



Haplotype-Based GWAS for Multi-Trait Seed-Borne Disease Resistance: A New Paradigm for Crop Breeding

*Jyoti, Harish Kumar and Jatin Singh

School of Agricultural Sciences, IIMT University, Meerut, India

*Corresponding Author's email: jyotisinghsst05@gmail.com

Seed-borne diseases continue to be a serious problem affecting crop productivity, quality and trade. Infected seed lots may carry pathogens within or on them and may carry the disease from one generation to the next and may cause infection before the crop is established. The development of resistant cultivars is therefore an important complement to seed testing, certification and disease-management. While these methods of conventional resistance breeding and single-nucleotide polymorphism (SNP)-based genome-wide association studies (GWAS) have produced significant progress, they have missed aspects of the genetic complexity of disease resistance. Many genes with small effects tend to regulate resistance against a pathogen and plant defense pathways may have multiple effects on several different pathogens, or multiple effects on the various stages of a disease. Haplotype-based GWAS is one approach to these limitations, as it focuses on linked sets of genetic variants, rather than on single SNPs. It can be used in combination with multi-trait statistical models, which allow the identification of allelic combinations and genomic regions associated with coordinated defence responses. This article discusses the principles, applications, limitations and prospects of incorporating multi-trait analysis in haplotype-based GWAS. This strategy could be used to improve the resistance of cultivars to seed-borne pathogens to be more widespread and persistent.

Seed-Borne Diseases Create a Distinct Breeding Challenge

Pathogens, which are fungi, bacteria, viruses and nematodes, are present in the seed. Seed-borne infections occur with the seed, and unlike many foliar diseases, are not evident until the crop is established and can be managed during the growing season. The pathogens can be present on seed surface, inside the seed tissue or inside the seed. Infected seed may be the first source of inoculum in a field after planting. This is a transmission mode that poses multiple risks. Disease can decrease the likelihood of germination, stunt early plant growth and decrease plant stand before other control measures are implemented. Pathogens can be transmitted by infected seed over generations and distances. Consequently, seed health is not just a yield concern, but also an international agriculture trade and safety concern of commercial seed systems. Seed certification and phytosanitary programmes have been tightened. Testing and treatment alone are not enough to eradicate the need for genetic resistance, however. Resistant cultivars can help prevent the build up of pathogens and may help to limit the likelihood of infection being carried with successive seed generations.

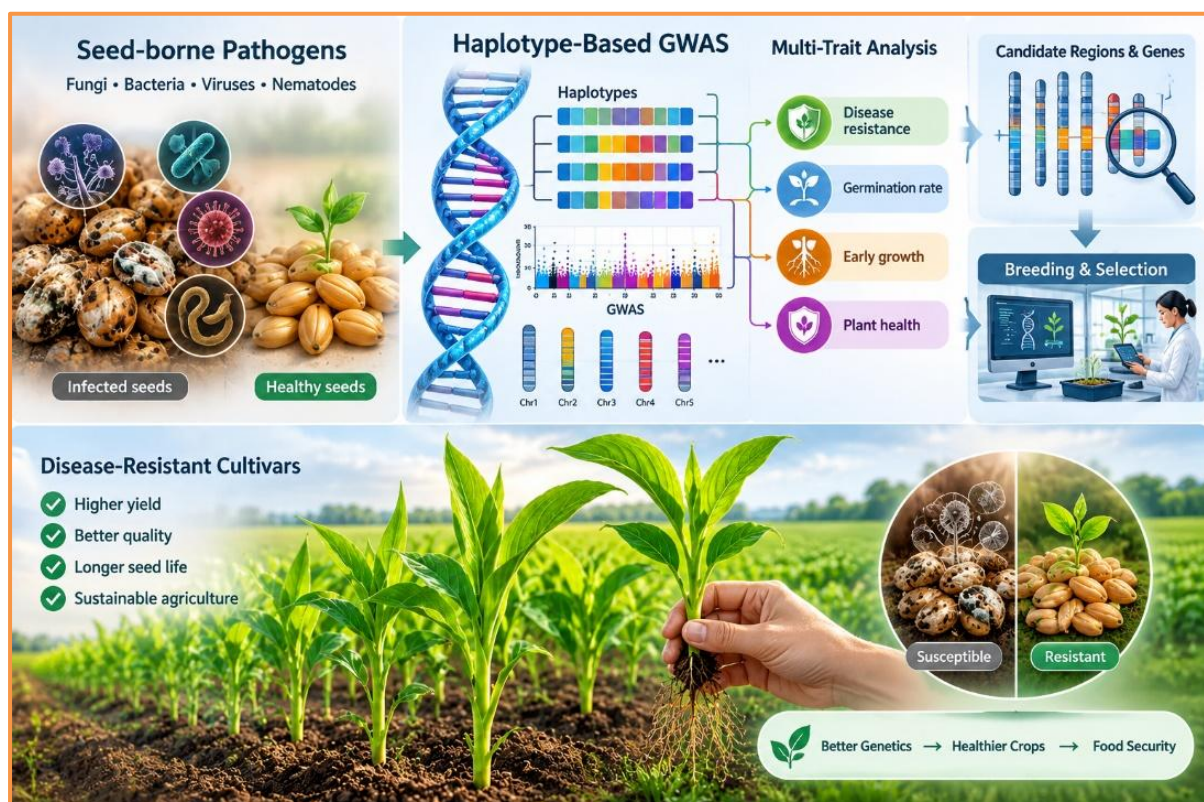
Why Single-Trait and Single-SNP Approaches Are Not Enough

