Part I
Genetic Mapping

The Handbook of Plant Genome Mapping.
Genetic and Physical Mapping. Edited by Khalid Meksem, Günter Kahl

ISBN 3-527-31116-5

1
Mapping Populations and Principles of Genetic Mapping

Katharina Schneider

Overview

Mapping populations consist of individuals of one species, or in some cases they derive from crosses among related species where the parents differ in the traits to be studied. These genetic tools are used to identify genetic factors or loci that influence phenotypic traits and to determine the recombination distance between loci. In different organisms of the same species, the genes, represented by alternate allelic forms, are arranged in a fixed linear order on the chromosomes. Linkage values among genetic factors are estimated based on recombination events between alleles of different loci, and linkage relationships along all chromosomes provide a genetic map of the organism. 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. This chapter illustrates the molecular basis of recombination, summarizes the different types of mapping populations, and discusses their advantages and disadvantages for different applications.

Abstract

In genetics and breeding, mapping populations are the tools used to identify the genetic loci controlling measurable phenotypic traits. For self-pollinating species, F_2 populations and recombinant inbred lines (RILs) are used; for self-incompatible, highly heterozygous species, F_1 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. Recombination frequencies between traits and markers reveal their genetic distance, and trait-linked markers can be anchored, when necessary, to a more complete genetic map of the species. For map-based cloning of a gene, populations of a large size provide the resolution required.

The Handbook of Plant Genome Mapping.
Genetic and Physical Mapping. Edited by Khalid Meksem, Günter Kahl

ISBN 3-527-31116-5

Due to intensive breeding and pedigree selection, genetic variability within the gene pools of relevant crops is at risk. Interspecific crosses help to increase the size of the gene pool, and the contribution of wild species to this germ plasm in the form of introgression lines is of high value, particularly with respect to traits like disease resistance. The concept of exotic libraries with near-isogenic lines, each harboring a DNA fragment from a wild species, implements a systematic scan of the gene pool of a wild species.

To describe the complexity of genome organization, genetic maps are not sufficient because they are based on recombination, which is largely different along all genomes. However, genetic maps, together with cytogenetic data, are the basis for the construction of physical maps. An integrated map then provides a detailed view on genome structure and enforces positional cloning of genes and ultimately the sequencing of complete genomes.

1.1 Introduction

Since Mendel formulated his laws of inheritance in 1865, it is a core component of biology to relate genetic factors to functions visible as phenotypes. At Mendel's time, genetic analysis was restricted to visual inspection of the plants. Pea (*Pisum sativum* [Fabaceae]) was already a model plant at the time, and Mendel studied visible traits such as seed and pod color, surface structure of seeds and pods (smooth versus wrinkled), and plant height. These traits are, in fact, the first genetic markers used in biology. In 1912 Vilmorin and Bateson described the first work on linkages in *Pisum*. However, the concept of linkage groups representing chromosomes was not clear in *Pisum* until 1948, when Lamprecht described the first genetic map with 37 markers distributed on 7 linkage groups (summarized in Swiecicki et al. 2000). Large collections of visible markers are today available for several crop species and for *Arabidopsis thaliana* (Koornneef et al. 1987; Neuffer et al. 1997).

In the process of finding more and more genetic markers, the first class of characters scored at the molecular level was isoenzymes. These are isoforms of proteins that vary in amino acid composition and charge and that can be distinguished by electrophoresis. The technique is applied to the characterization of plant populations and breeding lines and in plant systematics, but it is also used for genetic mapping of variants, as shown particularly in maize (Frei et al. 1986, Stuber et al. 1972). However, due to the small number of proteins for which isoforms exist and that can be separated by electrophoresis, the number of isoenzyme markers is limited.

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). This technique requires the hybridization of a specific probe to restricted genomic DNA of different genotypes. The whole genome can be covered

by RFLP and, depending on the probe, coding or non-coding sequences can be analyzed. The next generation of markers was based on PCR: rapid amplified polymorphic DNA (RAPD) (Williams et al. 1990; Welsh and McClelland 1990) and amplified fragment length polymorphism (AFLP) (Vos et al. 1995). Recently, methods have been developed to detect single nucleotide polymorphisms (summarized in Rafalski 2002). 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 selected and mapped. In this way function maps can be constructed (Chen et al. 2001; Schneider et al. 2002). Phenotypic data of the segregating population, correlated to marker data, prove or disprove potential candidate genes supporting mono- and 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) is acknowledged to best explain meiotic reciprocal recombination (Figure 1.1). 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 Büschges et al. (1997) 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 fre-

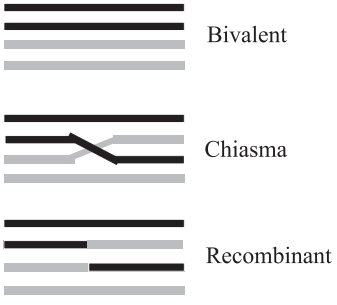

Figure 1.1. Generation of recombinants by chiasma formation. In the meiotic prophase, two sister chromatids of each parent (labeled in red and green, respectively) align to form a bivalent. A chiasma is formed by a physical strand exchange between two non-sister chromatids. Breakage and reunion of reciprocal strands leads to the generation of recombinants.

quency 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). Large amounts of segregation data are routinely processed by computer programs to calculate a genetic map; among the most popular are JoinMap (Stam 1993) and MAPMAKER (Lander et al. 1987).

