# Genetic mapping in a full-sib family of apple

# Genetische kartering in een nakomelingschap van een appel-kruising

Promotor: dr. ir. P. Stam
hoogleraar in de plantenveredeling, in het bijzonder
de selectiemethoden en duurzame resistentie

Co-promotor: dr. ir. J.W. van Ooijen
onderzoeker statistische genetica
Centrum voor Biometrie
Plant Research International

**Chris Maliepaard**

# Genetic mapping in a full-sib family of apple

**Proefschrift**

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
dr. C.M. Karssen,
in het openbaar te verdedigen
op 25 april 2000
des namiddags te 16.00 uur in de Aula

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG
Maliepaard, C.A.

Genetic mapping in a full-sib family of apple / Chris Maliepaard - [S.l.: s.n.]
Thesis Wageningen University. - With ref. - With summary in Dutch.
ISBN: 90-73384-58-3
NUGI: 835
Subject headings: quantitative genetics, genetic mapping, molecular markers, apple breeding, *Malus*

**Abstract**

**Maliepaard, C.A.,** 2000. Genetic mapping in a full-sib family of apple. Doctoral thesis, Wageningen University.

This doctoral thesis describes genetic analysis using molecular markers in a progeny of a cross between two apple cultivars. Theoretical aspects of linkage analysis in a full-sib family of an outcrossing species were investigated. A molecular linkage map of apple was constructed, using isozyme and DNA markers. Map positions were estimated for genes for scab resistance, resistance to rosy leaf curling aphid, self incompatibility, and fruit acidity. A genetic analysis was carried out for different fruit quality characters, and quantitative trait loci for these characters were mapped. Frequentist and Bayesian statistical methods for the analysis of quantitative trait loci in a full-sib family of an outbreeding species were compared, using the fruit firmness data of the apple progeny. Prospects for the use of molecular markers in apple breeding are discussed.

Cover design: Joost van Langen

*‘And all was for an appil,*
*an appil that he tok,*
*as clerkes finden*
*wreten in here book’*

*(Adam lay ibounden,*
*Anoniem, 15e eeuw)*

*Aan mijn ouders*

## Contents

# Chapter 1

# Introduction

## 1.1 Molecular markers in plant breeding

In current agricultural production, farmers have to cope with high demands for quality and availability of food products, industrial and ornamental crops. At the same time production costs must be continuously decreased in order to survive in a competitive economy where farmers' subsidies are no longer accepted practice. In addition to this, farmers are changing their production methods to respond to environmental and health issues and to the invariably stricter regulations in these fields. This requires not only the improvement of cultivation practices, but also the availability of improved and adapted cultivars.

For plant breeders this means increased demands for cultivars which have good production, good quality, good characteristics for storage and transportation, and higher levels of resistance to pests and diseases. Moreover, the dynamics of consumer preference, trade and industry, and governmental regulations require that improved cultivars are developed much faster and in anticipation of future needs. In order to meet these demands, the plant breeder uses his knowledge of the genetic make-up of the crop and his ability to use this knowledge to combine desired characteristics into new cultivars.

Relatively new tools in the plant breeder's toolbox are molecular markers. These are differences at the chromosomal level made visible, often as bands on a photographic film, by molecular techniques using enzymes or DNA molecules. They enable us to follow the chromosomal segments which are passed on from one generation to the next. Molecular markers have a simple Mendelian inheritance, which makes them ideal for studying genetics. Markers enable the identification of carriers of recessive alleles of genes, which otherwise could only be identified from the segregation in the progeny. Molecular markers can be used to verify the identity of cultivars, to distinguish and compare cultivars, but also to study segregation and linkage of genes and the origin and inheritance of heritable traits. For example, it is possible to trace back resistance genes to the wild progenitors from which they originated.

In breeding programmes markers can be used to increase the chance to select those individual plants from a population or a progeny of a cross that have the best possible combination of desired properties. Since for the determination of markers usually only a small amount of DNA is needed, which can be extracted from any plant tissue, markers offer the possibility to select for a character at a stage long before it is expressed in the plant. For example, fruit acidity in apple can only be

measured when the apple trees bear fruit, which means that selection of undesired trees has to wait until they are five or six years old. With molecular markers, fruit acidity can be selected for in the seedling stage and consequently save five years of cultivation costs of the eliminated trees. Another great advantage of molecular markers in selection is that they are not influenced by environmental noise to the trait and that they can be used even if the trait itself cannot be measured. For example, if markers are used to select for resistance to an insect pest, the marker test can be applied even when there are no insects available for a trial, or in a year and region when and where natural infection does not occur or is so high that differences between genotypes are obscured, or where the distribution of insects over the field is so heterogeneous that no reliable estimates of resistance can be made.

Another asset is the fact that marker tests can be applied at any time during the year. In field tests usually many traits have to be evaluated during the same period and there may be too many traits to evaluate in a single season. Markers offer the possibility to evaluate some of these characters in a different period. A breeder then can choose to evaluate some in the field during summer, while others can be evaluated in the laboratory during the winter before or after. This may increase the efficiency and speed up breeding.

Summarising, molecular markers can be used to predict the performance of individuals for certain traits, in an early stage and without much of the disadvantages attached to the evaluation of the traits themselves. However, before markers can be used for this purpose, the association between a trait and a set of markers has to be established, and this is done by performing linkage analysis, the construction of a linkage map and subsequent genetic analysis of the trait of interest.

## 1.2 Linkage analysis

Molecular markers can be used to construct genetic linkage maps. A linkage map is a representation of the genome, *i.e.* the set of chromosomes, of a biological species. In a diploid species every individual carries two genomes, one inherited from the mother, the other from the father. Corresponding chromosomes from these two genomes are called homologous chromosomes and each member of such a pair is called a homolog. Molecular markers correspond to physical locations of DNA on these chromosomes, *i.e.* the marker loci, and they allow the detection of differences between the homologs at the marker loci. These differences are called the marker alleles, or the allelic variants of the marker.

At meiosis, the process of dividing the nucleus with the genetic material (*i.e.* the DNA) when the reproductive cells are formed, the homologous chromosomes are paired. The DNA is duplicated, exchanged between homologs and subsequently divided over the reproductive cells. The exchange of genetic material and the division over the reproductive cells, giving rise to new combinations of marker alleles and/or genes, is called recombination. An individual in a progeny of a cross which

shows such a new combination for a marker pair, not present in either parent, is called a recombinant, and the term applies to this particular marker pair; for other marker pairs the individual may be non-recombinant.

The pattern of recombination can be reconstructed by observing the frequencies of marker alleles and the pairwise recombination frequencies over the individuals from a segregating progeny of a cross. From these recombination frequencies it can be inferred which markers are inherited together and which are inherited independently. Markers and/or genes which are inherited together are linked. A set of linked markers and/or genes is called a linkage group. The positions of marker loci and their relative distances on a linkage group can be estimated from the pairwise frequencies of the marker alleles. The division of markers over linkage groups and their estimated order and distances are all indicated in the genetic linkage map. In the ideal situation a linkage map consists of linkage groups which each individually correspond to the individual chromosome pairs, so that the number of linkage groups is equal to the number of chromosome pairs.

Although the linkage groups may correspond to the chromosomes, map distances on linkage maps do not represent accurately the physical distances on the chromosomes, since the amount of recombination between two loci does not necessarily correspond to the amount of DNA. Generally, there is a good agreement within a species between the map distances and the sizes of the chromosomes, although there may be so-called 'recombination hot spots', regions on a pair of homologous chromosomes where the amount of recombination is much greater than elsewhere in the genome. However, there are very large differences between species and the same map distance on linkage maps of different species may correspond to very different physical distances. This can be illustrated by the more than twenty-fold difference in amount of DNA per map unit between barley and *Arabidopsis thaliana*. (Nodari *et al.* 1993).

Linkage analysis and linkage maps are valuable for obtaining knowledge about the segregation and the genetic constitution of a species, or may be used for comparison within and across different species. When gene sequences or parts of gene sequences with a known biological function are used in mapping studies, this may reveal some pattern, for example clustering of functionally related genes. The linkage map is also a requirement as a starting point for positional cloning of genes. Within the context of this thesis, the most important application of linkage analysis is the detection and resolution of genetic factors which influence traits of interest, and finding the markers most strongly associated with these traits, so that these can be used for indirect selection.

## 1.3 Genetic analysis of monogenic and quantitative traits

The analysis of associations of markers and traits using molecular markers and linkage maps has been proved valuable, especially where monogenic traits are

concerned. If a trait is regulated by a single gene and if the level of environmental noise is not too high, the genotype may be directly observed from the phenotype, the outward manifestation of the genotype. In that case, there is a qualitative difference between the genotypes. Therefore, monogenic traits are sometimes also referred to as qualitative traits, although it may well be possible that a monogenic trait is influenced so much by environmental factors that the different genotypes cannot be distinguished on the basis of the phenotype. *Vice versa*, it may also be possible that traits regulated by two or even more genes are so well defined that all the genotype classes can be distinguished.

Whereas qualitative traits are often regulated by a single gene, quantitative traits are usually characterised by a more complex inheritance, *i.e.* of multiple genetic components, the polygenes or quantitative trait loci (QTLs). These cannot individually be distinguished by the phenotype and usually each contribute a small effect to the trait. Furthermore, the individual QTL effects are obscured by variation due to environmental conditions and by possible interactions among these individual QTLs and of the QTLs with the environment. Although the terms ‘polygenes’ and ‘QTLs’ have been used as synonyms, the two terms are sometimes distinguished, especially in the literature on animal breeding. There, the term ‘polygenes’ is used to denote only the small effect loci influencing a quantitative trait, as opposed to a QTL, which then indicates a major gene. In this thesis the terms are used as synonyms.

Markers linked to QTLs can be used to detect and characterise these QTLs. As early as 1923 Sax reported the association of a morphological marker, seed-coat pigmentation in beans, with a quantitative trait, seed size, and understood that this was caused by linkage of the single pigmentation gene with one or more genes controlling seed size (Sax 1923). Thoday (1961) suggested to employ such associations to characterise and map polygenes. The idea was that segregation of a single gene marker could be used to detect a linked polygene and to estimate its effect. If single gene markers were scattered through the genome, it should be possible to map and characterise all of the polygenes affecting a quantitative trait. However, there were only limited numbers of morphological markers available, and these often had detrimental effects on the phenotype. Therefore the practical use of morphological markers was very limited. Only with the advent of molecular markers, mostly since the late 1970s, large numbers of markers became available which were phenotypically neutral and which could be used for this purpose. The detection of quantitative trait loci (QTLs) with molecular markers, their localisation on a linkage map, and the estimation of the effects of individual QTLs is what we now denote as QTL analysis or QTL mapping.

Methods for QTL analysis with markers have received considerable attention since the pioneering work of Soller *et al.* (1976), Weller (1986), and especially since Lander and Botstein (1989) presented their method of interval mapping, of which an application in an interspecific cross of tomato had then already been published

(Paterson *et al.* 1988). In interval mapping a marker linkage map is used for inferring the map positions and effects of QTLs affecting a quantitative trait of interest. At regular distances in the genome, *i.e.* at certain positions in an interval between two flanking markers, one computes the likelihood of a possible segregating QTL. This likelihood is compared to the likelihood under the absence of a QTL. The odds ratio of the two likelihoods indicates the strength of evidence for the presence of a QTL. Usually the base 10 logarithm of the odds ratio is calculated for testing and for presentation purposes: this is called the LOD (Log of ODds) score; if the LOD score exceeds a pre-defined significance threshold, the presence of a QTL is inferred.

There are different methods for QTL analysis. The method mentioned above is a Maximum Likelihood based method of interval mapping. Also regression techniques can be employed in interval mapping (Haley and Knott 1992). For an overview of the development of QTL mapping methods, see Jansen (1994b). Recently, the availability of faster computers has allowed and stimulated the use of Bayesian methods for QTL mapping (Hoeschele and VanRaden 1993a, 1993b). In Bayesian methods a full probability model is constructed from the available data and all known and unknown parameters and the relationships among these. Unknown parameters are explicitly quantified by prior distributions, often quite 'diffuse' distributions within realistic ranges. Bayesian statistical inferences about unknown parameters are made by calculation and interpretation of the posterior distribution, the conditional probability distribution of the unobserved parameters, given the data. However, this requires complicated integration and in problems such as QTL analysis, this can not be done analytically. Instead, the conditional probability is approximated by drawing samples from the joint posterior density using computationally demanding Markov chain Monte Carlo techniques. In this way the marginal posterior density for a parameter of interest can be obtained, which takes into account uncertainty in all the other parameters.

Advantages of Bayesian methods for QTL analysis may be that (1) the number of QTLs on a chromosome can be modelled explicitly, (2) all available prior information can be used, (3) missing data can be dealt with by sampling, (4) in providing credible intervals for the position of a QTL (or for any other parameter of interest), uncertainties in other parameters are taken into account (while in other methods the uncertainties in other parameters are ignored).

## 1.4 Applications of QTL analysis

QTL analysis is being used in many plant and animal populations, including human pedigrees. Efficient designs for QTL mapping are segregating populations derived from an $F_1$ of a cross between two inbred lines which are fixed (homozygous) for alternative QTL and marker alleles. The advantages are that the $F_1$ is completely heterozygous for all markers and QTLs, that there is maximum linkage disequilibrium in the $F_1$ so that association between a QTL and a marker, *e.g.* in the $F_2$, can be

attributed almost entirely to linkage. Another advantage is that there are only two marker alleles present in the segregating population and for each of these the parental origin is unambiguous, in other words, the linkage phase of the markers is known. For these reasons, most of the early QTL mapping studies were in $F_2$ or backcross populations. However, for many crops and most animals inbreeding is impossible or can be done only for a limited number of generations, sometimes with large detrimental effects on the phenotype.

If segregating populations from inbred lines are at one end of the spectre, where straightforwardness in QTL mapping is concerned, then at the other end is a population structure of outbred progenies with small families per cross. In that case, QTL mapping may need to be done in multiple small full-sib families, or in a large half-sib family. This has considerable disadvantages: varying numbers of marker and QTL alleles may be segregating, some families may be uninformative with respect to a marker or a QTL, linkage phases between markers or between a marker and a QTL may vary (and cannot always be estimated very well in small families), and between-family effects on the phenotype may need to be taken into account. In half-sib families half of the contributions in terms of marker or QTL alleles is unknown. In humans, QTL analysis has to rely on pooled data over multiple families or on pedigree information over a number of generations.

Some of these disadvantages no longer apply if a large full-sib family can be produced from a single cross. Obviously in a single family no between-family effects have to be accounted for, and usually both parents and the full progeny can be genotyped. Only up to four marker and QTL alleles may segregate at a locus, and although the linkage phases between marker alleles may be unknown *a priori*, these can be estimated very well if the markers are not too widely spaced. Still, some other disadvantages have to be considered: QTLs which are homozygous in both parents, do not segregate, although they may add to the difference between the parents if they are homozygous for different alleles. Both markers and QTLs may have different segregation types, since one or both parents can be heterozygous and if they are both heterozygous they may have zero, one or two alleles in common. QTL analysis in apple has to deal with these aspects.

## 1.5 Apple

Cultivated apple (*Malus pumila* Mill. or *Malus* × *domestica* Borkh.) is one of a large number of apple species. It can be crossed easily with wild apple species. The name apple applies to the species of the genus *Malus* in the subfamily *Maloideae* of the Rosaceae. With regard to the evolutionary origin of the *Maloideae*, which also comprises pear, hawthorn and quince, many questions are still unanswered. It seems clear from the base chromosome number (x=17), which is high in comparison with other Rosaceous species (x=7, x=8 or x=9), that polyploidy must have occurred, but it has not yet been established whether this is autopolyploidy or allopolyploidy

(Phipps *et al.* 1991; Morgan *et al.* 1994). Although apple has a large genome in terms of the chromosome number, in terms of DNA content the genome is relatively small, as is the case for other Rosaceous species. Dickson *et al.* (1992) reported amounts of DNA per nucleus in the range of 1.6 pg/2C for diploid apple cultivars. Tomato (*Lycopersicon esculentum*) has about 2.0 pg/2C and has only 12 chromosomes per haploid genome (Arumuganathan and Earle 1991).

Despite its likely polyploid origin, apple has been shown to have a diploid segregation behaviour at meiosis. There are examples of diploid segregation of monogenic traits, *e.g.* of several resistance genes (Brown 1992). Lespinasse (1973) observed only bivalent pairing in diploid material studied. Chevreau *et al.* (1985) and Chevreau and Laurens (1987) showed disomic inheritance for pollen and leaf isozymes, confirming diploid segregation. Additionally, they found that several isozymes showed bigenic inheritance, indicating duplications within the apple genome.

Apple is an outbred species. It has a gametophytic self-incompatibility system, which prevents fertilisation by pollen carrying a self-incompatibility allele if the allele is also present in the tree being pollinated. Apple has a long juvenile period (five to six years from seed to flowering tree) but can be clonally propagated very easily by grafting or bud-grafting. It is possible to obtain a large full-sib family from a single cross. Since apple is a tree species, traits can be evaluated during a number of successive years on a single individual.

Typically, in apple breeding two cultivars or selections are crossed to produce a large progeny, in which selection is carried out for disease and pest resistances, tree habit, fruit production and fruit quality during a number of successive years. The two parents in a cross usually are both heterozygous at many loci so that the direct progeny, an $F_1$ consisting of full-sibs, will segregate for a large number of markers. Still, the number of different allelic variants at a locus may be small in modern breeding material. It has to be taken into account that in many breeding programmes all over the world the same founder cultivars have been used extensively, so that the same alleles can be expected to be present in many modern cultivars and selections.

Genetics of tree species, and this applies equally well to apple, is not a very popular topic of study among scientists, and for good reason too: results can only be obtained with great patience, due to the long generation cycle. In addition, this long generation cycle and the amount of land, labour and time required to grow and evaluate populations involve high costs. Furthermore, the large genome and the impossibility to produce inbred lines make matters even worse. Although apple is one of the major fruit crops in the non-tropical areas, its economical importance is relatively small in comparison with other crops and does not justify high input in breeding effort or genetical research.

Despite the disadvantages of genetic studies of apple, quite a number of single gene traits are known (Brown 1992). In the genus *Malus*, these include genes for resistance to scab (*Venturia inaequalis*), powdery mildew (*Podosphaera leucotricha*) and insect pests, and genes involved in self-incompatibility, tree habit and fruit quality. However, quantitative genetics of apple is still largely virgin territory. The availability of molecular markers has already facilitated genetic analysis of apple considerably, and it may be expected that further application will reveal more of the genetic components involved in fruit quality, fruit production, tree growth and resistances to diseases and pests. Additionally, because of the prospects for early selection of important traits, the use of markers in breeding programmes may be profitable for the very reasons that make apple such a nuisance for genetic analysis.

## 1.6 Outline of the thesis

In the underlying doctoral thesis, linkage analysis and QTL analysis in full-sib families is investigated and applied for apple. Theoretical aspects of linkage analysis in a full-sib family of an outcrossing plant species are addressed in chapter 2. Chapters 3 and 4 describe the mapping of the genes *Vf* and $Sd_1$, involved in resistance to apple scab and rosy leaf curling aphid, respectively. In chapter 5, a genetic linkage map of apple is constructed, using molecular markers. Chapter 6 describes the mapping of QTLs for fruit firmness traits. In chapter 7 these fruit firmness traits are used to make a comparison between two methods for QTL analysis, the maximum likelihood interval mapping method and a Bayesian approach. In the general discussion the application of molecular markers for QTL analysis in full-sib families, and more specifically for apple, and the prospects for marker-aided selection in apple are considered.

# Chapter 2

# Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications

C. Maliepaard, J. Jansen and J.W. van Ooijen[1]

## Abstract

Linkage analysis and map construction using molecular markers is far more complicated in full-sib families of outbreeding plant species than in progenies derived from homozygous parents. Markers may vary in the number of segregating alleles. One or both parents may be heterozygous, markers may be dominant or codominant and usually the linkage phases of marker pairs are unknown. Because of these differences, marker pairs provide different amounts of information for the estimation of recombination frequencies and the linkage phases of the markers in both parents, and usually these have to be estimated simultaneously. In this paper we present a complete overview of all possible configurations of marker pairs segregating in full-sib families. Maximum likelihood estimators for the recombination frequency and LOD score formulas are presented for all cases. Statistical properties of the estimators are studied analytically and by simulation. Specific problems of dominant markers, in particular with respect to the probability of detecting linkage, the probability of obtaining zero estimates, and the ability to distinguish linkage phase combinations, and consequences for mapping studies in outbred progenies are discussed.

## 2.1 Introduction

The application of molecular markers has become a major tool in genetic analysis. Genetic maps are available for a large number of plant and animal species and an increasing number of genes is being detected with the aid of these maps. Various types of markers are used: isozyme markers, restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs), (sequence tagged) microsatellites, etc. Apart from the techniques, these marker types differ in several respects: number of

[1] Genetical Research (1997) 70: 237-250.

loci which can be detected, degree of polymorphism within and between accessions and dominance characteristics. Usually, the choice of a particular marker type is based upon these aspects, the convenience of application and, not unimportantly, its costs.

Until recently, linkage analysis with molecular markers in plants has been applied mainly to populations derived from the $F_1$ of a cross between two fully homozygous diploid parents, *i.e.*, $BC_1$, $F_2$, RILs and doubled haploids. One of the reasons is that most important agricultural crops are either self-fertilising species or that inbreeding can be carried out without severe inbreeding depression. Another reason is that linkage analysis is more or less straightforward, while introgression of mapped genes can be done simply by repeated backcrossing.

The differences for linkage analysis between a progeny of a cross derived from inbred lines and a full-sib family of an outbreeding species are due to the number of segregating alleles per locus per parent and the linkage phase of the loci. Segregating populations like $BC_1$, $F_2$ or a set of RILs (in this paper $BC_1$, $F_2$ or RILs are considered to be derived from fully homozygous parents) are based on two non-identical inbred lines. Hence, all segregating loci will segregate for only two alleles, and all alleles from the same parent are in coupling phase in the $F_1$. Contrarily, a cross between two non-identical plants of an outbreeder may segregate for up to four alleles per locus, and this may vary between loci, while the linkage phases usually are unknown.

These differences complicate linkage analysis in a full-sib family (in this paper a full-sib family, or FS-family, is considered to be the progeny of a cross between two non-inbred plants of an outbreeding diploid species). There are a number of ways to circumvent these complications and enable the genetic analysis in outbreeders. The most straightforward is the so-called double (or two-way) pseudo-testcross, in which linkage analysis is carried out for each parent separately (Hemmat *et al.* 1994; Grattapaglia and Sederoff 1994; Grattapaglia *et al.* 1995). However, for crosses in which important alleles segregate in both parents, the integration of the individual parental linkage analyses remains problematic. Another method is to create a backcross progeny in order to simplify the segregation, resembling the $BC_1$ except for linkage phases, which may be unknown. For crop species with a long juvenile period such as tree species, tulip and lily, this is not a practical solution. Also, incompatibility may block this possibility, or otherwise cause severe selection or inbreeding depression in the progeny.

Linkage analysis using molecular markers in crosses with outbreeders is treated in a number of papers (Ritter *et al*. 1990; Arús *et al*. 1994; Ritter and Salamini 1996). The latter paper presents formulas useful for the estimation of recombination frequencies in nearly all situations. In some cases the formulas represent the actual estimators, whereas in others the formulas are likelihood equations which have to be implemented in numerical maximisation methods such

as Newton-Raphson. Unfortunately, two particular configurations were not treated in that paper, although with respect to the estimation, one of these is equivalent to another configuration mentioned. In this paper we present, from a genetic perspective, an overview of the whole range of situations of molecular markers in crosses with outbreeding species. Subsequently, we derive an estimator of the recombination frequency by applying an EM-algorithm to an example configuration. We do this without going into technical detail but completely by explaining the derivation in genetic terms, thereby making the EM-algorithm appear very natural. From this example we generalise the derivation to come to a new, general formula for the estimation of the recombination frequency applicable to all configurations. Using a few comprehensive tables we give a complete overview of the explicit or iterative estimators that were obtained by elaboration of the general formula. These can be implemented easily, even in a computer spreadsheet. A procedure for determining the linkage phases of the parents based upon the progeny is presented. In addition, the quality of the information obtained in the various situations is studied, both analytically and by simulation, and translated into consequences for the application of certain types of markers in linkage analysis for outbreeding species. Finally, we present a new and simple approximation to a confidence interval for a recombination frequency estimate that can be applied to all configurations.

## 2.2 Characteristics of the segregation of markers in FS-families

In the two diploid parents of an FS-family up to four different alleles may be present at a single locus; the number of alleles may vary over loci. For all molecular marker types the alleles are usually recognised as fragments with distinct molecular weights. In certain cases a marker detects one or more fragments in some genotypes, whereas it fails to detect a fragment in other genotypes. (Remark: In our terminology a marker is related to a locus, rather than to a single molecular fragment.) The allele corresponding to the absence of a fragment can be called a null-allele. Null-alleles in the parents of an FS-family lead to dominance, *i.e.*, two particular genotypes cannot be distinguished by phenotype. The so-called *segregation type* of a locus, *e.g.*, $ab \times cd$, describes the alleles present in the parents of an FS-family and hence the possible progeny genotypes: the two characters left of the '$\times$' represent the alleles of the first parent, the two characters on the right represent those of the second; each distinct allele is symbolised by a different character, and a null-allele with a '*0*'. Obviously, only segregation types where at least one of the parents is heterozygous are considered for linkage analysis.

In linkage analysis essentially one tries to detect recombination events between loci in both parental meioses. This can be done by reconstructing, for each homologue (or haplotype) of every individual in the offspring, which of the two

homologues of one parent contributed to its genotype: a recombination event has occurred if an allele at a certain locus is from one homologue of a parent and the allele at the next locus from the other. This reconstruction uses the phenotypes of the offspring, the parental phenotypes and possibly the grandparental phenotypes. In an FS-family it has to, or can, be done for both parents. If four distinct phenotypes are present in the offspring, the haplotypes which formed these phenotypes can be reconstructed completely, *i.e.*, for each parent the contributed haplotype is clear. This is the case for loci with four alleles (*ab* × *cd*, one of the four may be *'0'*). The segregation types with three non-null-alleles and heterozygous in both parents (*ab* × *ac*), or two null-alleles plus two other alleles and heterozygous in both parents (*a0* × *b0*), also allow the complete reconstruction. Therefore, these types are equivalent to the four allele type (*ab* × *cd*). For loci heterozygous in only one parent (two alleles: *ab* × *aa*, *aa* × *ab*, *'a'* may be *'0'*, *e.g.*, most RAPD markers; three alleles: *bc* × *aa*, *aa* × *bc*, one of the three may be *'0'*) the reconstruction can be done for one parent only; these three-allele types are equivalent to the respective two-allele segregation types. Of course, the configuration *a0* × *aa* does not segregate phenotypically and is not considered. For all other situations the reconstruction can be done only partially. Loci with two alleles and heterozygous in both parents (*ab* × *ab*) have two possible parental haplotype combinations for the heterozygous offspring *'ab':* the *'a'*-allele may have been derived from either parent and the *'b'* from the alternative. There are even more options for the dominant phenotype *'a-'* for the segregation type *a0* × *a0:* the genotype is either *'aa'* or *'a0'*, and in the latter case the *'a'*-allele may stem from either parent. This is the typical situation of a RAPD fragment present in both parents and segregating with an expected 3:1 ratio. Finally, the situation with three alleles in which the third allele is a null-allele and heterozygous in both parents (*ab* × *a0*, *a0* × *ab*), leaves open for the reconstruction two possibilities for the phenotype *'a-'* in the offspring: *'aa'* or *'a0'*, whereas for the phenotype *'ab'* in the offspring the reconstruction is complete. This situation may occur, for instance, when one of the molecular fragments of a three-allelic RFLP marker is too small to be detected.

Summarising, in an FS-family there are seven essentially distinct segregation types providing recombination information: (1) two alleles, one parent heterozygous (*ab* × *aa*), or (2) the other parent heterozygous (*aa* × *ab*), (3) two alleles, both parents heterozygous (*ab* × *ab*), (4) four alleles (*ab* × *cd*), (5) two alleles, of which one null-allele, both parents heterozygous (*a0* × *a0*), (6) three alleles, of which one null-allele (in one copy), two parents heterozygous, the null-allele in the one parent (*ab* × *a0*), or (7) in the other (*a0* × *ab*). The nine basic configurations of Ritter and Salamini (1996) correspond to these seven segregation types, since four of their configurations (A1A0×A2A0, A1A2×A3A0, A1A2×A1A3 and A1A2×A3A4) all have the same segregation type: *ab* × *cd*, while *ab* × *aa* and *aa* × *ab* are considered

**Table 2.1 Configuration numbers of all pairwise combinations of segregation types**

| Locus 1 | Locus 2 | | | | | | |
|---|---|---|---|---|---|---|---|
| | *ab* × *aa* | *aa* × *ab* | *ab* × *ab* | *ab* × *cd* | *a0* × *a0* | *ab* × *a0* | *a0* × *ab* |
| *ab* × *aa* | 1 | * | 2 | 3 | 4 | 5 | 6 |
| *aa* × *ab* | | (1) | (2) | (3) | (4) | (6) | (5) |
| *ab* × *ab* | | | 7 | 8 | 9 | 10 | (10) |
| *ab* × *cd* | | | | 11 | 12 | 13 | (13) |
| *a0* × *a0* | | | | | 14 | 15 | (15) |
| *ab* × *a0* | | | | | | 16 | 17 |
| *a0* × *ab* | | | | | | | (16) |

When no number is given the configuration is equivalent to the configuration with the loci exchanged.
When the number is given in parentheses, the configuration is equivalent to its reciprocal cross
* There is no information on recombination available

equivalent, as are *ab* × *a0* and *a0* × *ab*. The seven segregation types lead to a total of seventeen different combinations of loci (Table 2.1 and Table 2.2), some of which have been well studied, such as the $BC_1$ type of segregation (No. 1) or the $F_2$ type of segregation with codominant or dominant markers (Nos. 7, 9, 14). The exchange of either the loci or the parents leads to an equivalent situation.

A complicating factor in linkage analysis in crosses with outbreeders is that the linkage phase of the markers will often be unknown *a priori*, while knowledge of the phase is required for the detection of the recombination events. The linkage phase defines the configuration of alleles of a pair of heterozygous loci over the homologous chromosomes in a single parent. It has to be stressed that linkage phase is concerned with the allelic configuration, rather than the loci as such. Additionally, coupling of an allele at locus 1 with a certain allele at locus 2 also means repulsion with the other allele at locus 2. An important distinction from the standard segregating populations with inbred lines is that the linkage phases can be different for the two parents. Also the linkage phase can be undefined in one of the parents due to homozygosity, as in a $BC_1$. Hence, in an FS-family, we end up with the following linkage phase combinations: coupling (*c*) in the first parent ($P_1$) and undefined in the second parent ($P_2$), or *vice versa*, repulsion (*r*) in $P_1$ and undefined in $P_2$, or *vice versa*, coupling in both parents (*c* × *c*), repulsion in both parents (*r* × *r*), and coupling in $P_1$ and repulsion in $P_2$ (*c* × *r*), or *vice versa* (*r* × *c*).

**Table 2.2 Definitions of marker phenotype indicators**

| | | | | Phenotype indicator (*f*) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nr[a] | Locus | $P_1$[b] | $P_2$[b] | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 1 | L1: | *ab* | *aa* | *aa* | *aa* | *ab* | *ab* | | | | | | | | | | | | |
| | L2: | *ab* | *aa* | *aa* | *ab* | *aa* | *ab* | | | | | | | | | | | | |
| 2 | L1: | *ab* | *aa* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | | | | | | | | | | |
| | L2: | *ab* | *ab* | *aa* | *ab* | *bb* | *aa* | *ab* | *bb* | | | | | | | | | | |
| 3 | L1: | *ab* | *aa* | *aa* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | *ab* | | | | | | | | |
| | L2: | *ab* | *cd* | *ac* | *ad* | *bc* | *bd* | *ac* | *ad* | *bc* | *bd* | | | | | | | | |
| 4 | L1: | *ab* | *aa* | *aa* | *aa* | *ab* | *ab* | | | | | | | | | | | | |
| | L2: | *a0* | *a0* | *a-* | *00* | *a-* | *00* | | | | | | | | | | | | |
| 5 | L1: | *ab* | *aa* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | | | | | | | | | | |
| | L2: | *ab* | *a0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | | | | | | | |
| 6 | L1: | *ab* | *aa* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | | | | | | | | | | |
| | L2: | *a0* | *ab* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | | | | | | | |
| 7 | L1: | *ab* | *ab* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | *bb* | *bb* | *bb* | | | | | | | |
| | L2: | *ab* | *ab* | *aa* | *ab* | *bb* | *aa* | *ab* | *bb* | *aa* | *ab* | *bb* | | | | | | | |
| 8 | L1: | *ab* | *ab* | *aa* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | *ab* | *bb* | *bb* | *bb* | *bb* | | | | |
| | L2: | *ab* | *cd* | *ac* | *ad* | *bc* | *bd* | *ac* | *ad* | *bc* | *bd* | *ac* | *ad* | *bc* | *bd* | | | | |
| 9 | L1: | *ab* | *ab* | *aa* | *aa* | *ab* | *ab* | *bb* | *bb* | | | | | | | | | | |
| | L2: | *a0* | *a0* | *a-* | *00* | *a-* | *00* | *a-* | *00* | | | | | | | | | | |
| 10 | L1: | *ab* | *ab* | *aa* | *aa* | *aa* | *ab* | *ab* | *ab* | *bb* | *bb* | *bb* | | | | | | | |
| | L2: | *ab* | *a0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | | | | |
| 11 | L1: | *ab* | *cd* | *ac* | *ac* | *ac* | *ac* | *ad* | *ad* | *ad* | *ad* | *bc* | *bc* | *bc* | *bc* | *bd* | *bd* | *bd* | *bd* |
| | L2: | *ab* | *cd* | *ac* | *ad* | *bc* | *bd* | *ac* | *ad* | *bc* | *bd* | *ac* | *ad* | *bc* | *bd* | *ac* | *ad* | *bc* | *bd* |
| 12 | L1: | *ab* | *cd* | *ac* | *ac* | *ad* | *ad* | *bc* | *bc* | *bd* | *bd* | | | | | | | | |
| | L2: | *a0* | *a0* | *a-* | *00* | *a-* | *00* | *a-* | *00* | *a-* | *00* | | | | | | | | |
| 13 | L1: | *ab* | *cd* | *ac* | *ac* | *ac* | *ad* | *ad* | *ad* | *bc* | *bc* | *bc* | *bd* | *bd* | *bd* | | | | |
| | L2: | *ab* | *a0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | |
| 14 | L1: | *a0* | *a0* | *a-* | *a-* | *00* | *00* | | | | | | | | | | | | |
| | L2: | *a0* | *a0* | *a-* | *00* | *a-* | *00* | | | | | | | | | | | | |
| 15 | L1: | *a0* | *a0* | *a-* | *a-* | *a-* | *00* | *00* | *00* | | | | | | | | | | |
| | L2: | *ab* | *a0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | | | | | | | |
| 16 | L1: | *ab* | *a0* | *a-* | *a-* | *a-* | *ab* | *ab* | *ab* | *b0* | *b0* | *b0* | | | | | | | |
| | L2: | *ab* | *a0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | | | | |
| 17 | L1: | *ab* | *a0* | *a-* | *a-* | *a-* | *ab* | *ab* | *ab* | *b0* | *b0* | *b0* | | | | | | | |
| | L2: | *a0* | *ab* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | *a-* | *ab* | *b0* | | | | | | | |

The (dominant) phenotypes *'a-'* and *'b-'* can be of genotypes *'a0'* or *'aa'* and *'b0'* or *'bb'*, respectively. Reciprocal crosses have identical definitions.

[a] Configuration number according to Table 2.1

[b] The genotypes of the two parents ($P_1 \times P_2$) at the first (L1) and the second (L2) locus.

For example,

$$\frac{a}{b}\ \frac{a}{b} \times \frac{a}{b}\ \frac{b}{a}$$

depicts the $c \times r$ combination for a pair of markers with segregation type $ab \times ab$. A linkage phase combination has to be deduced from the segregation in the FS-family itself or from the grandparental genotypes, although this is not always possible.

## 2.3 Recombination frequency estimators, LOD scores and determination of linkage phases

Mather (1951), Allard (1956) and Weber and Wricke (1994) developed maximum likelihood estimators of the recombination frequency for a variety of genetic situations in BC1 and F2 populations. Ritter *et al.* (1990) developed estimators for most of the genetic situations in crosses between heterozygous parents. Arús et al. (1994) contributed the solution to two additional situations, Ritter and Salamini (1996) nearly completed the set, and here, we add one more estimator (Table 2.1, number 17) and mention a new configuration type (Table 2.1, number 6), so that now all combinations with molecular markers with two to four alleles (without epistasis) in an FS-family are covered, including segregation in one or both parents, dominance, and all linkage phase configurations.

In order to calculate the recombination frequency, one needs to know the number of recombination events in both parental meioses. If one knew the genotypes of the gametes, these could be counted easily. However, the marker genotypes of the gametes cannot always be deduced from the phenotypes of the individuals in the progeny. For example, for two $ab \times ab$ loci in linkage phase combination $c \times c$, nine marker phenotypes can be observed in the progeny (Table 2.2, No. 7). Marker phenotypes 1 and 9 are based on two non-recombinant gametes, phenotypes 2, 4, 6 and 8 each on a non-recombinant and a recombinant gamete, and phenotypes 3 and 7 on two recombinant gametes. So, for progeny individuals with one of these marker phenotypes, the number of recombinant gametes can be counted as $(n_1+n_9)0 + (n_2+n_4+n_6+n_8)1 + (n_3+n_7)2$, where $n_f$ is the number of individuals with marker phenotype $f$. However, for the double heterozygous phenotype (5), there are two possible haplotype combinations that cannot be distinguished: either two recombinant or two non-recombinant gametes. We do know, however, the expected proportion of these two combinations in terms of the recombination frequency $r$, $r^2 : (1-r)^2$. Suppose we knew $r$, then we would know the expected numbers for the two combinations. Using these numbers (and the other $n_f$), one can estimate the recombination frequency. With this new value the expected

numbers for the two combinations can be recalculated, which in turn can be used to estimate a subsequent value of *r*, and so on. This is an iterative procedure, which can be summarised in the following formula:

$$r_{i+1} = \frac{1}{2n}\left( (n_1 + n_9)\cdot 0 + (n_2 + n_4 + n_6 + n_8)\cdot 1 + (n_3 + n_7)\cdot 2 + n_5 \cdot \frac{(1-r_i)^2 \cdot 0 + r_i^2 \cdot 2}{(1-r_i)^2 + r_i^2} \right),$$

$$r_{i+1} = \frac{1}{2n}\left( n_2 + n_4 + n_6 + n_8 + 2\cdot(n_3 + n_7) + 2\cdot n_5 \frac{r_i^2}{(1-r_i)^2 + r_i^2} \right),$$

where $r_i$ is the value of *r* after iteration *i*. Using an initial value for the recombination frequency (*e.g.*, $r_0 = 0.25$), this formula can be iterated until a stable value is reached. Though it may not be obvious here, the previous formula is in fact an ML-estimator of *r* (Dempster *et al.* 1977; Lander and Green 1987).

In the following the above procedure will be formalised in a maximum likelihood context to develop a general formula for the estimators of the recombination frequency in all situations in an FS-family of outbreeders. Any given marker pair will segregate into *F* phenotypes, with $n_1$ to $n_F$ individuals in the *F* phenotype classes adding up to a total of *n* (Table 2.2). We define $p_f$ as the probability of (diploid) phenotype *f*, all $p_f$ are functions of the recombination frequency *r*. Then, the likelihood of the phenotype frequencies in the progeny is:

$$L = \begin{pmatrix} n \\ n_1 \quad \cdots \quad n_F \end{pmatrix} \prod_{f=1}^{F} p_f^{n_f}, \text{ so that: } \ln(\mathrm{L}) = \text{constant} + \sum_{\mathrm{f}=1}^{\mathrm{F}} n_f \ln(p_f).$$

To maximise the log-likelihood for *r* we need to solve the likelihood equation:

$$\frac{\delta \ln(\mathrm{L})}{\delta r} = \sum_{f=1}^{F} \frac{n_f}{p} \frac{\delta p_f}{\delta r} = 0. \qquad (2.1)$$

For configurations 1, 2, 3, 5, 6, 11, 13, 16 and 17 (Table 2.1) this likelihood equation can be solved readily leading to explicit ML-estimators (Table 2.3). For configurations 7 $c \times r$, $r \times c$, and 14 explicit ML-estimators can be derived by substituting $\theta$ for $r^2$, $r(1\text{-}r)$ or $(1\text{-}r)^2$ in the likelihood equation; in the legitimate range of *r* the maximum for $\theta$ will also be the maximum for *r* (Table 2.3). For all remaining configurations, however, the likelihood equations turn into finding zeros of higher-order polynomials, which is difficult. A much easier solution can be obtained by employing the EM-algorithm (Dempster *et al.* 1977). This approach, as used by Lander and Green (1987) for genetic maps in humans, can be used for all configurations.

Underlying a diploid phenotype of a marker pair is a combination of two haplotypes, *i.e.*, the gametes. Often there can be different haplotype combinations that lead to the same diploid phenotype, *e.g.*, think of linkage phase configurations as in the above example, or dominance. Thus, each of the marker phenotype probabilities, $p_f$, can be defined by the probabilities $\pi_{fh}$ of the $H_f$ underlying haplotype combinations:

$$p_f = \sum_{h=1}^{H_f} \pi_{fh}.$$

We can substitute this into

$$\frac{\delta \ln(\mathrm{L})}{\delta r} = \sum_{f=1}^{F} \frac{n_f}{p_f} \sum_{h=1}^{H_f} \frac{\delta \pi_{fh}}{\delta r} = \sum_{f=1}^{F} n_f \sum_{h=1}^{H_f} \frac{\pi_{fh}}{p_f} \frac{\delta \ln(\pi_{fh})}{\delta r} = 0. \qquad (2.2)$$

The probability of a haplotype combination is a simple function of the recombination frequency. A haplotype of two loci is either recombinant or non-recombinant. Recombination can only be observed if there is heterozygosity at both loci in a parent. We define the number of parents heterozygous at both loci to be $\gamma \in \{1,2\}$. Thus, a combination of two haplotypes may consist of zero up to $\gamma$ recombinants. If $\alpha_{fh}$ and $\beta_{fh}$ are the numbers of recombinant and non-recombinant haplotypes underlying the haplotype combination $h_f$, respectively, we obtain the constraint: $\alpha_{fh}+\beta_{fh}=\gamma$, with $\alpha_{fh}, \beta_{fh} \in \{0,1,2\}$. Accordingly, the probability for a haplotype combination is: $r^{\alpha}(1-r)^{\beta}$, multiplied by a constant. For the derivative of $\ln(\pi_{fh})$ to $r$ we obtain:

$$\frac{\delta \ln(\pi_{fh})}{\delta r} = \frac{\alpha_{fh}}{r} - \frac{\beta_{fh}}{1-r} = \frac{\alpha_{fh} - \gamma r}{r(1-r)}. \qquad (2.3)$$

Combining 2.2 and 2.3 gives:

$$\frac{\delta \ln(\mathrm{L})}{\delta r} = \frac{1}{r(1-r)} \sum_{f=1}^{F} n_f \sum_{h=1}^{H_f} \frac{\pi_{fh}}{p_f}\left(\alpha_{fh} - \gamma r\right) = 0. \qquad (2.4)$$

Now, since both $p_f$ and $\pi_{fh}$ are functions of $r$, solving this equation is hard, unless we employ the EM-algorithm. Suppose we know all ratios $\pi_{fh}/p_f$, *i.e.*, suppose we know the relative proportion of all underlying haplotype combinations for each phenotype (this is the expectation- or E-step), then we can solve (2.4) (this is the maximisation- or M-step):

$$\hat{r} = \frac{1}{\gamma n} \sum_{f=1}^{F} n_f \sum_{h=1}^{H_f} \frac{\alpha_{fh} \pi_{fh}}{p_f}. \qquad (2.5)$$

From this we get an estimate of $r$, and subsequently we can adjust the expectations of the haplotype proportions $\pi_{fh}/p_f$ and get a new estimate of $r$, and so on. This iterative procedure is an EM-algorithm (Dempster *et al.* 1977). In the E-step the ratio $\pi_{fh}/p_f$ is based upon the value of the recombination frequency of the last iteration, $r_i$, while in the first iteration usually 0.25 is a good starting value. Table 2.3 presents this iterative ML-estimator elaborated for the configurations for which an explicit ML-estimator could not be found. For several configurations all phenotypes have just a single underlying haplotype combination, *i.e.*, $H_f$=1 for all $f$, so that always $\pi_{fh}$=$p_f$, and thus (2.5) becomes an explicit estimator. These situations are special cases of (2.5) and result in estimators identical to the direct solutions of equation (2.1). The use of the EM-algorithm can be extended easily to other more complex situations sometimes encountered in practice, such as where a marker is scored as dominant in part of the progeny and as codominant in the remainder; here, the number of phenotype classes in (2.5) is simply extended.

To test whether a pair of markers is linked, *i.e.*, $r < 0.5$, the LOD score can be used as a test statistic. The LOD score is the logarithm to base 10 of the ratio of the likelihood under the estimated recombination frequency ($r$ =r) and the likelihood under the null-hypothesis of unlinked loci ($r$= 0.5): LOD = $\log_{10}( L(r\text{=r}) / L(r\text{=}0.5) )$. A LOD score of 3.0 is commonly used as the threshold for linkage (Morton 1955; Risch 1992). Table 2.3 lists the LOD score formulas for the different configurations.

The use of the estimators of Table 2.3 presumes that the linkage phase combination is known. However, unlike in crosses with inbred lines, this may not be the case in an FS-family. If the linkage phase combination cannot be determined from the grandparents, then the procedure is to apply the estimators for all possible linkage phases and subsequently deduce the actual linkage phase combination. The method and its success vary for the different configurations. The method depends on (a) the heterozygosity at both loci in both parents, (b) whether both loci have symmetric segregation types (*ab* × *ab*, *a0* × *a0*), and (c) whether both loci have a null-allele in the same parent. If the linkage phase combination of a pair cannot be (fully) determined based on this method, then the remaining option is to determine the phases indirectly through combinations with neighbouring loci with more informative segregation types.

Let us first consider the situation where only one of the parents is heterozygous for both markers (configurations 1 to 6). Always, $r_r = 1 - r_c$, with $r_r$ the estimate under repulsion and $r_c$ the estimate under coupling phase. Of course, only the estimate smaller than 0.5 is a legitimate value. If the LOD score is significant, the linkage phase with the legitimate estimate is chosen.

