Genetic analysis of maize by using the AFLP® method

Promotoren: dr. ir. P. Stam
hoogleraar in de plantenveredeling, i.h.b. selectiemethoden
en duurzame resistentie

dr. M. Zabeau
hoogleraar in de industriële biotechnologie
Vrije Universiteit Brussel, België

Co-promotor: dr. M.T.R. Kuiper
Plant Genetic Systems N.V., Gent, België

Marnik J.R. Vuylsteke

Genetic analysis of maize by using the AFLP® method

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van de Landbouwuniversiteit Wageningen,
Dr. C.M. Karssen,
in het openbaar te verdedigen
op woensdag 19 mei 1999
des namiddags te vier uur in de Aula

CIP-DATA KONINKLIJKE BIBILIOTHEEK, DEN HAAG

Vuylsteke, M.J.R.

Genetic analysis of maize by using the AFLP® method

Marnik J.R. Vuylsteke

Thesis Wageningen – With summary in Dutch

ISBN 90-5808-044-7

Subject headings: DNA-methylation, AFLP®, genetic mapping, genetic diversity, heterosis, maize

Voorwoord

Dit proefschrift zet een punt achter een leerrijke, drukke en unieke periode in mijn leven waarop ik met veel plezier terugkijk. Het leerrijke aspect omhelst het zelfstandig uitoefenen van wetenschappelijk onderzoek en het afsluiten hiervan met vier potentiële publicaties. Het drukke en tevens het unieke aspect van deze periode bestond uit het kombineren van de eerder vermelde inspanning met het vaderschap. Een betere cursus 'time-management' voor gevorderden kan ik niet bedenken.

Sommige kansen krijg je maar één keer in je leven; in mijn geval is de kans om te promoveren daar een mooi voorbeeld van. In 1993 bood Marc Zabeau mij een baan aan op Keygene N.V. en tegelijkertijd ook de mogelijkheid om te promoveren. Beste Marc, nogmaals dank voor deze toch wel unieke kans die je mij gegeven hebt. Het doet me trouwens ook veel plezier dat jij zelfs na je vertrek op Keygene, nu als promotor kan optreden.

Waar het allemaal begon om de rol van DNA-methylatie variatie in somaklonale variatie, werd een jaar na aanvang het promotie-onderzoek over een totaal andere boeg gegooid; 'heterosis' moest het worden. Met dit nieuwe onderwerp rolde ik van louter moleculaire biologie in de kwantitatieve genetica. Al vlug daarna werd Piet Stam gevraagd om ook als promotor op te treden, een taak die hij steeds met een grote inzet wist uit te voeren. Beste Piet, jou deskundigheid is essentieel geweest in het tot stand brengen van dit wetenschappelijk werk.

Een betere co-promotor dan Martin Kuiper kon ik me niet wensen. Onder zijn directe begeleiding heb ik mij in de 'high-density mapping'- en de heterosis-problematiek gestort. Ook na zijn vertrek op Keygene en in alle drukte van zijn nieuwe funktie bleef hij steeds bereikbaar voor allerlei discussies tijdens het opschrijven van de resultaten tot een publiceerbaar geheel. Beste Martin, hiervoor wil ik je nadrukkelijk bedanken.

Een woord van dank gaat zeker ook naar de Keygene-direktie, met in het bijzonder Pieter Vos. Vooral in het afronden en het publiceerbaar maken van het proefschrift heeft Pieter een niet onbelangrijke rol gespeeld.

Dank aan Cebeco Zaden, De Ruiter, ENZA Zaden, Rijk Zwaan en Van der Have, niet alleen omdat ze dit werk hebben meegefinancierd, maar ook omdat ze steeds een grote interesse aan de dag hebben gelegd voor dit werk.

Alle collega's van Keygene hebben bijgedragen tot een goede werkomgeving: de gezelligheid op het lab, de praatjes met het ondersteunend personeel, en de vele gezellige praatjes tijdens de koffie- en lunchpauzes hebben het werk er een stuk aangenamer op gemaakt.

Een woord van dank aan mijn ouders en schoonouders. Ondanks het feit dat het hun niet altijd even duidelijk was waarmee ik precies bezig was, hun betrokkenheid en interesse in de voortgang van mijn activiteiten was steeds merkbaar aanwezig. Vooral de bijdrage van mijn ouders kan niet sterk genoeg worden benadrukt: ze hebben mij destijds de vrijheid gegeven om te studeren. Dit proefschrift is dan ook een kroon op hun jarenlange inspanningen die ze hiervoor hebben geleverd.

Als laatst een woord aan mijn echtgenote Nicole, die er voor heeft gezorgd dat dit promotiewerk te combineren viel met een gezinsleven. Beste Nicole, alhoewel jouw rol zich hoofdzakelijk achter de schermen afspeelde, je liefde en steun waren onontbeerlijk.

Stellingen

1. Het effect van de variabiliteit in DNA methylatie op de segregatie van *PstI/MseI* AFLP® merkers in mais is verwaarloosbaar klein.

2. Een associatie-studie met een groot aantal gekarteerde merkers op maishybriden leidt tot identificatie van QTL.

3. 'Terug naar de gegevens op gel' is essentieel in het construeren van een genetische kaart.

4. Maar al te vaak wordt gemakshalve en op verkeerde wijze gekozen voor meer merkers ten koste van het aantal segregerende individuen, wat de resolutie van de genetische kaart en van QTL in meerdere loci niet ten goede komt.

5. Het gebruik van DNA-merkers in de plantenveredeling heeft zich tot nu toe hoofdzakelijk beperkt tot het in kaart brengen van het resultaat van decennia lang veredelen.

6. Een rechtstreeks gevolg van de Chinese 'één kind'-politiek is dat men straks in China meer zal aangewezen zijn op associatie-studies dan op koppelings-studies in het identificeren van genetische faktoren voor erfelijke ziektes.

7. Een allel met een groot effect wordt ten onrechte steeds gelijk gesteld aan een mutatie met een groot effect; de interactie tussen meerdere mutaties in een locus kan in eenzelfde effect resulteren. *(Stam, L.F. and Laurie C.C. 1996 Genetics 144, 1559-1564)*

8. De wereldbeschouwelijke gevolgen van de ontdekking van buitenaards leven zullen die van het copernicaanse wereldbeeld in de 16de eeuw en de darwiniaanse revolutie in de 19de eeuw overtreffen.

9. Vooral de advocatuur wordt rijk van de biotechnologie.

10. Biologisch gezien is vreemd gaan niet zo vreemd.

11. Alles kan beter.

Stellingen behorende bij het proefschrift "Genetic analysis of maize by using the AFLP® method" door Marnik Vuylsteke, in het openbaar te verdedigen op woensdag 19 mei 1999, te Wageningen.

Contents

1

Introduction

Maize (*Zea mays* L.), a worldwide grown and economically important crop, has a long history of intensive genetic and cytogenetic studies. Only part of the reasons that make maize an attractive organism for genetic and cytogenetic studies are a small chromosome number ($2n = 2x = 20$), easily observed hereditary characteristics, many recessive characters that are exposed through inbreeding or by mutagenic agents, the availability of a system for the efficient translocation of recessive mutants in the arm of a specific chromosome, known as the B-A translocation system, and the feasibility to perform either cross- or self-pollinations.

Especially the successful development in maize of the inbred-hybrid concept has stimulated genetic investigation in maize and made maize one of the earliest and most extensively studied organisms with respect to heterosis. The term heterosis, synonymous with hybrid vigor, was coined by Shull in 1908 (see Hayes 1992) and is usually described in terms of the increase in size or rate of growth of offspring over parents. However, although heterosis has been extensively examined and utilized on a large scale for production of selected hybrids of plants and animals for the past 75 years, the underlying genetic basis for the phenomenon has still not been satisfactorily explained. Since the early attempts to explain hybrid vigor in Mendelian terms, two principal hypotheses have been suggested as the genetic basis of heterosis: the dominance and the overdominance hypothesis. Because of the observed correlation between recessiveness and detrimental effects, the dominance hypothesis attributes the increased vigor of heterozygosity to dominant alleles (i.e. single loci at which alleles have dominant advantageous effects). The hypothesis still holds when dominance is not complete. The overdominance theory assumes the existence of single loci at which two alleles have the property

that the heterozygote is truly superior to either homozygote. These two theories hold important differences for practical breeders, since they propose two different methods for attaining maximum improvement. The dominance hypothesis suggests that is possible to fix heterosis by inbreeding and selection. However, if the number of favorable and unfavorable allele pairs differentiating the parents is large, the probability of recovering an F_2 segregant with all the desired favorable dominant alleles is extremely small. Heterosis due to overdominance or pseudo-overdominance (i.e. nearby loci at which alleles having dominant or partially dominant advantageous effects are in repulsion linkage phase) is not fixable by inbreeding. Other possible gene actions playing a role in heterosis are multigene interactions like epistasis.

Classical genetic studies in maize were originally performed using fairly gross morphological, anatomical or behavioural traits as genetic markers. However, single gene markers causing a discrete and visible change in the phenotype, posed major limitations on such research because only a limited number of such markers could be followed in any given cross, and the markers themselves often produced undesirable phenotypic effects or showed recessiveness, dominance or epistasis (Tanksley et al. 1989). In the 1950s, it became possible to look at subtle variation in the structure of polypeptides. Allelic forms of enzymes (often referred to as isozymes) can be seperated electrophoretically, detected histochemically and mapped genetically, independent of any phenotypic changes. However, the number of polymorphic loci that could be detected this way was still limited. In the 1980s, it became standard practice to explore variation at the DNA level itself, in both coding and non-coding regions of the genome. The advent of molecular (DNA) marker techniques has provided geneticists with an virtually unlimited supply of phenotypically neutral markers, opening new possibilities and horizons for genetic mapping.

Technological developments during the last 20 years have expanded the range of DNA polymorphism assays, the first of which was restriction fragment length polymorphism (RFLP, Botstein et al. 1980). RFLP analysis relies on the

hybridization between a probe and homologous DNA segment(s) within the genome. The RFLP assay, as replacement of isozymes, has been used successfully for a wide range of plant species including maize. RFLPs provide the maximum of information possible because they are usually co-dominant. However, the requirement of relatively large amounts of pure, high molecular weight genomic DNA for RFLP assays, the lack of polymorphism revealed by RFLP for some species, the time consuming and labor intensive nature of preparing RFLP probes and performing RFLP analyses, and the unamenability to automation, have prompted the search for more efficient marker systems. In the past few years, a generation of polymerase chain reactions (PCR)–based DNA marker systems emerged. These marker systems, such as simple sequence repeat polymorphism (SSR, Tautz 1989), random amplified polymorphic DNA markers (RAPD, Welsh and McClelland 1990; Williams et al. 1990), and the AFLP® (Vos et al. 1995) assay, rely on the exponential amplification of subsets of the total amount of DNA sequence variation in a genome using specific oligonucleotide priming sets. SSR loci are comprised of highly variable arrays of tandemly repeated two to six nucleotides of DNA sequences. To develop SSRs as genetic markers, the surrounding sequence must be obtained in order to design PCR primers for regions directly outside the repeat. The primers must amplify uniquely the desired region, which must also be polymorphic. As RFLPs, SSRs provide the maximum of information possible because they are multi-allelic.

Multiple uncharacterized annealing sites are used as PCR primer targets in the RAPD analysis. This technique often results in several markers per assay, which are generally dominant and therefore not as informative as RFLP and SSR markers. No prior sequence information is required so development costs reduce to screening the genome with the primers to identify those that give most polymorphisms per reaction. However, a major drawback has been the extreme difficulty to obtain reproducible results, making these types of markers difficult to transfer between laboratories (Jones et al. 1997).

The AFLP assay combines the restriction site variation used in RFLP with DNA length differences targeted with RAPD primers. The DNA is cut into defined

fragments using restriction endonucleases and adaptor oligonucleotides are ligated to the overhanging ends. These adaptors serve as recognition sequences for PCR primers. The complexity of the mixture of fragments is decreased using selective PCR amplification, by adding additional nucleotides to the 3' termini of the PCR primers. The number of additional bases is adjusted such that 50 – 100 DNA fragments are amplified during the PCR. The AFLP method usually detects only one of the two alleles at a heterozygous locus (dominant marker system). However, quantitative analysis of AFLP marker bands in order to differentiate heterozygotes from both homozygous classes, using proprietary software developed specifically for AFLP analysis, at Keygene N.V., makes co-dominant scoring of AFLP markers feasible. AFLP reactions are readily automated and AFLP markers transfer well between laboratories (Jones et al. 1997).
As with most isozymes, these DNA-based markers are phenotypically neutral. But unlike isozymes, the genetic variation is surveyed directly at the DNA level and thus can reveal more polymorphism.

With the advent of biochemical markers and, subsequently, molecular markers, several investigators have estimated the correlations between marker allele diversity and heterosis of single cross hybrids in maize. These correlation studies are based on the 'distance' model, assuming that heterosis expressed by a hybrid is related to the genetic divergence between its parental lines (Lee et al. 1989). Compared with pedigree information, or with assessment of distance between maize inbred lines based on a large set of morphological traits recorded at several stages of plant growth and chosen for their discrimination power (Smith and Smith 1989), DNA marker-based diversity estimates better reflect DNA differences among lines. The potential of this 'distance' model-strategy has been extensively tested in maize, where genetic distances were first computed from isozyme data on parental inbreds (Frei et al. 1986), later from RFLPs (Lee et al. 1989; Godshalk et al. 1990; Melchinger et al. 1990 a, b; Smith et al. 1990; Dudley et al. 1991; Melchinger et al. 1992; Boppenmaier et al. 1993; Burstin et al. 1995) and recently from PCR-based RAPDs (Lanza et al. 1997) and AFLP markers (Ajmone

Marsan et al. 1998). The general tendency found was that the prediction effiency of the 'distance model' is high when 1) hybrids between related lines (intraheterotic crosses) and 2) hybrids between both related and unrelated lines (intra- and interheterotic crosses) are considered. However, correlations between genetic distances of unrelated lines only and their respective interheterotic crosses, were of low practical predictive value. This tendency is in good agreement with quantitative-genetic expectations (Charcosset and Essioux 1994), ascribing the failure of the distance model for interheterotic crosses to the fact that linkage associations between markers and quantitative trait loci (QTL) generally differ randomly from one heterotic group to the other.

Beside assessing the genetic distances between parental inbred lines, accurately estimating the genetic diversity among inbreds is also helpful in assigning lines to heterotic groups, maintaining and broadening the genetic variation of the elite gene pool, and identifying and accurately describing new varieties for the purpose of plant variety protection.

The discovery that loci can be placed into ordered arrangements (chromosomes), based on the observation of recombination events in meiosis, has led to the development of genetic linkage maps for a number of species such as maize. The only requirements for the construction of linkage maps are the availability of distinguishable alleles at a locus, the ability to perform genetic crosses and to analyze the resulting progeny. The development of genetic maps can be traced back to the first genetic experiments of Sturtevant (1913), who succeeded to linearly arrange multiple traits on a linkage group in *Drosophila*. Current public maize genetic linkage maps (http://www.agron.missouri.edu) are an extension of the first complete maps prepared by Rhoades and presented in Emerson et al. (1935) and are among the best developed in plant species. Linkage maps for maize are easily generated because of its diploid nature, the presence of numerous homozygous inbred lines and the high degree of polymorphism present between inbred lines. Beside morphological marker data, cytogenetic data (i.e. knobs, heterochromatic regions, Carlson 1988), reciprocal translocations, both

between members of the A chromosome set and between the supernumerary B chromosomes and the A chromosomes (Becket 1987), and biochemical markers, maize linkage maps are primarily based upon RFLP markers (Helentjaris et al. 1986, 1988; Burr et al. 1988; Beavis and Grant 1991; Shoemaker et al. 1992; Gardiner et al. 1993; Matz et al. 1994; Coe et al. 1995). Recently, Senior et al. (1997) published the first maize linkage map based upon SSR markers.

Maize breeders were also among the first to utilize the quantitative genetics theory in the development of breeding methods. Most morphological and reproductive traits in maize, particularly those of economic importance, are classified as multigenic or quantitative. The quantitative trait that still receives the major emphasis in most maize breeding programs is grain yield. Maize has been used effectively as a model organism in the development and evaluation of molecular markers for the identification, mapping and manipulation of major genes affecting the expression of quantitative traits in plants (Stuber 1995). The basic concept of associating markers with quantitative traits was first proposed by Sax (1923), reporting the association of seed coat pigmentation (a discrete monogenic trait) with seed size (a quantitatively inherited character) in *Phaseolus vulgaris*. Subsequent studies by Rasmusson (1933) and Everson and Schaller (1955) reported linkages between single genetic markers and quantitative trait loci (QTL). It was Thoday (1961) who greatly elaborated upon the subject, and put forth the idea to use independent and easily recognizable single genetic marker loci as flags in order to identify the regions of chromosomes containing the QTL and to estimate the mean effects of the QTL alleles.

Today, in maize, based on molecular markers, many QTL for grain yield and yield components have been identified (Edwards et al. 1987; Zehr et al. 1992; Beavis et al. 1994; Veldboom et al. 1994; Ajmone Marsan et al. 1995; Austin and Lee 1996; Cockerham and Zeng 1996; Eathington et al. 1997; Austin and Lee 1998). Finding a large number of QTL is not surprising in view of the complex nature of the trait yield (photosynthesis, transpiration, metabolism, nutritive uptake, lodging and insect resistance, environment, ...). Although direct comparisons of QTL are

complicated by differences in parental lines, design of the cross, number of progeny and the environments in which the progeny was assessed, as well as by different marker loci and QTL detection methods, reports have identified identical regions to be associated with QTL for grain yield. QTL for grain yield heterosis were identified and mapped by Stuber et al. (1992); although the majority of the QTL supported the overdominance hypothesis, results could not distinguish true overdominance from pseudo-overdominance.

Maize seems to keep its prominent role as model plant in the new era of genomic analysis ("genomics"), launched since a few years. Genomics covers both the structural elucidation of the organisation of a genome (large-scale sequencing) and the assignment of the function(s) that are carried out by the full complement of genes. Large-scale sequencing of the maize genome have already been initiated by a number of companies and universities. The overwhelming amount of available sequence information will dramatically change experimental plant biology by providing researchers with many avenues for gene discovery. Especially the saturated collection of DNAs from approximately 40,000 maize plants each carrying multiple-inserted *Mu* family elements, established by Pioneer Hi-Bred International, Inc., referred to as the Trait Utility System for Corn (TUSC; Canada and Meeley 1996), combined with the high level of synteny observed between maize and the other cereals (Moore 1995), makes maize useful in the functional analysis of genes.

Objectives and outline of this thesis

At Keygene N.V. maize has acquired a prominent place by serving as model plant in many research projects. Maize played a crucial role in the development of the AFLP technique as well as in AFLP-related technology development. Maize F_2 populations served as model segregating populations in the software-aided conversion of dominant AFLP markers into co-dominant markers (i.e., heterozygotes can be differentiated from both homozygous classes). The first application projects (e.g. genetic mapping, assessment of genetic diversity,

backcross analysis) as well as further characterization of AFLP data, were primarily carried out in maize.

In framework of the AFLP-related technology development, a modification of the AFLP method was developed, called the methylation AFLP® method (**Chapter 2**). This chapter describes several features of the methylation AFLP technique and illustrates how the technique can be used 1) to estimate the extent of CpG and CpNpG methylation for maize, 2) to demonstrate that most moderately to highly repetitive DNA sequences in maize are strongly methylated and 3) to generate AFLP fragments originally bounded by a methylated restriction site, in order to study hypermethylated portions of the genome.

In support of application projects, the inheritance of AFLP markers, whether or not bounded by a methylated restriction site, relative to already known RFLP markers and isozymes, was investigated in two segregating populations of maize and outlined in **Chapter 3.** To my knowledge, this is the first detailed report of mapping C-methylation and its stable transmission from parent to offspring. The efficiency of generating high-density linkage maps using the AFLP technology was evaluated in terms of multiplex (*M*), effective multiplex (*EM*) and effective mapped multiplex (*EMM*) ratio. Both genetic maps of maize could be aligned on the basis of common AFLP markers and the allele-specificity of AFLP markers across both populations could be investigated. AFLP markers generated by CNG methylation sensitive and CNG methylation insensitive enzyme combinations and AFLP markers collected from hypomethylated and hypermethylated regions were compared for their genomic distribution and their position relative to the centromere.

Aiming at further characterization of AFLP data as tool for the breeder, AFLP markers associated with different enzyme combinations and originating from different methylated genomic regions were compared for their polymorphism information content (PIC), marker index (MI) and patterns of genetic diversity

among a representative sample of maize inbred lines (**Chapter 4**). Furthermore, the effect of reducing marker information redundancy, the choice of enzyme combination and the bootstrap sample size on the bootstrap sample variance of marker data in the estimation of genetic similarities among inbred lines, was determined.

In contrast with the chapters mentioned so far, where the application and evaluation of the AFLP and the methylation AFLP method have been restricted to general questions encountered in many crop species, **Chapter 5** addresses a more maize-specific issue. Identification of genetic factors contributing to hybrid performance and/or heterosis and finding a suitable method that could predict hybrid performance and/or heterosis with some accuracy before field evaluation of hybrid performance and heterosis for grain yield are of particular interest in maize breeding. In **Chapter 5** a novel approach towards the prediction of heterosis and hybrid performance is presented. This approach is based on 1) the assessment of associations between markers and hybrid performance across a number of hybrids and 2) the assumption that the joint effect of genetic factors (loci) determined this way can be obtained by addition. Since the map position of the selected markers is known, putative QTL affecting the trait of interest are identified.

Finally, in **Chapter 6** the implications of the results described in this thesis are discussed with regard to applications in marker-assisted breeding and further research.

2

Characterization of the level and target sites of cytosine methylation in plants using methylation AFLP®[1]

M. Vuylsteke, M. Zabeau and P. Vos

Abstract

Methylation AFLP® is a novel PCR-based method to detect methylation of restriction sites randomly over the genome. The power of the methylation AFLP technique resides in its positive display of the unmethylated and the native methylated sites jointly and separately. The technique is based on selective amplification of genomic restriction fragments obtained from native methylated DNA and from unmethylated DNA amplified in vitro. The advantages of the methylation AFLP method are manifold: no prior sequence knowledge is needed, a limited set of generic primers is used, a high multiplex ratio is obtained and the detection of DNA methylation is genome-wide. Using the methylation AFLP® technique, we have estimated the extent of CpG and CpNpG methylation for maize, tomato and oilseed rape. In addition we demonstrated that most moderately to highly repetitive DNA sequences are strongly methylated and that the methylation AFLP technique allows the exploitation of two forms of DNA polymophisms: (i) variation in the primary nucleotide sequence either in the restriction sites or in the fragment size and (ii) methylation polymorphism.

Key words: AFLP®, DNA methylation, *Zea mays* L., epi-alleles, repetitive DNA

[1] submitted for publication

Introduction

DNA methylation has been associated with several cellular processes, including gene expression, epigenetic silencing of plant transgenes, regulation of transposons and the inheritance of epigenetic states (Martienssen and Richards 1995). These multiple functions indicate an important role for DNA methylation in species where it occurs. For most plants, there is little or no direct information on the extent, pattern or inheritance of methylated DNA. A variety of techniques to assess the degree of DNA methylation is presently available, which can be divided into sequence-unspecific and sequence-specific methods. The first category can be used to analyse the different types of modified bases and to quantify them, but do not provide any information about the precise location of the modified site within a given nucleic acid sequence. Into this category fall immunological, chromatographic, electrophoretic and spectrophotometric procedures that follow a complete chemical or enzymatic hydrolysis of the target DNA (reviewed in Saluz and Jost 1993). The second category enables analysis of the precise location of methylated bases within a known DNA sequence. Into this category fall techniques based on the use of pairs of isoschizomeric restriction enzymes that differ in sensitivity to methylation, in combination with Southern-blot analysis (Bird and Southern 1978; Waalwijk and Flavell 1978) or PCR. PCR-based methods for detecting DNA methylation have also been reported. One of these PCR-based methods takes advantage of the fact that PCR amplification occurs only if the DNA between the two primer sites remains uncleaved by the methylation-sensitive restriction enzyme *Hpa*II (Singer-Sam et al. 1990; Heiskanen et al. 1994). Other PCR-based methods combine PCR with genomic sequencing to identify methylated cytosine residues (Maxam and Gilbert 1980; Pfeiffer et al. 1989), utilizing the Maxam and Gilbert chemical cleavage reactions carried out on genomic DNA with various additional procedures to enhance the signal from the sequence under investigation. The bisulfite genomic sequencing technique described by Frommer et al. (1992) circumvents the drawbacks of the latter methods by eliminating the chemical cleavage reactions and providing a positive identification of 5-methylcytosine residues. Additional methods have been

developed recently which utilize bisulfite treatment of DNA as a starting point for methylation analysis. These include methylation-specific PCR (MSP) (Herman et al. 1996), restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA (Sadri and Hornsby 1996; Xiong and Laird 1997) and bisulfite treatment of DNA followed by single nucleotide primer extension (Gonzalgo and Jones 1997).

This paper describes methylation AFLP®, a novel sequence-specific, PCR-based technique, to detect methylation of restriction sites randomly over the genome. In contrast with the bisulfite genomic sequencing technique and relatives, the methylation of a virtually unlimited number of loci can be detected by the methylation AFLP method. The technique is based on the use of a pair of restriction enzymes consisting of a methylation-sensitive rare cutter and a methylation-insensitive frequent cutter. Complete digestion of the unmethylated rare cutter sites preceeds nonselective amplification of the genomic restriction fragments containing the methylated rare cutter site using PCR. Positive display of the unmethylated and native methylated sites jointly and separately is obtained after selective amplification of the restriction fragments by PCR. Further characteristics of the methylation AFLP method are: no prior sequence knowledge is needed, a limited set of generic primers is used and high multiplex ratio is obtained. This paper describes several features of the methylation AFLP technique and illustrates how the technique can be used to estimate the extent of CpG and CpNpG methylation for maize, tomato and oilseed rape. We also demonstrate that the methylation AFLP technique detects variation in the primary nucleotide sequence as well as methylation polymorphism.