1.2 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. Based on these concepts, this section will be divided into seven parts:

1. mapping populations suitable for self-fertilizing plants,
2. mapping populations for cross-pollinating species,
3. two-step strategies for mapping mutants and DNA fragments,
4. chromosome-specific tools for mapping,
5. mapping in natural populations/breeding pools,
6. mapping genes and mutants to physically aligned DNA, and
7. specific mapping problems.

1.2.1 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 F_2 plants, recombinant inbred lines (RIL), backcross (BC) populations, introgression lines assembled in exotic libraries, and doubled haploid lines (DH).

1.2.1.1 F_2 Populations

The simplest form of a mapping population is a collection of F_2 plants (Figure 1.2). 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 (see Section 1.2.1.5). 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 F_1 generation. If the parental lines are true homozygotes, all individuals of the F_1 generation will have the same genotype and have a similar phenotype. This is the content of Mendel's law of uniformity. An individual F_1 plant is then selfed to produce an F_2 population that segregates for the traits different between the parents. F_2 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 F_2 populations cannot be easily preserved, because F_2 plants are frequently not immortal, and F_3 plants that result from their selfing are genetically not iden-

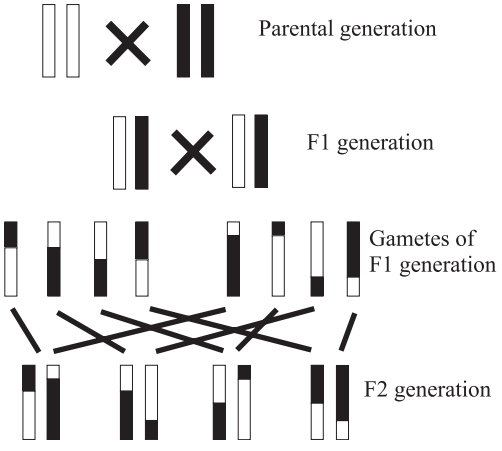

Figure 1.2. The generation of an F_2 population. Two chromosomes are shown as representatives of the diploid parental genome. In the parental generation, the genotypes are homozygous (represented by equal colors) and in the F_1 generation are heterozygous. For gamete formation the genetic material undergoes meiosis, leading to recombination events in F_1 gametes. Correspondingly, F_2 plants vary largely in their genetic constitution.

tical. For species like sugar beet, there is a possibility of maintaining F_2 plants as clones in tissue culture and of multiplying and re-growing them when needed. A particular strategy is to maintain the F_2 population in pools of F_3 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 F_2 individual with a common tester genotype (example given in Schneider et al. 2002). 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 F_2 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).

1.2.1.2 **Recombinant Inbred Lines**

Recombinant inbred lines (RILs) are the homozygous selfed or sib-mated progeny of the individuals of an F_2 population (Figures 1.2, 1.3). The RIL concept for mapping genes was originally developed for mouse genetics. In animals, approximately 20 generations of sib mating are required to reach useful levels of homozygosity. In plants, RI lines are produced by selfing, unless the species is completely self-incompatible. Because in the selfing process one seed of each line is the source for the next generation, RILs are also called single-seed descent lines. Self-pollination allows the production of RILs in a relatively short number of generations. In fact, within six generations, almost complete homozygosity can be reached. Along each chromosome, blocks of alleles derived from either parent alternate. Because recombination can no longer change the genetic constitution of RILs, further segregation in the progeny of such lines is absent. It is thus one major advantage that these lines constitute a permanent resource that can be replicated indefinitely and

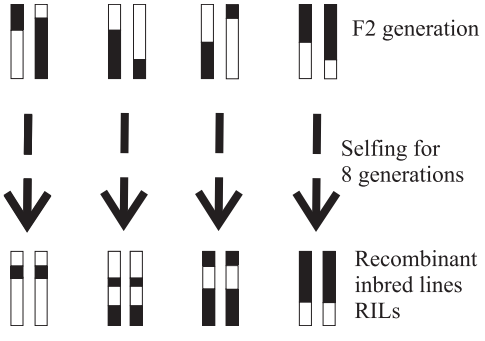

Figure 1.3. The generation of RILs. For the generation of RILs, plants of an F_2 population are continuously selfed. In each generation meiotic events lead to further recombination and reduced heterozygosity until completely homozygous RILs with fragments of either parental genome are achieved.

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 F_2 populations. Consequently, RIL populations show a higher resolution than maps generated from F_2 populations (Burr and Burr 1991), 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).