Next, consider the situations where both loci have a symmetric segregation type (conf. 7, 9, 14). Here, the $c \times r$ and the $r \times c$ estimators are identical, so that the choice between $c \times r$ and $r \times c$ can not be resolved; also $r_{c \times c} = 1 - r_{r \times r}$. If the phases are $c \times c$ or $r \times r$ for configurations 7 and 9, then the $c \times r$ (and $r \times c$) estimate is either imaginary (7) or about 0.5 (9) with a very low LOD score, while the $c \times c$ or $r \times r$ estimate, respectively, is legitimate. If the phases are $c \times r$ or $r \times$ c, then the $c \times c$ or $r \times r$ estimates are about 0.5 with a very low LOD score. Hence, for configurations 7 and 9 the phase combinations $c \times c$ and $r \times r$ can be distinguished from each other and from $c \times r$ or $r \times c$. Configuration 14 is worse, because in addition to being symmetrical, both loci have a null-allele in both parents. Here, if the phases are $c \times c$, then the $c \times r$ (and $r \times c$) estimate is imaginary while the $r \times r$ estimate is larger than 0.5. If, however, the phases are $c \times r$, $r \times c$ or $r \times r$, then all except the $c \times c$ estimate will be legitimate, with identical LOD scores as can be seen from Table 2.3. Hence, for configuration 14 only the phase combination $c \times c$ on the one hand can be distinguished from $c \times r$, $r \times c$ and $r \times r$ on the other, so that other linked markers, with more informative segregation types, are required to resolve the linkage phase combination.

Subsequently, we examine the non-symmetrical situations where both loci have a null-allele in the same parent (conf. 15, 16). Here, always $r_{c \times c} = 1 - r_{r \times r}$ and $r_{c \times r} = 1 - r_{r \times c}$. If the loci are in coupling in the first parent ($c \times c$, $c \times r$), then the estimate for the correct phases has the smallest value and by far the highest LOD score, while the other two estimates are larger than 0.5. If, however, the loci are in repulsion in the first parent ($r \times c$, $r \times r$), then the $r \times c$ and $r \times r$ estimates are approximately equal with similar LOD scores, while the other two estimates are larger than 0.5. Hence, the phase combinations $c \times c$ and $c \times r$ can be distinguished from each other and from $r \times c$ or $r \times r$. Although simulations of $r \times c$ and $r \times r$ phases (of conf. 15 and 16) showed that in more than 95% of the significant cases the correct phase combination was estimated, it would be prudent to verify the linkage phases through neighbouring loci (data not shown). This particular behaviour is caused by the typical characteristic of segregation type *ab* $\times$ *a0:* for the first parent the haplotype contributed to any phenotype in the offspring is always perfectly clear, *'a'* or *'b';* for the second parent this depends on the allele contributed by the first parent: if it is *'b'* then it is clear, but if it is *'a'*, then it cannot be resolved whether the second parent contributed the allele *'a'* or *'0'*. Now, suppose two closely linked loci have the segregation type *ab* $\times$ *a0* (conf. 16). When they are in coupling in the first parent, nearly half of the gametes will have a *'b'* allele on both loci, and thus the contribution of the second parent can be determined. For the rest of the gametes of the first parent there will be an *'a'* allele at one or both of the loci, thus blocking the determination of the contribution of the second parent. When, however, the loci are

**Table 2.3 ML-estimators of the recombination frequency and LOD score formulas**

| No.[a] | Phase[b] | Estimator | LOD score |
|---|---|---|---|
| 1 (1) | $c$ | $(n_2+n_3)/n$ | $(n_1+n_4)\log(2s) + (n_2+n_3)\log(2r)$ |
| | $r$ | exch[c] $(n_2, n_3 ; n_1, n_4)$ | |
| 2 (2) | $c$ | $(n_3+n_4)/(n_1+n_3+n_4+n_6)$ | $(n_1+n_6)\log(2s) + (n_3+n_4)\log(2r)$ |
| | $r$ | exch $(n_3, n_4 ; n_1, n_6)$ | |
| 3 | $c$ | $(n_3+n_4+n_5+n_6)/n$ | $(n_1+n_2+n_7+n_8)\log(2s) + (n_3+n_4+n_5+n_6)\log(2r)$ |
| | $r$ | exch $(n_3, n_4, n_5, n_6 ; n_1, n_2, n_7, n_8)$ | |
| (3) | $c$ | exch $(n_3, n_6 ; n_2, n_7)$ | |
| | $r$ | exch $(n_4, n_5 ; n_1, n_8)$ | |
| 4 (4) | $c$ | $[n_1r/(2-r) + n_2 + 2n_3r/(1+r)] / n$ | $n_1\log(2(1+s)/3) + n_2\log(2r) + n_3\log(2(1+r)/3) + n_4\log(2s)$ |
| | $r$ | exch $(n_1, n_2 ; n_3, n_4)$ | |
| 5 (5) | $c$ | $(n_2+n_3+n_4)/n$ | $(n_1+n_5+n_6)\log(2s) + (n_2+n_3+n_4)\log(2r)$ |
| | $r$ | exch $(n_2, n_3, n_4 ; n_5, n_6, n_1)$ | |
| 6 (6) | $c$ | $(n_3+n_5)/(n_2+n_3+n_5+n_6)$ | $(n_2+n_6)\log(2s) + (n_3+n_5)\log(2r)$ |
| | $r$ | exch $(n_3, n_5 ; n_2, n_6)$ | |
| 7[d] | $c \times c$ | $[n_2+n_4+n_6+n_8 + 2(n_3+n_7) + 2n_5r^2/(1-2rs)] / (2n)$ | $2(n_1+n_9)\log(2s) + (n_2+n_4+n_6+n_8)\log(4rs) + 2(n_3+n_7)\log(2r) + n_5\log(2(1-2rs))$ |
| | $r \times r$ | exch $(n_3, n_7 ; n_1, n_9)$ | |
| | $c \times r$ | $½ - \sqrt{[¼ - (n_1+n_3+n_5+n_7+n_9)/(2n)]}$ | $(n_1+n_3+n_5+n_7+n_9)\log(4rs) + (n_2+n_4+n_6+n_8)\log(2(1-2rs))$ |
| | $r \times c$[e] | exch none | |
| 8 | $c \times c$ | $[n_2+n_3+n_5+n_8+n_{10}+n_{11} + 2(n_4+n_9) + 2(n_6+n_7)r^2/(1-2rs)] / (2n)$ | $2(n_1+n_{12})\log(2s) + (n_2+n_3+n_5+n_8+n_{10}+n_{11})\log(4rs) + 2(n_4+n_9)\log(2r) + (n_6+n_7)\log(2(1-2rs))$ |
| | $c \times r$ | exch $(n_1, n_3, n_5, n_7, n_9, n_{11} ; n_2, n_4, n_6, n_8, n_{10}, n_{12})$ | |
| | $r \times c$ | exch $(n_1, n_2, n_5, n_6, n_9, n_{10} ; n_3, n_4, n_7, n_8, n_{11}, n_{12})$ | |
| | $r \times r$ | exch $(n_1, n_9 ; n_4, n_{12})$ | |
| 9[d] | $c \times c$ | $[2n_1r/(1+r) + 2n_2 + n_3r(1+r)/(1-rs) + n_4 + 2n_5/(2-r)] / (2n)$ | $n_1\log(4(1-r^2)/3) + 2n_2\log(2r) + n_3\log(4(1-rs)/3) + n_4\log(4rs) + n_5\log(4(1-s^2)/3) + 2n_6\log(2s)$ |
| | $r \times r$ | exch $(n_1, n_2 ; n_5, n_6)$ | |
| | $c \times r$ | $[(n_1+n_5)r(1+r)/(1-rs) + n_2+n_6 + 2n_3(1-s^2)/(1+2rs) + 2n_4r^2/(1-2rs)] / (2n)$ | $(n_1+n_5)\log(4(1-rs)/3) + (n_2+n_6)\log(4rs) + n_3\log(2(1+2rs)/3) + n_4\log(2(1-2rs))$ |
| | $r \times c$[e] | exch none | |

| | | | |
|---|---|---|---|
| 10 | $c \times c$ | $[(n_1+2n_4)r + n_2+n_6+n_8 + 2n_3 + 2n_5r^2/(1-2rs) + n_7(1+r)] / (2n)$ | $(n_1+2n_9)\log(2s) + (n_2+n_6+n_8)\log(4rs) + (2n_3+n_7)\log(2r) + n_5\log(2(1-2rs))$ |
| | $c \times r$ | exch ($n_2$, $n_5$, $n_8$ ; $n_3$, $n_6$, $n_9$) | |
| | $r \times c$ | exch ($n_1$, $n_2$, $n_3$, $n_5$ ; $n_7$, $n_9$, $n_8$, $n_6$) | |
| | $r \times r$ | exch ($n_1$, $n_3$ ; $n_7$, $n_9$) | |
| (10) | $c \times c$ | exch none | |
| | $c \times r$ | exch ($n_1$, $n_2$, $n_3$, $n_5$ ; $n_7$, $n_9$, $n_8$, $n_6$) | |
| | $r \times c$ | exch ($n_2$, $n_5$, $n_8$ ; $n_3$, $n_6$, $n_9$) | |
| | $r \times r$ | exch ($n_1$, $n_3$ ; $n_7$, $n_9$) | |
| 11 | $c \times c$ | $[n_2+n_3+n_5+n_8+n_9+n_{12}+n_{14}+n_{15}+ 2(n_4+n_7+n_{10}+n_{13})] / (2n)$ | $2(n_1+n_6+n_{11}+n_{16})\log(2s) + (n_2+n_3+n_5+n_8+n_9+n_{12}+n_{14}+n_{15})\log(4rs) + 2(n_4+n_7+n_{10}+n_{13})\log(2r)$ |
| | $c \times r$ | exch ($n_1$, $n_3$, $n_5$, $n_7$, $n_9$, $n_{11}$, $n_{13}$, $n_{15}$ ; $n_2$, $n_4$, $n_6$, $n_8$, $n_{10}$, $n_{12}$, $n_{14}$, $n_{16}$) | |
| | $r \times c$ | exch ($n_1$, $n_2$, $n_5$, $n_6$, $n_9$, $n_{10}$, $n_{13}$, $n_{14}$ ; $n_3$, $n_4$, $n_7$, $n_8$, $n_{11}$, $n_{12}$, $n_{15}$, $n_{16}$) | |
| | $r \times r$ | exch ($n_1$, $n_6$, $n_{10}$, $n_{13}$ ; $n_4$, $n_7$, $n_{11}$, $n_{16}$) | |
| 12 | $c \times c$ | $[2n_1r/(1+r) + 2n_2 + (n_3+n_5)r(1+r)/(1-rs) + n_4+n_6 + 2n_7/(2-r)] / (2n)$ | $n_1\log(4(1-r^2)/3) + 2n_2\log(2r) + (n_3+n_5)\log(4(1-rs)/3) + (n_4+n_6)\log(4rs) + n_7\log(4(1-s^2)/3) + 2n_8\log(2s)$ |
| | $c \times r$ | exch ($n_1$, $n_2$, $n_5$, $n_6$ ; $n_3$, $n_4$, $n_7$, $n_8$) | |
| | $r \times c$ | exch ($n_1$, $n_2$, $n_3$, $n_4$ ; $n_5$, $n_6$, $n_7$, $n_8$) | |
| | $r \times r$ | exch ($n_1$, $n_2$ ; $n_7$, $n_8$) | |
| 13 | $c \times c$ | $[n_2+n_6+n_7+n_9+n_{10}+n_{11}+2(n_3+n_5)]/ [n_1+n_4+n_7+n_{10}+ 2(n_2+n_3+n_5+n_6+n_8+n_9+n_{11}+n_{12})]$ | $(n_1+n_4+2(n_8+n_{12}))\log(2s) + (n_2+n_6+n_9+n_{11})\log(4rs) + (2(n_3+n_5)+n_7+n_{10})\log(2r)$ |
| | $c \times r$ | exch ($n_2$, $n_5$, $n_8$, $n_{11}$ ; $n_3$, $n_6$, $n_9$, $n_{12}$) | |
| | $r \times c$ | exch ($n_1$, $n_2$, $n_3$, $n_4$, $n_5$, $n_6$ ; $n_7$, $n_8$, $n_9$, $n_{10}$, $n_{11}$, $n_{12}$) | |
| | $r \times r$ | exch ($n_1$, $n_3$, $n_4$, $n_5$ ; $n_{10}$, $n_{12}$, $n_7$, $n_8$) | |
| (13) | $c \times c$ | exch ($n_4$, $n_5$ ; $n_7$, $n_8$) | |
| | $c \times r$ | exch ($n_1$, $n_2$, $n_3$, $n_5$, $n_9$ ; $n_{10}$, $n_8$, $n_6$, $n_{11}$, $n_{12}$) | |
| | $r \times c$ | exch ($n_2$, $n_4$, $n_5$, $n_6$, $n_{11}$ ; $n_3$, $n_7$, $n_9$, $n_8$, $n_{12}$) | |
| | $r \times r$ | exch ($n_1$, $n_3$ ; $n_{10}$, $n_{12}$) | |
| 14 | | $\theta = (n_1-2(n_2+n_3)-n_4)/(2n) + \sqrt{[(n_1-2(n_2+n_3)-n_4)^2/(2n)^2 + 2n_4/n]}$ | $n_1\log(4(2+\theta)/9) + (n_2+n_3)\log(4(1-\theta)/3) + n_4\log(4\theta)$ |
| | $c \times c$ | $1-\sqrt{\theta}$ ($\theta=s^2$) | |
| | $c \times r$ | $\frac{1}{2}-\sqrt{(\frac{1}{4}-\theta)}$ ($\theta=rs$) | |
| | $r \times c^e$ | $\frac{1}{2}-\sqrt{(\frac{1}{4}-\theta)}$ ($\theta=rs$) | |

| | | | |
|---|---|---|---|
| | $r \times r$ | $\sqrt{\theta}$ $(\theta=r^2)$ | |
| 15[d] | $c \times c$ | $[n_1r(3-r)/(2-r) + n_2r(1+r)/(1-rs) + 2n_3/(2-r) + n_4(1+r) + n_5] / (2n)$ | $n_1\log(2(2-r)/3) + n_2\log(4(1-rs)/3) + n_3\log(4(1-s^2)/3) + n_4\log(2r) + n_5\log(4rs) + 2n_6\log(2s)$ |
| | $c \times r$ | exch ($n_2$, $n_5$ ; $n_3$, $n_6$) | |
| (15) | $c \times c$ | exch none | |
| | $r \times c$ | exch ($n_2$, $n_5$ ; $n_3$, $n_6$) | |
| 15 | $r \times c$ | $[n_1r(3+r)/(1+r) + 2n_2r/(1+r) + n_3r(1+r)/(1-rs) + n_4r + 2n_5 + n_6] / (2n)$ | $n_1\log(2(1+r)/3) + n_2\log(4(1-r^2)/3) + n_3\log(4(1-rs)/3) + n_4\log(2s) + 2n_5\log(2r) + n_6\log(4rs)$ |
| | $r \times r$ | exch ($n_2$, $n_5$ ; $n_3$, $n_6$) | |
| (15) | $c \times r$ | exch none | |
| | $r \times r$ | exch ($n_2$, $n_5$ ; $n_3$, $n_6$) | |
| 16[d] | $c \times c$ | $(n_2+n_3+n_4+n_6+n_7+n_8) / [n_1+n_2+n_3+n_4+n_7 + 2(n_5+n_6+n_8+n_9)]$ | $(n_1+2(n_5+n_9))\log(2s) + (n_2+n_3+n_4+n_7)\log(2r) + (n_6+n_8)\log(4rs)$ |
| | $c \times r$ | exch ($n_5$, $n_8$ ; $n_6$, $n_9$) | |
| (16) | $c \times c$ | exch none | |
| | $r \times c$ | exch ($n_5$, $n_8$ ; $n_6$, $n_9$) | |
| 16 | $r \times c$ | $[n_1+n_5+n_9 + 2(n_6+n_8)] / [n_1+n_2+n_3+n_4+n_7 + 2(n_5+n_6+n_8+n_9)]$ | $(n_1+2(n_6+n_8))\log(2r) + (n_2+n_3+n_4+n_7)\log(2s) + (n_5+n_9)\log(4rs)$ |
| | $r \times r$ | exch ($n_5$, $n_8$ ; $n_6$, $n_9$) | |
| (16) | $c \times r$ | exch none | |
| | $r \times r$ | exch ($n_5$, $n_8$ ; $n_6$, $n_9$) | |
| 17 | $c \times c$ | $(n_3+n_6+n_7+n_8 + 2n_5) / [n_2+n_3+n_4+n_7 + 2(n_5+n_6+n_8+n_9)]$ | $(n_2+n_4+2n_9)\log(2s) + (n_3+n_7+2n_5)\log(2r) + (n_6+n_8)\log(4rs)$ |
| | $c \times r$ | exch ($n_4$, $n_5$, $n_6$ ; $n_7$, $n_8$, $n_9$) | |
| | $r \times c$ | exch ($n_2$, $n_5$, $n_8$ ; $n_3$, $n_6$, $n_9$) | |
| | $r \times r$ | exch ($n_2$, $n_4$, $n_5$ ; $n_3$, $n_7$, $n_9$) | |

For several linkage phase configurations the estimator and LOD score can be obtained by exchanging as indicated the phenotype frequencies ($n_f$) in the fully specified preceding formulas. If the estimator contains the recombination frequency itself (as $r$ or $s$), then it is an iterative estimator. $s = 1 - r$, $n$: total number of individuals, log: logarithm to base 10

[a] Configuration number according to Table 2.1

[b] For each configuration two or four linkage phase combinations are distinguished

[c] exch ($n_a$, $n_b$, ...; $n_w$, $n_x$, ...) means: exchange $n_a$, $n_b$, ...with $n_w$, $n_x$, ... respectively

[d] The order of the phase combinations and/or reciprocal crosses is changed for convenience

[e] $r \times c$ cannot be distinguished from $c \times r$

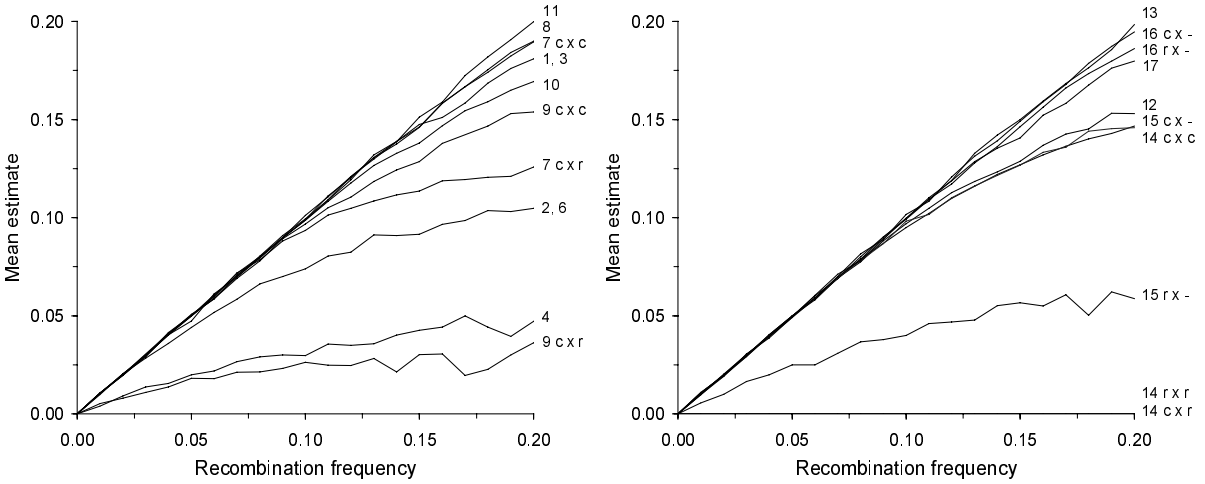



**Figure 2.1 Means of recombination frequency estimates with a significant LOD (> 3.0) over 1000 simulation runs for each value of the true recombination frequency (steps of 0.01) for a population of 50 individuals. For configurations 7 and 9 the linkage phase combinations *c* × *c* and *c* × *r* are equivalent to *r* × *r* and *r* × *c*, respectively, while for configurations 15 and 16 the combinations *c* × *c* and *r* × *c* are equivalent to *c* × *r* and *r* × *r*, respectively (indicated with *c* × - and *r* × - ). Configurations 14 *c* × *r* and 14 *r* × *r* coincide with the horizontal axis.**

in repulsion in the first parent, then most gametes will have an *'a'* allele at least at one locus, so that the contribution of the second parent cannot be determined for the majority of the phenotypes in the offspring. As a consequence, the phase determination is based on only a small minority of the offspring.

The remaining configurations (conf. 8, 10, 11, 12, 13, 17) supply sufficient information to resolve the linkage phase combination unambiguously. Here, always $r_{c \times c} = 1 - r_{r \times r}$ and $r_{c \times r} = 1 - r_{r \times c}$, leaving two legitimate estimates. The estimate with the correct phase practically always has the undoubtedly smallest value and highest LOD, whereas the other legitimate estimate is either close to 0.5 (conf. 8, 11, 12, 17) or in between the smallest estimate and 0.5 (conf. 10, 13).

## 2.4 Properties of the ML recombination frequency estimators

In the design of linkage experiments it is important to know the various statistical properties of the recombination frequency estimators for all situations. Bias and variance are important characteristics describing how close one can get to the true value. Another aspect is that segregation types differ in power with respect to detecting linkage; to obtain a complete linkage map it is necessary that linkage is detected for a sufficiently large number of markers at some significance level. Still, when linkage is detected between a pair of loci, this does not necessarily imply that the estimate is accurate. In some marker type combinations significant estimates are predominantly zero estimates, despite the presence of large numbers of recombination events.

In the simulation studies, individuals segregating for two loci were generated according to Mendelian inheritance at a given recombination frequency. Each study was based on 20,000 replicates of $F_1$-populations consisting of 50, 100, 150, 200 or 1000 individuals. The simulated recombination frequencies ranged from 0 to 0.5 with intervals of 0.001, 0.002 or 0.01. In each $F_1$ the recombination frequency and the LOD score were calculated using the formulas from Table 2.3 with the appropriate linkage phase.

### *(i) Bias*

For infinite population sizes the ML-estimators of all configurations are unbiased. This was proven analytically for some estimators; for others it was demonstrated by simulation, assuming the linkage phase combination was known, for populations of practically infinite sizes (not shown). However, in practice one deals with finite, sometimes small, population sizes. Here, linkage has to be tested and only recombination frequency estimates with a significant test statistic (the LOD score) are retained. In general, large estimates have small test statistics that are not significant, and as a consequence these large estimates are ignored. Thus, in finite populations, a downward bias is introduced in the set of estimates with a significant LOD score. This is illustrated for a population of 50 individuals in Fig. 2.1. Since the bias is caused by rejecting nonsignificant values, it is related to the variance of the estimators, which in turn depends largely on the configuration of the loci as well as on the population size (The variance is treated in the next section). In particular some of the configurations involving *a0* $\times$ *a0* loci are severely biased due to applying the LOD score significance threshold, even with population sizes of 100 or more individuals.

### *(ii) Variance*

The variance of a recombination frequency estimator comprises two components: (1) the number of recombination events that created the progeny sample, and (2) the (in)ability with which these events can be detected for a certain configuration of two loci. The first component is determined by the recombination frequency itself and the progeny size; the second by the segregation types of the loci and the linkage phases in the parents. For instance, from a pair of *ab* $\times$ *cd* loci all recombination events can be observed perfectly (apart from multiple recombination events); here the variance consists only of the sampling variance. In contrast, from a pair of *a0* $\times$ *a0* loci most of these events cannot be observed directly, but have to be estimated assuming Mendelian ratios. If each *ab* $\times$ *cd* locus were completely linked to an *a0* $\times$ *a0* locus, then the estimate based on the two *a0* $\times$ *a0* loci would be different from the estimate using the *ab* $\times$ *cd* loci in the same progeny sample.

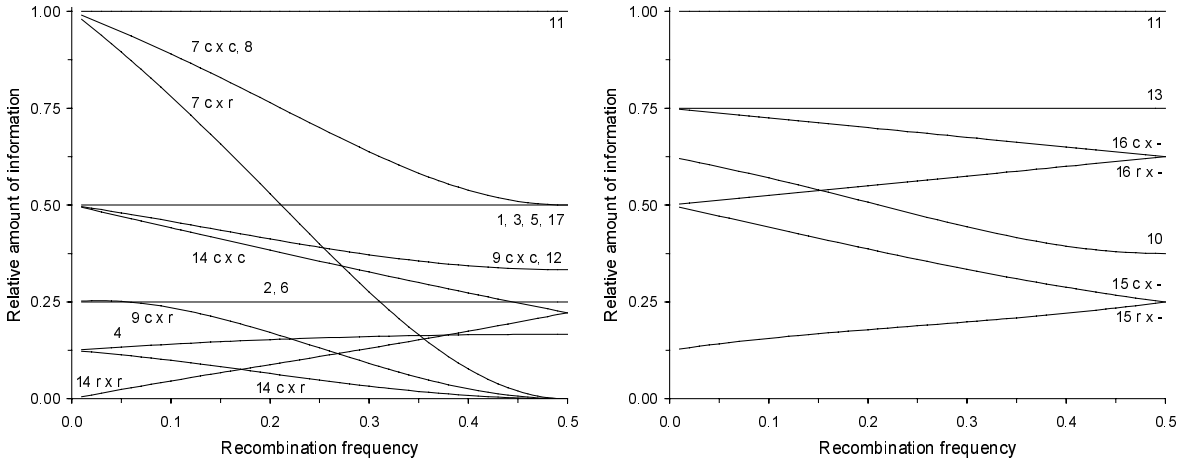



**Figure 2.2 Information functions relative to configuration 11 for all possible marker configurations in a full-sib family of outbred parents. For configurations 7 and 9 the functions for linkage phase combinations *c* × *c* and *c* × *r* are equal to *r* × *r* and *r* × *c*, respectively; for configuration 14 the functions for *c* × *r* and *r* × *c* are equal; for configurations 15 and 16 the combinations *c* × *c* and *r* × *r* are equal to *c* × *r* and *r* × *c*, respectively (indicated with *c* × - and *r* × -).**

The variance of ML-estimators is approximately equal to the inverse of Fisher's information, *i.e.* the expectation of minus the second derivative of the log-likelihood function. Several authors present the information functions of various configurations (Mather 1951; Allard 1956; Ritter *et al.* 1990; Weber and Wricke 1994; Ritter and Salamini 1996). The functions relative to the information of configuration 11 (two *ab* × *cd* loci) are depicted in Fig. 2.2. The information functions of configurations 6 and 17, not described previously, are equal to those of 2 and 1, respectively (equivalent to MCDs 9 and 1 in Ritter and Salamini 1996). Figure 2.2 shows that the combinations with *a0* × *a0* markers, especially configurations 14 *c* × *r* and *r* × *r*, provide a small amount of information. For configurations 2 and 6 (which are equivalent and have the same ML estimator after exchanging the corresponding phenotype frequencies), the reason for the relatively small amount of information is not so evident. In these configurations, according to expectation half of the progeny is not informative at all: the probabilities of two marker phenotype classes are independent of the recombination frequency ($p$=¼ each) (Ritter and Salamini 1996). Configuration 4, which is the dominant marker version of configuration 2, is even less informative: here, the non-informative half of the progeny is hidden behind the marker phenotype *'a-'* of the *a0* × *a0* marker and as such increases the variance of the recombination frequency estimate. In configurations 10 and 17 an expected quarter of the progeny is not informative with respect to the recombination frequency.

Since the inverse of Fisher's information is used only as an approximation for the variance, the variance was also investigated by simulation, assuming the linkage phase combination was known and not applying a LOD score threshold. In most instances the approximation was accurate. However, for configuration 14 $r \times r$, the variance estimated from the simulation results was smaller than the inverse information for small values of the recombination frequency. Only for the largest population size tested ($n$=1000, Fig. 2.3) did the results agree well with the estimate from the inverse information function. For $r$ approaching to 0, the variance estimated from the inverse information function approaches $1/n$. The discrepancy between calculation of the variance from Fisher's information and the simulation results is not well understood; presumably, this is due to the method being an approximation.

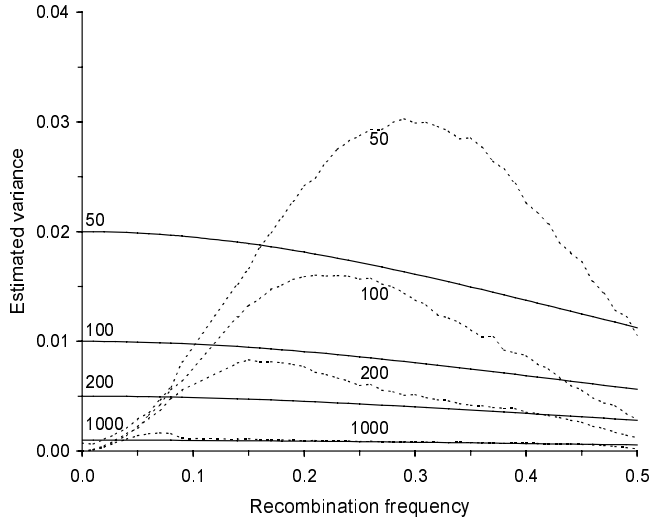



**Figure 2.3 Variance estimates for the recombination frequency in configuration 14 $r \times r$. Population sizes are 50, 100, 200 and 1000. The continuous lines indicate the variance estimated from the inverse information function, the dotted lines show the variance estimated from recombination frequency estimates over 20,000 simulation runs for each value of the true recombination frequency (steps of 0.01).**

## *(iii) Detection of linkage and recombination frequency estimation*

In the development of a linkage map one usually starts with a random set of markers of which no map positions are available. The first step in map construction is the separation of markers into linkage groups. A marker pair is considered to be linked when the realised marker frequencies in the progeny are significantly different from the expected frequencies in the absence of linkage ($r$ = 0.5). Several statistics can be used to test linkage, such as Mather's linkage test $\chi^2_L$ (Mather 1951), the contingency test for independence, or the LOD score test. The contingency test for independence is to be recommended, because the other two tests are affected by systematic segregation distortion (Garcia-Dorado and Gallego 1992). The LOD score test is adequate when there is no systematic segregation distortion, and at present it is possibly the most frequently used test. According to ML-theory, the LOD score follows approximately a chi-square distribution with one degree of freedom: LOD ~ 0.5 $\log_{10}(e)$ $\chi^2_{(1)}$ = 0.22 $\chi^2_{(1)}$ (*cf.* McCullagh and Nelder 1989). Often the value

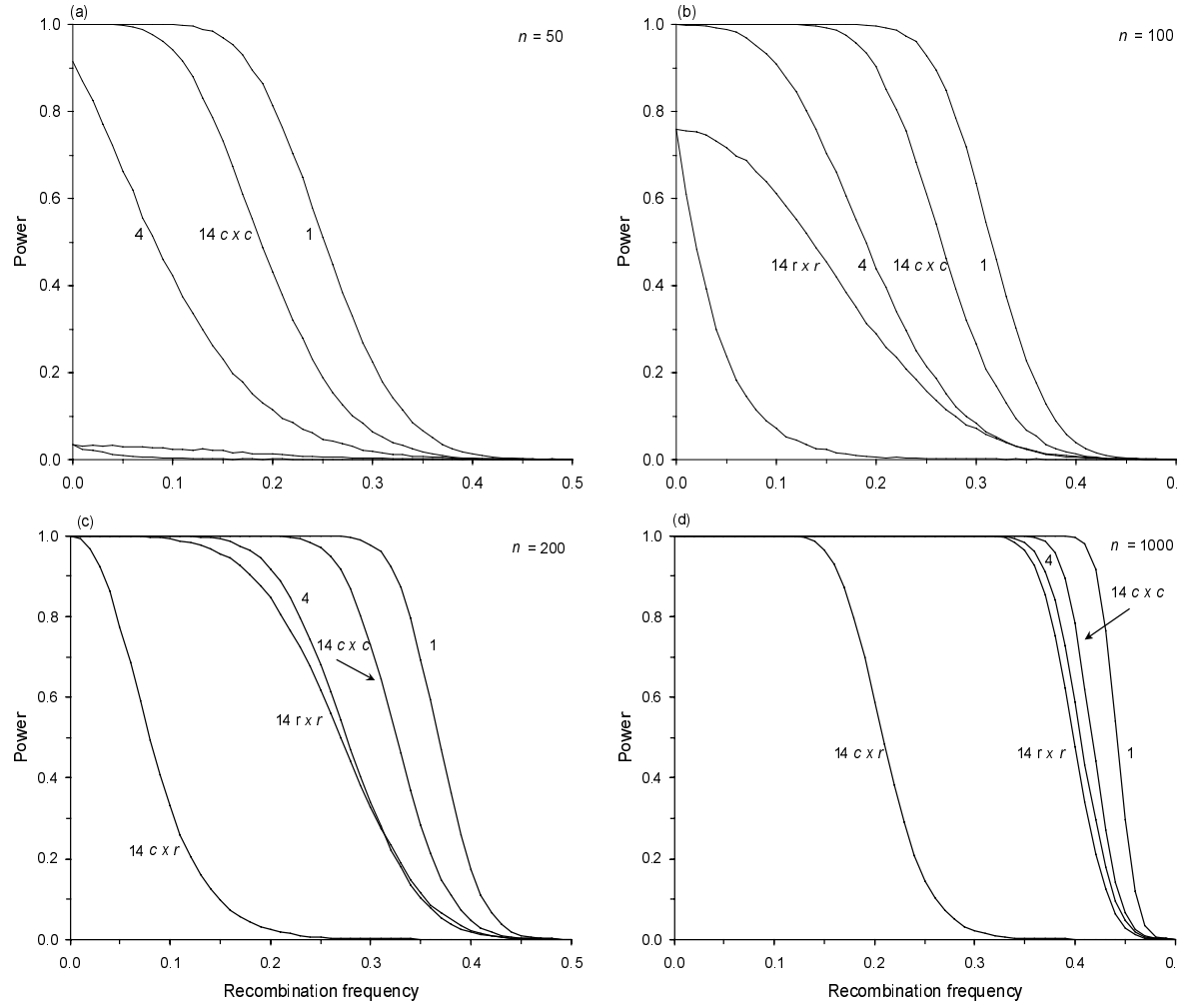



**Figure 2.4 The power of detecting linkage in configurations involving combinations of *ab* × *aa* (or *aa* × *ab*) and *a0* × *a0* type markers. (Fraction of 20,000 simulation runs where a LOD score > 3.0 was obtained). *(a)* population size *n* =50 *(b) n* =100 *(c) n* =200 *(d) n* =1000.**

three is used as the significance threshold, meaning linkage is 1000 times more likely than independent segregation. As a chi-square test, this value corresponds to a significance of 0.0002. This high stringency is needed because many pairs of markers are usually tested (*cf*. Morton 1955; Risch 1992). In the following example we illustrate some important phenomena related to the problems of detecting linkage and the estimation of recombination frequencies.

Suppose we want to construct a map based on RAPD markers determined in an FS-family. These markers would segregate as *ab* × *aa*, *aa* × *ab* or *a0* × *a0*, while pairs of markers would be of configurations 1, 4 or 14 in all possible linkage phase combinations (Table 2.1). Figure 2.4 shows that there are large differences between these configurations for the power of detecting linkage. These differences are related to differences in information functions (Fig. 2.2). For configuration 1, the detection of linkage would usually be no problem, even for recombination frequencies up to 0.3 at a population size of 100. This also holds for configuration 14 *c* × *c* with recom-

bination frequencies up to 0.2. For configuration 4 at a population size of 100, however, the probability of obtaining a significant LOD is larger than 0.9 only for recombination frequencies smaller than 0.1, and the power rapidly decreases beyond 0.1. The power is rather small for configuration 14 $r \times r$, and even dramatically small for configurations 14 $c \times r$ and $r \times c$. Since linkage between $ab \times aa$ and $aa \times ab$ markers cannot be established directly, their linkage has to be determined through $a0 \times a0$ markers, *i.e.* through configuration 4. Thus, in order to establish linkage between $ab \times aa$ and $aa \times ab$ markers one needs an $a0 \times a0$ marker closely linked to both an $ab \times aa$ and an $aa \times ab$ type marker and hence a large number of $a0 \times a0$ markers would be required; in practice these are not always available.

When significant LOD scores were obtained in our simulations for configurations 14 $r \times r$ and $c \times r$ (and $r \times c$), very often the corresponding estimate of the recombination frequency was zero, which can be understood from the small probability of finding visible recombinants in these configurations. Zero estimates were obtained for even quite large values of the recombination frequency. For instance, for a population size of 150 and a recombination frequency of 0.15, the fractions of the simulation runs that had a significant LOD score were 0.82 and 0.05 for $r \times r$ and $c \times r$, respectively, and the recombination frequency estimate was zero in 51% and 13% of those fractions, respectively. In a population size of 100 the fractions with a significant LOD were 0.45 and 0.02 and zero estimates were found in 87% and 75% of those fractions. In a population of size 50 significant LOD scores were hardly ever found and for zero estimates only.

A more remarkable though very rare phenomenon was the occurrence of non-zero estimates when the true recombination frequency was zero. This was observed in simulations of configurations 14 $r \times r$ and $c \times r$, 4, 9 $c \times r$ and 15 $r \times c$ and $r \times r$. In all cases the frequency of occurrence was below 2% for a population of size 50, and lower for larger populations. This occurred only for situations where there were large deviations from the expected segregation ratios. It can be proved that this cannot occur for configuration 14 $c \times c$.

Another aspect in this example is the accuracy of the estimates. Although a significant LOD score indicates linkage of a marker pair, it does not imply that the estimate of the recombination frequency is accurate. In the process of mapping we are not only interested in detecting linkage, but accurate estimates are needed to determine the order and distances of the markers. For configuration 1 exact confidence intervals can be given for the recombination frequency, since the number of recombinant genotypes in the progeny follows a binomial distribution with probability $r$ for recombination (Fig. 2.5). For an estimate of 0.10 and a population size of 100, the 95% confidence interval is [0.05, 0.18]. Although in the other configurations multinomial distributions might be used to construct exact confidence

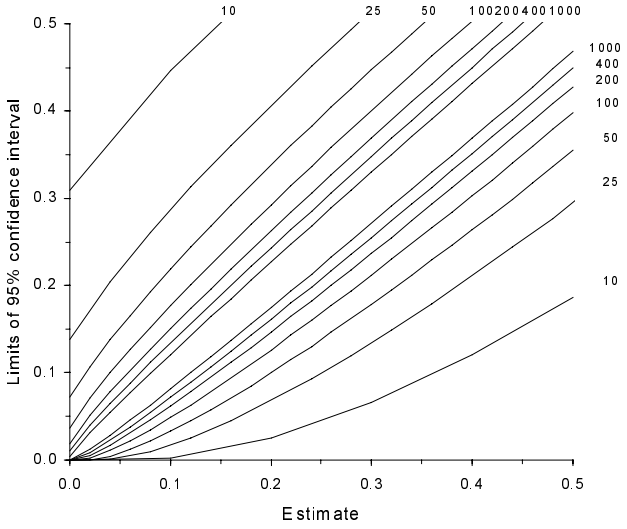



**Figure 2.5 Lower and upper limits for exact two-sided confidence intervals of the recombination frequency for estimates based on progeny sizes of 10, 25, ..., 1000 for configuration 1.**

intervals for the recombination frequency, this is quite laborious and these would have to be calculated for each situation separately. Instead, an indication of the accuracy can be obtained by using the relative amount of information from Fig. 2.2 to construct rough confidence intervals. For instance, for an estimate of 0.10 for configuration 4, the amount of information is a fraction 0.13/0.50 = 0.26 of the information in configuration 1 at $r$ = 0.10. So, the "effective population size" is a fraction 0.26 of the population size for markers in configuration 1. An approximate 95% confidence interval can now be found for a population of size 26 and is equal to [0.02, 0.28]. For situation 14 $c \times c$ an effective population size of 88 leads to an approximate confidence interval of [0.04, 0.19]. Similarly, approximate confidence intervals of [0.02, 0.32] and [0.01, 0.44] are found for 14 $c \times r$ and $r \times r$, respectively. The width of these rough confidence intervals indicates clearly that difficulties may be expected in the ordering of dominant markers. Although *a0* $\times$ *a0* markers can be used to combine the *ab* $\times$ *aa* with the *aa* $\times$ *ab* markers, their usefulness in establishing the correct marker order between these two groups will be very limited in small populations.

### *(iv) Linkage phase*

Prior to the detection of linkage the linkage phase combination has, of course, to be determined. The success of the methods described previously was tested by simulation. The choice for the linkage phase combination corresponding to a significant LOD score and a legitimate estimate of $r$ was correct in virtually all simulations for all configurations with normal population sizes ($n$ > 50), except where linkage phases cannot be distinguished according to theory. In just a few cases indirect estimation (or verification) of the linkage phase through more informative linked markers may be necessary. Of course, if the LOD score is not significant, the

choice of the linkage phase cannot be made reliably. From a theoretical point of view it may be interesting to develop a procedure for simultaneous estimation of recombination frequencies and linkage phase combinations over all linked markers. However, in most practical situations this will be redundant.

## 2.5 Concluding remarks

This paper describes the marker configurations found in segregating full-sib families of crosses of outbred parents. Seven distinct segregation types characterise the inheritance of individual markers. In practice, the determination of the segregation type of a marker is not always straightforward. For molecular markers this essentially means defining which molecular fragments are allelic. Two marker fragments present in only one parent can be regarded as alleles if either the one or the other is present in all progeny individuals. The probability that this occurs for unlinked loci is very small ($½^n$), even in small populations, and also for linked loci this probability ($r^n$ for coupling phase or $(1-r)^n$ for repulsion) decreases rapidly for increasing values of the recombination frequency $r$. For fragments from different parents, the inference of allelism cannot be made if the markers are based on short DNA sequences and yield large numbers of fragments, such as RAPDs. For RFLPs or sequence tagged sites, however, the inference of allelism will generally be easy since these techniques are based on homology of large segments of DNA and usually yield only a small number of molecular fragments. Still, it has to be realised that in allopolyploid species and species with a polyploid origin, homeology across the genome may impede such conclusions. On the other hand, our own experience on the apple has also shown that for RFLPs the comparison of restriction patterns with different restriction enzymes can be helpful (A.W. van Heusden, CPRO-DLO, Wageningen, the Netherlands, personal communication). If the parental phenotypes are missing and segregation may be distorted, the determination of the segregation type can become complex, *e.g.*, segregation types $a0 \times 00$, $00 \times a0$ and $a0 \times a0$ cannot be distinguished in the progeny since only presence or absence of the band can be scored in the progeny.

In this paper we demonstrated that the various marker pair configurations differ greatly in the accuracy of recombination frequency estimation, the power of detecting linkage and the (im)possibility to estimate the linkage phases in both parents. The information functions as presented in Fig. 2.2 are a good indication of such differences and may help in the planning of linkage experiments. Also, after collecting marker data the differences in accuracy of the recombination frequency estimates in the various configurations should be considered, so that the ordering of markers per linkage group and the calculation of marker distances may be optimised. After markers have been assigned to linkage groups, conflicting information with respect to the marker order is often provided by the different pairwise recombination frequency estimates. This can be due to missing marker

data, but also to random estimation errors in the recombination frequency inherent to the marker configurations. The knowledge of the (in)accuracy of the recombination frequency estimates should then be taken into account to solve such conflicts. For example, in the determination of the distance B-C in a group of four linked markers A-B-C-D, the combined (and weighted) information of the A-B, A-C, B-C, B-D and C-D estimates may well provide a more accurate distance estimate than the single and direct B-C estimate, especially when *e.g.* markers A and D are of type *ab* × *cd*, while B and C are of type *a0* × *a0*. For instance, in the computer program JoinMap® (Stam 1993; Stam and Van Ooijen 1995), this is done by using all pairwise recombination frequencies, weighted with the LOD scores, to simultaneously estimate the marker order and distances.

The (in)accuracy of recombination frequency estimates should further be borne in mind when a map resulting from a single cross is used for indirect selection. The upper bound of the confidence interval of the recombination frequency should give an idea of the maximum probability of breaking the linkage between marker and the gene of interest in the subsequent generations. In this respect it is good to note that apart from estimation errors there may also be genetic differences in the recombination frequency (and in the linkage phase combination) in different crosses, as there may be differences between male and female meioses. (*e.g.* Van Ooijen *et al.* 1994; Plomion and O'Malley 1996).

The possibility of constructing a single map for a cross, rather than two separate maps for the parents of the cross, depends upon the availability of allelic bridges (Ritter *et al.* 1990). Although in principle *a0* × *a0* markers could be used as allelic bridges, they will often provide little information so that RAPDs or AFLPs may be of limited use for combining the parental maps. For example, in the double pseudo-testcross populations of apple (Hemmat *et al.* 1994) and Eucalyptus (Grattapaglia and Sederoff 1994), where mostly dominant markers were used, separate maps for the individual parents in the cross could be constructed but the integration of these parental maps was difficult. When a mapping study is done with the intention of integrating the homologous linkage groups of the respective parents, multi-allelic markers, such as RFLPs or microsatellite markers, are recommended. Grattapaglia and Sederoff (1994) and Ritter and Salamini (1996) emphasised the power of such markers for mapping studies in outbred progenies. An extra advantage of these markers is the high probability that they can be used over a wide range of crosses. Another advantage is that, at least where the *ab* × *cd* type of markers is concerned, differences in recombination between the male and the female parent can be estimated directly, whereas *e.g.* in $F_2$ populations from inbred lines the recombination frequency has to be assumed equal in the male and female meioses and reciprocal backcross progenies are needed to detect possible differences. If a sufficient number of *ab* × *cd* markers is used in an FS-family of outbred parents, both options are available: either use the separate maps of both

parents, or, if the differences in recombination frequency are not too large, construct an integrated map for the cross.

# Chapter 3

## Multiple field and glasshouse assessments increase the reliability of linkage mapping of the *Vf* source of scab resistance in apple

G.J. King, F.H. Alston, L.M. Brown, E. Chevreau, K.M. Evans, F. Dunemann, J. Janse, F. Laurens, J.R. Lynn, C. Maliepaard, A.G. Manganaris, P. Roche, H. Schmidt, S. Tartarini, J. Verhaegh and R. Vrielink[1]

## Abstract

Apple scab, caused by the fungus *Venturia inaequalis* (Cke.) Wint., is an important disease in commercial apple production. A mapping population of 155 individuals, derived from a cross between the apple varieties 'Prima' (resistant) x 'Fiesta' (susceptible), was scored for response to the disease in replicated field and glasshouse trials throughout Europe. Twenty data sets were selected and cluster analysis was used to form a consensus score for the population fitting a 1:1 segregation ratio of resistance:susceptibility. The progeny were scored with molecular markers. A detailed map covering 54 cM of the 'Prima' linkage group containing the *Vf* gene of scab resistance was constructed using 24 molecular markers which are linked to the resistance gene. One isoenzyme marker (*Pgm-1*), 6 RFLP markers and 17 RAPD markers formed a linkage group with the consensus measure of resistance to scab. Four marker bridges were established with the corresponding 'Fiesta' linkage group with additional markers (one isozyme, one RFLP, three RAPD and one AFLP). A low chi-square value indicated a good fit of the marker ordering, which was in close agreement with previously reported linkage positions for some of the markers and *Vf*. Differences were observed in the ability of different scoring methods to resolve susceptible and resistant classes. The results obtained for the consensus classification of resistance to scab for the population may suggest the presence of virulent inocula at some sites, which could overcome the *Vf* gene for resistance. The consequences of relying on individual scoring occasions for studying *Vf* scab resistance are discussed in the context of linkage analysis, conventional breeding selection and marker assisted selection.

