# Utilization of Mapping Populations in Crop Improvement

Abhishek[1] , Satish D[2] and [3]Sandhyarani Nishani

[1]PhD Scholar, Genetics and Plant Breeding

[2]Professor, Genetics and Plant Breeding

[3]Professor Plant Biotechnology

University of Horticultural Sciences, Bagalkot

[*]Corresponding Author mail: abhisheks72048@gmail.com

## Introduction

Every new crop variety begins with a simple question: which genes make a plant yield more, resist disease, tolerate drought or taste better? For decades breeders answered this by growing thousands of plants and selecting the best ones by eye. Today they can often "read" the answer from the DNA itself. The tool that makes this possible is the **mapping population**, a carefully designed set of plants that links what we see in the field (the phenotype) with what lies in the genome (the genotype).

## What is a mapping population?

A mapping population is a group of individuals derived from a cross between two parents that differ clearly in the trait of interest, for example a resistant parent and a susceptible parent. Because the offspring inherit different combinations of parental DNA segments, breeders can ask a powerful question: *do plants that carry a particular DNA marker also show the trait?* If the answer is yes consistently, the marker lies close to the gene controlling that trait on the chromosome. This is the basic principle of genetic linkage mapping (Tanksley, 1993).

### The genetic principle behind mapping

During meiosis, homologous chromosomes exchange segments through crossing over. Two loci that lie far apart on a chromosome are frequently separated by a crossover, whereas loci that lie close together tend to be inherited as a block. The proportion of recombinant gametes, called the recombination frequency, is therefore a measure of distance and is expressed in centiMorgans (cM); one cM corresponds roughly to a 1% recombination frequency. Because double crossovers hide some recombination, observed frequencies are converted into map distances with mapping functions such as those of Haldane and Kosambi.

Linkage between two loci is declared statistically using the LOD (logarithm of odds) score. A LOD of 3 or more, meaning odds of 1000 to 1 in favour of linkage, is the usual threshold. Breeders also test every marker for segregation distortion using a chi-square test against the expected ratio (1:2:1 for a codominant marker in an F2, 3:1 for a dominant marker), because deviations can reveal gametic or zygotic selection and can distort the map.

Three factors decide the success of any mapping study: the amount of polymorphism between the parents, the size of the population, and the accuracy of phenotypic data. A well-chosen population with poor phenotyping gives a poor map of the trait, however many markers are used.

## Types of mapping populations

The choice of population depends on the crop, the time and resources available, and the objective. Table 1 summarises the commonly used types.

**Table 1. Major types of mapping populations and their features**

| Population | How it is developed | Nature | Main advantages | Main limitations | Typical use |
|---|---|---|---|---|---|
| F2 | Selfing or intercrossing F1 | Temporary, heterozygous | Quick; one season after F1; all genetic classes present | Cannot be replicated or preserved; limited trait precision | Rapid gene mapping, first-level linkage map |
| Backcross (BC1) | F1 crossed to a parent | Temporary | Simple genetics; good for dominant traits | Only one parent's recombination informative | Introgression, mapping dominant genes |
| Doubled haploid (DH) | Chromosome doubling of haploids from F1 gametes | Permanent, homozygous | Fully homozygous; replicable in multiple environments | Needs efficient tissue culture protocol | Cereals (barley, wheat, rice), Brassica |
| Recombinant inbred lines (RILs) | Single-seed descent from F2 for 6 to 8 generations | Permanent, homozygous | Replicated trials; accumulates recombination; reusable | Time consuming (several years) | QTL mapping for complex traits |
| Near isogenic lines (NILs) | Repeated backcrossing with marker selection | Permanent | Isolates a single QTL in a uniform background | Laborious to produce | Fine mapping, QTL validation |
| Bulked segregant analysis (BSA) | Pooling DNA of extreme phenotypes from any segregating population | Pooled samples | Fast, low cost; needs few markers | Detects only major genes | Quick tagging of major genes |
| MAGIC / NAM | Intercrossing several (4 to 16) or 25 or more founders | Permanent, multi-parent | High resolution; broad genetic diversity | Large and costly to build | Complex traits; modern genomics-assisted breeding |
| Association panel (GWAS) | Diverse germplasm or landraces | Natural population | No crossing needed; uses historical | Needs dense markers; false associations from | Gene discovery in diversity |

| | | | | | |
|---|---|---|---|---|---|
| | | | recombination | population structure | collections |

### $F_2$ and backcross populations

The $F_2$ is the quickest population to produce. Selfing a single $F_1$ plant gives a generation that segregates in a 1:2:1 genotypic ratio, so both additive and dominance effects can be estimated, and codominant markers can distinguish heterozygotes. Its main weakness is that every plant is genetically unique and cannot be reproduced; seed is limited, and the trait can be measured only once under one set of conditions. A backcross (BC1) population, obtained by crossing the $F_1$ to one of the parents, is simpler genetically and is useful for mapping dominant genes or for transferring a gene into an elite parent.

