

## Genomics-Driven Application of Plant Genetic Resources for Sustainable Agriculture

Rajeev K Varshney\*, Abhishek Bohra and Vanika Garg

Murdoch's Centre for Crop and Food Innovation, State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Murdoch, Western Australia, 6150, Australia.

Genebanks preserve the valuable genetic diversity in the form of huge collections of plant genetic resources (PGR). Dr Raj S. Paroda has spearheaded the modernization of one of the world's largest genebanks, the ICAR-National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India, which currently houses nearly 462,466 accessions. Dr Paroda's notable contributions in the field of PGR management have earned him the popular name of "Genebank Guru". As Dr Paroda turns 80 on August 28, 2022, we commemorate the occasion by synthesizing an article in line with Dr Paroda's forward-thinking vision for PGR management and utilization. The present article highlights how modern genetic technologies have expanded our capacity to characterize and exploit the unexplored genetic diversity from genebanks. Extensive molecular and phenotypic characterisation support efficient strategies to minimize large PGR collections to workable sizes, thus allowing targeted search of plant diversity for sustainable agriculture. As the technological innovations transform genebanks into "bio-digital resource centres", availability of detailed genotype-phenotype maps and genomic predictions for large PGR collections will guide selection and breeding decisions in crop improvement programmes. Effective preservation and use of PGR collections is imperative to impart climate change adaptation to crop production systems and safeguarding our future food supply.

**Key Words:** Accession, Genebank, Genome, Phenotype, Plant genetic resources, Trait

### Introduction

Plant genetic resources (PGR) harbour genetic variation that serves as raw material for breeding and selection decisions in crop improvement programmes. Padma Bhushan Dr Raj S. Paroda [Chairman, Trust for Advancement of Agricultural Sciences (TAAS) and Former Secretary DARE & DG, ICAR] has played a pivotal role in transformation of Indian national PGR system, represented by the ICAR-National Bureau of Plant Genetic Resources (NBPGR), New Delhi. To recognize his immense contribution in the field of PGR management, two genebanks have been named after him: 'RS Paroda Genebank' at ICRISAT (<https://www.genebanks.org/genebanks/icrisat/>) and genebank at the Agricultural Research Institute in Kazakhstan.

In recent years, evolving DNA sequencing and plant phenotyping have facilitated detailed characterisation and evaluation of PGR. For instance, we recently sequenced the genomes of 3,366 chickpea accessions from 'RS. Paroda Genebank' and provided genomic variations and haplotypes for future chickpea improvement

(Varshney *et al.*, 2021). Conforming to the vision of Dr Paroda, our article discusses how technological developments have enhanced efficiency of PGR management and their use in plant breeding and research. We begin with presenting a brief overview of major germplasm holdings across the globe. We then discuss the role of evolving DNA sequencing/genotyping and phenotyping systems in developing strategies for sustainable use of PGR. We also underscore the importance of targeted collection of under-represented germplasm from "hotspot" regions to bridge the conservation gaps in genebanks.

### Germplasm Holdings Across Major Genebanks Worldwide

Protection of plant genetic diversity is essential to sustainable increase in crop production and food security. *Ex situ* collection constitutes the major form of germplasm conservation that aims to conserve germplasm in controlled environments, away from their native habitats. An alternative germplasm conservation strategy

\*Author for Correspondence: Email-rajeev.varshney@murdoch.edu.au

stores accessions *in situ* in their natural habitats, allowing creation of new genetic variation because of continuous evolution. The growing realization of the urgent need of protecting genetic diversity for sustainable future has led to the establishment of the Svalbard Global Seed Vault by the Government of Norway (<https://www.seedvault.no/>). As a backup facility to the world's crop diversity, the global seed vault currently secures 1,165,041 seed samples from different genebanks including ICRISAT (Fig. 1 a, b).

According to the Second Report on The State of the World's Plant Genetic Resources for Food and Agriculture, a total of 7.4 million accessions are preserved in 1,750 genebanks worldwide. The 11 genebanks under the Consultative Group on International Agricultural Research (CGIAR) network contain 773,112 accessions. Though the accessions held in CGIAR genebanks represent 10% of the total accessions, the CGIAR network accounts for nearly 94% total germplasm shared in line with the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) framework. The Convention on Biological Diversity (CBD) and ITPGRFA constitute the mechanism of germplasm exchange across the globe. Besides, several national repositories also conserve germplasm accessions that protect genetic diversity for current and future use of breeders, researchers, farmers and other stakeholders. The other major national repositories National Plant