[1] Theoretical and Applied Genetics (1998) 96: 699-708
Reproduced with permission of Springer-Verlag, Berlin

## 3.1 Introduction

Apple scab is economically one of the most important diseases of apple trees (*Malus pumila* Mill.), especially in regions of cool, moist springs and summers. The causal organism of apple scab is the fungus *Venturia inaequalis* (Cke.) Wint. The disease is spread by dispersal of fungal spores and affects both leaves and fruit, with scab lesions on fruit reducing the marketable quality and value of the crop. Disease control is primarily achieved in commercial production systems by costly and repeated use of fungicide sprays; reduced fungicide inputs are seen to be beneficial to the grower, environment and consumer. Intensive use of fungicides over a prolonged period of time poses the potential threat of fungicide resistance in the pathogen population due to the selection of fungicide insensitive strains and a subsequent loss of control.

The introduction of scab resistance into apple cultivars has been the aim of many modern apple breeding programmes. The most widely used source of resistance has been derived from the small fruited ornamental species *M. floribunda* Siebold ex Van Houte, clone 821. This resistance, assigned to a gene called *Vf* (Williams *et al.* 1966), has been utilised in apple breeding programmes throughout the world for more than 40 years, and has been incorporated into a substantial number of apple cultivars, although none has been extensively planted (Crosby *et al.* 1992).

Until recently, *Vf* was considered to be the most effective source of resistance, because *Vf* cultivars had been free from scab for over 50 years in the different countries where they were grown. Parisi *et al.* (1993) then identified a pathotype (race 6), isolated in Germany, which was virulent against a number of cultivars or selections carrying *Vf*, whereas *M. floribunda* 821 itself was resistant. They concluded that there was an urgent need to diversify the sources of resistance to *V. inaequalis*. In addition it is likely that there are additional resistance loci in *M. floribunda* 821 which were not introgressed into the cultivars and selections which are susceptible to race 6. Roberts and Crute (1994) have since identified a pathotype of *V. inaequalis* (FL1, isolated in England) to which *M. floribunda* 821 and derived cultivars are susceptible.

Several molecular genetic markers have recently been identified which are linked to the introgressed segment of the *M. floribunda* 821 genome which confers *Vf* resistance to *V. inaequalis*. These include the isoenzyme locus Pgm-1 (Manganaris *et al.* 1994) and at least nine DNA-RAPD markers (Yang and Krüger 1994; Koller *et al.* 1994; Durham and Korban 1994; Tartarini 1996). All RAPD markers were identified following a bulked segregant approach (Michelmore *et al.* 1991), using bulks composed of either resistant or susceptible cultivars, or individuals from one or more segregating populations. The bulked segregant approach is relatively efficient in identifying markers linked to a major gene

phenotype, although the subsequent genetic analyses may be limited if these markers are not placed in the context of a linkage map. Gianfranceschi *et al.* (1996), Tartarini (1996) and Gardiner *et al.* (1996) have presented partial maps based on these RAPDs, for introgressed sections of the *Vf* linkage group. More detailed linkage information helps in the analysis of interactions with other genes and can be used to minimise linkage drag by selecting resistant plants that carry the shortest segment of alien chromosome.

The European Apple Genome Mapping Project was initiated in 1989 to consolidate the studies on apple genetics carried out by the major European apple breeding institutions (King *et al.* 1991). Replicated reference populations were established and distributed amongst partners in six countries (King 1994). A wide range of molecular marker and trait data have now been accumulated (King 1996), including comprehensive assessments for resistance to *V. inaequalis* under both field and glasshouse conditions. The data reported here represent the most detailed assessment of linkage relationships for this major source of apple scab resistance.

## 3.2 Materials and Methods

### *Plant material*

A cross between 'Prima' and 'Fiesta' was carried out at CPRO-DLO, Wageningen in 1988. 'Prima' (Dayton *et al.* 1970) is a variety selected in the USA Co-op programme (Crosby *et al.* 1992), where the *Vf* source of scab resistance has been introgressed over four generations. The scab-susceptible variety 'Fiesta' was selected at East Malling from a cross between 'Cox' and 'Idared'. The seedlings were raised in pots, and planted in the nursery at Wageningen in the winter of 1990/1991. Five replicate

**Table 3.1 Locations of field and glasshouse sites for screening 'Prima' x 'Fiesta' progeny for resistance to apple scab.**

| Site | Location | | Country | Propagation |
|---|---|---|---|---|
| | Grid reference | Altitude (m) | | |
| East Malling | 51°17’ N 00°27’ E glasshouse | 32 | England | M27 cordons M9 in pots |
| Elst | 51°55’ N 05°50’ E | 8 | The Netherlands | M27 staked |
| Angers | 47°30’ N 00°35’ W glasshouse | 57 | France | M27 staked MM106 in pots |
| Ahrensburg | 53°40’ N 10°15’ E | 46 | Germany | M27 staked |
| Cadriano | 44°32’ N 11°23’ E | 30 | Italy | M27 staked |
| Naoussa | 40°37’ N 22°07’ E | 121 | Greece | M27 staked |

trees of each genotype were obtained by bud-grafting wood of 155 seedlings onto M27 dwarfing rootstock in early 1992 and grown for another year. The trees were grown for another year and distributed to the six sites in early 1993 (Table 3.1), with one population being divided between Germany and Italy. Potted trees for glasshouse testing were obtained by grafting wood from the respective field plantations onto M9 (East Malling) and MM106 (Angers) rootstocks.

Fungicides were applied as sprays to the grafted trees at Elst and Cadriano throughout 1993 to establish healthy growth. At Angers and Ahrensburg, fungicides were applied until June 1993. At East Malling and Naoussa the trees were left unsprayed for the duration of the study.

## *Assessment of scab infection*

The segregating population was assessed for disease response on 41 occasions in three years over six field sites, and glasshouse trials at two sites. Different methods for scoring symptoms and incidence were used, some of which were based on those currently used in breeding selection. Twenty-two data sets were selected and cluster analysis used to form a 'consensus' score fitting a 1:1 segregation ratio over the population.

Glasshouse assessments were carried out at HRI-East Malling and INRA, Angers (Table 3.3). At East Malling plants were inoculated with the *E1* isolate of *V. inaequalis* (Kirkham 1957), whilst at Angers the inoculum was prepared from dried scabbed leaves collected from orchards around the INRA station at Angers. Previous characterisations of these scab populations showed that they are generally very aggressive and are primarily represented by race 1 spores. At East Malling a conidial suspension of $2x10^5$ spores $ml^{-1}$ was used for inoculation, whilst at Angers the concentration varied between $4x10^5$ and $6x10^5$ spores $ml^{-1}$. Inoculations were carried out by spraying and performed only on trees which were actively growing and healthy. To maintain a high humidity, plants were covered with transparent polythene for the first 48 hours following inoculation: at East Malling, plants were enclosed individually in polythene bags, whilst at Angers each tray of plants was covered. A humidity of 80-100% and a temperature of 18◦C was maintained throughout the experiments. Symptom development was recorded after two weeks using the Vi-GH-1 descriptor scale at both sites (Table 3.2). In addition at Angers, two replicate plants per individual genotype were screened. Trees were inoculated for a second time after 12 days and a second assessment of scab infection was made 15 days following the second inoculation using the Vi-GH-1 descriptor scale.

Field assessments of scab infection were made on several occasions (Table 3.3) at each site (Table 3.1). Only one plot had been subject to a fungicide spraying programme (Angers). Scoring was carried out using several different scoring methods or 'descriptors' described in Table 3.2. The level of field infection at each site was likely to be dependent upon several factors, including the population

structure of the pathogen and the environmental conditions. The relative abundance and virulence of the different pathotypes within a local population were expected to differ between sites and between years. For example, the virulent race 6 inoculum of *V. inaequalis* was known to be present at Ahrensburg (Parisi *et al.* 1993). In addition, environmental conditions suitable for scab infection varied from site to site as well as from year to year. For example at East Malling, the number of infection periods predicted by the Ventem$^{TM}$ infection warning system (Xu *et al.* 1995) for the critical period between March and June was 22 in 1993, 29 in 1994 and eight in 1995.

## *Molecular markers*

*Isoenzymes*

Young actively growing leaf tissue for isoenzyme analysis was collected from the field sites in early June and weighed samples were stored at -80°C until extraction and analysis. Protein extraction and starch electrophoresis were carried out according to Chevreau and Laurens (1987). The staining for Phosphoglucomutase (*Pgm* = EC 5.4.2.2) was carried out according to Wendel and Weeden (1989).

*DNA markers*

DNA was extracted using either a modified mini-prep CTAB-based method (Doyle and Doyle 1990) or a large-scale nuclei-isolation method (Van der Beek *et al.* 1992; Roche *et al.* 1997a). RAPD primers (Table 3.4) were obtained either from Operon (Alameda, California), the University of British Columbia or custom synthesised by a number of suppliers. PCR reaction conditions were based on a standard protocol (King 1994) adopted by participants in the European Apple Genome Mapping Project, or modified for the OPAM19 and OPAL07 assays (Tartarini 1996). RAPD data were scored in different participating laboratories and reproducibility was confirmed by replicating experiments in different laboratories. RFLP analysis followed the methods described in Roche *et al.* (1997a) as modified from Van der Beek *et al.* (1992). The pB610 clone used as an RFLP probe was kindly supplied by Dr Gavin S. Ross, HortResearch, Auckland, NZ.

A total of 208 markers were selected from a larger dataset for linkage analysis, based on population coverage and segregation ratio (Maliepaard *et al.* 1998)

## *Data management and analysis*

Scab infection, marker and related data were entered into the Apple-Store relational database (Hyne 1995) which was used to generate files for linkage analysis. Linkage analysis was performed with JoinMap version 2.0 (Stam 1993; Stam and Van Ooijen 1995) using a LOD score of 4.0 for grouping markers into linkage groups and the Kosambi mapping function to calculate map distances. The total chi-square value

**Table 3.2 Methods used to assess incidence and symptoms of scab occurrence in field (Vi-F-1, Vi-F-2, Vi-3, Vi-5) and glasshouse (Vi-GH-1) tests.**

| Descriptor | Type | | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|
| Vi-F-1; Vi-GH-1 | Symptom | leaf | No symptoms | Pin point pits | Chlorotic lesions, possibly small necrotic spots | Chlorotic and necrotic spots | Chlorotic and necrotic spots, sporulation present | Sporulation |
| Vi-F-2 | Incidence | leaf | No leaves attacked | 1-10 % leaves attacked | 11-35 % leaves attacked | 36-65 % leaves attacked | 66-90 % leaves attacked | 91-100 % leaves attacked |
| Vi-3 | Incidence | leaf | No infection | Very little infection | Clear spots of infection | - | - | - |
| Vi-5 | Symptom | leaf | No observation | Symptoms of resistance | Low susceptibility | Middle susceptibility | High susceptibility | Very high susceptibility |

was calculated by carrying out 297 pairwise recombination frequency estimates for the 25 markers, resulting in 273 degrees of freedom (df).

### *Scab trait data*

Inspection of many of the data sets showed them to have a unimodal distribution of scoring grades, with no strong relationship between pairs of data sets (Fig. 3.1 b,d). A subset of data sets did show a clear bimodality (Fig. 3.1 a,c). Within this subset, there was strong agreement of the classification of trees into each of the two modes. Since these observations fitted the expected model of a single dominant resistance gene with a 1:1 segregation ratio (Williams *et al.* 1966), this model was adopted in the subsequent analysis.

Forty-one sets of measurements relating to scab infection were examined. Each data set was inspected to determine the cut point which best divided the plants according to a 1:1 ratio. Twenty data sets which allowed a division into resistant and susceptible classes with a segregation ratio between 1:2 and 2:1 were selected, to form a 'consensus' classification by further analysis. These included 17 field and three glasshouse data sets. Of the data sets discarded, by further analysis only two had segregation ratios between 1:3 and 3:1, and many revealed no infection (*e.g.* at Naoussa). All but three of the sets retained had ratios between 2:3 and 3:2. Of the data sets retained the glasshouse assessments were scored using the symptomatic descriptor method Vi-GH-1 (Table 3.2) in two countries over two years, and the majority of the field assessments were scored using the descriptor methods Vi-F-1 (symptomatic) and Vi-F-2 (incidence) in three countries over two years. In total, field measurements included data sets from four countries using four scoring systems over three years.

These twenty data sets, with the plants classified as resistant or susceptible, were then combined using cluster analysis with the complete link clustering algorithm (Matula 1977) in the package Genstat 5 (1993). The classification was carried out for each data set individually with the cut point solely based on the assumption of a 1:1 segregation. Any discrepancies in relation to the actual scoring grade were not taken into account, but were addressed later in the analysis. Indeed, it was found that this approach had no significant effect on the composition of the consensus classification. Two distinct groups were formed and these were taken to represent the resistant and susceptible plants. This consensus classification was then subjectively assessed by comparing it to each of the original 41 data sets, to establish the degree to which there was disagreement.

To assess the impact of the arbitrary selection of the threshold values of 1:2 and 2:1 as the criterion for inclusion of data sets into the consensus-forming set, the clustering was repeated following either inclusion of additional sets or removal of existing sets, by setting the threshold values at either 3:1 and 1:3 or 2:3 and 3:2.

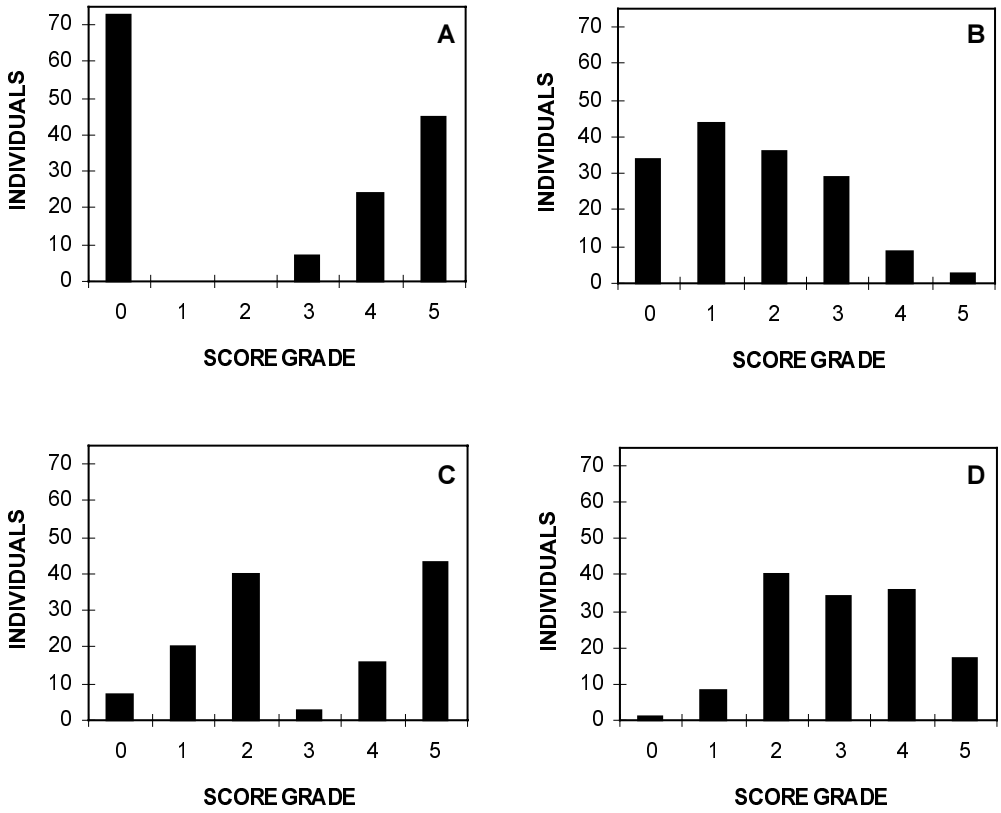



**Figure 3.1 Comparison of distribution of score grades in the segregating population 'Prima' x 'Fiesta' in field and glasshouse tests at different sites. Vi-F-2 incidence scores in the field A. at Elst in 1994; B. at East Malling in 1993; Vi-GH-1 symptomatic scores in glasshouse tests C. at Angers and D. at East Malling, both in 1994.**

The effect of changing the thresholds was assessed by comparing the output for the classification of individuals into each of the resistant and susceptible groups.

## 3.3 Results

### *Compilation and assessment of the consensus score*

There were differences between the threshold grade required to distinguish between resistant and susceptible individuals, in data sets scored using the same descriptor scale both between sites and within sites. In some data sets with a bimodal score distribution, where very few plants were assigned to the intermediate score grade, then the choice of threshold to form a 1:1 ratio resulted in a misclassification of that grade with respect to the consensus classification. For example, the Vi-F-1 data set from Elst in 1994 gives the optimal segregation ratio (76:72) when score grades 0-4 are considered resistant. All of the plants scoring grade '4' are, however, susceptible according to the consensus score, and indeed were susceptible in the opinion of the scorers of that occasion, and reclassifying in this way gives a segregation ratio of 71:77. In other data sets the fit to the consensus classification cannot be improved

by adjusting the threshold. The difference in thresholds assigned to different data sets must therefore be attributed either to different levels of attack, interactions with different *V. inaequalis* populations, patchy distribution of different pathotypes in field plots or discrepancies in classification by different assessors.

The clustering process carried out on the 20 sets of data to produce a consensus classification formed two groups at a similarity level of about 0.2. The resistant group consisted of 77 individuals, and the susceptible group 78. The robustness of the consensus classification to variations in the threshold for inclusion of data sets in the consensus-forming group was tested, as described in the Materials and Methods section. The consensus classification was unchanged by varying the datasets included in its formation. It showed a very high level of agreement with the field data sets from Elst, with only four classification discrepancies in individual scores over four scoring occasions. With glasshouse data from Angers, seven discrepancies occurred in two data sets.

The number of classification discrepancies present in the remaining non-consensus forming data sets was greater. In measurements taken in Naoussa where there was low infection pressure, and most trees showed no symptoms, susceptibility was observed in only four of the 150 trees present. In data sets from other sites discrepancies in the resistant:susceptible classification were present for up to 45 individuals (about 30% of the total). In some pairs of data sets from different dates in the same year at one site, or at the same time using different scoring methods, there were a large number of discrepancies in resistant:susceptible classification, with different individuals being responsible for the discrepancies in the different data sets. An inspection of the patterns of misclassifications failed to show any evidence of specific genotypes being consistently miss-scored.

## *Assessment of scab infection*

The ability of different assessors to agree on scoring grades was tested at Elst. This experiment is pertinent to the reliability of scoring systems used in breeding selection. Two teams of two people each carried out scoring within three days of each other in September 1995, using the Vi-F-2 descriptor scale. There was good agreement between the two teams, with 112 trees having identical scoring grades, 34 differing by one scoring grade, and only two plants differing by two scoring grades. Of the 112 trees which received identical scoring grades, 72 were scored as zero by both assessor teams. Even excluding these easily classified trees, the two assessor teams agreed on the scoring grades for 40 plants. In the cases where there was disagreement, one assessor team scored 14 plants at a lower grade, and 22 plants at a higher grade. For only one plant was the susceptibility assessed differently by the two assessor teams (grade '3' and grade '1').

**Table 3.3 Assessments of scab development carried out on the replicated progeny of the cross between 'Prima' x 'Fiesta' at different sites. Sets of scores selected to derive the 'consensus' phenotypes which were used in the linkage analysis. Vi-GH-1 is derived from the method described by Chevalier et al. (1991).**

| Descriptor | Site | Date month/year | Graded as resistant | Ratio (R:S) | $\chi^2$ 1:1 |
|---|---|---|---|---|---|
| *Vi-GH-1* | East Malling | 6/94 | 0 - 3 | 83 : 53 | 6.62 |
| | Angers | 4-5/94 | 0 - 2 | 67 : 63 | 0.12 |
| | Angers | 4-5/95 | 0 - 2 | 65 : 63 | 0.03 |
| *Vi-F-1* | East Malling | 6/93 | 0 - 3 | 69 : 86 | 1.86 |
| | East Malling | 5/94 | 0 - 4 | 66 : 85 | 2.39 |
| | East Malling | 8/94 | 0 - 4 | 68 : 83 | 1.49 |
| | Elst | 6/94 | 0 - 4 | 76 : 72 | 0.11 |
| | Ahrensburg | 7-9/94 | 0 - 4 | 31 : 36 | 0.37 |
| | Ahrensburg | 9/94 | 0 - 4 | 27 : 21 | 0.75 |
| | Ahrensburg | 95 | 0 - 4 | 34 : 37 | 0.13 |
| *Vi-F-2* | East Malling | 6/93 | 0 | 67 : 88 | 2.85 |
| | East Malling | 8/93 | 0 - 1 | 78 : 77 | 0.01 |
| | East Malling | 5/94 | 1 | 63 : 88 | 4.14 |
| | East Malling | 8/94 | 1 | 69 : 82 | 1.12 |
| | Elst | 9/94 | 0 | 73 : 76 | 0.06 |
| | Elst | 95 | 0 | 74 : 74 | 0 |
| | Elst | 95 | 0 - 1 | 75 : 73 | 0.03 |
| | Ahrensburg | 95 | 0 - 1 | 25 : 46 | 6.21 |
| *Vi-3* | Ahrensburg | 8/93 | 1 | 17 : 25 | 1.52 |
| *Vi-5* | Angers | 7/93 | 0 - 4 | 21 : 21 | 0 |

The effectiveness of the different scoring systems was compared. In field tests with low levels of scab inoculum many of the consensus susceptible individuals showed no symptoms. All of the data sets scored in the glasshouse show some consensus susceptible plants scored as resistant, and two of the three data sets show at least one consensus resistant plant scored as susceptible, possibly as a result of the increased infection pressure. The divergence from the consensus classification is higher in the data from East Malling than that from Angers.

Of the data sets scored in the field, the symptomatic and incidence descriptors Vi-F-1 and Vi-F-2 gave the same information about susceptibility at Elst, where the scores were in strong agreement with the consensus. However, on scoring occasions where there was deviation from the consensus, the two descriptors gave complementary information. At East Malling in 1994, for example, both Vi-F-1 and Vi-F-2 were scored in May and August. On both occasions the consensus

susceptible plants all exhibited sporulating scab, with all but four or five plants scoring a grade '5' (sporulating) on the Vi-F-1 scale. Fewer than one in five of the plants scored less than a grade '2' (up to 10% of leaves attacked) on the Vi-F-2 scale.

There was evidence that the resistance of the consensus "resistant" plants was partially overcome at East Malling. About one third of the consensus resistant plants displayed sporulating scab in May, and this had increased to about one half of the consensus resistant plants by August. About one half of the consensus resistant plants with sporulating scab scored a grade '5' on the Vi-F-1 scale, and only one in six scored more than a grade '1' on the Vi-F-2 scale. Comparing overall performance, on both occasions the Vi-F-1 scale alone misclassified between 15 and 20 plants, while the combined scores misclassified five or six. The Vi-F-2 scores alone misclassified 39 and 24 plants on the two occasions.

It had been observed that the consensus-forming group of data sets included data sets from some sites which consistently contained some individuals which displayed sporulation, whereas in the consensus they were classified as resistant. Removal of these data sets did not affect the consensus classification, but did raise the similarity level of the clustering from 0.2 to 0.3.

Only one false susceptible score was recorded at Angers during the two glasshouse tests, although several susceptible plants were not detected as such. In glasshouse tests at East Malling, with a single isolate, there were a larger number of susceptible plants which were not detected, and several false susceptible scores. The increased accuracy at Angers may be explained by the replication which allowed the rejection of the scores for plants giving inconsistent scores. The screening tests performed on unreplicated seedlings as part of a breeding programme might be expected to have an error rate more similar to that obtained at East Malling.

Additional preliminary scab records of presence or absence of sporulation from the nursery at Wageningen in 1991 agreed well with the consensus score, with the exception of 8 out of 157 plants. In five cases plants were scored as susceptible while the consensus score was resistant, and three were scored as resistant compared to a consensus of susceptible.

## *Linkage mapping of markers and consensus scab resistance scores*

Twenty-four linked markers (six RFLPs, one isoenzyme marker and 17 RAPD markers) were grouped together with the consensus scab resistance (Table 3.4). The map for this linkage group resulted in a total chi-square value of 74.5 for 273 df (mean chi-square of 0.273) indicating a good fit of the marker ordering. Double recombinant individuals were checked on the original score sheets and on autoradiographs. The most closely linked markers were the RFLP derived from M18-

**Table 3.4 Details of marker loci and the recombination percentages of these loci with the consensus score of scab resistance.**

| Locus | RAPD sequence or marker type | Map position (cM) | Recombination (%) | SE (%) |
|---|---|---|---|---|
| *'Prima'* | | | | |
| OPAF-12-2000 | GACGCAGCTT | 0.0 | 39.5 | 3.9 |
| OPAD-12-0510 | | 2.0 | 38.8 | 3.9 |
| MH876a | RFLP genomic | 9.1 | 34 | 4.9 |
| MC014a | RFLP cDNA | 16.5 | 29.1 | 4.3 |
| OPC-08-1100 | TGGACCGGTG | 23.0 | 19.7 | 3.2 |
| OPAB-19-1430 | ACACCGATGG | 23.7 | 19.1 | 3.2 |
| OPO-14-1700 | | 24.5 | 18.4 | 3.1 |
| UBC213-2100 | | 28.0 | 14.3 | 3 |
| OPAF-13-2100 | | 28.0 | 15.2 | 2.9 |
| OPA-11-2200 | CAATCGCCGT | 28.0 | 14.3 | 5.6 |
| OPC-09-0900 | CTCACCGTCC | 29.1 | 13.9 | 2.8 |
| OPD-20-0500 | ACCCGGTCAC | 35.2 | 7.5 | 2.2 |
| OPAG-05-1900 | CCCACTAGAC | 35.2 | 8 | 2.1 |
| MC110a | RFLP cDNA | 35.2 | 8 | 2.3 |
| pB610a | RFLP: EMBL | 35.2 | 7.8 | 2.3 |
| MC112a | RFLP cDNA | 35.2 | 7.7 | 2.2 |
| OPA-15-0900 | TTCCGAACCC | 38.1 | 6.2 | 2 |
| OPAM-19-2200 | CCAGGTCTTC | 42.7 | 0.7 | 0.7 |
| OPAL-07-0580 | CCGTCCATCC | 42.7 | 0.7 | 0.7 |
| Vf | Consensus | 43.4 | | |
| M18 | RFLP | 43.4 | 0 | 0.7 |
| OPM-18-0900 | CACCATCCGT | 43.4 | 0 | 0.7 |
| OPU-01-0400 | ACGGACGTCA | 46.8 | 3.3 | 1.5 |
| PGM-1 | isoenzyme E.C. 5.4.2.2 | 49.4 | 6.4 | 2.1 |
| OPAG-12-0800 | CTCCCAGGGT | 54.0 | 9.9 | 2.4 |
| *'Fiesta'* | | | | |
| MH876a | RFLP genomic | 0.0 | | |
| OPAD-18-1130 | | 11.4 | | |
| UBC249-2000 | | 15.5 | | |
| MC014a | RFLP cDNA | 15.5 | | |
| AFLP_F1 | AFLP | 30.9 | | |
| MC110a | RFLP cDNA | 34.4 | | |
| OPD-07-1600 | TTGGCACGGG | 34.4 | | |
| pB610a | | 34.4 | | |
| TPI-5 | Isoenzyme E.C. 5.3.1.1 | 34.4 | | |
| MC112a | RFLP cDNA | 42.4 | | |

CAPS (Gianfranceschi *et al.* 1996) and the original RAPD OPM18-0900 marker, with no recombination events between the markers and the resistance locus (from a total of 145 plants for which the marker was scored), and OPAL-07 and OPAM-19 (Tartarini 1996), which showed one recombinant plant each (on totals of 150 and 155 plants respectively). OPU-01 (Koller *et al.* 1994) showed five recombinants (of 158 plants) and OPD-20-0600 (Yang and Krüger 1994) showed 15 recombinants (out of 153 plants). Recombination percentages and standard errors of the markers with the consensus score for resistance were calculated (Table 3.4).

Four RFLP loci provided marker bridges with the corresponding 'Fiesta' linkage group. An additional isozyme, RFLP, three RAPD and one AFLP markers were present on the 'Fiesta' linkage group, which spanned 42.4 cM. The isozyme locus *Tpi*-5 which was detected by the activity of triose phosphate isomerase, co-segregates with the RFLP locus pB610a, which derives from a cloned sequence which has homology to triose phosphate isomerase (personal communication, G.S. Ross).The most likely marker ordering of OPU1, OPM18, *Vf* and OPD20 is in agreement with the ordering of Gianfranceschi *et al.* (1996), who also located OPM18 and OPU1 on one side of *Vf* and OPD20 on the other side. Our marker ordering (for *Pgm*-1, OPU01, *Vf*, OPM18, OPA15, OPD20) also corresponds with results from Gardiner *et al.* (1996) in a cross between Granny Smith and the *Vf*-heterozygote A679/2, with results from Hemmat *et al.* (1995) (for PGM-1, OPU1, OPM18, OPA15 and OPD20) in another 'Prima' cross and with Tartarini (1996) (for OPAM19, OPAL07, OPC09, OPAB19 and OPC08). However, since only one recombinant plant was found among the markers OPAM19, OPAL07 and OPM18 and the consensus score of resistance, other orders for this subgroup cannot be excluded from this study, and are most likely to be resolved by physical mapping with large insert genomic clones, as a prelude to map-based cloning. The ordering is also in agreement with an analysis of 50 resistant cultivars (King *et al.*, unpublished).

The recombination frequencies between *Vf* and OPD20, OPC09, OPU01 and OPM18 (Table 3.4) are within the standard errors of those reported by Gardiner *et al.* (1996) which are based on field or glasshouse screening. The recombination frequency for *Pgm*-1 is in agreement with the pooled data from four populations reported by Manganaris *et al.* (1994).

## *Effect of relying on data from one source to establish linkage positions*

The consequences of using only one disease assessment data set to determine the relative map position of the resistance gene was investigated. Mapping was repeated, using the same parameters, with a set of field data selected from a scoring of Vi-F-1 at East-Malling in May 1994. This data set was chosen as being typical of a reliable scab screening occasion where more virulent inocula may be present (Roberts and Crute 1994), where the infection rate in the plot was high and

where no fungicide treatment had been applied at the site for several years. Vi-F-1 is a symptomatic scale which may be expected to be more appropriate for distinguishing susceptible from resistant plants on the basis of sporulation.

There were 18 differences from the consensus score. For 14 individuals the consensus score indicated resistance while the East Malling test set indicated susceptibility, and in four cases the consensus score indicated susceptibility while the test set indicated resistance. The ordering of the markers remained identical but the resistance gene could not be inserted between the markers, due to the 'double recombinants' with the most closely linked markers, which now had to be accounted for. Adding the resistance locus based on the East Malling test set resulted in a map position for *Vf* at the end of the linkage group. However, a large increase in the chi-square value indicated strong discrepancies between this ordering and the original pair-wise recombination frequency estimates.

## 3.4 Discussion

Compilation of data from widely differing environmental conditions affecting disease development has demonstrated the difficulties involved in the reliable and accurate assessment and classification of resistant and susceptible plants. The use of cluster analysis as a means of forming a consensus classification was successful in identifying the segregation of a major source of resistance. The classification, based on a model which assumed the action of a major gene segregating in a 1:1 ratio, was validated by its close agreement with the marker segregation. A poorer fit of the consensus score to the marker segregation would have been observed if the 1:1 model had been incorrect. If there had been distortion in the segregation of the resistance score whilst the flanking markers did not show any evidence of distortion this would have caused difficulties in determining the position of the gene. The symptomatic score (Vi-F-1) based on sporulation symptoms was clearly the most reliable in assigning resistance classes, although supplementing this with the incidence of sporulation proved worthwhile at sites where the data suggest a partial breakdown of resistance. The agreement in assessment of the field incidence scale by two teams of assessors at one site indicates that the scale is well defined and robust. In addition, preliminary scab records of presence or absence of sporulation from the nursery at Wageningen in 1991 agreed well with the consensus score.

The results presented in Table 3.3 demonstrate that plants carrying the *Vf* resistance gene can exhibit different symptom classes. In glasshouse tests at East Malling it was apparent that plants having chlorotic and necrotic spots possessed *Vf*. The fact that plants displaying the same symptoms in glasshouse tests at Angers did not possess *Vf* may reflect the different origin and level of inocula used. Gardiner *et al.* (1996) also observed that plants classified into two symptomatic resistant classes (3A and 3B *sensu* Chevalier *et al.* 1991) possessed the *Vf* gene; these classes

correspond to classes 3 and 4 of both Vi-GH-1 and Vi-F-1. The field scores reported in this study appear to indicate that expression of the *Vf* resistance gene is affected by the environment. The resolution of classification in the field may be increased by combining symptomatic and incidence scores. The scanning electron microscope and histological studies carried out by Chevalier *et al.* (1991) demonstrated that there was a spectrum of resistant classes (1-4) over which the intensity of the host response decreased, while fungal development increased.

Some care is required in the interpretation of the same scoring scale at different sites. These results suggest that the race-composition of the natural field inoculum and environmental conditions at some sites appear more appropriate for the reliable and accurate selection of plants carrying the *Vf* gene. The evidence from Italy and Greece where scab infection was low highlights the problems associated with reliance on field infection in breeding selection.

The effect of more virulent races being present in natural inocula at some sites was detected by the interaction between local scores and the consensus set. At Ahrensburg, where the virulent Race 6 has been identified (Parisi *et al.* 1993), 25 of the 26 plants present in the resistant consensus were scored with sporulating scab on at least one occasion. However, since there were also consensus susceptible plants with no sporulating scab, indicating that the infection rate at this site was sporadic, no quantitative assessment of partial resistance was possible. Even at this site there was a significant relationship between sporulation and the consensus score. At East Malling the results suggested the presence of a more virulent race or races which partially overcame the *Vf* resistance. The FL1 race described by Roberts and Crute (1994), which overcomes resistance in *M. floribunda* 821, was found in the vicinity of East Malling.

In some cases repeated symptomatic measurements were taken at one site and inconsistencies found in which trees displayed discrepancies in the scores. This suggests the discrepancies may not arise from a host genetic factor. Some of the discrepancies such as those found between a field score early and late in the season may reflect turnover of leaf material due to factors such as summer pruning, different timing of bud burst or growth rate.

Reliance on assessments of resistance or susceptibility solely from one site, or use of only one method, may be misleading when attempting to assign an accurate linkage position to a major source of resistance. This study demonstrates that a different linkage position could have been calculated if the analysis had been based on only one scoring method or one site, in this case probably due to the presence of more virulent inocula.

Selection for resistance to scab in apple breeding programmes currently relies on glasshouse assessment and culling of young seedlings, followed by more extensive field assessments against local inocula at a range of trial sites. This process may be inefficient due to practical limitations imposed on replicating young

seedlings in glasshouse tests, and the geographical and climatic environmental variation in field screening. The ability to preselect individual seedlings reliably and accurately based on markers flanking characterised major sources of resistance may contribute to cost and time savings in such breeding programmes, especially when combined with markers for other important agronomic traits. Gianfranceschi *et al.* (1996) have demonstrated the application of *Vf*-linked markers to molecular pre-selection of individuals possessing resistance. Roche *et al.* (1997a) have discussed the requirements of a marker assisted selection (MAS) approach to resistance breeding in apple, where markers are now available flanking sources of scab and mildew resistance, as well as aphid resistances. From the work reported here, it is apparent that there would be advantages to using markers where there are either low levels of natural inocula, where virulent races exist, or where environmental conditions or glasshouse facilities are not optimal for scab development. In situations where virulent races (such as Race 6) exist, the advantages of using *Vf* as a component either of developing more durable resistance, or of integrated disease management have to be evaluated against the tendency to encourage development of those virulent races. However, there is scope to use the markers to combine *Vf* with other non-allelic or recessive sources of resistance to scab, as well as to use co-dominant markers to identify individuals homozygous for *Vf*. There is evidence that such homozygotes display a stronger resistance response (Gessler *et al.* 1997).

The data presented not only establish an accurate ordering of markers linked to the resistance from *M. floribunda* 821, but also provide an insight into the efficiency of methods currently used to select for scab resistance in European apple breeding programmes. The results enable confident use of flanking markers in order to recover and determine the length of introgressed regions conferring scab resistance in breeding selection. The results also suggest that great care is required when relying upon existing screening methods for breeding selection. Reliance on one test in one environment may lead to loss of valuable material from any one breeding programme. The use of the markers reported here in combination with suitably segregating material will enable accurate assessment of the relationships between pathotypes of *Venturia* and different sources of resistance in *Malus*. The availability of an accurate and saturated linkage map of the region closely flanking a resistance gene is also a pre-requisite for map-based cloning.

***Acknowledgements***

The large number of data sets required in this study could only be collected through extensive collaboration between partners brought about by EU funding. The 'Development of the European Apple Crop' project was funded by the Commission of the European Communities from the AIR-3 programme (CT920473), with matching funding provided from national ministries of agriculture, research councils and the Swiss National Science Foundation. We are grateful to D.A.C. Pink and K. Tobutt for valuable comments and suggestions on the manuscript.

# Chapter 4

# RFLP and RAPD markers linked to the rosy leaf curling aphid resistance gene ($Sd_1$) in apple

P. Roche, F.H. Alston, C. Maliepaard, K.M. Evans, R. Vrielink, F. Dunemann, T. Markussen, S. Tartarini, L.M. Brown, C. Ryder and G.J. King[1]

## Abstract

$Sd_1$ is a dominant gene for resistance to biotypes 1 and 2 of the rosy leaf curling aphid, *Dysaphis devecta* Wlk. which can cause economic damage to apple trees. This report describes the identification of three RFLP and four RAPD markers linked to $Sd_1$ in a cross between the *D. devecta* susceptible variety 'Prima' *($sd_1sd_1$)* and the resistant variety 'Fiesta' *($Sd_1sd_1$).* Potted trees were artificially infested in the glasshouse, and the ratio of resistant:susceptible plants supported the hypothesis that the resistance was under the control of a single dominant gene. The position of the gene was mapped to a single locus on a 'Fiesta' chromosome within 2 cM of three tightly linked RFLP markers (MC064a, 2B12a and MC029b); four RAPD markers were located further away (between 13 and 46 cM). This is the first report of molecular markers for an aphid resistance gene in tree fruit crops. The potential application of these markers in a marker-assisted resistance breeding programme is discussed.

## 4.1 Introduction

In the UK, the rosy leaf curling aphid *(Dysaphis devecta* Wik.) is a pest of cultivated and ornamental apples. In the absence of control, the aphid typically affects the same trees year after year, causing severe leaf curl with conspicuous red galls (Massee 1954; Gratwick 1992). The occurrence of the aphid has been reported further afield, including Germany, Hungary, Iran and Italy (*e.g.* Pfeifer 1994; Jenser and Balázs 1991; Rezwani and Radjabi 1987; Baronio and Briolini 1985).

The aphid is capable of causing economic damage to apple crops although, in general, it is well-controlled by routine aphicide applications. However, economic, regulatory and environmental pressures are changing attitudes towards the use of agrochemicals, and alternative control measures are being sought. In the case of

[1] Theoretical and Applied Genetics (1997) 94: 528-533

*D. devecta,* sources of genetic host-plant resistance are readily available in apple varieties.

Resistance to *D. devecta* was first reported by Dicker (1954) who observed that the aphid did not attack the variety 'Cox's Orange Pippin'. Alston and Briggs (1968) found that resistance in 'Cox' and 'James Grieve' was controlled by a single dominant gene. In a later report, Alston and Briggs (1977) identified three aphid biotypes and three functionally distinct resistance genes *($Sd_1$, $Sd_2$* and *$Sd_3$*). 'Cox's Orange Pippin' carries the *$Sd_1$* resistance gene and is resistant to biotypes 1 and 2. In the same report, Alston and Briggs suggested that a number of varieties (including 'Cox') were heterozygous for a precursor gene *($Sd_{pr}$* at an additional locus, without which any one of the three known *Sd* resistance genes is ineffective.

Currently, selection for resistance to *D. devecta* is dependent on insect-plant bioassays in the glasshouse or field. These techniques are simple and can give well-defined, reproducible results within one month. However, there are problems associated with conventional screening, including inoculation failures or contamination with aphid predators, which can result in an extension of the seasonally dependent screening programme. A marker-assisted approach conducted independently of such constraints could eliminate these problems.

Through the European Apple Genome Mapping Project (King *et al.* 1991; King 1994, 1996), a saturated linkage map is being constructed of DNA and isoenzyme markers in apple using a cross between 'Prima', susceptible to *D. devecta,* and 'Fiesta', which is resistant, carrying *$Sd_1$* from 'Cox'. The results of an investigation into the association of the *D. devecta* resistance gene (*$Sd_1$*) with DNA markers in this progeny are described in this paper.

## 4.2 Materials and Methods

### *Mapping population*

The population used in this investigation results from a 'Prima' *($sd_1sd_1$)* x 'Fiesta' (*$Sd_1sd_1$*) cross made at CPRO-DLO, Wageningen by Dr J. Janse and J. Verhaegh in 1988. A seedling population was established in the field (CPRO-DLO, Elst) and used as a source of propagating material and leaf material for marker screening. In 1993, 161 bud-grafted trees on M27 rootstocks were sent to East Malling and planted as cordon rows in an unsprayed plot. Potted trees suitable for glasshouse testing were obtained by bench-grafting wood from these trees, and the parental genotypes, onto M9 root- stocks. Replicate sets of trees were distributed to the other sites (HRI-W, IZZ and DCA) where they were used as sources of material for molecular marker analysis.

## *Aphid samples*

Aphid populations were scarce at the start of the 1994 season, and aphids obtained from the Royal Horticultural Society's Garden, Wisley (Surrey, England) were supplemented with aphids from Boughton Monchelsea, a fruit-growing region in Kent. The aphid was more abundant in 1995, and all the aphids were collected from the principal 1994 source at Wisley. The Wisley aphids are thought to be biotype 1 based on tests with differential hosts ('Fiesta' and 'Northern Spy') in 1995, but the biotype of the Kent aphids is unknown.

## *Resistance testing*

The 1994 and 1995 aphid screening experiments were conducted on single plants of 141 individuals and their parents, 'Prima' and 'Fiesta', in unheated insect-proof glasshouses during May and June. A single adult aphid or two nymphs were placed on a growing point of each potted tree; the reaction was assessed after 1 week using the scoring procedure of Alston and Briggs (1968). Individuals were scored as susceptible (conspicuous galling and reddening of the leaf), intermediate (development of small chlorotic lesions) or resistant (no symptoms). Those plants without aphid colonies were re-tested in weeks 2 and 3.

The presence of *D. devecta* colonies on the trees in the unsprayed East Malling field plot was also recorded in both years.

## *DNA extraction and restriction fragment length polymorphism (RFLP) analysis*

At CPRO-DLO, DNA extraction and RFLP analysis were based upon the method described by Van der Beek *et al.* (1992) with minor modifications. Frozen leaf tissue (2.5 g) was homogenised for 30 s in 15 ml STE-buffer (0.35 *M* sorbitol, 0.1 *M* TRIS-HCl, 5 m*M* EDTA, pH 7.5 and 20 m*M* Nabisulfite). The homogenate was filtered, rinsed with 20 ml cold STE-buffer, and centrifuged (2000 *g* for 15 min). The green pellet was rinsed with 20 ml cold STE-buffer with 0.4% Triton X-100 and centrifuged again. After a further round of rinsing and centrifugation, the pellet was resuspended in 1.25 ml STE-buffer and then mixed with 1.75 ml nuclear lysis buffer (0.2 *M* TRIS-HCl, 0.05 *M* EDTA, 2 *M* NaCl, 2% CTAB w/v, pH 7.5) and 0.6 ml of 5% Sarkosyl. The mixture was incubated at 65°C for 1 h, with occasional mixing, before being extracted with chloroform:isoamylalcohol (24:1); the DNA was precipitated with an equal volume of cold isopropanol. DNA was hooked out and rinsed with 76% EtOH/10 m*M* NH4Ac for 30 min, dried and dissolved in TE buffer (10 m*M* TRIS-HCl, 1 m*M* EDTA, pH 7.5). The DNA concentration was measured using a fluorometer (TKO 100; Hoefer Instruments).

One hundred and fifteen *Malus* cDNA clones were evaluated in a set of plants which included 'Prima', 'Fiesta', 'Cox's Orange Pippin' and ten of the 'Prima' x 'Fiesta' progeny, using the restriction enzymes *Dra*I*, Eco*RI*, Eco*RV*, Hae*III*, Hin*dIII and *Xba*I*;* 55 of these clones were then used to score the progeny plants. In addition, one

genomic DNA clone, 2B12 from IZZ, was screened on these plants after an initial screening had suggested linkage to the resistance allele.

### *Random amplified polymorphic DNA (RAPD) analysis*

Leaf material was harvested from the parental and segregating accessions at the various sites (DCA, IZZ, HRI-W). DNA minipreps were done following the CTAB method of Doyle and Doyle (1990) with minor modifications. Purified DNA was dissolved in water and adjusted to a final concentration of 10 $\mu$g $ml^{-1}$. Random decamer primers were purchased from Operon (Alameda, Calif.) and Genosys (Cambridge, UK). A standard RAPD polymerase chain reaction (PCR) procedure was followed (Williams *et al.* 1990). Hybaid OmniGene thermal cyclers were used in all laboratories.

Amplification products were separated on agarose gels, stained with ethidium bromide, illuminated with UV light and recorded with either a digital camera system or on Polaroid film.

### *Linkage analysis*

Linkage analysis was performed using JOINMAP version 2.0 with the Kosambi mapping function (Stam 1993; Stam and Van Ooijen 1995). JOINMAP 2.0 allows linkage analysis in a segregating progeny from heterozygous parents of an outcrossing species so that markers with different segregation types (segregating 1:1, 3:1, 1:2:1 and 1:1:1:1) can be integrated into a linkage map, and linkage phases can be estimated simultaneously. A LOD score of 4.0 was used for grouping markers into linkage groups.

