

Review Article

Genoplasmics: Advancing Plant Germplasm Research through Genomics

Abstract

Plant genetic resources (PGR) are crucial for crop improvement programs. The National Genebank (NGB) at ICAR-NBPGR is responsible for collecting, conserving, and facilitating the utilization of genetic diversity in crop plants. Plant germplasm forms the foundation for plant genetic improvement. Extensive collections of germplasm have been amassed and stored, presenting the challenge of effectively harnessing and exploiting this valuable resource. Genomics-based plant germplasm research (GPGR), also known as genoplasmics, is an emerging interdisciplinary field that applies genomic principles and methods to the study of germplasm. This article outlines the concept, strategy, and approach of GPGR, highlighting recent advancements in core collection creation, germplasm enhancement through core collections, and gene discovery from core collections. GPGR represents a significant milestone in germplasm research, ushering in a new era of scientific investigation and innovation.

Keywords: Genomics, Genoplasmics, Genebank, Plant Genetic Resources, Genomic Principles.

Introduction

In the past few decades, agriculture has undergone significant global growth due to technological advancements. However, it faces challenges related to sustainable soil management, environmental degradation, natural disasters, and the need to meet increasing food demands. To address these challenges, there is a pressing need for a new "green revolution" focused on higher productivity, lower investment, and improved human nutrition (Hari *et al.*, 2008).

Researchers working with plant genetic resources encounter challenges in conserving, managing, and utilizing germplasm stored in gene banks. These challenges include developing strategies for representative sampling of individuals, improving long-term conservation methods through genetic analysis, and characterizing and assessing the genetic relationships among stored

accessions. Currently, there is a global effort to characterize and evaluate conserved germplasm, identify novel genes and alleles through genome-wide association mapping, utilize climate data for germplasm selection and deployment, and integrate modern genomic approaches into breeding programs (Lanteriet *et al.*, 2006).

Specific priorities include developing genomic resources and creating core sets for lesser-explored crops of Indian origin, such as minor pulses, oilseeds, and cereals. These core sets will be evaluated to establish reference sets for important traits using genome-wide association mapping. Additionally, efforts are being made to integrate modern genomic approaches into breeding programs and use climate data to enhance the utilization of plant genetic resources. By addressing these priorities, researchers aim to unlock the potential of plant genetic resources for sustainable agriculture and food production (Tester *et al.*, 2010).

Challenges and opportunities of plant germplasm in crop improvement

The field of plant germplasm research presents both challenges and opportunities in today's context. These can be summarized as follows:

Challenges:

1. **Genetic Erosion:** The loss of genetic diversity in cultivated plant species due to modern agricultural practices and the displacement of traditional landraces poses a significant challenge. This reduction in genetic variation limits the ability to address future agricultural challenges, such as climate change, pests, and diseases.
2. **Conservation and Management:** The effective conservation and management of plant germplasm collections in gene banks require substantial resources and expertise. Maintaining the viability of seeds, ensuring proper storage conditions, and preventing genetic drift are ongoing challenges.
3. **Characterization and Utilization:** There is a need to characterize and understand the traits present in plant germplasm collections. This involves identifying useful genes and alleles that can contribute to crop improvement. However, the vastness of germplasm collections and limited resources hinder comprehensive characterization efforts.
4. **Access and Benefit-Sharing:** Developing fair and equitable mechanisms for accessing and sharing the benefits derived from plant germplasm is a complex challenge. Balancing the

interests of breeders, farmers, and indigenous communities is crucial to ensure the sustainable use and conservation of germplasm resources.

Opportunities:

1. **Genetic Diversity for Crop Improvement:** Plant germplasm provides a vast reservoir of genetic diversity that can be harnessed for crop improvement. Through the exploration and utilization of diverse germplasm, researchers can develop new crop varieties with improved traits, such as yield, disease resistance, and stress tolerance (Tester *et al.*, 2010).
2. **Advances in Genomics and Technologies:** Rapid advancements in genomics, high-throughput sequencing, and molecular marker technologies offer unprecedented opportunities for plant germplasm research. These tools enable researchers to analyze and characterize germplasm collections more efficiently and comprehensively, facilitating targeted breeding efforts (Lanier *et al.*, 2006).
3. **Trait Discovery and Gene Mining:** Studying plant germplasm allows for the identification of novel traits and the discovery of genes underlying important agronomic characteristics. This knowledge can contribute to the development of crop varieties with enhanced productivity, nutritional value, and resilience to biotic and abiotic stresses.
4. **Climate Change Adaptation:** Plant germplasm research plays a crucial role in addressing the challenges posed by climate change. By identifying germplasm resources with traits that confer resilience to changing environmental conditions, researchers can develop climate-smart crops that are better adapted to future climates.
5. **Collaboration and International Cooperation:** International collaboration among gene banks, research institutions, and breeding programs is essential for sharing knowledge, resources, and germplasm. Collaborative efforts can accelerate research outcomes, improve germplasm conservation, and facilitate the exchange of genetic materials for crop improvement (Khawaja *et al.*, 2023).

