association Study. *Front. Genet.* 13:811924. https://doi:10.3389/fgene.2022.811924.

Saroha, A., Pal, D., Kaur, V. Kumar S, Bartwal A, J. Aravind, J. Radhamani, S. Rajkumar, Kumar R, Gomashe SS, Sengupta A, Wankhede DP* (2022b) Agro-morphological variability and genetic diversity in linseed (*Linum usitatissimum* L.) germplasm accessions with emphasis on flowering and maturity time. *Genet Resour Crop Evo*. https://doi.org/10.1007/s10722-021-01231-3

Soto-Cerda, B. J., Aravena, G., and Cloutier, S. (2021). Genetic Dissection of Flowering Time in Flax (*Linum usitatissimum* L.) through Single- and Multi-Locus Genome-wide Association Studies. *Mol. Genet. Genomics* **296 (4):** 877–891. doi:10.1007/s00438-021-01785-y

Tripathi K, Kumari J, Gore PG, Mishra DC, Singh AK, Mishra GP, C G, Dikshit HK, Singh N, Semwal DP, Mehra R, Bhardwaj R, Bansal R, Rana JC, Kumar A, Gupta V, Singh K and Sarker A (2022) Agro-Morphological Characterization of Lentil Germplasm of Indian National Genebank and Development of a Core Set for Efficient Utilization in Lentil Improvement Programs. *Front. Plant Sci*. 12:751429. doi: 10.3389/fpls.2021.751429

Vikas, V.K., Pradhan, A.K., Budhlakoti, N. *et al.* Multi-locus genome-wide association studies (ML-GWAS) reveal novel genomic regions associated with seedling and adult plant stage leaf rust resistance in bread wheat (*Triticum aestivum* L.). *Heredity* **128**: 434–449 (2022). https://doi.org/10.1038/s41437-022-00525-1

Yadav B, Kaur V, Narayan OP, Yadav SK, Kumar A and Wankhede DP (2022) Integrated omics approaches for flax improvement under abiotic and biotic stress: Current status and future prospects. *Front. Plant Sci.* 13:931275. https://doi:10.3389/fpls.2022.931275.

Yadav S, Chet Ram, and Wankhede DP (2018) Biodiversity and Conservation of Plant Genetic Resources (*in-Situ* and *Ex-Situ*). In: Advanced Molecular Plant Breeding Meeting the Challenge of Food Security Eds. D. N. Bharadwaj. Apple Academic Press.

Yan W., Rutger J. N., Bryant R. J., Bockelman H. E., Fjellstrom R. G., Chen M.-H., et al. (2007). Development and Evaluation of a Core Subset of the USDA rice Germplasm Collection. *Crop Sci.* 47, 869–876. 10.2135/cropsci2006.07.0444

Zhang Y-M, Jia Z and Dunwell JM (2019) Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Front. Plant Sci*. **10:** 100. doi: 10.3389/fpls.2019.00100