Mapping populations for self-fertilizing crop plants

Sandeep Kumar Bangarwa, Ravi Kumawat, Heera Lal Barupal, Mukesh Kumar Yadav and Amardeep Kour

DOI: https://doi.org/10.22271/chemi.2020.v8.i4.ax.10261

Abstract

In genetics and breeding, mapping populations are the tools used to identify the genetic loci controlling measurable phenotypic traits. For self-pollinating species, F2 populations and recombinant inbred lines (RILs) are used; for self-incompatible, highly heterozygous species, F1 populations are mostly the tools of choice. Backcross populations and doubled haploid lines are a possibility for both types of plants. The inheritance of specific regions of DNA is followed by molecular markers that detect DNA sequence polymorphisms. For map-based cloning of a gene, populations of a large size provide the resolution required. The type of mapping population to be used depends on the reproductive mode of the plant to be analyzed. In this respect, the plants fall into the main classes of self-fertilizers and self-incompatibles. In this review, we describe the mapping population for self-fertilizing crops, types and problems in genetic mapping in detail.

Keywords: Homozygosity, mapping population, population, self-fertilizing crops

1. Introduction

In 1912 Vilmorin and Bateson described the first works on linkages in Pisum. However, the concept of linkage groups representing chromosomes was not clear in Pisum until 1948, when Lampricht described the first genetic map with 37 markers distributed on 7 linkage groups (summarized in Swiecicki et al., 2000) [35]. Large collections of visible markers are today available for several crop species and for Arabidopsis thaliana (Koornneef et al., 1987; Neuffer et al., 1997) [17, 26]. The advance of molecular biology provided a broad spectrum of technologies to assess the genetic situation at the DNA level. The first DNA polymorphisms described were restriction fragment length polymorphism (RFLP) markers (Botstein et al., 1980) [4]. Recently, methods have been developed to detect single nucleotide polymorphisms (Rafalski, 2002) [30]. Because these methods have the potential for automatization and multiplexing, they allow the establishment of high-density genetic maps. Whereas RAPD and AFLP analyses are based on anonymous fragments, RFLP and SNP analyses allow the choice of expressed genes as markers. Genes of a known sequence and that putatively influence the trait of interest can be based cloning of a gene, populations of a large size provide the resolution required. In this way function maps can be constructed (Chen et al., 2001; Schneider et al., 2002) [12]. Phenotypic data of the segregating population, correlated to marker data, prove or disprove potential candidate genes supporting monoand polygenic traits. The basis for genetic mapping is recombination among polymorphic loci, which involves the reaction between homologous DNA sequences in the meiotic prophase. Currently, the double-strand-break repair model (Szostak et al., 1983) [36] is acknowledged to best explain meiotic reciprocal recombination. In this model, two sister chromatids break at the same point and their ends are resected at the 5' ends. In the next step the single strands invade the intact homologue and pair with their complements. The single-strand gaps are filled in using the intact strand as template. The resulting molecule forms two Holliday junctions. Upon resolution of the junction, 50% of gametes with recombinant lateral markers and 50% non-recombinants are produced. In the non-recombinants, genetic markers located within the region of strand exchange may undergo gene conversion, which can result in nonreciprocal recombination, a problem interfering in genetic mapping. In plants, gene conversion events were identified by Buschges et al. (1997) [6] when cloning the Mlo resistance gene from barley.
The likelihood that recombination events occur between two points of a chromosome depends in general on their physical distance: the nearer they are located to each other, the more they will tend to stay together after meiosis. With the increase of the distance between them, the probability for recombination increases and genetic linkage tends to disappear. This is why genetic linkage can be interpreted as a measure of physical distance. However, taking the genome as a whole, the frequency of recombination is not constant because it is influenced by chromosome structure. An example is the observation that recombination is suppressed in the vicinity of heterochromatin: here, the recombination events along the same chromatid appear to be reduced, an observation called positive interference. It reduces the number of double recombinants when, for example, three linked loci are considered. Linkage analysis based on recombination frequency and the order of linked loci is evaluated statistically using maximum likelihood equations (Fisher, 1921; Haldane and Smith, 1947; Morton, 1955) [11, 14, 25]. Large amounts of segregation data are routinely processed by computer programs to calculate a genetic map; among the most popular are JoinMap (Stam, 1993) [34] and MAPMAKER (Lander et al., 1987) [18].