In traditional breeding, resistance to a single pathogen or disease-related phenotype is typically tested along with resistance to other ones. This approach can be useful if the resistance gene is robust and effective. It is less effective if quantitative resistance involves many genes or stages of infection or environmental factors. Plant defense is also linked with other plant defense. The same signaling elements can be involved in the reactions to multiple pathogens. One gene that may confer resistance to one disease can have a positive or negative

effect on resistance to another disease. These relationships cannot be completely captured by a one-trait analysis. GWAS is now a primary method of identifying regions of the genome that are linked to complex traits. A conventional GWAS investigates the association between individual SNPs and a phenotype. This is a very effective method but each SNP captures only a small proportion of the variation at a locus. This can therefore miss rare alleles, multi-variant effects, and interactions between linked variants. There are also numerous statistical tests, making it hard to identify loci that have small effects even when correcting for multiple tests. The restrictions apply particularly to resistance that is seed borne, which is usually polygenetic and has significant environmental variation.

Individual SNPs to Haplotypes

A haplotype is a set of alleles at close-by genetic markers that are inherited together. Nearby variants tend to be linked together because of linkage disequilibrium and so a haplotype is a more complete way of describing a genomic region than any single SNP. In haplotype-based GWAS the unit of analysis is changed from the allele to the haplotype. Whereas one marker is associated with resistance, a combination of linked markers are asked if they are associated with a trait. This shift can have a number of benefits. Firstly, haplotypes can include alleles that are not discernible in a biallelic SNP test. There can be several different haplotype variants within a genomic region, with different evolutionary histories and biological effects. These multi-allelic combinations can be tested and associations that are not obvious when the same region is represented by different SNPs can be identified. Second, haplotypes may carry multiple effects of a set of linked variants. If individual variants contribute to the results with small effects, then the analysis of the set of variants can enhance the power to detect the effect of the larger genomic region. Third, haplotypes carry information about historical recombinations, selections and ancestry. The presence of a resistance-associated haplotype could be a conserved defense attitude that remained due to its being an evolutionary benefit. This information can guide researchers to prioritize genomic regions for validation and breeding. Haplotype analysis is not a panacea for all statistical issues. The value of it lies in the accuracy of genotyping, block definition, sample size and control of population structure. However, it offers a biologically relevant approach to the treatment of each SNP as a separate signal.



The Case for Multi-Trait GWAS

There are no single biological processes associated with seed-borne disease resistance. This can be multiple measurable attributes such as pathogen incidence, pathogen load, germination of seeds, vigor of seedlings, disease severity and successful transmission to the next generation. More than one pathogen can also be present in a single seed lot. These phenotypes are modeled simultaneously in a multi-trait GWAS. The main concept is that there is a shared genetic architecture between genetically correlated traits. When the same region of the genome is important for resistance to several pathogens or is related to several steps of infection, information obtained from one trait can help the analysis of another. This information sharing can be valuable when individual traits are only moderately heritable, or when the contribution of a locus is too small to be significant in an analysis of a single trait. Multi-trait analysis can also be used to identify multi-trait loci, or regions throughout the genome that affect more than one trait. Pleiotropic loci could be used as good breeding targets. One haplotype may provide resistance to multiple diseases without compromising agronomic characteristics and thus contribute to broad-spectrum resistance through a single selection. Multi-trait analysis can also be used to detect genetic correlations that are not favorable. A gene that provides resistance to one pathogen may enhance susceptibility to another, for instance. Early identification of such trade-offs helps breeders to avoid pairing undesirable alleles together.