Materials and methods

Sources of genomic DNA and enzymes

Tomato DNA (Near Isogenic Lines (NILs) 83M-S and 83M-R obtained from De Ruiter zonen, The Netherlands; cv. Motelle and Mobox obtained from INRA, Montfavet, France; inbred line RC10 obtained from Enza Zaden, The

Netherlands; inbred line 52201 obtained from Rijk Zwaan, The Netherlands), maize DNA (inbred lines B73, Mo17 and A7 obtained from Dr. M. Motto, Instituto Sperimentale per La Cerealicoltura, Bergamo, Italy; inbred lines D102, DK105, D107, D118, D140, D503, CO125, W401, D44, D01, D403 and D406 obtained from Dr. W.G. Polmer, University of Hohenheim, Stuttgart, Germany; 88 recombinant inbred lines derived from the cross B73 × Mo17, obtained from Dr. C.W. Stuber, North Carolina State University, Raleigh, NC, USA); oilseed rape DNA (12 genotypes obtained from Van der have, The Netherlands) were isolated using a modified CTAB procedure described by Stewart and Via (1993). All restriction enzymes were purchased from Pharmacia (Pharmacia LKB Biotechnology AB, Uppsala, Sweden), except for the restriction enzyme *Mse*I, which was purchased from New England Biolabs Inc. (Beverly, MA, USA). TaqStart™ Antibody was obtained from Clontech (Clontech Laboratories, Palo Alto, CA, USA). T4 DNA ligase and T4 polynucleotide kinase were also obtained from Pharmacia. All PCR reagents and consumables were obtained from Perkin Elmer Corp. (Norwalk, CT, USA). Radioactive reagents were purchased from Amersham (Amersham International plc, Little Chalfont, Buckinghamshire, UK) or Isotopchim (Isotopchim SA, Ganagobie, France).

AFLP primers and adapters

All oligonucleotides were made on a Biotronic Synostat D DNA-synthesizer (Eppendorf GmbH, Maintal, Germany) or Milligen Expedite 8909 DNA-synthesizer (Millipore Corp. Bedford, MA, USA). The quality of the crude oligonucleotides was checked by end-labeling with polynucleotide kinase and [γ-^{33}P]ATP and subsequent electrophoresis on 18% denaturing polyacrylamide gels (Maxam and Gilbert 1980). Oligonucleotide adapters and primers for AFLP analysis were generally used without further purification when they proved to be >85% full length.

For the rare cutter site two different AFLP adapters were used: 1) the conventional AFLP adapter (e.g. *Pst*I-adapter) as target site for primer annealing and 2) an AFLP adapter serving as blocking agent (e.g. *Pst*I-adapter*). Both adapters consist of a

core sequence (CORE) and a site-specific sequence (SITE) (Zabeau and Vos 1993). The blocking adapters differ from the conventional adapters only in core sequence. This is illustrated below for *Pst*I and *Hpa*II- adapters.

	CORE	**SITE**
*Pst*I-adapter:	5'-CTCGTAGACTGCGTACA	**TGCA**-3'
	3'-CATCTGACGCATGT-5'	
*Pst*I-adapter*:	5'-GCATCAGTGCATGCG	**TGCA**-3'
	3'-GTAGTCACGTACGC-5'	
*Hpa*II-adapter:	5'-CTCGTAGACTGCGTACA-3'	
	3'-CATCTGACGCATGT	**GC**-5'
*Hpa*II-adapter*:	5'-GCATCAGTGCATGCG-3'	
	3'-GTAGTCACGTACGC	**GC**-5'

The conventional and blocking adapter for *Msp*I and *Cla*I are identical to those for *Hpa*II.

For the frequent cutter site *Mse*I, also two different AFLP adapters were used: 1) a *Mse*I-adapter only for nonselective amplification (called *Mse*I-adapter$^+$) and 2) the conventional AFLP *Mse*I-adapter (called *Mse*I-adapter) as annealing target site for nonselective and further selective amplification. The *Mse*I-adapters differ only in core sequence. This is illustrated below:

	CORE	**SITE**
*Mse*I-adapter:	5'-GACGATGAGTCCTGAG-3'	
	3'-TACTCAGGACTC	**AT**-5'
*Mse*I-adapter$^+$:	5'-CTCGTAGACTGCGTACC-3'	
	3'-CTGACGCATGG	**AT**-5'

AFLP primers consist of three parts, a core sequence, a site-specific sequence (SITE) and a selective extension (EXT) (Zabeau and Vos 1993). This is illustrated below for *Pst*I-, *Hpa*II- and *Cla*I-primers with three selective nucleotides (shown as NNN):

	CORE	SITE	EXT
*Pst*I-primer:	5'-GACTGCGTACA	TGCAG	NNN-3'
*Hpa*II-primer:	5'-GACTGCGTACA	CGG	NNN-3'
*Cla*I-primer:	5'-GACTGCGTACA	CGAT	NNN-3'

AFLP primers for *Msp*I and *Hpa*II have a similar architecture.
Note that the AFLP primers for *Pst*I, *Hpa*II, *Msp*I and *Cla*I are designed only for the conventional AFLP adapters.
The two *Mse*I-primers are distinguished in the same way as the two *Mse*I-adapters: the *Mse*I-primer has the *Mse*I-adapter as annealing target site, while *Mse*I-primer$^+$ has the *Mse*I-adapter$^+$ as annealing target site. The difference between the two *Mse*I-primers is shown below:

	CORE	SITE	EXT
*Mse*I-primer:	5'-GATGAGTCCTGAG	TAA	NNN-3'
*Mse*I-primer$^+$:	5'-GTAGACTGCGTACC	TAA-3'	

Modification of DNA and template preparation

The protocols A, B and C below describe the generation of A-,B- and C-templates for AFLP reactions using the enzyme combination (EC) *Pst*I/*Mse*I. Protocol C is equivalent to the standard AFLP protocol (Vos et al. 1995).

Protocol A:

Genomic DNA (0.5 µg) is incubated for 1 h at 37°C with 5 U *Pst*I and 5 U *Mse*I in 40 µl 10mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA (acetylated). Next, 10 µl of a solution is added, containing 5 pMol *Pst*I-adapter*, 50 pMol *Mse*I-adapter, 50 pMol *Mse*I-adapter$^+$, 1U T4 DNA-ligase, 1mM ATP in 10 mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA, and the incubation is continued for 3 h at 37°C. Adapters are prepared by adding equimolar amounts of both strands; adapters are not phosphorylated. After ligation, the reaction mixture is diluted to 250 µl with 10 mM Tris-HCl, 0.1 mM EDTA pH 8.0, and used for PCR amplification or stored at -20°C.

Protocol B:

Genomic DNA (0.5 µg) is incubated for 1 h at 37°C with only 5 U *Mse*I in 40 µl 10mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA (acetylated). Next, 10 µl of a solution is added, containing 50 pMol *Mse*I-adapter, 50 pMol *Mse*I-adapter$^+$, 1U T4 DNA-ligase, 1mM ATP in 10 mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA, and the incubation is continued for 3 h at 37°C. Adapters are prepared as above (Protocol A). After ligation, the reaction mixture is diluted to 250 µl with 10 mM Tris-HCl, 0.1 mM EDTA pH 8.0, and used for PCR amplification or stored at -20°C.

Protocol C:

Genomic DNA (0.5 µg) is incubated for 1 h at 37°C with 5 U *Pst*I and 5 U *Mse*I in 40 µl 10mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA (acetylated). Next, 10 µl of a solution is added, containing 5 pMol *Pst*I-adapter, 50 pMol *Mse*I-adapter, 1U T4 DNA-ligase, 1mM ATP in 10 mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA, and the incubation is continued for 3 h at 37°C. Adapters are prepared as above (Protocol A). After

ligation, the reaction mixture is diluted to 500 µl with 10 mM Tris-HCl, 0.1 mM EDTA pH 8.0, and used for PCR amplification or stored at -20°C.

Synthesis of unmethylated DNA by nonselective PCR amplification and further modification of DNA and template preparation

Synthesis of unmethylated A- and B-templates is performed by nonselective PCR amplification. Nonselective PCR amplification is performed with the following cycle profile for 7 cycles: a 30 s DNA denaturation step at 94°C, a 1 min annealing step at 56°C and a 2 min extension step at 72°C. Amplifications are performed in 20 µl containing 5 µl template-DNA, 30 ng *Mse*I-primer, 30 ng *Mse*I-primer$^+$, 0.4 U *Taq* polymerase, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 50 mM Kcl and 0.2 mM of all four dNTPs. After amplification, the reaction mixtures A and B are incubated again for 1 h at 37°C with 5 U *Pst*I in 40 µl 10mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA(acetylated). Next, 10 µl of a solution is added, containing 5 pMol *Pst*I-adapter, 1U T4 DNA-ligase, 1mM ATP in 10 mM Tris-HAc pH 7.5, 10 mM MgAc, 50 mM KAc, 5 mM DTT, 50 ng/µl BSA (acetylated), and the incubation is continued for 3 h at 37°C. After ligation, the reaction mixture is diluted to 1000 µl with 10 mM Tris-HCl, 0.1 mM EDTA pH 8.0, and used for PCR amplification or stored at -20°C.

Generation of A- and B-templates for AFLP reactions using methylation-sensitive rare cutters leaving a 5'-extension (e.g. *Hpa*II, *Msp*I, *Cla*I) involves inactivation of the remaining *Taq* polymerase after the nonselective amplification, to avoid incorporation of remaining dNTPs after restriction. This is achieved by adding 220 ng TaqStart™ Antibody/ U *Taq* polymerase to the amplification mixture, prior to further restriction and ligation.

AFLP reactions

Amplification reaction conditions are described using DNA templates (A, B and C) for the EC *Pst*I/*Mse*I. With other ECs, the procedure is identical except for the use of appropiate primers.

AFLP fingerprinting of large genomes generally involves an amplification in two steps. The first step of this amplification procedure, named pre-amplification, employes two AFLP primers, one aimed at the *Pst*I-ends and one at the *Mse*I-ends, each having a single selective nucleotide. These primers are not radioactively labelled. Amplifications are performed in 20 µl containing 5 µl template-DNA, 30 ng *Mse*I-primer, 30 ng *Pst*I-primer, 0.4 U *Taq* polymerase, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl and 0.2 mM of all four dNTPs.
After pre-amplification, the reaction mixtures are diluted 20-fold with 10 mM Tris-HCl, 0.1 mM EDTA pH=8.0, and used as templates for the second amplification reaction.

The second amplification reaction again employes two oligonucleotide primers, one aimed at the *Pst*I-ends and one at the *Mse*I-ends, each having two or three selective nucleotides. The *Pst*I-primer is radioactively end-labeled using [γ-^{33}P] ATP and T4 polynucleotide kinase. The labelling reactions are performed in 50 µl 25 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5mM DTT, 0.5mM spermidine-3HCl using 500 ng oligonucleotide primer, 100 µCi [γ-^{33}P] ATP (1000-3000Ci/mol) and 10 U T4 polynucleotide kinase. Amplifications are performed in 20 µl containing 5 µl DNA-template, 5 ng labeled *Pst*I-primer (0.5 µl from the labelling reaction mixture), 30 ng *Mse*I-primer, 0.4 U *Taq* polymerase, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl and 0.2 mM of all four dNTPs.

AFLP preamplification reactions are performed for 24 cycles (protocol A and B) and 20 cycles (protocol C) with the following cycle profile: a 30 s DNA denaturation step at 94 °C, a 1 min annealing step at 56°C and a 1 min extension step at 72°C. AFLP reactions with primers having two or three selective nucleotides are performed for 36 cycles with the following cyle profile: a 30 s DNA denaturation step at 94°C, a 30 s annealing step (see below) and a 1 min extension step at 72°C. The annealing temperature in the first cycle is 65 °C, and is subsequently reduced each cycle by 0.7 °C for the next 12 cycles, then continued at 56 °C for the remaining 23 cycles. All amplification reactions are performed in a PE-9600 thermocycler (Perkin Elmer Corp. Norwalk, CT, USA).

Gel analysis

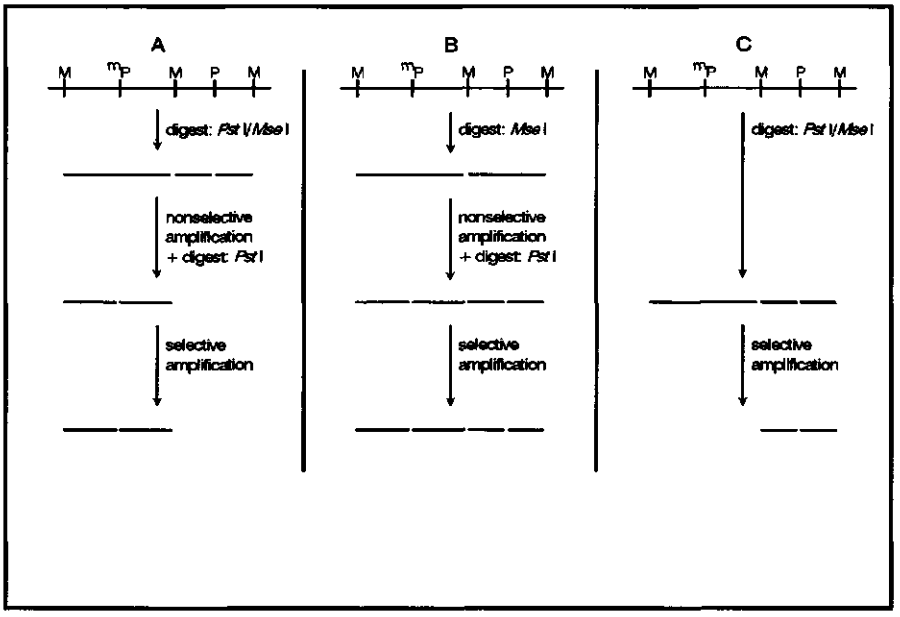
Following amplification, reaction products are mixed with an equal volume (20 µl) of formamide dye (98% formamide, 10 mM EDTA pH 8.0, and bromophenol blue and xylene cyanol as tracking dyes). The resulting mixture is heated for 3 min at 90 °C, then quickly cooled on ice. Each sample (2 µl) was loaded on a 5% denaturing (sequencing) polyacrylamide gel (Maxam and Gilbert 1980). The gel matrix is prepared using 5% acrylamide, 0.25% methylene bisacryl, 7.5 M urea in 50 mM Tris/50 mM Boric acid/1mM EDTA (pH 8.3). To 100 ml of gel solution 500 µl of 10% APS and 100 µl TEMED is added and gels are cast using a SequiGen 38x50 cm gel apparatus (Biorad Laboratories Inc., Hercules, CA, USA). 100 mM Tris/100mM Boric acid/2ml EDTA was used as running buffer. Electrophoresis is performed at constant power, 110W, for ~ 2 h. After electrophoresis, gels are fixed for 30 min in 10% acetic acid, rinsed with water for 10 min, dried on the glass plates and exposed to Fujix phosphorimage screens for 16 h. Fingerprint patterns are visualized using a Fujix BAS-2000 phosphorimage analysis system (Fuji Photo Film Company Ltd., Japan).

Results

Principle of the method

The methylation AFLP technique enables the discriminative detection of methylated and unmethylated sequences by separate display of restriction fragments originating from either methylated or unmethylated restriction enzyme sites. The technique employs a methylation sensitive restriction endonuclease in combination with methylation insensitive restriction enzyme for AFLP template preparation. Three types of genomic AFLP template DNA are used, denoted A, B and C, derived from either native or in vitro amplified genomic DNA. The general outline of the method is depicted in Figure 2.1 and is exemplified for the EC *Pst*I/*Mse*I. The three types of template DNA are prepared in the following way.

Figure 2.1 Schematic representation of the methylation AFLP technique for the EC *Pst*I/*Mse*I: 'nonselective amplification' stands for the nonselective *Mse*I/*Mse*I* amplification; 'selective amplification' stands for the selective *Pst*I/*Mse*I amplification. Left: generation of subset A-templates, representing restriction fragments with a native methylated *Pst*I site, followed by the selective *Pst*I/*Mse*I amplification procedure. Middle: generation of subset B-templates, representing the restriction fragments with a native methylated or unmethylated *Pst*I site, followed by the selective *Pst*I/*Mse*I amplification procedure. Right: generation of subset C-templates, representing the restriction fragments with a native unmethylated *Pst*I site, followed by the selective *Pst*I/*Mse*I amplification procedure.



For A templates the genomic DNA is digested with *Pst*I and *Mse*I, followed by ligation of ds *Pst*I- and two types of *Mse*I-adapters to the restriction fragments. Subsequently, a "whole genome amplification" step is carried out using two types of nonselective *Mse*I primers. In this way most *Mse*I fragments will be amplified except the fragments which have cleavable, i.e. unmethylated, *Pst*I sites within the *Mse*I fragment. The resulting MseI fragments are again digested with *Pst*I followed by ligation of *Pst*I-adapters (a different *Pst*I-adapter is used now). AFLP amplification of A templates will display the *Pst*I-*Mse*I fragments originating from native methylated *Pst*I sites.

For B templates the genomic DNA is digested with MseI only and two types of *Mse*I-adapters are ligated to the restriction fragments. Subsequently, a "whole genome amplification" step is carried out using two types of nonselective *Mse*I primers. In this way unmethylated genomic *Mse*I fragments will be generated. The resulting *Mse*I fragments are digested again with *Pst*I and *Pst*I-adapters are ligated. AFLP amplification of B templates will display the *Pst*I-*Mse*I fragments derived from native methylated and unmethylated PstI sites simultaneously.

C templates are standard genomic AFLP templates (Vos et al. 1995) and will display only PstI-MseI fragments from native unmethylated *Pst*I sites.

The methylation AFLP technique performs the parallel display of fingerprints from A, B and C templates detecting fragments derived from methylated *Pst*I sites (A), fragments derived from both methylated and unmethylated *Pst*I sites (B) and fragments derived from native unmethylated *Pst*I sites (C).

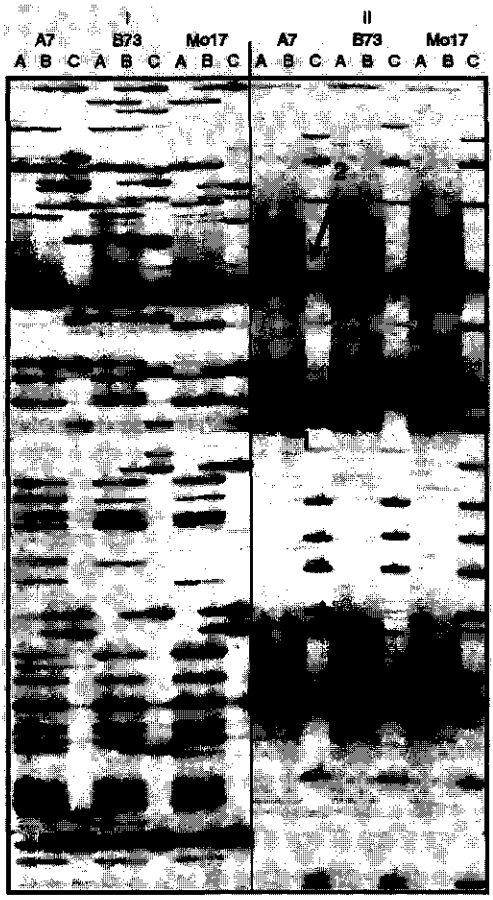
Methylation AFLP fingerprinting of large genomes

Initial experiments with methylation AFLP fingerprinting of plant genomic DNA indicated that the complexity of the template libraries is of the same order as in the standard AFLP protocol. For maize, however, the number of the fragments detected in the B reaction (~ A and C reaction jointly) by *Pst*I+2/*Mse*I+3 primer combinations (PCs) is substantially elevated, so 'tuning' to a slightly higher selectivity of the PCs is desirable and results in combined *Pst*I PCs, e.g. *Pst*I+AGW or *Pst*I+AGS, where W stands for A and T jointly, and where S stands for C and G jointly.

Figure 2.2 shows a number of typical methylation AFLP fingerprints obtained with the two step amplification strategy of genomic DNA from the three maize inbred lines A7, B73 and Mo17.

Although large plant genomes consist predominantly of unique AFLP fragments (Vos et al. 1995; Keim et al. 1997; Qi et al. 1998; Marques et al. 1998), the presence of repetitive DNA is reflected by fragments of moderate (arrow 1) or high (arrow 2) band intensities (Figure 2.2 PCII), in good accordance with the correlation that exists between AFLP band intensity and the original number of

Figure 2.2 Methylation AFLP fingerprints of genomic DNA from the three maize inbred lines A7, B73 and Mo17. The two panels show *Pst*I/*Mse*I fingerprints, corresponding with the following PCs (from left to right): I. *Pst*I+AGW/*Mse*I+CTT, and II. *Pst*I+AGS/*Mse*I+CTT. Lane A, B and C represent the corresponding A, B and C restriction fragments, referring to the native methylated state of the rare cutter sites. The molecular weight size range of the fingerprints is 200-500 nucleotides.



restriction fragments (Vos et al. 1995). Since these multicopy restriction fragments are more abundant in lane A than in lane C, we can conclude that most moderately to highly repetitive sequences are strongly methylated.

Theoretically, every band present in lane A or lane C should also be present in lane B, because lane B represents both the native methylated and unmethylated rare cutter sites jointly. In some cases, however, fragments present in lane C are

missing in lane B. There are two possible explanations for this: 1) short *Mse*I restriction fragments have a competitive advantage over the longer *Mse*I restriction fragments in the nonselective amplification. Consequently, unmethylated rare cutter restriction sites, on average residing more on long *Mse*I restriction fragments, may not be detected. Subsequent restriction and ligation of the underrepresented fragments will generate a reduced amount of templates; 2) co-amplification with highly repetitive restriction fragments makes it difficult for single copy fragments to reach the detection limit.

Estimating the extent of ^{5m}CpNpG as presented in PstI-sites of some large plant genomes

In plants the modified C at position 5 (^{m5}C) is not only confined to CpG dinucleotides as is in animals, but also occurs at a variety of other cytosine containing dinucleotides, all of which are part of the basic trinucleotide CpNpG where N can be any nucleotide (Gruenbaum et al. 1981). The restriction enzyme *Pst*I is sensitive to methylation at the C-residues at position 1 and 4 of its recognition site 5'-CTGCAG-3', containing two CpNpG trinucleotides (McClelland et al. 1994). A mean percentage ± standard error of methylated *Pst*I-sites in the nuclear DNA of tomato, maize and oilseed rape, are given in Table 2.1.

It is clear from Table 2.1 that variation in the percentage of methylated *Pst*I-sites is very low within a species. For tomato and maize, percentages of ^{5m}C residues in the nuclear DNA (expressed as ^{5m}C/^{5m}C +C) calculated as half the number of methylated *Pst*I-sites, are very similar to published percentages, based on HPLC analysis (Messeguer et al. 1991; Montero et al. 1992) (Table 2.1). These results suggest that, as assumed, only one C residue in the *Pst*I recognition site is methylated.

Table 2.1 Mean percentages ± standard error of methylated *Pst*I-sites and ^{5m}CpNpG[a] as presented in *Pst*I-sites in the nuclear DNA of tomato, maize and oilseed rape, and published mean percentages ^{5m}C residues (expressed as % ^{5m}C/^{5m}C +C (Messeguer et al. 1991; Montero et al. 1992)) in the nuclear DNA of tomato and maize. N= number of genotypes assayed. n = total number of *Pst*I/*Mse*I restriction fragments counted.

species	N	% *Pst*I-sites methylated	% CpNpG-sites methylated	n	% ^{5m}C/ ^{5m}C+C
tomato	6	39.75 ± 1.10	19.88 ± 0.55	219 - 248	22.8
maize	12	50.78 ± 1.33	25.39 ± 0.67	753 - 820	28.9
oilseed	12	33.12 ± 0.69	16.56 ± 0.35	548 - 609	

[a] in order to estimate the percentage of ^{5m}C residues in the nuclear DNA (expressed as ^{5m}C/^{5m}C +C) and to compare with published percentages, percentage of ^{5m}CpNpG-sites as presented in *Pst*I-sites were assumed to be half the number of methylated *Pst*I-sites.

Estimating the extent of ^{5m}CpG as presented in *Hpa*II-, *Msp*I- and *Cla*I-sites in some large plant genomes

*Msp*I and *Hpa*II (methylation isoschizomers) have the same recognition site 5'-CCGG-3', containing one CpG dinucleotide. *Msp*I is methylation sensitive only when the outer C is methylated, whereas *Hpa*II is sensitive to methylation at either C (McClelland et al. 1994). Therefore, *Msp*I and *Hpa*II are appropriate for estimating the extent of ^{5m}CpG. The extent of ^{5m}CpG methylation by using *Hpa*II and *Msp*I as methylation-sensitive rare cutters is measured as the difference in the number of bands counted in a *Msp*I and a *Hpa*II fingerprint in lane C; the difference in the number of bands counted in a *Msp*I and a *Hpa*II fingerprint in lane A must give the same result.

Another methylation-sensitive restriction enzyme *Cla*I, having the CpG-dinucleotide containing recognition site 5'-ATCGAT-3', and affected by ^{5m}CpG (McClelland et al. 1994) was also found to be a suitable rare cutter in the detection of ^{5m}CpG.

The percentages of ^{5m}CpG sites in the nuclear DNA of tomato and maize, as presented in *Hpa*II- and *Msp*I-sites, and the percentages of ^{5m}CpG as presented in

*Cla*I-sites in the nuclear DNA of tomato, maize and oilseed rape, were measured for only one genotype/species, and are given in Table 2.2.

Table 2.2 Percentages of ^{5m}CpG in the nuclear DNA of tomato, maize and oilseed rape, as presented in *Cla*I-sites, and percentages of ^{5m}CpG in the nuclear DNA of tomato and maize as presented in *Hpa*II- or *Msp*I-sites. n = total number of restriction fragments counted.

species	% ^{5m}CpG as presented in *Cla*I sites	n	% ^{5m}CpG as presented in *Hpa*II/*Msp*I sites	n
tomato (cv.52201)	57.9	2833	57.6	2169
maize (B73)	53.9	2694	39.6	1791
oilseed(T528)	48.4	2973		

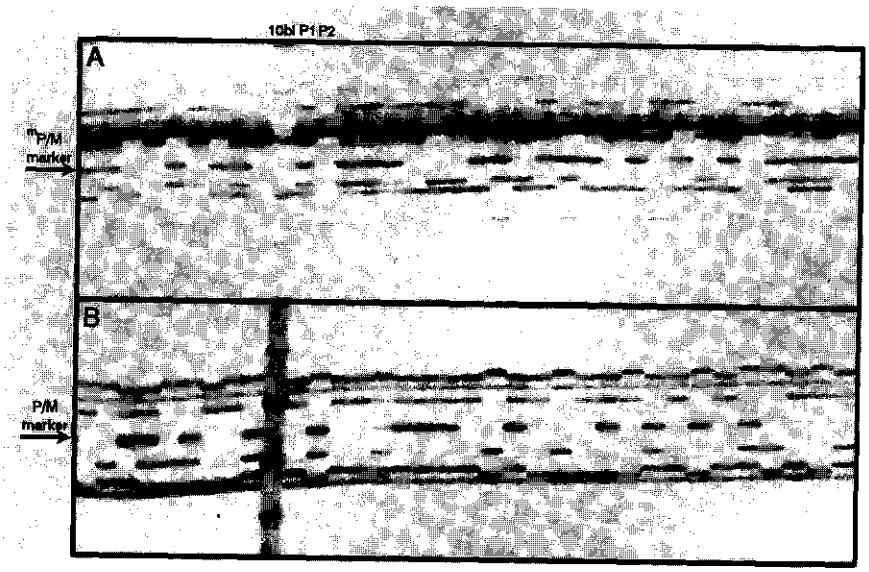
In tomato, a similar percentage of ^{5m}CpG as presented in *Cla*I-sites is obtained as presented in either *Hpa*II- or *Msp*I-sites, in contrast with maize. Estimations of the percentage of ^{5m}C residues as presented in *Cla*I-sites in the nuclear DNA of tomato, maize and oilseed rape is significantly higher than the corresponding percentages ^{5m}C residues as presented in *Pst*I-sites. This suggests that the majority of ^{5m}C residues in the tomato, maize and oilseed rape genome exist at CpG sites.