Germplasm System (NPGS), USA; ICAR-National Bureau of Plant Genetic Resources (NBPGR), India; Institute of Crop Germplasm Resources (ICGR)-Chinese Academy of Agricultural Sciences (CAAS), China; N.I. Vavilov All Russian Institute of Plant Genetic Resources (VIR), Russia; National Agriculture and Food Research Organization (NARO), Japan; Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Germany; and Canadian National Plant Germplasm System (CNPGRS), Canada collectively hold 2,282,718 accessions. The ICAR-National Bureau of Plant Genetic Resources (NBPGR), India represents the second largest genebank in the world that currently houses 462,466 accessions ([http://www.nbpgr.ernet.in/Research\\_Projects/Base\\_Collection\\_in\\_NGB.aspx](http://www.nbpgr.ernet.in/Research_Projects/Base_Collection_in_NGB.aspx)).

### Generation of Molecular Data on Large-scale Collections

Advancements in DNA sequencing technologies have led remarkable expansion in our capacity to perform in-depth molecular characterization of large PGR collections. In different crops, tens of thousands of genetic markers have been assayed on large PGR collections to elucidate the vast amount of genetic diversity they contain. Earlier, simple sequence repeat (SSR) markers were used to examine the genetic diversity of genebank accessions. For instance, the analysis of genotyping data and genetic structure of a total of 3,367 accessions from global



**Fig. 1.** a) Entrance to the Svalbard Global Seed Vault and b) Transferring seed boxes containing seeds of ICRISAT crops to the Seed Vault (Prof. Rajeev Varshney and Dr. Hari D. Upadhyaya, the then Research Program Director- Genetic Gains, and Head, Genebank, respectively at ICRISAT).. The global seed vault stores 1,165,041 backup seeds of 5947 species from 91 genebanks. The global seed vault provides seed backups of world's several genebanks. ICRISAT has deposited nearly 90% of its genebank seed as backup in the vault.

**Table 1. Characterization of diverse germplasm collections based on whole genome sequence information in some crop species**

Crop	Total accessions sequenced	Cultivated	Landraces	Wild	Others	Reference
Chickpea ( <i>Cicer arietinum</i> )	3,366	548	2,439	195	184	Varshney et al., 2021
	429	144	268	7	10	Varshney et al., 2019
Pigeonpea ( <i>Cajanus cajan</i> )	292	117	166	7	2	Varshney et al., 2017a
Soybean ( <i>Glycine max</i> )	2,898	1,747	1,048	103	-	Liu et al., 2020
	1,110	228	723	157	2	Bayer et al., 2022
Sunflower ( <i>Helianthus annuus</i> )	493	287	17	189	-	Hübner et al., 2019
Common bean ( <i>Phaseolus vulgaris</i> )	683	154	529	-	-	Wu et al., 2020
Pearl millet ( <i>Pennisetum glaucum</i> )	994	963	-	31	-	Varshney et al., 2017b
Rice ( <i>Oryza sativa</i> )	3,010	3,010	-	-	-	Wang et al., 2018
	1,143	1,143	-	-	-	Lv et al., 2020
Maize ( <i>Zea mays</i> )	103	60	23	19	1	Chia et al., 2012
Lettuce ( <i>Lactuca sativa</i> )	445	131	-	314	-	Wei et al., 2021
Sorghum ( <i>Sorghum bicolor</i> )	44	17	18	7	2	Mace et al., 2013
Tea ( <i>Camellia sinensis</i> )	81	58	20	3	-	Xia et al., 2020
Cotton ( <i>Gossypium</i> L.)	419	419	-	-	-	Ma et al., 2018
	318	271	35	12	-	Fang et al., 2017
Grape ( <i>Vitis vinifera</i> )	472	329	-	143	-	Liang et al., 2019
Spinach ( <i>Spinacia oleracea</i> )	305	295	-	10	-	Cai et al., 2021
Tartary Buckwheat ( <i>Fagopyrum tataricum</i> )	510	-	478	32	-	Zhang et al., 2021

composite germplasm collection represented majorly by landraces (89.5%) with 41 SSR markers allowed formation of ‘Reference Set’, a smaller representative subset of accessions for future use (Billot et al., 2013). A paradigm shift in marker genotyping assays and sequencing technologies has reflected in genetic profiling of plant genotypes (Garg et al., 2021). Table 1 shows the examples of genetic profiling large number of accessions as elucidated by whole genome resequencing. Next generation sequencing (NGS)-based methods based on reduced representation are particularly suitable for genetic profiling of large genebank collections. Notable examples include genetic profiling of 21,405 IPK barley accessions including domesticated and wild barley using genotyping-by-sequencing (GBS), thus revealing genetic variations in the form of 171,263 SNPs (Milner et al., 2019). Detailed genetic profiles of PGR offer opportunity to understand the molecular basis of