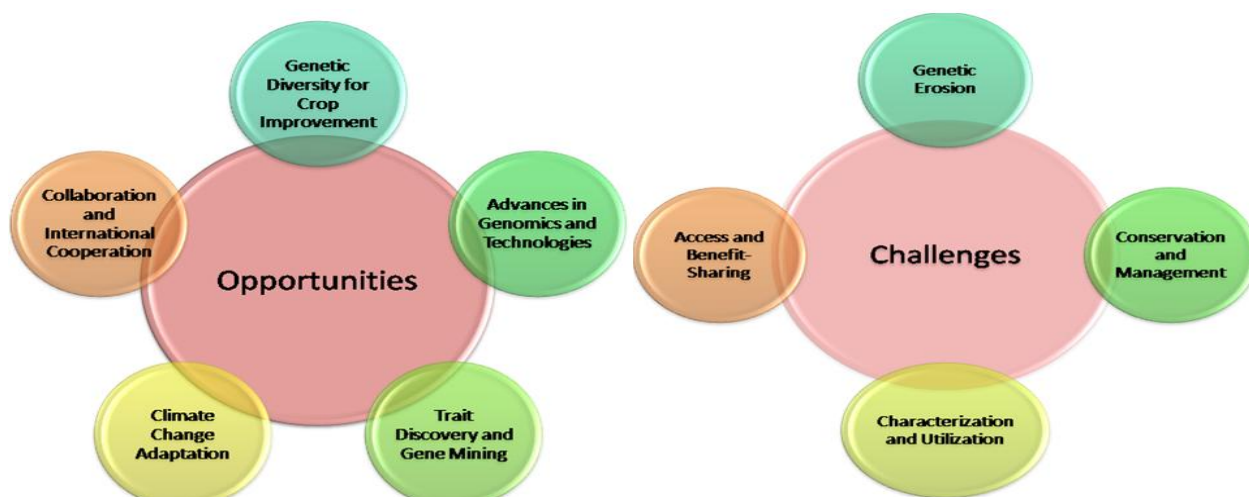


Figure 1. Challenges and opportunities of plant germplasm in crop improvement

Importance of genoplasmicsin crop improvement

Plant germplasm resources play a pivotal role in crop improvement across various areas of research and cultivation. The initial green revolution achieved success by utilizing dwarf germplasm sources in wheat (*Triticum aestivum*) and rice (*Oryza sativa*). Since then, the global recognition of collecting and preserving germplasm has grown, with over 6 million collections cataloged and conserved by 2001. However, despite extensive collection and conservation efforts, yield growth has been limited. While the number of germplasm accessions continues to increase, the overall genetic diversity of cultivated lands has declined, and crop yields have plateaued. Although germplasm resources harbor substantial genetic diversity, the identification of useful targets for marker-assisted selection (MAS) and transgenic breeding remains limited. These bottlenecks pose significant barriers to biotechnological solutions for enhancing crop productivity (Jia *et al.*, 2017).

In the past two decades, plant genomics has made rapid progress. Genetic and physical maps have been constructed for major crops, and complete genome sequences of *Arabidopsis thaliana* (Fu *et al.*, 2009), rice (*Oryza sativa*), corn (*Zea mays*), soybean (*Glycine max*), and wheat (*Triticum aestivum*) have been published. Despite these advancements, the impact on yield improvement has been modest. The main reason for this discrepancy is that genomic approaches typically focus on a limited number of accessions, while successful breeding maximizes allelic

variation by utilizing multiple accessions simultaneously. To harness specific alleles, a novel integration of genomics, germplasm research, and crop breeding is required (Jia *et al.*, 2017).

The combination of genomics and traditional genetics has given rise to various new disciplines such as comparative genomics, bioinformatics, evolutionary genomics, and developmental genomics. Genome-based plant germplasm research, also known as genoplasmics or GPGR, represents the convergence of genome and germplasm research and offers insights into genome content and strategies. These strategies include: (1) detecting genomic diversity within germplasm, (2) conserving and safeguarding germplasm based on knowledge of genomic diversity, (3) utilizing diversity information to establish representative core collections (CCs), (4) enhancing germplasm through the use of CCs, and (5) discovering novel alleles/genes within CCs. These five components are interconnected and synergistic, providing a holistic understanding of plant germplasm (Zou *et al.*, 2016), (Lanteriet *et al.*, 2006).