C-methylation polymorphism

Since DNA methylation occurs at defined target sequences (i.e. CpG and CpNpG) in specific genomic regions, methylation polymorphism may represent a potentially important form of DNA polymorphism. The methylation AFLP technique detects two forms of DNA polymophisms: 1) variation of the primary nucleotide sequence either in the restriction sites (point mutation) or in the fragment size (deletion/insertion), resulting in ^{m}AFLP markers; and 2) allele-specific methylation (methylation polymorphism) resulting in ^{asm}AFLP markers. Since DNA methylation is also the only source of allelic difference between epi-alleles, ^{asm}AFLP markers are epiallelic markers. To identify a AFLP and a ^{m}AFLP marker as a possible ^{asm}AFLP marker pair, the following criteria are applied: 1) both AFLP

fragments have exactly the same size; 2) both AFLP fragments are derived from different parents, with the same PC; and 3) both AFLP fragments map to the same locus (complementary segregation). Figure 2.3A shows the segregation of a number of ^{m}AFLP markers in a Recombinant Inbred (RI) population derived from the cross B73 × Mo17, while Figures 2.3A and 2.3B show the complementary segregation of a pair of ^{asm}AFLP markers.

Figure 2.3 The segregation of (A) a number of *Pst*I/*Mse*I ^{m}AFLP markers and (B) *Pst*I/*Mse*I AFLP markers in RI population derived from the cross B73 × Mo17: 10bl is the 10 base ladder, P1 and P2 are the parental lines B73 and Mo17, respectively. The two marked markers in A and B are a pair of *Pst*I/*Mse*I ^{asm}AFLP markers.



Discussion

Methylation AFLP is a DNA fingerprinting technique that detects genomic restriction fragments and in that respect resembles the AFLP technique. The major difference is that methylation AFLP method detects native methylated sites, but only

if they are located in the recognition site of the methylation-sensitive rare cutter. As for the AFLP technique, the multiplex ratio is high (50-100 restriction fragments) and is a function of (i) the cleavage frequency of the methylation-sensitive rare cutter enzyme and (ii) the number and nature of the selective bases of the specific primer set.

In this paper, we have shown the use of the methylation AFLP method to estimate the extent of ^{5m}CpNpG and ^{5m}CpG in the genomes of a few plants, as presented in *Pst*I-, *Cla*I-, *Hpa*II/*Msp*I-sites. The methylation AFLP method can provide valuable estimations of the extent of ^{5m}CpNpG and ^{5m}CpG in any genome, although generalization of these estimates to percentages ^{5m}CpNpG and ^{5m}CpG in the genome, may in some cases not be fully representative as methylation of C can be biased by the nature of its flanking nucleotides

(e.g. recognition site), or by the percentage G+C of the genomic region it resides (e.g. CG-islands). Another potential use of the technique described is the estimation of the low prevalence of ^{5m}C-residue in non-symmetrical dinucleotides (e.g. ^{5m}CpC), (Selker et al. 1993; Meyer et al. 1994), or modified nucleotides other than C (e.g. ^{m6}A) (McClelland et al. 1994), dependent upon the availability of the appropriate restriction enzymes. We have also demonstrated that most moderately and highly repetitive sequences are strongly methylated. Methylation of repetitive sequences, of which some are transposableelements, is thought to serve as a genome-defense mechanism that guards against the deleterious effects of multicopy transposable elements. Methylation of repetitive sequences might also suppress recombination between repeats in different genomic positions and guard the genome against aberrant gene duplications (reviewed by Bender 1998).

The methylation AFLP technique allows the exploitation of two forms of DNA polymorphisms: (i) polymorphism reflecting variation in the primary nucleotide sequence of the methylated restriction site and/or variation in the restriction fragment size, and (ii) methylation polymorphism. By monitoring the segregation of both forms of DNA polymorphism in a RI population, it was shown that DNA methylation inherits in a Mendelian fashion to the offspring and co-segregates in perfect accordance with the primary target site (Chapter 3). Of a total of 673

mapped $^{(m)}$AFLP markers, generated by a set of 14 *PstI/MseI* PCs, ^{m}AFLP markers account for 44.6%, which is in good accordance with the 50.8% methylated *PstI*-sites in the maize genome (Table 2.1). Only 1% of the marker alleles behave like epi-alleles, indicating that methylation polymorphism at the CpNpG target site is not common in maize (Chapter 3). This low percentage shows the high accuracy of the transmission of a given pattern of ^{5m}C, and can account for the specificity of the pattern that can be highly conserved between individuals.

The technique described here has many more potential uses. ^{m}AFLP markers corresponding to unique positions in the genome, can be exploited as landmarks in and as bridging tools between genetic and physical maps (Vos et al. 1995), and used to distinguish native methylated sites from native unmethylated sites on a physical map. Therefore, the methylation AFLP technique has the potential to shed a light on the distribution of DNA methylation on the genetic map (Chapter 3) as well as on the physical map. Finally, this method can be applied, due to its sensitive, reliable and quantitative nature, to identify variable, tissue-specific levels of DNA methylation at specific gene loci and to trace epigenetic changes at the DNA methylation level upon environmental stress.

Acknowdledgements

We would like to thank Jerina Pot for artwork and Martin Kuiper and Michiel van Eijk for critically reading the manuscript. This work was financed by Keygene N.V. and 6 Dutch breeding companies: Cebeco Zaden, De Ruiter, De ZPC, ENZA Zaden, Rijk Zwaan and Van der Have.

3

Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers [1]

M. Vuylsteke, R. Mank, R. Antonise, E. Bastiaans, M.L. Senior, C.W. Stuber, A.E. Melchinger, T. Lübberstedt, X.C. Xia, P. Stam, M. Zabeau and M. Kuiper

Abstract

This study demonstrates the relative ease of generating high-density linkage maps using the AFLP® technology. Two high-density AFLP linkage maps of Zea mays L. were generated based on: 1) a B73 × Mo17 Recombinant Inbred population; and 2) a D32 × D145 Immortalized F_2 population. Although AFLP technology is in essence a dominant marker system, markers can be scored quantitatively and used to deduce zygosity. AFLP markers were generated using the enzyme combinations *Eco*RI/*Mse*I and *Pst*I/*Mse*I. A total of 1539 and 1355 AFLP markers have been mapped in the two populations, respectively. Among the mapped *Pst*I/*Mse*I AFLP markers we have included fragments bounded by a methylated *Pst*I site ("AFLP markers). Mapping these "AFLP markers shows that the presence of C-methylation segregates in perfect accordance with the primary target sequence, leading to Mendelian inheritance. Simultaneous mapping of *Pst*I/*Mse*I AFLP and *Pst*I/*Mse*I "AFLP markers allowed us to identify a number of epi-alleles, showing allelic variation in the CpNpG methylation only. However, their frequency in maize is low. Map comparison shows that, despite some rearrangements, most of the AFLP markers that are common in both populations, map at similar positions. This would indicate that AFLP markers are predominantly single locus markers. Changes in map order occur mainly in marker-dense regions. These marker-dense regions, representing clusters of mainly *Eco*RI/*Mse*I AFLP and *Pst*I/*Mse*I "AFLP

[1] accepted for publication in *Theoretical and Applied Genetics*

markers, co-localize well with the putative centromeric regions of the maize chromosomes. In contrast, *Pst*I/*Mse*I markers are more uniformly distributed over the genome.

Key words: *Zea mays* L., AFLP®, methylation AFLP®, genetic map, DNA methylation

Introduction

High-density genetic maps are becoming increasingly useful in fundamental and applied genetic research. They serve to 1) locate genes of interest, 2) facilitate marker-assisted breeding and map-based cloning and 3) provide the framework towards understanding the biological basis of complex traits. In genome projects, high-density genetic maps are central to localizing a large portion of the loci in the germplasm of interest and to top-down anchoring of physical maps.

Until recently, genetic maps of many plant species such as maize (*Zea mays* L.) were primarily based upon segregating restriction fragment length polymorphism (RFLP) markers (Helentjaris et al. 1986, 1988; Burr et al. 1988; Beavis and Grant 1991; Shoemaker et al. 1992; Gardiner et al. 1993; Matz et al. 1994; Coe et al. 1995). The disadvantages of RFLPs include 1) large quantities of DNA are required, 2) analyses of large populations are costly and 3) the technique is difficult to automate. This has prompted the search for more efficient marker systems. Of these, the randomly amplified polymorphic DNA (RAPD) assay (Welsh and McClelland 1990; Williams et al. 1990) has been widely used in plant breeding and genetics (Waugh and Powell 1992). However, problems with the reproducibility of RAPD amplification were reported (Demeke et al. 1997; Karp et al. 1997). Simple sequence repeat polymorphisms or microsatellites (SSR) (Tautz et al. 1989), a marker system first made popular in mammalian genetics, has generated considerable interest among plant geneticists. SSR markers were developed for many plant species, including maize (Senior et al. 1997). However,

the high development and application costs may hinder their application in the large numbers needed to study, e.g., a large germplasm collection.

Another efficient polymerase chain reaction (PCR)-based method, called AFLP® has been developed (Vos et al. 1995), combining the restriction site recognition element of RFLP with the exponential amplification aspects of PCR-based DNA marker systems. The major advantages of AFLP are 1) a high multiplex ratio, 2) a limited set of generic primers is used, and 3) no need for sequence information. To date, the AFLP technique has been succesfully applied to identify markers linked to disease resistance loci (Meksem et al. 1995; Thomas et al. 1995; Cervera et al. 1996; Sharma et al. 1996; Brigneti et al. 1997; Simons et al. 1997; Vos et al. 1998), in germplasm analyses (Hill et al. 1996; Powell et al. 1996; Maughan et al. 1996; Milbourne et al.1997; Paul et al. 1997; Schut et al. 1997; Zhu et al. 1998; Cervera et al. 1998; Barrett et al. 1998; Pejic et al. 1998), and in mapping barley (Becker et al. 1995; Qi et al. 1998; Castiglioni et al. 1998), eucalypts (Marques et al. 1998), potato (van Eck et al. 1995), rice (Maheswaran et al. 1997; Zhu et al. 1998), soybean (Keim et al. 1997) and sugar beet (Schondelmaier et al. 1996). AFLP publications in maize are today limited to Ajmone Marsan et al. (1998), Melchinger et al. (1998a) and Pejic et al. (1998).

In the present study, two high-density AFLP linkage maps of *Zea mays* L. were generated. The aim of this study was to 1) confirm the relative ease of generating high-density maps using the AFLP technology, 2) evaluate the efficiency of the AFLP technology in terms of multiplex (*M*), effective multiplex (*EM*) and effective mapped multiplex (*EMM*) ratio for linkage analysis, 3) study the transmission of C-methylation from parent to offspring and 4) to investigate the consistency of AFLP markers across populations. In addition, the data obtained enable a comparison of the genomic distribution of AFLP and ^{m}AFLP markers and of their position relative to the centromere.

Materials and methods

Plant material and DNA isolation

A Recombinant Inbred (RI) population (Senior et al. 1997) and an Immortalized F_2 (IF_2) population (Xia et al. 1998) involving four different inbred lines of *Zea mays* L. were used for this study. The parents of the RI population were B73, a central corn belt line derived directly from Iowa Stiff Stalk Synthetic (BSSS), and Mo17, a central corn belt line derived from Lancaster and Krug germplasm. The RI population was set up as follows: an F1 was produced from a cross between B73 and Mo17. Selfing for 2 generations produced a set of 264 F_3 lines. Two hundred and eight RI lines were developed from these F3 plants by single seed descent for 3 generations. DNA of the RIs was extracted by a modified CTAB procedure described by Saghai Maroof et al. (1984).

The parents of the IF_2 population were D32, a sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV) resistant European Dent line having an Iodent and Illinois High Protein (IHP) background, and D145, a SCMV and MDMV susceptible European Flint with Lancaster background (Melchinger et al. 1998b). The IF_2 population has been developed at the University of Hohenheim, Stuttgart, Germany and was set up as follows: 1) an F1 was produced from a cross between D32 and D145; 2) selfing for 2 generations produced a set of 220 F_3 families. Per F_3 family, a pool of 20 F_3 plants was chosen to generate the IF_2 lines; 3) random crosses were performed between 10 F_3 plants as female and the 10 F_3 plants as male; and 4) seed obtained from the 10 females was subsequently pooled, representing an IF_2 line. Thus, 220 IF_2 lines were obtained. At the DNA level, each IF_2 line represented a pool of 60 sib-mated F_3 plants. DNA of the pooled sib-mated F_3 plants was extracted using a modified CTAB procedure described by Stewart et al. (1993).

AFLP and methylation AFLP analysis

The AFLP analysis was performed according to Vos et al. (1995), using the enzyme combinations (ECs) *Eco*RI/*Mse*I and *Pst*I/*Mse*I. Methylation AFLP® analysis was

performed according to Chapter 2, using the EC *Pst*I/*Mse*I. The adaptor sequences specific for these enzymes were synthesized according to Zabeau and Vos (1993) and are as follows:

*Eco*RI-adapter:	5'-CTCGTAGACTGCGTACC-3' 3'-CATCTGACGCATGGTTAA-5'
*Pst*I-adapter:	5'-CTCGTAGACTGCGTACATGCA-3' 3'-CATCTGACGCATGT-5'
*Pst*I-adapter*:	5'-GCATCAGTGCATGCGTGCA-3' 3'-GTAGTCACGTACGC-5'
*Mse*I-adapter:	5'-GACGATGAGTCCTGAG-3' 3'-TACTCAGGACTCAT-5'
*Mse*I-adapter$^+$:	5'-CTCGTAGACTGCGTACC-3' 3'-CTGACGCATGGAT-5'

The non-selective amplification of the restriction fragments during the methylation AFLP analysis is performed using both of the *Mse*I-primers shown below.

A two-step amplification strategy was followed in the methylation AFLP as well as in the AFLP analysis: in a selective preamplification, the restriction fragments were amplified with AFLP primers both having a single selective nucleotide. In the second step, further selective amplification was carried out using primers having two (*Pst*I-primer) or three (*Eco*RI- and *Mse*I-primer) selective nucleotides. The AFLP primers were designed based on the adapter sequence and restriction sites of *Eco*RI, *Pst*I and *Mse*I, and have the following sequences:

*Pst*I-primer	5'-GACTGCGTACATGCAG...NN-3'
*Eco*RI -primer	5'-GACTGCGTACCAATTC...NNN-3'
*Mse*I-primer	5'-GATGAGTCCTGAGTAA...NNN-3'
*Mse*I-primer$^+$	5'-GTAGACTGCGTACCTAA-3'

Hereinafter *Eco*RI/*Mse*I and *Pst*I/*Mse*I ECs will be referred to as E/M and P/M ECs and *Eco*RI+3/*Mse*I+3 and *Pst*I+2/*Mse*I+3 PCs will be referred to as E/M PCs and P/M PCs. The P/M PCs used in the methylation AFLP analysis will be referred to as ^{m}P/M PCs.

AFLP marker nomenclature

Each polymorphic AFLP fragment was identified by: 1) the code referring to the corresponding PC (see Table 3.1) followed by 2) the estimated molecular size of the DNA fragment in nucleotides and 3) a code indicating the parental origin of the fragment (RI population: P1=B73; P2=Mo17. IF_2 population: P1=D32; P2=D145). SequaMark™ (Research Genetics, Huntsville, AL, USA) was implemented as size standard to assign molecular weights to the AFLP fragments. Fragments and markers detected by E/M, P/M or ^{m}P/M PCs will be referred to as E/M, P/M or ^{m}P/M fragments and markers, respectively. Markers detected by ^{m}P/M PCs will also be referred to as AFLP markers, except when AFLP and ^{m}AFLP markers need to be distinguished.

Analysis of gel images

The AFLP technology usually detects only one of the two alleles at a heterozygous locus is detected. However, since product concentration directly reflects initial template concentration, the expected difference between a heterozygous locus and a homozygous locus is an approximately two-fold difference in intensity of a band (= a reflection of the fragment quantity). This phenomenon is exploited to quantitatively analyze AFLP marker bands in order to deduce zygosity. Thus, AFLP markers can in principle be scored quantitatively as co-dominant markers, i.e., heterozygotes can be differentiated from both homozygous classes. For the analysis of complex AFLP fingerprint patterns, we have used proprietary software developed specifically for AFLP analysis, at Keygene. This software allows the display and analysis of pixel images of X-Ray scans or phosphorimager scans. For analysis of pixel images, the software has tools to navigate through the image and individual band signals, and to size and quantify the AFLP bands with great

precision. Each band of a specific marker is classified with respect to its intensity, using a mixture model of normal distributions, described by Jansen (1993). The basic idea behind quantifying band intensities is that the observed intensities of a marker are mixtures of two (RI lines) or three (F_2 plants) normal distributions. The estimated proportions, means and variances of the mixture components form the basis of band classification and of determination of genotypes. The algorithm can be set to identify either two classes (RI lines) or three classes (F_2 individuals) of intensities among the bands. Finally, genotypic data are exported to a file for each of the samples, for each marker.

The absolute metrics multiplex ratio (*M*), effective multiplex ratio (*EM*) (Powell et al. 1996) and effective mapped multiplex ratio (*EMM*) define the number of fragments, polymorphic fragments and mapped polymorhic fragments, respectively, simultaneously analyzed in a single assay. To calculate *M*, the AFLP fragments in the two parental lanes had to meet the following criteria: 1) fragment size ranges from 50bp to 500bp; and 2) the mean intensity of the two parental bands must be higher than an intensity minimum (empirically defined by visual inspection). To calculate *EM*, polymorphic bands were discriminated from non-polymorphic bands by a two-fold difference (empirically determined) in intensity between the parental bands. The three metrics *M*, *EM* an *EMM* are suited in order to facilitate selection of an appropriate EC or appropriate PCs for a given application. Although these metrics are influenced by the number of selective nucleotides at the 3' ends of the PCR primers and can be manipulated by combining PCs in a multiplex reaction approach, it is useful to compare the *M*, *EM* and *EMM* afforded by the PCs in their standard implementation. Especially *EMM* is suited for selection of an appropriate EC or appropriate PCs for mapping.

Linkage analysis and segregation distortion tests

Linkage analyses and segregation distortion tests were performed with the software package JoinMap version 2 (JM) (Stam 1993; Stam and van Ooijen 1995), using the appropriate mapping population type; option RI6 for the RI population and F2 for the IF_2 population. Linkage groups were assigned to the

corresponding chromosomes of *Zea mays* L. by inclusion of segregation data of isozymes, RFLPs and SSRs obtained previously on the RI lines (Senior et al. 1997) and/or on the IF_2 lines (Xia et al. 1998). No order was forced during map construction, except for linkage group 1 of the RI population: a fixed order of four markers (*phi056*, *bnl5.62a*, *umc157(chn)* and *umc76*) belonging to bins 1.01 and 1.02 was forced in order to preserve their relative positions as given in Senior et al. (1997) and on the UMC 1998 molecular marker map of maize (Davis et al. 1998). The recombination frequencies were converted to Kosambi centiMorgans (cM) (Kosambi 1944). Maps were drawn using proprietary software (see Figure 3.1).

Distribution of AFLP markers over the genome

Equal representation of genomic regions in the map and genome coverage are a function of the distribution of markers over the linkage maps of chromosomes. In order to get information on the distribution of E/M-, P/M- and ^{m}P/M markers over linkage maps of chromosomes, their distribution has been determined statistically. The Kolmogorov assay was used to test the null-hypothesis:

$$H_0: F(x) = F_0(x),$$

where F(x) represents the observed distribution function of the interval (expressed in cM) between two adjacent markers, either two E/M-, two P/M- or two ^{m}P/M-markers; $F_0(x)$ represents the corresponding distribution function under the null hypothesis (H_0); in this case we hypothesize that marker positions are independent and uniformly distributed over linkage maps of chromosomes. This implies an exponential distribution of inter-marker distances, i.e., $F_0(x) = 1\text{-}e^{-x/u}$, where u is the mean interval length.

Contingent on the rejection of H_0, the one-sided alternative hypothesis H_1: $F(x) > F_0(x)$, for at least one value for x, is accepted. The test statistic D_n is defined as the largest difference between F(x) and $F_0(x)$ ($D_n = \max(F(x)\text{-}F_0(x))$.

The minimal size of the interval in which D_n is measured is determined by the resolution of the mapping population: 0.5 cM and 1.0 cM for the RI and the IF_2 population, respectively.

Since biallelic AFLP markers map to the same locus, they lead to an overestimation of the number of intervals of zero length and hence, to erroneous rejection of the H_0. Therefore, biallelic AFLP markers are represented by only one marker in this analysis.

All calculations were carried out using the Genstat programme (Genstat-5-Committee 1993)

Co-localization of hypothetical centromeres and AFLP marker clusters

Map positions of putative centromeric regions were assigned using RFLP markers, segregating in one or both populations, which are known to map to the centromeres of the maize chromosomes, with the exception of chromosome 8 (Matz et al. 1994). Localization and visualization of clusters was assessed by scanning the linkage groups, using a 5 cM window, for the largest cluster of AFLP markers.

All calculations were carried out using the Genstat programme (Genstat-5-Committee 1993).

Results

Polymorphism rates among B73 and Mo17, and among D32 and D145

The first step required in the assembly of a linkage map, is to screen the two parental lines for polymorphism. The two parental lines B73 and Mo17 of the RI population were screened for polymorphism with 84 E/M, 36 P/M and 36 ^{m}P/M PCs, listed in Table 3.1. This screening was designed to identify those PCs which fulfilled the 'optimizing and minimizing effort' criteria in that they: 1) gave patterns containing 50 to 100 fragments, 2) revealed a high polymorphism between the two parental lines, and 3) did not amplify heavily repeated restriction fragments, visible as very intense bands. Fourty one E/M, 21 P/M and 14 ^{m}P/M PCs

fulfilled these criteria (see Table 3.1). Some E/M PCs were combined in a multiplex reaction approach, resulting in a total of 36 E/M PCs. For the parental lines D32 and D145, the same set of 71 PCs was used to ensure a high polymorhism rate and the maximum overlap of markers in the two maps.

Table 3.1 Overview of the 84 *Eco*RI/*Mse*I (E/M), 36 *Pst*I/*Mse*I (P/M) and 36 ᵐ*Pst*I/*Mse*I (ᵐP/M) primer combinations (PCs) the two parental lines B73 and Mo17 of the RI population were screened with for polymorphism. The 41 E/M and 21 P/M PCs fulfilling the 'optimizing and minimizing effort' criteria are marked by '×'; the 14 ᵐP/M PCs are marked by ⊗ as a subfraction of P/M PCs.

		M47 CAA	M48 CAC	M49 CAG	M50 CAT	M51 CCA	M54 CCT	M55 CGA	M58 CGT	M59 CTA	M60 CTC	M61 CTG	M62 CTT
E32	**AAC**		×	×	×			×			×		
E33	**AAG**	×			×	×	×		×		×	×	×
E35	**ACA**				×	×	×		×	×			
E38	**ACT**	×				×					×		
E39	**AGA**	×	×	×	×	×	×	×		×	×	×	×
E42	**AGT**	×			×						×	×	
E45	**ATG**		×	×							×	×	
P12	**AC**	⊗	⊗	⊗	⊗					⊗		⊗	⊗
P13	**AG**	×	×	⊗	×					×	×	×	×
P18	**CT**		⊗	⊗						⊗	⊗	⊗	⊗

Table 3.2 gives an overview of the total number of AFLP fragments and polymorphic AFLP fragments counted for both pairs of inbred lines. The percentages of polymorphism reflect that the lines D32 and D145 are genetically slightly more divergent than the lines B73 and Mo17. ᵐP/M PCs reveal the highest percentage of polymorphism, while E/M PCs reveal the lowest degree of polymorphism. It is clear from Table 3.2 that E/M and P/M PCs on average yield the highest *M* and *EM*, respectively.

Construction of the B73 × Mo17 and D32 × D145 high-density AFLP linkage maps

Using the 36 E/M and 21 P/M PCs, allelic segregation data were recorded for a first set of 90 RI lines. Using the 14 ᵐP/M PCs, segregation data were recorded for a second set of 90 RI lines, having 75 individuals in common with the first set of RI

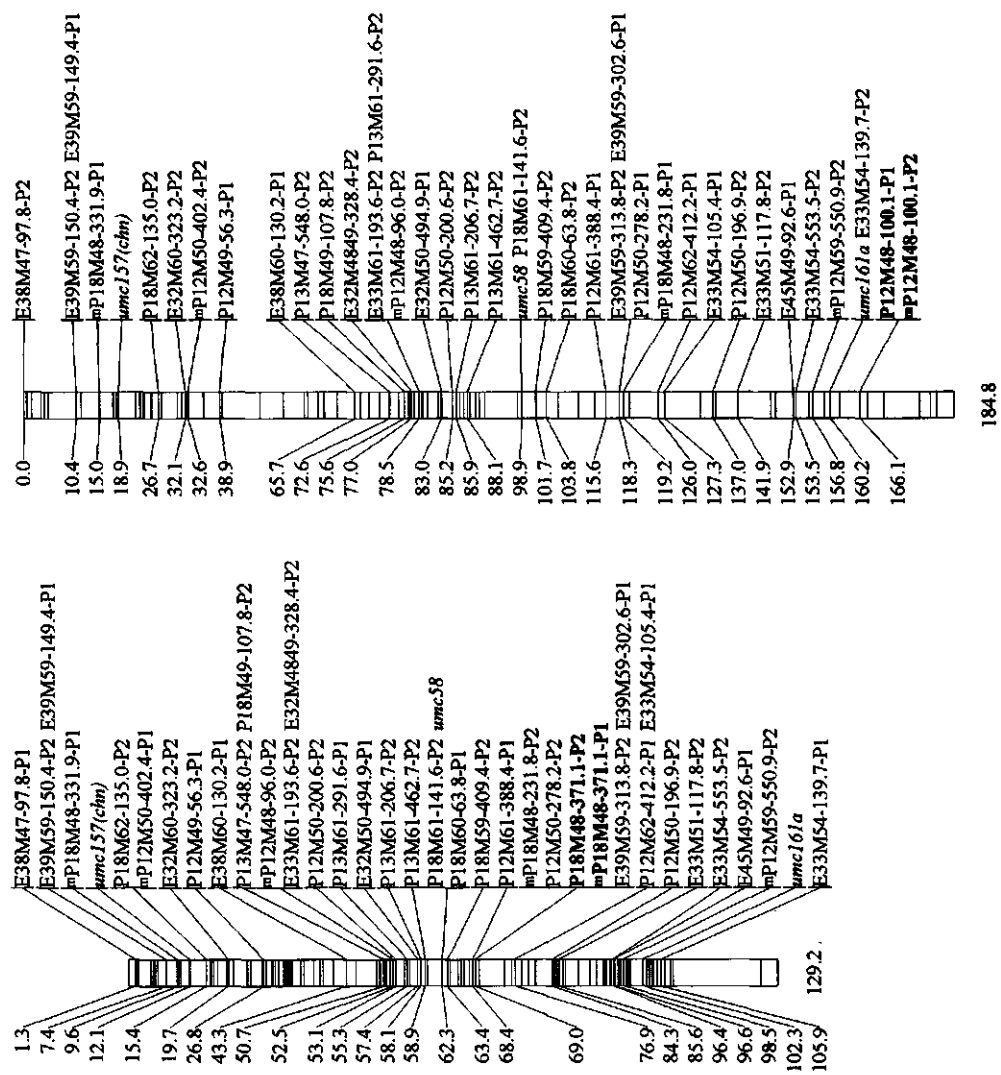
lines. This resulted in a total set of 105 RI lines, genotyped either with E/M, P/M and ^{m}P/M PCs, or with ^{m}P/M PCs only.