## 4.3 Results

The genotypes tested in 1994 were classified as either resistant, intermediate or susceptible. Those genotypes classified as intermediate in 1994 were re-classified as either resistant or susceptible after re-testing in 1995, when only resistant or susceptible symptoms were observed. Eleven susceptible genotypes were identified in the unsprayed field plot during 1994 and 1995, despite the low level of natural aphid colonisation in both years. These records are in full agreement with the glasshouse screening tests, where the same 11 genotypes were also found to be susceptible.

The field and glasshouse scores for 1994 and 1995 were combined. Of the 141 plants that were tested, there were 75 resistant and 62 susceptible plants while 4 plants remained unclassified since they appeared susceptible in 1994 and resistant in 1995. There was no evidence to suggest any statistically significant departure from a simple 1:1 segregation (chi-squared = 1.64, P = 0.27). When the data were prepared for linkage analysis, the four unclassified individuals were excluded. The analysis was repeated, assigning either resistant or susceptible scores to all the

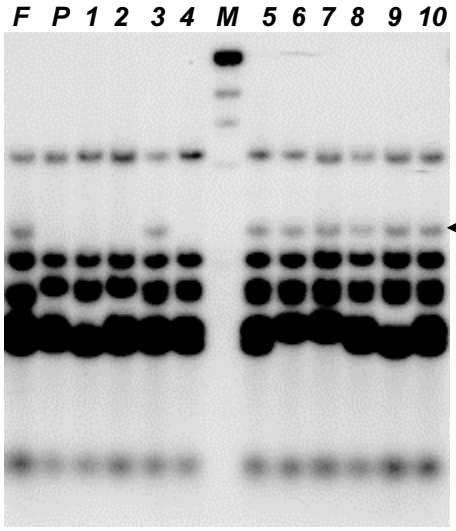



**Figure 4.1 Southern hybridisation autoradiograph of *Hae*III-digest DNAs probed with MC029. *F* resistant parent 'Fiesta', *P* susceptible parent 'Prima', lanes 1-10 ten progeny plants, *M* λ/*Hin*dIII molecular weight marker. Of the progeny plants, seven are resistant (lanes 3 and 5-10) and three are susceptible (lanes 1, 2 and 4). The 2.8-kb fragment linked to the *D. devecta* resistance is shown (⟵)**

unclassified plants in order to evaluate the influence of these genotypes on the map position of the gene.

Three RFLP markers and four RAPD markers were linked to the resistance gene, with recombination frequencies ranging from 0.015 to 0.346. One of the two RFLP markers obtained with MC029/*Hae*III had an <aaxab> segregation type (a = 4.7 kb and b = 2.8 kb). The 2.8-kb fragment, which was present in 'Cox' and 'Fiesta' and absent in 'Prima', was linked with the resistance gene (Table 4.1). This locus was denoted MC029b, and a sample autoradiograph is shown (Fig. 4.1). Two RFLP markers were also obtained for MC064/*Eco*RV; one of these loci (MC064a) gave three allelic fragments (a = 4.1 kb, b = 6.2 kb, c = 2.5 kb) with an <abxac> segregation type. The 2.5-kb fragment, present in 'Cox' and 'Fiesta' and absent in 'Prima', was linked with the resistance (Table 4.1). Clone 2B12 was first analysed at IZZ on a subset of 40 plants with *Hin*dIII, and the 4-kb fragment was found to be linked to the resistance gene. When analysed at CPRO-DLO with DraI, two RFLP markers were obtained. One of these markers (2B12a) displayed an <abxcd> segregation type (a = 3.2 kb + 2.0 kb, b = 2.5 kb, c = 2.5 kb + 2.0 kb, d = 2.5 kb + 1.5 kb). The 1.5-kb fragment, present in 'Cox' and 'Fiesta' and absent in 'Prima', was linked with the resistance (Table 4.1). The three other markers described for MC029, MC064 and 2B12 corresponded to unlinked loci. The four RAPD markers linked to the resistance, listed in order of increasing distance from the $Sd_1$ gene, were OPC-08-1700, OPT-09-1200, OPA-10-1000 and GE80-19-0550 (Table 4.1). In all cases, the RAPD fragment was present in 'Fiesta' and absent in 'Prima'.

The unclassified plants were coded as missing values for the resistance locus in the first analysis, and in 2 plants recombination events were identified between the resistance locus and two of the RFLP markers (MC064a and 2B12a, Table 4.1). These plants were also identified as recombinants for MC029b, together with an

**Table 4.1 Segregation analysis of the 'Prima' × 'Fiesta' progeny by molecular marker and reaction to *D. devecta*. The recombination frequency and standard error values are shown.**

| Locus | Segregation type | Banding pattern | Resistant | Susceptible | Number of plants | Recombination frequency | Standard error |
|---|---|---|---|---|---|---|---|
| MC029b | <aa × ab> | aa | 1 | 59 | | 0.022 | 0.013 |
| | | ab | 73 | 2 | | | |
| | | Total | 74 | 61 | 135 | | |
| MC064a | <ab × ac> | aa | 0 | 30 | | 0.015 | 0.010 |
| | | ab | 1 | 30 | | | |
| | | ac | 35 | 1 | | | |
| | | bc | 38 | 0 | | | |
| | | Total | 74 | 61 | 135 | | |
| 2B12a | <ab × cd> | ac | 0 | 30 | | 0.015 | 0.010 |
| | | bc | 1 | 29 | | | |
| | | ad | 35 | 1 | | | |
| | | bd | 38 | 0 | | | |
| | | Total | 74 | 60 | 134 | | |
| OPC-08-1700 | <aa × ab> | aa | 8 | 46 | | 0.147 | 0.032 |
| | | ab | 64 | 11 | | | |
| | | Total | 72 | 57 | 129 | | |
| OPT-09-1200 | <aa × ab> | aa | 12 | 47 | | 0.189 | 0.035 |
| | | ab | 56 | 12 | | | |
| | | Total | 68 | 59 | 127 | | |
| OPA-10-1000 | <aa × ab> | aa | 42 | 18 | | 0.276 | 0.040 |
| | | ab | 17 | 50 | | | |
| | | Total | 59 | 68 | 127 | | |
| GE80-19-0550 | <aa × ab> | aa | 26 | 40 | | 0.346 | 0.041 |
| | | ab | 47 | 20 | | | |
| | | Total | 73 | 60 | 133 | | |

additional plant. The marker genotype of the four unclassified plants for the three RFLP markers suggests that these plants are probably resistant, as was observed in the 1995 single-biotype trials. Including these four plants as resistant instead of missing made no difference in the marker order and minimal changes in the marker distances since no extra recombination events had to be accounted for. However, when these four plants were included as susceptible, the resistance gene was placed at the end of the linkage group but with a higher mean chi-square value for the resulting map, indicating more conflicts between the estimated map distances and the original pairwise estimates.

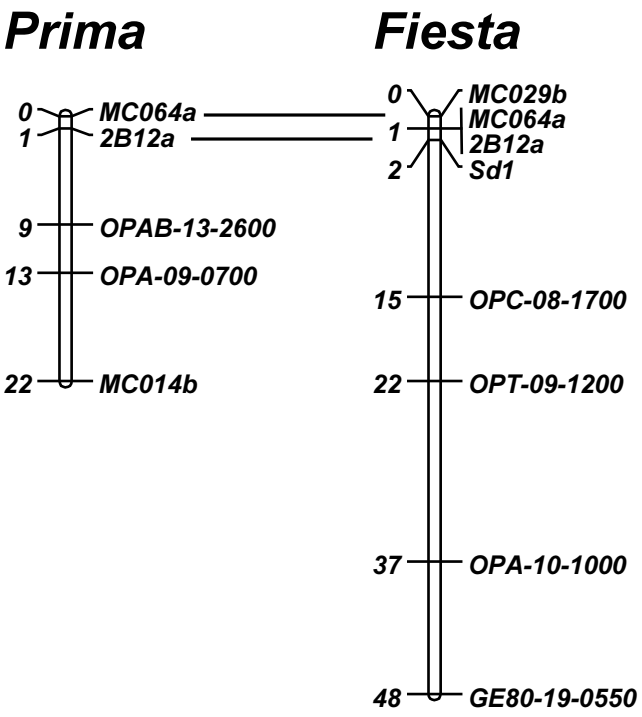



**Figure 4.2 Homologous 'Prima' and 'Fiesta' linkage groups with markers linked to the *Dysaphis devecta* resistance locus $Sd_1$. Distances in centi-Morgans**

Allelic bridges to the 'Prima' homologue of this linkage group were provided by the RFLP markers MC064a and 2B12a, for which both parents were heterozygous. Markers linked to MC064a and 2B12a on the 'Prima' homologue included RAPD markers OPAB-13-2600, OPA-09-0700 and RFLP marker MC014b. Figure 4.2 shows the 'Fiesta' linkage group with $Sd_1$ and the 'Prima' homologue; the four unclassified plants were coded as missing values for the resistance locus in this analysis.

## 4.4 Discussion

We have located the position of the $Sd_1$ gene for resistance to biotypes 1 and 2 of the rosy leaf curling aphid in the context of seven molecular markers (three RFLPs and four RAPDS); this is the first report of molecular markers for an aphid resistance gene in tree fruit crops.

Three RFLP markers have been mapped at less than 2 cM from $Sd_1$, each marker displaying a fragment which was linked in coupling phase to the resistance. In the best fitting marker order, the resistance gene was flanked by the RFLP markers on one side and the RAPD marker OPC-08-1700 on the other side.

This resistance gene has several characteristics which make it suitable for use in breeding programmes. Certain high quality varieties, including 'Cox' and some of its derivatives (*e.g.* 'Fiesta'), carry the $Sd_1$ gene. They are readily available to the breeder and have been widely used as parents in many breeding programmes. In addition, the resistance appears to be readily transferred to the progeny in a distinct, simply inherited fashion. In this particular study, the resistance gene segregated in a simple 1:1 fashion as would be expected if one of the parents was homozygous for the precursor gene $Sd_{pr}$. Nevertheless, many varieties are heterozygous for the precursor gene, without which the $Sd_1$ gene is ineffective (Alston and Briggs 1977). In a cross between two such heterozygous parents *($Sd_{pr}sd_{pr}$),* only 75% of the plants carrying the $Sd_1$ gene will be resistant since the remaining quarter will lack the precursor gene (*i.e.* $sd_{pr}\,sd_{pr}$ genotypes). This hypothesis may now be tested by using appropriate RFLPs to analyse a progeny where the precursor gene is segregating; it may then be possible to map the precursor gene.

Now that the $Sd_1$ gene has been mapped with closely linked markers, a marker-assisted selection (MAS) approach becomes a possibility. It has the potential to enhance the process of breeding for resistance to *D. devecta* by reducing the need for insect-plant bioassays; the magnitude of this benefit is dependent upon the precursor gene status of the parents at present. However, the identification of an appropriate molecular marker for the precursor gene could overcome this problem.

A molecular breeding approach is most appropriate with a range of markers that are cost-effective, reliable, robust, easy to score and suitable for screening large numbers of individuals. These factors may explain why marker-assisted selection techniques have been discussed widely in the literature but there are comparatively few practical examples. When compared to genes for resistance to diseases, there have been relatively few reports describing the mapping of genes for resistance to insects in plants. DNA markers associated with resistance to insects have been reported for several crops either as single genes, *e.g.* mungbean (Young *et al.* 1992) and rice (Mohan *et al.* 1994), or as quantitative traits, *e.g.* maize (Schön *et al.* 1993), potato (Bonierbale *et al.* 1994), barley (Nieto-Lopez and Blake 1994) and tomato (Nienhuis *et al.* 1987; Maliepaard *et al.* 1995). MAS can be restricted to a small part of a linkage group for a single dominant gene like $Sd_1$, where the expression of resistance is largely unaffected by environmental variation. Linkage drag can be controlled more easily under these circumstances, and the testing of extra progeny is often unnecessary.

In this study, linkage has been demonstrated in only one progeny but, if the most tightly linked RFLP markers prove to be reliable, for screening progenies they

also need to be easy to use and cost-effective. RFLP markers are not readily suited to routine screening programmes. However, the application of MAS techniques could be facilitated by converting candidate RFLPs into allele-specific PCR-based assays (*e.g.* Penner *et al.* 1995) which may be amenable to the automated processing of samples. These markers could then be used for the routine screening of progenies where the $Sd_1$ gene has been introgressed. Nevertheless, the economics of a marker-assisted screening strategy need to be evaluated and compared with the conventional glasshouse screening approach.

There are additional non-economic factors to consider; MAS has the potential to eliminate the problems associated with conventional screening experiments, which are complicated by the occurrence of different aphid biotypes, inoculation failures and the activities of predators of aphids. Indeed, differences in the aphid biotypes may explain the presence of the unclassified and "Intermediate" plants, since a supplementary source of aphids of unknown biotype was used in the 1994 experiments.

As the genetic distance between a marker and an agronomic trait increases, the breeder will become more inclined to confirm the MAS result with a plant-based test. However, the gene for resistance to *D. devecta* is associated with three tightly linked markers, and any individual which contains these markers could be selected with a reasonable degree of confidence. For routine screening purposes, it may be possible to rely on one marker, MC064a or 2B12a, for example, both of which correctly classified more than 98% of the 'Prima' x 'Fiesta' population. The MAS result may have to be confirmed by an insect-based test at a later stage of the breeding programme: bioassays would be vital for progenies where the $Sd_{pr}$ precursor gene may be absent in some individuals. The benefits of MAS become even more attractive when selection for more than one locus, for example, $Sd_1$ aphid resistance, *Vf* scab resistance (Gianfranceschi *et al.* 1996) and $Pl_1$ mildew resistance (Markussen *et al.* 1995) in apple, can be achieved by multiplexing of PCR reactions and gel separation.

It may be possible to devise a strategy that avoids total reliance on the $Sd_1$ resistance gene, which can be broken down by the rare biotype 3 (Alston and Briggs 1977). As markers for the $Sd_2$, $Sd_3$ *and* $Sd_{pr}$ genes become available, MAS techniques could be used to combine two or more functionally distinct resistance genes in a single individual. For example, pyramiding the $Sd_1$ and $Sd_3$ genes would confer resistance to all three reported aphid biotypes. However, the relative importance of these biotypes is unclear: 'Cox' has been grown widely in the UK for many years, and there have been no reports of the breakdown of resistance in the field.

As the linkage map for the 'Prima' x 'Fiesta' progeny becomes more saturated, further markers flanking the resistance gene may be identified. This could be a

useful start for linkage drag studies, map-based gene cloning and an increased understanding of the biochemical basis of the resistance mechanism.

***Acknowledgements***

We thank our colleagues for their help and support and J. Arbury and Dr A. Halstead at the Royal Horticultural Society's Garden, Wisley, and D. Vaughan of F.A.S.T. for providing aphid samples. The work of the European Apple Genome Mapping Project is funded as a shared-cost project as part of the CEC AIR-3 programme (project CT920473), with matching funding from government and other sources within the EC, and from the Swiss Federal Ministry of Technology.

# Chapter 5

# Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers

C. Maliepaard, F.H. Alston, G. van Arkel, L.M. Brown, E. Chevreau, F. Dunemann, K.M. Evans, S. Gardiner, P. Guilford, A.W. van Heusden, J. Janse, F. Laurens, J.R. Lynn, A.G. Manganaris, A.P.M. den Nijs, N. Periam, E. Rikkerink, P. Roche, C. Ryder, S. Sansavini, H. Schmidt, S. Tartarini, J.J. Verhaegh, M. Vrielink-van Ginkel and G.J. King[1]

## Abstract

Linkage maps for the apple cultivars 'Prima' and 'Fiesta' were constructed using RFLP, RAPD, isozyme, AFLP, SCAR and microsatellite markers in a 'Prima' × 'Fiesta' progeny of 152 individuals. Seventeen linkage groups, putatively corresponding to the seventeen haploid apple chromosomes, were obtained for each parent. These maps were aligned using 67 multi-allelic markers that were heterozygous in both parents. A large number of duplicate RFLP loci was observed and, in several instances, linked RFLP markers in one linkage group showed corresponding linkage in another linkage group. Distorted segregation was observed mainly in two regions of the genome, especially in the male parent alleles. Map positions were provided for resistance genes to scab and rosy leaf curling aphid (*Vf* and $Sd_1$, respectively) for the fruit acidity gene *Ma* and for the self-incompatibility locus *S*. The high marker density and large number of mapped codominant RFLPs and some microsatellite markers makes this map an ideal reference map for use also in other progenies and a valuable tool for mapping of quantitative trait loci.

## 5.1 Introduction

Although apple (*Malus pumila* Mill.) has been cultivated for centuries and is one of the main fruit tree species in the world, genetic studies and breeding have always been hampered by the long generation cycle, the space, time and cost involved in screening and maintaining populations, the high chromosome number (2*n*=34) and its outbreeding mode of reproduction. These same considerations have stimulated the interest in molecular genetics, since molecular markers provide tools to detect

[1] Theoretical and Applied Genetics (1998) 97: 60-73

genes for economically important traits and for early selection of these traits in breeding programmes.

Backcrossing is an unrealistic option for an outcrossing tree species due to the long juvenile period and self-incompatibility. Therefore, genetic analysis in apple is typically performed in the full-sib progeny of a single cross, which is also the base population for selection in breeding. Both parents of a cross are expected to display a high level of heterozygosity allowing markers to be found that are heterozygous in one or both parents. This type of analysis has been named double pseudo-testcross or two-way pseudo-testcross (Hemmat *et al.* 1994; Grattapaglia and Sederoff 1994). This terminology may, however, be confusing, since markers with a 3:1, 1:2:1 and 1:1:1:1 type of segregation (if both parents are heterozygous) can be used in addition to markers with a 1:1 segregation (if one parent is heterozygous).

Various types of markers are being used in studying apple genetics: markers for important genes were already provided by isozymes: AAT-1 (GOT-1) was shown to be linked to the self-incompatibility locus *S*, ACP-1 and ENP-1 to the pale green lethal gene *l*, LAP-2 to the mildew resistance gene $Pl_w$ and PGM-1 to the scab resistance gene *Vf* (Manganaris and Alston 1987; Manganaris and Alston 1988; Manganaris and Alston 1992; Manganaris *et al.* 1994). More recently random amplified polymorphic DNA (RAPD) markers were used to find closer linkage to *Vf* (Yang and Krüger 1994; Koller *et al.* 1994; Durham and Korban 1994; Gardiner *et al.* 1996; Tartarini 1996; Yang *et al.* 1997a, 1997b). RAPD markers were also found for the mildew resistance gene $Pl_1$ (Markussen *et al.*1995), $R_f$ for fruit skin colour (Cheng *et al.* 1996), *Tb* for terminal bearing, *Rbb* for initial bud break, *Rs* for rootsucker formation (Weeden *et al.* 1994) and *Co* for columnar tree habit (Hemmat *et al.* 1997). Linkage of restriction fragment length polymorphism (RFLPs) markers to *Vf* and $Sd_1$ for resistance to rosy leaf curling aphid was reported by King *et al.* (1998) and Roche *et al.* (1997a), respectively. Some of these markers have been transformed successfully into sequence characterised amplified regions (SCARs) or cleavage amplified polymorphic sites (CAPSs) which are more efficient than RAPDs or RFLPs for selection in apple breeding (Markussen *et al.* 1995; Gianfranceschi *et al.* 1996; Yang and Korban 1996; Yang *et al.* 1997a, 1997b; Roche *et al.* 1997b; Tartarini *et al.*1999). In a recent paper, characterisation of microsatellites in apple has also been reported (Guilford *et al.* 1997). One microsatellite marker was developed from a marker containing a simple sequence repeat (SSR), linked to the *Co* gene for columnar tree habit (Hemmat *et al.* 1997).

A first molecular genetic linkage map for apple covering over 400 markers was published by Hemmat *et al.* (1994). Fifteen homologous pairs of linkage groups were identified using markers that were heterozygous in both parents. Additionally, nine and six single linkage groups were formed for the respective parents. Since this map is mainly based on RAPD markers, it cannot easily be transported to other progenies. The relatively small progeny size is the likely cause that several parts of

linkage groups remained unlinked. Further RAPD based maps were constructed for three cultivars using two larger progenies (Conner *et al.* 1997) and homologies between all three maps were established for thirteen linkage groups. Isozyme and morphological markers also allowed the identification of six linkage groups homologous to the map of Hemmat *et al.* (1994). Since these maps are mainly based on RAPD markers, their usefulness in new progenies is very limited. Conner *et al.* (1997) therefore recommended using additional codominant markers.

This paper presents the construction of a linkage map containing more than 200 RAPD, isozyme, RFLP, SSR and amplification fragment length polymorphism (AFLP) markers. The map is one of the results of the European Apple Genome Mapping Project which was started in 1993 to study the genetics and genotype-by-environment interaction of a wide range of resistance, tree habit, fruit quality and fruit production characters (King *et al.* 1991; King 1996)

## 5.2 Materials and Methods

### *Plant Material*

A cross between the apple cultivars ‘Prima’ and ‘Fiesta’, using ‘Fiesta’ as the pollen parent, was made at CPRO-DLO, the Netherlands, in 1988. The petals of unopened ‘Prima’ flower buds were removed and, after pollination, the clusters were bagged to prevent foreign pollination. ‘Prima’ is a scab resistant cultivar carrying the *Vf* gene from *Malus floribunda* 821 and selections derived from ‘Prima’ are used in breeding programmes throughout the world to transfer this resistance. ‘Fiesta’ (syn. ‘Red Pippin’), derived from a cross between ‘Cox’s Orange Pippin’ and ‘Idared’, has a high productivity and good fruit quality and shelf life. The germination rate of the seeds was 73%. A seedling population of 161 trees was established and used as a source for propagation and leaf material for marker screening. Replicate sets of trees were propagated by grafting onto M27 rootstock. One set of grafted trees and the original seedlings were established in the field at CPRO-DLO. Other (sub-)sets were distributed to six sites in Europe (HRI East Malling, UK; HRI Wellesbourne, UK; IZZ Ahrensburg, Germany; INRA Angers, France; DCA Bologna, Italy and NAGREF, Naoussa, Greece) for DNA and protein extraction and phenotypic observations. On site, trees were re-propagated if this was necessary for resistance tests. The population was screened in greenhouse and field for resistance to diseases and pests, fruit quality and fruit production parameters and tree habit.

During marker analyses, mostly from isozyme and RFLP markers, it became evident that eight individuals in the mapping population were outcrosses. In addition, isozyme and RFLP analyses suggested that one individual was a triploid. These nine individuals were excluded from linkage analysis.

## *Markers*

*Notation*

The notation of segregating markers follows Maliepaard *et al.* (1997) and uses different characters to denote different alleles of a marker locus including *'0'* for a null-allele. Left of the '×' is the genotype of the mother, at the right the genotype of the father. Segregation types *ab*×*cd* and *ab*×*ac* are used for markers segregating into four marker phenotypes in the progeny, *ab*×*ab* is used for markers when parents are both heterozygous for the same two alleles and the expected segregation is 1:2:1 as in an $F_2$ of a selfing species. If a dominant marker is present in both parents and segregates in the progeny with an expected 3:1, the notation *a0*×*a0* is used for segregation of the fragment. Segregation types *ab*×*aa* and *aa*×*ab* are used for markers which are heterozygous in the female and male parent, respectively. Linkage groups are indicated with 'L' for linkage group followed by a number. 'Prima' and 'Fiesta' homologues of a pair of linkage groups are indicated with 'Pr' and 'Fi'. For example, L01-Pr refers to the Prima homologue of linkage group L01.

*RAPDs*

Over 600 decamer primers were screened for markers that were polymorphic between 'Prima' and 'Fiesta'. Primers were obtained from Operon (OP), Genosys (GE) and the University of British Columbia (UBC). Initial experiments were carried out to determine the reproducibility of RAPD marker generation and scoring between four laboratories. Identical reaction conditions and DNA samples were used, and all reactions were carried out in Hybaid OmniGene thermal cyclers (tube control). Following comparison of replicated experiments, it was found that primers which produced many or faint bands were inherently less reproducible. Primers producing a few very bright bands were used for systematic and reproducible mapping. It was also found preferable to select primers that produced one or more non-polymorphic bands flanking the polymorphic bands in the track, as these acted as internal controls of the efficiency of the amplification reaction (King 1994).

*RFLPs*

A cDNA library was made by Invitrogen Ltd. from leaf material collected from M27 rootstock trees. Clones from this library are indicated with MC (*Malus* cDNA). In addition apple cDNA clones were obtained from a library constructed by M. Lay Yee and selected as RFLP probes for apple at HortResearch. These are indicated as 'LY'. Genomic clones from two libraries constructed at Ahrensburg were used and are indicated with 2B10, 2B11 etc. (one as 7BC7) and 'MRC', respectively. At HRI, twelve genomic clones from a sub-genomic library, developed by S. Tartarini, were used, indicated with 'MH' (*Malus Hin*dIII). Known *Prunus* gene sequences Oleosin (EMBL accn: X78118) and Extensin (EMBL accn: X65718) as well as cDNA and genomic clones from almond ('AC' and 'AG', respectively) were also screened

(Viruel *et al.* 1995). Probes designated LBA from a *Prunus avium* cDNA library were kindly supplied by Dr. Paul Hand, HRI. Known gene sequences were also used as clones in RFLP analysis: pB610 (homologous to TPI=triose phosphate isomerase; Dr. G.S. Ross, personal communication), pAP4 (EMBL accn: X61390, ACC synthase), pAP79, pAP260 (kindly obtained from Dr. N. Weeden, Cornell University), pADH32 (kindly obtained from Dr. G.S. Ross, HortResearch, New Zealand) and pS6 (EMBL accn: D11080; NADP-dependent D-sorbitol-6-phosphate dehydrogenase; kindly obtained from Dr. Shohei Yamaki, Nagoya University, Japan). One clone also used for RFLP analysis was M18, kindly obtained from Dr. L. Gianfranceschi. This is a cloned fragment from RAPD marker OPM-18-0900, closely linked to *Vf*, from which Gianfranceschi *et al.* (1996) developed a CAPS marker.

At CPRO-DLO, DNA extraction, southern blotting, clone labelling and hybridisation were as in Roche *et al.* (1997a). Clones were first tested on the parents and a subset of 10 progeny plants, using six restriction enzymes (*Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, and *Xba*I). Clones were selected for the analysis of the entire progeny if the banding pattern fitted an *ab*×*ab*, *ab*×*ac*, or *ab*×*cd* type of segregation or if multiple markers could be scored. In many cases segregation of two marker loci was evident but both loci could not always be scored using a single restriction enzyme. Therefore, in a number of cases, two restriction enzyme/probe combinations were scored. In modification to CPRO-DLO, the enzymes used at HRI were *Eco*RI, *Xba*I and *Dra*I, whilst IZZ used *Eco*RI and *Hind*III.

In total, 245 clones were used to screen the parents and a small subset (usually 10 plants) of progeny for segregating bands with a number of restriction enzymes. Ninety-two clones were used to screen the larger progeny.

*Isozyme markers*

Isozyme analysis was performed on extracts from fresh leaves, in acrylamide and starch gels, according to protocols described in Chevreau and Laurens (1987) and Manganaris and Alston (1987). Staining was performed according to Wendel and Weeden (1989) except for catechol oxidase (CO) which was stained according to Schwennesen *et al.* (1982). Fifty-five markers from 21 enzyme systems (AAT, ACO, ACP, ADH, CO, DIA, ENP, EST, FDH, GDH, IDH, LAP, MDH, ME, PGD, PGI, PGM, PRX, SKD, SOD, TPI ) were screened in the progeny.

*Microsatellite markers*

Four microsatellite markers developed by HRI, were used to screen the progeny. A further five microsatellite markers were obtained from HortResearch (Guilford *et al.* 1997). One microsatellite marker, based on the SSR present in a marker for the *Co* gene, was also used (Hemmat *et al.* 1997). Using the specific primers and optimised annealing temperatures, PCR was carried out using $\gamma$-$^{33}$P ATP labelled forward

primers. The products were separated on 7.5M urea denaturing 6% polyacrylamide gels which were then dried and visualised following exposure against Fuji X-ray film.

*AFLP markers*

AFLP (Amplified Fragment Length Polymorphism) is a molecular marker technique based on selective PCR amplification of restriction fragments (Vos *et al.* 1995). 500 ng genomic DNA was digested with restriction enzymes *Eco*RI and *Mse*I. Restriction fragments were ligated with double strand adapters and a preamplification was done using the appropriate primers with one added selective nucleotide. The reaction mix was diluted 1/40 and 10 Tl was used for the final amplification with two primers each having three selective nucleotides. One of these primers was end-labelled with $\gamma$-$^{33}$P. As an initial study of the use of AFLP markers in apple, one primer combination was used: 5' GAC TGC GTA CCA ATT CAC A 3' (E35) and 5' GAT GAG TCC TGA GTA ACA A 3' (M47), the three selective nucleotides are underlined. Amplification products were separated on 5% denaturing polyacrylamide gels. The gels were dried and exposed to Kodak XOMAT AR X-ray film for 10-14 days at room temperature. Segregating AFLPs were coded as E35/M47/P1 to E35/M47/P6 for 'Prima' fragments and E35/M47/F1 and E35/M47/F2 for 'Fiesta' fragments in order of decreasing fragment size. One segregating fragment present in both parents was labeled E35/M47/PF1. Eighty-two plants were scored for presence or absence of these fragments.

*SCAR marker*

The BC226 SCAR, which was found to be linked to a gene for fruit skin colour, was used with no modification (Cheng *et al.* 1996).

## *Linkage analysis*

Single locus analysis, grouping of markers and mapping were performed with JoinMap version 2.0 which permits linkage analysis in outbred progenies involving markers with different segregation types (Stam 1993; Stam and Van Ooijen 1995). The Kosambi mapping function was used for the calculation of map distances. Markers were first divided into linkage groups using a LOD score threshold of 4.0. This LOD score calculated in JoinMap version 2.0 is based on the chi-square test for independence of segregation and is different from the usual LOD score in linkage analysis. The latter LOD score is affected by distorted segregation, while the test of independence is not. In this paper the LOD score is the JoinMap 2.0 LOD score unless specified differently. After being divided into linkage groups, markers were separated for the individual parents, using where possible only the alleles from that particular parent to calculate recombination frequencies. Obviously, this was not possible for *ab*×*ab* and *a0*×*a0* type markers. The *a0*×*a0* type markers were assigned to linkage groups but not used in the construction of the linkage map since such markers contribute little information to the map and recombination frequency

estimates obtained with such markers are typically inaccurate (Maliepaard *et al.* 1997).

In a few cases a marker was discarded during the mapping stage if its presence caused inconsistencies in the map. Such markers were identified by discrepancies (for multiple marker intervals) between the original distances with the final map distances after integrating all markers of a linkage group.

## *Monogenic traits*

Scab resistance was evaluated in field and glasshouse tests in 1993, 1994 and 1995 at sites in France, Germany, Greece, Italy, the Netherlands and the UK. A clear bimodality and a strong agreement with respect to the resistance classification was observed in a number of data sets. These observations fitted a 1:1 segregation, as was expected for *Vf* which is a single dominant gene. Cluster analysis was then used to combine data sets and divide the progeny into two clusters representing consensus resistant and consensus susceptible plants (King *et al.* 1996). Using this consensus score, the *Vf* gene for resistance to apple scab (*Venturia inaequalis* Cke.) was mapped as a single dominant gene.

Fruit acidity was evaluated using bromocresol and pH indicator paper to distinguish between fruits with higher and lower acidity. pH values < 3.8 were considered to indicate the presence of the *Ma* gene. Genotypes with fruit pH values > 3.8 were considered to be of the *mama* recessive genotype (Visser and Verhaegh 1978). The parents were assumed to be heterozygous *Mama*. In total, 140 genotypes were tested at one or more of three sites.

The gene $Sd_1$ for resistance to two biotypes of rosy leaf curling aphid (*Dysaphis devecta* Wlk.), present in 'Fiesta', was mapped on the basis of field and glasshouse resistance tests at HRI East Malling. The tests and mapping of this gene have been described in Roche *et al.* (1997a).

For mapping the self-incompatibility locus, the parents were screened using the allele specific primer method described by Janssens *et al.* (1995), indicating the 'Fiesta' alleles to be $S_3$ and $S_5$, and the 'Prima' alleles to be $S_2$ and an undetermined *S* type. Further investigation by W. Broothaerts (personal communication) revealed the undetermined allele to be $S_{10}$ and allowed the progeny to be scored with an additional allele specific marker, thereby detecting all four alleles. In addition to this method, an RNase assay was performed. RNase activity has been associated with the self-incompatibility reaction and was detected following electrophoretic separation of stylar extracts (Bošković and Tobutt 1996). For the assay, flowers were collected from the field at late balloon stage. Stigmas were dissected and styles collected in microfuge tubes. Stylar extracts were prepared and the protein products were separated by gel electrophoresis. RNase alleles were detected following the methods described by Bošković and Tobutt (1996). Four alleles segregated in the population. Of 51 individuals tested with both the RNase and DNA

allele specific assays, there was cosegregation for 50 individuals. For one individual, the DNA assay indicated an $S_2S_5$ genotype, whilst the RNase assay indicated $S_{10}S_5$. For the same set of 51 individuals there was complete cosegregation between the RNase assay and AAT-1.

## 5.3 Results

### *Marker segregation*

*RAPD markers*

In total, 168 RAPD markers from 88 primers were scored. Thirteen markers were present in both parents and an *a0*×*a0* (3:1) type of segregation was assumed. However, five of these markers were discarded due to difficult reproducibility and strongly skewed segregation. Two of the markers were scored on a limited set of progeny and no LOD scores larger than 4.0 were found. Therefore, only six of the *a0*×*a0* segregating markers could be assigned to linkage groups (Fig. 5.1). Eighteen RAPD markers with an *ab*×*aa* or *aa*×*ab* type of segregation were removed from the analysis for different reasons (strong segregation distortion, non-random distribution of recombinants with other markers, interaction with other markers from the same primer). Four more RAPD markers were discarded during mapping because of inconsistencies in recombination frequency estimates with other markers. In the end, 133 RAPD markers were positioned on the linkage map.

One primer, OPT-09, generated two bands with a codominant pattern of segregation: there were no individuals for which both bands were absent. The segregation of the combined marker did not deviate from the expected 1:2:1 ratio. Although this marker could be assigned to linkage group L05, including this marker in the map resulted in inconsistencies with other markers and the marker was discarded. Another primer, OPAC-15 generated two bands in the female parent (sizes 2000 and 2050 bp) for which always either one was present in the progeny, thus behaving as alleles.

Markers OPAE-01-1210 and OPAE-01-1190 were mapped to the ‘Prima’ and ‘Fiesta’ homologues, respectively, of linkage group L09 and possibly also represent two alleles of a single locus.

*RFLP markers*

In total, 124 RFLP markers from 86 probes were mapped. Thirty-seven probes generated one or more markers including one with an *ab*×*cd* or an *ab*×*ac* segregation type, 17 generated one or more markers including one with an *ab*×*ab* type of segregation. Thirty-two probes generated one or more markers with an *aa*×*ab* or an *ab*×*aa* type of segregation. Two probes (LY27 and MC023) yielded markers for three, another (MC228) for four loci.

*Microsatellite markers*

Ten SSR markers were used in this study. Each generated only a single marker locus; all ten markers were mapped. With the exception of two markers which are linked on L10, all SSR markers were mapped on different linkage groups.

*Isozyme markers*

Twenty-four segregating markers from 13 isozyme systems were scored in total. Seventeen isozyme markers from 12 isozyme systems were mapped. PRX-2 and PRX-3 showed no recombinants. Ten of the isozyme markers were mapped near the distal ends of the linkage groups (Fig. 5.1). With the exception of PRX-2 and PRX-3, markers from one system were always located on different linkage groups.

*AFLP markers*

Nine fragments showed segregation and were scorable: six 'Prima' markers, two 'Fiesta' markers and one marker present in both parents. Seven more fragments segregated but could not be scored reliably. An extra 60 bands were clear but did not segregate. No significant LOD scores were obtained for 'Prima' marker E35/M47/P3. A second 'Prima' marker (E35/M47/P6) showed distorted segregation and, although the marker could be assigned to L05, was discarded since no recombination frequencies smaller than 0.19 were obtained with any of the other markers. One of the 'Fiesta' markers was assigned to linkage group L05 but was discarded at the mapping stage due to inconsistencies in the estimated recombination frequencies. Two of the 'Prima' markers were linked in repulsion and showed no recombinants, thus behaving as alleles (E35/M47/P4 and E35/M47/P5).

*SCAR marker*

The BC226 marker was scored as the presence or absence of the $a^2$ allele segregating from the 'Fiesta' parent (Cheng *et al.* 1996). The other 'Fiesta' allele was $A^2$, as was the only detectable allele in the 'Prima' parent. As the $A^2$ allele was detectable in all screened progeny, it is reasonable to assume that 'Prima' is homozygous at this locus. The segregating allele is thought to be linked to the absence of red anthocyanin pigmentation, but is overridden in all progeny by the dominant presence of the pigment linked to the $A^2$ allele. This is consistent with red coloration of all the fruit in this progeny. BC226 was mapped to linkage group L09-Fi.

*Segregation types*

The segregation types of mapped markers are shown (Table 5.1). The numbers of markers indicate a higher level of heterozygosity for 'Prima' than for 'Fiesta', although this may be somewhat biased for the RAPD markers because of specific searches for markers in coupling phase with *Vf*. RFLP markers were preferably selected if they were heterozygous in both parents.

## *Single locus analysis*

Thirty-three of the marker loci displayed distorted segregation ($P < 0.05$; chi-square test), and are indicated with asterisks in Fig. 5.1. The 'Fiesta' and 'Prima' alleles showed distorted segregation in 21 and 10 cases, respectively, while in two cases *ab*×*ab* type markers showed segregation distortion, so that for these the parent generating the distortion is unknown. The distorted markers were found mainly in two regions on linkage groups L02 and L10. These were also the regions with the most severely distorted segregation ratios. As can be seen in Fig. 5.1, *e.g.* for linkage group L02-Fi, markers with a high significance value for the chi-square test occur in close linkage with markers with lower significance values. This is due to differences in segregation types and in the numbers of individuals genotyped. For instance, *MC116b* and *MC029a* on L02 both have an *ab*×*ab* segregation type and are tested against the 1:2:1 ratio whereas the other markers are tested against the 1:1 ratio. For *aa*×*ab* markers, a higher frequency of one of the 'Fiesta' alleles is directly observed in the higher abundance of either the *aa* or the *ab* genotype, whereas for *ab*×*ab* markers the distortion is observed only in overabundance of one homozygote at the cost of the other while the frequency of the heterozygote is as expected.

**Table 5.1 Segregation types of markers mapped in the 'Prima' × 'Fiesta' progeny**

| | *ab*×*cd* | *ab*×*ac* | *ab*×*ab* | *ab*×*aa* | *aa*×*ab* | Total |
|---|---|---|---|---|---|---|
| RFLP | 14 | 25 | 18 | 37 | 30 | 124 |
| RAPD | 0 | 0 | 0 | 80 | 53 | 133 |
| SSR | 4 | 5 | 0 | 0 | 1 | 10 |
| Isozyme | 1 | 0 | 0 | 6 | 10 | 17 |
| AFLP | 0 | 0 | 0 | 3 | 1 | 4 |
| CAPS-RFLP | 0 | 0 | 0 | 1 | 0 | 1 |
| SCAR | 0 | 0 | 0 | 0 | 1 | 1 |
| Total | 19 | 30 | 18 | 127 | 96 | 290 |

## *Map construction*

Initial grouping of markers, based on a JoinMap LOD score of 4.0, resulted in 16 linkage groups, all groups combining markers heterozygous in the female parent, the male parent and both parents. One large linkage group consisted of two subgroups (L01 and L12), held together by a single marker pair. Further investigation demonstrated that only 25 individuals had been scored for both markers, that the scoring of one of these markers was difficult and that the JoinMap LOD score of 4.2 deviated strongly from the traditional LOD score for linkage which was 1.3. No other markers suggested linkage of the two groups. The problematic marker was discarded and the two groups were separated so that 17 linkage groups remained.

These were divided into separate linkage groups for each of the parents. Sixty-seven markers (57 RFLPs, nine microsatellite markers and one isozyme marker) were heterozygous in both parents (Table 5.1) and allowed the identification of homologous pairs of linkage groups of the 'Prima' and 'Fiesta' parents. The seventeen pairs of linkage groups have been numbered from L01 to L17 (Fig. 5.1).

An integrated map was also calculated (not shown). In some instances there were differences in the estimated marker order between the individual parental maps and/or the integrated map. For instance, for linkage group L02 the orders of markers heterozygous in 'Prima' and 'Fiesta' were identical. However, the order of the 'Fiesta' markers in the integrated map deviated from the L02-Fi order, due to the differences in recombination frequency estimates in 'Prima' and 'Fiesta' for markers heterozygous in both. For linkage group L06-Pr the initial analysis resulted in a marker order with MC023 and MC034 swapped in comparison with the 'Fiesta' order. Using all markers from both parents resulted in the 'Fiesta' order. Using a JoinMap option to enforce a fixed order, the 'Fiesta' marker order was enforced for the 'Prima' linkage group. This resulted in only a slightly smaller likelihood. For L04 the marker order of MC013, NZ05g8 and MC019 was estimated differently in 'Prima' and 'Fiesta'. Enforcing the 'Fiesta' order for the 'Prima' map resulted in unresolvable negative marker distances, since many linked markers confirmed the originally estimated order. Enforcing the 'Prima' order on the 'Fiesta' map resulted in a solution which was slightly suboptimal but acceptable (mean chi-square less than 2.0), so that this order was chosen. For L01 and L11 the initially estimated marker order for the integrated map differed from both the 'Prima' and the 'Fiesta' order, but enforcing these orders for the integrated map resulted in a better solution.

## *Comparison of recombination frequencies*

The level of recombination was compared between 'Prima' and 'Fiesta' for non-overlapping segments of the linkage groups. For this, segments were used that are flanked by one or two *ab*×*cd* or *ab*×*ac* type markers and possibly on one side by an *ab*×*ab* type marker. To test whether recombination was smaller in 'Prima' than in 'Fiesta', the signed-rank test of Wilcoxon was applied (*e.g.* Lehmann 1975). For eight out of 48 segments the recombination frequencies were identical for both parents, so that the test was applied to 40 segments. In 12 cases the 'Fiesta' estimate was smaller than the 'Prima' estimate. For the remaining 28 segments the 'Prima' estimate was smaller. This indicated that in general recombination was smaller in 'Prima' than in 'Fiesta' ($P < 0.05$). However, there were some differences between linkage groups: on linkage groups L12 and L17 some recombination frequencies were smaller for 'Fiesta'. On linkage group L10 the 'Prima' recombination frequencies were smaller for the top (where the segregation was distorted for the 'Prima' alleles), whereas 'Fiesta' recombination frequencies were smaller for the bottom of this linkage group (where the segregation was distorted for

the ‘Fiesta’ alleles). On this linkage group some differences in the recombination frequency estimates were quite large. For example, the direct recombination frequency of 2E11 and MC227 was 0.20 for the ‘Prima’ alleles, whereas the recombination frequency in ‘Fiesta’ was larger than 0.5. However, in between these two markers there are several other markers confirming linkage.

## *Monogenic traits*

The *Vf* gene for scab resistance was mapped to the distal end of linkage group L01-Pr as described in King *et al.* (1996). The gene $Sd_1$ for resistance to (two biotypes of) the rosy leaf curling aphid (*D. devecta*) was mapped to the distal end of linkage group L07-Fi closely linked to three RFLP markers (Roche *et al.* 1997a).

Assuming a genetic model *Mama* × *Mama* for the fruit pH data the *Ma* gene for fruit acidity was mapped to the distal end of linkage group L16.

The self-incompatibility locus was mapped closely linked to AAT-1 on linkage group L17. The *S* combination $S_3$ $S_5$ for ‘Fiesta’ confirmed results by Batlle *et al.* (1995). The red fruit gene *Rf*, not segregating in this cross, can be assigned to linkage group L09 through the linked SCAR BC226. The gene for columnar tree habit, *Co*, can be assigned to linkage group L10 through the linked SSR marker USA-SSR11. The $Pl_w$ gene from the ornamental apple ‘White Angel’ for resistance to powdery mildew (*Podosphaera leucotricha*) can be assigned to linkage group L08 on the basis of previously detected linkage to LAP-2 (Manganaris and Alston 1992).

## *Duplicate markers*

Thirty-seven of the probes used in this study generated markers for two or more loci. Van Heusden showed that a majority of cDNA clones detects two loci in apple (unpublished).

It was not only observed that certain clones detected two loci on different linkage groups, but, moreover, that in six instances, a set of RFLP loci linked on one linkage group also showed a set of corresponding RFLP loci, from the same clones, linked on another linkage group (Fig. 5.1). On linkage groups L02 and L07 markers of clones MC029, MC064 and MC116 are linked. Markers MC019 and MC034 are linked on L04 and L06, MC040 and BQ7 on L04 and L14, while MC023 has markers on L04, L06 and L14. MC041, MC001, LY05 and MC221 are linked on L13 and L16. Eight markers linked on linkage group L05 have corresponding markers on linkage group L10. Markers on linkage group L14 have corresponding markers both on L06 and L12. The marker orders are not always identical, *e.g.* the order of seven markers is identical on L05 and L10, but LY29 is at different positions on these linkage groups. On L02 and L07 the order of MC029 and MC064 is reversed with respect to MC116, but here it should be taken into account that there are hardly any recombinants between the two markers, so that the likelihood of a reverse marker

order on either of the two linkage groups is only slightly smaller. In most cases of linked duplicate markers the map distances are similar.

### *Map length and density*

The total map length for 'Prima' is 842 cM, for 'Fiesta' 984 cM. Considering the numbers of markers in 'Prima' (194) and 'Fiesta' (163) this corresponds to a mean density of one marker every 4.3 cM for the 'Prima' map and one every 6.0 cM for the 'Fiesta' map. However, markers are not uniformly distributed over the maps. Comparison of the map lengths of linkage groups of the two parents and the positions of markers present in both parents suggests that part of the ends of L01-Fi (bottom), L03-Pr (top), L10-Pr (bottom) L11-Pr (bottom), L15-Fi (bottom) and L17-Pr (top) are missing. In addition, there are widely spaced marker intervals (> 20 cM) at L02-Pr/Fi, L03-Pr/Fi, L04-Fi, L05-Fi, L08-Fi, L09-Pr/Fi, L16-Fi and L17-Pr/Fi. An attempt to provide (RAPD) markers for two of these intervals (L05 and L16), using BSA based on the markers flanking the intervals, failed. No polymorphisms between the bulks were observed in 165 clear bands from 36 primers for L05 and 309 bands from 62 primers for L16.