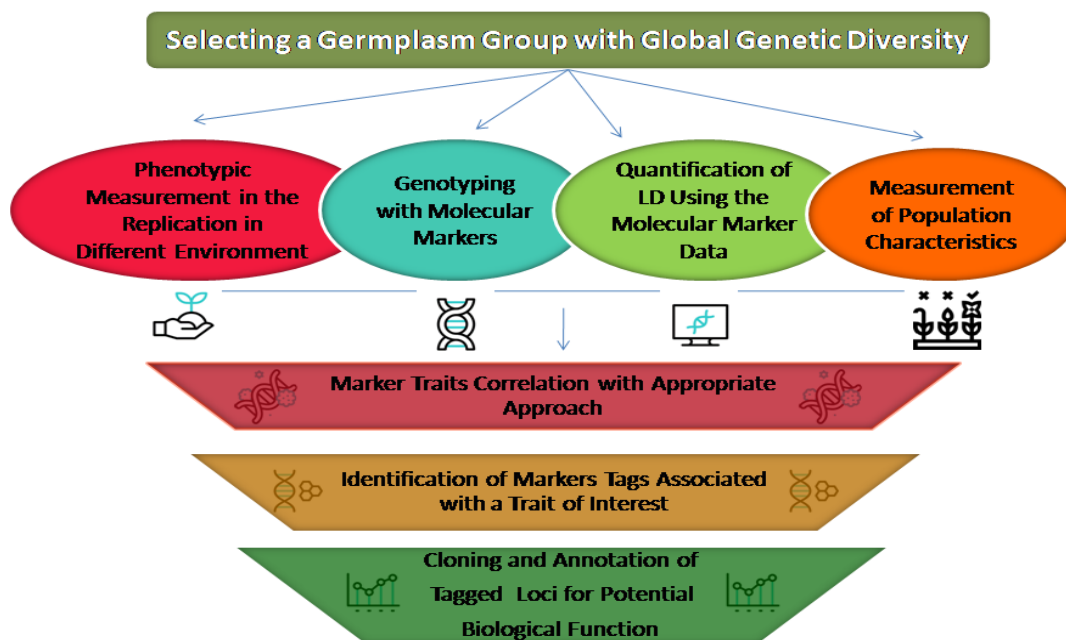


Figure 2. Selecting a germplasm assemblage characterized by worldwide genetic variation.

Leveraging genomic data, a plant core collection was meticulously crafted

Hariet *et al.*, (2008) developed a standardized approach for constructing core and mini core collections in plant germplasm research. This methodology involves the stratification of the entire collection based on taxonomic groups and country of origin. Accessions from neighboring

countries with similar agricultural climates are grouped together, resulting in reduced accessions within each group. Standard data, which eliminates scale differences, are then hierarchically subordinated (Hariat *et al.*, 2010). Ward's (1963) clustering algorithm is employed, as it optimizes an objective function by minimizing the number of clusters and maximizing the sum of squares between them. Within each cluster, approximately 10 percent of the accessions are randomly selected for inclusion in the core collection. In cases where there are fewer than ten accessions in a cluster, at least one accession is included (Hariat *et al.*, 2008). The reliability of this approach was tested using various measures such as mean, variance, Shannon-Weaver diversity index (H'), and feature frequency distribution. These measures were used to compare the entire collection with the core collection, ensuring that genetically controlled phenotypic associations were maintained (Hariat *et al.*, 2010).

A similar method was employed to create mini core collections, using the core collections as a basis. These mini core collections underwent further evaluation for agronomic and nutritional properties, and the resulting data were subjected to statistical analysis as described above. The core and mini core collections have been successfully established for various crops including chickpea, finger millet, foxtail millet, peanuts, pigeon pea, pearl millet, and sorghum. Additionally, core collections have been developed for barnyard millet, kodo millet, and little millet (Hariat *et al.*, 2008; 2010). The core collection typically represents around 10 percent of the accessions in the entire collection, while the mini core collection represents 1 percent of the entire collection or 10 percent of the accessions in the core collection. These collections aim to capture the diversity present in both the core and entire collections of a particular species preserved in a gene bank (Vetriventhan *et al.*, 2020).

Process of germplasm enhancement was revolutionized through the integration of genomics

Genomics-based germplasm enhancement involves using genomics tools and techniques to improve the quality and performance of plant germplasm (Lanteriet *et al.*, 2006). It integrates genomic information with traditional breeding methods to enhance the efficiency and precision of crop improvement programs. By leveraging genomic data, genetic markers, and high-throughput genotyping, specific genes or genomic regions associated with desirable traits are identified and manipulated. Genome-wide association studies (GWAS) and marker-assisted selection (MAS) aid in discovering genomic regions or candidate genes responsible for desired

traits and selecting plants with desired characteristics. Advanced genetic engineering techniques like CRISPR-Cas9 are employed to directly modify genes or introduce novel genes for enhancing traits. This integration of genomics into germplasm enhancement efforts expedites the development of improved crop varieties with higher yield, resistance to pests and diseases, and adaptability to changing environmental conditions. Germplasm improvement is crucial for increasing breeding and genetic/genomic stock, and the Mini Core Collection (MCC) enrichment strategy helps address challenges related to genome conservation by combining the development of elite germplasm with variety improvement (Lanteriet *al.*, 2006). The resulting multi-allelic genomic introgression lineage (GMA-IL) population inherits genetic diversity, serves as a platform for gene discovery, and facilitates the integration of plant genomics, breeding, and germplasm research (Zou et al., 2016).