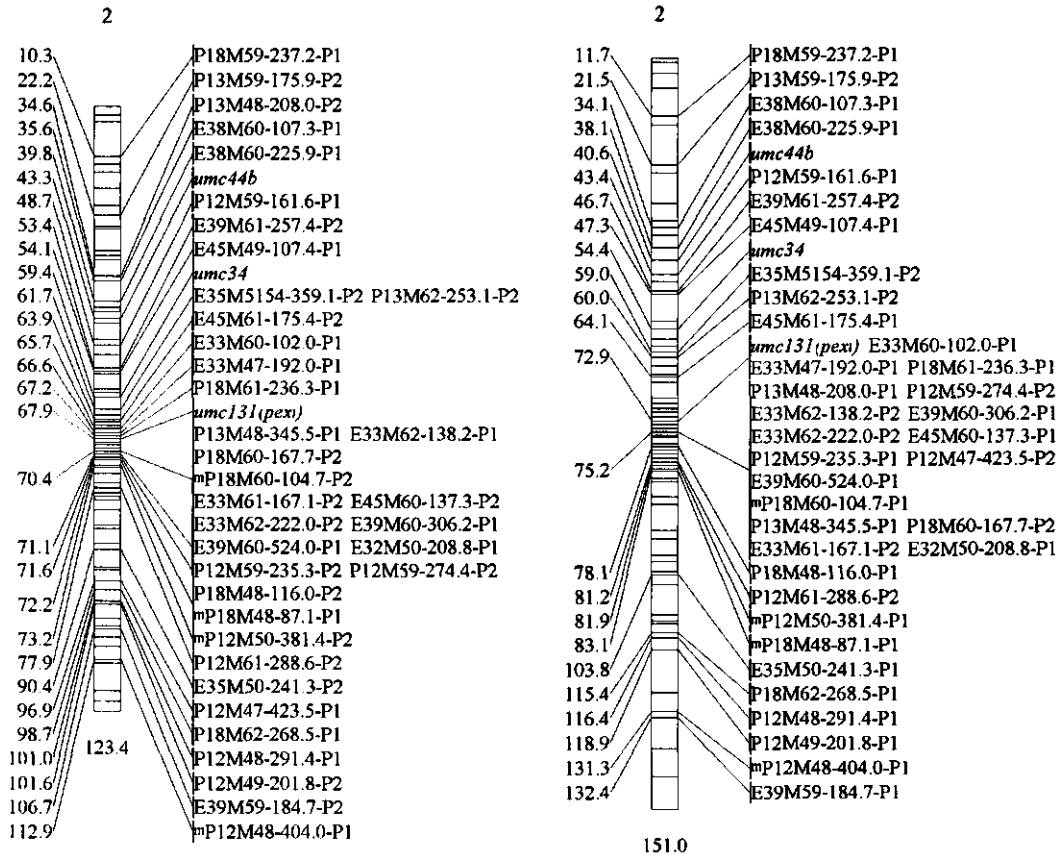
Inclusion of segregation data of 184 isozymes, RFLPs and SSRs obtained previously on the set of 105 RI lines (Senior et al. 1997) totalizes the dataset to 1723 markers. At a LOD=6, these 1723 markers were split into 10 major and 6 minor linkage groups, each containing at least 1 anchor marker to assign to the 10 maize chromosomes. The same set of 71 PCs has been used to genotype a set of 88 IF_2 lines, yielding 1355 AFLP markers. Inclusion of segregation data of 47 RFLPs and SSRs obtained on the same set of 88 IF_2 individuals (Xia et al. 1998) totalizes the dataset to 1402 markers. At a LOD threshold grouping value of 6.0, these 1402 markers were split into 11 linkage groups, 10 major and 1 minor of 6 markers. The 10 major linkage groups were assigned to the 10 maize chromosomes, while the minor linkage group was assigned to chromosome 7 since it contained the core marker *asg8(myb)*.

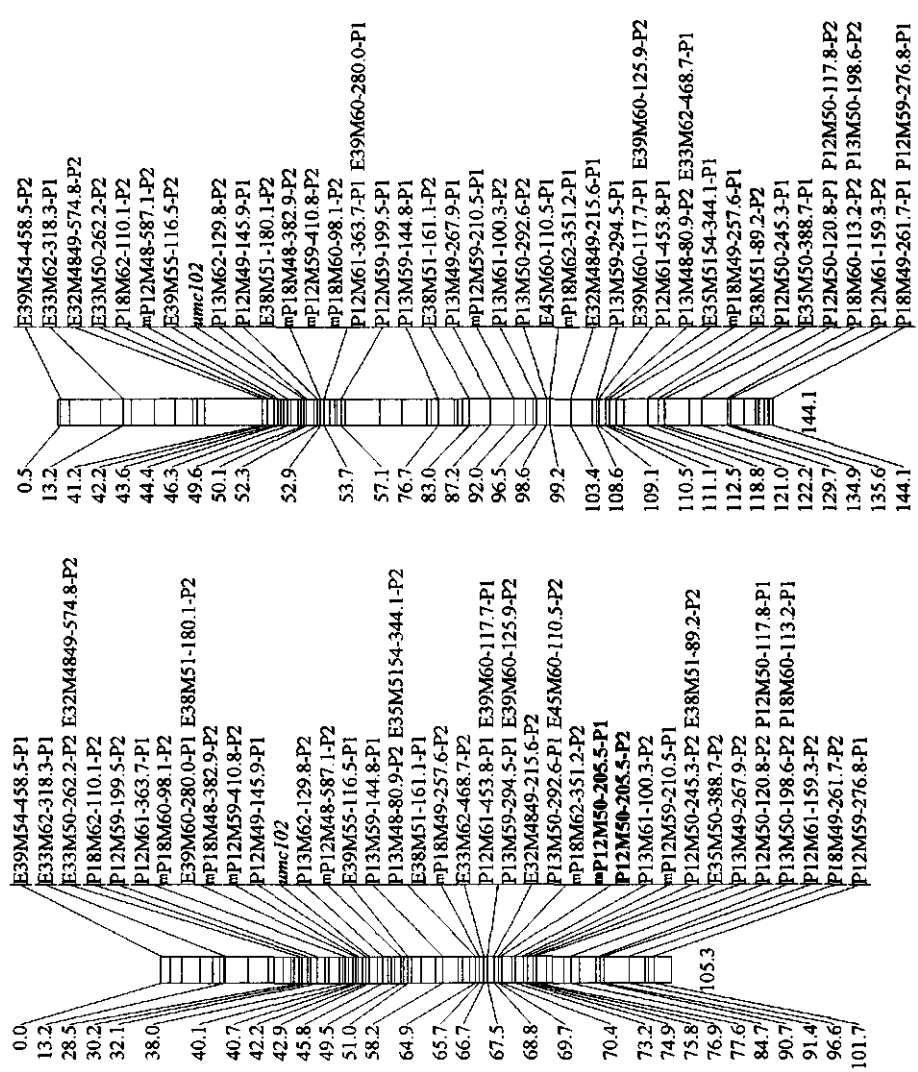
Table 3.2 Overview of the total number of 1) AFLP fragments, 2) polymorphic AFLP fragments, 3) mapped polymorphism and 4) the corresponding Multiplex ratio (*M*), Effective Multiplex ratio (*EM*) and Effective Mapped Multiplex ratio (*EMM*) per enzyme combination (*Eco*RI/*Mse*I (E/M), *Pst*I/*Mse*I (P/M) and ^{m}*Pst*I/*Mse*I (^{m}P/M)) in the Recombinant Inbred (RI) and the Immortalized F_2 (IF_2) mapping population.

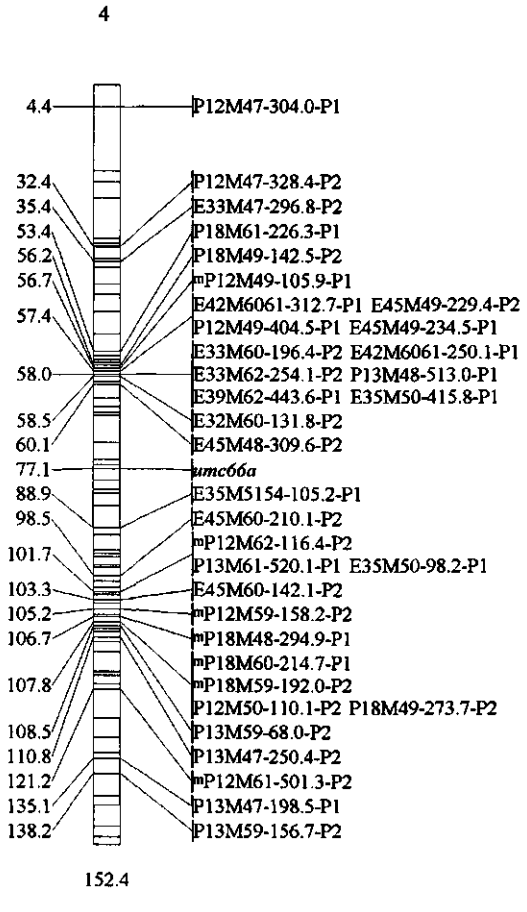
		E/M	P/M	^{m}P/M	Total
	Total	3297	1824	1083	6204
	M	92	87	77	87
RI	Polymorphic	1197(36.3%)	746(40.9%)	483(44.6%)	2425(39.1%)
	EM	33	36	35	34
	Mapped	670(20.3%)	565(31.0%)	304(28.1%)	1539(24.8%)
	EMM	19	27	22	22
	Total	3182	1894	1023	6099
	M	88	90	73	86
IF_2	Polymorphic	1137(35.7%)	871(46.0%)	491(48.0%)	2499(41.0%)
	EM	32	42	35	35
	Mapped	587(18.4%)	550(29.0%)	218(21.3%)	1355(22.2%)
	EMM	16	26	16	19

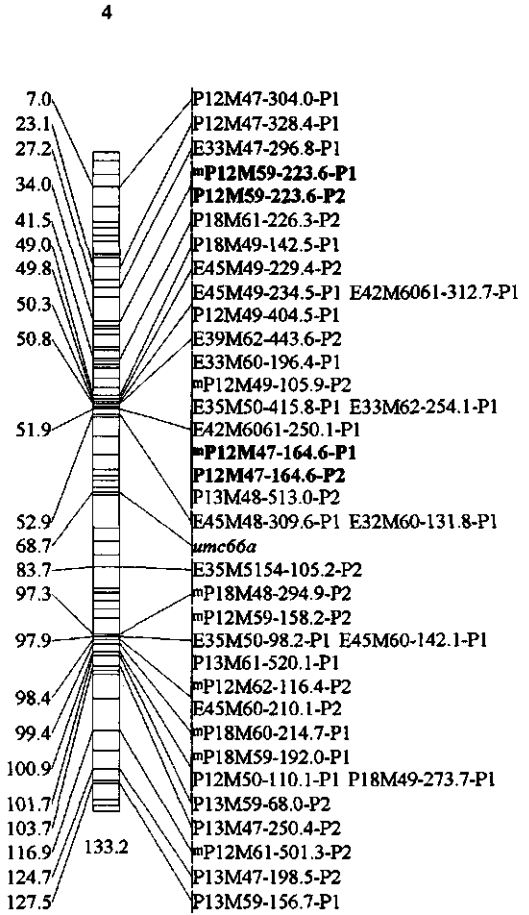
Figure 3.1 Two high-density linkage maps of maize based on 105 recombinant inbred (RI) lines from a B73 × Mo17 cross (left map) and 88 immortalized F_2 (IF_2) lines from a D32 × D145 cross (right map). Distances are given in Kosambi centiMorgans. For simplicity, only common markers, including those that map to different chromosomes in both maps, and pairs of allele specific methylation AFLP markers (^{asm}AFLP markers) are shown. The AFLP markers are named with 1) the code referring to the corresponding primer combination (see Table 3.1) followed by 2) the estimated molecular size of the DNA fragment in nucleotides and 3) a code indicating the parental origin of the fragment (RI population: P1=B73; P2=Mo17. IF_2 population: P1=D32; P2=D145). The pairs of ^{asm}AFLP markers, located on chromosome 1, 3 and 6 and on 1 and 4 in the RI and IF_2 map, respectively, are indicated in bold.

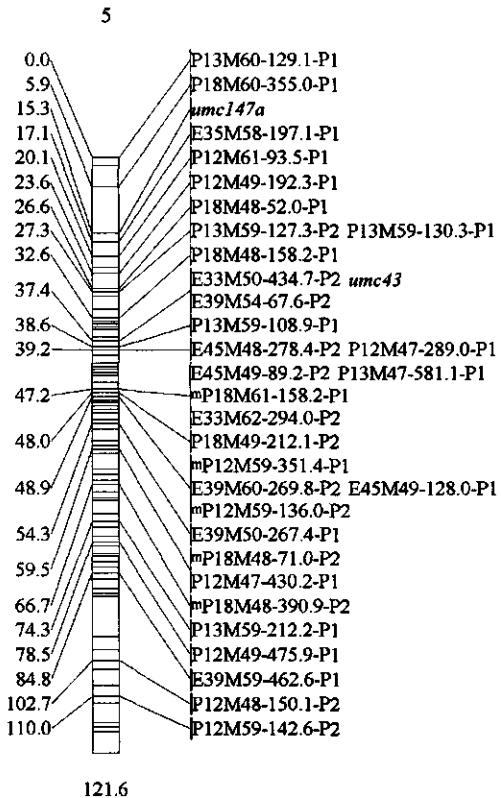

1
1.3 E38M47-97.8-P1
7.4 E39M59-150.4-P2 E39M59-149.4-P1
9.6 mP18M48-331.9-P1
12.1 *umc157(chm)*
15.4 P18M62-135.0-P2
mP12M50-402.4-P1
19.7 E32M60-323.2-P2
26.8 P12M49-56.3-P1
43.3 E38M60-130.2-P1
50.7 P13M47-548.0-P2 P18M49-107.8-P2
mP12M48-96.0-P2
52.5 E33M61-193.6-P2 E32M4849-328.4-P2
53.1 P12M50-200.6-P2
55.3 P13M61-291.6-P1
57.4 E32M50-494.9-P1
58.1 P13M61-206.7-P2
58.9 P13M61-462.7-P2
62.3 P18M61-141.6-P2 *umc58*
63.4 P18M60-63.8-P1
68.4 P18M59-409.4-P2
P12M61-388.4-P1
mP18M48-231.8-P2
69.0 P12M50-278.2-P2
P18M48-371.1-P2
mP18M48-371.1-P1
76.9 E39M59-313.8-P2 E39M59-302.6-P1
84.3 P12M62-412.2-P1 E33M54-105.4-P1
85.6 P12M50-196.9-P2
96.4 E33M51-117.8-P2
96.6 E33M54-553.5-P2
98.5 E45M49-92.6-P1
102.3 mP12M59-550.9-P2
105.9 *umc161a*
129.2 E33M54-139.7-P1
1
0.0 E38M47-97.8-P2
10.4 E39M59-150.4-P2 E39M59-149.4-P1
15.0 mP18M48-331.9-P1
18.9 *umc157(chm)*
26.7 P18M62-135.0-P2
32.1 E32M60-323.2-P2
32.6 mP12M50-402.4-P2
38.9 P12M49-56.3-P1
65.7 E38M60-130.2-P1
72.6 P13M47-548.0-P2
75.6 P18M49-107.8-P2
77.0 E32M4849-328.4-P2
78.5 E33M61-193.6-P2 P13M61-291.6-P2
mP12M48-96.0-P2
83.0 E32M50-494.9-P1
85.2 P12M50-200.6-P2
85.9 P13M61-206.7-P2
88.1 P13M61-462.7-P2
98.9 *umc58* P18M61-141.6-P2
101.7 P18M59-409.4-P2
103.8 P18M60-63.8-P2
115.6 P12M61-388.4-P1
118.3 E39M59-313.8-P2 E39M59-302.6-P1
119.2 P12M50-278.2-P1
mP18M48-231.8-P1
126.0 P12M62-412.2-P1
127.3 E33M54-105.4-P1
137.0 P12M50-196.9-P2
141.9 E33M51-117.8-P2
152.9 E45M49-92.6-P1
153.5 E33M54-553.5-P2
156.8 mP12M59-550.9-P2
160.2 *umc161a* E33M54-139.7-P2
166.1 **P12M48-100.1-P1**
mP12M48-100.1-P2
184.8

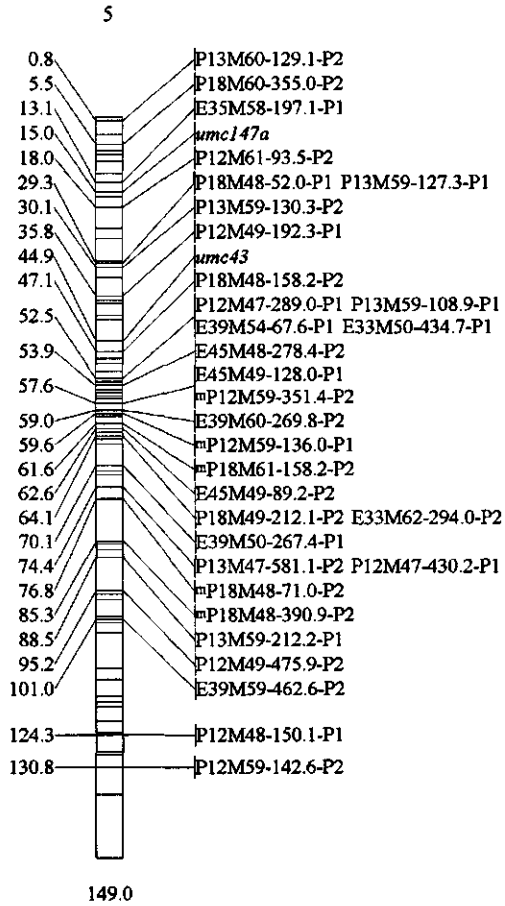

2
10.3
22.2
34.6
35.6
39.8
43.3
48.7
53.4
54.1
59.4
61.7
63.9
65.7
66.6
67.2
67.9
70.4
71.1
71.6
72.2
73.2
77.9
90.4
96.9
98.7
101.0
101.6
106.7
112.9
123.4
P18M59-237.2-P1
P13M59-175.9-P2
P13M48-208.0-P2
E38M60-107.3-P1
E38M60-225.9-P1
umc44b
P12M59-161.6-P1
E39M61-257.4-P2
E45M49-107.4-P1
umc34
E35M5154-359.1-P2 P13M62-253.1-P2
E45M61-175.4-P2
E33M60-102.0-P1
E33M47-192.0-P1
P18M61-236.3-P1
umc131(pexi)
P13M48-345.5-P1 E33M62-138.2-P1
P18M60-167.7-P2
mP18M60-104.7-P2
E33M61-167.1-P2 E45M60-137.3-P2
E33M62-222.0-P2 E39M60-306.2-P1
E39M60-524.0-P1 E32M50-208.8-P1
P12M59-235.3-P2 P12M59-274.4-P2
P18M48-116.0-P2
mP18M48-87.1-P1
mP12M50-381.4-P2
P12M61-288.6-P2
E35M50-241.3-P2
P12M47-423.5-P1
P18M62-268.5-P1
P12M48-291.4-P1
P12M49-201.8-P2
E39M59-184.7-P2
mP12M48-404.0-P1
2
11.7
21.5
34.1
38.1
40.6
43.4
46.7
47.3
54.4
59.0
60.0
64.1
72.9
75.2
78.1
81.2
81.9
83.1
103.8
115.4
116.4
118.9
131.3
132.4
151.0
P18M59-237.2-P1
P13M59-175.9-P2
E38M60-107.3-P1
E38M60-225.9-P1
umc44b
P12M59-161.6-P1
E39M61-257.4-P2
E45M49-107.4-P1
umc34
E35M5154-359.1-P2
P13M62-253.1-P2
E45M61-175.4-P1
umc131(pexi) E33M60-102.0-P1
E33M47-192.0-P1 P18M61-236.3-P1
P13M48-208.0-P1 P12M59-274.4-P2
E33M62-138.2-P2 E39M60-306.2-P1
E33M62-222.0-P2 E45M60-137.3-P1
P12M59-235.3-P1 P12M47-423.5-P2
E39M60-524.0-P1
mP18M60-104.7-P1
P13M48-345.5-P1 P18M60-167.7-P2
E33M61-167.1-P2 E32M50-208.8-P1
P18M48-116.0-P1
P12M61-288.6-P2
mP12M50-381.4-P1
mP18M48-87.1-P1
E35M50-241.3-P1
P18M62-268.5-P1
P12M48-291.4-P1
P12M49-201.8-P1
mP12M48-404.0-P1
E39M59-184.7-P1

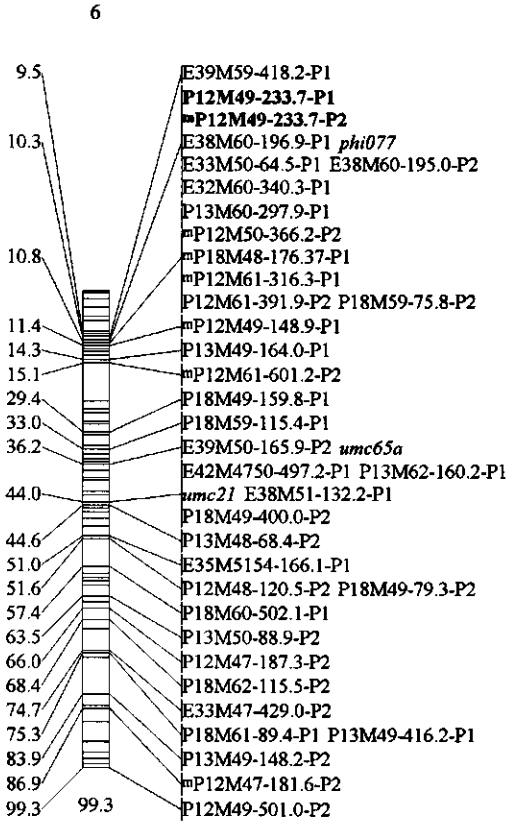

3
0.0
13.2
28.5
30.2
32.1
38.0
40.1
40.7
42.2
42.9
45.8
49.5
51.0
58.2
64.9
65.7
66.7
67.5
68.8
69.7
70.4
73.2
74.9
75.8
76.9
77.6
84.7
90.7
91.4
96.6
101.7
105.3
E39M54-458.5-P1
E33M62-318.3-P1
E33M50-262.2-P2
E32M4849-574.8-P2
P18M62-110.1-P2
P12M59-199.5-P2
P12M61-363.7-P1
mP18M60-98.1-P2
E39M60-280.0-P1
E38M51-180.1-P2
mP18M48-382.9-P2
mP12M59-410.8-P2
P12M49-145.9-P1
umc102
P13M62-129.8-P2
mP12M48-587.1-P2
E39M55-116.5-P1
P13M59-144.8-P1
P13M48-80.9-P2
E35M5154-344.1-P2
E38M51-161.1-P1
mP18M49-257.6-P2
E33M62-468.7-P2
P12M61-453.8-P1
E39M60-117.7-P1
P13M59-294.5-P1
E39M60-125.9-P2
E32M4849-215.6-P2
P13M50-292.6-P1
E45M60-110.5-P2
mP18M62-351.2-P2
mP12M50-205.5-P1
P12M50-205.5-P2
P13M61-100.3-P2
mP12M59-210.5-P1
P12M50-245.3-P2
E38M51-89.2-P2
E35M50-388.7-P2
P13M49-267.9-P2
P12M50-120.8-P2
P12M50-117.8-P1
P13M50-198.6-P2
P18M60-113.2-P1
P12M61-159.3-P2
P18M49-261.7-P2
P12M59-276.8-P1
3
0.5
13.2
41.2
42.2
43.6
44.4
46.3
49.6
50.1
52.3
52.9
53.7
57.1
76.7
83.0
87.2
92.0
96.5
98.6
99.2
103.4
108.6
109.1
110.5
111.1
112.5
118.8
121.0
122.2
129.7
134.9
135.6
144.1
144.1
E39M54-458.5-P2
E33M62-318.3-P1
E32M4849-574.8-P2
E33M50-262.2-P2
P18M62-110.1-P2
mP12M48-587.1-P2
E39M55-116.5-P2
umc102
P13M62-129.8-P2
P12M49-145.9-P1
E38M51-180.1-P2
mP18M48-382.9-P2
mP12M59-410.8-P2
mP18M60-98.1-P2
P12M61-363.7-P1
E39M60-280.0-P1
P12M59-199.5-P1
P13M59-144.8-P1
E38M51-161.1-P2
P13M49-267.9-P1
mP12M59-210.5-P1
P13M61-100.3-P2
P13M50-292.6-P2
E45M60-110.5-P1
mP18M62-351.2-P1
E32M4849-215.6-P1
P13M59-294.5-P1
E39M60-117.7-P1
E39M60-125.9-P2
P12M61-453.8-P1
P13M48-80.9-P2
E33M62-468.7-P1
E35M5154-344.1-P1
mP18M49-257.6-P1
E38M51-89.2-P2
P12M50-245.3-P1
E35M50-388.7-P1
P12M50-120.8-P1
P12M50-117.8-P2
P18M60-113.2-P2
P13M50-198.6-P2
P12M61-159.3-P2
P18M49-261.7-P1
P12M59-276.8-P1