In *Arabidopsis thaliana*, 300 RILs have become a public mapping tool (Lister and Dean 1993). *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.

1.2.1.3 **Backcross Populations**

To analyze specific DNA fragments derived from parent A in the background of parent B, a hybrid F_1 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 % (see Figure 1.4). Molecular markers help to monitor this process and to speed it up. In an analysis of the chro-

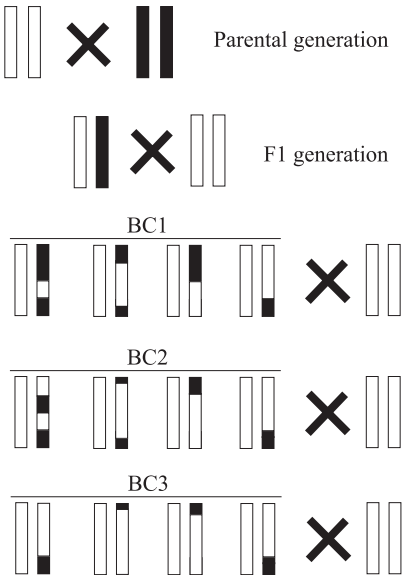

Figure 1.4. The generation of advanced backcross lines (BC). BC lines originate from an F_1 plant that is repeatedly backcrossed to the same recurrent parent. With each round of backcrossing, the number and size of genomic fragments of the donor parent are reduced until a single DNA fragment distinguishes the BC line from the recurrent parent (white: recurrent parent; black: donor parent).

mosomal segments retained around the *Tm-2* locus of tomato, it was estimated that marker-assisted selection reduced the number of required backcrosses from 100 – in the case of no marker selection – to two (Young and Tanksley 1989). The progeny of each backcross is later screened for the trait introduced by the donor. In the case of dominant traits, the progeny can be screened directly; in the case of recessive traits, the selfed progeny of each backcross plant has to be assessed.

Lines that are identical, with the exception of a single fragment comprising one to a few loci, are called nearly isogenic lines (NILs). The generation of NILs involves several generations of backcrossing assisted by marker selection. To fix the donor segments and to visualize traits that are caused by recessive genes, two additional rounds of self-fertilization are required at the end of the backcrossing process. If two NILs differ in phenotypic performance, this is seen as the effect of the alleles carried by the introgressed DNA fragment. The procedure is quite helpful in the functional analysis of the underlying genes. The strategy is particularly valuable for those species for which no transformation protocol is established to produce transgenics for the alleles of interest. A further advantage is that in NILs genomic rearrangements, which may happen during transformation, are avoided.

Backcross breeding is an important strategy if a single trait, such as resistance, has to be introduced into a cultivar that already contains other desirable traits. The only requirement is that the two lines be crossable and produce fertile progeny. Lines incorporating a fragment of genomic DNA from a very distantly related species are called introgression lines, whereas lines incorporating genetic material from a different variety are indicated as intervarietal substitution lines.

1.2.1.4 **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), loci from the wild tomato species *Lycopersicon 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 (summarized in Zamir 2001). An example is

the introgression lines derived from a cross between the wild green-fruited species *L. pennellii* and the tomato variety M82 (Eshed and Zamir 1995). 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-than-expected fragments are retained during backcross breeding (Young and Tanksley 1989). 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).

1.2.1.5 **Doubled Haploid Lines**

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. In cultivated barley it is possible to induce the generation of haploid embryos by using pollen from the wild species *Hordeum bulbosum*. During the first cell divisions of the embryo, the chromosomes of *H. bulbosum* are eliminated, leaving the haploid chromosomal set derived from the egg cell. 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 of their use in wheat, barley, and rice are found in Chao et al. (1989), Heun et al. (1991), and McCouch et al. (1988).

1.2.2 Mapping Populations for Cross-pollinating Species

If pure lines cannot be generated from a species due to self-incompatibility or inbreeding depression, heterozygous parental plants are used to derive mapping populations such as F_1 and backcross lines (BC). This is the case for several tree species such as apple, pear, and grape and for potato. For the tree species crosses between different cultivars are used to produce F_1 progenies to be genotyped (Maliegaard et al. 1998; Yamamoto et al. 2002; Grando et al. 2003). In potato, the heterozygosity of parental lines used for one cross was evaluated to correspond to 57–59 % (Gebhardt et al. 1989). In the foundation cross population, different alleles are contributed from either parent to individual F_1 plants. The linkage among markers is assessed by the production of a genetic map for either parent. In potato, Gebhardt et al. (1989, 1991) reported the construction of a backcross population in which an individual F_1 plant was pollinated with one parent. To maintain the identity of the F_1 genotypes of the mapping population, parental lines and each of their F_1 progenies were propagated clonally.

1.2.3 Two-step Strategies for Mapping Mutants and DNA Fragments

Mapping mutants always requires the construction of a segregating population. In a first step, tightly linked or co-segregating markers are selected. In a second step, the map position of these markers is determined. Whereas the first step requires a population segregating for the trait of interest, the linked marker can be anchored in a reference population for which a dense marker framework is available. Such a method is particularly applicable when a large set of mutants needs to be mapped in a limited time.