crop domestication and evaluation. The high-density genotyping datasets can be combined with historical evaluation records to develop genotype-phenotype maps for the PGR collections and genetically dissect the variation in agriculturally important traits using genome-wide association studies (GWAS). Examination of the genetic similarities based on genome-wide marker data may help revealing the redundancies within and between collections besides aiding in identification of collection gaps and mislabelling of biological status in the historical records in genebanks across the globe (Bohra et al., 2022a). Recent genome sequencing of 3,366 chickpea accessions allowed identification and correction in labelling of the chickpea accessions ICC 16369 (Varshney et al., 2021). The mislabelling of ICC 16369 as cultivated was evident owing to the presence of wild-specific allele of the *SHATTERPROOF2* homolog.

## Next-generation Phenotyping to Assist PGR Characterization

Characterization and evaluation of PGR is crucial to realize their true potential for genetic research and breeding. Plant phenotyping has always remained a great challenge in crop management and improvement. The traditional methods of monitoring phenotypic changes in plant morphological traits are cumbersome, destructive, costly, and inaccurate. Furthermore, the substantial influence of genotype  $\times$  environment interaction on phenotypic expression challenges the efficacy of the traditional methods of PGR characterization and evaluation. In recent years, advances in imaging and sensor technology have enabled automation of measurements of anatomical, physiological and biochemical changes in plants. Availability of such high-throughput phenotyping platforms have facilitated rapid plant phenotyping at different growth stages in accurate, cost-effective and non-invasive manner. Different kinds of imaging platforms including RGB, thermal, fluorescence, laser scanning, tomography, multispectral and hyperspectral offer plant phenotyping at varying scales i.e. microscope, laboratory, glasshouse, field and satellite (Nguyen *et al.*, 2022). Several national (<https://www.plantphenotyping.org/>; <https://www.plantphenomics.org.au/>) and international networks (<https://www.plantphenotyping.org/>) have been established across the globe to allow scientists/stakeholders in academia and industry accessing cutting-edge tools, facilities and analytics of plant phenomics. In the wake of growing sequence information, the acquisition of high-quality data opens exciting avenues for identification of beneficial alleles preserved in the large genebanks. Association of the high-throughput phenotypes with genetic loci has been demonstrated in various crops including rice, wheat, maize, barley, soybean, sorghum, cotton, rapeseed etc (Xiao *et al.*, 2022).

## Enhancing PGR use for Future Genetic Research and Crop Improvement

### Generation of Trait-specific Subsets for Efficient Use

Huge collections of PGR accessions held in genebanks are the reservoir of valuable traits, which may be essential to impart climate adaptation and stress tolerance to future crop varieties. For example, Isleib *et al.* (2001) documented the economic value of the tomato spotted wilt virus (TSWV) resistance introgressed from PI

203396 into peanut (*Arachis hypogaea*) cultivars in the USA, which amounted to more than \$200 million annually. Despite this, the use of PGR from the large genebank collections remains limited largely to the poor morphological and molecular characterization of the germplasm. Less than 10% of the germplasm accessions held in genebanks have been used so far in breeding programs (Nguyen and Norton 2020). Lack of trait-specific subsets of crop germplasm has also been a major limitation in the selection and use of appropriate accessions in pre-breeding and varietal development. Strategies based on the development of manageable germplasm diversity subsets have been proposed to enhance the use of PGR from genebanks. Brown' core collection concept aimed to capture 70% diversity in a smaller collection comprising 10% of the total accessions. Core collections have been developed in various crops including rice, maize, wheat, soybean. Later, the concept of mini core collection gained popularity given the fact that the size of core collections still poses challenge in terms of evaluation and management, especially in crops having large germplasm holdings (Anglin *et al.*, 2018). Phenotyping data in combination with genetic information has allowed construction of mini core collections comprising 1% of the total accessions in different crops including chickpea, pigeonpea, and several other crops of semi-arid regions.