Mapping Populations
The trait to be studied in a mapping population needs to be polymorphic between the parental lines. Additionally, a significant trait heritability is essential. It is always advisable to screen a panel of genotypes for their phenotype and to identify the extremes of the phenotypic distribution before choosing the parents of a mapping population. It is expected that the more the parental lines differ, the more genetic factors will be described for the trait in the segregating population and the easier their identification will be. This applies to monogenic as well as to polygenic traits. A second important feature to be considered when constructing a mapping population is the reproductive mode of the plant. There are two basic types. On the one hand are plants that self naturally, such as Arabidopsis thaliana, tomato, and soybean, or that can be manually selfed, such as sugar beet and maize; on the other hand are the self-incompatible, inbreeding-sensitive plants such as potato. Self-incompatible plants show high genetic heterozygosity, and for these species it is frequently not possible to produce pure lines due to inbreeding depression. Usually only self-compatible plants allow the generation of lines displaying a maximum degree of homozygosity. In conclusion, the available plant material determines the choice of a mapping population. Other factors are the time available for the construction of the population and the mapping resolution required.

Mapping Populations Suitable for Self-fertilizing Plants
If pure lines are available or can be generated with only a slight change of plant vigor, the mapping populations that can be used consist of F2 plants, recombinant inbred lines (RIL), backcross (BC) populations, introgression lines assembled in exotic libraries, and doubled haploid lines (DH).

Recombinant inbred lines (RIL)
The causative genetic loci underlying phenotypic traits can be mapped and studied using recombinant inbred lines (RILs) (Bailey, 1971). RILs are a collection of strains derived from a cross of genetically divergent parent strains. RILs are the products of successive inbreeding over several generations to develop true breeding lines. In self-pollinated crops, it is usually achieved by a SSD approach which involves continuous selfing of individuals from an initial F2 population until the desired level of homozygosity is achieved. F8 RILs are often used for genetic studies, where 99% homozygosity is expected (Seymour et al., 2012) [33]. The expected genetic segregation ratio of RILs and the overall frequency of alleles for both dominant and codominant marker are 1:1. RILs are often used to map traits that differ between the parental lines. As a product of many meiotic cycles, RILs are expected to have higher recombination rate, which is not only important for QTL mapping but also very useful to identify tightly linked markers. RILs are difficult to develop in crops that exhibit high levels of inbreeding depression (Madhusudhana, 2015) [22]. Distorted segregation can also be observed in RILs due to genetic factors resulting in gametic and zygotic selection (Liu et al., 2010) [20]. The major advantage is these lines constitute a permanent resource that can be replicated indefinitely and be shared by many groups in the research community. A second advantage of RILs is that because they undergo several rounds of meiosis before homozygosity is reached, the degree of recombination is higher compared to F2 populations. Consequently, RIL populations show a higher resolution than maps generated from F2 populations (Burr and Burr, 1991) [5], and the map positions of even tightly linked markers can be determined. In plants, RILs are available for many species, including rice and oat (Wang et al., 1994; O’Donoughue et al., 1995) [19, 27]. In Arabidopsis thaliana, 300 RILs have become a public mapping tool (Lister and Dean, 1993) [19]. Arabidopsis RILs were constructed by an initial cross between the ecotypes Landsberg erecta and Columbia, and a dense marker framework was established. Every genomic fragment that displays a polymorphism between Landsberg erecta and Columbia can be mapped by molecular techniques.