This concept was implemented by Castiglioni et al. (1998), who mapped mutations in barley by an AFLP-based procedure. The procedure takes advantage of the very high diversity index of AFLP markers that allows the screening of a whole genome with a limited number of PCR primer combinations. In the cited work, a genetic map comprising 511 AFLP markers was derived from the cross between the lines Proctor × Nudinka. To map morphological mutations, mutants were crossed to Proctor and to Nudinka, respectively, and 30–50 mutant plants were selected for AFLP analysis. The specific presence or absence of AFLP bands in the mutant population was correlated to identical fragments mapped in the Proctor × Nudinka cross, thus inferring their map position and, consequently, anchoring the mutant locus to the same map.

A different approach to enrich for linked markers is based on the concept of bulked segregant analysis (BSA), introduced by Michelmore et al. (1991). BSA requires only a population segregating for the trait of interest. Two bulks or pools of segregating genotypes are selected in which either the mutant or the wild-type phenotype is present and homozygous. This implies that within each pool the individuals are identical with respect to the genomic region in which the responsible

gene maps, but the genetic constitution of the rest of the genome is random, while between the pools the selected region is molecularly dissimilar. The bulks are screened for polymorphisms by molecular techniques such as RFLP, RAPD, or AFLP. In their first BSA analysis, Michelmore et al. (1991) identified markers linked to a gene conferring resistance to downy mildew in lettuce. They generated bulks of 17 F_2 individuals homozygous for alternate alleles of the resistance locus DM5/8 and analyzed them with 100 arbitrary RAPD primers to detect around 900 loci. Three RAPD markers linked to the resistance locus were identified. This work shows that markers can be reliably identified in a 0- to 25-cM window to either side of the locus of interest. The method can be applied iteratively, in the sense that new bulks are constructed based on each new marker linked more closely to the gene. The linkage of each marker with the tagged locus is verified by analyzing single plants of the segregating population. In the work cited, the BSA pools were from the mapping population. However, the BSA markers could also be anchored in a different population, according to the two-step procedure outlined before.

1.2.4 Chromosome-specific Tools for Mapping

Chromosome-specific tools allow a segregating population to be genotyped in a way that each chromosome is directly scanned for linkage. The first tools of this kind were mutant lines with one or more visible mapped mutations. For *Arabidopsis thaliana* such multiple marker lines are available: the line W100, for example, contains mutations identifying each arm of the five chromosomes (Koornneef et al. 1987). The marker lines with the genotypes *aa*, *bb*, to *zz* are crossed to the mutant line with the genotype *mm*. The progeny is selfed to generate an F_2 population. In this population the frequency of double mutants *aa*/*mm*, *bb*/*mm*, to *zz*/*mm* is counted. It is expected that in the case of recessive inheritance of both mutations, 1/16 of all progeny are double mutants. If there are less than the expected number of double mutants, this is taken as evidence of reduced recombination due to the linkage between the locus *m* and one of the markers tested. The more the two loci are in physical proximity, the fewer the recombination events are.

In the past decade, numerous molecular marker sets have been developed. For a defined mutation in a given genetic background, a cross to a contrasting genotype needs to be performed, and an F_2 is generated from which plants displaying the phenotype are selected for molecular analysis. In *Arabidopsis*, different ecotypes such as Columbia and Landsberg *erecta* are well established, and marker sets for cleaved amplified polymorphic sequence (CAPS) and simple sequence repeat (SSR) analysis have been developed (Konieczny and Ausubel 1993; Bell and Ecker 1994). Marker loci covering the entire genome are tested in the mutant F_2 plants, and, again, reduced recombination between the mutant allele and a marker allele indicates linkage to the marker locus.

1.2.5 Mapping in Natural Populations/Breeding Pools

The natural variation between individuals of one species can be exploited for mapping. In the case of crop plants, sets of different breeding lines can fulfill this purpose. This approach is suited to map complex traits that are influenced by the action of many genes in a quantitative way. Such loci are defined as quantitative trait loci (QTL). It is important that such a collection of different accessions contain a whole spectrum of phenotypes for a given trait; in particular, the availability of extreme phenotypes is advantageous. The underlying idea is that genomic fragments naturally present in a particular genotype are transmitted as non-recombining blocks and that markers, like single-nucleotide polymorphisms (SNPs) and insertions/deletions, can easily follow the inheritance of such blocks. These units are also called haplotypes, and their existence reveals a state of linkage disequilibrium (LD) among allelic variants of tightly linked genes. The existence of haplotypes has clearly been shown for maize and sugar beet (Ching et al. 2002, Schneider et al. 2001).

