

Role of Genebanking in the Era of Genomics

Veena Gupta and Sherry Rachel Jacob*

Division of Germplasm Conservation, ICAR-National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi-110012, India

“Cherish variations, for without it life will perish” – Sir Otto Frankel

The global system for conservation of plant genetic resources (PGR) is a pan-continental network integrating more than 2000 national and international genebanks out of which 130 genebanks hold more than 10,000 accessions and eight have more than 100,000 accessions. As per the PGR Genesys platform, accessions conserved in *ex situ* genebanks comprise more than 4 million maintained by over 450 institutes across the globe. Accumulation of huge quantity of germplasm accessions of unknown value and genetic composition within genebanks made its inclusion in crop breeding, economically unmanageable. This article provides some of the new age tools and methodologies being applied for more effective management of genebanks.

Key Words: Digital Sequence Information, Genome-wide association studies, Genomic selection, Machine learning, Pan-genome studies

Introduction

Plant genetic resources (PGR) are the forerunners of crop evolution and they occupy the most diverse ecological habitats on this planet. They have defined and determined the survival and adaptation of all living beings, including mankind and has generously provided for their sustenance. Subsequent crop improvement efforts in selected crops led to preferential cultivation of improved varieties in a very limited number of species. The narrow genetic base in elite varieties resulted in significant diversity loss and the non-preferred species faced extinction. Plant biologists soon realized the impending threat and put forth the need to address ‘diversity’ in a more technical perspective. Starting from 1970s, there were concerted efforts to promote an international network of genetic resources centres to further collection, conservation, documentation, evaluation and use of germplasm (Frankel and Bennett, 1970). Today, the global system for conservation of PGR is a pan-continental network integrating more than 2000 national and international genebanks out of which 130 genebanks hold more than 10,000 accessions and 8 have more than 100,000 accessions. The PGR Genesys platform which hosts compiled data on accessions conserved in *ex situ* genebanks, reports a total of more

than 4 million accessions maintained by over 450 institutes across the globe (<https://www.croptrust.org/>). Over the decades, the accumulation of huge quantity of germplasm accessions of unknown value and genetic composition within genebanks made its inclusion in crop breeding, economically unmanageable. The primary strategy adopted by Genebank researchers to overcome this hindrance was to undertake redundancy assessment and subsequently, identify a priority set of germplasm accessions that can effectively capture the allelic diversity, to the extent required. In the case of Crop Wild Relatives (CWRs) and landraces, the priority was to minimize barriers in mobilizing genetic variation into elite background. The tools and techniques employed for meeting this objective have evolved along with the advances in science and today, PGR management and crop improvement are making parallel progress by harnessing the benefits of cutting edge technologies.

Genomic Tools for Genetic Dissection of Complex Traits

The two major forward genetics approaches used for understanding the genetic basis of traits in germplasm collections are linkage mapping or QTL mapping and association mapping or linkage disequilibrium (LD) mapping (Cadic *et al.*, 2013). In large *ex situ* gene bank

*Author for Correspondence: Email-sherry.jacob@icar.gov.in

collections, an ideal approach for genetic dissection is the association analysis *viz.*, genome-wide association studies (GWAS), which is based on linkage disequilibrium (LD) or the non-random association of alleles at different loci (Kumar *et al.*, 2022). The fundamental principle of GWAS is to analyse and validate marker-phenotype association within a very large, diverse set of individuals, as in the case of germplasm collections. Such collections have preserved historical recombinations and also account for larger number of alleles, thus enabling high-resolution mapping (Alqudah *et al.*, 2020). GWAS is being used extensively to identify allelic variants associated with multiple traits of adaptive and agronomic importance (Akram *et al.*, 2021).

The concept of Genomic Selection (GS) has also gained prominence, w.r.t to understanding the genetic basis of minor alleles. In GS, the genetic value of a set of accessions (breeding population) is predicted based on Genomic Estimated Breeding Value (GEBV) generated on a Training Population (TP) that is related to BP and is both genotyped and phenotyped. Using this prediction model, the marker effects are estimated across the breeding population, without phenotyping or QTL mapping (Heffner *et al.*, 2009; Calus, 2010). As compared to MAS, GS captures larger amount of genetic variation of the specified trait since it involves high density markers with both minor and major effects and distributed throughout the genome (Desti and Ortiz, 2014; Crossa *et al.*, 2017). The developments in the field of Machine Learning (ML) has enhanced genomic prediction accuracy through use of high precision tools for identifying subsets of SNPs having direct link to candidate genes affecting the trait of interest (Li *et al.*, 2018). The GS approach has been attempted in the ex situ genebank collection of CIMMYT, Mexico, for genomic prediction within Mexican and Iranian landrace accessions. The prediction accuracy implied potential scope of this technique in rapid introgression of exotic material including CWRs, into elite background, for effective germplasm enhancement (Crossa *et al.*, 2016).