Doubled haploid lines (DH)
Doubled haploid lines contain two identical sets of chromosomes in each cell. They are completely homozygous, as only one allele is available for all genes. Doubled haploids can be produced from haploid lines. Haploid lines either occur spontaneously, as in the case of rape and maize, or are artificially induced. Haploid plants are smaller and less vital than diploids and are nearly sterile. It is possible to induce haploids by culturing immature anthers on special media. Haploid plants can later be regenerated from the haploid cells of the gametophyte. A second option is microspore culture. Occasionally in haploid plants the chromosome number doubles spontaneously, leading to doubled haploid (DH) plants. Such lines can also be obtained by colchicine treatment of haploids or of their parts. Colchicines prevent the formation of the spindle apparatus during mitosis, thus inhibiting the separation of chromosomes and leading to doubled haploid cells. If callus is induced in haploid plants, a doubling of chromosomes often occurs spontaneously during endomitosis and doubled haploid lines can be regenerated via somatic embryogenesis. However, in vitro culture conditions may reduce the genetic variability of regenerated materials to be used for genetic mapping. Doubled haploid lines constitute a permanent resource for mapping purposes and are ideal crossing partners in the production of mapping populations because they have no residual heterozygosity. Examples wheat, barley, and rice (Chao et al., 1989, Heun et al., 1991 and McCouch et al., 1988) [7, 15, 23].
The expected genotypic segregation ratio is 1:1, irrespective of whether the marker is dominant or codominant. DH plants are fully fertile and if suitable can be used as parents or released as a cultivar by breeding programs. DHs have been widely used for cultivar development, genetic mapping, mutagenesis, and gene function studies (Ferrie and Mollers, 2011; Hussain et al., 2012) [10, 16]. However, distorted segregation ratios can be observed, reducing the accuracy of genetic maps. This may be due to several causes: (1) genetic factors due to gametic or zygotic selection for pollen tube competition, preferential fertilization, chromosome translocation, etc. (Liu et al., 2010) [20]; (2) the genotype-dependency of DH, i.e., the different responses of the cross parents to DH method (Tanhuuani et al., 2008) [37]; (3) somaclonal variation arising during DH production resulting in aneuploid production (Oleszczuk et al., 2011) [29]; and (4) high frequencies of clones via androgenesis (Oleszczuk et al., 2014) [28].

Haploids have long been proved to be invaluable materials for basic genetic studies and they can be used for quick generation of quadruple, quintuple, sextuple, or higher order multiple mutant combinations, production of homozygous maternal gametophyte lethal mutants, and detection of gene conversion events during meiosis. (Wijnker et al., 2013; Ravi et al., 2014; Fulchar and Riha, 2016) [40, 31, 12]. A new breeding concept “reverse breeding” using DH has recently been proposed and successfully demonstrated in Arabidopsis, in which the approach also enabled the quick generation of a series of chromosome substitution lines (Wijnker et al., 2012) [41].

Backcross Populations

To analyze specific DNA fragments derived from parent A in the background of parent B, a hybrid F1 plant is backcrossed to parent B. In this situation, parent A is the donor of DNA fragments and parent B is the recipient. The latter is also called the recurrent parent. During this process two goals are achieved: unlinked donor fragments are separated by segregation and linked donor fragments are minimized due to recombination with the recurrent parent. To reduce the number and size of donor fragments, backcrossing is repeated and, as a result, so-called advanced backcross lines are generated. With each round of backcrossing, the proportion of the donor genome is reduced by 50%. Molecular markers help to monitor this process and to speed it up.