Two strategies exist for searching for haplotypes specifically associated with extreme phenotypes. On the one hand, the whole genome can be scanned for phenotype-marker associations, if sufficient sequence information is available for marker development. However, as the linkage disequilibrium often does not extend for more than 2000 bases (reviewed in Buckler and Thornsberry 2002), this approach is generally too time-consuming and costly. Alternatively, the focus is on candidate genes that, upon their predicted physiological function, are likely to influence the trait under investigation. The latter approach has been successfully applied in maize. An association between a marker and a trait exists if one marker allele or haplotype is significantly associated with a particular phenotype when studied in unrelated genotypes. In maize this has been shown for *dwarf 8*, a locus influencing plant height (Thornsberry et al. 2001), and for six genes of the starch metabolism influencing kernel quality (Whitt et al. 2002). The advantage of this approach is that it does not require the construction of experimental populations. Particularly for self-pollinating species, inbred individuals of natural ecotypes are practically immortal, and phenotyping needs to be performed only once. Natural populations are particularly informative because usually more than two alleles exist for each marker locus. For crop species, different breeding lines representing extreme phenotypes can be used for the same purpose.

As unrelated lines of natural populations are genetically separated by many generations, the corresponding large number of meiotic events leads to a high rate of recombinations. Therefore, and with the limit that LD blocks still exist, trait-supporting loci can be mapped with high precision, largely exceeding the resolution of F_2 populations. Association mapping can thus greatly accelerate QTL positional cloning approaches. However, it requires thorough statistical assessment to investigate the relatedness of the lines and the overall population structure. Only if the population structure is homogenous can an association between a haplotype and a phenotype be considered realistic.

1.2.6
Mapping Genes and Mutants to Physically Aligned DNA

The distances in genetic maps are based on recombination frequencies. However, recombination frequencies are not equally distributed over the genome. In heterochromatic regions such as the centromeres, recombination frequency is indeed quite reduced. In these cases, cytogenetic maps can provide complementary information because they are based on the fine physical structure of chromosomes. Chromosomes can be visualized under the microscope and be characterized by specific staining patterns, e. g., with Giemsa C, or can be based on morphological structures such as the centromeres, the nucleolus-organizing region (NOR), the telomeres, and the so-called knobs, heritable heterochromatic regions of particular shape. Cytogenetic maps allow association of linkage groups with chromosomes and determination of the orientation of the linkage groups with respect to chromosome morphology. In species such as maize, wheat, barley, and *Arabidopsis thaliana*, lines carrying chromosome deletions, translocation breakpoints, or trisomics can be generated as valuable tools for the cytogenetic approach (Helentjaris et al. 1986; Weber and Helentjaris 1989; Sandhu et al. 2001; Künzel et al. 2000; Koornneef and Vanderveen 1983). Numerical aberrations in chromosome number, together with marker data generated, e. g., by RFLP analysis, can clearly identify chromosomes. Defined translocation breakpoints can also localize probes to specific regions on the arms of chromosomes.

More recently, techniques have been developed to localize nucleic acids *in situ* on the chromosomes. During the pachytene, a stage during the meiotic prophase, the chromosomes are 20 times longer than at mitotic metaphase. They display a differentiated pattern of brightly fluorescing heterochromatin segments. It is possible to identify all chromosomes based on chromosome length, centromere position, heterochromatin patterns, and the positions of repetitive sequences such as 5S rDNA and 45S rDNA visualized by fluorescence *in situ* hybridization (FISH) (Hanson et al. 1995), as shown in *Medicago trunculata* by Kulikova et al. (2001). In tomato this approach has been successful in mapping two genes near the junction of euchromatin and pericentromeric heterochromatin (Zhong et al. 1999). Refined multicolor FISH even allows the mapping of single-copy sequences (Desel et al. 2001). In this context, cytogenetic maps based on FISH provide complementary information for the construction of physical maps to position BAC clones (Islam-Faridi et al. 2002) and other DNA sequences along the chromosomes (see Figure 1.5).

1.2.7
Specific Mapping Problems

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). When tomato was introduced from Latin America to Europe by Spanish explorers, presumably only lim-