Another strategy for efficient use of large PGR collection is the focused identification of germplasm strategy (FIGS) that uses environmental information for customization of the germplasm sets to a workable size (Bhullar *et al.*, 2009). FIGS is based on the premise that suitable germplasm for adaptive traits can be sampled from the sites experiencing selection pressures for the particular trait (Bohra *et al.* 2022b, c). In wheat, authors could narrow down on 1,320 landraces from the large collection of 16,089 accessions through focusing on 323 geographic sites having strong selection pressure for powdery mildew (Bhullar *et al.*, 2009). This FIGS-customized set of wheat facilitated targeted search for useful genetic variation for the resistance (*Pm3* gene) against powdery mildew disease. More recently, Haupt and Schmid (2020) made 'trait-focused' panels of 183 and 366 accessions in soybean for environmental adaptation by combining core collection and FIGS in a collection of >17,000 accessions, representing introduced landraces from the USDA Soybean Germplasm Collection. A combination of targeted approaches could prove highly

useful for identification of functional diversity for adaptation traits.

### **Genomic Predictions to ‘Turbocharge’ Genebanks**

Availability of genotype and phenotype data on germplasm sets may help train prediction models to compute the genetic worth of millions of accessions archived in the genebanks. Genomic selection bypasses the need of repeated phenotyping, and predicts the worth of an unobserved individual based on the models trained on individuals having both genotype and phenotype scores. Yu *et al.*, 2016 recently demonstrated a proof-of-concept for genomic prediction as a promising strategy to aid selection and breeding decisions. The high density GBS genotyping enabled molecular characterization of sorghum reference set comprising 962 accessions with 340,496 single nucleotide polymorphisms (SNPs). The authors selected a set of 299 accessions as training population to predict yield of the 633 untested lines of the reference set. The study recorded high prediction accuracies for the untested lines of the reference set and a 200-accession validation set. Relatively poor prediction accuracies for an independent set of 580 exotic sorghum accessions suggested less representation of alleles from exotic lines in the training population derived from the reference set. The advances in genotyping and phenotyping would help extending the genomic prediction strategy from small subsets of accessions to whole genebank collection to enable efficient use of PGR.

### **Analysis of the Existing Gaps to Inform Future Collections**

A systematic analysis of the representativeness of different taxa and species distributions in large *ex situ* collections is required for identification of the critical conservation gaps and setting of collection priorities (Ramírez-Villegas *et al.*, 2010). Advances in Geographic Information Systems (GIS) technologies have supported conservation planning efforts. With the help of GIS-modelled distribution, Ramírez-Villegas *et al.* (2010) identified *Phaseolus* wild relatives that are underrepresented in *ex situ* collections and the authors suggested ‘hot spot’ regions that should be targeted for future collection of high-priority germplasm.

### **Conclusions and Future Directions**

Approaches to accessing the suitable germplasm are greatly benefited by the growing acquisition of molecular information and phenotyping data on large-

germplasm collections. Extensive phenotypic and genetic characterization will be crucial to unlock the full potential of PGR held in genebanks via revealing the genomic regions controlling traits that are important to biotic and abiotic stress adaptation (Haupt and Schmid, 2020). The high-volume and complexity of the phenotyping data offers a new set of challenges in analysis; such as the need for data/image processing algorithms to analyse the data acquired from various imaging systems. The adoption of the new-generation plant phenotyping will rely on standardization of the high-volume datasets, which in turn highlights the growing requirement of new digital datasets to remain adhered with the guidelines or the checklists provided in the Findability, Accessibility, Interoperability, and Reuse (FAIR; <https://www.go-fair.org/fair-principles/>) and Minimum Information About a Plant Phenotyping Experiment (MIAPPE; <https://www.miappe.org/>). The rising availability of big data on PGR resulting from latest innovations in genetic and phenotyping technologies poses new challenges in terms of networking and equitable sharing of the resources and information. In this context, coordinated efforts are needed to address the access and benefit sharing issues particularly in relation to the developing countries.

### **Acknowledgements**

Authors are thankful to Food Futures Institute, Murdoch University (Australia) for financial support. RKV is thankful to all colleagues at ICRISAT and its collaborators, and donors for their contributions to germplasm research during his tenure at ICRISAT.

### **References**

- Anglin NL, A Amri, Z Kehel, and D Ellis (2018) A case of need: linking traits to genebank accessions. *Biopreserv. Biobank.* **16**: 337-349.
- Bayer PE, B Valliyodan, H Hu, JI Marsh, Y Yuan, TD Vuong, *et al.* (2022) Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *Plant Genome* **15**: e20109.
- Bhullar NK, K Street, M Mackay, N Yahiaoui, and B Keller (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 9519-9524.
- Billot C, P Ramu, S Bouchet, J Chanterreau, M Deu, L Gardes, *et al.* (2013) Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PLoS One* **8**: e59714.