Introgression Lines: Exotic Libraries

The breeding of superior plants consists of combining positive alleles for desirable traits on the elite cultivar. One source for such alleles conferring traits such as disease resistance or quality parameters is distantly related or even wild species. If the trait to be introduced is already known, the introgression can be performed in a direct way supported by marker-assisted selection. However, the potential of wild species to influence quantitative traits often is not yet assessed. In this case, backcross breeding is a method to identify single genetic components contributing to the phenotype. NILs are constructed by an advanced backcross program, and their phenotypic effects are assayed. For example, in the work of (Tanksley et al., 1996) [38], loci from the wild tomato species Lycopersicum pimpinellifolium were shown to have positive effects on tomato fruit size and shape. To assess the effects of small chromosomal introgressions at a genome-wide level, a collection of introgression lines, each harboring a different fragment of genomic DNA, can be generated. Such a collection is called an exotic library, which is achieved by advanced backcrossing. This corresponds to a process of recurrent backcrossing (ADB) and marker-assisted selection for six generations and to the self-fertilization of two more generations to generate plants homozygous to the introgressed DNA fragments (Zamir, 2001) [45]. An example is the introgression lines derived from a cross between the wild green-fruit species L. pennelli and the tomato variety M82 (Eshed and Zamir, 1995) [9]. The lines, after the ADB program, will resemble the cultivated parent, but introgressed fragments with even subtle phenotypic effects can be easily identified. In other words, phenotypic assessment for all traits of interest will reveal genomic fragments with positive effects on measurable traits. The introgressed fragments are obviously defined by the use of molecular markers. In this context, it should be noted that recombination is reduced in interspecific hybrids with respect to intraspecific ones because differences in DNA sequence lead to reduced pairing of the chromosomes during meiosis. This, in turn, causes a phenomenon called linkage drag, which describes the situation when larger-thanexpected fragments are retained during backcross breeding (Young and Tanksley, 1989) [42]. The following example illustrates this concept. For the Tm2a resistance gene introgressed into tomato from the distantly related Lycopersicum peruvianum species, the ratio of physical to genetic distance is more than 4000 kb cM–1, whereas the average ratio in the cultivated species is about 700 kb cM–1 (Ganal et al., 1989) [13].

F2 Populations

The simplest form of a mapping population is a collection of F2 plants. This type of population was the basis for the Mendelian laws (1865) in which the foundations of classic genetics were laid. Two pure lines that result from natural or artificial inbreeding are selected as parents, parent 1 (P1) and parent 2 (P2). Alternatively, doubled haploid lines can be used to avoid any residual heterozygosity. If possible, the parental lines should be different in all traits to be studied. The degree of polymorphism can be assessed at the phenotypic level (e.g., morphology, disease resistance) or by molecular markers at the nucleic acid level. For inbreeding species such as soybean and the Brassicaceae, wide crosses between genetically distant parents help to increase polymorphism. However, it is required that the cross lead to fertile progeny. The progeny of such a cross is called the F1 generation. If the parental lines are true homozygotes, all individuals of the F1 generation will have the same genotype and have a similar phenotype. This is the content of Mendel’s law of uniformity. An individual F1 plant is then selfed to produce an F2 population that segregates for the traits different between the parents. F2 populations are the outcome of one meiosis, during which the genetic material is recombined. The expected segregation ratio for each codominant marker is 1:2:1 (homozygous like P1: heterozygous: homozygous like P2). It is a disadvantage that F2 populations cannot be easily preserved, because F2 plants are frequently not immortal, and F1 plants that result from their selfing are genetically not identical. For species like sugar beet, there is a possibility of maintaining F2 plants as clones in tissue culture and of multiplying and re-growing them when needed. A particular strategy is to maintain the F2 population in pools of F1 plants. Traits that can be evaluated only in hybrid plants, such as quality and yield parameters in sugar beet or maize, require the construction of testcross
plants by crossing each F₂ individual with a common tester genotype (Schneider et al., 2002) [32]. Ideally, different common testers should produce corresponding results to exclude the specific effects of one particular tester genotype. To produce a genome-wide map as an overview, a population of around 100 F₂ individuals is recommended as a compromise between resolution of linked loci and cost/feasibility. For mapping quantitative trait loci (QTLs), Monte Carlo simulations have shown that at least 200 individuals are required (Bevis, 1994). For higher resolution, as required for positional cloning of selected genes, progenies of several thousand plants are developed. For example, more than 3400 individuals were analyzed to obtain a detailed map around a fruit weight locus in tomato (Alpert and Tanksley, 1996) [11].