### Recombinant inbred lines (RILs)

RILs are produced by taking $F_2$ plants and advancing each one separately by single-seed descent until it is almost homozygous, normally at the $F_6$ to $F_8$ generation. Each line is a fixed mosaic of the two parental genomes. Since the lines are homozygous, seed can be multiplied and the same population can be tested in many seasons and locations, which is ideal for traits with strong genotype by environment interaction. Several generations of selfing also give extra opportunities for recombination, so the map is expanded to roughly twice the $F_2$ distance, which helps to separate closely linked markers. The population can be shared among laboratories, so the data of many researchers can be added to a single map, as with the maize recombinant inbreds used by Burr et al. (1988). The disadvantage is the time needed, which is several seasons or years.

### Doubled haploids (DH)

Doubled haploid lines are made by culturing anthers or microspores from $F_1$ plants, or by wide crossing followed by chromosome elimination, and then doubling the chromosome number with colchicine. Complete homozygosity is reached in a single generation, so a permanent population is obtained within one or two seasons. As they pass through only one round of meiosis, map distances are the same as in an $F_2$. DH populations are common in barley, wheat, rice, rapeseed and maize, but they depend on an efficient haploid induction protocol, which is not available for every species or genotype.

### Near isogenic lines (NILs) and chromosome segment substitution lines

NILs are developed by repeated backcrossing, with selection for the target region, so that the line differs from the recurrent parent only for a small chromosome segment. Any difference in phenotype between the NIL and its recurrent parent can be traced to that segment, which makes NILs the standard tool for confirming a QTL and for fine mapping. Libraries of introgression lines or chromosome segment substitution lines (CSSLs) cover the whole donor genome in overlapping segments; the tomato introgression lines of Solanum pennellii developed by Eshed and Zamir (1995) are a well-known example, and they exposed yield-associated QTLs from a wild relative.

### Bulked segregant analysis (BSA)

BSA is not a population type but a sampling strategy applied to an existing segregating population. DNA from 10 to 20 plants with the extreme phenotypes, for instance highly resistant and highly susceptible, is pooled into two bulks. Markers that differ between the bulks lie near the gene of interest, since the bulks are expected to be random for all other regions of the genome (Michelmore et al., 1991). The method is quick and inexpensive and is mainly suited to traits controlled by one or a few major genes.

### Multi-parent populations: MAGIC and NAM

Two-parent populations sample only two alleles per gene and give wide QTL intervals. Multi-parent advanced generation inter-cross (MAGIC) populations intercross several founders, often four to sixteen, for a number of generations before inbreeding. The extra recombination and allelic diversity increase mapping resolution (Cavanagh et al., 2008). The nested association mapping (NAM) population in maize was designed by crossing 25 diverse founder lines to a common parent, B73, and deriving about 200 RILs from each cross, which gave about 5,000 lines in total. It combines the power of linkage mapping with the resolution of association mapping (Yu et al., 2008; McMullen et al., 2009).

### Natural populations for association mapping

Germplasm collections, landraces and breeding lines carry the recombination accumulated over thousands of years of evolution. Genome-wide association studies (GWAS) use this historical recombination to link markers to traits without making a cross (Mackay and Powell, 2007; Huang et al., 2010). The resolution can be at the level of a single gene, but population structure and relatedness must be corrected for statistically, or false associations will be reported.

## How a mapping population is used: step by step

Figure 1 shows the general workflow, from choosing parents to using markers in a breeding programme.