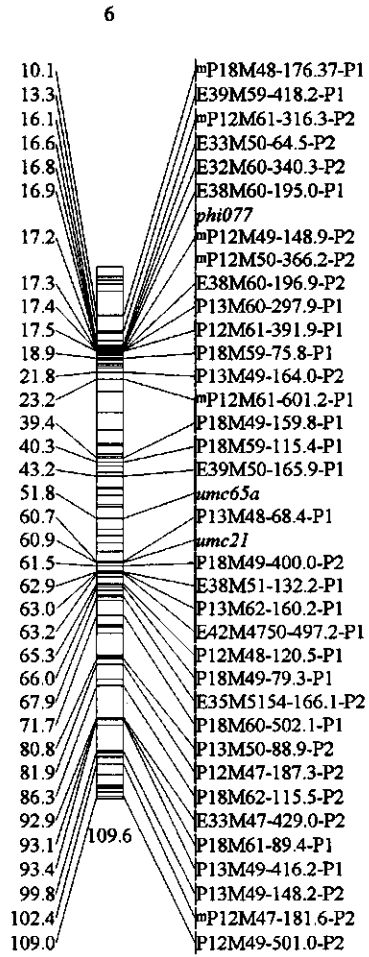

4
4.4 P12M47-304.0-P1
32.4 P12M47-328.4-P2
35.4 E33M47-296.8-P2
53.4 P18M61-226.3-P1
56.2 P18M49-142.5-P2
56.7 mP12M49-105.9-P1
57.4 E42M6061-312.7-P1 E45M49-229.4-P2
P12M49-404.5-P1 E45M49-234.5-P1
E33M60-196.4-P2 E42M6061-250.1-P1
58.0 E33M62-254.1-P2 P13M48-513.0-P1
E39M62-443.6-P1 E35M50-415.8-P1
58.5 E32M60-131.8-P2
60.1 E45M48-309.6-P2
77.1 umc66a
88.9 E35M5154-105.2-P1
98.5 E45M60-210.1-P2
mP12M62-116.4-P2
101.7 P13M61-520.1-P1 E35M50-98.2-P1
103.3 E45M60-142.1-P2
105.2 mP12M59-158.2-P2
106.7 mP18M48-294.9-P1
mP18M60-214.7-P1
107.8 mP18M59-192.0-P2
P12M50-110.1-P2 P18M49-273.7-P2
108.5 P13M59-68.0-P2
110.8 P13M47-250.4-P2
121.2 mP12M61-501.3-P2
135.1 P13M47-198.5-P1
138.2 P13M59-156.7-P2
152.4


4
7.0 P12M47-304.0-P1
23.1 P12M47-328.4-P1
27.2 E33M47-296.8-P1
34.0 mP12M59-223.6-P1
P12M59-223.6-P2
41.5 P18M61-226.3-P2
49.0 P18M49-142.5-P1
49.8 E45M49-229.4-P2
50.3 E45M49-234.5-P1 E42M6061-312.7-P1
P12M49-404.5-P1
50.8 E39M62-443.6-P2
E33M60-196.4-P1
mP12M49-105.9-P2
51.9 E35M50-415.8-P1 E33M62-254.1-P1
E42M6061-250.1-P1
mP12M47-164.6-P1
P12M47-164.6-P2
52.9 P13M48-513.0-P2
68.7 E45M48-309.6-P1 E32M60-131.8-P1
umc66a
83.7 E35M5154-105.2-P2
97.3 mP18M48-294.9-P2
mP12M59-158.2-P2
97.9 E35M50-98.2-P1 E45M60-142.1-P1
P13M61-520.1-P1
98.4 mP12M62-116.4-P2
E45M60-210.1-P2
99.4 mP18M60-214.7-P1
100.9 mP18M59-192.0-P1
P12M50-110.1-P1 P18M49-273.7-P1
101.7 P13M59-68.0-P2
103.7 P13M47-250.4-P2
116.9 mP12M61-501.3-P2
124.7 P13M47-198.5-P2
127.5 P13M59-156.7-P1
133.2

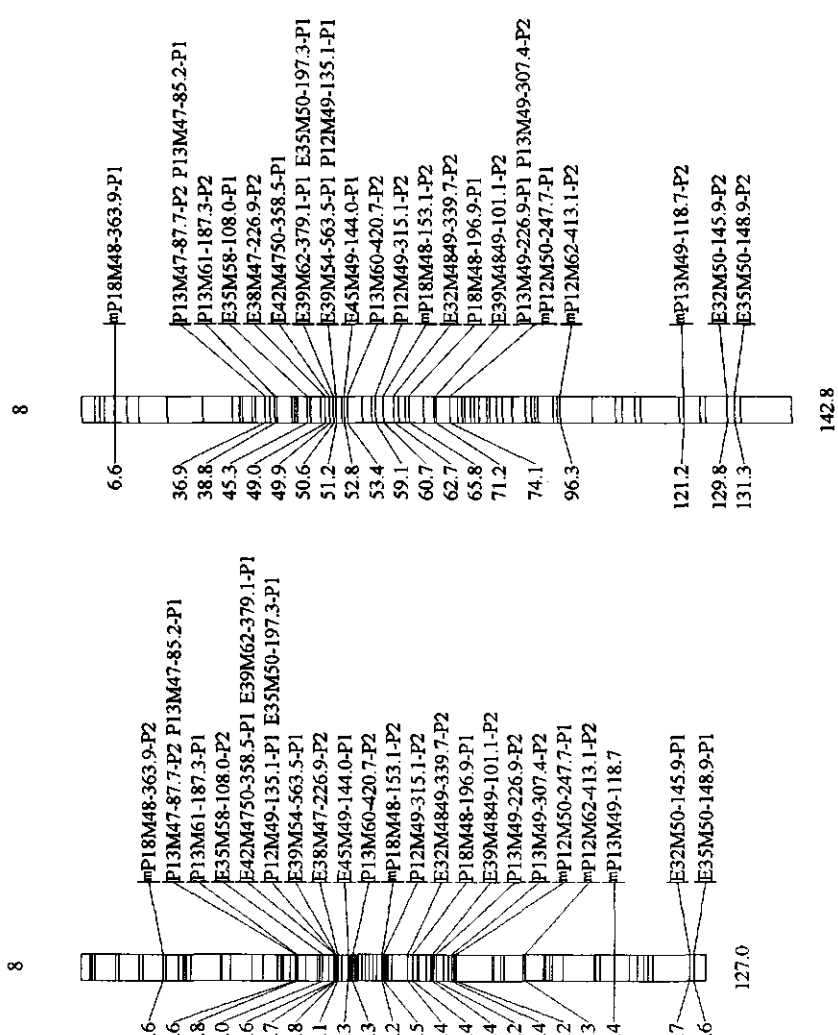

5
0.0 P13M60-129.1-P1
5.9 P18M60-355.0-P1
15.3 umc147a
17.1 E35M58-197.1-P1
20.1 P12M61-93.5-P1
23.6 P12M49-192.3-P1
26.6 P18M48-52.0-P1
27.3 P13M59-127.3-P2 P13M59-130.3-P1
32.6 P18M48-158.2-P1
37.4 E33M50-434.7-P2 umc43
E39M54-67.6-P2
38.6 P13M59-108.9-P1
39.2 E45M48-278.4-P2 P12M47-289.0-P1
E45M49-89.2-P2 P13M47-581.1-P1
47.2 mP18M61-158.2-P1
E33M62-294.0-P2
48.0 P18M49-212.1-P2
mP12M59-351.4-P1
48.9 E39M60-269.8-P2 E45M49-128.0-P1
mP12M59-136.0-P2
54.3 E39M50-267.4-P1
59.5 mP18M48-71.0-P2
P12M47-430.2-P1
66.7 mP18M48-390.9-P2
74.3 P13M59-212.2-P1
78.5 P12M49-475.9-P1
84.8 E39M59-462.6-P1
102.7 P12M48-150.1-P2
110.0 P12M59-142.6-P2
121.6

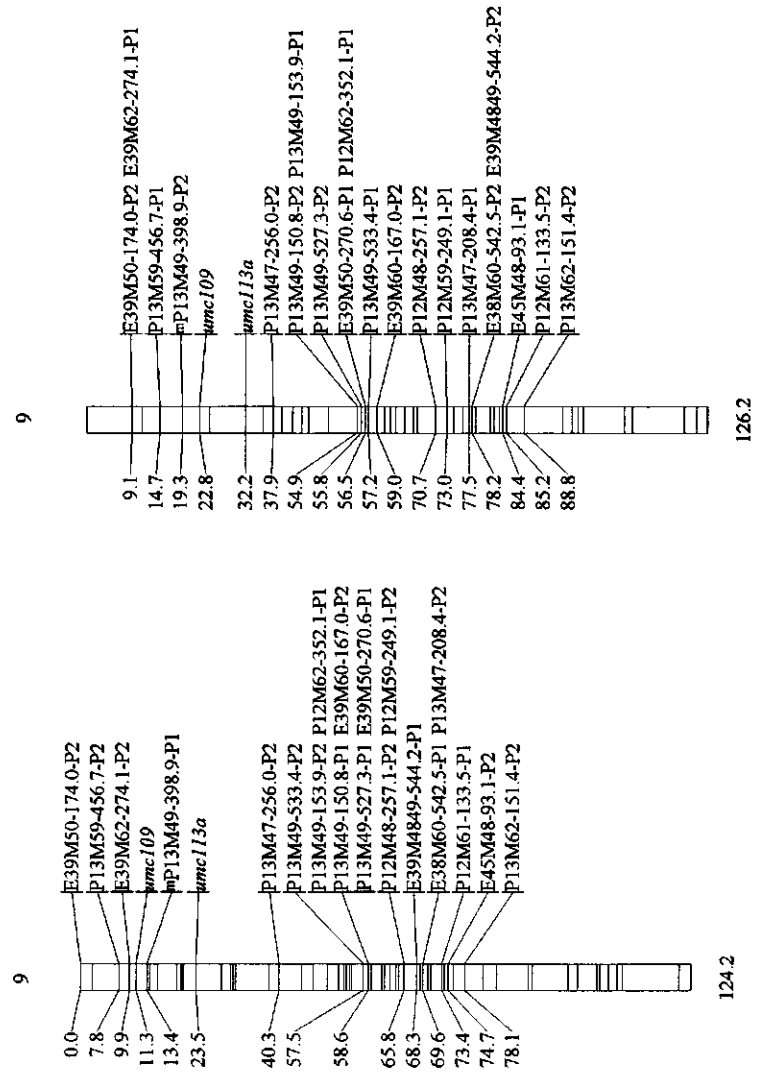

5
0.8 P13M60-129.1-P2
5.5 P18M60-355.0-P2
13.1 E35M58-197.1-P1
15.0 umc147a
18.0 P12M61-93.5-P2
29.3 P18M48-52.0-P1 P13M59-127.3-P1
30.1 P13M59-130.3-P2
35.8 P12M49-192.3-P1
44.9 umc43
47.1 P18M48-158.2-P2
52.5 P12M47-289.0-P1 P13M59-108.9-P1
E39M54-67.6-P1 E33M50-434.7-P1
53.9 E45M48-278.4-P2
57.6 E45M49-128.0-P1
mP12M59-351.4-P2
59.0 E39M60-269.8-P2
59.6 mP12M59-136.0-P1
61.6 mP18M61-158.2-P2
62.6 E45M49-89.2-P2
64.1 P18M49-212.1-P2 E33M62-294.0-P2
70.1 E39M50-267.4-P1
74.4 P13M47-581.1-P2 P12M47-430.2-P1
76.8 mP18M48-71.0-P2
85.3 mP18M48-390.9-P2
88.5 P13M59-212.2-P1
95.2 P12M49-475.9-P2
101.0 E39M59-462.6-P2
124.3 P12M48-150.1-P1
130.8 P12M59-142.6-P2
149.0

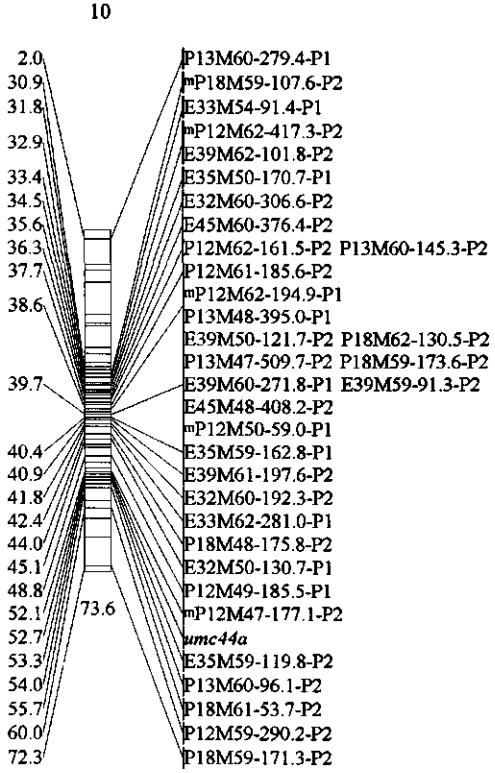

6
9.5
10.3
10.8
11.4
14.3
15.1
29.4
33.0
36.2
44.0
44.6
51.0
51.6
57.4
63.5
66.0
68.4
74.7
75.3
83.9
86.9
99.3
99.3
E39M59-418.2-P1
P12M49-233.7-P1
mP12M49-233.7-P2
E38M60-196.9-P1 phi077
E33M50-64.5-P1 E38M60-195.0-P2
E32M60-340.3-P1
P13M60-297.9-P1
mP12M50-366.2-P2
mP18M48-176.37-P1
mP12M61-316.3-P1
P12M61-391.9-P2 P18M59-75.8-P2
mP12M49-148.9-P1
P13M49-164.0-P1
mP12M61-601.2-P2
P18M49-159.8-P1
P18M59-115.4-P1
E39M50-165.9-P2 umc65a
E42M4750-497.2-P1 P13M62-160.2-P1
umc21 E38M51-132.2-P1
P18M49-400.0-P2
P13M48-68.4-P2
E35M5154-166.1-P1
P12M48-120.5-P2 P18M49-79.3-P2
P18M60-502.1-P1
P13M50-88.9-P2
P12M47-187.3-P2
P18M62-115.5-P2
E33M47-429.0-P2
P18M61-89.4-P1 P13M49-416.2-P1
P13M49-148.2-P2
mP12M47-181.6-P2
P12M49-501.0-P2

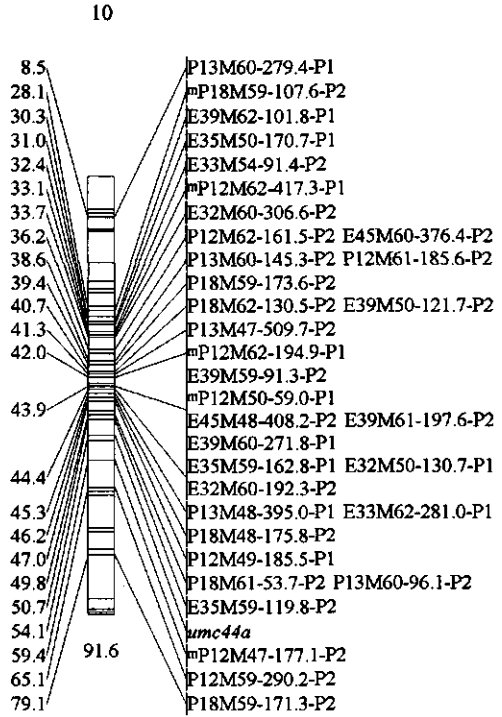

6
10.1
13.3
16.1
16.6
16.8
16.9
17.2
17.3
17.4
17.5
18.9
21.8
23.2
39.4
40.3
43.2
51.8
60.7
60.9
61.5
62.9
63.0
63.2
65.3
66.0
67.9
71.7
80.8
81.9
86.3
92.9
93.1
93.4
99.8
102.4
109.0
109.6
mP18M48-176.37-P1
E39M59-418.2-P1
mP12M61-316.3-P2
E33M50-64.5-P2
E32M60-340.3-P2
E38M60-195.0-P1
phi077
mP12M49-148.9-P2
mP12M50-366.2-P2
E38M60-196.9-P2
P13M60-297.9-P1
P12M61-391.9-P1
P18M59-75.8-P1
P13M49-164.0-P2
mP12M61-601.2-P1
P18M49-159.8-P1
P18M59-115.4-P1
E39M50-165.9-P1
umc65a
P13M48-68.4-P1
umc21
P18M49-400.0-P2
E38M51-132.2-P1
P13M62-160.2-P1
E42M4750-497.2-P1
P12M48-120.5-P1
P18M49-79.3-P1
E35M5154-166.1-P2
P18M60-502.1-P1
P13M50-88.9-P2
P12M47-187.3-P2
P18M62-115.5-P2
E33M47-429.0-P2
P18M61-89.4-P1
P13M49-416.2-P1
P13M49-148.2-P2
mP12M47-181.6-P2
P12M49-501.0-P2

7
24.0 asg34a(msa)
25.9 E33M60-298.6-P2
29.0 P13M59-469.9-P2
32.2 E33M61-188.5-P2
32.9 mP12M50-216.8-P1
33.0 E45M48-272.3-P1
33.1 P18M62-360.7-P1
33.6 E33M47-211.9-P1 P13M59-120.2-P2
33.8 E33M47-219.0-P2
33.9 E35M50-247.1-P1
34.0 mP12M47-137.8-P2
34.2 P12M62-310.8-P1 E42M4750-517.4-P1
34.4 E39M62-117.0-P2
35.1 E35M49-223.3-P1
35.4 E33M61-396.4-P1
36.6 P13M61-512.1-P1
47.3 E33M61-131.3-P2 P18M49-200.6-P2
47.7 E35M49-321.5-P2
48.3 P18M48-79.0-P1
51.3 P12M61-200.0-P2
53.4 P12M47-244.4-P2
54.1 P13M50-79.0-P1
56.0 P12M59-115.2-P2
60.4 P13M48-455.0-P2
63.1 E39M50-463.7-P1
66.0 umc254
70.9 mP12M50-573.3-P2
E39M61-108.7-P1
119.9 umc168
122.1
7
7.6 P13M59-118.5-P2
49.3 asg34a(msa)
54.7 E33M60-298.6-P2
55.7 P13M59-469.9-P2
58.9 E45M48-272.3-P2
59.2 P18M62-360.7-P1
59.7 E33M61-188.5-P2
61.5 P13M61-512.1-P2
61.6 mP12M50-216.8-P2
61.9 P12M62-310.8-P2
62.1 E42M4750-517.4-P2
62.2 E33M47-211.9-P2 E39M62-117.0-P1
62.3 P13M59-120.2-P1
62.5 E33M47-219.0-P1
62.7 E35M49-223.3-P2
66.8 E35M50-247.1-P2
67.6 mP12M47-137.8-P2
70.7 E33M61-396.4-P1
70.9 E33M61-131.3-P2
93.1 P18M49-200.6-P2
94.0 P18M48-79.0-P1
97.7 E35M49-321.5-P2
100.9 P12M61-200.0-P2
101.3 umc254
102.2 mP12M50-573.3-P2
103.8 E39M61-108.7-P1
104.1 E39M50-463.7-P1
104.5 P13M48-455.0-P2
104.7 P12M47-244.4-P2
105.8 P12M59-115.2-P2
143.3 P13M50-79.0-P1
143.3 umc168


8
mP18M48-363.9-P1
P13M47-87.7-P2 P13M47-85.2-P1
P13M61-187.3-P2
E35M58-108.0-P1
E38M47-226.9-P2
E42M4750-358.5-P1
E39M62-379.1-P1 E35M50-197.3-P1
E39M54-563.5-P1 P12M49-135.1-P1
E45M49-144.0-P1
P13M60-420.7-P2
P12M49-315.1-P2
mP18M48-153.1-P2
E32M4849-339.7-P2
P18M48-196.9-P1
E39M4849-101.1-P2
P13M49-226.9-P1 P13M49-307.4-P2
mP12M50-247.7-P1
mP12M62-413.1-P2
mP13M49-118.7-P2
E32M50-145.9-P2
E35M50-148.9-P2
6.6
36.9
38.8
45.3
49.0
49.9
50.6
51.2
52.8
53.4
59.1
60.7
62.7
65.8
71.2
74.1
96.3
121.2
129.8
131.3
142.8
8
mP18M48-363.9-P2
P13M47-87.7-P2 P13M47-85.2-P1
P13M61-187.3-P1
E35M58-108.0-P2
E42M4750-358.5-P1 E39M62-379.1-P1
P12M49-135.1-P1 E35M50-197.3-P1
E39M54-563.5-P1
E38M47-226.9-P2
E45M49-144.0-P1
P13M60-420.7-P2
mP18M48-153.1-P2
P12M49-315.1-P2
E32M4849-339.7-P2
P18M48-196.9-P1
E39M4849-101.1-P2
P13M49-226.9-P2
P13M49-307.4-P2
mP12M50-247.7-P1
mP12M62-413.1-P2
mP13M49-118.7
E32M50-145.9-P1
E35M50-148.9-P1
16.6
43.6
43.8
49.0
51.6
51.7
51.8
52.1
54.3
55.3
61.2
61.5
66.4
67.4
71.4
72.2
75.4
76.2
90.3
108.4
123.7
124.6
127.0


9
0.0 E39M50-174.0-P2
7.8 P13M59-456.7-P2
9.9 E39M62-274.1-P2
11.3 umc109
13.4 mP13M49-398.9-P1
23.5 umc113a
40.3 P13M47-256.0-P2
57.5 P13M49-533.4-P2
P13M49-153.9-P2 P12M62-352.1-P1
58.6 P13M49-150.8-P1 E39M60-167.0-P2
P13M49-527.3-P1 E39M50-270.6-P1
65.8 P12M48-257.1-P2 P12M59-249.1-P2
68.3 E39M4849-544.2-P1
69.6 E38M60-542.5-P1 P13M47-208.4-P2
73.4 P12M61-133.5-P1
74.7 E45M48-93.1-P2
78.1 P13M62-151.4-P2
124.2
9
9.1 E39M50-174.0-P2 E39M62-274.1-P1
14.7 P13M59-456.7-P1
19.3 mP13M49-398.9-P2
22.8 umc109
32.2 umc113a
37.9 P13M47-256.0-P2
54.9 P13M49-150.8-P2 P13M49-153.9-P1
55.8 P13M49-527.3-P2
56.5 E39M50-270.6-P1 P12M62-352.1-P1
57.2 P13M49-533.4-P1
59.0 E39M60-167.0-P2
70.7 P12M48-257.1-P2
73.0 P12M59-249.1-P1
77.5 P13M47-208.4-P1
78.2 E38M60-542.5-P2 E39M4849-544.2-P2
84.4 E45M48-93.1-P1
85.2 P12M61-133.5-P2
88.8 P13M62-151.4-P2
126.2


10
2.0
30.9
31.8
32.9
33.4
34.5
35.6
36.3
37.7
38.6
39.7
40.4
40.9
41.8
42.4
44.0
45.1
48.8
52.1
52.7
53.3
54.0
55.7
60.0
72.3
73.6
P13M60-279.4-P1
mP18M59-107.6-P2
E33M54-91.4-P1
mP12M62-417.3-P2
E39M62-101.8-P2
E35M50-170.7-P1
E32M60-306.6-P2
E45M60-376.4-P2
P12M62-161.5-P2 P13M60-145.3-P2
P12M61-185.6-P2
mP12M62-194.9-P1
P13M48-395.0-P1
E39M50-121.7-P2 P18M62-130.5-P2
P13M47-509.7-P2 P18M59-173.6-P2
E39M60-271.8-P1 E39M59-91.3-P2
E45M48-408.2-P2
mP12M50-59.0-P1
E35M59-162.8-P1
E39M61-197.6-P2
E32M60-192.3-P2
E33M62-281.0-P1
P18M48-175.8-P2
E32M50-130.7-P1
P12M49-185.5-P1
mP12M47-177.1-P2
umc44a
E35M59-119.8-P2
P13M60-96.1-P2
P18M61-53.7-P2
P12M59-290.2-P2
P18M59-171.3-P2


10
8.5
28.1
30.3
31.0
32.4
33.1
33.7
36.2
38.6
39.4
40.7
41.3
42.0
43.9
44.4
45.3
46.2
47.0
49.8
50.7
54.1
59.4
65.1
79.1
91.6
P13M60-279.4-P1
mP18M59-107.6-P2
E39M62-101.8-P1
E35M50-170.7-P1
E33M54-91.4-P2
mP12M62-417.3-P1
E32M60-306.6-P2
P12M62-161.5-P2 E45M60-376.4-P2
P13M60-145.3-P2 P12M61-185.6-P2
P18M59-173.6-P2
P18M62-130.5-P2 E39M50-121.7-P2
P13M47-509.7-P2
mP12M62-194.9-P1
E39M59-91.3-P2
mP12M50-59.0-P1
E45M48-408.2-P2 E39M61-197.6-P2
E39M60-271.8-P1
E35M59-162.8-P1 E32M50-130.7-P1
E32M60-192.3-P2
P13M48-395.0-P1 E33M62-281.0-P1
P18M48-175.8-P2
P12M49-185.5-P1
P18M61-53.7-P2 P13M60-96.1-P2
E35M59-119.8-P2
umc44a
mP12M47-177.1-P2
P12M59-290.2-P2
P18M59-171.3-P2

The percentage of mapped polymorphic fragments equals 24.8% and 22.2% for the RI and IF_2 linkage maps, respectively (Table 3.2). This means that 63.5% and 54.2% of the clearly visable polymorphic fragments of the RI and and IF_2 fingerprints, respectively, were of sufficient quality to allow quantitative scoring and reliable mapping.

It is clear from Table 3.2 that by chosing P/M PCs, with a mean *EMM* of 27, the mapping efforts are minimized.

Map consistency

Although the number of mapped polymorphic AFLP fragments in the IF_2 map was lower than for the RI linkage map (Table 3.2), the IF_2 map spans 1376 cM and the length of most of its linkage groups are systematically longer than in the RI map (1178 cM) (Table 3.3 and Figure 3.1), and is more comparable with the lengths reported for other maps. For both maps, the position of RFLP, SSR and isozyme markers on this map were consistent with previously published *Zea mays* L. maps.

Distorted segregation ratios

Segregation ratios of the two homozygous classes at each marker mapped in the RI map were tested for the 1:1 expected proportion at the 5% level of significance. Chromosome 3 showed distorted segregation over almost its entire length with an excess of B73 alleles (Figure 3.2), while on chromosomes 1, 2, 5 and 10 only minor regions showed distorted segregation. When the significance level was raised to 1%, chromosome 3 still showed distorted segregation over its total length. The largest distortions at chromosome 3 reached the 3:1 magnitude, while the remainder showed only a 2:1 ratio.

