

Dissecting the Genetic Basis of Salt Tolerance in Indian Mustard Using GWAS

Vijayata Singh, Vivek Singh, Ankit Kumar, Zeetendra Singh and Jogendra Singh

¹ICAR-Central Soil Salinity Research Institute, Karnal-132001, India

²ICAR-Central Soil Salinity Research Institute, Regional Research Station, Lucknow Uttar Pradesh, India

Jogendrasingh82@gmail.com

Globally the total area of saline soil is 397 million ha and 434 million ha of sodic soil. Of the 230 million irrigated fields, 45 million ha (19.5%) are salt-affected and almost 1500 million ha of arable agriculture, 32 million (2.1%) salt-affected. Of the world's salt-affected land, an estimated 6.73 million ha are in India (Singh et al., 2014). Extreme events, climatic aberrations and anthropogenic interventions are likely to aggravate the extent of these degraded soils further. Despite the importance of salinity on crop production worldwide, and the abundance of knowledge gathered on genes/mechanisms involved in salinity tolerance, there has been surprisingly little effort to breed for improved salinity tolerance. Therefore, it is crucial to enhance the productivity of less productive lands, including salt-affected areas, in addition to increasing crop yields. Developing salt tolerant crop varieties is critical because the increased water scarcity will demand alternate sources like those having salinity and residual alkalinity.

Indian mustard [*Brassica juncea* (L.) Czern & Coss] is a major oilseed crop in these areas. However, salt affects as much as 50–90% of worldwide yield reduction. Salt tolerance is a very complex factor controlled by many independent and/or interdependent mechanisms and genetic modifications that lead to many changes in physiology and biochemistry at the cellular level. The classical methods of plant breeding for salt tolerance involve the widespread use of inter and intraspecific variations in the available germplasm which is essential for any crop development program. This large germplasm is then tested under various salt levels in microplots, which is a quick, reliable, reproducible and inexpensive method of salt tolerance. Genotypes that have demonstrated better stress tolerance without significant yield reduction are considered tolerant and used as potential donors in breeding programs. The six Brassica species have the highest agricultural importance and are known as "Brassicas". The U triangle by Nagaharu (1935) shows the relationship among Brassicas. Out of

the nine major oilseeds crops, Rapeseed-Mustard; contributes more than 24% of the total oilseeds production in the country (Anonymous, 2018-19). Inter and intraspecific variations in Brassica species show the potential of molecular breeding to select salinity-tolerant genotypes. This variation can be exploited extensively from the germplasm of Brassicas by association mapping. Using modern molecular tools like NGS, GWAS, OMICs, and Genome editing, along with classical breeding systems, further aids in developing salt-tolerant mustard varieties in a short period. The kinship between genotypic polymorphism and the phenotypic variation within a species is of fundamental agricultural interest. To predict risk factors at the genetic level for agronomically important traits like yield, biotic and abiotic stress in crops need a mutual understanding of both the genetic architecture and phenotype of a trait. This association between the phenotype and genotype has been of the keystone at least since the father of genetic "Mendel" postulated the existence of 'internal factors' that are inherited by the forthcoming generation. In this approach, we consider an advanced technique for interconnecting the mustard genotype-salinity phenotype data, which is Genome-Wide Association Mapping.

Genome-wide association study for salinity tolerance in Mustard

Genome-wide association study (GWAS) is an observational study that examines a genome-wide set of genetic variations in numerous individuals to determine if any variation is associated with a specific trait (Shikha et al., 2021). This technique helps to identify the underlying genetics of a particular trait, by targeting relationships between SNPs (single base-pair changes or mutations within DNA) and the desired traits. The effectiveness of association mapping has been enhanced with the advent of Next-Generation Sequencing (NGS) based genotyping platforms, which can capture SNPs using genotyping technologies such as customized, chip-based microarray technology, Illumina, Affymetrix, or other NGS technologies. GWAS was initially

established for the analysis of the human genome, but it has now been successfully employed in other organisms such as Arabidopsis, rice and many other crops for a variety of traits (Korte et al., 2013). Despite its success, QTL mapping has two main limitations. Firstly, only allelic diversity that segregates between the parents of the particular mapping population can be assayed, and secondly, the amount of recombination that occurs during the creation of the mapping population has a low resolution of polygenic complex traits. The information gathered from GWAS resolves these problems and improves the efficiency of marker-assisted selection (MAS) in mustard breeding (Singh and Singh, 2015).

Molecular mapping

Molecular mapping is a method used to represent genetic markers in a chromosome, along with the genetic distance between them. This method involves two types of mapping techniques: linkage mapping and association mapping. Linkage mapping measures the recombination between genetic markers and an unknown gene, referred to as linkage. It involves testing only two alleles and uses a biparental or multi-parent mapping population that is produced by crossing. Association mapping, on the other hand, measures the correlation between marker alleles and alleles in a population. This technique does not require crossing to prepare a mapping population, and it works with existing germplasm. More than two alleles can be tested in association mapping.