mapping population workflow · 8 steps

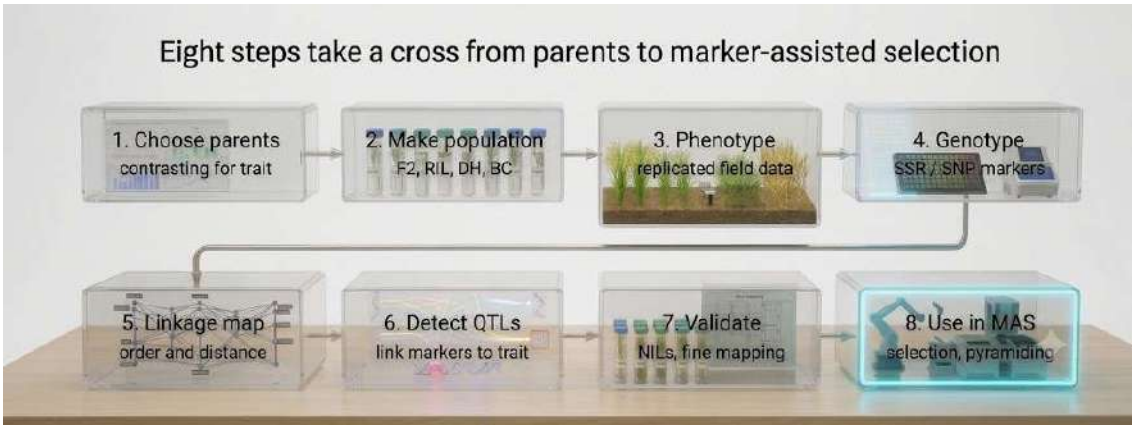



**Figure 1.** General workflow for developing and using a mapping population in crop improvement.

1. **Select contrasting parents** for the target trait.
2. **Develop the population** (F2, RIL, DH or other type).
3. **Phenotype** the population carefully, ideally across seasons and locations, since quantitative traits are strongly influenced by environment.
4. **Genotype** the same plants using markers such as SSRs, SNPs or DArT arrays (Semagn et al., 2006).
5. **Build a linkage map** by calculating how often markers are inherited together.
6. **Detect QTLs** (quantitative trait loci) by statistically linking marker data with trait data, for example using interval mapping (Lander and Botstein, 1989).
7. **Validate and fine map** the QTL in NILs or new populations.
8. **Apply** the linked markers in marker-assisted selection (MAS) and gene pyramiding.

## Phenotyping, markers and QTL detection in detail

### Phenotyping

The quality of a QTL study is limited by the quality of the phenotype. Experiments should use replications and a suitable design such as an alpha-lattice or augmented design, be repeated across at least two seasons or locations, and measure the trait by a precise and repeatable method. For horticultural crops this may mean fruit or pod traits recorded on several plants, biochemical assays for quality traits such as vitamin C, carotenoids or total phenolics, and standard disease scoring scales for resistance. Where the environment influences the trait strongly, the best linear unbiased predictor (BLUP) or mean over environments is analysed rather than raw data from a single season.

### Molecular markers

Markers are landmarks along the chromosome that are not influenced by environment. The marker system selected affects the cost, the speed and the information obtained (Semagn et al., 2006). Table 2 compares the main types.

**Table 2. Commonly used molecular markers in mapping studies**

| Marker | Inheritance | Throughput | Main advantages | Main limitations |
|---|---|---|---|---|
| RFLP | Codominant | Low | Highly reproducible; transferable between related species | Needs large amounts of DNA; laborious and slow |
| RAPD | Dominant | Medium | No sequence information needed; cheap | Poor reproducibility; cannot separate heterozygotes |
| AFLP | Dominant | High | Many loci per reaction | Technically demanding; dominant scoring |
| SSR (microsatellite) | Codominant | Medium | Multi-allelic, locus specific, robust and easy to use | Development needs sequence information; gel-based scoring is slow |
| SNP | Codominant | Very high | Abundant across the genome; automated; suitable for sequencing and array platforms | Biallelic; needs equipment or sequencing service |

Today, SNPs generated by genotyping-by-sequencing or SNP arrays are the preferred marker for new mapping projects, while SSRs remain popular where laboratories have limited resources.

### Constructing the linkage map

After genotyping, markers are grouped into linkage groups using a LOD threshold, ordered within each group and assigned distances in cM. Software such as MapMaker, JoinMap, R/qtl (Broman et al., 2003) and IciMapping is commonly used. The number of linkage groups should match the haploid chromosome number of the species; if more groups are obtained, the map still has gaps that need additional markers.

### Detecting QTLs

Several statistical approaches are used, each with its own strengths:

1. **Single-marker analysis** (t-test, ANOVA or regression) compares the trait means of the marker classes. It is simple and needs no map, but it cannot separate QTL position from effect.
2. **Interval mapping** estimates the likelihood of a QTL at every position between two flanking markers, and is more precise (Lander and Botstein, 1989).
3. **Composite interval mapping** adds markers elsewhere in the genome as cofactors to remove the influence of other QTLs, giving better resolution and fewer false positives (Zeng, 1994).
4. **Multiple-environment and multi-QTL models** test QTL by environment interaction and epistasis, which helps decide whether a QTL is stable or specific to a location.