## 5.4 Discussion

### *Marker detection*

Large numbers of segregating markers could be found with all marker types. RAPD primers yielded on average 2.0 scorable and segregating markers. About half of the isozyme markers tested yielded a marker which segregated in this progeny. From 133 'MC' RFLP clones only five did not generate a polymorphism with any of the six restriction enzymes (Van Heusden *et al.* personal communication). Where the RFLPs are concerned, some choice of the segregation type was possible, depending on the restriction enzyme used, and where possible the probe/enzyme combination which yielded the most informative type(s) of segregation (*ab*×*cd* or *ab*×*ac*) in the progeny was chosen. Nine of the ten microsatellite markers used also displayed the most informative segregation types. This illustrates that finding markers for mapping studies in apple is not difficult, as was also observed by Hemmat *et al.* (1994) and Conner *et al.* (1997). Chaparro *et al.* (1994) reported much lower levels of polymorphism (for RAPD markers) in peach, which is autogamous. Rajapakse *et al.* (1995) found 50% of the primers to produce polymorphic fragments in peach. Polymorphism in almond, which is only preferentially autogamous, was found to be much higher than in peach, but not so high as in apple (Viruel *et al.* 1995).

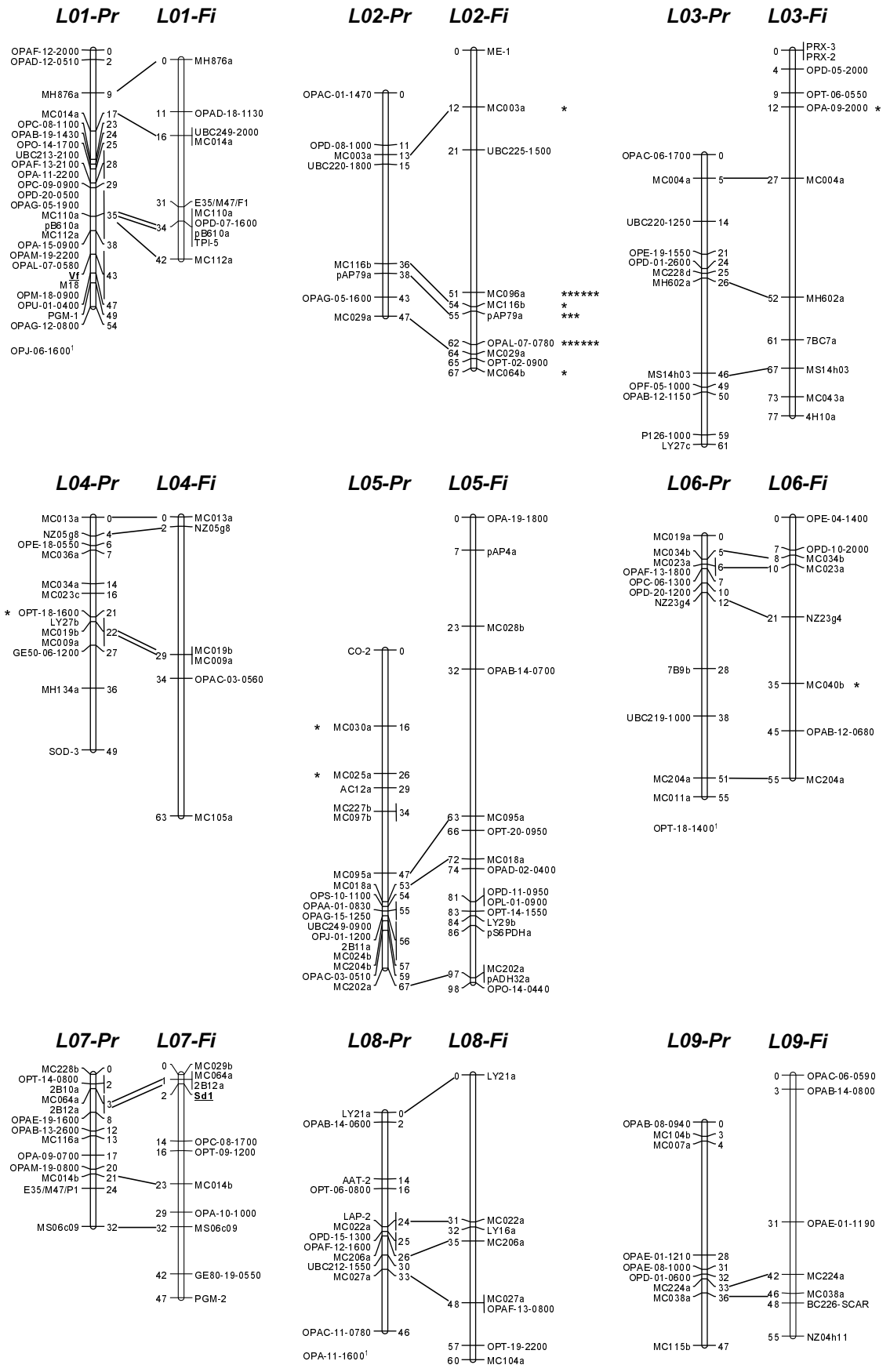
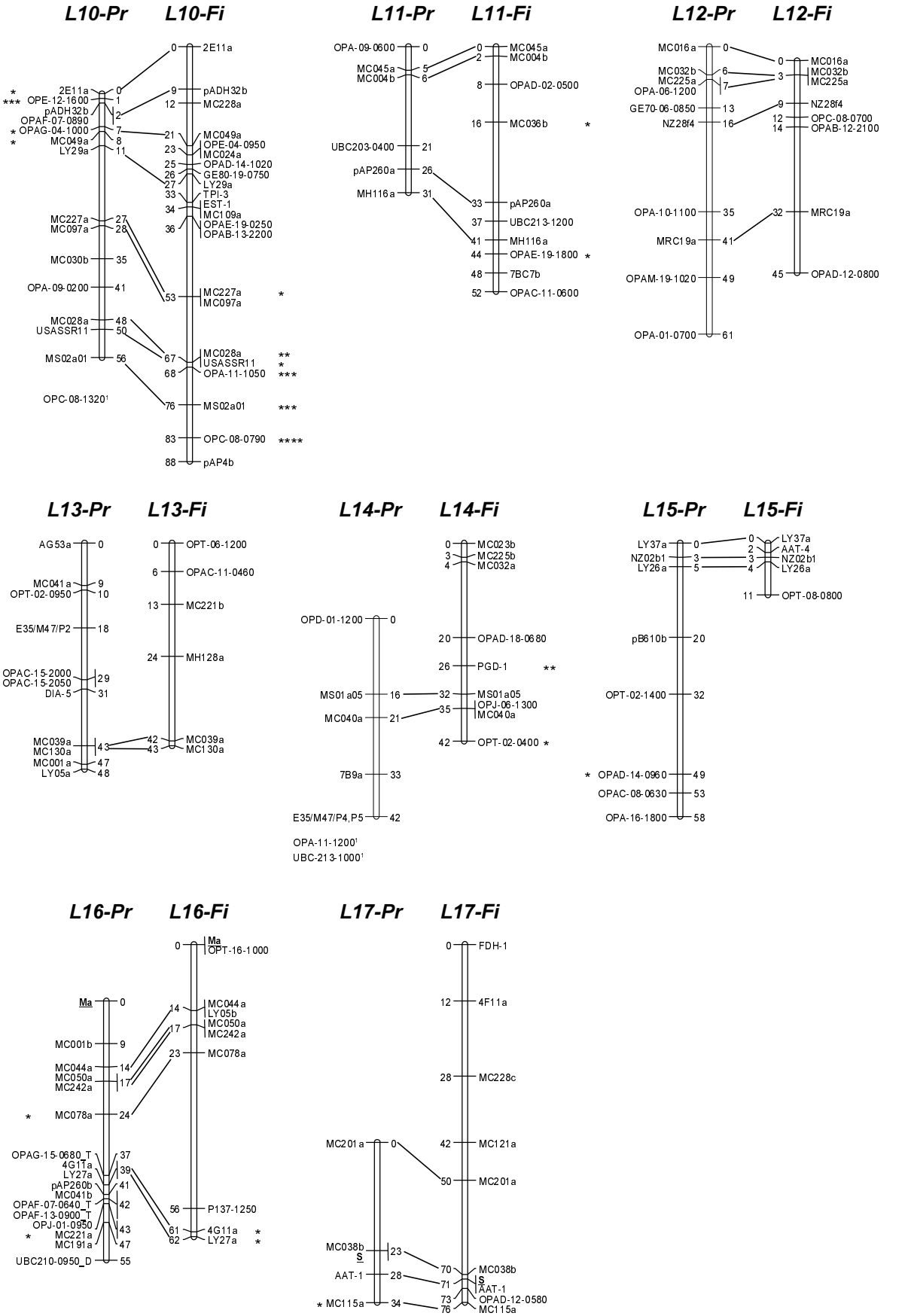



**Figure 5.1 Genetic linkage map of apple. 'Prima' (Pr) and 'Fiesta' (Fi) linkage groups are numbered from *L01* to *L17*. Allelic bridges are indicated by lines connecting Pr and Fi linkage groups.**




**Asterisks indicate distorted segregations of markers (chi-square test). * P=0.05, ** P=0.01, *** P=0.005, **** P=0.001, ***** P=0.0005, ****** P=0.0001. [1] *a0*×*a0* segregating markers were assigned to linkage groups but no map positions were determined.**

## *Map length and density*

This linkage map consists of seventeen pairs of homologous linkage groups from 'Prima' and 'Fiesta' and may well represent the seventeen pairs of chromosomes, considering the progeny size and the result that most markers which were genotyped on at least 50 plants could be placed on this linkage map. Markers which nonetheless could not be mapped usually displayed a complex banding pattern and/or skewed segregation ratios. In those cases, multiple bands may be overlapping. For at least one of the mapped markers (MC038) we also identified bands with similar sizes from two different loci.

The average map size per linkage group, 50 cM for 'Prima', 58 cM for 'Fiesta', is considerably smaller than the 100-150 cM which is commonly found in agricultural crops. Other tree species, such as *Eucalyptus*, *Theobroma cacao* and *Pinus radiata* also have considerably larger linkage groups (Grattapaglia and Sederoff 1994; Byrne *et al.* 1995; Lanaud *et al.* 1995; Devey *et al.* 1996). As indicated, some parts of linkage groups are probably missing. Discarded markers may also partly explain smaller linkage groups, but not to the extent as is observed here. The small map is in agreement with the 950 cM for the crab-apple 'White Angel' (Hemmat *et al.* 1994) and with map sizes in the range of 692 to 898 cM for the single parent maps in Conner *et al.* (1997). Moreover, it should be noted that the small linkage groups of apple are not unique within the Rosaceae: comparable sizes were observed and discussed for peach and almond, which yielded a total map length of 400 cM for eight linkage groups (Chaparro *et al.* 1994; Viruel *et al.* 1995). Dickson *et al.* (1992) reported that the nuclear DNA content of the Rosaceae is low among angiosperms, and although DNA content per map unit is known to vary widely across plant species (*e.g.* Nodari *et al.* 1993), the consistency between *Malus* and *Prunus* may indicate that in Rosaceae both DNA content and map size per haploid genome are small. However, a general conclusion cannot be inferred until more linkage maps of Rosaceous species have become available.

## *Integration of parental maps*

The availability of codominant markers allowed not only the identification of homologous linkage groups, but also the integration of both parental maps. For most linkage groups the integrated map was consistent with the marker orders in the individual parental linkage groups, although some minor differences with respect to marker orders were observed, usually with small differences in the likelihoods of these marker orders. However, where large differences of the recombination frequencies in both parents were observed, such as for L02, the integration of both parental maps caused marker orders to change. The integrated map combines markers segregating in one or the other parent with those segregating in both parents. For marker pairs heterozygous in both parents the combined recombination frequency estimate is an average over the recombination frequencies in the male

and the female meioses. This combined estimate may differ from the single parent estimates and thereby cause a changed marker order in the integrated map in comparison with the single parent map. Since differences in the estimated distances of both maps may reflect real differences in the recombination frequencies of both parents, these can best be presented separately. Both parental maps can be used separately to investigate quantitative trait loci (QTLs) segregating from a single parent. However, if QTLs may be present in both parents and for studying the different allelic combinations at QTLs, it is better to use the integrated map with an all-marker mapping approach (Knott and Haley 1992; Maliepaard and Van Ooijen 1995).

## *Duplicate markers*

The origin of the Maloideae subfamily of the Rosaceae has been subject to some debate (*e.g.* Chevreau *et al.* 1985; Chevreau and Laurens 1987; Phipps *et al.* 1991; Morgan *et al.* 1994). The basic chromosome number, x=17, suggests a polyploid origin since other Rosaceae have x=7, 8 or 9. Autopolyploidy was suggested by Darlington and Moffet (1930), but more recent studies showed only bivalent pairing of the chromosomes in meiosis (Lespinasse 1973) and monogenic or bigenic disomic inheritance of isozymes, in agreement with diploid behaviour of single or duplicated genes, supporting an amphidiploid origin (Chevreau *et al.* 1985; Chevreau and Laurens 1987; Weeden and Lamb 1987). Phipps *et al.* (1991) discussed an allopolyploid origin with Amygdaloideae (x=8) and Spiraeoideae (x=9) as ancestors. Morgan *et al.* (1994) provided evidence that at least one progenitor was a Spiraeoid and suggested that x=17 resulted from aneuploid reduction from x=18 either by autopolyploidy or allopolyploidy. Van Heusden *et al.* (personal communication) showed that a majority of cDNA clones corresponded to two loci with a disomic inheritance.

The present results demonstrate not only that duplicate RFLP markers are abundant in the apple genome, but also that in a number of cases, linked sequences on one linkage group, as detected by RFLP clones, can also be found at another linkage group. Considering the proposed polyploid origin of apple this is not unexpected. However, the results suggest differences in the amount of homology across linkage groups: L05 and L10, for example, have eight markers in common, suggesting sequence homology across large portions of these linkage groups. Other pairs of linkage groups do not show such a high amount of homology. L04 and L06, for example, share a pair of markers from two clones but other markers on L06 have counterparts on L14 and L05, while one marker has counterparts on both L04 and L14. Also, markers on L14 do not only have counterparts on L06 but also on L12. These results suggest that large chromosomal regions are conserved between and possibly also within the basic genomes underlying the apple genome. This is consistent with the proposed amphidiploid origin of apple and suggests that the

ancestors of the original hybrid were closely related. It is worth noting that Viruel *et al.* (1995) reported that a majority of cDNA clones in *Prunus* (x=8) identified single loci and that the (genomic and cDNA) clones that detected two loci did not provide evidence for the existence of duplicated chromosomes or of large duplicated chromosomal parts. The SSR markers used in this study did not detect multiple loci across the apple genome, but Guilford *et al.* (1997) observed that approximately 25% of their (GA)-repeat markers showed complex banding patterns consistent with two loci and they verified independent bigenic inheritance for one of these.

## *Mapping of monogenic traits and QTLs*

The apple map presented in this study combines different marker types and has, overall, a high marker density, although there still are a few regions with wide marker intervals. Map positions were provided for *Vf* for scab resistance, $Sd_1$ for resistance to rosy leaf curling aphid, *Ma* for the presence of malic acid and the self-incompatibility locus *S*. This map can be used for QTL mapping purposes and, considering the large number of codominant markers, especially RFLPs, can probably be transported to other progenies. A higher number of microsatellites per linkage group would greatly facilitate this. At present, microsatellite markers for apple are evaluated by the different mapping groups (Guilford *et al.* 1997; Ryder *et al.* personal communication; Weeden, personal communication).

## *Comparison with other maps*

There are some markers in common between this map and the maps published by Hemmat *et al.* (1994) for 'Rome Beauty' (RB) and 'White Angel' (WA) and by Conner *et al.* (1997) for 'Wijcik McIntosh' (WM), and the 'Prima' × 'Spartan' selections NY-67 and NY-58. Corresponding markers identifying probably homologous linkage groups are indicated in Table 5.2.

It can be concluded that our L01 corresponds to USA-L08, L02 to USA-L14, L03 to USA-RB5/WA8=USAnewL10 (RB-5 and WA-8 are now considered to be homologous and are coded here as RB-10new and WA-10new; RB-10 was found to be homologous to WA-5; Weeden, personal communication), L07 to USA-L09, L08 to USA-L07, L09 to USA-L03, L10 to USA-L06, L14 to USA-L02, L17 to USA-L01 (and maybe also to USA-L17).

In addition, RFLP clones pAP79 and pAP260 were used in both studies. One marker from clone pAP79 was mapped to linkage group L02 and may correspond to pAP79 on linkage group WA-13, although ME-1 suggests that our linkage group L02 corresponds to RB-14. pAP260 generated two markers, one on linkage group L11 and one on linkage group L16. One of these may correspond to pAP260 on WA-15 of Hemmat *et al.* (1994). The *Co* gene was mapped to linkage group 10 in Conner *et al.* (1997), corresponding to our linkage group L10. This map position of the *Co* gene

was confirmed in a 'Fiesta' × 'SA572/2' progeny which also segregates for the columnar tree habit.

Our map positions of *Vf* and *Ma* on two different linkage groups, L01 and L16, respectively, are in contrast with those of Conner *et al.* (1997), who placed *Vf* and *Ma* on a single linkage group. However, this was based on the inference of homology of a small WM linkage group and a NY-67 linkage group through only a single allelic bridge. Such a pairing should be viewed with caution, as was also pointed out by Conner *et al.* (1997).

## *Prospects for comparative mapping*

This map combining large numbers of RFLP, SSR, RAPD and isozyme markers offers possibilities for comparative mapping with other important Rosaceous genera, such as *Pyrus*, *Prunus*, *Rosa* and *Fragaria*. For isozyme markers, a first comparison between apple and pear has been made by Chevreau *et al.* (1997). At present markers are being exchanged among *Prunus* mapping groups and the European, New Zealand and USA mapping groups of apple to establish a basis for comparative genomic studies.

**Table 5.2 Markers in common between our map and the maps of Hemmat *et al.* (1994) and Conner *et al.* (1997) (microsatellite positions from personal communications).**
**WA= White Angel linkage group. RB=Rome Beauty linkage group.**

| Marker | Hemmat or Conner designation | Our linkage group | Hemmat or Conner linkage group |
|---|---|---|---|
| PGM-1 | Pgm-p1 | L01 | RB-8 |
| TPI-5 | Tpi-c2 | L01 | RB-8 |
| ME-1 | Me | L02 | RB-14 |
| PRX-2 | Prx-A | L03 | RB-5old/RB-10new[1] |
| MS14H03 | | L03 | WA-8old/WA-10new[1] |
| PGM-2 | Pgm-2 | L07 | WA-9 |
| AAT-2 | Aat-p | L08 | WA-7 |
| NZ04H11 | | L09 | 3[1] |
| BC226 | BC226 | L09 | 3 |
| EST-1 | Est-1 | L10 | WA/RB-6 |
| MS02A01 | | L10 | 6[1] |
| USA-SSR11 | Co SSR | L10 | 6[1] (10 in Conner et al.) |
| PGD-1 | Pgd-p | L14 | WA-2 |
| MS01A05 | | L14 | 2[1] |
| AAT-1 | Aat-c | L17 | WA-1 |
| FDH-1 | Fdh | L17 | WA-17 |

[1] personal communication from Dr. N.F. Weeden

## *Prospects for marker assisted selection*

Molecular markers for important characters in apple could greatly reduce the amount of space and time required for breeding. In an apple breeding programme, marker-assisted selection should be applied before seedlings are planted in the field. Preferably, multi-allelic reproducible markers should be used which require minimal amounts of DNA and a minimum of DNA isolation and purification steps, so that large numbers of plants can be screened in a short period of time. Considering these aspects, SCARs are the markers of choice at present and these have been developed for a number of traits. For *Vf* based scab resistance, SCARs are available. Although seedlings carrying *Vf* can normally be selected accurately with traditional screening methods, markers for *Vf* resistance may be highly efficient for pyramiding monogenic resistance genes or for combining *Vf* with genes for partial resistance. These markers may also be useful for selecting for *Vf* resistance in the absence of the pathogen, *e.g.* if no inoculum is present in certain years or at some location, or in the presence of races which have overcome *Vf*, such as race 6 (Parisi *et al.* 1993). Furthermore, these markers can be useful for eliminating the portion of *Malus floribunda* genome around the resistance gene and for identifying *Vf* homozygotes.

In contrast with scab, resistance to rosy leaf curling aphid can more readily be scored with markers than in field or greenhouse tests, but the economic importance of the pest is relatively small. Markers for fruit characters would have a high efficiency since these phenotypic traits can be evaluated only after five or six years, once the tree has passed through its juvenile phase. For instance, a marker specific for the presence of *Ma* could save considerable costs in apple breeding programmes where many crosses involve two heterozygous parents. In such crosses an expected quarter of the progeny yields tasteless fruits without malic acid. A marker for *Ma* would allow selection at an early seedling stage before field planting and could save up to six years of tree care. Markers can also be used for other purposes than selection: Janssens *et al.* (1995) suggested that the *S*-allele specific markers may be helpful in the choice of fully compatible pollinators to provide optimal fruit set for new cultivars.

This linkage map provides an important tool not only for the detection of simple major genes but also for more complex QTLs and for providing breeders and researchers with markers for these genes. The map may also be helpful in studying the interaction of genes or establishing evidence of allelism. In the future, this map may provide an essential tool for map-based cloning techniques. Improvement of popular apple cultivars with a single gene, impossible with conventional breeding techniques, is likely to become possible once important genes, such as scab resistance genes, have been cloned.

Ideally a reference linkage map should contain at least a backbone of codominant markers, such as RFLPs or SSRs which are reproducible, can be

transported to another progeny and which can then be supplemented with RAPD or AFLP markers to saturate the more interesting regions of the genome. The present map fulfils these requirements and is therefore an ideal core map for apple genetic research.

***Acknowledgements***

We gratefully acknowledge our colleagues Dr. N. Weeden from Cornell University, USA, Dr. P. Arús from the European *Prunus* mapping group, Dr. G.S. Ross from HortResearch, New Zealand, Dr. P. Hand from HRI, UK, and Dr. L. Gianfranceschi from ETH Switzerland, for providing markers. We acknowledge Dr. J. Brace for early work on RFLP optimisation at HRI, Dr. W. Broothaerts (KUL, Belgium) for help with the *S*-locus allele specific assays and R. Bošković (HRI) for testing with RNase isozyme. This work, part of the project 'The Development of the European Apple Crop' was funded by the EU (AIR-3: CT920473) with matching funding from the national ministries of agriculture, research councils and the Swiss National Science Foundation.

# Chapter 6

# Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.)

G.J. King and C. Maliepaard, J.R. Lynn, F.H. Alston, C.E. Durel, K.M. Evans, B. Griffon, F. Laurens, A.G. Manganaris, E. Schrevens and S. Tartarini[1]

## Abstract

Texture is a major component of consumer preference for eating quality in apple. A quantitative genetic analysis of traits associated with fruit flesh firmness was carried out. This was based on segregation in an unselected mapping population replicated in six sites and harvested over two years. Different methods of assessment were compared, and a principal components analysis carried out. Instrumental measures used were Magness-Taylor penetrometer readings, stiffness by acoustic resonance, and a range of sensory descriptors assessed by a trained panel. There were good correlations between some measures, although stiffness was correlated poorly. Whilst genotype by environment effects were large, significant effects were attributable to the genotype, and these were used to detect QTLs. Significant QTLs were detected on seven linkage groups, with large effects on linkage groups L01, L10 and L16. Whilst there was poor correlation between acoustic stiffness and other measures, the significant and suggestive QTLs detected for stiffness on linkage group L10 did represent a subset of significant QTLs detected for the penetrometer measure. The use of sensory assessment proved valuable in detecting QTLs representing different attributes of fruit texture. The possibility of interaction between significant QTLs for fruit texture and other strongly selected traits such as scab resistance and fruit acidity is addressed.

## 6.1 Introduction

Fruit texture is a major component of consumer preference for eating quality, with appearance and flavour being the other major components. There is considerable demand for high quality, firm apples that will maintain or achieve their optimal

[1] Theoretical and Applied Genetics (2000), in press. Reproduced with permission of Springer-Verlag, Berlin. G.J. King and C. Maliepaard are recognised as jointly contributing to first authorship.

texture properties following harvest, storage, distribution and retailing. In tests conducted by marketing inspectors at harvest, and in breeding selection programmes, texture has often been equated with quality in apples, with firm, crisp apples being the ideal. Existing cultivars vary considerably in their innate textural properties, and in their response to various environmental factors. However, it is apparent from breeding programmes, sensory trials and universal anecdotal observation that there is a considerable cultivar-specific, and therefore genetic contribution to fruit texture.

Several approaches have been developed to quantify the variation in apple fruit texture, for breeding selection and as an indication of fruit maturity in the production and marketing chain. The fruit of apples consist of an epidermal layer, covering a relatively uniform cortex of parenchyma cells which surrounds a complex inner core of the pericarp and seed. Assessment of texture tends to be focused on the cortex tissue, although this may be influenced by skin, core or fruit size depending on the technique adopted. The physical methods developed include penetrometer readings and non-invasive acoustic resonance measurements to assess fruit stiffness. Sensory evaluation has been carried out either by trained or untrained panels, or by individual breeders and growers.

The sensory evaluation of apple fruit has played a significant role in pre-release trialling and in the comparison of modern and existing cultivars. The perception of fruit flesh texture has been studied in various sensory trials as comprising components of firmness, chewing response and juiciness (Watada and Abbott 1985; Daillant-Spinnler *et al.* 1996). Due to the large contribution of environmental factors in fruit texture variation, such studies are often difficult to interpret where fruit originate from different orchards or regions, where they differ in harvest time, or when different or untrained sensory panels are employed. Prior to the work reported here, a preliminary study with a trained and stable sensory panel was carried out in 1994 and 1995 (Daillant-Spinnler *et al.* 1996). This demonstrated that the panel was able to rank different cultivars consistently, and use an agreed vocabulary and scoring scale. The same panel was used in the present study.

A range of physical measures has been developed in an attempt to reflect and standardise the sensory perceptions of fruit texture. Apple firmness has traditionally been measured as the maximum force required to push a manually operated Magness-Taylor (MT) fruit firmness probe (penetrometer), of specified shape and with an 11 mm (in some market regions, 8 mm) diameter tip, 7.9 mm into peeled cortex tissue on opposite sides of the fruit equator (Magness and Taylor 1925; Bourne 1974; Lehman-Salada 1996). Force is applied perpendicular to the cut surface in a smooth motion in 2 to 3 sec, with depth of penetration in manual tests controlled by the operator. The MT measurement is accepted as the standard firmness measurement in the apple industry (Abbott 1994). The data obtained are expressed in terms of the force required to rupture cortex parenchyma cells, and

thus represent a compound of many cellular and macro-cellular properties including cell turgor and wall strength. Many variables affect the reliability of destructive firmness measurements. Lower readings are associated with slow insertion speed, shallow probe penetration or with apples that are large, water-cored, bruised or warm (Blanpied *et al.* 1978). Due to the relative simplicity of obtaining data with hand-held MT penetrometers, readings are also used as a measure of maturity and ripeness in commercial production.

Techniques for the non-destructive measurement of apple texture have been developed using sonic or vibrational methods (Abbott *et al.* 1992, Chen and De Baerdemaeker 1993). The indices these detect are essentially the coefficient of elasticity of stretching, or stiffness of intact fruit. Stiffness may thus be assessed in a non-destructive manner by measuring the acoustic resonance of fruit which have been struck with a light object (Chen and De Baerdemaeker 1993).

To date there have been relatively few studies relating to the genetic basis of the firmness of apple, or indeed for any fruit. Genetic linkage maps for *Malus* are now available (Hemmat *et al.* 1994; Maliepaard *et al.* 1998). The European Apple Linkage Map (Maliepaard *et al.* 1998) was constructed from segregation data of 290 markers scored on 152 individuals from a replicated reference mapping population. The population derives from a cross between the cultivars ‘Prima’ and ‘Fiesta’.

Data sets were accumulated over two years from six different sites, in order to maximise the possibility of detecting genotype × environment interactions. The individuals were grown on six sites in five countries, covering 13 degrees of latitude (King 1996). The experiment design allowed comparison of two different physical methodologies, comparison over different sites and comparison of physical measurements with sensory perceptions. Quantitative trait analyses were carried out to study the genetic basis of fruit firmness in apple, by identifying loci in the context of the existing linkage map.

## 6.2 Materials and Methods

### *Plant material*

A cross between ‘Prima’ and ‘Fiesta’ was carried out at CPRO-DLO, Wageningen in 1988, using ‘Prima’ as the female parent. Unselected seedlings were raised in pots, and planted in the nursery at Wageningen in the winter of 1990/1991, and later planted on their own roots at Elst. Six replicate trees of each genotype were obtained by bud-grafting wood of 152 seedlings onto M27 dwarfing rootstock in early 1992. The trees were grown for another year. One set was grown on rootstock at Elst and the remainder were distributed to the five additional sites in early 1993 (Table 6.1). At Cadriano only one-half of the population was grown.

**Table 6.1 Locations of field sites for harvesting fruit from the replicated progeny of the 'Prima' × 'Fiesta' family.**

| Site | Country | Location | | Propagation |
|---|---|---|---|---|
| | | Grid reference | Altitude (m) | |
| Wellesbourne | England | 52°12' N 01°36' W | 49 | M27 staked |
| East Malling | England | 51°17' N 00°27' E | 32 | M27 cordons |
| Elst | The Netherlands | 51°55' N 05°50' E | 8 | M27 staked own roots |
| Angers | France | 47°30' N 00°35' W | 57 | M27 staked |
| Cadriano | Italy | 44°32' N 11°23' E | 30 | M27 staked |
| Naoussa | Greece | 40°37' N 22°07' E | 121 | M27 staked |

The trees were grown either in rows 2 m apart or at East Malling in cordons, 1 m apart. Up to five replicate trees of the parents were planted at each site, but these replicate trees were often not assessed separately.

The relative positions of the trees was randomised between Elst, Naoussa, East Malling and Cadriano. At the remaining sites, the trees were in the same order as the original seedlings at Elst. Normal cultural practices were followed. Trees were sprayed with insecticides and fungicides during the years that they were being evaluated for fruit characters, to prevent fruit quality being affected by pests and diseases.

## *Harvesting and sampling*

Fruit were harvested up to twice a week as they reached the tree-ripe stage in 1995 and 1996 (Table 6.3). Fruit which were damaged by hail, insects or birds, or which were asymmetrical, were rejected. In general, fruit of a median size were picked, with selection against particularly small or large fruit. Fruit were over-sampled in the field, to allow for possible damage or decay in transit, and were placed in standard apple trays, with smaller fruit wrapped individually in tissue paper. Where the apples were not assessed at the site where they were harvested, they were stored at 4 °C for up to 3 days, and transported by air or land courier as required.

For penetrometer measurements, fruit from Wellesbourne were assessed at East Malling. For other sites, fruit were assessed locally.

For acoustic resonance measurements, in 1995 fruit from CPRO, the Netherlands, were collected from the trees on their own roots and sent for assessment in Leuven. In the 1996 trial, fruit-set data across all sites were collated in June and a sampling strategy devised. Due to limitations on sample-throughput for the acoustic resonance measurements, we maximised population coverage from two sites (Wellesbourne and Angers, Table 6.3). To obtain data indicative of the genotype × environment (G×E) effect, eight fruit from up to ten individuals replicated from six site/year combinations were selected.

For sensory evaluation, fruit from East Malling and Wellesbourne were sent to Sensory Dimensions, Reading, England. In 1995 two batches of fruit, and in 1996 eight batches of fruit, were sent. Each batch consisted of 16 separate genotypes per batch and included reference population segregants and fruit of both parent cultivars. The different batches were assessed separately.

## *Trait assessments*

*Penetrometer readings*

A small area (up to 15 mm diameter) of peel was removed from each apple fruit with a knife. Fruit was then placed on a stand, and the resistance of the flesh determined using a mechanised 8 mm MT probe (FFF-1, Table 6.2). At Angers an automated electronic penetrometer (Penelaupe) was used with an 11 mm (FFF-2) MT probe. Resistance was expressed in g force. The arithmetic mean of readings from three separate fruit was recorded.

**Table 6.2 Methods used to assess fruit firmness and texture. Sensory descriptors of first-bite texture and texture during chewing were scored on a scale 0-100 based on an arbitrary scale of nil to extreme.**

| Class | Descriptor | | Units |
|---|---|---|---|
| Firmness | FFF-1 | Fruit flesh firmness by penetrometer with 8 mm probe; 3 readings | kg |
| | FFF-2 | Fruit flesh firmness by penetrometer with 11 mm probe; 20 fruit | g |
| Stiffness | FST-RES-1 | Fruit stiffness determined by acoustic resonance | $Hz^2 \cdot g^{2/3}$ |
| First bite texture | Hardness | Take one bite from the segment with front teeth | nil to extreme |
| Texture during chewing | Crispness | How crisp the apple seems during chewing - hard but brittle, makes a characteristic crunchy noise when chewing. | nil to extreme |
| | Granularity | Disintegrates into small granules when fruit is chewed | nil to extreme |
| | Spongy texture | Pulpy/fluffy type of texture | nil to extreme |
| | Slow breakdown | Speed of breakdown of apple flesh (ignoring skin) in the mouth until state ready for swallowing | nil to extreme |
| | Juiciness | Amount of juice produced during chewing | nil to extreme |
| | Overall liking | How much the apple was liked overall | nil to extreme |

*Acoustic resonance*

The acoustic resonant frequency of fruit was determined using the Acoustic Response Technique (Chen and De Baerdemaeker 1993) and recorded in Hz; the mass of each fruit was determined and the stiffness calculated from Stiffness = (Resonant Frequency corrected for size/weight) [$Hz^2.g^{2/3}$]. Readings were taken on receipt of the fruit (day 0) and at 7, 14 and 21 days during storage at 20 °C, to reflect shelf-life changes which would occur during transport or retailing of the fruit.

*Sensory analysis*

The fruit were evaluated by a trained sensory panel of eleven individuals which had been used in the two previous years. As part of an initial standardisation process the panellists scored a comprehensive range of attributes on fruit from twelve commercial varieties. These scores were subjected to principal component analysis, as well as analysis of and a non-parametric test of rank interaction. Following this, a non-redundant set of descriptor terms was selected such that each descriptor showed significant differences between the varieties. Analysis of the panellist x variety interactions showed that although interactions were present for most of the descriptors, the interactions never affected rank order (Daillant-Spinnler *et al.* 1996). It was therefore concluded that the mean scores for each variety given by the panellists for each descriptor could be considered satisfactory estimates of the sensory profiles of the varieties.

Each fruit was peeled and then assessed for a number of attributes relating to internal odour, first-bite texture, internal appearance, texture and flavour during chewing. Descriptors for first bite texture and texture during chewing were included in this study, along with an overall liking score (Table 6.2).

In 1995, the samples included ‘Prima’, ‘Fiesta’ and 27 segregant genotypes harvested from Wellesbourne. In 1996 the samples included ‘Prima’, ‘Fiesta’ and a total of 105 genotypes harvested from East Malling or Wellesbourne, of which 7 were sampled from both sites. The total number of genotypes sampled was 115. Each descriptor was scored individually by 11 panel members on a scale of 0-100, denoting nil to extreme. The mean of these 11 scores was then taken. Due to the limited sample availability, one tasting of each sample was made by each assessor. Fruit from individual replicate trees of each parent, ‘Prima’ and ‘Fiesta’, were kept separate.

*Statistical analysis*

The trait data were analysed using REML (Patterson and Thompson 1971) in the statistical package Genstat 5 (Payne *et al*. 1993). REML is an analysis of variance also suitable for unbalanced designs. Data were analysed first with all factors as random to obtain the variance components. Estimates of the proportions of the variance contributed by various factors were calculated as the ratio of the variance

**Table 6.3 Assessments of fruit texture carried out on the replicated progeny of the cross between 'Prima' × 'Fiesta' at different sites and years. Fruit for FST-RES-1 were transported to Leuven, Belgium for assessment. Fruit for sensory measures were sent to Reading, UK for assessment. Note that the numbers of fruit per genotype given are the largest in a sample. Some genotypes yielded fewer fruit than this.**

| Descriptor | Site | Fruit per genotype | Year | Genotypes sampled |
|---|---|---|---|---|
| *FFF-1* | East Malling | 3 | 1995 | 37 |
| | East Malling | 3 | 1996 | 105 |
| | Wellesbourne | 3 | 1995 | 35 |
| | Wellesbourne | 6 | 1996 | 108 |
| | Elst | 3 | 1996 | 129 |
| | Cadriano | 3 | 1995 | 43 |
| | Cadriano | 3 | 1996 | 82 |
| | Naoussa | 3 | 1995 | 81 |
| | Naoussa | 3 | 1996 | 145 |
| *FFF-2* | Angers | 20 | 1996 | 74 |
| *FST-RES-1* | Wellesbourne | 7[1] | 1996 | 106 |
| | Elst | 25 | 1995 | 75 |
| | Elst | 12 | 1996 | 8 |
| | Cadriano | 12 | 1996 | 12 |
| | Naoussa | 12 | 1996 | 10 |
| | Angers | 8[2] | 1996 | 112 |
| *Sensory descriptors* | East Malling | 11[3] | 1996 | 49 |
| | Wellesbourne | 11[3] | 1995 | 27 |
| | Wellesbourne | 11[3] | 1996 | 63 |

[1]Together with a further five fruit per genotype from 11 genotypes
[2]Together with a further four fruit per genotype from 10 genotypes
[3]Eleven panel members assessed each genotype, using sectors from one or more fruit

component for that factor to the sum of all variance components (the proportion of variance contributed by genotype is referred to as heritability). In order to obtain also the estimates of genotype means for QTL analysis, the analysis was repeated, first with genotype as fixed to obtain the overall genotype estimates, then with genotype and genotype × environment fixed, to obtain the genotype estimates per environment. Data were weighted according to how many fruit of a tree had been scored.

Penetrometer data were analysed with crossed effects of genotype and environment and an individual tree effect nested within this (although the only trees replicated on any site were the parents). The environment effect was represented by a single factor having a different level for different sites or years. In the analysis of penetrometer data, only datasets scored using descriptor FFF-1 were analysed

(Table 6.2). Attempts to include datasets scored using FFF-2 resulted in substantially increased error estimates, even after scaling.

The analysis of stiffness data from acoustic resonance measurements was broadly similar. A preliminary analysis considered the effect of time in store as randomised, and indicated that although there was a substantial linear trend of stiffness with time in store, there were no apparent effects of interest, such as interactions with genotype (results not presented). Consequently, in order to avoid the difficulties of a repeated measures analysis on these unbalanced data, the formal analysis was confined to the data from acoustic resonance measurements prior to storage.

For the analysis of the sensory data, the effects included were that of the batch of fruit, genotype with site nested within it, and a residual term.

The overall genotype estimates of the penetrometer readings, stiffness and the sensory descriptors were then analysed by principal component analysis. Due to the differences in scale, correlations rather than variances were analysed. The components accounting for most of the variance were subjected to QTL analysis.

*QTL analysis*

QTL analysis of the REML estimates for genotype, per environment and over environments, was performed using the Maximum Likelihood based interval mapping approach of MapQTL$^{TM}$ ver. 3.0 software (Van Ooijen and Maliepaard 1996). This version of the program enables QTL analysis of a full-sib family of a cross-pollinating species with four QTL alleles per segregating QTL. The integrated linkage map of ‘Prima’ and ‘Fiesta’ was used (Maliepaard *et al.* 1998). This map consisted of 290 markers (124 RFLP, 133 RAPD, 10 SSR, 17 isozyme, four AFLP, one CAPS-RFLP and one SCAR). Of these markers 127 were ‘Prima’ markers, 96 were ‘Fiesta’ markers and 67 markers were heterozygous in both parents. Of the heterozygous markers 49 segregated for three or four alleles and thus were fully informative, and 18 segregated for two alleles. Linkage phases between markers were known from linkage analysis. Since the integrated linkage map consists of markers with different segregation types, the all-markers mapping approach (Knott and Haley 1992; Maliepaard and Van Ooijen 1994) was used to upgrade marker information. In this method markers from neighbouring intervals, as well as flanking markers, are used to calculate the probabilities of QTL alleles. Five neighbouring intervals were employed, except for linkage group L1, where ten neighbouring intervals were used. This was done as the bottom end of L1 was covered mostly by markers which were informative for ‘Prima’ alleles only. By employing ten neighbouring intervals, information from ‘Fiesta’ alleles could also be included. A 1 cM step size was used.

For interval mapping, a LOD score threshold of 3.0 was used to indicate evidence for a QTL (suggestive linkage). This threshold corresponds to a per linkage

group error rate of 5% for the average linkage group length, which was 63 cM. A threshold of 4.5 was used to indicate significant linkage, which corresponds to a genome-wide error rate of 5% (Van Ooijen, personal communication). Interval mapping results were checked against results from QTL analysis using the regression approach of Haley *et al.* (1994), allowing for 4 alleles of a QTL, and from the non-parametric Kruskal-Wallis test, performed per marker (*e.g.* Lehmann 1975).

Multiple QTL Model (MQM) analysis (Jansen 1994a) was performed for selected sensory data, using this feature in MapQTL (Van Ooijen and Maliepaard 1996).

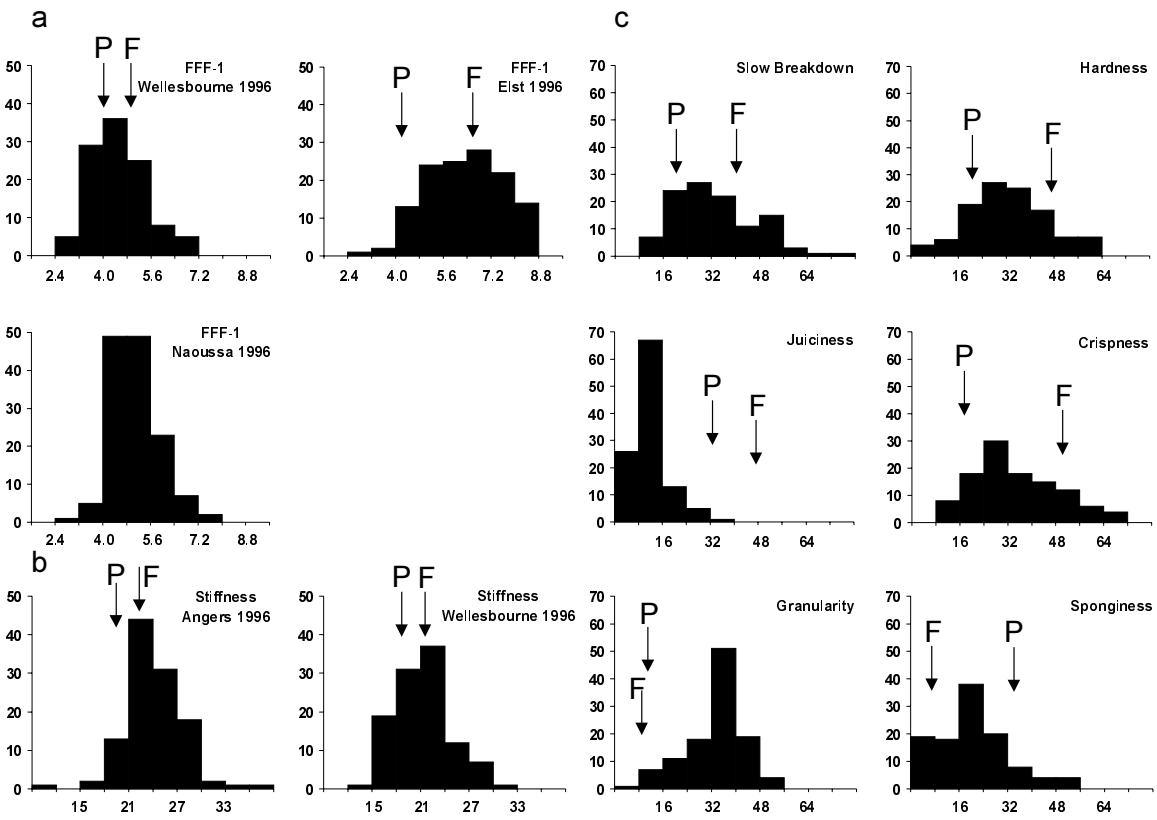



**Figure 6.1 Histograms of genotype estimates for traits relating to fruit texture scored on individuals from the segregating population derived from 'Prima' × 'Fiesta'. Values were calculated using REML. Y axis = number of genotypes in class. Mean parental values are indicated with arrows: 'P' = 'Prima'; 'F' = 'Fiesta'. a) Histogram of Fruit Firmness scores recorded in 1996; b) Stiffness as measured by resonant frequency in 1996; c) Sensory measures assessed in 1996.**

## 6.3 Results

### *Trait distributions*

The distribution of trait values across the population differed between environments and occasion, with evidence of transgressive segregation in all cases. All measures indicated that fruit of 'Prima' were less firm than 'Fiesta' (Fig. 6.1a, b, c). For penetrometer readings and stiffness measured by acoustic resonance, there were differences in the amount and direction of skewness in the distributions.

### *Variance components*

For the penetrometer data, 43% of the variability is accounted for by the site/year (=environment) combination. A further 25% is accounted for by the genotype × environment interactions, with a smaller (20%) though still highly significant ($p<0.001$) effect of genotype. For the stiffness data, the variability accounted for by the genotype was 14%, and that for genotype × environment was 9%.

The genotype × environment interaction was not significant ($p=0.05$) for any of the sensory descriptors. The effect of the batch of fruit was significant ($p<0.05$) only for slow breakdown and hardness.

Heritability estimates for the sensory traits ranged from 14% (granularity) to 57% (crispness). In decreasing order the estimates were: hardness (52%), slow breakdown (51%), sponginess (48%), juiciness (46%) and overall liking (28%).