Why the Integrated Approach Fits Seed-Borne Disease Biology

The haplotype-based GWAS and multi-trait analysis are biologically similar to seed-borne disease resistance in three ways.

It captures polygenic resistance- Quantitative resistance is frequently polygenic with each locus contributing a small percentage of the phenotype. Haplotype analysis can combine the effects of linked variants, and deliver a more powerful signal for the region of the genome.

It reflects interconnected defense pathways- An allele that confers resistance is best when it offers resistance to a range of pathogens, environments and stages of crop development. Integrated analysis can separate broad protection from narrow or contradictory effects haplotypes. This important distinction is needed for creating long-lasting cultivars instead of varieties that are resistant to one pathogen but susceptible to another.

It distinguishes broad resistance from trade-offs- An allele that confers resistance is best when it offers resistance to a range of pathogens, environments and stages of crop development. Integrated analysis can separate broad protection from narrow or contradictory effects haplotypes. This important distinction is needed for creating long-lasting cultivars instead of varieties that are resistant to one pathogen but susceptible to another.

Implementing Haplotype-Based Multi-Trait GWAS

Haplotype-Based Multi-Trait GWAS. Fixed physical windows, linkage-disequilibrium windows, or identity-by-descent windows can be used in research. Fixed windows are easy to perform but are not necessarily the windows of biological interest. Linkage-disequilibrium methods segment the genome based on how the various variants are inherited together at a given locus in the population. Methods of identity by descent will place greater emphasis on shared ancestry and can be helpful when multiple populations are analyzed. The associations obtained may vary, depending on the method selected. The stability of important results over reasonable block definitions should therefore be investigated rather than relying on one algorithm or one setting of the parameters. Once the haplotypes are built, they can be assessed for several different phenotypes by using multivariate mixed models, multi-trait GWAS approach, etc. These models must take into consideration population structure, kinship, experimental design and environmental effects. If these controls are not in place, genetic associations may be confused with differences in ancestry or conditions of the trials. Genetic-network and multi-omics analyses can be applied to further extend the analysis. Candidate haplotypes can be compared to gene-expression, metabolite, protein or co-expression networks during the infection. This integration may be useful to relate statistical

signals to biological mechanisms and to select the most appropriate genes for functional validation.

Translating Associations into Breeding Tools

The first is that the construction of haplotypes is dependent on the method being used. There can be variations in the results when using different block sizes, linkage-disequilibrium thresholds, and minor-allele-frequency filters. Such choices should be made transparent and the researchers should run sensitivity analyses. The second is statistical power. Haplotype methods can enhance the identification of regional effects, but are not a substitute for a sufficient sample size or proper phenotyping. It is when traits are genetically correlated that multi-trait models are most useful. When traits have little or no common genetic basis, it may be hard to benefit from information sharing or information sharing may be difficult to interpret. Population structure is the third. Geographic or breeding subpopulations may vary in their haplotype frequencies. A lack of modelling of structure may create erroneous correlations. Such a problem may be more difficult to handle using multi-allelic haplotypes rather than individual SNPs. The fourth being a gap between association and causation. There can be multiple variants within a significant haplotype and the observed effect could be due to one variant, multiple interacting variants, or a variant within a different "linked" region outside the tested block. It is therefore essential to conduct functional experiments to identify mechanism.

Limitations That Require Careful Attention

There are some restrictions that need careful attention before the immediate application of this framework. The first is that the construction of the haplotype is method-dependent. Various block sizes, minor-allele-frequency filters and linkage-disequilibrium thresholds may lead to varying results. Such decisions should be reported openly by the researchers and run sensitivity analyses. The second is statistical power. Haplotype methods can enhance the ability to detect regional effects but cannot substitute for proper sample size or good phenotyping. Multi-trait models are best used when there are genetic correlations among traits. However, if there is little or no genetic overlap, then information sharing would be limited benefit, or could be difficult to interpret. The third is population structure. Geographic or breeding subpopulations may have different frequencies of haplotypes. Excessive correlations can result from inappropriate structural modeling. Multi-allelic haplotypes may be more challenging to manage than individual SNPs. The fourth is the difference between association and causation. There can be multiple variants within a significant haplotype and multiple interacting mutations can cause the observed effect, or a variant in a different region of the block may be linked to the effect. Therefore, functional experiments are required to establish mechanism.