Gene discovery propelled by genomics advancements

Genomics-based gene discovery refers to the process of identifying and characterizing genes using genomics tools and techniques. It involves the analysis of genomic data to identify specific genes or genomic regions associated with particular traits or functions of interest. Genomics has revolutionized the field of gene discovery by enabling the study of entire genomes, including coding and non-coding regions, and providing insights into gene structure, regulation, and function. Genomics-based gene discovery involves sequencing the organism's genome, annotating genes, comparing genomes, studying gene expression, understanding gene function, and analyzing genomic data (Lanteriet *al.*, 2006). It has greatly enhanced our understanding of genes, genetic variation, and molecular mechanisms underlying traits and diseases. This knowledge benefits areas like crop improvement and personalized medicine. Gene discovery methods include map-based, association-based, allele extraction, and relative genome-based approaches. The availability of diverse genomic technologies and populations like the CC has accelerated gene discovery and integration of genomics into germplasm research (Lanteriet *al.*, 2006).

Gene discovery and cloning through map-based genomics approaches

In 1992, Arondel et al. achieved a significant milestone in plant research by isolating a gene responsible for omega-3 fatty acid desaturation from *Arabidopsis* using map-based gene

cloning (Arodelet *et al.*, 1992). Since then, numerous agronomically important genes, including 63 in rice, 17 in wheat, and 10 in maize, have been successfully cloned using this method (Jia *et al.*, 2017). The establishment of genetic linkage maps for various crop species has provided a valuable framework for gene discovery. Gene mapping can infer allelic relationships and identify novel genes. Around 50% of genes have been identified through genetic mapping, and this approach is particularly effective for de novo gene discovery. Genetic mapping, especially using Introgression Lines (ILs), has facilitated the identification of useful genes from wild relatives or land races. For instance, IL mapping in tomato led to the discovery of the fw2.2 gene, which contributes to increased fruit weight. While map-based gene discovery has been successful, it often taps into only a small fraction of the genetic diversity in germplasm collections (Zhanget *al.*, 2015). To overcome this limitation, the use of large numbers of mapping populations, particularly those derived from Core Collections (CCs), is a promising strategy (Zou *et al.*, 2016), (Jia *et al.*, 2017). Expanding reference genomes and advances in sequencing and genotyping techniques will further enhance the isolation of functional genes through map-based cloning.

Utilizing genome-wide association mapping, gene discovery is achieved through association-based approaches

Historically, linkage disequilibrium (LD) or genome-wide association studies (GWAS) have been widely utilized in medical genetics and have recently gained popularity in plant research due to the decreasing cost of sequencing and genotyping (Yanget *al.*, 2015). GWAS is particularly valuable in germplasm studies because it allows for simultaneous testing of multiple allelic loci in populations with diverse phenotypes, enabling assessment of numerous traits and providing high mapping resolution. This approach also allows for the evaluation of natural and artificial selection effects and the discovery of agronomically important genes. Recent research has focused on LD in various major and minor crops, revealing species-dependent variations in LD decay. For example, autogamous species like rice and foxtail millet exhibit LD decay within approximately 100kb (Zhanget *al.*, 2015), while allogamous species like maize display decay within 1kb. Genome-wide scans for selection signatures offer an alternative approach to identify novel genes of agronomic importance, although gene function and phenotype relevance may be uncertain (Zhanget *al.*, 2015). Analysis of single nucleotide polymorphisms (SNPs) has

suggested that around 2-4% of maize genes have experienced artificial selection, potentially impacting approximately 1200 genes in the maize genome. Resequencing analysis has identified 928 and 1106 candidate genes significantly affected by artificial selection during soybean domestication and genetic improvement, respectively (Jia *et al.*, 2017). GWAS analysis has also successfully identified specific genes, such as OsSPL13, a grain-size gene in cultivated rice (Yanget *al.*, 2015). Advancements in GWAS include extreme-phenotype GWAS (XP-GWAS) for variant identification using pooled individuals from a diversity panel, and bulked sample analysis (BSA) using extreme phenotypes from diverse populations (Ritchie *et al.*, 2015), (Jia *et al.*, 2017). The mixed linear model (MLM) or linear mixed model (LMM) is currently the preferred method for QTL detection, enabling analysis of large-scale omics data, including GxG epistasis and GxE environment interaction effects, in a timely manner (Ritchie *et al.*, 2015). (Jia *et al.*, 2017) have implemented mixed linear models to simultaneously dissect various genetic effects in data sets containing up to 120,000 individuals.