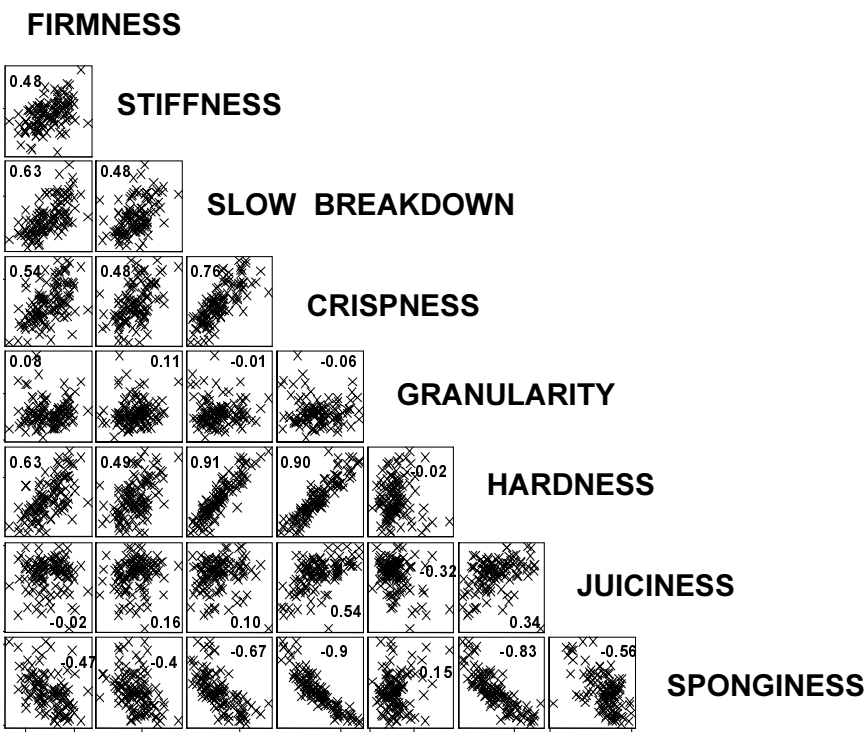



**Figure 6.2 Scatter plot matrix of mechanical and sensory descriptors. Correlation values have been added to the plots. Individual points correspond to genotype estimates using REML.**

## *Relationships between measurements*

A scatter plot matrix of the different measures is shown in Fig. 6.2. The penetrometer measurement FFF-1 has a correlation of 0.63 with the sensory descriptors of hardness and slow breakdown, and of 0.54 with crispness. The stiffness descriptor is not correlated very highly with any of the other measurements, having a correlation of just below 0.5 with the penetrometer value, slow breakdown, crispness and hardness. Among the sensory descriptors, hardness has a correlation over 0.9 with slow breakdown and crispness, although that between crispness and slow breakdown is only 0.76. Crispness is correlated negatively (-0.9) with sponginess. Juiciness has a correlation of 0.54 with crispness, and is also correlated positively with hardness, and correlated negatively with sponginess (-0.83) and granularity. The first principal component accounts for 55% of the total variability. It appears to represent an overall firmness measure as it comprises approximately equal weights of hardness, crispness, negative sponginess and slow breakdown together with penetrometer readings and stiffness, and a smaller component of juiciness. The second principal component accounts for a further 19% of the variability, and consists mostly of a contrast between juiciness and granularity, with smaller components of crispness and negative sponginess associated with the juiciness, and of penetrometer reading and stiffness with the granularity. The third principal component, explaining just under 10% of the variability, consists of a contrast between juiciness, granularity, and a smaller component of stiffness with penetrometer reading and slow breakdown. We have used negative sponginess in this description since sponginess is generally of the opposite sense to the other measures (Fig. 6.2).

## *QTL analysis*

QTL mapping results are presented in Table 6.4. For the penetrometer data (FFF-1), LOD scores greater than 3.0 in at least one environment were observed for 12 of the 17 linkage groups. A possible QTL on linkage groups L01 is observed in five of the ten site/year combinations, in five out of the six sites, all of them in 1996. The LOD score for the sixth site (Angers) was 2.9. A possible QTL on linkage group L10 was observed also for five of the ten site/year combinations. Significant linkage (LOD score greater than 4.5) was observed on linkage groups L01, L06, L10 and L12 in different environments. Results of regression based interval mapping confirmed those of the maximum likelihood approach, the absolute LOD score difference usually being less than 0.5. Only for L06 was the LOD score much smaller with regression than with the maximum likelihood approach (LOD score of 3.7 compared to 4.8). For the overall estimates (over environments) linkage groups L01, L08 and L10 yielded LOD scores greater than 4.5.

**Table 6.4 QTLs associated with penetrometer readings, stiffness and sensory descriptors, detected with the interval mapping technique. Linkage groups are shown where maximum LOD scores greater than 3.0 were detected (suggestive linkage). Linkage groups with LODs greater than 4.5 are indicated in boxes, with percentage of variance explained by the putative QTLs indicated below. Kruskal-Wallis tests were carried out for sensory descriptors, and significance levels are indicated by asterisks: * p < 0.005; ** p < 0.001; *** p < 0.0005; **** p < 0.0001.**
**Ninf = number of individuals for which data for the quantitative trait were available.**

| | | Ninf | L01 | L02 | L03 | L04 | L05 | L06 |
|---|---|---|---|---|---|---|---|---|
| FFF-1 | Naoussa 1995 | 79 | | | | | | |
| | Naoussa 1996 | 132 | 3.0 | 3.1 | | | | |
| | East Malling 1995 | 35 | | | | | | |
| | East Malling 1996 | 101 | 3.3 | | | | | 4.8 *28%* |
| | Wellesbourne 1995 | 34 | | | 3.2 | | | |
| | Wellesbourne 1996 | 106 | 3.7 | | | | | |
| | Elst 1996 | 125 | 4.6 *16%* | | | 4.1 | | |
| | Cadriano 1995 | 42 | | | 3.1 | | | 3.0 |
| | Cadriano 1996 | 79 | 4.1 | 3.4 | | | | 3.9 |
| FFF-2 | Angers 1995 | 73 | | | | | | |
| FFF | Overall | 152 | 6.5 *19%* | | 3.6 *17%* | | | |
| Stiffness | W'bourne '96, day 0 | 104 | 3.7 | | | | | |
| | Elst 1995, day 0 | 72 | 4.0 | | | | | |
| | Angers 1996, day 0 | 111 | 3.4 | | | | | 3.1 |
| | Stiff-Overall, day 0 | 144 | 4.4 | | | | | |
| Sensory | Slow breakdown | 113 | * | | | | 3.6 ** | |
| | Crispness | 113 | 3.5 * | | | | ** | * |
| | Granularity | 113 | | 5.1 *24%* | | | | |
| | Hardness | 113 | | | | | 3.7 *** | 3.4 ** |
| | Juiciness | 113 | 3.6 | | | | | |
| | Sponginess | 113 | 3.3 | | | | | |
| | Overall liking | 113 | | | | | | |
| PCP | 1st principal comp | 110 | 3.4 | | | | 3.2 | 3.1 |
| | 2nd principal comp | 110 | 4.0 | | | | | |
| | 3rd principal comp | 110 | | | | | | |

| L08 | L09 | L10 | L12 | L13 | L15 | L16 |
|---|---|---|---|---|---|---|
| | | 4.0 | | | | |
| | | 4.9 | | | 3.1 | |
| | | *18%* | | | | |
| | | | 5.9 | | | |
| | | | *60%* | | | |
| 4.4 | | 3.6 | | | | 3.7 |
| | | | | 3.1 | 3.0 | |
| | | 7.1 | | | | |
| | | *29%* | | | | |
| 4.0 | 3.1 | 3.0 | 3.0 | | | |
| | | | | | | 3.2 |
| | | | | | 4.0 | |
| 4.7 | | 7.4 | | | 3.5 | |
| *16%* | | *22%* | | | *10%* | |
| | | 4.6 | | | | 3.2 |
| | | *21%* | | | | |
| | | 4.5 | | | | |
| | | *16%* | | | | |
| | | 3.5 | 3.1 | | 3.1 | |
| *** | | * | | | | |
| | | 3.2 | 3.9 | | | 6.0 |
| | | | | | | *24%* |
| ** | | | * | * | | **** |
| 3.0 | | | | | | 3.0 |
| | | | | * | | |
| | | 3.7 | | | 3.2 | 3.8 |
| *** | | * | * | | | *** |
| | | | | | | 14.8 |
| | | | | | | 46% |
| | | | | | 3.1 | 7.7 |
| | | | | | | 30% |
| | | | | | | 11.3 |
| | | | | | | *38%* |
| | | | | * | | **** |
| | | 4.2 | 3.4 | | 3.4 | 5.1 |
| | | | | | | *22%* |
| 3.0 | | | | | | 7.8 |
| | | | | | | *28%* |

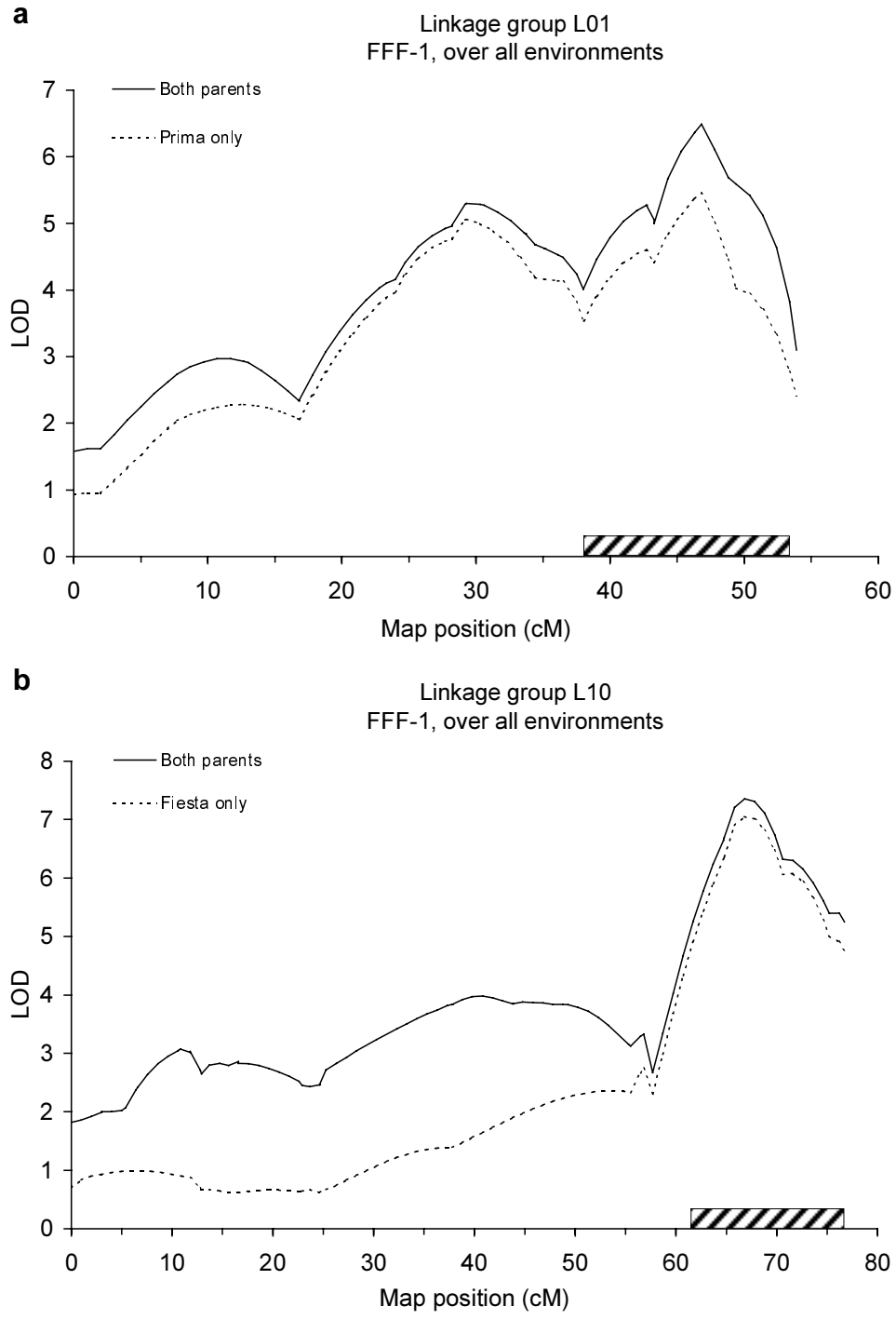



**Figure 6.3 LOD plots resulting from interval mapping of REML estimates of the FFF-1 trait data over all environments. Solid lines indicate results where marker data from both parents were used. a) QTL detection on linkage group L01. Dashed line indicates results from using only the 'Prima' map and 'Prima' markers. b) QTL detection on linkage group L10. Dashed line indicates results from using only the 'Fiesta' map and 'Fiesta' markers. Two-LOD support intervals for the map position of the QTL are indicated by bars along the x-axes with diagonal hatching.**

For the resonant frequency data, LOD scores greater than 3.0 were observed for four of the 17 linkage groups. LOD scores greater than 4.5 were observed for linkage group L10 only. A possible QTL on L01 for the combined data for day 0 was just below the threshold of 4.5.

The QTL on linkage group L01 for the penetrometer and resonant frequency data was primarily as a contrast of between the two 'Prima' alleles. The QTL maps to a region introgressed from a wild apple species, *M. floribunda* 821 and which includes the *Vf* gene, which confers scab resistance (King *et al.* 1998). The QTL allele which contributes to firmer fruit is in coupling phase with *Vf*. The penetrometer overall mean was analysed again, using only the 'Prima' map and 'Prima' markers to verify whether the QTL effect was due to a 'Prima' allele. The results of both analyses are presented in Fig. 6.3a. The QTL from linkage group L10 was mostly a contrast between the 'Fiesta' alleles. Here also the data were re-analysed using only 'Fiesta' markers and the 'Fiesta' map to verify this (Fig. 6.3b). The differences between the LOD graphs are small, indicating that in each case the QTL is an effect of a single allele from a single parent, *i.e.* a contrast between the 'Prima' alleles in the case of L01, a contrast between the 'Fiesta' alleles in the case of L10.

For the sensory data sets, LOD scores greater than or equal to 3.0 were found for nine linkage groups for one or more of the sensory descriptors. LOD scores greater than 4.5 were observed for L02 and for L16 for three of the sensory descriptors. The high LOD score for L02 for granularity was not confirmed by regression. Linkage groups L01 and L10 showed LOD scores greater than 3.0 for slow breakdown and hardness (L10 only), crispness (both), juiciness and sponginess (L01 only). Although less clear, here also the QTL on L01 seem to be mainly effects of one 'Prima' allele in coupling phase with *Vf*. For juiciness, one allelic combination of the QTL seemed to give less juicy fruits than the other three allelic combinations. The QTL on L16 gave high LOD scores for crispness, juiciness, sponginess and overall liking. This QTL was mapped to the region of *Ma*, the malic acid gene (Maliepaard *et al.* 1998). The effect was a combined effect of alleles from both parents, showing a distinct contrast between individuals in the progeny with the dominant *Ma* and the recessive *ma* phenotype, the *Ma* phenotype (more acidic) being the more favoured in all three cases.

For crispness, juiciness, sponginess and overall liking, a marker closely linked to *Ma* was used as a cofactor in MQM analysis, so that this marker (OPT-16-1000) would absorb variation due to the QTL in this region. For crispness, possible QTLs (LOD > 3.0) were observed on L5 and L13, in addition to those previously observed on L1, L10, L12 and L16 (Table 6.4). For juiciness, a possible QTL was observed on L12, and significant linkage (LOD score >4.5) was now observed for the QTL on L1. For sponginess, possible QTLs were observed on L5 and L6, and for overall liking, on L12.

In the QTL analysis of the first two principal components, LOD scores greater than 4.5 were found for L16, while LOD scores greater than 3.0 were found for L01, L05, L06, L08, L10, L12 and L15. No LOD scores greater than 3.0 were found for the third component.

The 2-LOD support intervals were calculated for the QTLs for the overall mean of the penetrometer data, for linkage groups L01, L08 and L10. Simulation studies have shown that a 2-LOD support interval usually has a probability over 95% of containing the QTL (Van Ooijen 1992). The 2-LOD support interval for L01 was [39.0-53.4 cM], for L08 the 2-LOD support interval was the interval [32.1-53.5 cM] and for L10 the interval [61.7-76.2 cM]. The 2-LOD support intervals for L01 and L10 are indicated in Fig. 6.3.

## 6.4 Discussion

The trait distributions for penetrometer readings, stiffness and some of the sensory descriptors show transgression of the progeny towards softer fruit and firmer fruit. In all cases ‘Prima’ had softer fruit than ‘Fiesta’, and in all sensory descriptors ‘Fiesta’ was favoured over ‘Prima’. The measures juiciness and granularity had skewed distributions, with the parents clustered towards one end. This may reflect strong selection for juiciness in marketed cultivars.

Large environmental effects and large genotype × environment interactions were observed for physical measurements of fruit firmness. These effects complicate the selection in one environment for performance in different environments, if traditional breeding methods are used. Rather small heritability estimates were obtained for penetrometer readings and stiffness as measured with acoustic resonance, with those for stiffness being smaller than for penetrometer readings. Despite these high genotype × environment interactions and low heritabilities, a number of QTLs for fruit firmness characters were detected in this study. Apparently some of these were effective in multiple environments, and perhaps even in those environments where the LOD score in this study did not exceed the significance threshold. Indeed some other elevated LOD scores for linkage groups L01 and L10 were observed, but these remained below the thresholds, due perhaps to larger environmental bias or smaller population size. This indicates that marker-assisted selection, at least for these main QTLs, may be feasible and effective for multiple environments, whereas the effectiveness of traditional selection methods without the additional information provided by markers might be hampered by the environmental noise and interaction effects.

There are concerns about the relevance or reliability of penetrometer tests since they are not always considered to reflect accurately the textural components "crisp", "tough" or "mealy" (Lurie and Nussinovitch 1996). In the current study, low to moderate correlations were found between physical measurements of fruit firmness and between physical and sensory measurements. These correlations were deter-

mined across 147 different genotypes. Although Lurie and Nussinovitch (1996) had found much higher correlations for penetrometer and sensory descriptors, their data were from only two cultivars, and correlations were within cultivars. Our observations confirm recent findings based on inter-varietal comparisons (Barreiro *et al.* 1998).

Correlations among sensory descriptors were high for hardness, crispness, slow breakdown and negative sponginess. Correlations of these descriptors with juiciness were smaller, and granularity showed only very small correlations with the other sensory descriptors. The higher correlations may be attributed to perceptual interactions. The relationship between crispness and juiciness is likely to arise from cortex tissue of very crisp apple fruit possessing stronger inter-cellular bonds in the middle lamella of adjacent cell walls. When such fruit are bitten, the cells are more likely to fracture and release juice. A weaker middle lamella leads to fracture between cells and thus gives rise to the perception of a mealy or granular texture. This is reflected in the negative correlation of crispness and juiciness with granularity. It is also possible that the detection of juiciness is emphasised by the co-segregation of acidity, which also contributes to an overall liking descriptor which was scored by the same panel. However, recent studies based on three commercial cultivars, each presumably of normal acidity, have also shown a high correlation between crispness and juiciness (Barreiro *et al.* 1998). They also observed higher correlations of these descriptors with instrumental measurements, these being confined compression of fruit cylinders and acoustic impulse response.

Despite the moderate correlations between penetrometer readings, stiffness and sensory data, there was evidence for QTLs for each of these on linkage groups L01 and L10. In the sensory evaluation, L01 and L10 show up for those data which are expected to reflect firmness most strongly: crispness, hardness and slow breakdown. More QTLs were detected for penetrometer readings than for stiffness or individual sensory descriptors. This may be explained by higher number of replicate observations of the genotypes and the higher estimated heritability for penetrometer readings. Also, the possibility cannot be excluded that the penetrometer reading comprises more and different underlying genetic components than those involved in stiffness and sensory descriptors. As a consequence of the lower heritability estimate and the somewhat lower correlations with sensory descriptors, the value of stiffness measured by acoustic resonance may be rather limited in breeding selection. For the breeder, the ease of determination of fruit firmness with a penetrometer is also likely to continue to favour use of the latter.

The two most significant QTLs for penetrometer readings were shown to be derived from the respective parents. The QTL on L01 appeared to be a contrast of the 'Prima' alleles, with no significant effect from the 'Fiesta' alleles (Fig. 6.3a) whereas, with the QTL on L10 the effects were reversed. This demonstrates that each parent possesses different QTL alleles for penetrometer readings, and that even 'Prima', which has softer fruit than 'Fiesta', is able to contribute to increased

fruit firmness. This is in agreement with previous findings from breeding studies. It also emphasises the value of carrying out QTL analysis simultaneously for both parents in a cross. For example, a QTL analysis which was based on one parent expected to contribute to greater fruit firmness would overlook possible QTL contrasts segregating from the other parent. A similar situation has been exploited in tomato, where advanced backcross QTL analysis has been able to detect 25 alleles from a wild relative which improve traits from a horticultural perspective, despite the fact that overall the donor is phenotypically inferior to the elite parent (Bernacchi *et al.* 1998a). It is also in agreement with the distributions shown in Fig. 6.1a and 6.1b which illustrate transgressive segregation in the progeny, towards firmer fruits and softer fruits.

The QTL identified on linkage group L01 of 'Prima' is linked to the introgressed region originating from *Malus floribunda* 821 carrying the *Vf* locus which confers scab resistance (King *et al.* 1998). The allele contributing to firmer fruit is in coupling phase with *Vf*. Although there was a small gap in between two LOD peaks on this linkage group, there was no evidence for a second QTL on this linkage group. This was verified by using a marker as a cofactor on this linkage group and testing for the presence of another QTL. It was also confirmed in Bayesian analysis (Maliepaard and Sillanpää *et al.* submitted).

The sensory analysis attempted to address the ability of different human perceptions to resolve variation in aspects of apple fruit texture. The use of a trained panel to assess large populations of fruit in a genetic analysis is rare. We have been able to demonstrate that such data may be used to detect genetic effects. The fact that relatively few QTLs were detected for the sensory descriptors may indicate that these do not reflect very strongly the underlying genetic factors, perhaps due to large environmental effect, or due to masking of these genetic factors by other genes which may overrule their expression in human perception. However, it is also possible that the sensory descriptors represent simple traits. The crispness QTL on group L16 has a high LOD score (6.0), and accounts for 17% of the variance. The major juiciness QTL also locates to this region, with a LOD score of 14.8, and accounts for 46% of the variance. For sponginess, a LOD score of 7.7 was obtained in this region and this QTL accounts for 30% of the variance. QTLs for hardness and granularity are also suggested in the same region. These sensory QTLs map to a region known to contain the acidity locus *Ma* (Maliepaard *et al.* 1998). Three aspects emphasise the correspondence of these sensory QTLs with the *Ma* locus. Firstly, they are apparently linked rather closely, as shown by the LOD peaks at the *Ma* locus. Secondly, the segregation of the sensory QTL appears to be identical to the segregation at the *Ma* locus: of the four possible QTL genotypes for each sensory descriptor, the estimated mean of one QTL genotype is in the direction of less firm, less juicy and more spongy fruit, whereas the other three QTL genotype means are approximately equal. In each case, the less favoured QTL genotype corresponds to

the *mama* genotype in the progeny. This suggests dominance of a QTL allele present in both parents, identical to the situation of *Ma*. Finally, this favoured and dominant QTL allele is in coupling phase with *Ma*.

This association of sensory texture QTLs on L16 with the *Ma* locus may be due to perceptual interactions. The marked lack of taste of apples from *mama* genotypes may override other positive perceptions, whereas the better tasting fruits from *MaMa* and *Mama* genotypes may mask some other negative attributes. The co-segregation with the *Mama* locus may also indicate that they are pleiotropic effects of a gene at the *Ma* locus. Finally, a cluster of different genes co-localised in the same region may exist. In peach, QTLs for several fruit quality traits including acidity and soluble solid content, measured instrumentally, have been located recently in the same regions of just two linkage groups (Dirlewanger *et al.* 1999). In tomato a region on chromosome 6 contained a QTL for fruit mass, pH and soluble solid concentration (Paterson *et al.* 1991).

The experiment design was developed when little initial information about site-to-site variation existed for the particular measures. These and other results from the same research programme (King 1996; King *et al.* 1998) indicate that a modified design may be more appropriate in future. For accurate positioning of QTLs it is important, within a given size of experiment, to maximise the number of recombinants whilst providing inter- and intra-site replication of only a relatively small proportion of the population to obtain accurate estimates of variance for site-to-site and occasion-to-occasion variation (Lynn 1998).

The identification of chromosomal regions contributing to major attributes of fruit texture is the first stage in developing selectable markers for the early selection of desirable genotypes. Several issues are raised. One relates to the accuracy of QTL position and the ability to predict effects in subsequent generations. The accuracy of mapped QTLs is still problematic. Some improvement can be made by using marker cofactors (Jansen 1993; Zeng 1994), further improvement can be made by using only fully informative markers, larger progenies or more generations, and by reducing missing values for markers and trait. In apple, where further pedigree testing is difficult or impossible, there are no major remedies. However, for marker-assisted early selection, it may be enough to increase considerably the probability of selecting the favourable genotype and this is feasible even for a QTL of which the position is known with little accuracy. This becomes more problematic when there are disadvantageous genes linked to the positive QTL allele. The second issue relates to the conservation of the position of functional alleles in different genetic backgrounds. The results from these experiments would need to be validated in crosses or pedigree analysis involving cultivars other than ‘Prima’ and ‘Fiesta’ before universal statements could be applied.

## *Acknowledgements*

This work was funded by the CEC Framework III research and technical development programme in Agriculture and Agricultural Industries, as part of the project CT920473, 'Development of the European Apple Crop', with matching funding from the UK Ministry of Agriculture, Fisheries and Food, and the Biotechnology and Biological Sciences Research Council and other national governments of the EC. We are grateful to Dr Pam Beyts of Sensory Dimensions Ltd, Reading for carrying out the sensory analysis.

# Chapter 7

# Bayesian versus frequentist analysis of multiple quantitative trait loci with an application to an outbred apple cross

C. Maliepaard and M.J. Sillanpää, J.W. van Ooijen, R.C. Jansen and E. Arjas[1]

## ABSTRACT

Two methods, following different statistical paradigms for mapping of multiple quantitative trait loci (QTLs), were compared. The first is a frequentist, the second a Bayesian approach. Both methods were applied to experimental data from an outbred progeny of a single cross between two apple cultivars (*Malus pumila* Mill.). These approaches were compared with respect to 1) the models used, 2) the number of putative QTLs, 3) their estimated map positions and accuracies thereof, and 4) the choice of cofactor markers. In general, the strongest evidence for QTLs, provided by both methods, was for the same linkage groups, and for similar map positions. However, some differences were found with respect to evidence for QTLs on other linkage groups. The effect of using cofactor markers was also somewhat different between both methods.

## 7.1 Introduction

The analysis of quantitative trait loci (QTLs) using molecular markers has become routine for genetic studies in many plant and animal species (see Tanksley 1993; Haley 1995; Doerge *et al.* 1997; Hoeschele *et al.* 1997; Kearsey and Farquhar 1998). QTL detection is often based on the simple interval mapping (SIM) method which was first introduced for inbred lines and which uses a mixture model or a linear regression model (Lander and Botstein 1989; Haley and Knott; 1992). Parameters are estimated with a maximum likelihood (ML) or least squares (LS) approach. Phenotypic variation caused by other QTLs is not taken into account. This single QTL model was later extended to the composite interval mapping method (CIM) also known as multiple QTL models (MQM) mapping (Jansen 1992, 1993, 1994a; Zeng 1993, 1994; Jansen and Stam 1994; Kao and Zeng 1997). This method employs an approximate multiple QTL model which also considers some

[1] Submitted for publication. C. Maliepaard and M. Sillanpää are recognised as jointly contributing to first authorship

QTLs outside the tested interval, indirectly, through marker cofactors. The power of detecting QTLs is thus increased by reducing the error variance. When fitting multiple QTLs on a single chromosome the choice of the cofactors on that chromosome is critical: different choices will have to be tested to prevent picking up a so-called 'ghost QTL' (Martínez and Curnow 1992).

Haley and Knott (1992) and Martínez and Curnow (1992) investigated the use of an exact two-QTL model in regression for mapping linked QTLs. Jansen (1992) described how the equations arising from the exact multiple QTL models can be solved using mixture models. However, computation may become time consuming when the number of QTLs increases. This problem can be alleviated by omitting those QTL genotype combinations with small probabilities (Jansen 1995). Recently, Kao *et al.* (1999) proposed a stepwise procedure for model selection in exact multiple QTL models, but they pointed out that the issue of determining appropriate critical values still needed to be solved.

The use of Markov Chain Monte Carlo (MCMC) methods has made it practical to consider exact multiple QTL models and highly incomplete data numerically. By applying MCMC to the Bayesian framework, Thaller and Hoeschele (1996), Uimari *et al.* (1996a), and Uimari and Hoeschele (1997) estimated the number of linked QTLs by using fixed models with a linkage indicator to control the status (either linked or unlinked) of each QTL. Satagopan *et al.* (1996) used the Bayes factor for choosing between models that involved different numbers of QTLs. In the Bayesian framework, the number of QTLs can be treated as an unobserved random variable which can be estimated simultaneously with the other parameters (see Satagopan and Yandell 1996; Heath 1997; Uimari and Hoeschele 1997; Stephens and Fisch 1998; Sillanpää and Arjas 1998, 1999).

Sillanpää and Arjas (1998, 1999) modelled multiple QTLs on one chromosome while, for computational reasons, the QTLs on other chromosomes were accounted for by the cofactor markers, comparable to what is done in MQM. The exact consideration of all chromosomes simultaneously in the model would have been more in line with the principles of Bayesian modelling and analysis. For example, Stephens and Fisch (1998) considered multiple chromosomes simultaneously, but they did not consider missing values in the data.

In this paper we focus on QTL analysis in a full-sib family, the direct progeny of a single cross of two outbred parents. This situation differs in several respects from the situation in progenies derived from homozygous parents like backcross or $F_2$: (1) at any marker locus or QTL, two to four alleles may be segregating in the progeny. (2) segregation types may vary from marker to marker (see Maliepaard *et al.* 1997). (3) the linkage phase of a marker pair may be different in the two parents (*e.g.* coupling phase in the mother and repulsion phase in the father) and can vary from interval to interval.

As a consequence, the information content with respect to the estimation of QTL genotype probabilities varies across the genome and simple interval mapping may result in discontinuities in the profile of the test statistic. Knott and Haley (1992) showed that the test statistic may be higher in the intervals with a higher information content, even when the true position of the QTL is in a different but less informative interval. Therefore, the maximum amount of information with respect to the QTL genotype probabilities should be used by employing not only the markers flanking an interval but preferably all markers of a linkage group simultaneously (Knott and Haley 1992; Maliepaard and Van Ooijen 1994; Haley *et al.* 1994; Jansen 1996; Sillanpää and Arjas 1998, 1999). This procedure, denoted as 'multipoint linkage' or 'all-markers mapping' can be used both for a QTL and a cofactor marker.

There are only few papers in the literature that compare Bayesian and classical QTL mapping methods. Scheler *et al.* (1998) compared the 'Bayesian' approach to the traditional SIM analysis in an inbred line cross situation with simulated data. However, their 'Bayesian' method is not strictly Bayesian; it is a likelihood method, where all QTL positions are integrated out of the likelihood expression. As a consequence, no location estimates are provided. Moreover, their comparison was focused entirely on frequentist test theory and on the asymptotics of test statistics. Uimari *et al.* (1996) compared their Bayesian method to LS and residual maximum likelihood (REML) methods using simulated and experimental data in an outbred livestock population with a granddaughter design. They found a good agreement between the methods in their location estimates. Sillanpää and Arjas (1998, 1999) compared their Bayesian methods to SIM and CIM using simulated data sets and inbred and outbred experimental designs. In their view, the main advantages in using Bayesian methods were in probabilistic inference and in probabilistic summary statistics. Vieland (1998) compared the Bayesian posterior probability of linkage and the LOD score theoretically and provided some considerations of a more general nature. Differences between LOD support intervals and Bayesian credible regions were discussed in Dupuis (1996) and Dupuis and Siegmund (1999). Very recently, Shoemaker *et al.* (1999) reviewed properties of Bayesian and frequentist methods used in genetics. They focused on the differences in the underlying paradigms.

The purpose of this paper is to compare the frequentist methods of SIM and MQM mapping with the Bayesian multiple QTL analysis. For this comparison we use experimental data from a single large full-sib (FS) family derived from a cross between two apple cultivars. The methods are compared with respect to 1) the models used 2) the number of QTLs mapped, 3) the estimated map positions of the QTLs and their accuracies, and 4) the choice of marker cofactors. The same data set has been previously analysed using SIM in King and Maliepaard *et al.* (2000).

## 7.2 Materials and Methods

### *Experimental data*

A cross between the apple cultivars 'Prima' and 'Fiesta' was carried out at CPRO-DLO, the Netherlands, in 1988, using 'Prima' as the female parent. A full-sib progeny consisting of 152 genotypes from this cross was vegetatively propagated and replicate sets or subsets of this progeny plus parents were planted at seven sites in six countries in Europe in 1993 (King *et al.* 1991; King 1996). In 1995 and 1996 apples from trees from six sites were analysed for fruit firmness using two test methods: 1) acoustic resonance frequency (RF) (Abbott *et al.* 1992; Chen and De Baerdemaker 1993) and 2) hand penetrometer (PEN) (Magness and Taylor 1925; Bourne 1974). For both sets of phenotypic measurements, data over sites and years were analysed using REML in Genstat 5, weighting with the number of apples per tree measured and taking as random factors the site/year combination and trees of the same genotype within a site/year; the genotype was taken as fixed. In this way estimates for each genotype were obtained and these were used for QTL analysis. For the resonance frequency the estimates were over three site/year combinations; for the penetrometer readings the estimates were over nine site/year combinations. Figure 7.1 shows a scatter plot of the PEN estimates plotted against the RF estimates.

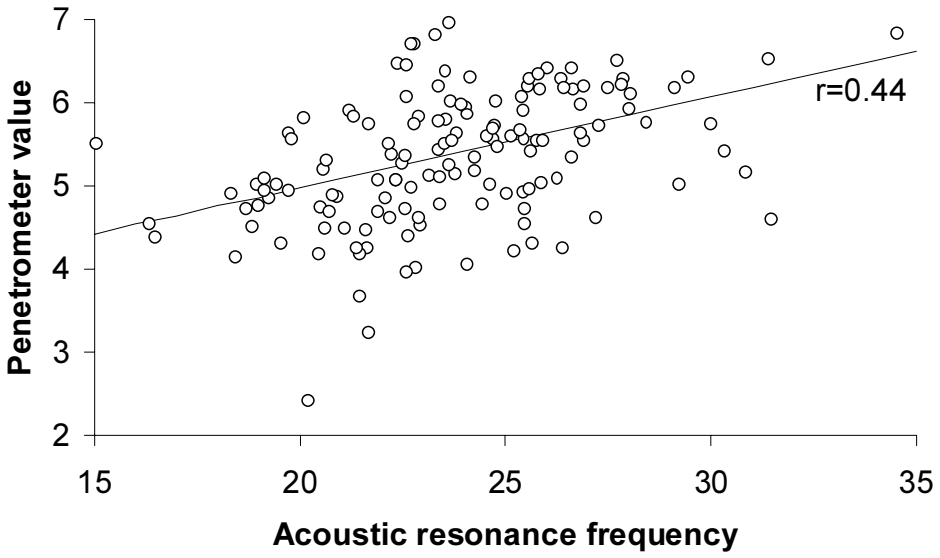



**Figure 7.1 Scatter plot of penetrometer estimates by resonance frequency estimates of the 'Prima' × 'Fiesta' progeny. Each circle indicates an individual in the progeny.**

## *Information content*

The information content of each linkage group was calculated, using the experimental genotype data, and taking into account the numbers of missing values and the upgrading of marker information if markers were only partly informative or if there were missing values. The information content was first defined, for each individual, as the maximum of the four QTL genotype probabilities at a map position. These maxima were then averaged over the individuals in the progeny. The information content was calculated with steps of 1 cM and varied from linkage group to linkage group, with a minimum value of 0.58 and an average of 0.86 over the whole genome. The information content of linkage group L01, using a maximum of ten neighbouring markers to upgrade the information, is shown in Figure 7.3.

## *QTL analysis*

Linkage maps of 'Prima' and 'Fiesta' were constructed using 290 of mostly RFLP, RAPD, isozyme and SSR markers to genotype both parents and the progeny (Maliepaard *et al.* 1998). Sixty-seven multi-allelic markers for which both parents were heterozygous allowed the two parental maps to be integrated. The linkage map consisted of 17 linkage groups, which corresponds to the haploid chromosome number of apple. Linkage phase combinations between marker pairs were estimated using JoinMap™ version 2.0 (Stam and Van Ooijen 1995). The integrated linkage map (293 markers, including mapped genes *Vf*, *Sd*$_1$ and *Ma*) was used for QTL mapping. The marker order and distances on the linkage map, and the parental linkage phases, were assumed known, in both the frequentist and the Bayesian QTL analysis.

In a full-sib family a QTL or a marker can segregate for four distinct alleles, *i.e.* parental mating type $Q_1Q_2 \times Q_3Q_4$, producing four different genotypes. Therefore, in both the frequentist and Bayesian approach, three effects (deviances from the effect of the first genotype) are modelled for a QTL and each cofactor. As usual only additive QTL terms (no epistasis) were considered in the models.

## *Frequentist QTL analysis*

In SIM and MQM mapping, standard statistical procedures are used for testing hypotheses: under the null hypothesis (no QTL) a likelihood is constructed from a single normal distribution with mean and variance equal to the population mean and variance. Under the alternative hypothesis (presence of a QTL at a given map position) the likelihood is constructed from a mixture of normal distributions. The iterative EM algorithm (Dempster *et al.* 1977) was used to estimate QTL genotype means, cofactor effects and residual variance, and to calculate the likelihood of a QTL at a certain map position. For interval mapping this estimation procedure has been described in detail by Van Ooijen (1992). In MQM the cofactor effects were estimated simultaneously with the QTL effects (Jansen 1994a).

The LOD profiles over each of the linkage groups were used to determine the map positions of QTLs. These were estimated as the position with the maximum LOD score on a linkage group. Uncertainty of the map position was indicated by a 2-LOD support interval (Conneally *et al.* 1985; Van Ooijen 1992). The number of QTLs was inferred from the number of LOD peaks exceeding the significance threshold.

In addition to the determination of the 2-LOD support intervals, also bootstrapping (Efron 1979, 1982) was used to obtain approximate central 95% confidence intervals for QTL positions (Visscher *et al.* 1996b) which were expected to be more comparable to the Bayesian credible intervals of Sillanpää and Arjas (1998). This was done only for selected linkage groups and only for SIM. Observations, consisting of a phenotypic value and the marker genotypes for all markers, were resampled simultaneously from the original data set as in Visscher *et al.* (1996b). Each bootstrap sample was analysed with SIM in an identical way as was the original data set. The position with the maximum LOD score was retained after each bootstrap analysis.

Suitable threshold values for different population types have been obtained through computer simulation by Van Ooijen (1999). Threshold values can also be obtained by permutation tests (Churchill and Doerge 1994). In this study, as in King and Maliepaard *et al.* (2000), LOD scores greater than 3.0 were considered as evidence of a QTL. LOD scores greater than 4.5 were considered significant. These values correspond to a per linkage group error rate of 5% for the average linkage group length (63 cM), and a genome-wide error rate of 5%, respectively (Van Ooijen 1999).

Flanking markers were used to calculate the probabilities of the four QTL genotypes at a given map position. For missing marker data and for markers that were not completely informative with respect to the four possible parental allelic combinations, also marker genotype information from neighbouring markers was used ('all-markers mapping'). Usually up to five neighbouring markers on both the left and right hand side of the marker interval were used to upgrade the genotype information. Only for linkage group L01, ten neighbouring markers were used, since there was a group of markers at one end of the linkage group which provided information with respect to one parent only. For cofactors the same procedure was applied in MQM mapping.

MapQTL$^{TM}$ 3.0 (Van Ooijen and Maliepaard 1996) was used both for SIM and MQM mapping in this outbred full-sib progeny. On average, nine seconds per linkage group were required when performing SIM (without cofactors) on a 200 MHz PC with a Pentium processor, running under MS DOS 6.2 (5 neighbouring markers, 1 cM steps). For MQM analyses with 3 cofactors (only on the other linkage groups), approximately 2'30'' were required per linkage group. Results from SIM were checked against results obtained with the LS approach of Haley *et al.* (1994) and against the non-parametric test of Kruskal-Wallis, performed on the markers.

## *Bayesian QTL analysis*

Following Sillanpää and Arjas (1998,1999) an exact multiple QTL model was used for one chromosome while the other chromosomes were controlled by using a preselected set of cofactor markers. In Bayesian analysis, the QTL mapping problem is not formulated in a sequential (hypothesis) testing framework as in frequentist methods. All results of the analysis can be expressed in terms of the posterior distribution of the unknown variables/parameters in the model, given the data. Convenient summary measures for variables of interest, such as the number of QTLs and the QTL positions, can be defined by considering suitable marginals of the posterior distribution or corresponding expectations. Bayesian posterior credible intervals for these parameters can be constructed from the marginals of the posterior distribution. In principle, any interval can be taken as a credible interval. The main advantage of these intervals is in their straightforward interpretation in terms of conditional probabilities of containing the unknown parameters given the data.

In the numerical estimation of the model parameters in the adopted Bayesian hierarchical model (Sillanpää and Arjas 1999), the Metropolis-Hastings-Green algorithm was used (Metropolis *et al.* 1953; Hastings 1970; Green 1995). In each round of the estimation, the QTL genotype probabilities were determined for each individual in the offspring given current values for completed fully informative markers and/or QTLs. In each round, incomplete marker data were completed for missing genotypes and linkage phases and coded according to their grandparental origin by using all other markers in the same linkage group. With equal probabilities, this block-updating was conducted for the whole family and separately at each marker point or for the entire haplotype and separately for each individual (see Appendix). Missing cofactor genotypes were augmented by assuming marker independence and by using M-H where acceptance of the imputed values was always conditional on the current parameter values. Initially in each MCMC run, three QTLs (which was also the maximum allowed in each chromosome) were placed along each linkage group to be analysed. A truncated Poisson distribution was used with mean equal to 2 as the prior for the number of QTLs. The prior for the residual variance was uniform over the range [0.0, 3.35] in the RF data and over [0.0, 0.69] in the PEN data. The right endpoints of these ranges were set equal to the variance estimates from the corresponding data. The prior of the regression intercept was taken to be uniform over [-100, 100] in the RF data and over [-13, 13] in PEN data. The prior of QTL genotype coefficients was N(0,100) in both data sets and the prior of cofactor coefficients was uniform over [-13, 13]. The prior of the QTL location was taken to be uniform over the entire length of the particular linkage group. The random walk proposal ranges in the MCMC analyses were chosen to be 2.0 (location), 1.0 (intercept), 0.2 (residual S.D.), 1.5 (QTL coefficients), and 2.0 (cofactor coefficients). The proposal distribution for new QTL effects was N(0.0, 0.5).

The burn-in period was not deleted, since high numbers of MCMC cycles (from 2,500,000 to 5,000,000) were run in all analyses. In the estimation (Monte Carlo averaging), the MCMC samples were thinned, using only every tenth iteration, because of the limited storage capacity. Credible intervals for the positions of QTLs were constructed from the posterior QTL intensities as in Sillanpää and Arjas (1998).

The posterior distribution of the number of QTLs in a mapped chromosome can be used as an initial summary measure of the analysis. Based on this measure, chromosomes showing some QTL activity can then be investigated further by looking at their posterior QTL intensities along the chromosome.

It is important to note here that the whole Bayesian analysis for experimental data was conducted independently from the mapping results obtained by MQM. All MCMC calculations (using Multimapper/OUTBRED software http:/www.rni.helsinki.fi/~mjs)) were performed on a DEC ALPHA 21164/437MHz processor in the Center of Scientific Computing of Finland. When there was no extra load on the computer, chromosomal run times in the Bayesian analyses varied from 46 min to 1 h 25 min / $10^6$ iterations, depending on the length of the linkage group. Running with or without cofactors did not have much influence on the run times. An overview of QTL mapping software, including the packages used in this study for frequentist and Bayesian multiple QTL analysis is described in Manly and Olson (1999).

## *Selection of cofactors, frequentist analysis*

A first round of simple interval mapping (SIM) was applied. Cofactors were selected from regions where the LOD was greater than 3.0 and a subsequent MQM analysis was then performed using these markers as marker cofactors. This low threshold was chosen since in SIM the error variance still comprises genetic variance from other segregating QTLs and therefore the full power of the test is not yet used. In subsequent rounds of MQM mapping, marker cofactors were added or dropped according to this 3.0 threshold. Because of the 'all-markers mapping' approach any marker could be chosen as a cofactor, regardless of the segregation type (informativeness) or the number of missing values. Initially no marker cofactors were used on the linkage group where a QTL was fitted. However, for linkage groups with evidence of a QTL, MQM mapping was also done using cofactors on those same linkage groups, in order to check for the possibility of having detected a ghost QTL.

## *Selection of cofactors, Bayesian analysis*

In the Bayesian approach a preliminary analysis was performed using the multiple QTL model, allowing for up to three QTLs on the linkage group under investigation, but without selecting marker cofactors from other linkage groups. Based on this single analysis, cofactors were then chosen from linkage groups in the regions showing higher than 0.2 posterior probability for single or multiple QTLs. In fact, a rather sharp distinction between linkage groups with and without evidence of a QTL

was observed. For the selected linkage groups, cofactors were chosen from the regions showing high and condensed posterior QTL intensities.

Note that in the Bayesian approach no cofactors were chosen from the linkage group to be analysed because a multiple QTL model was used. Following Sillanpää and Arjas (1999) cofactor genotypes were augmented without using genotype information from neighbouring markers, although in principle this can be done the same way as for the QTL genotypes. Instead, the most informative marker in a region (distinguishing four genotypes in the progeny) was chosen as a cofactor, or a set of two or more cofactor markers, in order to maximise the information content.

## 7.3 Results

### *Preliminary analysis (no cofactors used):*

*Frequentist QTL analysis (SIM)*

For the PEN data the threshold was exceeded on five linkage groups, including L01 and L10. For the RF data, the LOD score threshold of 3.0 was exceeded on linkage groups L01 and L10 (Table 7.1). The LOD score graph for PEN for L01 clearly showed a double peak, the graph of RF was rather irregularly shaped, but also showed multiple peaks (Fig. 7.2). The 2-LOD support intervals for the PEN and RF data for L01 and L10 are presented in Table 7.3 and indicated in Fig. 7.2.

*Bayesian QTL analysis*

The preliminary Bayesian analysis of the PEN data provided evidence of a single QTL on linkage groups L01 and L10. On all other linkage groups the posterior probability of one or more QTLs was less than 0.04. The analysis of the RF data resulted in three linkage groups where the posterior probability of at least one QTL exceeded 0.5. For linkage group L15 the posterior probability for 1 QTL was 0.26. On all other linkage groups the posterior probability of one or more QTLs was less than 0.08 (Table 7.1). The data did not support the existence of more than a single QTL on any of the linkage groups, although for linkage group L01 a strong bimodality was observed, both for the PEN and the RF data (Fig. 7.2). The credible regions for the PEN and RF data are indicated in Table 7.3, together with the posterior probabilities of containing at least one QTL in these respective regions.