Problems in Genetic Mapping
A loss in genetic diversity inevitably causes problems for the breeding of new varieties. The genus Lycopersicon, which comprises modern tomato cultivars, is an example of this development (Miller and Tanksley, 1990) [24]. A second problem in genetic mapping is distorted segregation. This term describes a deviation from the expected Mendelian proportion of individuals in a given genotypic class within a segregating population (Lyttle, 1991) [21].

References
1. Alpert KB and Tanksley SD. High-resolution mapping and isolation of a yeast artificial chromosome contig containing fwa2.2: A major fruit weight quantitative trait locus in tomato. PNAS USA, 1996; 93:15503-15507.
2. Bailey DW. Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation. 1971; (11):325-327.
3. Bevis WD. The power and deceit of QTL experiments: lessons from comparative QTL studies. Proc. 49th Ann. Corn and Sorghum Ind. Res. Conf., Washington, DC American Seed Trade Association, 1994.
4. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 1980; (32):314-331.
5. Burr B, Burr FA. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends in Genet. 1991; (7):55-60.
6. Buschges R, Holrichter K, Panstruga R. The barley Mlo gene: a novel control element of plant pathogen resistance. Cell 1997; (88):695-705.
7. Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD. RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. 1989; (78):495-504.
8. Chen X, Salamini F, Gebhardt C. A potato molecular-function map for carbohydrate metabolism and transport. Theor. Appl. Genet. 2001; (102):284-295.
9. Eshed Y, Zamir D. An introgression line population of Lycopersicon pennelli in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics. 1995; (141):1147-1162.
10. Ferrie AMR, Mollers C. Haploids and doubled haploids in Brassica Spp. for genetic and genomic research. Plant Cell Tissue Organ Cult. 2011; (104):375-386. doi: 10.1007/s11240-010-9831-4
11. Fisher RA. On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. London, Ser. A. 1921; (122):309-368.
12. Fulcher N, Riha K. Using centromere mediated genome elimination to elucidate the functional redundancy of candidate telomere binding proteins in Arabidopsis thaliana. Front. Genet. 2016; (6):349. doi: 10.3389/fgen.2015.00349
13. Galan MW, Young ND, Tanksley SD. Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the Tom-2a region of chromosome 9 in tomato. Mol. Gen. Genet. 1989; (215):395-400.
14. Haldane JBS, Smith CAB. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eugen. 1947; (14):10-31.
15. Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrells ME, Tanksley. Construction of a restriction fragment length polymorphism map for barley (Hordeum vulgare). Genome. 1991; (34):437-447.
16. Hussain B, Kha M, Ali Q, Shaukat S. Double haploid production is the best method for genetic improvement and genetic studies of wheat. Int. J. Agro Vet. Med. Sci. 2012; (6):216-228. doi: 10.5455/ijavms.169.
17. Koornneef M, Hanhart CJ, von Loenen Martinet EP, Peeters AJM and van der Veen JH. Trisomics in Arabidopsis thaliana and the location of linkage groups. Arabidopsis Inf. Serv. Frankfurt. 1987: (23):46-50.
18. Lander ES, Green P, Abrahamson J, Barlow A, Daly M, Lincoln SE, Newburg L. MAPMAKER: an interactive computer package for constructing primary linkage maps of experimental and natural populations. Genomics. 1987; (1):174-181.
19. Lister C, Dean C. Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana. Plant J. 1993; (4):745-750.
20. Liu X, Guo L, You J, Liu X, He Y, Yuan J. Progress of segregation distortion in genetic mapping of plants. Res. J. Agron. 2010; (4):78-83. doi: 10.3923/rjagr.2010.78.83
21. Lyttle TW. Segregation distorters. Annu. Rev. Genet. 1991; (25):511-557.
22. Madhusudhana R. “Linkage mapping,” in Sorghum Molecular Breeding, eds. R. Madhusudhana, P. Rajendrakumar, and J. V. Patil (New Delhi: Springer), 2015, 47-70.
23. McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR et al. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 1988; (76):815-829.
24. Miller JC, Tanksley SD. RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theor Appl Genet. 1990: (80):437-448.
25. Morton N. Sequential tests for the detection of linkage. Am. J. Hum. Genet. 1955; (7):277-318.
26. Neuffer MG, Coe E, Wessler S. Mutants of Maize. Cold Spring Harbor Laboratory, New York, 1997.
27. O’Donoughue LS, Kiamian SF, Rayapati PJ, Penner GA, Sorrells ME. Molecular linkage map of cultivated oat (Avena byzantina X A. sativa cv. Ogle). Genome. 1995; (38):368-380.
28. Oleszczuk S, Tytka M, Zimny J. The origin of clones among androgenic regenerants of hexaploid triticale. Euphytica 2014; (198):325-336. doi: 10.1007/s10681-014-1109-1
29. Oleszczuk S, Zimny J, Rabiz-Swider J, Lukaszewski AJ. Aneuploidy among androgenic progeny of hexaploid