Because thousands of tests are made along the genome, the significance threshold should be set by a permutation test rather than by an arbitrary LOD value (Churchill and Doerge, 1994). Each QTL is reported with its position, LOD score, the percentage of phenotypic variance explained (PVE or R2), the additive effect, and the allele-contributing parent. QTLs that explain more than about 10% of the variance and are found in several environments are called major or stable QTLs, and these are the best candidates for marker-assisted selection. A QTL identified in only one environment, or with a very small effect, should be validated before being used in breeding.

## Uses in crop improvement

**1. Constructing genetic linkage maps.** Maps built from RILs, DH and F2 populations provide the framework for locating genes along chromosomes. Early maps in maize used recombinant inbreds to place genes precisely (Burr et al., 1988).

**2. Discovering genes and QTLs for important traits.** Yield, quality, drought tolerance and disease resistance are usually polygenic. Mapping populations break this complexity down into individual QTLs, each with a measurable effect (Tanksley, 1993; Collard et al., 2005).

**3. Rapid tagging of major genes.** Bulked segregant analysis lets breeders find markers linked to a resistance gene by testing just two DNA pools, saving time and cost (Michelmore et al., 1991).

**4. Marker-assisted selection.** Once a marker is tightly linked to a gene, seedlings can be screened in the laboratory long before the trait appears in the field. This is especially valuable for traits expressed late, such as fruit quality, or traits that are difficult to score, such as root architecture (Collard and Mackill, 2008).

**5. Gene pyramiding.** Several resistance genes or QTLs can be combined in one variety using markers, which makes resistance more durable against changing pathogens.

**6. Fine mapping and gene cloning.** NILs and large segregating populations help narrow a QTL to a small chromosome region, leading to map-based cloning of the underlying gene.

**7. Multi-parent and association approaches.** MAGIC and NAM populations combine the benefits of linkage mapping with greater diversity and mapping resolution (Cavanagh et al., 2008; Yu et al., 2008). Association panels extend this to natural diversity collections without making any crosses (Mackay and Powell, 2007; Huang et al., 2010).

## From QTL to variety: marker-assisted breeding schemes

Mapping is of value to the breeder only when its results are converted into a better variety. Three main schemes are used.

**Marker-assisted backcrossing (MABC).** The target gene or QTL from a donor parent is moved into an elite variety through repeated backcrossing. In foreground selection, markers closely linked to the target locus confirm that every selected plant carries it. In background selection, markers spread across the genome identify the plants that have the largest proportion of the recurrent parent genome. Without selection, the recurrent parent contribution rises only from 75% at BC1 to 87.5% at BC2 and 93.75% at BC3, but with background selection a recovery of more than 95% can be reached by BC2 or BC3, saving several generations. Flanking markers on both sides of the target region also reduce linkage drag, the unwanted donor segment that comes with the target gene.

**Gene pyramiding.** Genes for resistance to the same disease are difficult to combine by phenotypic selection, because one gene masks the presence of the others in disease tests. Markers allow several genes or QTLs to be stacked in one genotype, producing broad-spectrum and more durable resistance (Collard and Mackill, 2008).

**Marker-assisted recurrent selection and genomic selection.** For traits controlled by many small-effect QTLs, selecting only a few significant markers captures little of the variation. Genomic selection instead uses a dense marker set to predict the breeding value of each individual from a model trained on a population that has been both phenotyped and genotyped (Meuwissen et al., 2001). Mapping populations are often used as the training population in the first cycles of such programmes.

## A success story:

### *Sub1A* in rice

Flash floods can destroy a rice crop within two weeks. A major QTL for submergence tolerance, mapped using a mapping population derived from a tolerant landrace, was eventually cloned as the *Sub1A* gene (Xu et al., 2006). Using linked markers, breeders transferred this gene into popular high-yielding varieties by marker-assisted backcrossing, producing "Swarna-Sub1" and similar varieties now grown by millions of farmers in flood-prone South Asia.

### Tomato fruit size and wild alleles

The domesticated tomato produces fruits that are many times heavier than those of its wild ancestors. A QTL on chromosome 2, named fw2.2, explains a large share of this difference. It was located by mapping in a population derived from a cross between cultivated and wild tomato and later cloned, showing that a major quantitative trait can be caused by a single gene whose expression level is changed (Frary et al., 2000). The introgression lines mentioned earlier have also been used to discover wild alleles that raise yield and soluble solids in elite tomato backgrounds (Eshed and Zamir, 1995). These examples show that mapping populations can bring back useful variation from wild relatives that were lost during domestication.