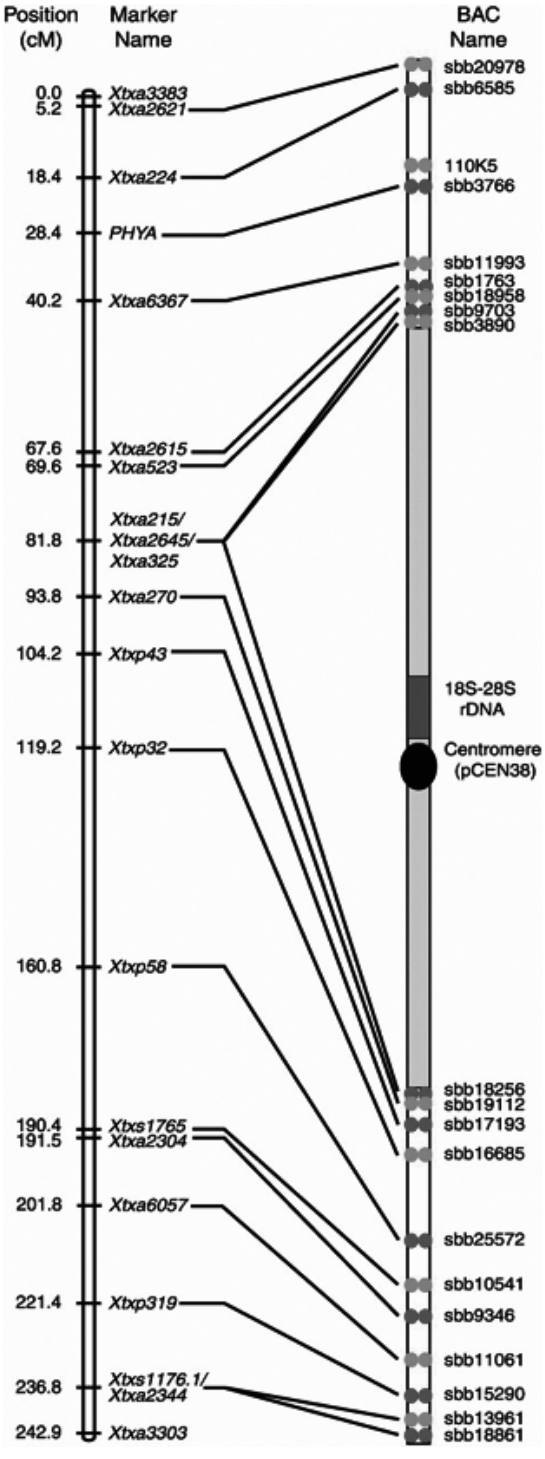

Figure 1.5. Alignment of the genetic and physical map of sorghum chromosome 1 (figure taken from Islam-Faridi et al. 2002, courtesy of *Genetics*). A diagrammatic representation of the cytogenetic locations of 20 sorghum BACs, 18S-28S rDNA, and CEN38 (a centromere-associated clone; on right).

ited numbers of seeds (and accessions) were transferred and became the basis of today's modern cultivars. This created a bottleneck. Breeding methods such as single-seed descent and pedigree selection also promote genetic uniformity. As the tomato cultivars are generally self-compatible, this contributes even further to a decrease DNA polymorphism. For RFLP analysis, the degree of polymorphic probes has been found to be exceptionally low in tomato crosses between modern cultivars (Miller and Tanksley 1990). A further indication of the low genetic diversity of tomato genotypes is the small number of microsatellite alleles in a set of tomato varieties (Areshchenkova and Ganal 2002). Self-incompatible species of the *Lycopersicon* genus show a much higher genetic distance within and between accessions, indicating the role of the mode of reproduction in the maintenance of genetic variability. The use of landraces that are not genetically uniform is one option to increase genetic polymorphism. Given that almost all species of the *Lycopersicon* genus can be crossed with cultivated tomato, the construction of inter- rather than intraspecific crosses and populations is essential for introducing new genetic factors into the breeding pool of this crop.

A second problem that is often encountered 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). That one allelic class can be underrepresented due to a dysfunction of the concerned gametes is well known for plants (Xu et al. 1997). This can occur in pollen, in megaspores, or in both organs and can be explained either by the selective abortion of male and female gametes or by the selective fertilization of particular gametic genotypes. A selection process during seed development, seed germination, and plant growth can also be active. Gametophyte loci leading to a distorted segregation have been identified in rice (summarized in Xu et al. 1997). They are supposed to be responsible for the partial or total elimination of gametes carrying one of the parental alleles. A marker locus linked to a gametophyte locus, also referred to as a gamete eliminator or pollen killer, can also show distorted segregation. Self-incompatibility loci preventing self-pollination are also a direct cause of distorted segregation, as is discussed for potato (Gebhardt et al. 1991). Breeding programs that aim at the generation of specific recombinants are directly affected if one locus is close to a region affected by segregation distortion.

1.3 Discussion

In the field of plant breeding, genetic mapping is still the most valuable approach to identifying the genetic factors that underlie particularly quantitatively inherited traits. Genetically linked markers can be used in marker-assisted breeding to identify individuals with the desirable level of relevant characters at an early stage. High-throughput technologies based on SNP detection (Rafalski 2002) allow the scoring of thousands of data points in a short time. This has the potential to reduce

greatly the size of field trials. Therefore, it is important to have mapping tools at hand as described in this chapter.

Important factors to be balanced in the experimental design are the mapping resolution, the required time, and human resources. RILs, NILs, and introgression lines are very laborious and time-consuming to construct, although they allow applied assessments with high precision. Alternatively, local maps of the regions harboring mutations are produced in F_2 populations and cross-linked to populations with high mapping resolution (Castiglioni et al. 1998). Working with natural populations is an alternative that circumvents the construction of experimental mapping populations and still maintains a high resolution because it takes advantage of the multiple meiotic events that occurred during plant evolution (Rafalski 2002). Any marker in the region of linkage disequilibrium that surrounds a genetic factor responsible for a trait may be indicative of the level of the expression of a trait.