Moving Ahead of Genomes

The advances in various sequencing technologies have led to generation of high quality reference genomes in several major crops and even in the wild relatives of rice, wheat, barley, soybean and tomato. But, in the context of PGR, this whole genome sequence assemblies had

the serious shortcoming of not addressing Structural Variations (SV) like Copy Number variations (CNVs) which are a major source of genetic variation during domestication and further diversification of crop taxa (Lye and Purugganan, 2019) that can have a pronounced phenotypic impact (Alkan *et al.*, 2011, Mahmoud *et al.*, 2019). The emerging area of pan-genome analysis, effectively addresses this crucial requirement, thus holding significant implications for breakthroughs in genebank germplasm utilization. In pan-genome approach, several reference genome sequences are used as template to map the sequencing data, whereby all possible natural variations get detected, including those lost during domestication and crop improvement (Bohra *et al.*, 2022).

The pangenome work on cultivated Asian rice conserved at IRRRI Genebank could successfully unravel the full repertoire of genetic diversity that exists in the largest ex situ rice Genebank collection (Zhou *et al.*, 2020). Similar work has been reported in hexaploid bread wheat (Montenegro *et al.*, 2017). Recently, the scope of pan-genome studies has been further expanded by including wild relatives, landraces and cultivars in multiple crops (Liu Y *et al.*, 2020). As science progresses in leaps and bounds, the idea of pan-genome has further moved forward into the concept of ‘super-pangenome’, which involves a genus-level approach, by assembling multiple species specific reference genomes. Shang *et al.*, 2022 reports the development of a *Oryza* super pangenome comprising both cultivated and wild species of Asian and African rice that can identify lineage-specific haplotypes for trait-associated genes and can also provide fundamental understanding on the genetic basis of environmental adaptation and domestication of rice. The super pan-genome of genus *Glycine* has also made a remarkable contribution to PGR utilization by unleashing genetic potential from the perennial gene pool for soybean improvement (Zhuang *et al.*, 2022). This strategy will facilitate in fully reaping the benefits of CWRs conserved in genebanks.

Genebank Documentation at the Genebank-Genomics Interphase

Genomics has propelled PGR management into the arena of ‘big data’ science, where data bases are no longer the supportive components, but rather, forms a highly specialized information inter phase that technically

regulates the access to resources. This Digital Sequence Information (DSI) now calls for defined guidelines and management strategies for ensuring fair benefit sharing without restricting access (Aubrey, 2019). Contracting parties to the Convention on Biological Diversity (CBD) and ITPGRFA are starkly divided in their opinions with regard to the rules for accessing PGR DSI (Scholz *et al.*, 2022) and it is currently challenging the basic fundamentals of PGR governance as envisaged in case of physical germplasm collections. Several solutions are being proposed for enhancing transparency in sequence data access. One example is the compulsory tagging of provenance information with the sequence information that has been proposed by the International Nucleotide Sequence Database Collaboration (INSDC), a core infrastructure for storage of global sequence data (Wilkinson *et al.*, 2016). Such endeavours will prove beneficial only if it is effectively complemented with the complete passport data of national genebanks. Thus, as the debate continues, the onus is on genebanks to ensure effective mechanism for extensive documentation of their sovereign genetic resources and also make in-house provisions for feedback and tracking of germplasm utilization by diverse consumers.