Segregation ratios of the three genotype classes of each marker mapped in the IF_2 map were tested for the 1:2:1 expected proportion at the 5% level and the 1% level of significance. Again chromosome 3 showed distorted segregation over a major part of its total length. The distortions for chromosome 3 ranged from magnitude 1:4:1 to 1:5:1

Table 3.3 Comparison of the genetic length and numbers of *Eco*RI/*Mse*I (E/M), *Pst*I/*Mse*I (P/M), ^{m}*Pst*I/*Mse*I (^{m}P/M) and non-AFLP markers (RFLPs, isozymes and SSRs) mapped per linkage group, of the Recombinant Inbred (RI) and the Immortalized F_2 (IF_2) mapping population.

RI map

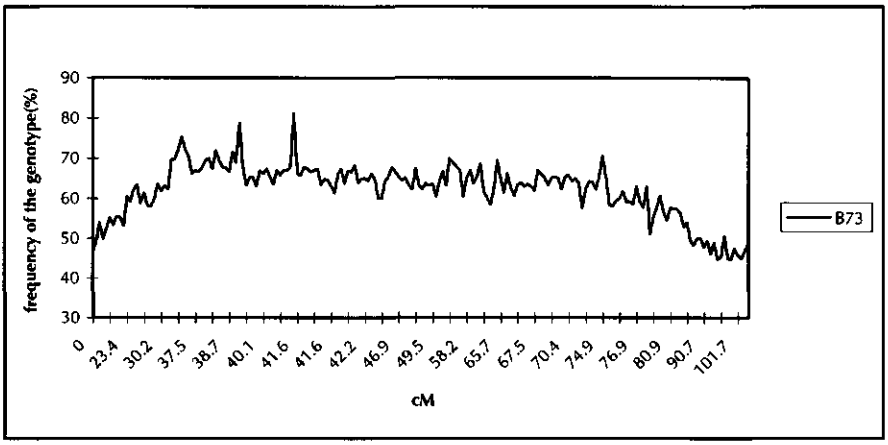
chrom	*cM*	*E/M*	*P/M*	*^{m}P/M*	*non-AFLP*	*# markers*
1	129.2	104	87	57	25	273
2	123.4	67	66	32	13	178
3	105.3	77	64	38	13	192
4	152.4	71	55	33	24	183
5	121.6	57	64	31	29	181
6	99.3	67	54	34	17	172
7	122.1	50	43	21	12	126
8	127.0	58	47	22	17	144
9	124.2	72	41	15	17	145
10	73.6	47	44	21	17	129
total	**1178.1**	**670**	**565**	**304**	**184**	**1723**

IF_2 map

chrom	*cM*	*E/M*	*P/M*	*^{m}P/M*	*non-AFLP*	*# markers*
1	184.8	65	68	31	6	170
2	151.0	76	70	27	9	182
3	144.1	74	64	24	6	168
4	133.2	75	59	33	1	168
5	149.0	55	75	18	8	156
6	109.6	44	50	15	6	115
7	143.3	61	34	11	4	117
8	142.8	55	57	16	2	130
9	126.2	38	33	13	3	87
10	91.6	44	40	23	2	109
total	**1375.6**	**587**	**550**	**218**	**47**	**1402**

($P<0.01$) for a smaller region (40-62 cM), and from magnitude 3:4:1 to 6:8:1 ($P<0.01$) (excess of D32 alleles) for a larger genomic region (79-117 cM), respectively. Minor regions showing distorted segregation ($P<0.05$) were found in only few cases for other chromosomes (1, 2, 5 and 6) and were different from those on the identical chromosomes in the RI map.

Figure 3.2 Distorted segregation on chromosome 3 in the Recombinant Inbred (RI) linkage map.



Allelism of AFLP fragments and epi-alleles

Due to occasional length polymorhism, some of the AFLP markers could be used as co-dominant markers. These allelic AFLP fragments have to meet the two following criteria: i) they originate from a different parent and amplify with the same PC; ii) they map to the same locus (complementary segregation). Putative allelic AFLP fragments were observed in both AFLP linkage maps. The majority of the biallelic markers differ in size by only a few bp (1- 20 bp), probably the result from a small insertion/deletion, while a few pairs of markers were of a large size difference (100-487 bp), reflecting neighbouring restriction sites. In terms of biallelic pairs of markers, the P/M EC predominates: 48 (17.1%) and 32 (11.7%) pairs of mapped P/M markers, in contrast with only 36 (10.9%) and 23 (7.8%) pairs of mapped E/M markers and 11 (7.3%) and 10 (9.2%) pairs of mapped ^{m}P/M markers for the RI and IF_2 maps, respectively.

Allele-specific methylation (methylation polymorphism) results in ^{asm}AFLP markers. Since DNA methylation is the only source of allelic difference between epi-alleles, some of the scored P/M and ^{m}P/M fragments might be epiallelic. To identify a AFLP and a ^{m}AFLP marker as a possible ^{asm}AFLP marker pair, the following criteria are applied: 1) both AFLP fragments have exactly the same size; 2) both AFLP

fragments are derived from different parents, with the same complementary PCs; and 3) both AFLP fragments map to the same locus (complementary segregation). Only three pairs of epi-alleles were identified in both the RI and IF_2 map (Figure 3.1). This means that only 1% of a total of 673 and 595 mapped $^{(m)}$P/M markers (obtained with the 14 complementary P/M PCs), respectively, showed allelic variaton in CpNpG methylation.

Co-linearity between the RI and the IF_2 high-density AFLP linkage maps

The two maps had 353 AFLP markers of identical size and amplified by the same PC, representing 23% and 26% of the AFLP markers mapped in RI and the IF_2 linkage map, respectively. For 327 out of these 353 AFLP markers, linkage maps of chromosomes were moderately (chromosomes 3, 4, 5, 6, 7 and 10) to highly (chromosomes 1, 2, 8, and 9) co-linear (Figure 3.1). Rearrangements occur in map order of some markers, predominantly those residing in or flanking marker-dense regions. Considering the RI map in Figure 3.1, major rearrangements concerned chromosome 3 (38.0-51.0 cM; 64.9-77.6 cM), chromosome 4 (98.5-106.7 cM), chromosome 5 (47.2-48.9 cM), chromosome 6 (9.5-10.8 cM) and chromosome 7 (33.6-35.4 cM; 51.3-70.9 cM). Alternative marker orders, involving inversion of two adjacent markers or segments of more than two markers, without any increase in total chromosomal map length, were evaluated. The most likely map location of a marker was given as a mean rank order and its variance. The variance of the rank order of markers residing in marker-dense regions were among the highest (data not shown), indicating that different map orders in these specific chromosomal regions can be produced with the same data set.

A total of 26 (7.4%) of the AFLP markers maps to two different chromosomes. To elucidate the nature of this behaviour, thirteen pairs of these 'common' markers from the two populations were chosen (estimated molecular size > 100 bases) for comparison. For three pairs of markers, the SeqmanII™ module of Lasergene (DNASTAR, INC. Madison, WC, USA) could assemble a contig for each set, indicative for sequence identity. Although the distal sequences (restriction sites + selective nucleotides) are similar for the remaining ten pairs of markers, they could

not be build in contigs by SeqmanII™, suggesting a low homology between the pairs of sequences.

Distribution of the AFLP markers over the maize genome

Distribution of the E/M, P/M and "P/M markers was determined using the Kolmogorov-assay (Table 3.4). E/M markers are not uniformly distributed on any of the 10 linkage groups ($P<0.05$) in either map. In contrast, P/M markers are randomly distributed on 6 out of the 10 linkage groups of the RI map ($P<0.05$), and on 8 out of the 10 linkage groups of the IF_2 map ($P<0.05$). The distribution of "P/M markers is intermediate, with a tendency to non-uniform distribution: "P/M markers are not uniformly distributed ($P<0.05$) on 9 and 6 out of the 10 linkage groups of the RI map and the IF_2 map, respectively.

Scanning for the largest cluster of AFLP markers revealed that for some linkage groups of the RI linkage map the largest cluster co-localized well with RFLP markers residing on the hypothetical centromeric region. The IF_2 map did not contain sufficient centromeric RFLPs to perform an identical analysis as described above. Nevertheless, clustering of AFLP markers on the IF_2 and RI map co-localized well. Taken together the results suggest that indeed large clusters of AFLP markers occur in the regions of the hypothetical centromeres of the maize chromosomes.

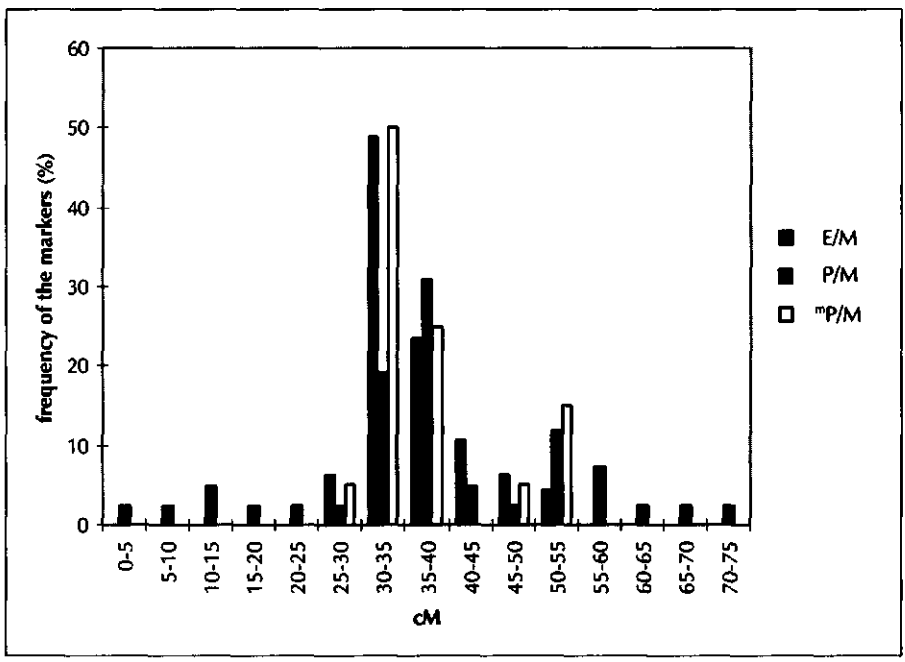
On the RI map, 44% of the E/M and 43% of the "P/M markers reside within the 5 cM windows co-localizing with the putative centromeres, whilst only 21% of the P/M markers reside on these clusters. Figure 3.3 shows the distribution of E/M, P/M and "P/M markers over the total length of linkage group 10 of the RI map and illustrates the strong clustering of E/M and "P/M markers in the centromeric region. The same pattern of distribution of E/M, P/M and "P/M markers is found for the 9 other chromosomes of the RI map, although to a lesser extent.

Table 3.4 Statistical determination of the distribution of the *Eco*RI/*Mse*I (E/M), *Pst*I/*Mse*I (P/M) and *m*Pst*I*/*Mse*I (*m*P/M) markers over the Recombinant Inbred (RI) and the Immortalized F_2 (IF_2) linkage maps of chromosomes. The Kolmogorov assay was used to test whether marker positions are independent and uniformly distributed over linkage maps of chromosomes.

		RI map		**IF_2 map**	
		number of		*number of*	
chrom	*EC*	*intervals*	D_n	*intervals*	D_n
1	E/M	102	0.210**	61	0.189*
	P/M	80	0.194*	65	0.089
	*m*P/M	57	0.291**	32	0.185
2	E/M	67	0.286**	73	0.252**
	P/M	61	0.084	68	0.195**
	*m*P/M	31	0.373**	26	0.334**
3	E/M	76	0.197**	71	0.215**
	P/M	54	0.056	58	0.307**
	*m*P/M	39	0.256**	25	0.481**
4	E/M	64	0.404**	71	0.245**
	P/M	52	0.184*	58	0.139
	*m*P/M	32	0.342**	31	0.385**
5	E/M	54	0.226**	55	0.347**
	P/M	58	0.103	66	0.135
	*m*P/M	31	0.231*	18	0.372**
6	E/M	64	0.365**	44	0.285**
	P/M	52	0.174	48	0.150
	*m*P/M	34	0.314**	16	0.238
7	E/M	47	0.348**	60	0.311**
	P/M	38	0.154	33	0.192
	*m*P/M	19	0.340*	17	0.484**
8	E/M	51	0.210*	52	0.181*
	P/M	42	0.204*	51	0.136
	*m*P/M	22	0.207	16	0.077
9	E/M	66	0.346*	38	0.237*
	P/M	40	0.229*	30	0.153
	*m*P/M	16	0.356*	13	0.267
10	E/M	48	0.326**	42	0.401**
	P/M	43	0.072	37	0.188
	*m*P/M	21	0.391**	23	0.324**

**, *: Significant at the 0.01 and 0.05 probability level, respectively

Figure 3.3 Distribution of *Eco*RI/*Mse*I (E/M), *Pst*I/*Mse*I (P/M) and ^{m}*Pst*I/*Mse*I (^{m}P/M) markers over the total length (expressed in centiMorgan (cM)) of chromosome 10 of the Recombinant Inbred (RI) linkage map; clustering of E/M and ^{m}P/M markers in the centromeric region.



Discussion

Construction of the RI and the IF_2 high-density AFLP maps of maize

The high multiplex ratio of the AFLP technique, combined with the high level of polymorphism of maize was exploited to generate a large number of markers with relative ease. It is clea from the *EMM* that P/M PCs, rather than ^{m}P/M and E/M PCs, are to be chosen to minimize the mapping efforts while maximizing the number of markers to be mapped. Ideally, *EMM* equals *EM*. However, bands, identified as polymorphic and taken up in the calculation of *EM*, can be too close to allow reliable quantification and mapping. This causes a drop in *EMM* relative to *EM*. Furthermore, it should be emphasized that the technical skills that the performer

displayed in slabgel electrophoresis can highly influence the metric *EMM*. Any deviation from good laboratory practice may cause a drop in the *EMM*.

There are two additional explanations for the lower than expected *EMM* of ^{m}P/M markers, based on their high *EM*: 1) although ^{m}P/M PCs were screened and selected against highly abundant restriction fragments, numerous repetitive restriction fragments remain, leading to a poorer quality of the ^{m}P/M fingerprints; and 2) an additional amplification step makes the methylation AFLP technique more complex than the AFLP technique. This affects the linear relation between band-intensity and initial template concentration, broadening the variances of the mixture components, and finally leads to inaccurately band classification and rejection of the marker.

Differences in genetic map length

Despite the larger number of markers, the total genome coverage and the length of the individual linkage groups in both maps are systematically shorter than the ones already published (Helentjaris et al. 1986, 1988; Burr et al. 1988; Beavis and Grant 1991; Shoemaker et al. 1992; Gardiner et al. 1993; Matz et al. 1994; Causse et al. 1995; Coe et al. 1995; Senior et al. 1997). This observation does not support the assumption that the complete coverage of the maize genome is approached asymptotically as the number of mapped markers increases. On the other hand, it doesn't show evidence that the larger the number of mapped markers, the more false recombinants are induced, resulting in inflation of the total genetic map length.

A difference in map function used can hardly be the only explanation for the observed discrepancy in map length. A difference in mapping algorithm is a more plausible explanation. The maps reported above, with exception of Causse et al. (1995) were constructed with MAPMAKER (MM) (Lander et al. 1987), whereas in this study the JM package was used. To estimate the distance between a pair of adjacent markers, MM only uses the information of that pair of markers, whereas JM uses all pairwise recombination estimates in a dataset simultaneously. The other difference is the way mapping functions are applied. In the primary

estimation procedure MM assumes absence of interference, and only afterwards adjacent recombination frequencies are translated to map distances with a given mapping function. In JM, on the contrary, all calculations are based on map distances that are obtained by applying a mapping function to recombination estimates, thus accounting for a given level of interference. Due to these differences in estimation procedure, JM will produce shorter maps than MM, whenever the assumed level of interference by Kosambi is less than the true degree of interference (P. Stam, personal results).

Despite the fact that both maps are produced with JM, the total genome coverage and the length of most linkage groups in the RI map are systematically shorter than in the IF_2 map. In addition to differences in true recombination rates and environmental conditions, differences in reliability of the data may affect the observed recombination rates. It is well known that even small error rates in genotyping leads to map inflation, especially in high-density maps (Lincoln and Lander 1992). The AFLP fingerprint patterns of the IF_2 population are more complex by nature (three zygosity classes), making genotyping more prone to misclassification. In the RI population, there is a more clear-cut difference in band intensities. This might explain the difference in total length between the IF_2 and RI map. So, population types with two instead of three genotype classes and a higher level of recombination are preferred to generate a high-density AFLP linkage map.

Distorted segregation ratios

It is not unusual to find distorted Mendelian segregation ratios in populations where moderate numbers of markers were analyzed, but the observed conformity in segregation distortion across major genomic regions in chromosome 3 in both high-density AFLP linkage maps is striking and can hardly be explained by sampling bias.

The excess of D32 alleles at the major genomic region in chromosome 3 in the IF_2 map is in good agreement with the excess of B73 alleles at chromosome 3 in the RI map: D32 is partially composed of BSSS genetic material, while B73 is a BSSS inbred line. Also Lübberstedt et al. (1998), analyzing a KW1265 × D145 and a

D145 × KW1292 F_4 population, found underrepresentation of the D145 allele in a region of chromosome 3, comprising *umc10* and *bnl6.06*. In contrast, neither Senior et al. (1997), genotyping 192 RI lines of the same B73 × Mo17 RI population, nor Beavis and Grant (1991), analyzing a B73 × Mo17 F_2 population, found evidence for segregation distortion on chromosome 3. Only Gardiner et al. (1993) found distorted segregation towards a heterozygote excess on chromosome 3 between markers *umc92* and *bnl5.37*.

Co-linearity between the RI and the IF2 linkage maps

Although the two populations have no parent in common, and the four parental lines are not highly related lines, sufficient common AFLP markers were generated to align the two maps. The comparison between the two linkage maps, based on the 92.6 % of common markers, shows that linkage maps of chromosomes were, despite some rearrangements, moderately to highly co-linear (Figure1). Rearrangements in map order of some markers, predominantly those residing in or flanking marker-dense regions, were also observed by Castiglioni et al. (1998). Mapping using JoinMap results in the most likely marker order according the parameter settings. However, especially when the number of informative recombination events in a particular region is low, e.g. centromeric region, alternative marker orders with an equivalent goodness of fit, are possible.

For the 7.4% 'common' AFLP markers that map to different chromosomes in both maps, there are two plausible explanations: 1) co-incidental co-migration of two non-related AFLP fragments; or 2) areas of genomic duplications in the maize genome. Helentjaris et al. (1988) observed that 29% of their cloned maize sequences hybridized to at least two different genomic regions. These duplicate loci suggested that the maize genome either contains a partial duplication or is tetraploid in origin. Recent analyses confirmed the tetraploid nature of the maize genome, possibly being derived from the hybridization of two parents with different arrangements of rice linkage segments constituting their chromosomes (Moore 1995). However, for two unrelated maize crosses, AFLP markers of equal size and generated by the same PC map in > 90% to the same map location,

making transferring of AFLP markers of one population to the other feasible with minor risk. The expectations are that for two more related maize crosses, or for two maize crosses with one common parent, the frequency of these 'ambiguous' common markers will decrease.

Distribution of the AFLP markers over the maize genome

Considering that the underlying basis for AFLP polymorphisms (point-mutation or insertion/deletion) is evenly distributed over the DNA, and that rare-cutter sites are also randomly distributed, one could assume an even sampling of the physical genome. However, the observed tendency is that genetically there is an overrepresentation of the centromeres. So, clustering of AFLP markers around the centromere again raises the question as to whether recombination is predominantly confined to the distal regions with the centromeric regions being recombination 'cold spots'. Investigation of the genetic location of AFLP markers obtained with different restriction enzymes and with different levels of methylation revealed that clusters which co-localize with the putative centromeric regions of maize are enriched especially by E/M, while the P/M markers are shown to be more randomly spread over the genome. Surprisingly, mP/M markers behave in a same way as E/M markers.

The clustering of E/M markers also appears in other plant AFLP linkage maps, such as potato (van Eck et al. 1995), barley (Becker et al. 1995; Powell et al. 1997; Qi et al. 1998), soybean (Keim et al. 1997) and Arabidopsis (Alonso-Blanco et al. 1998). In Arabidopsis it was shown that pericentromeric heterochromatin fluoresces brightly when stained with the fluorochrome DAPI (Ross et al. 1996), known to show preference for AT-rich DNA (Sumner 1990). This is a plausible explanation of the enrichment of E/M AFLP markers in the Arabidopsis centromeres (Alonso-Blanco et al. 1998) and possibly in other plant genome centromeres, since the restriction enzymes *Eco*RI and *Mse*I have AT-rich target sequences (*Mse*I recognizes 5'-TTAA-3', while *Eco*RI recognizes 5'-GAATTC-3').

In maize the heterochromatin, enriched with methyl groups, is concentrated in the centromeric regions, the nucleolus organizer region, telomeres and knobs,

mainly consisting of particular regions that are not transcribed. So, the lower frequency of P/M markers and the clustering of mP/M markers in the centromeric regions is consistent with the enrichment of methyl groups in heterochromatin.
The good agreement between the prevalence of P/M markers in the distal genomic regions, which harbour the gene spaces, and the prevalence of P/M markers among the biallelic markers can be explained as follows: maize transposable elements, which can easily induce allelism by integration and deletion, show a preference for gene-rich transcriptionally active regions. Thus, P/M PCs give not only a better genome coverage with less markers, but also plausible landmarks for genes.

Conclusion

The high multiplex ratio of the AFLP technique, combined with the high polymorphism rate of maize, was exploited to generate a large number of markers. The high effective mapped multiplex ratio of AFLP in maize has the potential to improve the efficiency of genetic map construction of maize and to generate high-density maps around loci that control commercially important traits. To our knowledge this is the first detailed report of mapping C-methylation and epi-alleles. Although the rules governing the transmission of methylation from one generation to the next are still unclear, it is shown that C-methylation can be inherited in a Mendelian way. mAFLP markers are also of practical use in genome research. Like AFLP markers, most mAFLP markers correspond to unique positions in the genome, and, hence, can be exploited as landmarks in and as bridging tools between genetic and physical maps. Native methylated sites are present on cloned DNA segments, e.g. yeast artificial chromosomes (YACs) and bacterial artifical chromosomes (BACs) as unmethylated sites. Hence, native methylated sites can not be distinguished from native unmethylated sites on a physical map. However, lining up the physical map with a genetic map containing mAFLP markers, may help to identify native methylated sites on the physical map (Chapter 2).

Beside top-down anchoring of physical maps and mapping commercially important traits like grain yield, these linkage maps can be used 1) for map-based AFLP fingerprinting of maize inbred lines in order to determine the levels of genetic diversity in different regions of the maize genome, and 2) as main framework in a unified AFLP linkage map for maize. Although the relative map position of markers in an integrated linkage map is less reliable due to statistical errors associated with the recombination estimates and to differences in recombination frequency among crosses, an integrated AFLP linkage map for maize is an inexhaustible resource of markers, encouraging the use of the AFLP technique in maize breeding.

Acknowledgements

We would like to thank L. Heijnen and R. Hoogers for the sequencing effort. This work was financed by Keygene N.V. and 5 Dutch breeding companies: Cebeco Zaden, De Ruiter, ENZA Zaden, Rijk Zwaan and Van der Have.

Note: AFLP® is a registered trademark of Keygene N.V.

Note: The methylation AFLP® method is subject to a patent application filed by Keygene N.V.

4

Further characterization of AFLP® data as tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines

M. Vuylsteke, R. Mank, B. Brugmans, P. Stam and M. Kuiper

Abstract

AFLP® markers generated by CNG methylation sensitive (*Pst*I/*Mse*I) and CNG methylation insensitive (*Eco*RI/*Mse*I) enzyme combinations (ECs) and AFLP markers collected from hypomethylated (*Pst*I/*Mse*I) and hypermethylated (m*Pst*I/*Mse*I) regions have been compared for their efficiency of polymorphism information content (PIC), marker index (MI), sampling variance and patterns of genetic diversity among a representative sample of 33 inbred lines of maize (*Zea mays* L.). We have demonstrated that P/M or mP/M markers generate significantly higher mean PIC values (0.38 and 0.38, respectively) than E/M markers (0.33), and sets of P/M PCs generate higher MI (22.5) than sets of E/M or mP/M PCs (18.9 and 17.2, respectively). In addition, to achieve a mean standard deviation (*sd*) of 5% in the estimation of genetic distance among the 33 inbreds used in this study, the required P/M and mP/M marker sets (135 and 129 markers, respectively) are clearly smaller than the E/M marker set (173 markers). Thus, the efficiency of polymorphism detection and marker index of AFLP markers, and the sampling variance of AFLP data in the estimation of genetic similarities can be adjusted by altering the ECs. In order to investigate the effect of redundancy of information carried by the markers (over-sampling of certain regions of the genome) on the estimation of genetic relationships among inbred lines, sampling or weighing markers conditionally on their genetic map position, whether combined or not with further selection or weighing conditionally on their PIC values, was performed. Compared to random marker sets, the obtained non-random marker

sets did not reveal significant differences in genetic relationships or dendrogram topologies. However, reducing marker information redundancy by selecting markers evenly distributed over the each chromosome is effective in minimizing the sampling variance: only 106 AFLP markers evenly spread over the genome are required for a mean *sd* of 5% in the estimation of genetic distances among all pairs of the 33 maize inbred lines.

Key words: genetic diversity, AFLP®, methylation AFLP®, *Zea mays* L., linkage map

Introduction

Accurate estimates of genetic diversity levels among elite breeding materials are becoming increasingly useful in crop improvement. In maize (*Zea mays* L.) breeding, knowledge of genetic diversity levels among inbreds is helpful in 1) assigning lines to heterotic groups, 2) maintaining and broadening the genetic variation of the elite gene pool, and 3) identifying and accurately describing new varieties for the purpose of plant variety protection. During the last twenty years, DNA marker systems have become extremely useful tools for assessing genetic diversity levels among germplasm. Compared with pedigree information, DNA marker-based diversity estimates better reflect DNA differences among lines since selection pressure and genetic drift are accounted for. Until recently, genetic diversity assessments of many plant species such as maize were primarily based upon restriction fragment length polymorphism (RFLP) markers (Smith et al. 1990; Smith et al. 1991; Melchinger et al. 1991; Messmer et al. 1991; Ajmone Marsan et al. 1992; Livini et al. 1992; Messmer et al. 1993; Mumm and Dudley 1994; Dillmann et al. 1997). Although RFLP-based estimates of genetic diversity levels among elite breeding materials have been shown to be useful, and increasing the number of marker loci may further improve the discriminative power of RFLP, the labor intensive and time consuming nature of the RFLP assay may not make this feasible.