## Special considerations for horticultural and perennial crops

In fruit, vegetable and tree crops, long juvenile periods, high heterozygosity and self-incompatibility make RILs and DH lines difficult or impossible to produce. Breeders therefore rely on F1 pseudo-testcross populations, F2 and backcross populations where feasible, bulked segregant analysis and association panels built from germplasm collections. Dense SNP markers from low-cost sequencing now make even small populations of such crops useful for mapping.

## Challenges and the way forward

Mapping populations are not a magic solution. QTL effects often vary with the genetic background and environment, small populations overestimate effects, and accurate phenotyping remains the most

common bottleneck. The future lies in combining mapping populations with high-throughput phenotyping, genotyping-by-sequencing, genomic selection and genome editing. Together these will shorten breeding cycles and deliver climate-resilient, nutritious varieties more quickly.

### Limitations in detail

- **Population size.** Small populations of fewer than about 100 individuals tend to detect only the largest QTLs and overestimate their effects, while missing many smaller ones. Larger populations, and replicated data, give more reliable estimates, and fine mapping typically needs thousands of recombinant individuals.
- **Wide QTL intervals.** In a two-parent population, a QTL interval often covers 10 to 20 cM and hundreds of genes. Fine mapping, multi-parent populations or association mapping is needed to narrow it.
- **Background and environment effects.** A QTL found in one cross may not appear in another because the other parent carries the same allele or because epistatic partners are missing. QTL by environment interaction can also make a QTL useful in one region only.
- **Linkage drag and cost.** Markers close to a gene are not the gene itself, and donor segments transferred with it can reduce yield or quality. Marker development, genotyping and phenotyping are expensive, so mapping is best focused on traits that are important, difficult to select phenotypically and reasonably heritable.
- **Limited transfer to varieties.** Many QTLs are reported in publications, but only a small fraction are validated and used in breeding programmes. Validation in elite backgrounds should be planned from the beginning.

### Key points

- A mapping population is built from parents that differ for the trait, and its value comes from the link between marker data and trait data.
- Permanent populations (RIL, DH, NIL) allow replicated testing; temporary ones (F2, BC) are faster; multi-parent and association populations give higher resolution.
- Accurate multi-environment phenotyping is as important as dense markers.
- Major, stable QTLs are the best candidates for marker-assisted backcrossing and pyramiding; polygenic traits are better handled by genomic selection.
- Perennial and heterozygous horticultural crops need modified designs such as F1 pseudo-testcross populations and association panels.

## Conclusion

Mapping populations are the bridge between the genome and the field. By choosing the right population and combining it with careful phenotyping and modern markers, breeders can locate valuable genes, select them efficiently and pyramid them into improved varieties. For a country such as India, with diverse crops and a wide range of stresses, this approach is central to sustaining productivity and farmer income.

## References

Broman, K. W., Wu, H., Sen, S. and Churchill, G. A. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19: 889-890.

Burr, B., Burr, F. A., Thompson, K. H., Albertson, M. C. and Stuber, C. W. 1988. Gene mapping with recombinant inbreds in maize. *Genetics*, 118: 519-526.

Cavanagh, C., Morell, M., Mackay, I. and Powell, W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*, 11: 215-221.

Churchill, G. A. and Doerge, R. W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics*, 138: 963-971.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. and Pang, E. C. K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, 142: 169-196.

Collard, B. C. Y. and Mackill, D. J. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B*, 363: 557-572.

Eshed, Y. and Zamir, D. 1995. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, 141: 1147-1162.

Frary, A., Nesbitt, T. C., Frary, A., et al. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289: 85-88.

Huang, X., Wei, X., Sang, T., et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42: 961-967.

Lander, E. S. and Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121: 185-199.

Mackay, I. and Powell, W. 2007. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, 12: 57-63.

McMullen, M. D., Kresovich, S., Villeda, H. S., et al. 2009. Genetic properties of the maize nested association mapping population. *Science*, 325: 737-740.

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819-1829.

Michelmore, R. W., Paran, I. and Kesseli, R. V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis. *Proceedings of the National Academy of Sciences USA*, 88: 9828-9832.

Semagn, K., Bjornstad, A. and Ndjiondjop, M. N. 2006. An overview of molecular marker methods for plants. *African Journal of Biotechnology*, 5: 2540-2548.

Tanksley, S. D. 1993. Mapping polygenes. *Annual Review of Genetics*, 27: 205-233.

Xu, K., Xu, X., Fukao, T., et al. 2006. Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature*, 442: 705-708.

Yu, J., Holland, J. B., McMullen, M. D. and Buckler, E. S. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178: 539-551.

Zeng, Z. B. 1994. Precision mapping of quantitative trait loci. *Genetics*, 136: 1457-1468.