Gene discovery rooted in allele mining methodologies

Phenotypic variations arise from genetic modifications within coding regions, introns, and promoters. For instance, a single nucleotide change in Rht1 leading to a stop codon results in dwarfism in wheat, contributing to the green revolution. Similarly, alterations in the promoter region, such as nucleotide substitutions, can give rise to phenotypes like fw2.2. Detection of allelic variations in germplasm collections has been facilitated through the removal of tagged sequences. Resequencing and EcoTILLING are the main methods employed for allele mining. Resequencing has identified multiple allelic versions of genes like RPP13 and GmF3'H/GmF3'5'H in Arabidopsis and soybean, respectively. EcoTILLING allows for efficient simultaneous detection of variations in many individuals and has been utilized for large-scale studies of genetic variation in species like *Populus trichocarpa*. In rice, EcoTILLING has been applied to genotype accessions and discover allelic variations in key salt-related genes (Negrao *et al.*, 2013). To enhance allele mining efficiency, the development of NILs and GMA-IL is underway, enabling large-scale exploration of germplasm collections.

Gene discovery facilitated by comparative genomics analyses

Comparative genomics is a field that compares gene structure, sequence, and function across different organisms (Jia *et al.*, 2017). Two approaches are used for gene discovery in comparative genomics. The first approach utilizes comparative genetic mapping to guide map-based cloning, leading to the identification of genes like Vrn1, Vrn2, Ph1, and Rpg1 in wheat and barley (Jia *et al.*, 2017). The second approach relies on the knowledge of gene function and sequence in a model species, leading to the isolation of genes such as Rht1 and Rht2 in wheat (Jia *et al.*, 2015). Comparative genomics has revealed that genes involved in disease resistance often share similar structures, with families like NBS-LRR, LRR-TM, PK, and LRR-TM-PK being recognized (Negrao *et al.*, 2013), (Jia *et al.*, 2015). The candidate gene approach based on conserved motifs in these families has been successful in cloning disease-resistant genes. However, this approach may not always yield novel genes (Negrao *et al.*, 2013). Functional divergence detection is expected to play a larger role in future comparative genomics-based gene discovery (Jia *et al.*, 2017). Interspecific comparisons using data from multiple genome sequencing projects have been valuable for improving gene annotations, automating annotation processes, and identifying novel coding regions, splice variants, microRNA precursors, and small peptide-encoding open reading frames. These extrinsic strategies combine gene prediction, expression, and homology data to identify conserved gene candidates (Cibrian *et al.*, 2010)

Enhancing germplasm through comprehensive panomics strategies

PANOMICS involves integrating various types of data, such as genome, RNA, proteins, metabolites, and phenome, to create models for predicting complex traits (Lanteriet *et al.*, 2006). This approach helps reduce false positives and improves genotype-phenotype predictions compared to using single data sources alone. Three strategies based on high-throughput technologies are employed: (1) combining genome selection with environment-dependent PANOMICS analysis and deep learning to enhance the accuracy of trait performance prediction using markers, (2) utilizing PANOMICS at sub-tissue, cellular, and sub-cellular levels to gain insights into the functions of selected markers, and (3) integrating PANOMICS with genome editing and speed breeding techniques for efficient large-scale validation of trait-specific precision breeding (Weckwerth *et al.*, 2020). An important example of PANOMICS application is the successful development of submergence-tolerant rice varieties through marker-assisted

backcrossing, specifically targeting the SUBMERGENCE 1 (SUB1) locus from FR13A into modern high-yielding rice varieties (Balley *et al.*, 2010).