The next step is the identification of the genetic factor itself. Unless a gene is tagged by the candidate gene approach, positional cloning involving the partial construction of a physical map is the method of choice. This requires the construction of contigs based on large-insert clones such as BACs, which are assembled according to fingerprinting data. Local, but also genome-wide, physical maps are in progress or have already been obtained for plants with model character such as *Arabidopsis thaliana*, as well as for crops such as rice and sorghum (Mozo et al. 1999; Tao et al. 2001; Klein et al. 2003). In plants for which the whole genome has been sequenced, such as *Arabidopsis thaliana* (The Arabidopsis Genome Initiative 2000) and rice (Yu et al. 2002), this is becoming increasingly easier. Alternatively, new genetic markers located in underrepresented regions of the genome can be developed from the complete DNA sequence and applied in mapping. Once cloned, a gene becomes subject to transgenic approaches. Complementation of a phenotype missing a specific trait is the ultimate proof of gene function, and superior varieties can be created accordingly. Alternatively, association studies for the gene identified can be performed to detect superior alleles of the locus linked to specific markers.

In summary, mapping populations are the basic tools for understanding the effect of selected genetic factors and the organization of the genome of a species as a whole. They are the backbone of genomics research that aims to decipher large, complex genomes at the physical or even sequence level.

Acknowledgments

The author would like to thank Francesco Salamini for fruitful discussions in writing this chapter.

References

Alpert KB, Tanksley SD (1996) High-resolution mapping andisolation of a yeast artificial chromosome contig containing fw2.2: A major fruit weight quantitative trait locus in tomato. PNAS USA 93, 15503–15507.

The Arabidopsis Genome Initiative (2000) Nature 408, 796–815.

Areshchenkova T, Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104, 229–235.

Beavis WD (1994) 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.

Bell C, Ecker J (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics 19, 137–144.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5:107–111.

Burr B, Burr FA (1991) Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends in Genet. 7, 55–60.

Büschges R, Hollrichter K, Panstruga R et al. (1997) The barley Mlo gene: a novel control element of plant pathogen resistance. Cell 88, 695–705.

Castiglioni P, Pozzi C, Heun M, Terzi V, Müller KJ, Rohde W, Salamini F (1998) AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics 149: 2039–2056.

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. 78, 495–504.

Chen X, Salamini F, Gebhardt C (2001) A potato molecular-function map for carbohydrate metabolism and transport. Theor. Appl. Genet. 102, 284–295.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3, 19.

Desel C, Jung C, Cai D, Kleine M, Schmidt T (2001) High-resolution mapping of YACs and the single-copy gene Hs1(pro-1) on Beta vulgaris chromosomes by multi-colour fluorescence in situ hybridization. Plant Mol. Biol. 45:113–122.

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.

Fisher, RA (1921) On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. London, Ser. A 122, 309–368.

Frei OM, Stuber CW, Goodman MM (1986) Use of allozymes as genetic markers for predicting performance in maize single cross hybrids. Crop Sci. 26, 37–42.

Ganal MW, Young ND, Tanksley SD (1989) Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the Tm-2a region of chromosome 9 in tomato. Mol. Gen. Genet. 215: 395–400.

Gebhardt C, Ritter E, Debener T, Schachtschabel U, Walkemeier B, Uhrig H, Salamini F (1989) RFLP mapping and linkage analysis in Solanum tuberosum. Theor. Appl. Genet. 78, 65–75.

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufmann H, Thompson RD, Bonierbale MW, Ganal MW, Tanksley SD, Salamini F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome. Theor. Appl. Genet. 83, 49–57.

Grando MS, Bellin D, Edwards KJ, Pozzi C, Stefanini M, Velasco R (2003) Molecular linkage maps of Vitis vinifera L. and Vitis riparia Mchx. Theor. Appl. Genet. 106, 1213–1224.

Haldane JBS and Smith CAB (1947) A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eugen. 14, 10–31.

Hanson RE, Zwick M, Choi S, Islam-Faridi MN, McKnight TD (1995) Fluorescent in situ