An efficient PCR-based method, called AFLP® has been developed (Vos et al. 1995) combining the restriction site variation used in RFLP with the exponential amplification aspects of PCR-based marker systems. The major advantages of the AFLP technique are: 1) the amplification has a high multiplex ratio; 2) a limited set of generic primers is used; 3) there is no need for sequence information; and 4) an almost unlimited number of markers can be generated. The AFLP method is highly efficient compared to other DNA marker systems like RFLP, RAPD and SSR (Powell et al. 1996; Russell et al. 1997; Pejic et al. 1998), although its Polymorphism Information Content (PIC) value is amongst the lowest, due to its predominantly dominant nature. To date, the AFLP technique has been succesfully used to assay diversity within many plant genomes including those of maize (Ajmone Marsan et al. 1997; Pejic et al. 1998), barley (Russell et al. 1997; Schut et al. 1997), wheat (Barrett et al. 1998), lettuce (Hill et al. 1996), tea (Paul et al. 1997), grapevine (Cervera et al. 1998), potato (Milbourne et al. 1997), rice (Zhu et al. 1998) and soybean (Maughan et al.1996). Additionally, the methylation AFLP® method (Chapter 2) offers the opportunity to generate AFLP fragments originally bounded by a methylated restriction site. Thus, genetic diversity assessments for hypermethylated portions of the genome can be made. Due to the high multiplex ratio nature of the AFLP assay and the high polymorphism rate of maize, large numbers of available markers are accumulating rapidly with a relatively low effort. Although larger numbers of random markers will provide an increasingly more precise estimate of genetic relationships, over-sampling certain regions of the genome arises, causing higher sampling variance in the estimation of genetic relationships. Over-sampling certain genomic regions is induced by the non-random genomic distribution of markers (Chapter 3). Therefore, it would be desirable to estimate genetic relationships using a subset of AFLP markers uniformly distributed over the linkage groups in order to minimize the over-sampling of the certain genomic regions and, subsequently, the sampling variance.

Aiming at further characterization of AFLP data as tool for the breeder, the objectives of this study were: 1) to compare AFLP markers generated by CNG

methylation sensitive (*Pst*I/*Mse*I) and CNG methylation insensitive (*Eco*RI/*Mse*I) enzyme combinations (ECs) and AFLP markers collected from hypomethylated (*Pst*I/*Mse*I) and hypermethylated (*^m^Pst*I/*Mse*I) regions for their efficiency of polymorphism detection, marker utility and patterns of genetic diversity among a representative sample of maize inbred lines; 2) to compare genetic diversity estimates generated using AFLP marker sets selected or weighted according to their position on the genetic map and/or efficiency of polymorphism detection, in order to estimate the effect of reducing the marker information redundancy or weighing marker information; and 3) to determine and compare the bootstrap sampling variance associated with a) bootstrap sample size, b) the use of different ECs and c) the use of random versus non-random AFLP marker sets.

Materials and methods

Plant material and DNA isolation

A total of 33 elite maize inbreds of *Zea mays* L. were chosen, representing inbred lines and their relatives largely used in the production of hybrid seed and maize breeding programs, particularly in Italy. Information regarding pedigree is listed in Table 4.1. Based on breeding information and on their heterotic behaviour in crosses, fifteen inbreds are assigned to the Iowa Stiff Stalk Synthetic (BSSS) and sixteen inbreds to the Lancaster Sure Crop (LSC) heterotic group; two inbreds (H55 and Pa91) are from miscellaneous origin and could not be assigned to any well defined heterotic group. Seeds were obtained from Dr. M. Motto, Istituto Sperimentale per La Cerealicoltura, Bergamo, Italy. Genomic DNA was extracted from a bulk of ten young plants from each line using a modified CTAB procedure described by Stewart et al. (1993).

Table 4.1 Classification and source of maize inbred lines used in this study

Inbred	Pedigree	heterotic group †
B14A	Cuzco × B14[8]	BSSS
B37	BSSS (HT) C0	BSSS
B73	BSSS (HT) C5	BSSS
B68	(41250B × B14[3]) Selection	BSSS
B84	BSSS (HT) C7	BSSS
Lo950	Pioneer P3183 SELF	BSSS
Lo951	Pioneer P3183 SELF	BSSS
Lo999	(B73 × Teosinte) × B73[n]	BSSS
N28	Nebraska Stiff Stalk Synthetic	BSSS
A1	50% B14	BSSS
A2	50% A1	BSSS
A3	commercial hybrid	BSSS
A4	commercial hybrid	BSSS
A5		BSSS
A8	commercial hybrid	BSSS
C103	Lancaster Sure Crop	LSC
C123	(C102 × C103) selection	LSC
Lo881	Synthetic C103 C0	LSC
Lo924	H99[2] × Mo17	LSC
Lo976	Mo17[2] × La16215	LSC
Lo977	Mo17[2] × H99	LSC
Mo17	CI187-2 × C103	LSC
Va22	Va17 × C103[2]	LSC
Va26	Oh43 × K155	LSC
Va35	C103 × T8[2]	LSC
Va59	(C103 × T8[2]) × (K4 × C103[2])	LSC
Va85	Virginia Long Ear Synthetic	LSC
H99	Illinois Synthetic	LSC
A6	75% Oh43	LSC
A7	75% Oh43	LSC
A10	75%Oh43	LSC
H55	Illinois High Yield[2] × Mo21A	MO
Pa91	(Wf9 × Oh40B)S4 × (Ind28-11[2] × L317)S4	MO

† BSSS = Iowa Stiff Stalk Synthetic; LSC = Lancaster Sure Crop; MO = miscellaneous origin

[a] number of backcrosses to the parent

AFLP® and methylation AFLP® analysis

The AFLP analysis was performed according to Vos et al. (1995), using the enzyme combinations (ECs) *Eco*RI/*Mse*I and *Pst*I/*Mse*I. Methylation AFLP analysis was performed as outlined in Chapter 2, using the EC *Pst*I/*Mse*I.

The adaptor sequences specific for these enzymes were synthesized according to Zabeau and Vos (1993) and are as follows:

*Eco*RI-adapter:	5'-CTCGTAGACTGCGTACC-3' 3'-CATCTGACGCATGGTTAA-5'
*Pst*I-adapter:	5'-CTCGTAGACTGCGTACATGCA'-3 3'-CATCTGACGCATGT'-5
*Pst*I-adapter*:	5'-GCATCAGTGCATGCGTGCA-3' 3'-GTAGTCACGTACGC-5'
*Mse*I-adapter:	5'-GACGATGAGTCCTGAG-3' 3'-TACTCAGGACTCAT-5'
*Mse*I-adapter^{+}:	5'-CTCGTAGACTGCGTACC-3' 3'-CTGACGCATGGAT-5'

The nonselective amplification of the restriction fragments during the methylation AFLP analysis is performed using the two different *Mse*I-primers shown below.

A two-step amplification strategy was followed in the methylation AFLP as well as in the AFLP analysis: in a selective preamplification, the restriction fragments were amplified with AFLP primers both having a single selective nucleotide. In the second step, further selective amplification was carried out using primers having two (*Pst*I-primer) or three (*Eco*RI- and *Mse*I-primer) selective nucleotides. The AFLP primers were designed based on the adapter sequence and restriction sites of *Eco*RI, *Pst*I and *Mse*I, and have the following sequences:

*Pst*I-primer	5'-GACTGCGTACATGCAG...NN-3'
*Eco*RI -primer	5'-GACTGCGTACCAATTC...NNN-3'

*Mse*I-primer	5'-GATGAGTCCTGAGTAA...NNN-3'
*Mse*I-primer^{+}	5'-GTAGACTGCGTACCTAA-3'

Hereinafter *Eco*RI/*Mse*I and *Pst*I/*Mse*I ECs will be referred to as E/M and P/M ECs and *Eco*RI+3/*Mse*I+3 and *Pst*I+2/*Mse*I+3 PCs will be referred to as E/M PCs and P/M PCs. The P/M PCs used in the methylation AFLP analysis will be referred to as ^{m}P/M PCs.

AFLP marker nomenclature

Each polymorphic AFLP fragment was identified by: 1) the code referring to the corresponding PC (Table 3.1) followed by 2) the estimated molecular size of the DNA fragment in nucleotides. SequaMark™ (Research Genetics, Huntsville, AL, USA) was implemented as size standard to assign molecular weights to the AFLP fragments. Fragments and markers detected by E/M, P/M or ^{m}P/M PCs will be referred to as E/M, P/M or ^{m}P/M fragments and markers, respectively. Markers detected by ^{m}P/M PCs will also be referred to as AFLP markers, except when AFLP and ^{m}AFLP markers need to be distinguished.

Marker information

Polymorphism Information Content (PIC) (Smith et al. 1997) of a marker, defined as the probability that two alleles taken at random from a population can be distinguished using the marker in question, is a measure of allele diversity at a locus and is equal to

$$1 - \sum_{i=1}^{n} f_i^2$$

where f_i is the frequency of the *i*th allele. The PIC value provides an estimate of the discriminatory power of a marker by taking into account not only the number of alleles at a locus, but also the relative frequencies of those alleles in the population under study. Since the AFLP technology produces dominant markers, only two states (present and absent) can be distinguished at each band position.

Therefore, assuming that each band position corresponds to a locus with two alleles, presence and absence of the band, respectively, the highest PIC value obtained by an AFLP marker equals 0.5.

In order to compare PCs for their efficiency of detecting polymorphisms in the germplasm and between any two lines taken at random from that pool, an arithmetic mean

$$PIC_{av} = \sum_{j=1}^{N} PIC_j / N$$

was calculated, where PIC_j is the PIC value of the jth AFLP marker and N is the number of AFLP markers generated by a PC. Loci that are non-polymorphic in the germplasm of interest were excluded from this calculation since there PIC = 0. Since most loci in the germplasm under study are polymorphic, the PIC_{av} for a set of markers is a slight overestimation of the real PIC_{av}.

Other metrics to compare marker sets are the multiplex ratio (*M*) and the effective multiplex ratio (*EM*) (Powell et al. 1996), which define the number of fragments and polymorphic fragments, respectively, simultaneously analyzed in a single assay. Although these metrics are influenced by the number of selective nucleotides at the 3' ends of the PCR primers and can be manipulated by combining PCs in a multiplex reaction approach, they are suited in order to facilitate selection of an appropriate EC or appropriate PCs for a given application.

Finally, the utility of a subset of PCs is a balance between the level of polymorphism detected (information content) and the extent to which an assay can identify multiple polymorphisms. Therefore, a convenient metric for marker utility, called marker index (*MI*) (Powell et al. 1996), may be defined as the number of polymorphic markers per gel lane for a pair of genotypes being compared. Its value can be calculated as

$$MI = EM \times PIC_{av}$$

The genome coverage (GC) of a marker is a function of the map distance between its flanking markers. So, high and low GC values will be assigned to markers residing in under- or overrepresented genomic regions, respectively. Consider a map with marker order ABCD. Then, the GC of a marker B (GC_B) is calculated as

$$GC_B = md_{AC} / 2 \times TML$$

where md_{AC} is the map distance (expressed in cM) between marker A and C, and *TML* is the total map length (cM). For marker A, which is telomeric and only flanked by marker B, GC_A is calculated as

$$GC_A = mp_B / 2 \times TML$$

where mp_B is the map position (expressed in cM) of marker B. Suppose markers B and C map on the same position, then

$$GC_B = GC_C = md_{AD} / 4 \times TML$$

Computation of PIC and GC values and selection of AFLP markers based on their PIC and GC values were performed using the Genstat programme (Genstat-5-Committee 1993).

Computation of genetic distances

As already mentioned above, because of the dominant nature of the AFLP markers, only two states (present and absent) can be distinguished at each band position. However, shared absence of a band is not taken as evidence of similarity in the computation of genetic similarity (GS), since it may lead to errors, particularly in the case of distantly related individuals. Therefore, shared presence of a band only, averaged over loci, was used as the measure of GS, originally deviced by Jaccard (1908) and calculated by the following equation:

$$GS_{ij} = N_{ij} / [N_i + N_j + N_{ij}]$$

where N_{ij} is the total number of bands common to lines i and j, and N_i and N_j are the number of bands only present in i and j, respectively. In the case of a missing observation for a marker in genotype i and/or j, this marker was not included in the calculation of GS_{ij}. The complement to the Jaccard similarity coefficient was used in this study and termed genetic distance (GD). Values of GD may range from 0 (identical profiles for all markers in the two inbreds) to 1 (no bands in common).

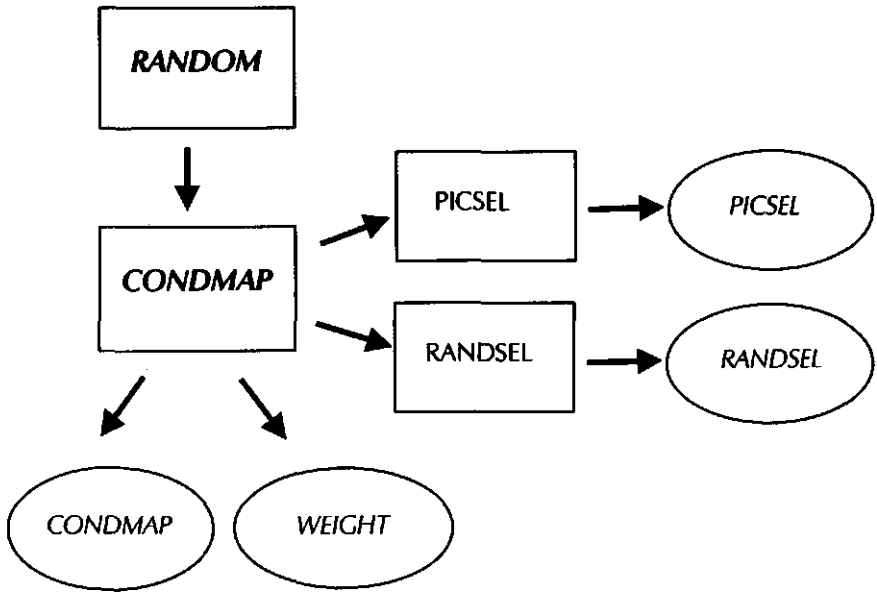
Cluster analyses were based on similarity matrices using the unweighted pair group arithmetic average (UPGMA) (Rolf 1992) and relationships between accessions were visualized as dendrograms. Differences between dendrograms were tested by generating cophenetic values for each dendrogram and the assembly of a cophenetic matrix for the different subsets of AFLP markers. The Mantel matrix (Mantel 1967) correspondence test was used to compare similarity and cophenetic matrices.

Unweighted genetic similarities were calculated, cluster analysis and Mantel tests were performed, and dendrograms were visualized using the NTSYS-PC (Rolf 1992) version 1.70. Weighted genetic similarities were performed using the Genstat programme (Genstat-5- Committee 1993).

Random versus non-random sampling

AFLP markers generated by a random set of PCs can be considered as a random sample of the genome. In contrast, non-random samples of the genome can be obtained by reducing biallelic markers or markers clustered on the genetic map, to a single representative, in order to reduce the marker redundancy. Figure 4.1 clearly outlines how different non-random marker sets and similarity matrices are obtained. A first non-random subset of markers, called CONDMAP, extracted from a random data set, contains E/M, P/M and ^{m}P/M markers mapped on the B73 × Mo17 Recombinant Inbred (RI) high-density AFLP linkage map (Chapter 3). The

Figure 4.1 Data flow chart representing the different marker sets (represented by boxes) and similarity matrices (represented by ovals) and how they are obtained. RANDOM MARKER SET contains the total amount of *Eco*RI/*Mse*I (E/M), *Pst*I/*Mse*I (P/M) and *m*Pst*I*/*Mse*I (*m*P/M) markers the 33 inbred lines are scored for. CONDMAP contains those E/M, P/M and *m*P/M markers mapped on the B73 × Mo17 Recombinant Inbred (RI) high-density AFLP linkage map (Chapter 3). The two similarity matrices constructed, based on the CONDMAP markers, unweighted and weighted by their Polymorphism Information Content (PIC) and Genome Coverage (GC) values, have been called CONDMAP and WEIGHT, respectively. RANDSEL and PICSEL were deduced from CONDMAP. Additional to single markers residing in a 1 cM window, RANDSEL and PICSEL comprise markers selected out of two or more adjacent markers residing in a window of 1 cM, at random and based on their highest PIC value in the collection of the 33 inbred lines, respectively. The corresponding similarity matrices will also be referred to as RANDSEL and PICSEL.



two similarity matrices constructed, based on the CONDMAP markers, unweighted and weighted by their PIC and GC values, have been called CONDMAP and WEIGHT, respectively. Two subsets of non-randomly sampled AFLP markers, called RANDSEL and PICSEL, were deduced from CONDMAP in the following way: linkage groups were scanned, using a 1 cM window, for single

markers or clusters of markers. So, additional to single markers residing in a 1 cM window, RANDSEL and PICSEL comprise markers selected out of two or more adjacent markers residing in a window of 1 cM, at random and based on their highest PIC value in the collection of the 33 inbred lines, respectively. As a result, RANDSEL and PICSEL represent a more even distribution of markers on each chromosome than CONDMAP. The corresponding similarity matrices will also be referred to as RANDSEL and PICSEL.

Estimation of sampling variance of marker data using the bootstrap procedure

In order to empirically estimate the sampling variance of marker data in the estimation of genetic similarities for the different marker sets, a computer program was written to execute a bootstrap sampling procedure, which is similar to the bootstrap procedure of Tivang et al. (1994). The program was designed to execute bootstrap sampling treating markers as independent sampling units. For a given number of *N* markers, thousand bootstrap samples were taken. The *N* markers, with *N* ranging from 50 to 200 markers in steps of 50, were selected at random with replacement out of the whole data sets. The GS was calculated between all (528) inbred pairs using the Jaccard similarity index. The variability among 1000 bootstrap samples for each pair of inbreds was measured in standard deviation (*sd*) units. Subsequently, a mean percentage *sd* of the 528 inbred pairs for a sample size *N* was plotted against the sample size. Natural log transformations of the percentage standard deviation resulted in a linear relationship between *sd* and sample size *N*. This linear relationship was used to estimate the number of markers required to achieve, e.g., a mean *sd* of 5% in the estimation of genetic distance among the 33 inbreds. Calculations were performed in the Genstat programme (Genstat-5- Committee 1993).

Results

Polymorphism, efficiency of polymorphism detection and marker index

The thirty-three inbred lines were surveyed for polymorphism using 36 E/M, 21 P/M and 14 ^{m}P/M PCs (Table 3.1). In total, 2047 E/M, 1250 P/M and 639 ^{m}P/M polymorphic bands were identified. The fraction of polymorphic bands (β) was calculated for each PC. The highest mean β values were calculated for P/M and ^{m}P/M PCs and were very similar (0.80 and 0.81, respectively) (Table 4.2); in contrast, the fraction of polymorphic markers detected by E/M PCs was 0.72. PIC and PIC_{av} values were calculated for each marker and for each PC, respectively. PIC_{av} values calculated for P/M and ^{m}P/M PCs are not significantly different from each other, but are significantly higher ($P<0.0001$, standard two-sample *t*-test) than the PIC_{av} value calculated for E/M PCs (Table 4.2). The MI for each PC was calculated as the product of the PIC_{av} and the *EM*. The highest mean MI was calculated for P/M PCs (22.5), which is, in comparison with the E/M PCs, due to a higher mean PIC_{av} component and fraction of polymorphic bands, and, in comparison with the ^{m}P/M PC, due to a higher multiplex ratio component (Table 4.2).

Table 4.2 Comparison of the multiplex ratio *M*, the fraction of polymorphic bands β, the effective multiplex ratio *EM*, the average Polymorphism Information Content PIC_{av} and its standard deviation (sd), and the marker index *MI* for each enzyme combination (EC), calculated on the basis of experimental data obtained from 33 maize inbred lines (Table 4.1).

EC	number of PCs	number of bands	M	β	EM	PIC_{av}	sd of PIC_{av}	MI
E/M	36	2843	79	0.72	57	0.33	0.024	18.77
P/M	21	1562	74	0.80	60	0.38	0.019	22.50
^{m}P/M	14	789	56	0.81	46	0.38	0.021	17.23

Genetic relationships between maize genotypes based on random sampled markers

Similarity matrices were constructed based on shared presence of a band and revealed that the estimates for average genetic similarity between genotypes for 36 E/M, 21 P/M and 14 "P/M PCs were very similar to one another. The Mantel matrix correspondence test was used to compare the similarity matrices and the correlation coefficients are presented in Table 4.3. All the correlation coefficients are high and statistically significant ($P<0.01$). Genetic similarity estimates obtained from each EC data set were used to derive three dendrograms. The dendrograms (Figure 4.2a-c) clearly discriminate between the two heterotic groups (BSSS and LSC) and only minor differences in their topologies are present. Note that Oh43-related lines A6 and A7 were positioned within the BSSS group, although Oh43 is usually considered a LSC type. The similarity in dendrogram topology is reflected in the overall highly significant ($P<0.01$) cophenetic correlation coefficients.

Table 4.3 Correlations between cophenetic matrices (above diagonal) and similarity matrices (below diagonal) obtained with different marker sets. Cophenetic correlation coefficients for the unweighted pair group arithmetic average (UPGMA) dendrograms shown in Figure 4.2a-g are given in bold on the diagonal.

	E/M	P/M	"P/M	CONDMAP	RANDSEL	PICSEL	WEIGHT
E/M	**0.938**	0.929	0.905	0.943	0.923	0.908	0.928
P/M	0.936	**0.912**	0.903	0.954	0.93	0.954	0.933
"P/M	0.869	0.891	**0.880**	0.927	0.905	0.895	0.896
CONDMAP	0.885	0.956	0.892	**0.865**	0.977	0.949	0.973
RANDSEL	0.857	0.937	0.868	0.971	**0.863**	0.927	0.96
PICSEL	0.884	0.949	0.873	0.977	0.98	**0.876**	0.964
WEIGHT	0.87	0.941	0.859	0.967	0.973	0.974	**0.889**

The cophenetic correlation coefficients provided by each EC (Table 4.3) indicate the extent to which the clustering of genotypes depicted in the dendrograms accurately represents the estimates of genetic similarity of the genotypes obtained with that EC. In order to quantify the extent of any differences between the three

dendrograms, a cophenetic matrix was constructed for each assay and compared using the Mantel matrix correspondence test (Table 4.3).

Genetic relationships between maize genotypes based on non-random sampled markers

From a total of 1385 AFLP markers (592 E/M, 532 P/M and 261 ^{m}P/M markers) the map position on the B73 x Mo17 Recombinant Inbred (RI) high-density AFLP linkage map (Chapter 3) is known. These 1385 markers are well dispersed throughout the maize genome and cover 93.7 % of total genome length (1178 cM) of the RI linkage map. Two hundred twenty-nine markers represent a unique map position, while the remaining 1156 markers were clustered into 263 clusters of at least 2 markers in a 1cM window. Comparison of the CONDMAP, WEIGHT, RANDSEL and PICSEL similarity matrices and the corresponding correlation coefficients reveal that the genetic similarity estimates between genotypes are very similar to one another (Table 4.3). All the similarity correlation coefficients are high and statistically significant ($P<0.01$). Genetic similarity estimates were used to derive dendrograms. The four dendrograms clearly discriminate between the two heterotic groups (BSSS and LSC), and show only some minor differences in topology (Figure 4.2d-g). Also here the inbred lines A6 and A7 are positioned within the BSSS group. The similarity in dendrogram topology is reflected in the overall highly significant ($P<0.01$) cophenetic correlation coefficients.

Comparison of genetic relationships between maize genotypes based on random and non-random sampled markers

Comparison of the similarity matrices and dendrogram topology based on either the random marker sets E/M, P/M and ^{m}P/M or the non-random marker sets CONDMAP, WEIGHT, RANDSEL and PICSEL, and the corresponding similarity and cophenetic correlation coefficients, reveal that 1) all similarity and cophenetic correlation coefficients are high and statistically significant ($P<0.01$), and 2) the highest correlations between similarity matrices as well as dendrogram topology are obtained between the random P/M marker set and the four non-random data sets (Table 4.3).