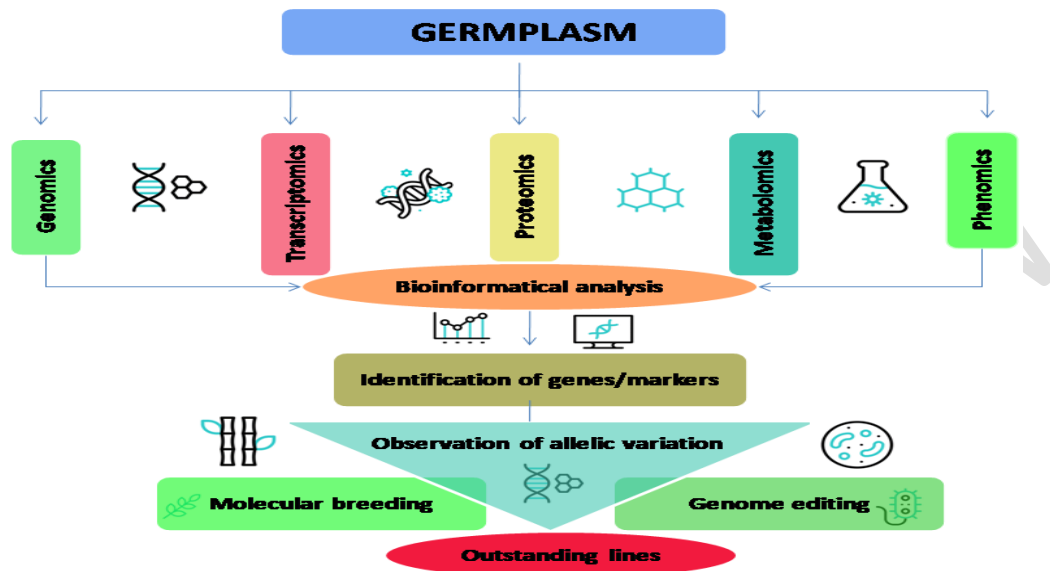


Figure 3. Utilizing diverse omics methodologies to cultivate exceptional lineages

Utilizing metabolic GWAS for advancing germplasm enhancement (Weckwerth *et al.*, 2020)

The combination of genome wide association studies (GWAS) with metabolomics has become a valuable approach to investigate the genetic and biochemical factors in crop plants (Weckwerth *et al.*, 2020). By analyzing intermediate traits that reflect the plant's biochemical and physiological status, researchers can gain insights into the genetic determinants present in the vast range of intraspecific variation (Keurentjes *et al.*, 2009). This strategy has been facilitated by advancements in high-throughput mass spectrometry-based analytical platforms (GC-MS/LC-MS) and genome sequencing technologies (Yanget *et al.*, 2015). Initially applied in *Arabidopsis thaliana*, GWAS combined with metabolomics (mGWAS) (Weckwerth *et al.*, 2020), (Chanet *et al.*, 2010) has been extended to numerous crop species. A recent study conducted by (Li *et al.*, 2019) identified 65 primary metabolites in different tissues with distinct patterns (Jie *et al.*, 2015). They examined 350 quantitative trait loci (QTLs) associated with these metabolites (Fanget *et al.*, 2019), which were unevenly distributed across the genome and included two QTL hotspots (Weckwerth *et al.*, 2020). mGWAS, focusing on individual metabolite content and ratios under varying conditions, not only enhances our understanding of metabolic diversity in response to changing

environments (Yanget *al.*, 2015), but also aids in identifying key regulators involved in stress responses (both biotic and abiotic) (Weckwerth *et al.*, 2020)..

Leveraging panomics-guided genome editing to achieve precision breeding, enhancing both climate resilience and nutritional value within germplasm(Weckwerth *et al.*, 2020).

Genome editing technologies (Weckwerth *et al.*, 2020). have revolutionized the precise manipulation of genomic sequences, enabling various modifications such as point mutations, gene knockouts, gene activation or repression, and epigenetic changes. These advancements, combined with the integration of PANOMICS and genome editing tools, hold great potential for the development of precision breeding (Lanteriet *al.*, 2006). By employing genome editing, it becomes possible to understand the genetic basis of different phenotypic groups at the population level, analyzing multiple target genes in parallel to study the loss of function and resulting trait alterations. The combination of genome editing and speed breeding approaches further facilitates the validation of incorporated genes without the need for complex in vitro manipulations (Kamburovaet *al.*, 2017). An important application of PANOMICS in germplasm collections is the improvement of nutritional value in crops (Weckwerth *et al.*, 2020), balancing stress resistance and productivity. This can be achieved through the integration of multi-omics characterization, considering the production of proteins, carbohydrates, fats, vitamins, minerals, and genomic selection (Weckwerth *et al.*, 2020). Moreover, this approach has the potential to enhance the production of other cereal crops like wheat, sorghum, and pearl millet, developing elite lines and improving existing germplasm for enhanced stress tolerance and nutritional value (Vetriventhanet *al.*, 2020). However, the successful implementation of precision breeding relies on the development of improved infrastructure, adherence to ethical norms, and the establishment of more powerful computational tools as routine resources.