— 3936 —
triticale (×Triticosecale Wittmack). Plant Cell Rep. 2011; (30):575-586. doi: 10.1007/s00299-010-0971-0
30. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Curr. Opinion in Plant Biol. 2002; (5):94-100.
31. Ravi M, Marimuthu MP, Tan EH, Maheshwari S, Henry IM, Marin- Rodriguez B. A haploid genetics toolbox for Arabidopsis thaliana. Nat. Commun. 2014; (5):5334. doi: 10.1038/ncomms6334
32. Schneider K, Schfer-Pregl R, Borchardt DC, Salamini F. Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor. and Appl. Genet. 2002; (104):1107-1113.
33. Seymour DK, Filiault DL, Henry IM, Monson-Miller J, Ravi M, Pang A. Rapid creation of arabidopsis doubled haploid lines for quantitative trait locus mapping. Proc. Natl. Acad. Sci. U.S.A. 2012; (109):4227-4232. doi: 10.1073/pnas.1117277109
34. Stam P. Construction of integrated linkage maps by means of a new computer package: JOINMAP. The Plant Journal 1993; (5):739-744.
35. Swiecicki WK, Wolko B, Weeden NF. Mendel’s genetics, the Pisum genome and pea breeding. Vortr. Pflanzenz.chtg. 2000; (48):65-76.
36. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The Double-Strand-Break Repair Model for Recombination. Cell. 1983; (33):25-35.
37. Tanhuanpää P, Kalendars R, Schulman AH, Kiviiharju E. The first doubled haploid linkage map for cultivated oat. Genome 2008; (51):560-569. doi: 10.1139/G08-040
38. Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V et al. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. Pimpinellifolium. Theor. Appl. Genet. 1996; (92):213-224.
39. Wang GL, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics. 1994; (136):1421-1434.
40. Wijnker E, Velikkakam Jame G, Ding J, Becker F, Klasen JR, Rawat V. The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana. Elife 2013; (2):e01426 doi: 10.7554/eLife.01426
41. Wijnker E, Van Dun K, De Snoo CB, Lelivelt CL, Keurentjes JJ, Naharudin NS. Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. Nat. Genet. 2012; (44):467-470. doi: 10.1038/ng.2203
42. Young ND, Tanksley SD. RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backcross breeding. Theor. Appl. Genet. 1989; (77):353-359.
43. Zamir D. Improving plant breeding with exotic genetic libraries. Nature Rev. 2001; (2):983-989.