General steps in the association mapping

A general outline of the GWAS procedure (Fig. 1) must be involved the following steps

i. Selection of an association panel: A panmictic or randomly mating diploid population with at least 200 lines and without any formal organization is required for a Genome-Wide Association Study (GWAS) panel (Galli et al., 2020).

ii. Phenotyping of panel: Phenotyping is a crucial step in any molecular approach, and it should include replicated trials conducted over different locations and years under salt and control conditions. A suitable experimental design like augmented design, Randomized Block Design (RBD), nested design, etc. should be used to reduce environmental effects.

iii. Genotyping: The panel is tested with evenly distributed molecular markers over the entire genome, and the markers should be unlinked, i.e., more than 40 cM apart in the mustard genome.

iv. Structure and kinship analysis: The molecular data is analysed using the STRUCTURE package to estimate population

structure, and the TASSEL program analyses kinship among individuals of the population.

v. Genotypic for Linkage Disequilibrium analysis: Linkage disequilibrium (LD) or gametic phase disequilibrium is the non-random association of alleles at linked loci. It is a measure of the tendency of some alleles to be inherited together on a haplotype descended from the ancestral chromosome. There are different measures of LD like D, D², D*, and r². The panel is also genotyped with a large number of SSR and SNP markers so that LD between the marker and loci of interest can be estimated, which determines the number of markers needed for adequate coverage of the whole genome.

vi. Association mapping and LD analysis: Different software is used to detect an association between the phenotype and genotype data to minimize false association. Different models are developed to balance the confounding effects of the structure. Statistical free software packages like Trait Analysis by aSSociation, Evolution and Linkage" (TASSEL), PowerMarker, Graphical overview of linkage disequilibrium (GOLD), GeneAMap, FaST-LMM, STRUCTURE, R, MTDFREML and GAPIT. Commercial packages like GeneStat, ASREML, JMP Genomics, and SAS are also used (Glaubitz et al., 2014; Singh and Singh, 2015).



Figure 1. General outline of genome-wide association studies

The haplotype map, or "HapMap," is a modern technique that enables researchers to identify genes and genetic variations that are associated with health and disease. The HapMap has reduced the number of SNPs (single nucleotide polymorphisms) required to analyse the entire genome for a relationship with a phenotype from 10 million to approximately 500,000 tag SNPs, making it easier to study genetic variations and their impact on human health. (Dakouri et al., 2021).

Role in salinity tolerance

Genome-wide association studies (GWAS) can be used to understand the molecular pathways involved in mitigating salinity, identify new candidate genes, eliminate genetic variation at the molecular level, explore gene-environment

interactions, and identify alleles responsible for salt stress phenotype. This information can help plant breeders make better decisions to feed the growing population, improve crop yield, food quality, and increase resistance to salt stress. (Sandhu et al., 2019).

Limitations

Genome-wide association studies (GWAS) produce results that require confirmation in independent samples from diverse populations, which can be both expensive and time-consuming. These studies identify associations, not necessarily causation, and only highlight specific genetic locations rather than entire genes. Furthermore, GWAS focus on common variants and many of the associated variants may not be causal. The costs associated with collecting and pooling DNA samples are also high. Unfortunately, funding agencies for GWAS research are not readily available in India (Korte et al., 2013).

Conclusion

The modern Brassica cultivar has a limited genetic base and struggles to adapt to harsh environmental conditions, like salinity stress. However, there is hope for breeding salt-tolerant Brassica plants due to the significant variability within and between Brassica species. To achieve this, a proper breeding system assisted by markers is necessary to connect the salt tolerance factor and the QTL site to the chromosome. By using genome-wide association mapping, we can exploit the variability present in Brassica germplasm and link phenotypic and genotypic data.

References

- Anonymous. (2019). Department of Agriculture, Cooperation & Farmers Welfare. www.agricoop.nic.in. 94.
- Dakouri, A., Lamara, M., Karim, M. M., Wang, J., Chen, Q., Gossen, B. D., Strelkov, S. E., Hwang, S., Peng, G., and Yu, F. (2021). Identification of resistance loci against new pathotypes of *Plasmodiophora brassicae* in *Brassica napus* based on genome-wide association mapping. *Scientific reports*, 11(1), 1-11.
- Galli, G., Alves, F. C., Morosini, J. S. and Fritsche-Neto, R. (2020). On the usefulness of parental lines GWAS for predicting low heritability traits in tropical maize hybrids. *PLoS one*, 15(2), e0228724.
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. and Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS one*, 9(2), e90346.
- Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, 9(1), 1-9.
- Sandhu, S. K., Pal, L., Kaur, J. and Bhatia, D. (2019). Genome wide association studies for yield and its component traits under terminal heat stress in Indian mustard (*Brassica juncea* L.). *Euphytica*, 215(11), 1-15.
- Shikha, K., Shahi, J. P., Vinayan, M. T., Zaidi, P. H., Singh, A. K. and Sinha, B. (2021). Genome-wide association mapping in maize: status and prospects. *3 Biotech*, 11(5), 1-10.

- Singh, B. D. and Singh, A. K. (2015). Association mapping. In *Marker-Assisted Plant Breeding: Principles and Practices* (pp. 217-256). Springer, New Delhi.
- Singh, J., Sharma, P.C., Sharma, S.K. and Rai, M. (2014). Assessing effect of salinity on oil quality parameters of Indian mustard [*Brassica juncea* (L.) Czern & Coss.] using Fourier Transform Near-Infrared Reflectance (FT-NIR) spectroscopy. *Grasas y Aceites*. 65: e009.