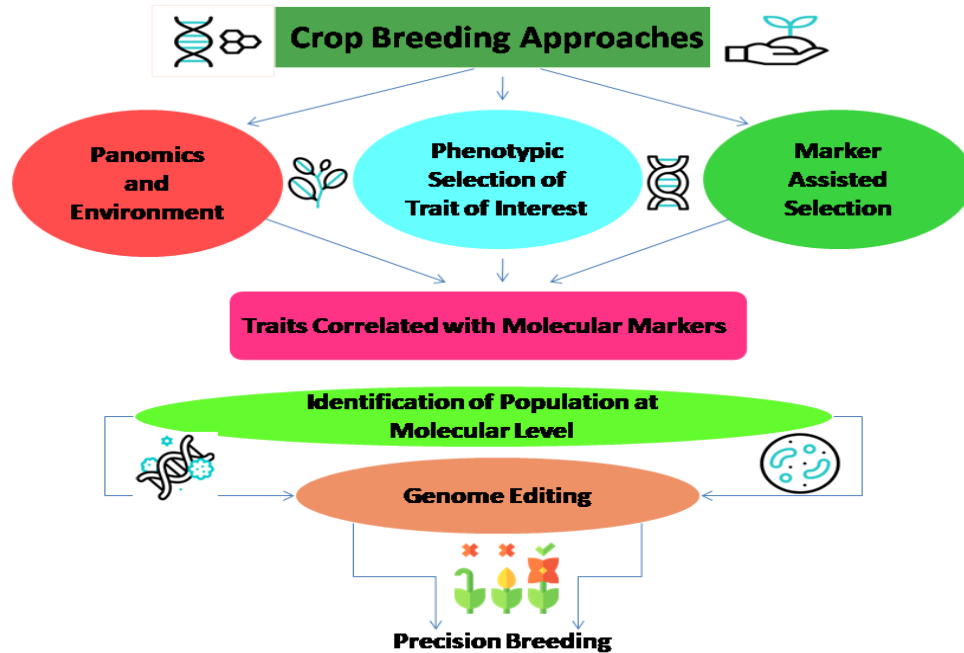


Figure 4. Strategies in crop breeding employed to achieve precision in genetic improvement

DNA banks play a pivotal role in enabling the utilization of genomics in plant germplasm applications(Kole, 2013)

DNA banks play a crucial role in facilitating the application of genomics to plant germplasm by providing access to a large number of plant accessions for high-throughput analysis of plant genomes(Kole, 2013). These banks make DNA from diverse plant materials readily available, enabling gene discovery and marker identification in germplasm collections. The availability of genetically diverse material is essential for gaining insights into within-species diversity and improving our understanding of plant genetic resources (Vetriventhan *et al.*, 2020). Marker discovery, made possible through DNA banks, is highly valuable for the management, sustainable utilization, and conservation of various species. Additionally, markers can be utilized to enhance breeding programs and, in certain cases, for biotechnological applications (Kamburova *et al.*, 2017). It is essential to have internationally coordinated efforts to develop and maintain DNA collections, ensuring their continual growth as a crucial resource for plant biology.

Employing molecular markers to profile and safeguard crop plant germplasm diversity(Varshney, 2009)

The advent of genomic DNA-based marker assays has greatly transformed our ability to characterize genetic variation and facilitate genetic selection in crop plants (Varshney, 2009). Molecular markers have emerged as highly effective and reliable tools for studying genome architectures and investigating gene polymorphisms in crops (Varshney, 2009). Techniques such as RFLP and PCR-derived markers have found extensive applications in plant genetics and breeding, enabling the mapping of Mendelian genes and QTLs (Zhanget *al.*, 2015). By utilizing molecular markers, it becomes possible to identify the presence of valuable traits, genes, and alleles, which in turn helps inform decisions regarding the multiplication of accessions and maintenance of seed stocks to meet the anticipated demand for plant materials (Varshney, 2009). Implementing the use of molecular markers for characterizing and conserving genetic resources allows for the addition of genotypes with known and beneficial genes and alleles into core collections, thereby enhancing their value and usability for breeders. A novel approach that holds promise is the establishment of collections based on the knowledge of valuable genes and traits, enabling more targeted and valuable utilization of crop plant germplasm resources (Kamburovaet *al.*, 2017).

Conclusion

The progress made in plant genomics has opened up exciting avenues for studying plant germplasm, leading to the introduction of the term GPGR or genoplasmics to describe this fusion. As genomics continues to advance rapidly, it will soon be necessary to sequence the complete genomes of numerous crucial plant species. Additionally, functional genomics is expected to provide valuable insights into genetic functions. Mapping populations, particularly NIL and CC-IL, have been developed from germplasm, contributing to gene discovery and the preservation of diversity (Vetriventhanet *al.*, 2020).The 21st century is characterized by significant advancements in life sciences, with omic technologies such as genomics, transcriptomics, proteomics, and metabolomics leading the way. The integration of these omic technologies with germplasm will give rise to novel plasmic technologies such as metaboloplasmics and proteoplasmics. Omics technology is widely employed in various domains, including human, microbial, and animal studies, and its continued progress will drive further research in plant germplasm. We can envision a future where our understanding of plant germplasm becomes deeper and more comprehensive, enabling its efficient collection,

utilization, and storage in agricultural gene banks. This will bring immediate and long-term benefits to humanity.

References :

Arodel V., Lemieux B., Hwang I., Gibson S., Goodman H. M., Somerville C. R. Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. *Science*. 1992;258:1353–1355.