hybridization of a bacterial artificial chromosome. Genome 38, 646–657.
Helentjaris T, Weber DF, Wright S (1986) Use of monosomics to map cloned DNA fragments in maize. Proc. Natl. Acad. Sci. USA 83, 6035–6039.
Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrells ME, Tanksley (1991) Construction of a restriction fragment length polymorphism map for barley (Hordeum vulgare). Genome 34, 437–447.
Islam-Faridi MN, Childs KL, Klein PE, Hodnett G, Menz MA, Klein RR, Rooney WL, Mullet JE, Stelly DM, Price HJ (2002) Molecular cytogenetic map of sorghum chromosome 1: fluorescence in situ hybridization analysis with mapped bacterial artificial chromosomes. Genetics 161, 345–353.
Klein PE, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignments of sorghum chromosome 3 and rice chromosome reveals extensive conservation of gene order and one major chromosomal rearrangement. The Plant J. 34, 605–521.
Konieczny A, Ausubel FM (1993) A procedure for mapping Arabidopsis mutations using co-dominant; ecotype-specific PCR-based markers. The Plant J. 4, 403–410.
Koornneef M, Vanderveen HJ (1983) A marker line, that allows the detection of linkage on all Arabidopsis chromosomes. Genetica 61, 41–46.
Koornneef M, Hanhart CJ, von Loenen Martinet EP, Peeters AJM, van der Veen JH (1987) Trisomics in Arabidopsis thaliana and the location of linkage groups. Arabidopsis Inf. Serv. Frankfurt 23, 46–50.
Kulikova O, Gualtieri G, Geurts R, Kim DJ. Cook D, Huguet T, de Jong HJ, Fransz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of Medicago truncatula. The Plant J. 27, 49–58.
Künzel G, Korzun L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154, 397–412.
Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary linkage maps of experimental and natural populations. Genomics 1, 174–181.
Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana. Plant J. 4, 745–750.
Lyttle TW (1991) Segregation distorters. Annu. Rev. Genet. 25, 511–557.
Maliepaard C et al. (1998) Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers. Theor. Appl. Genet. 97, 60–73.
McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76, 815–829.
Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. USA 88, 9828–9832.
Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theor Appl Genet 80, 437–448.
Morton N (1955) Sequential tests for the detection of linkage. Am. J. Hum. Genet. 7, 277–318.
Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martinssen R, Meier-Ewert S, Altmann T (1999) A complete BAC-based physical map of the Arabidopsis thaliana genome. Nat. Genet. 22, 271–275.
Neuffer MG, Coe E, Wessler S (1997) Mutants of Maize. Cold Spring Harbor Laboratory, New York.
O'Donoughue LS, Kianian SF Rayapati PJ Penner GA Sorrells ME et al. (1995) Molecular linkage map of cultivated oat (Avena byzantina X A. sativa cv. Ogle). Genome 38:368–380.
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr. Opinion in Plant Biol. 5:94–100.
Sandhu D, Champoux JA, Bondareva SN, Gill KS (2001) Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes. Genetics 157, 1735–1747.
Schneider K, Weisshaar B, Borchardt DC, Salamini F (2001) SNP frequency and allelic haplotype structure of Beta vulgaris expressed genes. Mol. Breeding 8, 63–74.

Schneider K, Schäfer-Pregl R, Borchardt DC, Salamini F (2002) Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor. and Appl. Genet. 104, 1107–1113.

Stam, P (1993) Construction of integrated linkage maps by means of a new computer package: JOINMAP. The Plant Journal 5, 739–744.

Stuber CW, Goodman MM, Moll RH (1972) Improvement of yield and ear number resulting from selection of allozyme loci in a maize population. Crop Sci. 22, 737–740.

Swiecicki WK, Wolko B, Weeden NF (2000) Mendel's genetics, the Pisum genome and pea breeding. Vortr. Pflanzenzüchtg 48, 65–76.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The Double-Strand-Break Repair Model for Recombination. Cell 33, 25–35.

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beckbunn T, (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium. Theor. Appl. Genet. 92:213–224.

Tao QZ, Chang YL. Wang JZ, Chen HM, Islam-Faridi MN, Scheuring C. Wang B, Stelly DM, Zhang HB (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158, 1711–1724.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarfs polymorphisms associate with variation in flowering time. Nat. Genet. 28, 286–289.

Vos P, Hogers R, Bleeker M, Reijans M, van der Lee T, Fornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414

Wang GL, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136:1421–1434.

Weber D, Helentjaris T (1989) Mapping RFLP loci in maize using B-A translocations. Genetics 121, 583–590.

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18, 7213–7218.

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. PNAS 99, 12959–12962.

Williams JGK, Kubelik AR, Livak, KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531–6535.

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2 backcross, doubled haploid and recombinant inbred populations in rice (Oryza sativa L). Mol Gen Genet 253, 535–545.

Yamamoto T, Kimura T, Shoda M, Imai T, Saito T, Sawamura Y, Kotobuki K, Hayashi K, Matsuta N (2002) Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. Theor. Appl. Genet. 106, 9–18.

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backross breeding. Theor. Appl. Genet. 77, 353–359.

Yu J et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296, 79–92.

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nature Rev. 2, 983–989.

Zhong XB, Bodeau J, Fransz PF, Williamson VM, van Kammen A, de Jong JH, Zabel P (1999) FISH to meiotic pachytene chromosomes of tomato locates the root-knot nematode resistance gene Mi-1 and the acid phosphatase gene Aps-1 near the junction of euchromatin and pericentromeric heterochromatin of chromosome arms 6S and 6L, respectively. Theor. Appl. Genet. 98, 365–370.