**Table 7.1 Preliminary analysis (without cofactors).**
**Maximum LOD score and the posterior distribution of the number of QTLs in linkage groups with evidence of a QTL in simple interval mapping and/or Bayesian analysis with penetrometer and acoustic resonance frequency data.**

| Linkage group | *Frequentist:* | *Bayesian:* | | |
|---|---|---|---|---|
| | Maximum LOD | Posterior distribution | | |
| | | Number of QTLs | | |
| | | 0 | 1 | 2 or 3 |
| Penetrometer | | | | |
| L1 | 6.5 | 0.56 | 0.43 | 0.00 |
| L3 | 3.6 | 1.00 | 0.00 | 0.00 |
| L8 | 4.7 | 0.96 | 0.03 | 0.00 |
| L10 | 7.4 | 0.09 | 0.91 | 0.00 |
| L15 | 3.5 | 0.98 | 0.01 | 0.01 |
| others | <3 | > 0.98 | < 0.01 | < 0.02 |
| Resonance Frequency | | | | |
| L1 | 4.4 | 0.26 | 0.73 | 0.01 |
| L10 | 4.5 | 0.31 | 0.68 | 0.00 |
| L11 | 1.5 | 0.25 | 0.73 | 0.03 |
| L15 | 2.6 | 0.73 | 0.26 | 0.01 |
| others | <3 | > 0.92 | < 0.08 | < 0.03 |

## *Choice of cofactors and final analysis:*

*Frequentist QTL analysis*

Based on the results of the preliminary analysis, cofactor markers were chosen on L01, L03, L08, L10 and L15 for the PEN data. Using these cofactors in MQM mapping, the maximum LOD score of linkage group L03 decreased to a value below 3.0, so that the marker on L03 was dropped as a cofactor. No new regions with LOD scores greater than 3.0 were obtained in the next round. The final model for the PEN data included cofactors on linkage groups L01, L10 and L15. On these linkage groups the LOD significance threshold of 4.5 was exceeded. The LOD score for linkage group L08 (4.3) was just below this threshold (Table 7.2). Compared to the SIM results, the 2-LOD support interval on L01 was just a little bit smaller, on L10 the interval was identical (Table 7.3). For the RF data, two cofactor markers were chosen, on L01 and L10.

**Table 7.2 Final analysis (including cofactors).**
**Maximum LOD score and the posterior distribution of the number of QTLs in linkage groups with evidence of a QTL in MQM mapping and/or Bayesian analysis with penetrometer and acoustic resonance frequency data.**

| Linkage group | *Frequentist:* | *Bayesian:* | | |
|---|---|---|---|---|
| | Maximum LOD | Posterior distribution | | |
| | | Number of QTLs | | |
| | | 0 | 1 | 2 or 3 |
| Penetrometer | | | | |
| L1 | 7.6 | 0.22 | 0.78 | 0.00 |
| L3 | 2.6 | 1.00 | 0.00 | 0.00 |
| L8 | 4.3 | 0.99 | 0.01 | 0.00 |
| L10 | 9.2 | 0.05 | 0.95 | 0.00 |
| L15 | 5.8 | 0.79 | 0.21 | 0.00 |
| others | <3 | > 0.98 | < 0.01 | < 0.03 |
| Resonance Frequency | | | | |
| L1 | 4.2 | 0.41 | 0.59 | 0.01 |
| L10 | 5.0 | 0.43 | 0.56 | 0.01 |
| L11 | 1.5 | 0.94 | 0.06 | 0.00 |
| L15 | 4.0 | 0.37 | 0.62 | 0.01 |
| others | <3 | > 0.91 | < 0.09 | < 0.01 |

In the analysis of the RF data with these cofactors, a LOD score of 4.0 was obtained on linkage group L15. LOD scores for the markers on L01 and L10 remained above 3.0. A cofactor marker on linkage group L15 was added and in the next round no new genome regions with LODs over 3.0 were found. The final model for the RF data included only the cofactor marker on linkage group L10. LOD scores for the other linkage groups were below the significance threshold of 4.5 (Table 7.2). For the RF data also, the 2-LOD support interval for L01 was a bit smaller than in SIM, but much larger than the intervals estimated for the PEN data. The L10 interval was practically identical to the situation in SIM and also to the intervals estimated for the PEN data (Table 7.3). The effect on linkage group L01 was mainly a contrast between the alleles from ‘Prima’, on L10 mainly from ‘Fiesta’ (King and Maliepaard *et al.* 2000).

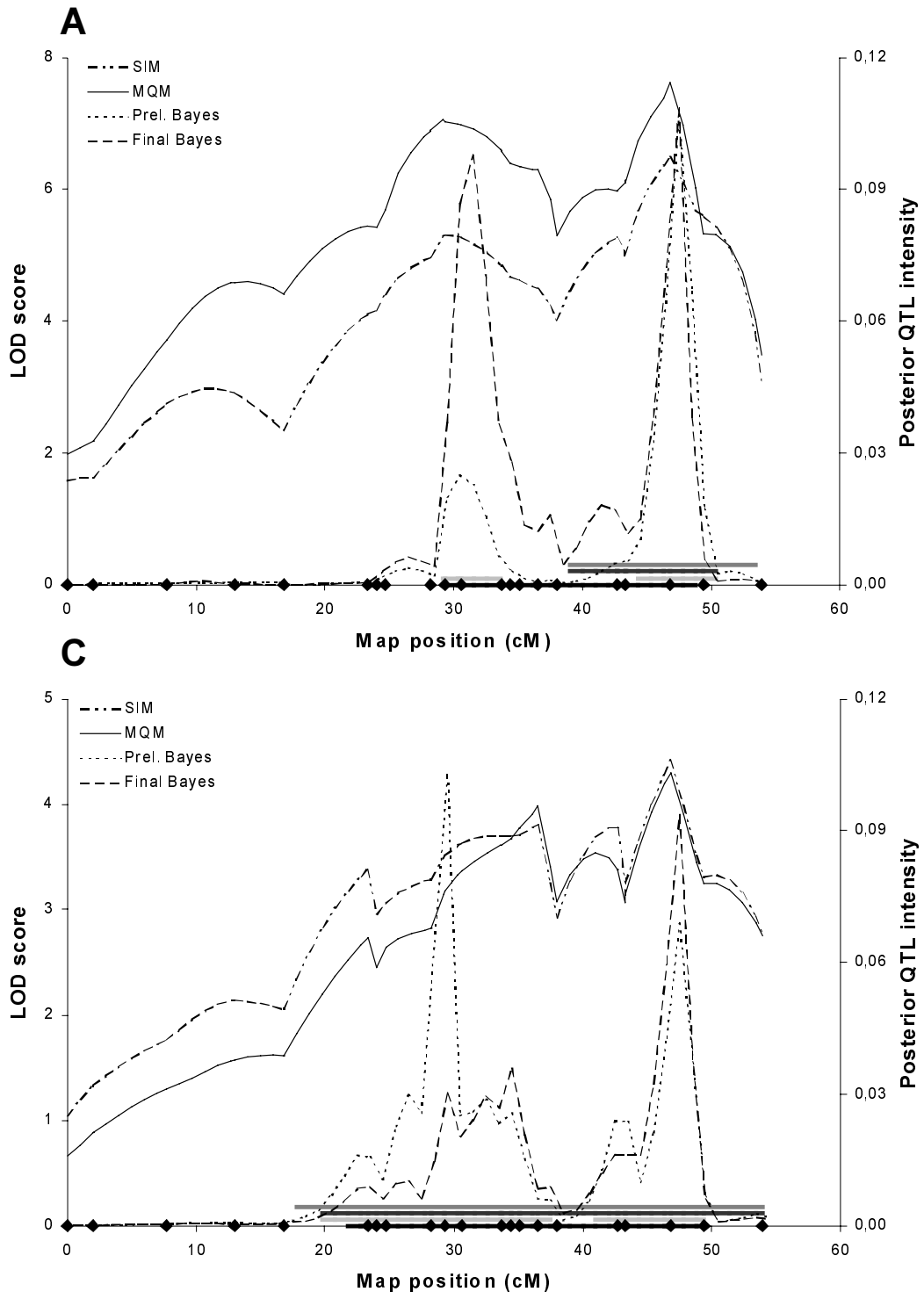
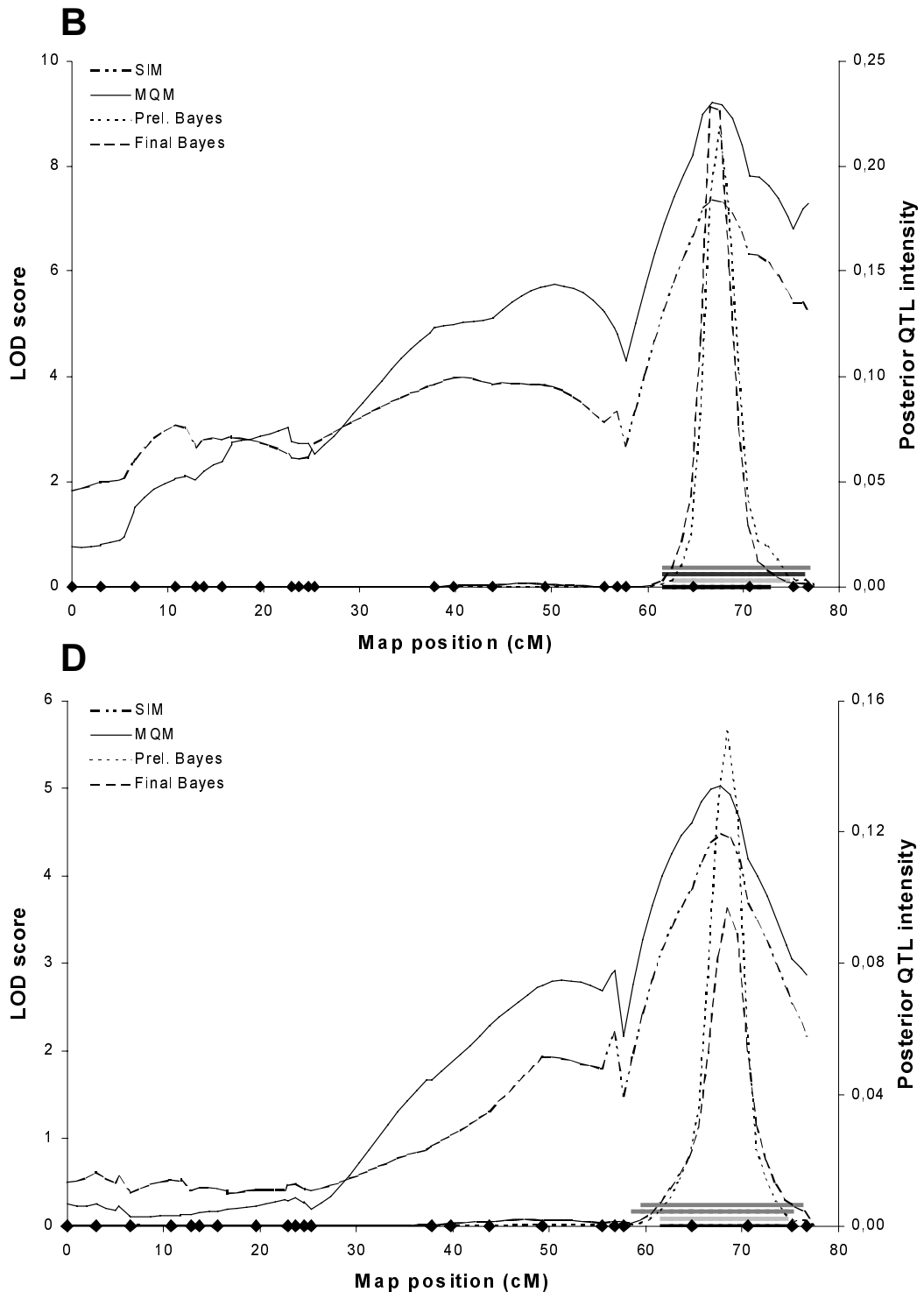



**Figure 7.2 LOD score graphs for interval mapping (SIM) and MQM mapping and Bayesian posterior QTL intensity graphs of preliminary and final analyses of linkage groups which showed evidence of a QTL. Posterior QTL intensities with bin length 1 cM are represented as frequency polygons. Explanations of the different line types are given in the upper left corner of each panel. Bars at the bottom of each graph indicate 2-LOD support intervals and Bayesian credible intervals for the four situations, in the same order (top to bottom) as the legend.**




**Symbols at the horizontal axes indicate marker positions. The left (right) y-axis corresponds to the LOD score (posterior QTL intensity). Labelling of the four panels is as follows: A. Linkage group L01, PEN data B. Linkage group L10, PEN data C. Linkage group L01, RF data D. Linkage group L10, RF data.**

**Table 7.3 2-LOD support intervals in frequentist QTL analysis (SIM and MQM) and credible regions for Bayesian analysis (preliminary and final analysis) with Bayesian posterior probabilities of containing at least one QTL in the corresponding credible regions, for linkage groups L01, L10 and L15.**

| Linkage group | Preliminary analysis (without cofactors) | | | Final analysis (including cofactors) | | |
|---|---|---|---|---|---|---|
| | *Frequentist:* SIM | *Bayesian:* Preliminary | Posterior Probability | *Frequentist:* MQM | *Bayesian:* Final | Posterior Probability |
| Penetrometer | | | | | | |
| L1 | [39.0, 53.4] | [29.1, 33.5] ∪ [44.2, 49.9] | 0.32 | [39.0, 50.4] | [29.9, 49.3] | 0.51 |
| L10 | [61.7, 76.2] | [63.1, 74.3] | 0.59 | [61.7, 76.2] | [62.3, 72.7] | 0.60 |
| L15 | | | | [0.0, 17.4] | [0.0, 9.1] | 0.14 |
| Resonance Frequency | | | | | | |
| L1 | [17.8, 53.9] | [20.1, 37.5] ∪ [40.8, 49.9] | 0.51 | [23.3, 53.9] | [22.1, 49.3] | 0.43 |
| L10 | [59.7, 76.2] | [62.2, 73.4] | 0.49 | [58.7, 76.2] | [61.6, 74.6] | 0.41 |
| L15 | | [0.0, 2.1] ∪ [7.3, 9.5] ∪ [25.4, 37.0] | 0.16 | [0.0, 32.9] | [0.0, 17.4] ∪ [23.8, 30.9] | 0.43 |

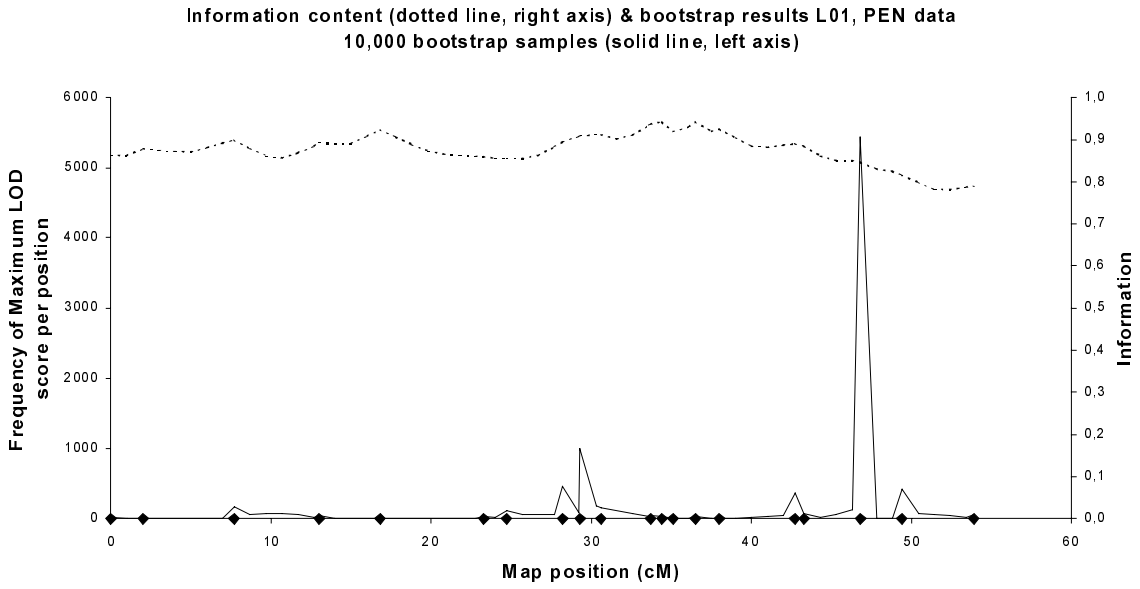



**Figure 7.3 Information content for linkage group L01, using 10 neighboring marker intervals for upgrading linkage information for missing or partially missing marker genotypes (right axis, upper graph) and empirical frequency distributions of the estimated positions of QTLs represented as frequency polygons (right axis, lower graph). Results from interval mapping of bootstrap samples of the PEN data from 152 individuals in the 'Prima' × 'Fiesta' progeny, linkage group L01. A step size of 1 cM was used. Symbols at the horizontal axes indicate marker positions. The left y-axis corresponds to the frequency of the maximum LOD score per map position.**

*Bayesian QTL analysis*

For the PEN data three cofactor markers were chosen: two on linkage group L01 and one on linkage group L10. For the RF data a total of seven cofactors were chosen on linkage groups L01, L10, L11 and L15. These consisted of two markers on L01 to cover a larger area of this linkage group (to account for the two peaks), two on each of linkage groups L11 and L15 to cover for all possible allelic combinations and also a larger area, and one informative marker on L10.

For the PEN data there was strong evidence for a QTL on linkage groups L01 and L10, some evidence of a QTL on L15 and hardly any evidence for a QTL on other linkage groups. In the analysis of the RF data with cofactors, most linkage groups did not present any evidence of the presence of one or more QTLs. There was evidence of a single QTL on linkage groups L01, L10 and L15 (Table 7.2). There was no support for more than a single QTL on any of these linkage groups. The posterior probability of a QTL on linkage group L11 decreased considerably when cofactors were used. Figure 7.2 shows some of the results from both analyses with and without cofactors. Credible regions for the PEN and RF data are given in Table 7.3, together with the posterior probabilities for the presence of at least one QTL in these respective regions.

## 7.4 Discussion

### *Comparison of the frequentist with the Bayesian approach: correspondence*

The frequentist and Bayesian approach for multiple QTL analysis were both applied to two data sets, collected from a single full-sib family of apple. The data sets, PEN and RF, were different in the way the fruit firmness phenotypes were measured. Phenotypes in both data sets were genotype means over different sites and years. King and Maliepaard *et al.* (2000) have analysed the individual data sets per site and year previously, together with additional sensory measurements of fruit firmness. These individual data sets indicated possible QTLs on linkage groups L01 and L10 and some evidence for QTLs on L15. Also in the present study evidence was found for QTLs on these linkage groups, and for these there was a good agreement between the results of the frequentist and the Bayesian method. Both methods indicated a single QTL for fruit firmness (both for the PEN data and the RF data) on linkage groups L01 and L10, and perhaps L15 (in the frequentist approach stronger evidence for L15 came from the PEN data, whereas in the Bayesian approach it was from the RF data). For these linkage groups there also was a very good correspondence of the estimated QTL positions, as indicated by the LOD score and Bayesian QTL intensity peaks (Fig. 7.2).

There was no indication of a second QTL on either of these linkage groups, even though the graphs for linkage group L01 showed bimodality for the Bayesian intensity for both the PEN and the RF data. This was also visible for the LOD score but less extreme than in the Bayesian analysis. This bimodality could not be explained by the variation in the information content, since the information content is very high throughout this linkage group and there is no visible decrease in information in the region where the posterior density or the LOD score drops (Fig. 7.3). Upon inspection of the marker data, it was found that there were three double recombinants in the region between 36 cM and 43 cM. However, these could not explain the decrease in the LOD score curve as was verified by more detailed inspection (results not shown). When studying the single recombinants in the region between 29 cM and 43 cM, it was observed that the phenotypic values of these individuals indeed could explain a decrease in significance going from 29 to 38 cM, and a subsequent increase in significance from 38 to 43 cM. It seems likely that the sampling bias among the recombinants is responsible for the observed bimodality of the curves. It can not be excluded that there are some errors in the marker data or in the map order and/or the map distances.

## *Comparison of the frequentist with the Bayesian approach: differences*

Although the results of the Bayesian and the frequentist approach agreed very well on QTLs on linkage groups L01 and L10, still some differences between the methods were observed. In the preliminary frequentist analysis of the PEN data, LOD scores larger than 3.0 were also found on linkage groups L03, L08 and L15. The preliminary Bayesian analysis did not provide evidence for QTLs on these linkage groups. On the other hand, the preliminary Bayesian analysis resulted in an elevated posterior density for a QTL on linkage group L11 (RF data), whereas with the frequentist approach a maximum LOD score of only 1.5 was obtained (Table 7.1). Additionally, the non-parametric test of Kruskal-Wallis was used to verify the results without being necessitated to assume normality of the data. This test also indicated possible presence of QTLs on linkage groups L03 and L08 and slight evidence for a single marker on linkage group L11.

In this study rather diffuse priors for parameters in the model were used, and therefore we do not believe that posterior inferences were noticeably influenced by the priors. It is also unlikely that the use of the ML procedure gave rise to the high LOD scores in linkage groups L03, L08 and L15, since the LS method (Haley *et al.* 1994) resulted in almost equal LOD scores. Apparently, the difference between the Bayesian and frequentist method is also not just a matter of a difference in power: evidence for a QTL is found with one method and not with the other, as well as *vice versa*. Therefore, we do not have a satisfactory explanation for these differences yet.

## *Cofactor choice and its effect*

In both methods the choice of cofactors was based on the results from the preliminary analysis, so marker cofactors were chosen only in those regions with elevated LOD scores or high QTL intensities. Note that this does not necessarily provide us with the optimal set of cofactors. For example, in MQM when two QTLs are present on a linkage group, the highest LOD score may be found in between these two (as a ghost QTL). Choosing a cofactor at that position may absorb most of the genetic effects generated by the two QTLs, so that these will not be detected in subsequent rounds. This could be prevented by performing a backward elimination procedure to select cofactors on a linkage group of interest (Jansen 1993). In this study possible ‘ghost QTLs’ were checked for on hindsight, by using different pairs of cofactors on those linkage groups with evidence of a QTL. In the Bayesian multiple QTL model, this ‘ghost QTL’ behaviour is not expected. Choice of an incorrect cofactor may also occur when the information content is rather variable across the linkage group. In smaller data sets there is also the danger that, due to chance, a major QTL has distorted segregation within the marker classes of a marker on a different linkage group, or partial cosegregation with an unlinked

marker. When this occurs, another type of 'ghost QTL' may be detected on the latter through association with the real QTL. In fact, partial cosegregation was observed between a set of marker pairs on linkage groups L01 and L08 and this may explain the decrease in the LOD score for L08 when a cofactor on L01 was used. However, currently it is not yet feasible to compare and evaluate efficiently different possible sets of cofactors. The logical solution to this problem would be to consider the entire genome in a single multiple-QTL analysis.

Although the choice of cofactors was based on the same principles in both methods, the cofactors were used differently in the models, and also the choices were different. As a consequence of this, the effect of using the marker cofactors seemed different in the two methods. In the frequentist method the effect was generally an increase of the LOD scores, while the shape of the LOD score graph remained very similar. In the Bayesian analysis the differences were more notable. For L11 the change was rather drastic, since the posterior probability for a single QTL on this linkage group decreased from 0.73 down to 0.06. The increase for L15 (from 0.26 to 0.62) was also rather large. For the Bayesian analysis, the shape also changed. For example, for the PEN data the bimodality on L01 became stronger when cofactors were used, and the intensity in the region around 30 cM increased so that the two peaks became almost equally high.

## *Position estimates and their accuracies*

The estimated positions of QTLs were very similar, for both methods and data sets and regardless of cofactors chosen. Especially for linkage group L10, the correspondence is striking. The QTL position estimates obtained with the Bayesian approach visually appear to be more accurate than the results with the frequentist approach. This is because they show sharper peaks, whereas the LOD curve is rather flat. The chosen credible regions, however, are not very different from the 2-LOD support intervals. Note that the results cannot be compared directly. With respect to the visual appearance of the peaks, the results from the Bayesian analysis were expected to be more comparable to results from bootstrapping. Indeed, these were more similar (Fig. 7.3). A part of the difference is explained by the logarithmic scale of LOD scores, whereas results from the Bayesian analysis and from bootstrapping are based on frequency distributions. In the bootstrapping results we observed a similar bias as Walling *et al.* (1998), resulting in higher frequencies of the maximum LOD score at the marker positions, especially near the estimated QTL position. Where the LOD support intervals are concerned, these are defined differently than confidence intervals. Van Ooijen (1992) demonstrated, both for BC and F2 populations, that 2-LOD support intervals may be conservative only if the QTL effect is large. Dupuis and Siegmund (1999) showed with simulations that 1-LOD and 1.5-LOD support intervals provided a QTL coverage probability of approximately 90% and 95%, respectively, for dense maps (markers at every 1 cM)

and an even greater percentage for sparse maps. These authors also compared confidence regions in simulations with a single QTL and concluded that the coverage probabilities of LOD support regions and Bayesian credible intervals were roughly comparable in large samples.

## *Multiple linked loci*

In this study we found no evidence for more than a single QTL on any linkage group, so that we were not able to compare the performance of the two methods when more QTLs are present. In general, the Bayesian method seems to be well suited to detect multiple QTLs on a linkage group since these are modelled explicitly. This is supported by simulation studies (Sillanpää and Arjas 1998, 1999). Although MQM mapping can also be used to detect multiple QTLs on a single linkage group, long computation time may be required if there are also cofactors on other linkage groups and if the 'all-markers mapping' approach is applied to upgrade marker information for all cofactors and the fitted QTL. This may be solved by using more multi-allelic markers, and by omitting those QTL genotype combinations which have a probability close to zero (Jansen 1995).

## *Environmental cofactors*

In neither the frequentist nor the Bayesian method we included environmental cofactors such as the site/year combinations in the model, although this is certainly possible and has been done *e.g.* by Jansen *et al.* (1995), Tinker and Mather (1995), and Korol *et al.* (1998a). This would also be more in agreement with the Bayesian paradigm of using all prior information and of including uncertainties rather than using point estimates. The use of estimated means over sites and years for QTL mapping may have undesirable effects since some genetic effects may be lost by the adjustment for environmental cofactors. It would be preferable to include also these environmental cofactors into the analysis and estimate all effects simultaneously. However, this would be computationally more demanding and the sample size in this case would not allow for the reliable estimation of all main and interaction effects.

## *Conclusion*

Both methods provided evidence for the main QTLs on the same linkage groups, and with similar map positions. However, there were also some differences, with respect to evidence for QTLs on other linkage groups. The response to adding cofactor markers was also somewhat different. The shape of the graphs of the LOD score and Bayesian posterior intensity were found to differ as well. Neither method provided evidence for more than a single QTL on any linkage group.

## 7.5 Appendix

Here we comment briefly on the order in which the updating of the MCMC algorithm takes place. There are two strong dependency relations in the offspring data: The vertical dependency between parents and their offspring, and the horizontal dependency between adjacent loci in each individual. If single-site updating is applied, the sampler can easily get stuck in some part of the sample space because of these dependencies (Sheehan and Thomas 1993; Janss *et al.* 1995; Jensen and Sheehan 1998). To facilitate movement in the sample space (mixing) of the MCMC sampler, especially in cases in which a large proportion of the data is missing and the markers are very close to each other, the following two-directional blocking scheme was implemented to the sampling algorithm of the Multimapper/OUTBRED program of Sillanpää and Arjas (1999).

This modified program version is currently available on the web (http:/www.rni.helsinki.fi/~mjs) and it was applied in all cases in this study.

Step 2 (APPENDIX A) in the sampling scheme of Sillanpää and Arjas (1999) is modified in the following way:

With probabilities 0.5, the sampler does either a (1) FAMILY BLOCK-UPDATE or an (2) INDIVIDUAL UPDATE.

(1) FAMILY BLOCK-UPDATE is similar as before, except for Step 2.5. In the new version, the grandparental origins are determined for offspring alleles having a heterozygous parent, but are proposed directly from the prior (Equation 4) for alleles inherited from homozygotes. The acceptance ratio is then modified accordingly.

(2) INDIVIDUAL UPDATE: Proposals covering the entire chromosome (all markers jointly) of each offspring are constructed similarly as in Step 2 of APPENDIX A, but their acceptance is tested separately for each haplotype proposal of each individual. The acceptance ratio is again modified accordingly.

### *Acknowledgements*

We are grateful to Dr. Marco Bink for helpful comments to the manuscript. The experimental data were kindly provided by the partners in the CEC project CT920473, 'Development of the European Apple Crop', which was funded by the CEC Framework III research and technical development programme in Agriculture and Agricultural Industries and national governments of the EU.

# General discussion

In this thesis the use of molecular markers in genetic analysis of full-sib families of outbreeding (plant) species has been investigated. A genetic linkage map of apple has been constructed and this map was used for the detection of loci involved in monogenic and quantitative traits. Methods for the analysis of quantitative trait loci (QTLs) were compared using experimental data from the progeny of an apple cross, grown at different sites in Europe.

## Current situation

The genetic analysis of apple has always been hampered by its strictly outbred mode of reproduction, its long juvenile period and the high costs involved in maintaining breeding populations. Molecular markers have provided a possibility to detect and locate genes of interest to the breeder, and a possibility to select for those traits in an early phase of the breeding program. Most work using molecular markers in apple has been focused on genes for resistance to scab, such as *Vf* (Gardiner *et al.* 1996; Gianfranceschi *et al.* 1996; Hemmat *et al.* 1995; King *et al.* 1998; Koller *et al.* 1994; Tartarini 1996; Tartarini *et al.* 1999; Yang and Krüger 1994; Yang and Korban 1996; Yang *et al.* 1997a, 1997b) and *Vr* (Cheng *et al.* 1996) and *Vm* (Cheng *et al.* 1998). However, also genes for resistance to powdery mildew (Markussen *et al.* 1995), insect resistance (Roche *et al.* 1997a, 1997b; Bus *et al.* personal communication), fruit skin colour (Cheng *et al.* 1996) and tree type (Hemmat *et al*. 1997) have been studied. Allele-specific isozyme and PCR markers have been developed and investigated for a number of self-incompatibility alleles of apple (Bošković and Tobutt 1999; Broothaerts *et al.* 1995; Janssens *et al*. 1995). Recently the array of marker types in apple has been extended from isozymes, RAPDs, RFLPs to AFLPs and simple sequence repeats (SSRs) (Gianfranceschi *et al.* 1998; Guilford *et al.* 1997; Hokanson *et al.* 1998; Maliepaard *et al.* 1998). Isozymes, RFLPs and SSRs have been used to align linkage maps of different apple cultivars (Chevreau *et al.* 1999; Conner *et al.* 1997; Gianfranceschi *et al.* 1998; Hemmat *et al.* 1994; Maliepaard *et al.* 1998). These studies have yielded selectable markers, such as SCARs, which can be used in breeding programs in addition to the existing selection methods (Cheng *et al*. 1996; Gianfranceschi *et al.* 1996; Roche *et al*. 1997b; Tartarini *et al.* 1999). A few studies have indicated correspondences between the apple genome and genomes of related species (Chevreau *et al.* 1997; Chevreau *et al.* 1999; Joobeur *et al.* 1998). Recently, also quantitative trait loci in apple using molecular markers have been studied (Chapters 6 and 7; Conner *et al.* 1998; King and Maliepaard *et al.* 2000) and a bacterial artificial chromosome (BAC) library has been constructed (Vinatzer *et al.* 1998). Currently, activities are underway for physical mapping and map-based cloning of *Vf* (Vinatzer *et al.* 1998; Patocchi *et*

*al.* 1999) and for physical mapping and comparative genome analysis of the *Sd1* region (Cevik and King, personal communication).

Despite this considerable progress where genetical research in apple is concerned, molecular markers are not very widely used in apple breeding programs. In two or three breeding programs in the world, markers are now used for pyramiding resistance genes, or for the selection of *Vf* homozygotes (Bus *et al.* personal communication; Kellerhals *et al.* personal communication). For some simply inherited resistance genes, the possibilities of marker-assisted selection are now being evaluated within the regular breeding program (Bus *et al.* personal communication).

This limited use of markers in breeding can partly be ascribed to the time lag expected between experimental research on marker-trait associations and introduction of markers into the breeding program. However, there are more serious limitations to be considered. In part, these are of a more general nature, partly these are specific to outbreeding species or to tree breeding.

## Monogenic traits

For monogenic traits the detection of closely linked markers has been proven successful in many crops and for many traits. If the phenotype can be scored unambiguously it is usually no problem to find linked markers and to map the gene. If the phenotype (or the markers) cannot be scored unambiguously, this will obviously affect the precision with which the gene is mapped. Even so, the heritability of such traits will usually be very high and despite some imprecision, it can be expected that closely linked markers can be found, which, usually after being transformed into allele-specific PCR markers, can be used for subsequent selection. Using a pair of markers flanking the gene at small distances, the selection of the gene by markers is straightforward. In selfing species, marker-assisted introgression of single genes has become routine.

However, there still may be an important problem: a marker which co-segregates with a gene in one progeny need not segregate at all in another, or may be in repulsion phase to the favourable allele. Therefore the linkage phase and segregation may need to be assessed again in (small parts of) new progenies. Obviously, this may be strongly disadvantageous if the marker was meant to be used for early selection in the breeding program. In some cases, for example when marker and gene were introgressed simultaneously into the breeding material, *e.g.* from wild relatives, the marker may be useful across large parts of the breeding program although recombinants may also be present (King *et al.* 1999). In selfing species it is, of course, possible to introgress the desired allele into elite germplasm by recurrent backcrossing. Strategies for marker-assisted selection in recurrent backcross programs of selfing species have been described, *e.g.* by Hospital and Charcosset (1997), Hospital *et al.* (1992, 1997), Frisch *et al.* (1999a, 1999b) and Visscher *et al.* (1996a). By doing this, the association between marker and desired allele is

transferred to the improved elite line which can be used subsequently for new crosses. In strictly outbreeding species the recovery of recurrent parent genome through backcrossing is not possible. It is possible, however, to use an observed association between a marker and a desired allele, in subsequent generations, or in crosses with the same parent as in the original mapping experiment.

Another aspect to be considered is that for highly heritable traits conventional selection is usually quite efficient and in those cases marker-assisted selection (MAS) will only be profitable if the trait itself is expressed late in the development of a plant (*e.g.* fruit quality traits), or when the conventional test for the determination of the phenotype is expensive, unreliable, or impossible to perform (*e.g.* no pathogen available for plant resistance tests). Even in those cases the combined expenses of finding and/or developing suitable markers and of subsequent MAS must be weighed against the costs of the available alternatives (Luby and Shaw, submitted). Obviously, costs may be reduced if marker tests for other traits are already being done as well.

## Quantitative traits

Where quantitative traits are concerned the situation is much more complex. Beavis (1998) recently concluded that 'despite hundreds of QTL experiments during the last 10 years, molecular markers have had little impact on crop improvement', and ascribed this to inadequate experiments (yielding little consistency in identified QTL in different experiments and imprecise and biased estimates of genomic locations and genetic effects) and failure to integrate QTL mapping methods into existing breeding programs. He was referring mainly to simple interval mapping in progenies of moderate size, and expected improvements from MQM methods which incorporate cofactors representing other QTLs, and methods incorporating multiple related crosses and/or larger progenies. Such methods would be better suited for bridging the gap between QTL mapping experiments and the wider breeding program.

Where selfing species are concerned, experiences with and prospects of marker-assisted breeding are quite favourable. Marker-assisted introgression of QTLs from wild species has been proven successful in some species (*e.g.* Bernacchi *et al.* 1998a, 1998b plus erratum in 1998c). In a backcrossing program, markers can be used not only for the introgression of a desired allele but also for the simultaneous recovery of elite parent genome. It has been demonstrated that this recovery of elite parent genome can be accelerated considerably by using markers (Hospital and Charcosset 1997; Hospital *et al.* 1992, 1997). Van Berloo and Stam (1998) have demonstrated in a simulation study that for the purpose of pyramiding QTLs marker-assisted selection may in some cases be preferable to conventional selection procedures. Vuylsteke (1999) used molecular markers for identifying QTLs associated with hybrid performance and heterosis in maize. However, it should be

realised that all these examples concern breeding methods in species allowing selfing, and these cannot be generalised to strictly outbreeding species.

## *Methodological concerns*

Existing QTL mapping methods have rather small power for detection of QTLs in the population sizes generally used (Beavis 1994; Van Ooijen 1992). Therefore, in a given QTL mapping experiment, possibly only a subset of QTLs will be detected. The power decreases with the heritability, which leads to the paradox, as was pointed out *e.g.* by Knapp (1998), that QTL detection power is smallest for those traits where MAS would have the greatest expected impact in comparison with phenotypic selection (Lande and Thompson 1990)

The estimated effects of significant QTLs will be upwardly biased. This follows from theoretical considerations (Lande and Thompson 1990) but was also observed in empirical QTL verification studies (Austin and Lee 1996a, 1996b; Melchinger *et al.* 1998). This can be ascribed to the significance threshold. Significant QTLs will have, on average, positive bias due to sampling and effects not taken into account such as QTL $\times$ environment interactions, epistatic effects, and due to confounding of recombination frequency estimates, or of errors in the map distances of markers, with the estimated QTL effects (this may cause bias in both directions, but here also, significant QTLs will usually be the ones biased upward).

The position of detected QTLs cannot be determined very precisely: even when mapping populations are sufficiently large to detect a number of QTLs, the precision of locating a QTL is usually not very high in experimental populations, even when sophisticated methods for QTL mapping, such as MQM are used (Kearsey and Farquhar 1998). One of the important limitations in QTL resolution is the low chiasma frequency per chromosome (Kearsey and Pooni 1996), others being missing values, errors in genotyping and phenotyping, linked QTLs, and undetected QTLs on other linkage groups which are not accounted for (Lee 1995). Very large populations are needed to decrease the size of confidence regions for QTLs (Van Ooijen 1992). Alternatively, the number of generations can be increased to obtain higher numbers of recombination events per chromosome. This is being done in inbred species by using recombinant inbred lines (RILs), but obviously more time is required to produce these, and the method cannot be applied in strictly outbreeding species. A comparable method for outbreeding species, employing full-sib intercross lines (FSILs) was proposed by Song *et al.* (1999), but this requires an initial cross and intercrossing of full-sibs during two or more generations. In the case of apple, this would require considerable resources, due to the long generation cycle and the requirement of maintaining unselected progenies.

A consequence of these problems for MAS is that only a limited part of all QTLs can be selected for, that it will be difficult to determine which are the most important ones, and that the combined effect of the QTLs selected for is expected to be

smaller in later progenies than was originally estimated. Another consequence for marker-assisted breeding is that rather large parts of the linkage group containing a QTL would need to be selected, in order to ascertain a high probability of selecting for the desired QTL allele. As a consequence of linkage drag, undesirable genes in this region would then be selected as well. On the other hand it must be realised that these need not be present. Moreover, other favourable alleles might also be present in the region.

### *Generalisation of detected QTL-marker associations*

However, there are even more serious limitations to be considered and these have to deal with Beavis's second argument: failure of making the transition from QTL mapping experiments to the breeding program (Beavis 1998). In a QTL mapping experiment we usually deal with a specifically designed segregating population, usually from a single cross, sometimes from a very wide cross, which is evaluated usually for only a subset of the traits that the breeder is interested in. Marker-phenotype associations detected in this one progeny do not necessarily hold true in progenies from other crosses: in those other progenies, the marker or the QTL may not segregate, or a marker allele which was in coupling phase with a positive QTL allele in the mapping progeny may be in repulsion phase in another progeny. Therefore, associations between markers and traits of interest would need to be verified in other crosses. However, it must be realised that if the heritability of these QTLs is small, verification cannot be done in a small subset of these new progenies, in other words, a whole new QTL experiment would be needed. Rather than smaller populations, such QTL verification would demand unselected and therefore larger populations than the breeder would normally use, and marker assessment would be needed additional to phenotype scoring, so that costs would be much higher than in traditional selection.

Even when the marker-QTL association holds true in another progeny, the effect of the QTL may be much smaller, first of all because of the upward bias in the QTL mapping experiment, but also because of the different genetic background. In fact, in a new progeny other QTLs not detected in the original mapping progeny (and possibly not even segregating there), may even be the more important ones.

### *Unpredicted effects on other traits*

Where a breeder needs to consider a large number of traits simultaneously, QTL experiments are usually devoted to only a small subset of traits. The effect of marker-assisted selection for the QTL on the other traits cannot be predicted, but due to linkage drag and drift in the selected subset of the population, progress for other traits may be smaller, so that possible gains for the trait under selection must be weighed against a possibly smaller gain, or even loss, for other traits of importance. On the other hand, it should also be realised that even though in traditional breeding a large number of traits are considered simultaneously for

selection, the knowledge about the genetics of these traits, especially quantitative traits, is often very poor. QTL experiments provide a possibility of accumulating genetic knowledge which in traditional breeding would not very likely become available to the breeder. A stepwise approach with a first selection based on markers linked to important QTLs and subsequent phenotypic selection for other traits, may still be profitable, depending on (a) the extra costs involved in the production of a larger progeny and the marker tests, (b) gains expected from more effective selection of the QTLs, and (c) possible losses for some other traits, which, it should be realised, might also have occurred with traditional selection procedures.

## *Problems specific to outbred crosses*

With respect to the possibilities of pyramiding favourable QTL alleles, one must realise that introgression breeding by recurrent backcrossing, which is possible in inbred species, is not feasible in most outbreeding species, and certainly not practicable in apple. Also, the construction of QTL substitution lines, as in inbred species, is not possible. A 'once and for all' fixation of positive QTL alleles across the breeding program will also not be feasible in a strictly outbreeding species, since in every new cross homozygosity at a QTL can be lost. However, markers linked to a QTL do provide a better possibility to maintain the desired QTL allele into the next generation, even though this would require some marker tests for each new progeny. With traditional selection methods this would be difficult, since it would hardly be possible to ascertain the presence of a specific QTL allele based on the phenotype of a single plant. With markers it may also be possible to make a better choice of parents for new crosses, based on previously collected information on QTL alleles present in these parents.

Although substitution lines cannot be created in strictly outbreeding species, it may be worthwhile to consider a possibility to investigate specific parts of the genome by comparing selections sharing a part of the genome, derived from a common ancestor, with selections which do not share that particular part of the genome. In apple breeding where typically a restricted set of founders is used extensively through the full width of the breeding program, it may very well be possible to map 'founder QTLs' by such a type of analysis.

## *Costs and logistics*

Costs of molecular markers are still very high and a breeding program wide use of molecular markers would require high input in terms of expertise and finances. It remains to be seen whether cost effectiveness can be achieved for marker-assisted selection, especially in the case of quantitative traits, even when markers allow early selection for some traits in a fruit breeding program. Luby and Shaw (submitted) point out that the logistics of marker-assisted selection should be carefully considered: marker-assisted selection would be most profitable when undesirable seedlings can be selected against before these are planted in the nursery or the

orchard. This would require screening large numbers of seedlings in the short period that these are in trays or pots, in the greenhouse. If only a subset of the progeny, or only the potential parents, were to be tested, it can be questioned whether the savings will be enough to justify implementation of the technology.

## Conclusions

Currently, marker-assisted selection for QTLs in apple is not yet feasible, at least not in the full width of the breeding program, and for the time being the question remains whether it can be cost effective at all. The use of QTL analysis as a research tool for the dissection and characterisation of quantitative traits remains highly valuable, beyond any doubt, and this may indeed provide the breeder with extra tools for a more effective selection than is currently practised in traditional breeding. Where simply inherited traits are concerned, marker-assisted selection can be cost effective. As Mehlenbacher (1995) also pointed out, the most likely application of MAS in fruit crops will be for early selection of traits that can be evaluated only after a long juvenile phase, such as fruit quality traits, selection of traits that are difficult or expensive to measure (*e.g.* some insect resistances), or for pyramiding, *e.g.* of resistance genes. In addition, MAS can also be used to select for homozygous rather than heterozygous individuals for traits controlled by a dominant gene (Gianfranceschi *et al.* 1998).

It can be expected that in the near future more easy-to-use PCR markers closely linked to single genes become available for traits such as scab resistance, mildew resistance, insect resistance and maybe also fruit acidity, in addition to the ones currently available. Combination of two markers in a single test has already been proven successful (Roche *et al.* 1997b), and can possibly be extended to a higher number of traits to further reduce costs. When costs of marker technology are decreased and more markers for simple traits become available, it is to be expected that marker-assisted selection for such traits will become feasible within commercial breeding programs.

## Prospects

So where does this leave us? Undoubtedly, the analysis of (quantitative) traits using molecular markers has significantly increased our knowledge about plant genetics and provided us with much better tools for studying the genomes that we are dealing with. However, the early optimism concerning prospects for improving plant breeding has obviously been nuanced to some degree. Spelman and Bovenhuis (1998) recently remarked that ‘examples of successful marker-aided selection of QTLs in practical breeding programs have rarely been published’ and that these ‘are still at the theoretical level where outbred species are concerned’.

## *Improvement of current methodology*

Where the methodological problems are concerned, it must be mentioned that over the last years methods for the detection of QTLs have been improved considerably, for example by accounting for different environments (Jansen *et al.* 1995; Korol *et al.* 1998a; Tinker and Mather 1995), accounting for genetic covariables such as unmapped QTLs, QTLs on other linkage groups and multiple QTLs on a linkage group (Jansen 1994a, 1996; Jansen and Stam 1994; Kao *et al.* 1999; Sillanpää and Arjas 1998, 1999; Uimari *et al.* 1996a; Zeng 1993, 1994), and by the combined analysis of correlated traits (Korol *et al.* 1995, 1998b). Some of these improvements have already found their way into software for mapping QTLs (for an overview, see Manly and Olson 1999), and it is to be expected that further improvements will be implemented also.

The effect of using these will be that the detection of QTLs will be improved, that QTL effects may possibly be estimated more accurately and that QTLs can be located with a higher precision. It is also to be expected that we will acquire a better understanding of phenomena such as QTL x environment interaction, epistasis, and more generally the effect of the genetic background. However, it should be realised that the estimation of higher numbers of interaction parameters will also require larger progenies.

For the most important crops, high density maps will become available within the next few years. For those crops, QTL mapping will become a matter of multiple regression of a trait on the marker set, and methodological problems will then be reduced to the familiar concerns regarding model selection.

## *Fine mapping populations*

In addition to improvements in the methodology, the use of populations specifically aimed at fine mapping of QTLs will also help to bridge the gap between QTL experiment and breeding practice (Darvasi 1997, 1998; Darvasi and Soller 1995; Hill 1998; Song *et al.* 1999; Tanksley and Nelson 1996). In tree species with a long generation cycle such as apple, it has not yet been established what the best strategy would be for fine mapping. Obviously, even though it may be costly, large populations can be generated from a single cross. In addition, replication of genotypes by vegetative propagation is possible. This does not increase the number of recombinants, but it does provide a means of more accurate estimation of phenotypic values. A fine mapping approach employing large and/or replicated populations should be compared to approaches using multiple generations or multiple crosses.

## *New methods*

The concerns relating to extending the use of markers linked to QTLs to other progenies are more serious. Beavis (1998) pointed out that for the implementation of MAS for QTLs it would be necessary to develop QTL detection methods which are

directed more toward the situation in a practical breeding program and which can be applied in multiple related crosses.