Balley-Serres J., Fukao T., Ronald P., Ismail A., Heuer S., Mackill D. Submergence tolerant rice: SUB1's Journey from landrace to modern cultivar. *Rice*. 3(2): 2010; 138-147.

Chan E. K., Rowe H. C., Hansen B.G., Kliebenstein D. J. The complex genetic architecture of the metabolome. *Plos Genet*. 2010; 6:1001198.

Cibrian-Jaramillo A., De la Torre-Barcena J. E., Lee E. K., Katari M. S., Little D. P., Stevenson D. W., DeSalle R. Using phylogenomic patterns and gene ontology to identify proteins of importance in plant evolution. *Genome biology and evolution*. 2010; 2: 225-239.

Fang C. Y., Fernie A. R., Luo J. Exploring the diversity of plant metabolism. *Trends Plant Sci*. 2019; 24: 83-98.

Fu J., Keurentjes J. J., Bouwmeester H., America T., Verstappen F. W., Ward J. L., Beale M. H. System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nat. Genet*. 2009; 41: 166-167.

Hari D. Upadhyaya., Devvart Yadav., Naresh Dronavalli C. L. L., Gowda., Sube Singh. Mini core germplasm collections for infusing genetic diversity in plant breeding programs. *Electronic journal of plant breeding*. 2010; 1(4): 1294-1309.

Hari D. Upadhyaya, R. P. S. Pundir., C. L. L. Gowda V., Gopal Reddy., S Singh. "Establishing a core collection of foxtail millet to enhance the utilization of germplasm of an underutilized crop", *Plant Genetic Resources*. 2008.

Jia J.Z., Gao L.F., Zhao G.Y. Crop genomics and crop science revolutions. *Sci. Agric. Sin*. 2015;48: 3316–3332.

Jia J., Li H., Zhang X., Li Z., Qiu L. Genomics-based plant germplasm research (GPGR). *The Crop Journal*. 2017; 5(2): 166-174.

Jie Luo. Metabolite-based genome-wide association studies in plants. *Curr Opin Plant Biol*. 2015; 24:31-38.

Kamburova V. S., Nikitina E. V., Shermatov S. E., Buriev Z. T., Kumpatla S. P., Emani C., Abdurakhmonov I.Y. Genome editing in plants: an overview of tools and applications. *Int J. Agronomy*. 2017; 1-15.

Keurentjes J. J. Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* 2009; 12: 223-230.

Kole C. Genomics and breeding for climate-resilient crops. New York: Springer. 2013; 441.

Lanteri S, Barcaccia G. Molecular markers based analysis for crop germplasm preservation. In: Ruane J, Sonnino A. The role of biotechnology in exploring and protecting agricultural genetic resources, *FAO*, Rome. 2006; 55-66.

Li K., Wen W., Alseekh S., Yang X., Guo H., Li W., Wang L. Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement. *Plant J.* 2019; 99(2): 216-230.

Tester M., Langridge P. Breeding technologies to increase crop production in a changing world. *Science*. 2010; 327: 818–822.

Negrao S., Almadanim M. C., Pires I. S., Abreu I. A., Maroco J., Courtois B., Gregorio G. B., McNally K. L., Oliveira M. M. New allelic variants found in key rice salt-tolerance genes: an association study. *Plant Biotechnol. J.* 2013; 11: 87-100

Ritchie M. D., Holzinger E.R., Li R.W., Pendergrass S. A., Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat rev Genet.* 2015; 16: 85-97.

Varshney R. K. Gene-based marker systems in plants: high throughput approaches for marker discovery and genotyping. *Molecular Techniques in Crop Improvement: 2nd Edition*. 2009; 119-142.

Vetriventhan M., Vania C. R., Azevedo H. D. Upadhyaya, A. Nirmalakumari et al. "Genetic and genomic resources, and breeding for accelerating improvement of small millets: current status and future interventions", *The Nucleus*. 2020.

Ward, H. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963; 58: 236-244.

Weckwerth W., Ghatak A., Bellaire A., Chaturvedi P., Varshney R. K. (2020). PANOMICS meets germplasm. *Plant Biotechnology Journal*. 2020; 18(7): 1507-1525.

Weckwerth W. Green Systems Biology From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J Proteomics*. 2011; 75: 284-305.

Yang J. L., Jiang H. Y., Yeh C. T., Yu J. M., Jeddeloh J. A., Nettleton D., Schnable P. S. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. *Plant J*. 2015; 84: 587–596.

Zhang F. T., Zhu Z. H., Tong X. R., Zhu Z. X., Zhu J. Mixed linear model approaches of association mapping for complex traits based on omics variants. *Sci. Rep*. 2015; 5: 10298.

Zou C., Wang P. X., Xu Y. B. Bulk sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J*. 2016; 10: 1941–1955.

UNDER PEER REVIEW