References

- Akram S, Arif MAR and Hameed A (2021) A GBS-based GWAS analysis of adaptability and yield traits in bread wheat (*Triticum aestivum* L.). *J. Appl. Genet.*, **62(1)**: 27-41.
- Alkan C, BP Coe and EE Eichler (2011) Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12(5)**: 363-376.
- Alqudah AM, A Sallam, PS Baenziger, A Börner, (2020) GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review, *J. Adv. Res.*, **22**: 119-135, <https://doi.org/10.1016/j.jare.2019.10.013>.
- Aubry S (2019) The Future of Digital Sequence Information for Plant Genetic Resources for Food and Agriculture, *Frontiers in Plant Sci.* 10, DOI=10.3389/fpls.2019.01046
- Bohra A, B Kilian, S Sivasankar, M Caccamo, C Mba, SR McCouch, RK Varshney (2022) Reap the crop wild relatives for breeding future crops. *Trends in Biotechnology.* **40(4)**: 412-431, ISSN 0167-7799, <https://doi.org/10.1016/j.tibtech.2021.08.009>.
- Cadic E, M Coque, F Vear, B Grezes-Besset, J Pauquet, J Piquemal, Y Lippi, P Blanchard, M Romestant, Pouilly N, Rengel D (2013) Combined linkage and association mapping of flowering time in Sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **126(5)**: 1337-1356. doi: 10.1007/s00122-013-2056-2.
- Calus MP (2010) Genomic breeding value prediction: methods and procedures. *Animal: an international journal of animal bioscience*, **4(2)**:157-64.
- Crossa J, D Jarquín, J Franco, P Pérez-Rodríguez, J Burgueño, C Saint-Pierre, P Vikram, C Sansaloni, C Petrolí, D Akdemir, C Sneller (2016) Genomic prediction of gene bank wheat landraces. *G3 (Bethesda, Md.)* **6(7)**: 1819-1834
- Crossa J, González-Camacho, S Pérez-Elizalde, Y Beyene, S Dreisigacker, R Singh, X Zhang, M Gowda, M Roorkiwal, J Rutkoski, R K. Varshney (2017) Genomic selection in plant breeding: Methods, Models, and Perspectives. *Trends in Plant Sci.* **22(11)**: 961-975
- Desta ZA and R Ortiz (2014) Genomic selection: genome-wide prediction in plant improvement, Genomic selection: genome-wide prediction in plant improvement, *Trends in Plant Sci.* **19(9)**: 592-601, <https://doi.org/10.1016/j.tplants.2014.05.006>.
- Frankel OH, and E Bennett (1970) Genetic resources in plants-their exploration and conservation. (Frankel, O.H. and Bennett, E., eds.), pp. 7–17. Blackwell, Oxford
- Heffner EL, ME Sorrells, and JL Jannink (2009) Genomic selection for crop improvement. *Crop Sci.* **49(1)**: 1-2. <https://doi.org/10.1186/s13059-019-1828-7>
- Kumar S, Jacob SR, MR Rouf, VK Vikas, K Pawan, C Tilak, K Satinder, K Uttam, K Suneel, SSHailendra, S Ravinder, P Sai, SA Mahendru, SA Kumar, K Jyoti, MS Saharan, BS Chander, P Manoj, K Sanjay, S Kuldeep (2022) Indian wheat genomics initiative for harnessing the potential of wheat germplasm resources for breeding disease-resistant, nutrient-dense, and climate-resilient cultivars, *Front Genet* **(13)** DOI=10.3389/fgene.2022.834366
- Li B, N Zhang, YG Wang, AW George, A Reverter, and Y Li (2018) Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front Genet.* **(9)** DOI=10.3389/fgene.2018.00237
- Liu Y, H Du, P Li, Y Shen, H Peng, S Liu, GA Zhou, H Zhang, Z Liu, M Shi, and X Huang (2020) Pan-genome of wild and cultivated soybeans. *Cell.* **182(1)**: 162-176. e13, ISSN 0092-8674, <https://doi.org/10.1016/j.cell.2020.05.023>.
- Lye ZN and Purugganan MD (2019) Copy number variation in domestication. *Trends Plant Sci.* **24(4)**: 352-365.
- Mahmoud M, N Gobet, DI Cruz-Dávalos, N Mounier, C Dessimoz, and FJ Sedlazeck (2019). Structural variant calling: the long and the short of it. *Genom. Biol.* **20(1)**: 1-4
- Montenegro JD, AA Golicz, PE Bayer, B Hurgobin, H Lee, CK Chan, P Visendi, K Lai, J Doležel, J Batley and D Edwards (2017) The pangenome of hexaploid bread wheat. *The Plant J.* **90(5)**: 1007-1013.
- Scholz AH, J Freitag, CH Lyal, R Sara, ML Cepeda, I Cancio, S Sett, AL Hufton, Y Abebaw, K Bansal, and H Benbouza (2022) Multilateral benefit-sharing from digital sequence information will support both science and biodiversity conservation. *Nat Commun.* **13(1)**: 1-5.
- Shang L, X Li, H He, Q Yuan, Y Song, Z Wei, H Lin, M Hu, F Zhao, C Zhang and Y Li (2022) A super pan-genomic landscape of rice. *Cell Res.* **12**: 1-9. <https://doi.org/10.1038/s41422-022-00685-z>

Wilkinson MD and IJ Aalbersberg et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. **3(1)**: 1-9

Zhou Y, D Chebotarov, D Kudrna, V Llaca, S Lee, S Rajasekar, N Mohammed, N Al-Bader, C Sobel-Sorenson, P Parakkal, and LJ Arbelaez (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. *Sci. Data*. **7(1)**: 113 <https://doi.org/10.1038/s41597-020-0438-2>

Zhuang Y, X Wang, X Li, J Hu, L Fan, JB Landis, SB Cannon, J Grimwood, J Schmutz, SA Jackson, and JJ Doyle (2022) Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nat. Plants*. **8(3)**: 233-44. <https://doi.org/10.1038/s41477-022-01102-4>.