In apple breeding, where a small number of founder parents have been used extensively, and where the number of generations between these founders and the current modern cultivars is small, it can be expected that it will be possible to trace back the origin of larger and smaller chromosomal regions in the modern cultivars to these individual founders. Monitoring of new crosses, using pedigree and marker information simultaneously across multiple families, would perhaps allow the assignment of QTLs to these specific ‘founder regions’. These could then be used within the wider scope of a breeding program. However, it is to be expected that large numbers of multi-allelic markers such as SSR markers would be required for this purpose and that larger numbers of individuals would need to be evaluated first for markers and traits. Success of this method would depend upon the answer to the question whether identical marker alleles in the breeding population can really be assigned unambiguously to progenitors. Even so, it can be envisaged that large expenses will be needed to set up and evaluate such a method within a breeding program. However, it must also be remarked that in this way, information can be accumulated over generations and transferred or shared between different breeding programs.

# Samenvatting

Kruisbevruchtende gewassen lenen zich niet zo gemakkelijk voor genetisch onderzoek als gewassen die zelfbevruchtend zijn, of waar zelfbevruchting op zijn minst een mogelijkheid is. Als een kruisbevruchter dan ook nog eens een lange generatiecyclus heeft en kruisingspopulaties veel ruimte vergen, zoals dat voor appel het geval is, dan is het opbouwen van genetische kennis een kwestie van hoge kosten en een lange adem. Voor de veredelaar is deze kennis echter van groot belang omdat deze hem of haar beter in staat stelt, gewenste eigenschappen in nieuwe rassen samen te brengen. Methoden om in een veredelingsprogramma in een vroeg stadium te selecteren of met een grotere mate van zekerheid de eigenschappen van nakomelingen van een kruising te voorspellen, kunnen een belangrijke bijdrage leveren aan de effectiviteit van de veredeling, en het ontwikkelen van nieuwe rassen versnellen.

DNA-technieken die in de jaren ’80 en ’90 ontwikkeld zijn hebben het mogelijk gemaakt om variatie in het erfelijk materiaal, bijvoorbeeld van twee kruisings-ouders, zichtbaar te maken, meestal als bandjespatronen op een fotografische film. Deze zichtbare varianten, de bandjes, worden moleculaire merkers genoemd. De merkers corresponderen met vaste plaatsen op de chromosomen en uit de aanwezigheid en afwezigheid van (combinaties van) merkers in de nakomelingen, kan worden nagegaan hoe de merkers verdeeld zijn over de chromosomen en wat hun volgorde op de chromosomen is. Dit wordt weergegeven op een (genetische) kaart.

Een belangrijke toepassing van moleculaire merkers is de kartering van genen die betrokken zijn bij de eigenschappen waarin we geïnteresseerd zijn. Sommige eigenschappen worden bepaald door één enkel gen, en vaak leveren de varianten van zo’n gen (de allelen) duidelijk waarneembare verschillen op in de nakomelingschap, zodat nakomelingen met het ene allel onderscheiden kunnen worden van nakomelingen die het andere allel geërfd hebben. We spreken dan van kwalitatieve verschillen. Met moleculaire merkers is het mogelijk de plaats van zo’n gen op het chromosoom (het zogenaamde ‘locus’) vast te stellen.

Veel eigenschappen, zoals ziekte-resistentie en vruchtkwaliteit, worden echter bepaald door meer genen en worden daarnaast ook nog sterk beïnvloed door omgevingsfactoren, zodat in de nakomelingschap niet een kwalitatief verschil zichtbaar is, maar een continue variatie. Dit zijn kwantitatieve eigenschappen en de bijdragen van afzonderlijke genen (de ‘quantitative trait loci’ of QTLs) zijn niet direct waarneembaar. Biometrici hebben echter statistische methoden ontwikkeld om met behulp van moleculaire merkers QTLs op te sporen, en de bijdrage van elk QTL te schatten (d.i. kartering van QTLs of ‘QTL mapping’). Als eenmaal de posities van QTLs of kwalitatieve genen op chromosomen bekend zijn, en er moleculaire merkers gevonden zijn die nauw gekoppeld zijn met deze genen, kunnen deze merkers vervolgens gebruikt worden om te selecteren op de aanwezigheid van gewenste

allelen in nakomelingschappen van kruisingen, of om gericht bepaalde genen te combineren in een nieuwe kruising. Omdat moleculaire merkers al in een heel vroeg stadium bepaald kunnen worden (er is slechts een geringe hoeveelheid DNA voor nodig, die al uit de eerste blaadjes gehaald kan worden), is het mogelijk de erfelijke eigenschappen al te voorspellen voordat deze in de plant tot uiting komen. Voor appel is dit een belangrijk voordeel, omdat het vaak lang duurt voor bepaalde eigenschappen, bijvoorbeeld vrucht-eigenschappen, waargenomen kunnen worden. Door met behulp van moleculaire merkers in een vroeg stadium te selecteren, hoeven alleen die planten opgekweekt te worden die met een grotere mate van waarschijnlijkheid gunstige eigenschappen in zich gecombineerd hebben. Daarnaast hebben bepalingen van moleculaire merkers het voordeel dat ze ook gedaan kunnen worden als de eigenschap zelf niet waarneembaar is. Zo kan er bijvoorbeeld op resistentie tegen ziekten geselecteerd worden in jaren dat de ziekte zelf niet optreedt, of kunnen verschillende resistentiegenen gecombineerd worden die elk afzonderlijk niet waargenomen kunnen worden in een ziektetoets.

Dit proefschrift beschrijft methoden en theoretische aspecten die van belang zijn bij kartering van genen in de nakomelingschap van twee ouders (een ‘full-sib’ familie) van een kruisbevruchtend gewas, en toepassingen in de genetische analyse van een kruising van twee appelrassen.

In hoofdstuk 1 worden het onderzoeksgebied en de probleemstelling toegelicht.

In hoofdstuk 2 worden theoretische aspecten bij een koppelingsstudie met moleculaire merkers in kruisbevruchtende gewassen onderzocht. De eigenschappen van verschillende typen moleculaire merkers en de consequenties voor het maken van genetische kaarten en koppelingsstudies worden beschreven. Alle combinaties van uitsplitsingen van merkers in een full-sib familie worden gepresenteerd met schatters voor de recombinatie-frequentie, informatiegehalte voor deze schatters, en formules voor de berekening van de LOD score. Daarnaast worden eigenschappen van dominante merkers bekeken.

Hoofdstuk 3 beschrijft het karteren van het *Vf* gen voor resistentie tegen appelschurft. Gegevens over schurft-aantasting werden verzameld op zes locaties in vijf Europese landen gedurende de periode 1993-1995, aan de nakomelingen van een kruising van het resistente appelras ‘Prima’ met een vatbaar ras, ‘Fiesta’. Naast de gegevens verkregen van waarnemingen op het veld door middel van natuurlijke infectie, werden ook waarnemingen gebruikt van kastoetsen met kunstmatige infectie. Een gedetailleerde kaart van het gen en de gekoppelde merkers wordt gepresenteerd. Enkele merkers zijn zeer nauw gekoppeld aan het gen en kunnen gebruikt worden voor merker-gestuurde selectie op het gen.

Hoofdstuk 4 beschrijft het karteren van een andere monogene resistentie ($Sd_1$), resistentie tegen de bloedvlekkenluis *Dysapha devecta* Wlk., afkomstig uit

Fiesta. Ook hier werden merkers gevonden die nauw gekoppeld zijn aan het gen en die gebruikt kunnen worden bij merkergestuurde selectie.

In hoofdstuk 5 wordt met bijna 300 zeer uiteenlopende typen moleculaire merkers, een genetische kaart van appel gemaakt aan de hand van de directe nakomelingschap (een full-sib familie) van de kruising Prima × Fiesta. Deze genetische kaart omvat 17 paar koppelingsgroepen, mogelijk overeenkomend met de 17 chromosoomparen die appel bezit.

Op deze kaart worden de posities aangegeven van vier bekende genen, waaronder naast de bovengenoemde resistentiegenen *Vf* en *Sd1* ook het gen voor zelf-incompatibiliteit *S* en het gen voor appelzuur, *Ma*. Daarnaast wordt de recombinatie in de twee ouders met elkaar vergeleken, en worden gebieden aangegeven waar merkers scheve uitsplitsing in de nakomelingschap lieten zien. Waargenomen duplicaties van merkers in het appel-genoom worden geïnterpreteerd vanuit de theorie dat in de evolutie de subfamilie *Maloideae*, waar appel en peer deel van uitmaken, vermoedelijk ontstaan is door hybridisatie van twee nauw verwante voorouders.

In hoofdstuk 6 wordt een analyse van QTLs voor vruchtstevigheid beschreven. Vruchten van de Prima × Fiesta nakomelingschap zijn, in verschillende jaren en afkomstig van verschillende locaties, door middel van mechanische en smaak-testen beoordeeld voor stevigheid en andere kwaliteits-eigenschappen. Twee QTLs voor vruchtstevigheid werden gekarteerd op de koppelingsgroepen L01 en L10, de eerste in de buurt van *Vf*. Voor andere kwaliteits-eigenschappen werd een significante associatie met merkers in de buurt van het *Ma* gen op koppelingsgroep L16 gevonden.

In hoofdstuk 7 worden verschillende methoden voor analyse van QTLs met elkaar vergeleken. De ene is een methode die gebaseerd is op een Bayesiaanse aanpak, de andere op de ‘klassieke’ frequentistische statistiek. Hiervoor zijn opnieuw de gegevens voor vruchtstevigheid uit de Prima × Fiesta kruising gebruikt. Waar het de belangrijkste resultaten betreft, voor de koppelingsgroepen met een QTL, en de posities van QTLs, bleken de beide methoden goed overeen te komen. Er waren echter ook koppelingsgroepen waarvoor de ene methode wel de aanwezigheid van een QTL suggereerde en de andere niet, en *vice versa*.

# References

Abbott JA (1994) Firmness measurement of freshly harvested 'Delicious' apples by sensory methods, sonic transmission, Magness-Taylor, and compression. J Amer Soc Hort Sci 19: 510-515.

Abbott JA, Affeldt HA, Liljedahl LA (1992) Firmness measurement of stored 'Delicious' apples by sensory methods, Magness-taylor, and sonic transmission. J Amer Soc Hort Sci 117: 590-594.

Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24: 235-278.

Alston FH, Briggs JB (1968) Resistance to *Sappaphis devecta* (Wik.) in apple. Euphytica 17: 468-472.

Alston FH, Briggs JB (1977) Resistance genes in apple and biotypes of *Dysaphis devecta*. Ann Appl Biol 87: 75-81.

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9: 208-218.

Arús P, Olarte C, Romero M, Vargas F (1994) Linkage analysis of ten isozyme genes in $F_1$ segregating almond progenies. J Amer Soc Hort Sci 119: 339-344.

Austin DF, Lee M (1996a) Comparative mapping in $F_{2:3}$ and $F_{6:7}$ generations of quantitative trait loci for grain yield and yield components in maize. Theor Appl Genet 92: 817-826.

Austin DF, Lee M (1996b) Genetic resolution and verification of quantitative trait loci for flowering and plant height with recombinant inbred lines of maize. Genome 39: 957-968.

Baronio P, Briolini G (1985) Afidi delle Pomacee. Ital Agric 122: 110-121.

Barreiro P, Ortiz C, RuizAltisent M, DeSmedt V, Schotte S, Andani Z, Wakeling I, Beyts PK (1998) Comparison between sensory and instrumental measurements for mealiness assessment in apples. A collaborative test. J Texture Studies 29: 509-525.

Batlle I, Alston FH, Evans KM (1995) The use of the isoenzymic marker gene *GOT-1* in the recognition of incompatibility *S* alleles in apple. Theor Appl Genet 90: 303-306.

Beavis WD (1994) The power and deceit of QTL experiments: Lessons from comparative QTL studies. Proc 49th Annu Corn Sorghum Indus Res Conf Am Seed Trade Assoc. Washington, DC. 250-266.

Beavis WD (1998) QTL analyses: power, precision and accuracy. p145-162 *in*: *Paterson AH (ed.) Molecular dissection of complex traits*. CRC press LLC. pp. 305.

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998a) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor Appl Genet 97: 381-397.

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998b) Advanced backcross QTL analysis in tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-

alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97: 170-180.

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998c) Advanced backcross QTL analysis in tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97: 1191-1196.

Blanpied GD, Bramlage WJ, Dewey DH, LaBelle RL, Massey Jnr, LM, Mattus GE, Stiles WC, Watada AE (1978) A standardized method for collecting apple pressure test data. N.Y. Food & Life Sci Bul 74. pp 8.

Bonierbale M, Plaisted RL, Pinedo O, Tanksley SD (1994) QTL analysis of trichome-rnediated insect resistance in potato. Theor Appl Genet 87: 973-987.

Bošković R, Tobutt KR (1996) Correlation of stylar ribonuclease zymograms with incompatibility alleles in sweet cherry. Euphytica 90: 245-250.

Bošković R, Tobutt KR (1999) Correlation of stylar ribonuclease isoenzymes with incompatibility alleles in apple. Euphytica 107: 29-43.

Bourne MC (1974) Comparison of results from the use of the Magness-Taylor pressure tip in hand- and machine-operation. J Texture Studies 5: 105-108.

Broothaerts W, Janssens GA, Proost P, Broekaert WF (1995) cDNA cloning and molecular analysis of two self-incompatibility alleles from apple. Plant Mol Bio 27: 499-511.

Brown SK (1992) Genetics of apple. Plant Breeding Reviews 9: 333-366.

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91: 869-875.

Chaparro JX, Werner DJ, O'Malley D, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological, isozyme, and RAPD markers in peach. Theor Appl Genet 87: 805-815.

Chen H, De Baerdemaeker J (1993) Effect of apple shape on acoustic measurements of firmness. J Agric Eng Res 56: 259-266.

Cheng FS, Weeden NF, Brown SK (1996) Identification of co-dominant RAPD markers tightly linked to fruit skin color in apple. Theor Appl Genet 93: 222-227.

Cheng FS, Weeden NF, Brown SK, Aldwinckle HS, Gardiner SE, Bus VG (1998) Development of a DNA marker for *Vm*, a gene conferring resistance to apple scab. Genome 41: 208-214.

Chevalier C, Lespinasse Y, Renaudin S (1991) A microscopic study of the different classes of symptoms coded by the *Vf* gene in apple for resistance to scab (*Venturia inaqualis*) Plant Pathology 40: 249-256.

Chevreau E, Laurens, F (1987) The pattern of inheritance in apple (*Malus* × *domestica* Borkh.): further results from leaf isozyme analysis. Theor Appl Genet 75: 90-95.

Chevreau E, Lespinasse Y, Gallet M (1985) Inheritance of pollen enzymes and polyploid origin of apple (*Malus* × *domestica* Borkh.). Theor Appl Genet 71: 268-277.

Chevreau E, Leuliette S, Gallet M (1997) Inheritance and linkage of isozyme loci in pear (*Pyrus communis* L.) Theor Appl Genet 94: 498-506.

Chevreau E, Manganaris AG, Gallet M (1999) Isozyme segregation in five apple progenies and potential use for map construction. Theor Appl Genet 98: 329-336.

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

Conneally PM, Edwards JH, Kidd KK, Lalouel J-M, Morton N, Ott J, White R (1985) Report of the committee on methods of linkage analysis and reporting. Cytogenet Cell Genet 40: 356-359.

Conner PJ, Brown, SK, Weeden NF (1997) Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. J Amer Soc Hort Sci 122: 350-359.

Conner PJ, Brown WK, Weeden NF (1998) Molecular-marker analysis of quantitative traits for growth and development in juvenile apple trees. Theor Appl Genet 96: 1027-1035.

Crosby JA, Janick J, Pecknold PC, Korban SS, O'Connor PA, Ries SM, Goffreda J, Voordeckers A (1992) Breeding apples for scab resistance: 1945-1990. Fruit Var J 46: 145-166.

Daillant-Spinnler B, MacFie HJH., Beyts PK, Hedderley D (1996) Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. Food Qual Pref 7: 113-126.

Darlington CD, Moffet AA (1930) Primary and secondary chromosome balance in *Pyrus*. Journal of Genetics 22: 129-151.

Darvasi A (1997) Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. Mammalian Genome 8: 163-167.

Darvasi A (1998) Experimental strategies for the genetic dissection of complex traits in animal models. Nature Genetics 18: 19-23.

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141: 1199-1207.

Dayton DF, Mowry JB, Hough LF, Bailey CH, Williams EB, Janick J, Emerson FH (1970) 'Prima', an early fall apple with resistance to scab. Fruit Var Hort Dig 24: 20-22.

Dempster AP, Laird, NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Statist Soc B 39: 1-38.

Devey ME, Bell JC, Smith DN, Neale DB, Moran GF (1996) A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. Theor Appl Genet 92: 673-679.

Dicker GHL (1954) The apple, pear, and quince aphids. Rep E Malling Res Stn for 1953, East Malling, pp 213-217.

Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the Rosaceae. Am J Bot 79: 1081-1086.

Dirlewanger E, Moing A, Rothan C, Svanella L, Pronier V, Gye A, Plomion C, Monet R (1999) Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch). Theor Appl Genet 98: 18-31.

Doerge RW, Zeng Z-B, Weir B (1997) Statistical issues in the search for genes affecting quantitative traits in experimental populations. Stat Science 12: 195-219.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15.

Dupuis J (1996) Statistical aspect of trait mapping using a dense set of markers: a partial review. In: Speed T, Waterman MS (eds.): *Genetic Mapping and DNA sequencing*. Springer-Verlag, New York. pp. 111-131.

Dupuis J, Siegmund D (1999) Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics 151: 373-386.

Durham RE, Korban SS (1994) Evidence of gene introgression in apple using RAPD markers. Euphytica 79: 109-114.

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Statist 7: 1-26.

Efron B (1982) *The jackknife, the bootstrap and other resampling plans*. Society for Industrial and Applied Mathematics, Philadelphia.

Frisch M, Bohn M, Melchinger AE (1999a) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39: 967-975.

Frisch M, Bohn M, Melchinger AE (1999b) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39: 1295-1301.

Garcia-Dorado A, Gallego A (1992) On the use of the classical tests for detecting linkage. The Journal of Heredity 83: 143-146.

Gardiner SE, Bassett HCM, Noiton DAM, Bus VG, Hofstee ME, White AG, Ball RD, Forster RLS, Rikkerink EHA (1996) A detailed linkage map around an apple scab resistance gene demonstrates that two disease resistance classes both carry the *Vf* gene. Theor Appl Genet 93: 485-493.

Genstat (1993) *Genstat 5 Reference Manual*. Oxford Scientific Publications, Clarendon Press Oxford.

Gessler C, Patocchi A, Kellerhals M, Gianfranceschi L (1997) Molecular markers applied to apple breeding and map-based cloning of resistance genes. IOBC/WPRS Bulletin 20: 105-109.

Gianfranceschi L, McDermott JM, Seglias N, Koller B, Kellerhals, M, Gessler C (1994) Towards a marker assisted breeding for resistance against apple scab. Euphytica 77: 93-96.

Gianfranceschi L, Koller B, Seglias N, Kellerhals M, Gessler C (1996) Molecular selection in apple for resistance to scab caused by *Venturia inaequalis*. Theor Appl Genet 93: 199-204.

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) simple sequence repeats for the genetic analysis of apple. Theor Appl Genet 96: 1069-1076.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121-1137.

Grattapaglia D, Bertolucci FL, Sederoff RR (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. Theor Appl Genet 90: 933-947.

Gratwick M (1992) *Crop pests in the UK*. Collected edition of MAFF leaflets, 1st edn. Chapman and Hall, London.

Green PJ (1995) Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82: 711-732.

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in *Malus* × *domestica* (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94: 249-254.

Haley CS (1995) Livestock QTLs - bringing home the bacon ? Trends in Genetics 11: 463-524.

Haley CS, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315-324.

Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136: 1195-1207.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57: 97-109.

Heath SC (1997) Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61: 748-760.

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. Journal of Heredity 85: 4-11.

Hemmat M, Weeden NF, Brown SK (1995) Molecular markers for the scab resistance (*Vf* ) region in apple. HortSci 30: 855.

Hemmat M, Weeden NF, Conner PJ, Brown SK (1997) A DNA marker for columnar growth habit in apple contains a simple sequence repeat. J Amer Soc Hort Sci 122: 347-349.

Hill WG (1998) Selection with recurrent backcrossing to develop congenic lines for quantitative trait loci analysis. Genetics 148: 1341-1352.

Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM (1997) Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics 147: 1445-1457.

Hoeschele I, VanRaden PM (1993a) Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. Theor Appl Genet 85: 953-960.

Hoeschele I, VanRaden PM (1993b) Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. Theor Appl Genet 85: 946-952.

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* x *domestica* Borkh. core subset collection. Theor Appl Genet 97: 671-683.

Hospital F, Charcosset A (1997) Marker-assisted introgression of quantitative trait loci. Genetics 147: 1469-1485.

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132: 1199-1210.

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95: 1181-1189.

Hyne V (1995) The Apple-Store database. European Apple 3: 2-3. August 1995, HRI, Wellesbourne.

Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. Theor Appl Genet 85: 252-260.

Jansen RC (1993) Interval mapping of quantitative trait loci. Genetics 135: 205-211.

Jansen RC (1994a) Controlling the type I and type II errors in mapping quantitative trait loci. Genetics 138: 871-881.

Jansen RC (1994b) Mapping of quantitative trait loci by using genetic markers: an overview of biometrical models used. In: Van Ooijen JW, Jansen J (Eds.). *Biometrics in plant breeding: applications of molecular markers. Proceedings of the ninth meeting of the EUCARPIA section biometrics in plant breeding, 6–8 July 1994, Wageningen, the Netherlands.* pp. 116-124.

Jansen RC (1995) *Genetic mapping of quantitative trait loci in plants - a novel statistical approach*. PhD Thesis. Wageningen, the Netherlands. ISBN 90-73771-14-5.

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics. 136: 1447-1455.

Jansen RC, Van Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype by environment interaction in genetic mapping of multiple quantitative trait loci. Theor Appl Genet 91: 33-37.

Jansen RC (1996) A general Monte Carlo method for mapping multiple quantitative trait loci. Genetics 142: 305-311.

Janss LL, Thompson GR, Van Arendonk JAM (1995) Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. Theor Appl Genet 91: 1137-1147.

Janssens GA, Goderis IJ, Broekaert WF, Broothaerts W (1995) A molecular method for *S*-allele identification in apple based on allele-specific PCR. Theor Appl Genet 91: 691-698.

Jensen CS, Sheehan N (1998) Problems with determination of noncommunicating classes for Monte Carlo Markov Chain applications in pedigree analysis. Biometrics 54: 416-425.

Jenser G, Balázs K (1991) The possibilities and difficulties in introducing integrated pest management in apple orchards. Növényvédelem 27: 97-102.

Joobeur T, Viruel MA, de Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Battle I, Quarta R, Dirlewanger E, Arús P (1998) Construction of a saturated linkage map for *Prunus* using an almond x peach $F_2$ progeny. Theor Appl Genet 97: 1034-1041.

Kao C-H, Zeng Z-B (1997) General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. Biometrics 53: 653-665.

Kao C-H, Zeng Z-B, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152: 1203-1216.

Kearsey MJ, Farquhar GL (1998) Short review: QTL analysis in plants; where are we now? Heredity 80: 137-142.

Kearsey MJ, Pooni HS (1996) *The genetical analysis of quantitative traits*. Chapman & Hall, London. pp 381.

King GJ, Alston FH, Batlle I, Chevreau E, Gessler C, Janse J, Lindhout P, Manganaris AG, Sansavini S, Schmidt H, Tobutt KR (1991) 'The European Apple Genome Mapping Project' — developing a strategy for mapping genes coding for agronomic characters in tree species. Euphytica 56: 89-94.

King GJ (1994) Progress in mapping agronomic genes in apple. Euphytica 77: 65-69.

King GJ (1996) Progress of apple genetic mapping in Europe. HortSci 31: 1108-1111.

King GJ, Alston FH, Brown LM, Chevreau E, Evans KM, Dunemann F, Janse J, Laurens F, Lynn JR, Maliepaard C, Manganaris AG, Roche P, Schmidt H,

Tartarini S, Verhaegh J, Vrielink R. (1998) Multiple field and glasshouse assessments increase the reliability of linkage mapping of markers flanking the *Vf* source of scab resistance in apple. Theor Appl Genet 96: 699-708.

King GJ, Maliepaard C, Lynn JR, Alston FH, Durel CE, Evans KM, Griffon B, Laurens F, Manganaris AG, Schrevens E, Tartarini S (2000) Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.) Theor Appl Genet (in press)

King GJ, Tartarini S, Brown LM, Gennari F, Sansavini S (1999) Introgression of the *Vf* source of scab resistance and distribution of linked marker alleles within the *Malus* gene pool. Theor Appl Genet 99: 1039-1046.

Kirkham DS (1957) Relationships between cultural characters and pathogenicity in *Venturia inaequalis* and *Venturia pirina*. J Gen Mic 16: 360-373.

Knapp SJ (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci 38: 1164-1174.

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132: 1211-1222.

Koller B, Gianfranceschi L, Seglias N, McDermott J, Gessler C (1994) DNA markers linked to the *Malus floribunda* 821 scab resistance. Plant Mol Biol 26: 597-602.

Korol AB, Ronin YI, Kirzhner VM (1995) Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics 140: 1137-1147.

Korol AB, Ronin YI, Nevo E (1998a) Approximate analysis of QTL-environment interaction with no limits on the number of environments. Genetics 148: 2015-2028.

Korol AB, Ronin YI, Nevo E, Hayes PM (1998b) Multi-interval mapping of correlated trait complexes. Heredity 80: 273-284.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756.

Lanaud C, Risterucci AM, N'Goran AKJ, Clement D, Flament MH, Laurent V, Falque M (1995) A genetic linkage map of *Theobroma cacao* L. Theor Appl Genet 91: 987-993.

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Lander ES, Green P (1987) Construction of multilocus genetic maps in humans. Proc Natl Acad Sci USA 84: 2363-2367.

Lee M (1995) *DNA markers and plant breeding programs*. In: Advances in agronomy 55: 265-344.

Lehmann EL (1975) *Nonparametrics: statistical methods based on ranks*. McGraw-Hill, New York.

Lehman-Salada (1996) Instrument and operator effects on apple firmness readings. HortSci 31: 994-997

Lespinasse Y (1973) Chromosomic observations in *Malus* and *Pyrus*. In: *Proc Eucarpia Fruit Sect Symp V. Canterbury*, pp. 75-83.

Luby and Shaw (2000) Does marker-assisted selection make dollars and sense in a fruit breeding program? HortScience: submitted.

Lurie S & Nussinovitch A (1996) Compression characteristics, firmness, and texture perception of heat treated and unheated apples. Intl J Food Sci Tech 31: 1-5.

Lynn JR (1998) Variability in a multi-site apple trial for genetic analysis. Acta Horticulturae 484: 281-284

Magness JR, Taylor GF (1925) *An improved type of pressure tester for the determination of fruit maturity*. USDA Dept Circ No 350. Washington, D.C.: US Dept of Agriculture.

Maliepaard C, Van Ooijen JW (1994) QTL mapping in a full-sib family of an outcrossing species. In: Van Ooijen JW, Jansen J (eds) *Biometrics in plant breeding: applications of molecular markers*, *Proc.of the ninth meeting of the EUCARPIA section biometrics in plant breeding, 6-8 July 1994, Wageningen, The Netherlands*, pp140-146.

Maliepaard C, Bas N, Van Heusden S, Kos J, Pet G, Verkerk R, Vrielink R, Zabel P, Lindhout P (1995) Mapping of QTLs for glandular trichome densities and *Trialeurodes vaporariorum* (greenhouse whitefly) resistance in an $F_2$ from *Lycopersicon esculentum* x *Lycopersicon hirsutum* f. *glabratum*. Heredity 75: 425-433.

Maliepaard C, Jansen J, Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Gen Research 70: 237-250.

Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97: 60-73.

Manganaris AG, Alston FH (1987) Inheritance and linkage relationships of glutamate oxaloacetate transaminase isoenzymes in apple. 1. The gene *GOT-1*, a marker for the *S* incompatibility locus. Theor Appl Genet 74: 154-161.

Manganaris AG, Alston FH (1988) The Acid Phosphatase gene ACP-1 and its linkage with the Endopeptidase gene *ENP-1* and the pale green lethal gene *l* in apple. Acta Horticulturae 224: 177-184.

Manganaris AG, Alston FH (1992) Genetics of leucine aminopeptidase in apple. Theor Appl Genet 83: 345-352.

Manganaris AG, Alston FH, Weeden NF, Aldwinckle HS, Gustafson HL, Brown SK (1994) Isozyme locus *Pgm-1* is tightly linked to a gene (*Vf*) for scab resistance in apple. J Amer Soc Hort Sci 119: 1286-1288.

Manly KF, Olson JM (1999) Overview of QTL mapping software and introduction to Map Manager QTL. Mammalian Genome 10: 327-334.

Markussen T, Krüger J, Schmidt H, Dunemann F (1995) Identification of PCR-based markers linked to the powdery-mildew-resistance gene $Pl_1$ from *Malus robusta* in cultivated apple. Plant Breed 114: 530-534.

Martínez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85: 480-488.

Massee AM (1954) *The pests of fruits and hops*, 3rd edn. Crosby Lockwood, London.

Mather K (1951) *The measurement of linkage in heredity*. London: Methuen & Co.

Matula DW (1977) Graph theoretic techniques for cluster analysis algorithms. In : Van Ryzin J ed. *Classification and Clustering: proceedings of an advanced seminar conducted by the Mathematics Research Center, the University of Wisconsin at Madison, May 3-5, 1976*, Academic Press, New York NY, USA, pp 95-129.

McCullagh P, Nelder JA (1989) *Generalized linear models*. In *Monographs on tatistics and Applied Probability 37*. London: Chapman & Hall.

Mehlenbacher SA (1995) Classical and molecular approaches to breeding fruit and nut crops for disease resistance. HortScience 30 (3): 466-477.

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149: 383-403.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21: 1087-1092.

Michelmore RW, Paran I, Kesseli TV (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.

Mohan M, Nair S, Bentur JS, Rao UP, Bennett J (1994) RFLP and RAPD mapping of the rice Gm2 gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theor Appl Genet 87: 782-788.

Morgan DR, Soltis DE, Robertson KR (1994) Systematic and evolutionary implications of *rbc*L sequence variation in Rosaceae. Am J Botany 81: 890-903.

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7: 277-318.

Nienhuis J, Helentjaris T, Slocum M, Ruggero B, Schaefer A (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tornato. Crop Sci 27: 797-803.

Nieto-Lopez RM, Blake TK (1994) Russian wheat aphid resistance in barley: inheritance and linked molecular markers. Crop Sci 34: 655-659.

Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean 2. Development of an RFLP-based linkage map. Theor Appl Genet 85: 513-520.

Parisi L, Lespinasse Y, Guillaumes J, Krüger J (1993) A new race of *Venturia inaequalis* virulent to apples with resistance due to the *Vf* gene. Phytophathology 83: 533-537.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335: 721-726.

Paterson AH, Damon S, Hewitt JD, Samir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato - comparison across species, generations, and environments. Genetics 127: 181-187.

Patocchi A, Gianfranceschi L, Gessler C (1999) Towards the map-based cloning of *Vf:* fine and physical mapping of the *Vf* region. Theor Appl Genet 99: 1012-1017.

Patterson HD, Thompson R. (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58: 545-554.

Payne RW, Lane PW, Digby PGN, Harding SA, Leech PK, Morgan GW, Todd AD, Thompson R, Tunnicliffe Wilson G, Welham SJ & White RP. (1993) *Genstat 5 Release 3 Reference Manual*, Oxford University Press, Oxford.

Penner GA, Stebbing J-A, Legge B (1995) Conversion of an RFLP marker for the barley stem rust resistance gene *Rpg1* to a specific PCR-amplifiable polymorphism. Mol Breed 1: 349-354.

Pfeifer B (1994) Erste Eindrücke von Schorf-Mehltauresistenten Apfelsorten an der SLVA in Ahrweiler. Obstbau 19: 236-240.

Phipps JB, Robertson KR, Rohrer JR (1991) Origins and Evolution of Subfam. Maloideae (Rosaceae). Systematic Botany 16: 303-332.

Plomion C, O'Malley DM (1996) Recombination rate differences for pollen parents and seed parents in *Pinus pinaster*. Heredity 77: 341-350.

Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. Theor Appl Genet 90: 503-510.

Rezwani A, Radjabi G (1987) The species of Aphididae family injuring rosaceous fruit trees in Iran. Entomol Phytopathol Appl 54: 67-68, 165-178.

Risch N (1992) Genetic linkage: interpreting LOD scores. Science 255: 803-804

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125: 645-654.

Ritter E, Salamini F (1996) The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. Gen Research 67: 55-65.

Roberts AL, Crute IR (1994) apple scab resistance from *Malus floribunda* 821 (*Vf*) is rendered ineffective by isolates of *Venturia inaequalis* from *Malus floribunda*. Norwegian J Agric Sci Suppl 17: 403-406.

Roche P, Alston FH, Maliepaard C, Evans KM, Vrielink R, Dunemann F, Markussen T, Tartarini S, Brown LM, Ryder C, King GJ (1997a) RFLP and RAPD markers linked to the rosy leaf curling aphid resistance gene ($Sd_1$ ) in apple. Theor Appl Genet 94: 528-533.

Roche P, Van Arkel G, Van Heusden AW (1997b) A specific PCR assay for resistance to biotypes 1 and 2 of the rosy leaf curling aphid in apple based on an RFLP marker closely linked to the $Sd_1$ gene. Plant Breeding 116: 567-572.

Satagopan JM, Yandell BS (1996) Estimating the number of quantitative trait loci via Bayesian model determination. Special Contributed Paper Session on Genetic Analysis of Quantitative Traits and Complex Diseases, Biometric Section, Joint Statistical Meetings, Chicago, IL., USA (available at ftp://ftp.stat.wisc.edu/pub/yandell/revjump.html).

Satagopan JM, Yandell BS, Newton MA, Osborn C (1996) A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. Genetics 144: 805-816.

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolis vulgaris*. Genetics 8: 552-560.

Scheler P, Mangin B, Goffinet B, Le Roy P, Boichard D, Elsen JM (1998) Properties of a Bayesian approach to detect QTL compared to the flanking markers regression method. J Anim Breed Genet 115: 87-95.

Schön CC, Lee M, Melchinger AE, Guthrie WD, Woodman WL (1993) Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. Heredity 70: 648-659.

Schwennesen J, Mielke EA, Wolfe WH (1982) Identification of seedless table grape cultivars and a bud sport with berry isozymes. HortSci 17: 366-368.

Sheehan N, Thomas A (1993) On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme. Biometrics 49: 163-175.

Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics. A guide for the uninitiated. Trends in Genetics 15 (9): 354-358.

Sillanpää MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148: 1373-1388.

Sillanpää MJ, Arjas E (1999) Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. Genetics 151: 1605-1619.

Soller M, Brody T, Genizi A (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47: 35-39.

Song JZ, Soller M, Genizi A (1998) The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. Genet Res 73: 61-73.

Spelman R, Bovenhuis H (1998) Genetic response from marker assisted selection in an outbred population for differing maker bracket sizes and with two identified quantitative trait loci. Genetics 148: 1389-1396.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant J 3: 739-744.

Stam P, Van Ooijen JW (1995) JoinMap$^{TM}$ version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen.

Stephens DA, Fisch RD (1998) Bayesian analysis of quantitative trait locus data using Reversible Jump Markov Chain Monte Carlo. Biometrics 54: 1334-1347.

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27: 205-233.

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92: 191-203.

Tartarini S (1996) RAPD markers linked to the *Vf* gene for scab resistance in apple. Theor Appl Genet 92: 803-810.

Tartarini S, Gianfranceschi L, Sansavini S, Gessler C (1999) Development of reliable PCR markers for the selection of the *Vf* gene conferring scab resistance in apple. Plant Breeding 118: 183-186.

Thaller G, Hoeschele I (1996) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. Methodology. Theor Appl Genet 93: 1161-1166.

Thoday JM (1961) Location of polygenes. Nature 191: 368-370.

Tinker NA, Mather DE (1995) Methods for QTL analysis with progeny replicated in multiple environments. J QTL 15: 1. http://probe.nalusda.gov:8000/otherdocs/jqtl/index.html

Thompson E (1994) Monte Carlo likelihood in genetic mapping. Statist Sci 9: 355-366.

Uimari P, Thaller G, Hoeschele I (1996a) The use of multiple markers in a Bayesian method for mapping quantitative trait loci. Genetics 143: 1831-1842.

Uimari P, Zhang Q, Grignola F, Hoeschele I, Thaller G (1996b) Analysis of QTL workshop I Granddaughter design data using least-squares, residual maximum likelihood and Bayesian methods. J QTL 1996: 2, art. 7. http://probe.nalusda.gov:8000/otherdocs/jqtl.

Uimari P, Hoeschele I (1997) Mapping linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. Genetics 146: 735-743.

Van Berloo R, Stam P (1998) Marker-assisted selection in autogamous RIL populations: a simulation study. Theor Appl Genet 96: 147-154.

Van der Beek JG, Verkerk R, Zabel P, Lindhout P (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf*$_9$ (resistance to *Cladosporium fulvum*) on chromosome 1. Theor Appl Genet 84: 803-811.

Van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84: 803-811.

Van Ooijen JW (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. Heredity 83: 613-624.

Van Ooijen JW, Maliepaard C (1996) MapQTL$^{TM}$ version 3.0: Software for the calculation of QTL positions on genetic maps. CPRO-DLO, Wageningen.

Van Ooijen JW, Sandbrink JM, Vrielink M, Verkerk R, Zabel P, Lindhout P (1994) An RFLP linkage map of *Lycopersicon peruvianum*. Theor Appl Genet 89: 1007-1013.

Vieland V (1998) Bayesian linkage analysis, or: How I learned to stop worrying and love the posterior probability of linkage. Am J Hum Genet 63: 947-954.

Vinatzer BA, Zhang H-B, Sansavini S (1998) Construction and characterization of a bacterial artifical chromosome library of apple. Theor Appl Genet 97: 1183-1190.

Viruel MA, Messeguer R, de Vicente MC, Garcia-Mas J, Puigdomènech P, Vargas F, Arús P (1995) A linkage map with RFLP and isozyme markers for almond. Theor Appl Genet 91: 964-971.

Visscher PM, Haley CS, Thompson R (1996a) Marker-assisted introgression in backcross breeding programs. Genetics 144: 1923-1932.

Visscher PM, Thompson R, Haley CS (1996b) Confidence intervals in QTL mapping by bootstrapping. Genetics 143: 1013-1020.

Visser T, Verhaegh JJ (1978) Inheritance and selection of some fruit characters of apple. I. Inheritance of low and high acidity. Euphytica 27: 753-760.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23 (21): 4407-4414.

Vuylsteke M, Kuiper M, Stam P (1999) Chromosomal regions involved in hybrid performance and heterosis: their AFLP-based identification and practical use in prediction models. In: *Genetic analysis of maize by using the AFLP method*. PhD thesis Wageningen. M. Vuylsteke. pp. 91-115.

Walling GA, Visscher PM, Haley CS (1998) A comparison of bootstrap methods to construct confidence intervals in QTL mapping. Genet Res 71: 171-180.

Watada AE, Abbott JA (1985) Apple quality: influences of pre- and postharvest factors and estimation by objective methods. In: *Evaluation of quality of fruits and vegetables*, Ed. H.E. Pattee. AVI Publishing Co, Wesport, Connecticut. pp. 63-81.

Weber WE, Wricke G (1994) *Genetic markers in plant breeding. Advances in plant breeding 16*. Berlin: Parey Scientific Publishers.

Weeden NF, Lamb RC (1987) Genetics and linkage analysis of 19 isozyme loci in apple. J Amer Soc Hort Sci 11 (2): 865-872.

Weeden NF, Hemmat M, Lawson DM, Lodhi M, Bell RL, Manganaris AG, Reisch BI, Brown SK, Ye G-N (1994) Development and application of molecular marker linkage maps in woody fruit crops. Euphytica 77: 71-75.

Weller I (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42: 627-640.

Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes. In: Soltis DE and Soltis PS (eds.) *Isozymes in Plant Biology*, Dioscorides Press, pp 5-45.

Williams EB, Dayton DF, Shay JR (1966) Allelic genes in *Malus* for resistance to *Venturia inaequalis*. Proc Amer Soc Hort Sci 88: 52-56.

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

Xu X-M, Butt DJ, Santen G (1995) A dynamic model simulating infection of apple leaves by *Venturia inaequalis*. Plant Path 44: 865-876.

Yang H, Korban SS (1996) Screening apples for OPD20/600 using sequence-specific primers. Theor Appl Genet 92: 263-266.

Yang H, Krüger J (1994) Identification of a RAPD marker linked to the *Vf* gene for scab resistance in apples. Plant Breed 112: 323-329.

Yang HY, Korban SS, Krüger J, Schmidt H (1997a) A Randomly Amplified Polymorphic DNA (RAPD) marker tightly linked to the scab-resistance gene *Vf* in apple. J Amer Soc Hort Sci 122: 47-52.

Yang HY, Korban SS, Krüger J, Schmidt H (1997b) The use of a modified bulk segregant analysis to identify a molecular marker linked to a scab resistance gene in apple. Euphytica 94: 175-182.

Young ND, Kumar L, Menancio-Hautea D, Danesh D, Talekar NS, Shanmugasundarum S, Kim D-H (1992) RFLP rnapping of a major bruchid resistance gene in mungbean (*Vigna radiata*, L. Wilczek). Theor Appl Genet 84: 839-844.

Zeng Z-B (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90: 10972-10976.

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

# Dankwoord

Met dit proefschrift sluit ik een periode af waarin ik met veel plezier met verschillende mensen heb samengewerkt. Een groot deel van het beschreven onderzoek is uitgevoerd binnen een project van de EU, 'The Development of the European Apple Crop'. I would like to thank my colleagues in Belgium, France, Germany, Greece, Italy, the United Kingdom and Switzerland for a pleasant collaboration and fruitful discussions about the genetics and breeding of apple. I would like to thank Dr. Mikko Sillanpää for our collaboration on QTL mapping methodologies.

Het werk beschreven in dit proefschrift ben ik begonnen op de toenmalige afdeling Groente- en Fruitgewassen. Die periode heb ik als erg plezierig ervaren, met veel humor, relativeringsvermogen en een samenbindende afkeer van te veel bureaucratie. Voor die stimulerende werksfeer wil ik de collega's uit die periode heel hartelijk bedanken. In het bijzonder wil ik hier graag mijn waardering uiten voor Sjaak, Ria, Joke, Jac, en Gert: een beter team had ik mij niet kunnen wensen. Jullie waren altijd bereid om met me mee te denken en in te springen daar waar het nodig was.

Daarnaast wil ik graag Willem, Remmelt, Karin, Eric en Bert bedanken voor de prettige samenwerking. Mijn collega's in de kwekerij en in de boomgaard bedank ik voor de goede zorg over het plantmateriaal, met name Teus en Cees. Ton, bedankt voor je toegankelijkheid, je humor, en de manier waarop je, naar mijn idee, een zekere mate van anarchie binnen je afdeling wel wist te waarderen.

Ook in mijn huidige 'eenheid', het Centrum voor Biometrie, bevind ik me in een plezierige werkomgeving, en daar wil ik mijn collega's graag voor bedanken. Ik kan altijd met mijn vragen bij jullie terecht en voel me thuis in de sfeer waarin we zonder veel afstand met elkaar omgaan. In het bijzonder wil ik Hans, Ritsert en Johan noemen, met wie ik inmiddels al een aardig tijdje optrek: veel van wat ik weet, heb ik van jullie geleerd, en jullie zijn zeer zeker mijn 'naaste collega's'. Ik prijs me gelukkig met jullie betrokkenheid bij mijn werk en persoon. Wies, jij die als kamergenoot in de letterlijke zin mijn 'naaste collega' bent: ik hoop dat het jou even goed bevalt als mij.

Johan, jou wil ik graag in het bijzonder bedanken: jij bent als mijn eerdere kamergenoot, collega en co-promotor echt bij al mijn werk betrokken geweest, en zonder jouw enorme inzet, je kritische en intensieve begeleiding, je precisie, en je stimulans zou dit boekje er misschien nooit gekomen zijn. Hulde!

Piet, ik ben erg blij om juist bij jou te kunnen promoveren. Mede via jou ben ik 'QTLs gaan mappen', eerst in tomaat en daarna in appel. Ook toen ik nog in Ton z'n toko werkte, voelde ik me al even goed thuis bij jou en 'jouw' populatiebiologen. Als promotor heb je me niet zo vaak op je kamer gezien. Dat lag niet aan jou maar

vooral aan mijzelf en aan de uitstekende begeleiding van Johan. Met je vragen heb je me regelmatig aan het denken gezet. Ik hoop dat je dat zult blijven doen.

Mijn paranimfen Marloes en Arjan wil ik heel graag bedanken voor al het werk dat ze me uit handen genomen hebben en het enthousiasme waarmee ze me door de promotie stuwen. Met organisatoren van jullie kaliber hoef ik me eigenlijk geen moment zorgen te maken over het verloop van de promotiedag.

Op deze plek wil ik ook graag mijn beide ouders bedanken: aan hen draag ik dit proefschrift op. Zij hebben mij een echt thuis gegeven, een liefdevolle en evenwichtige omgeving om onbezorgd in op te groeien en de mogelijkheid om in Wageningen te gaan studeren.

Lieve Joost, jou wil ik heel graag twee keer bedanken: voor de prachtige kaft die je voor me gemaakt hebt, maar meer nog voor jouw niet aflatende werk achter de schermen, je geduld, je zorg, je goede raad: je bent er altijd voor me.

*'Blessed be the time*
*that appil take was,*
*therefore we moun singen*
*Deo gracias'*

*(Adam lay ibounden,*
*Anoniem, 15$^{e}$ eeuw)*

# Curriculum vitae

Christoffel Aren (Chris) Maliepaard was born in Stad aan ‘t Haringvliet on July 24, 1965. In 1983 he finished secondary education at the Athenaeum of the ‘Rijksscholengemeenschap Goeree-Overflakkee’ in Middelharnis and started his studies of plant breeding at Wageningen University. He finished his studies in 1989, with main subjects in the areas of plant breeding, crop protection, and applied statistics. After his studies, he worked as a teacher at the Department of Mathematics and as a quantitative geneticist at the Department of Genetics of Wageningen University. In 1993 he started work at the Centre for Plant Breeding and Reproduction Research (CPRO), now a part of Plant Research International. His current position is quantitative geneticist at the Centre for Biometry of Plant Research International. His main expertise is in the fields of genetic mapping, QTL analysis and applied statistics in plant breeding. Most of the research described in this dissertation was carried out from 1993 to 1997 in the context of a project of the European Union, ‘Development of the European Apple Crop’.