Future Directions

With the advent of pangenome resources, the horizon of haplotype-based analysis will be likely to broadened. Structural variants and presence-absence variants and other sources of diversity that are important for disease resistance may not be reflected in a single reference genome. Pangenomes can offer a more comprehensive portrait of these variants and enable the analyses of resistance-associated genomic configurations in a variety of germplasm. Functional genomics will also increase the ability to interpret. Measurements of transcriptomes and metabolomes throughout infection time series can help to identify the effect of candidate haplotypes on defense activation. The integration of these data, expression quantitative trait loci, co-expression networks and gene-regulatory analysis could assist the prioritization of candidate genes for experimental testing. The evolutionary history of haplotypes may be included more explicitly in statistical models in the future. Information on ancestry and selection, for instance, could be used to better estimate the effects with Bayesian approaches and to separate conserved resistance haplotypes from population-specific signals. Careful research on seed-borne pathogens is of special importance. Seed transmission

genetics may be different than foliar or soil borne disease resistance. There is a need to specifically investigate interactions of the seed microbiome, spread of pathogens into reproductive tissues, vertical transmission, and seed infection to seedling disease.

Conclusion

Haplotype based GWAS integrated with the multi-trait approach. It represents sets of coupled variants, maintains a record of evolutionary history and mimics genetic correlations between defence measures across pathogens and infection stages. The approach is not an alternative to good experimental design, size and/or useful validation. It's best used when combined with other sources of information. Multi-omics data in combination with pangenome resources and comprehensive field testing can be used in a multi-traits analysis approach to discover a broad, stable, and practical resistance mechanism with the support of the haplotypes. The final goal is to use these genomic discoveries to produce better seed, better establishment, less pathogen spread and enhanced protection of agriculture in the face of evolving diseases and climate scenarios.

References

1. Munkvold, G., du Toit, L., & Dunkle, R. (2025). Seed pathology: Challenges and advances in ensuring a safe global seed supply. *Annual Review of Phytopathology*, 63, 43–62. <https://doi.org/10.1146/annurev-phyto-121423-093855>
2. Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, 9, Article 29. <https://doi.org/10.1186/1746-4811-9-29>
3. Bhat, J. A., Yu, D., Bohra, A., Ganie, S. A., & Varshney, R. K. (2021). Features and applications of haplotypes in crop breeding. *Communications Biology*, 4, Article 1266. <https://doi.org/10.1038/s42003-021-02782-y>
4. Lee, C. H., Shi, H., Pasaniuc, B., Eskin, E., & Han, B. (2021). PLEIO: A method to map and interpret pleiotropic loci with GWAS summary statistics. *The American Journal of Human Genetics*, 108(1), 36–48. <https://doi.org/10.1016/j.ajhg.2020.11.017>
5. Cowger, C., & Brown, J. K. M. (2019). Durability of quantitative resistance in crops: Greater than we know? *Annual Review of Phytopathology*, 57, 253–277. <https://doi.org/10.1146/annurev-phyto-082718-100016>
6. Alemu, A., Batista, L., Singh, P. K., Ceplitis, A., & Chawade, A. (2023). Haplotype-tagged SNPs improve genomic prediction accuracy for *Fusarium* head blight resistance and yield-related traits in wheat. *Theoretical and Applied Genetics*, 136, Article 92. <https://doi.org/10.1007/s00122-023-04352-8>
7. Jayakodi, M., Lu, Q., Pidon, H., Rabanus-Wallace, M. T., Bayer, M., Lux, T., Guo, Y., Jaegle, B., Badea, A., Bekele, W., Brar, G. S., Braune, K., Bunk, B., Chalmers, K. J., Chapman, B., Jørgensen, M. E., Feng, J.-W., Feser, M., Fiebig, A., et al. (2024). Structural variation in the pangenome of wild and domesticated barley. *Nature*, 636(8043), 654–662. <https://doi.org/10.1038/s41586-024-08187-1>
8. Yang, Y., Saand, M. A. S., Huang, L., Abdelaal, W. B., Zhang, J., Wu, Y., Li, J., Sirohi, M. H., & Wang, F. (2021). Applications of multi-omics technologies for crop improvement. *Frontiers in Plant Science*, 12, Article 563953. <https://doi.org/10.3389/fpls.2021.563953>