Figure 4.2 Dendrograms produced from genetic similarity matrices calculated using the Jaccard similarity coeffficient

Fig. 4.2a *Eco*RI/*Mse*I

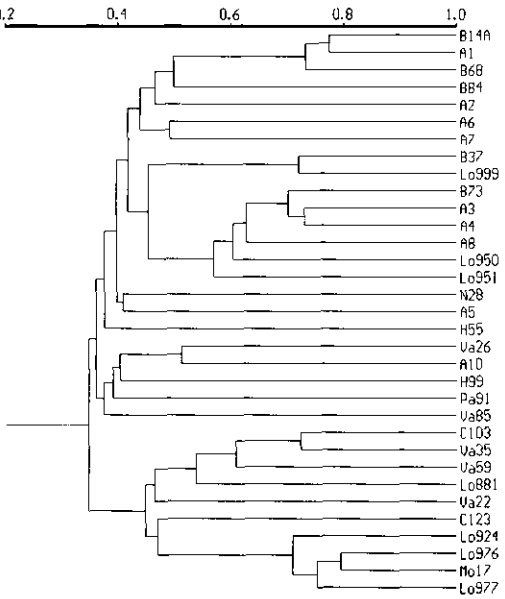


Fig. 4.2b *Pst*I/*Mse*I

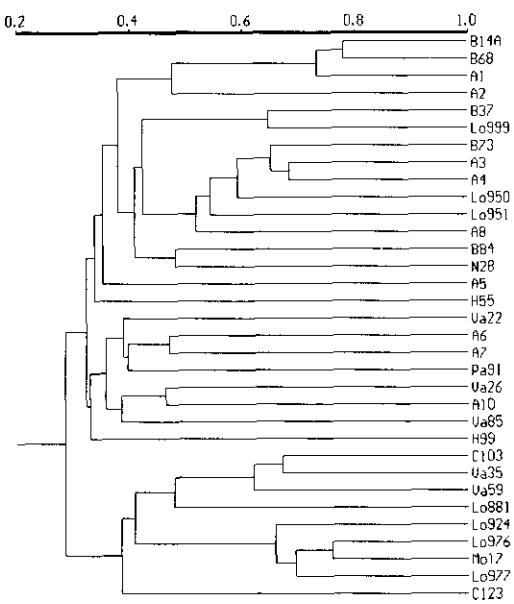


Fig. 4.2c "*Pst*I/*Mse*I

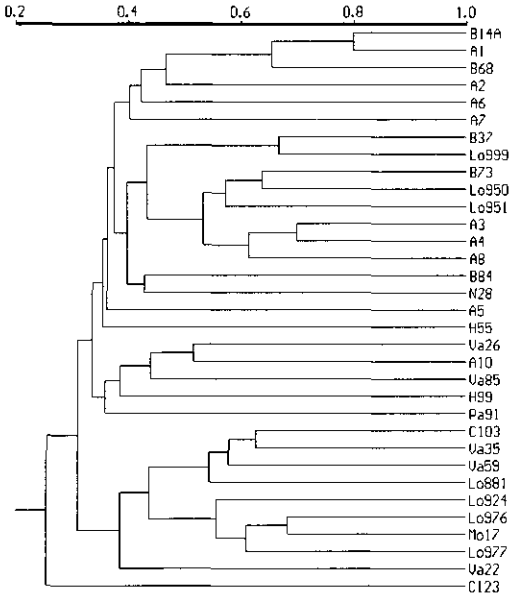


Fig. 4.2d CONDMAP

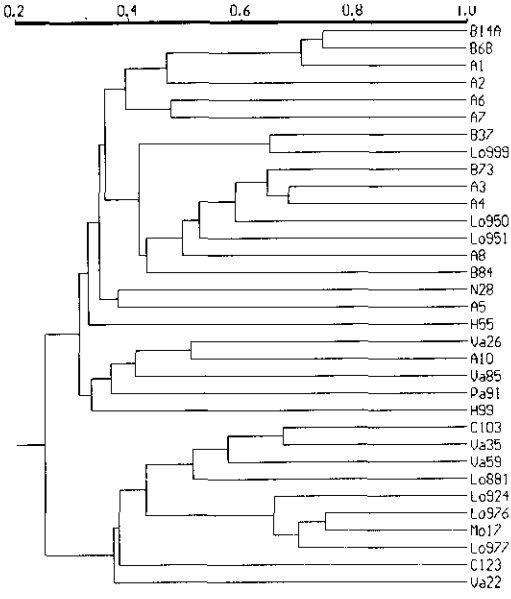


Fig. 4.2e WEIGHT

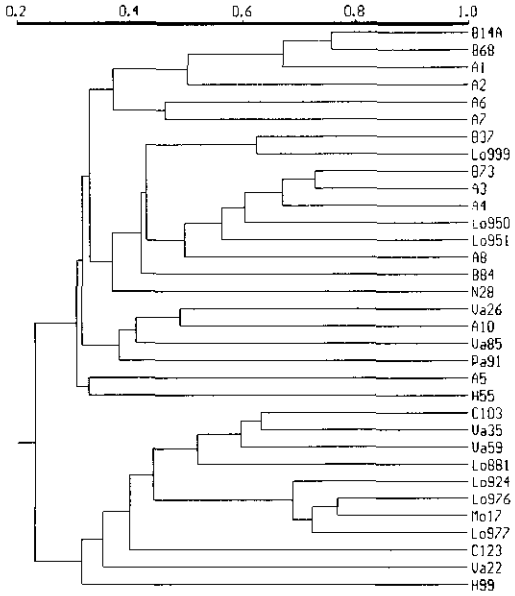


Fig. 4.2f RANDSEL

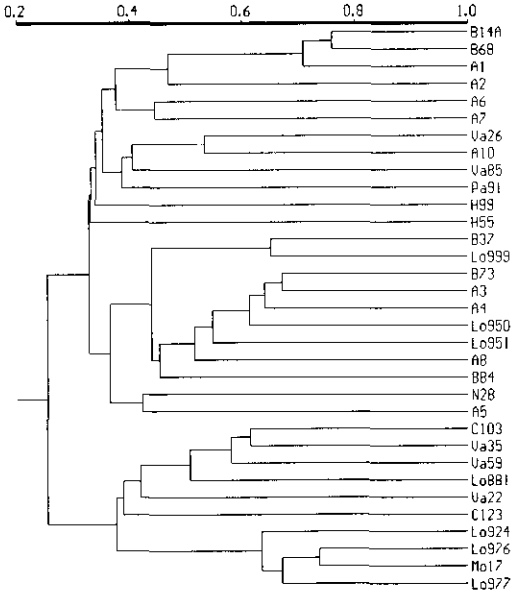
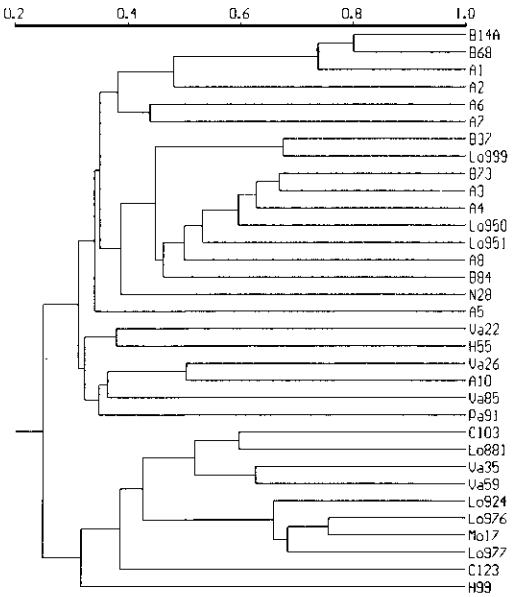


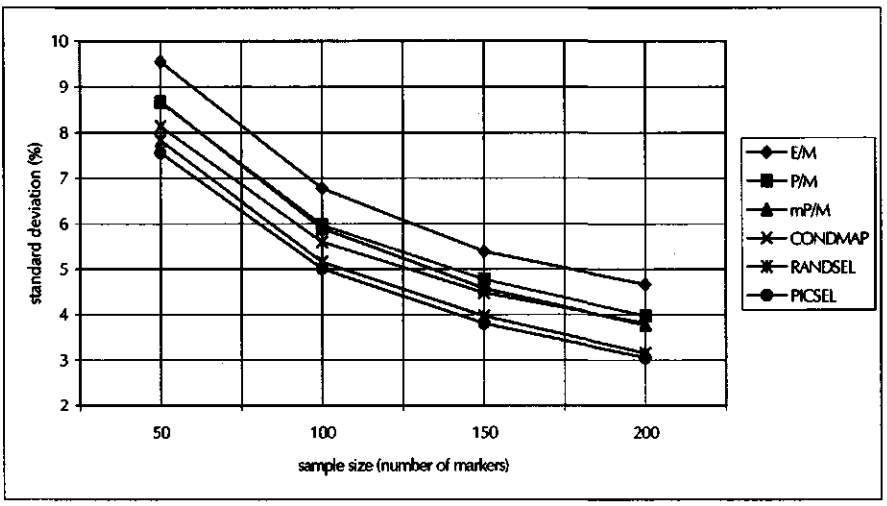
Fig. 4.2g PICSEL



Sampling variance of genetic similarities

Relationships between sample size and sampling variance in the estimation of genetic similarity among all inbred pairs for random and non-random marker sets is depicted in Figure 4.3. As expected, the sampling variance, measured in percentage standard deviation, is inversely proportional to the sample size. Natural log transformations of the percentage *sd* resulted in a linear relationship between *sd* and sample size (data not shown). Y-intercepts ranged from 2.27 (PICSEL) to 2.44 (E/M), while the rate of reduction in mean *sd* (slope) ranged from 0.0048 (E/M) to 0.0060 (PICSEL). The number of markers required to achieve a mean *sd* of 5% in the estimation of genetic distance among the 33 inbreds were estimated using the linear relationship between *sd* and sample size: 173, 135, 129, 122, 106 and 101 for the E/M, P/M, mP/M, CONDMAP, RANDSEL and PICSEL marker set, respectively.

Figure 4.3 Plot of the relationships between the mean standard deviation (*sd*) (%) and the sample size for the random (*Eco*RI/*Mse*I (E/M), *Pst*I/*Mse*I (P/M) and ^{m}*Pst*I/*Mse*I (^{m}P/M)) and non-random (CONDMAP; RANDSEL; PICSEL) marker sets.



Discussion

Our results from AFLP analyses of the set of 33 inbred lines confirm that maize exhibits a high degree of DNA polymorphism, even within commercially germplasm pools (Livini et al. 1992; Mumm and Dudley 1994; Smith et al. 1997; Senior et al. 1998; Pejic et al. 1998). The dendrograms (Figure 4.2a-c) clearly discriminate between the two heterotic groups (BSSS and LSC). The similarity in dendrogram topology is reflected in the overall highly significant ($P<0.01$) cophenetic correlation coefficients. Not all lines with defined affiliation to one of the two heterotic groups were assigned to their specific main clusters, e.g., the Oh43-related lines A6 and A7 were positioned within the BSSS group, although Oh43 is usually considered a LSC type. Also Pejic et al. (1998) noted that Oh43-related inbreds were grouped apart from the LSC cluster and merged with the BSSS lines. In addition, Mumm and Dudley (1994) classified Oh43 separately from the "pure" LSC.

AFLP-detectable DNA polymorphisms are predominantly single nucleotide type mutations. However, since there are indications that some species, including maize (SanMiguel et al. 1996), contain more active transposon systems than others, the proportion of insertion/deletion type mutations detected by AFLP in maize is not negligible. Due to a high sensitivity and level of resolution for differences in band size, the AFLP technique is capable to detect small insertion/deletion type mutations of only a few base pairs long. However, despite of the high resolution of the AFLP technique, erroneous scoring of non-homologous bands of similar mobility, which may increase when the allelic diversity in the group is large, can not be excluded. The frequency of occurrence of these 'erroneous' bands in maize was estimated at 7.4% (Chapter 3), which may lead to a slight overestimation of similarity.

The highest fractions of polymorphic bands were obtained with P/M and ^{m}P/M markers. A higher degree of polymorphism detected by the P/M EC than by the E/M EC can be ascribed to the higher rate of mutation in the *PstI* recognition site (containing two CpNpG trinucleotides) due to $^{5m}C \rightarrow T$ transitions. According to Yang et al. (1996), there is a 10- to 40-fold increase in the rate of transitions at methylated versus unmethylated cytosines.

We have demonstrated that the efficiency of polymorphism detection by the AFLP method can be adjusted by altering the ECs: sets of P/M or ^{m}P/M PCs generate significantly higher PIC_{av} (0.38 and 0.38, respectively) than sets of E/M PCs (0.33). Due to the dominant nature of AFLP markers, the highest PIC value obtained by an AFLP marker equals 0.5, which is low in comparison with multi-allellic marker systems like RFLP and SSR. Senior et al. (1998) assayed 94 elite maize inbred lines for polymorphism at 70 SSR loci. PIC values for these SSR loci ranged from 0.17 to 0.92, with an average of 0.59. This result was consistent with the result of Smith et al. (1997) who assayed 58 inbred lines for polymorphism at 131 SSR and 80 RFLP loci. Mean values for PIC for SSRs and RFLPs were similar, approximately 0.62. Dubreuil et al. (1996) reported a mean PIC value of 0.60 for 63 RFLP loci used within a set of 116 maize inbred lines.

Powell et al. (1996) highlighted in their report the distinctive nature of the AFLP assay, with an MI almost an order of magnitude higher than for other marker systems like RFLP, RAPD and SSR examined for a soybean germplasm. The high MI of the AFLP method is due to its very high EM component in the calculation. We have demonstrated that the MI of the AFLP method can be adjusted by altering the ECs: sets of P/M PCs generate, for a pair of genotypes being compared, 4 to 5 polymorphic fragments more per assay than sets of E/M or "P/M PCs. The MI for the three ECs are in good accordance with the Effective Mapped Multiplex Ratio (*EMM*) (number of polymorphic fragments between two parental lines in a single assay, of which the segregational pattern could be scored and quantified in such a way that they could be reliably mapped) of the three ECs, calculated for two maize mapping populations in (Chapter 3).

Not only the metrics PIC_{av} and MI highlight the distinctive nature of the P/M EC, but also the sampling variance in the estimation of genetic similarity among inbred pairs. The sampling variance was used to determine how many AFLP markers are required to provide a given level of precision. To achieve a mean *sd* of 5% in the estimation of genetic distance among the 33 inbreds used in this study, the P/M and "P/M marker sets (135 and 129 markers, respectively) are clearly smaller than the E/M marker set (173 markers). Accordingly, the average number of P/M, "P/M and E/M PCs required to obtain a 5% precision in the GD estimate are 2-3, 3-4 and 3-4, respectively. Similarly, Pejic et al. (1998), investigating in a set of 33 maize inbred lines the sampling variance of a E/M AFLP data set, found that the standard deviation of the estimate was not longer significantly reduced when more than 150 E/M markers were analysed. Accordingly, the average number of RFLP, RAPD and SSR assays to attain an equal precision was 30-40 clone-enzyme RFLP combinations, 40-50 RAPD primers and 20-30 SSR primers, respectively. In contrast, Messmer et al. (1993) reported that at least 100 independent polymorphic RFLP clone-enzyme combinations are required to keep the standard error below 0.05 in the estimation of genetic diversity among 29 maize inbreds. Tivang et al. (1994), investigating in a set of 37 maize the sampling variance of a RFLP data set, found that the number of bands required for a coefficient of

variation of 10% was 388, 150 and 38 for closely-, intermediately- and distantly-related inbreds, respectively.

So, thanks to their high PIC_{av} and MI value, their lower sampling variance in the GD estimation among inbreds and their being uniformly distributed over the genome (Chapter 3), P/M markers offer a distinctive advantage over E/M and ^{m}P/M markers when financial and laboratory efforts, marker utility, genome coverage and minimizing sampling variances in GD estimations are simultaneous constraints.

Reducing marker information redundancy by sampling or weighing markers conditionally on their genetic map position, to obtain a more even distribution of markers on each chromosome, whether combined or not with further selection or weighing conditionally on their information content, did not reveal significant differences in genetic relationships or dendrogram topologies. The highest values for similarity and cophenetic correlation coefficients between a random marker set and the four non-random marker sets are obtained by the P/M marker set. These high correlations can be attributed to a more even distribution of P/M than E/M and ^{m}P/M markers on each chromosome, combined with the on average higher PIC values for P/M markers.

Non-random marker sets differ from random sets since they allow minimizing the number of markers required for a given level of precision in the estimation of genetic distances among maize inbreds. This discrepancy can be attributed to redundancy of information in the markers belonging to a non-random marker set. Reducing marker information redundancy by selecting markers evenly distributed over the each chromosome is highly effective in minimizing the sampling variance: 106 (RANDSEL) instead of 122 (CONDMAP) markers are required for a mean *sd* of 5%. Representing clusters of markers by their highest informative marker instead of by a random selected marker, is of a low effectiveness. It is important to realize, however, that in general AFLP markers are produced on a random basis, and not as a subset of AFLP markers uniformly distributed over the linkage groups. Nevertheless, these results underscore the need to perform an analysis based on 'map-based' AFLP markers in order to minimize the number of markers required

for a given level of precision in the estimation of genetic distances among individuals, or in order to estimate genetic relationships among individuals more precisely for a given number of markers.

Acknowledgements

We would like to thank Dr. M. Motto from Istituto Sperimentale per La Cerealicoltura at Bergamo, Italy, for providing us seeds of the 33 inbred lines. This work was financed by Keygene N.V. and 5 Dutch breeding companies: Cebeco Zaden, De Ruiter, ENZA Zaden, Rijk Zwaan and Van der Have.

5

Chromosomal regions involved in hybrid performance and heterosis: their AFLP®-based identification and practical use in prediction models

M. Vuylsteke, M. Kuiper and P. Stam

Abstract

In this paper, a novel approach towards the prediction of hybrid performance and heterosis is presented. Here, we describe an approach based on: 1) the assessment of associations between AFLP® markers and hybrid performance and specific combining ability (SCA) across a set of hybrids; and 2) the assumption that the joint effect of genetic factors (loci) determined this way can be obtained by addition. Estimated gene effects for grain yield varied from additive, partial dominance to overdominance. This procedure was applied to 53 interheterotic hybrids out of a 13 by 13 half-diallel among maize inbreds, evaluated for grain yield. The hybrid value, representing the joint effect of the genetic factors, accounted for up to 62.4% of the variation in the hybrid performance observed, whereas the corresponding efficiency of the SCA model was 36.8%. Efficiency of the prediction for hybrid performance was evaluated by means of a cross-validation procedure for grain yield of 1) the 53 interheterotic hybrids and 2) 16 hybrids partly related to the 13 by 13 half-diallel. Comparisons in prediction efficiency with the 'distance' model were made. Since the map position of the selected markers is known, putative quantitative trait loci (QTL) affecting grain yield, in terms of hybrid performance or heterosis, are identified. Some QTL of grain yield detected in the present study were located in the vicinity of loci reported earlier as having quantitative effects on grain yield.

Key words: hybrid performance, heterosis, Zea *mays* L., AFLP®, QTL

Introduction

Heterosis is still the main cause of the success of the commercial maize (Zea *mays* L.) industry. Therefore, identification of genetic factors contributing to hybrid performance (HP) and/or heterosis and a suitable method that could predict HP and/or heterosis with some accuracy before field evaluation of test hybrids are of particular interest. In maize, the main strategy that has been followed towards new ways of hybrid prediction during the last decade is based on the 'distance' model: heterosis, defined and measured as the superiority of the hybrid over the midparent (the average performance of the two parents of the hybrid), is related to the genetic divergence between its parental lines (Lee et al. 1989). The potential of this 'distance' model-strategy has been extensively tested in maize, where genetic distances were first computed from isozyme data on parental inbreds (Frei et al. 1986), later from RFLPs (Lee et al. 1989; Godshalk et al. 1990; Melchinger et al. 1990 a, b; Smith et al. 1990; Dudley et al. 1991; Melchinger et al. 1992; Boppenmaier et al. 1993; Burstin et al. 1995) and recently from PCR-based RAPDs (Lanza et al. 1997) and AFLP® markers (Ajmone Marsan et al. 1998). The general tendency found was that the prediction effiency of the 'distance' model is high when 1) hybrids between related lines (intraheterotic crosses) and 2) hybrids between both related and unrelated lines (intra- and interheterotic crosses) are considered. However, correlations between genetic distances of unrelated lines only and their respective interheterotic crosses, were of low practical predictive value. This tendency is in good agreement with quantitative-genetic expectations (Charcosset and Essioux 1994), ascribing the failure of the 'distance' model for interheterotic crosses to the fact that linkage associations between markers and quantitative trait loci (QTL) generally differ randomly from one heterotic group to the other. Since only the interheterotic crosses are of commercial importance and of interest to the breeder, the practical value of the 'distance' model-approach is limited.

Other, more recent strategies for predicting HP, especially between unrelated lines, were proposed by Bernardo (1994) and Charcosset et al. (1998). The first method, based on best linear unbiased prediction (BLUP), uses covariances between HPs, estimated with marker data on parental inbreds, to predict the performance of an untested hybrid from the performance of related, tested hybrids. The second method is based on the principle that two hybrids with parents similar at the marker level, should display similar specific combining ability (SCA) values. Markers are used to generate covariates for SCA by means of principle component analysis.

In maize, many studies have been conducted to identify and map QTL for grain yield (GY) and yield components (Edwards et al. 1987; Stuber et al. 1992; Zehr et al. 1992; Beavis et al. 1994; Veldboom et al. 1994; Ajmone Marsan et al. 1995; Austin and Lee 1996; Cockerham and Zeng 1996; Eathington et al. 1997; Austin and Lee 1998). These studies strongly suggested that there are multiple QTL affecting GY throughout the genome. The results are generally in favor of the hypothesis of dominance of favorable alleles to explain the observed heterosis in GY, although overdominance at individual QTL (Stuber et al. 1992) and epistasis can not be ruled out.

In this paper, we present a novel approach towards the prediction of HP and heterosis. This approach is based on (1) the assessment of associations between AFLP markers and HP, resp. SCA across a set of hybrids and (2) the assumption that the joint effect of genetic factors determined this way can be obtained by addition. The chromosomal position of the loci involved in HP or heterosis is assumed to be in tight linkage with the marker locus since loosely trait locus-marker associations will be broken up due to accumulated recombination events during the establishment of the inbred lines. At the same time, since the map position of the selected markers is known, putative QTL affecting the trait of interest are identified.

Materials and methods

Plant material

Six inbred lines from Iowa Stiff Stalk Synthetic (BSSS), five from Lancaster Sure Crop (LSC) and two of miscellaneous origin (MO) were chosen as parents for a half-diallel mating design (Figure 5.1). The 6 BSSS and the 5 LSC inbred lines from the half-diallel, were also chosen to be tested against B14A and B73 (BSSS testers), and Lo881 and C103 (LSC testers), respectively. Another 16 single crosses were obtained by testing 8 BSSS inbreds (Lo999, N28, A1, A2, A3, A4, A5 and A8) against Lo881 and C103. The pedigree backgrounds of all inbreds are given in Table 4.1.

Figure 5.1 Schematic representation of the half-diallel mating design, involving six inbred lines from Iowa Stiff Stalk Synthetic (BSSS), five inbred lines from Lancaster Sure Crop (LSC) and two of miscellaneous origin (MO) chosen as parents. Intraheterotic crosses are marked by '×', interheterotic crosses are marked by '⊗'.

		BSSS						LSC					MO	
		B14A	B37	B73	B84	Lo950	Lo951	C103	C123	Lo881	Mo17	Va59	H55	Pa91
BSSS	B14A		×	×	×	×	×	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	B37			×	×	×	×	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	B73				×	×	×	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	B84					×	×	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	Lo950						×	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	Lo951							⊗	⊗	⊗	⊗	⊗	⊗	⊗
LSC	C103								×	×	×	×	⊗	⊗
	C123									×	×	×	⊗	⊗
	Lo881										×	×	⊗	⊗
	Mo17											×	⊗	⊗
	Va59												⊗	⊗
MO	H55													⊗
	Pa91													

Field trials

The 78 single cross hybrids from the half-diallel and the 38 parental testcrosses were evaluated in 1994 at three different environments (Bergamo, Luignano and Turano) for GY (t ha^{-1} at 15.5% moisture). The experimental design is described in Ajmone Marsan et al. (1998).

Data handling

The half-diallel and the testcrosses had 22 interheterotic F_1 data in common. These duplicate and reciprocal F_1 data were averaged. For the application of the 'distance' model, all the 78 F_1 data were considered. Diallel analysis was performed according to Griffing (1956) Model I of Method 4 excluding parents and reciprocals, partitioning the performance of the hybrid (Y_{ij}) between inbreds *i* and *j* classically as:

$$Y_{ij} = \mu + gca_i + gca_j + sca_{ij}$$

where μ is the mean of the HPs, gca_i and gca_j are the GCAs of the inbreds *i* and *j*, respectively, and sca_{ij} is the SCA between inbreds *i* and *j*. GCA and SCA variances were highly significant ($P<0.01$) and of a similar order of magnitude.

Since the breeder is only interested in interheterotic crosses, and to remove the influence of group effects (intra- versus interheterotic groups) on further analyses, e.g. marker selection, the intraheterotic group crosses (BSSSxBSSS; LSCxLSC) are excluded, reducing the data set to 53 hybrids (Figure 5.1). A simple ANOVA-test revealed a significant difference ($P<0.001$) between the intra- and interheterotic HP and SCA for GY (data not shown).

Considering the 53 interheterotic F_1 data only, the GCAs of the parental lines were adjusted for the contribution of the other lines to the mean of the line in question, since there are a small number of parental lines (Falconer 1989):

$$gca_i = \left(\frac{n-1}{n-2}\right)(\bar{i}-\mu)$$

where gca_i is the GCA of the inbred *i*, *n* is the number of lines crossed with inbred *i*, $\bar{i}$ is the mean performance of parental line *i* and μ is the mean of the HPs of 1) BSSS × LSC and BSSS × MO crosses, when inbred *i* is a BSSS inbred line, 2) LSC × BSSS and LSC × MO crosses, when inbred *i* is a LSC inbred line, and 3) BSSS × MO and LSC × MO crosses, when inbred *i* is from miscellaneous origin (Figure

5.1). If there is no dominance or epistasis, the performance of the hybrid from a cross between the *i*th female and the *j*th male is predicted by,

$$EY_{ij} = \mu + gca_i + gca_j$$

Any significant deviation from the observed Y_{ij} must be due to dominance or epistatic effects. These deviations, specific to individual crosses, are referred to as SCA.

AFLP® and methylation AFLP® analysis

The 13 inbred lines were assayed for their respective AFLP and methylation AFLP profiles as described in Chapter 4. A total of 1385 AFLP markers (592 *Eco*RI/*Mse*I (E/M), 532 *Pst*I/*Mse*I (P/M) and 261 ^{m}*Pst*I/*Mse*I (^{m}P/M) markers) out of 1539 AFLP markers mapped on the B73 × Mo17 Recombinant Inbred (RI) high-density AFLP linkage map (Chapter 3), were chosen for further analysis.

'Distance' model

Genetic Distances (GD) between pairs of inbred lines were calculated from AFLP data for all possible pairs of inbreds by the following equation:

$$GD_{ij} = 1 - [N_{ij}/(N_i + N_j + N_{ij})]$$

where N_{ij} is the total number of bands common to lines *i* and *j*, and N_i and N_j are the number of bands only present in *i* and *j*, respectively. This distance measure is equal to one minus the genetic similarity coefficient originally deviced by Jaccard (1908). Values of GD may range from 0 (identical profiles for all markers in the two inbreds) to 1 (no bands in common).

In a same way as the 78 F_1 HP data were partitioned into a mean value, a GCA and SCA component, the GD values associated with 78 F_1 hybrids were partitioned as:

$$GD_{ij} = \mu + ggd_i + ggd_j + sgd_{ij}$$

where μ is the mean of the GDs, ggd_i and ggd_j are the general genetic distances (GGD) of the inbreds *i* and *j*, respectively, and sgd_{ij} is the specific genetic distance (SGD) between the inbreds *i* and *j* (Melchinger et al. 1990b). Linear correlations were calculated for various combinations of HP, SCA, GD and SGD for the 78 F_1 hybrids.

Selection of markers using the Kruskal-Wallis test

To find markers that are, across the 53 F_1's, significantly associated with HP, a nonparametric statistical method has been used. Here, the rank sum test of Kruskal-Wallis, one of the three QTL mapping methods handled by MapQTL (van Ooijen and Maliepaard 1996) has been used as nonparametric statistical method. The molecular genotypes of the 53 hybrid combinations were predicted on basis of the parental genotypes and converted into the following genotype codes: A for homozygous absence, H for heterozygosity and B for homozygous presence of the marker allele. The actual genotype information for each locus and for all hybrids was structured in this way in order to meet the input file structure of the software used (MapQTL; van Ooijen and Maliepaard 1996). For each locus, the F_1 is classified as either of three classes, i.e. A, H or B. Beside the genotype information at each locus, the map position of the loci and the quantitative data are needed as input for the Kruskal-Wallis test as performed by MapQTL. The output of the Kruskal-Wallis test lists for every locus (sorted according the map) the name of the locus and its map position, the number of informative individuals, the Kruskal-Wallis test statistic and corresponding p-value, and, subsequently, for each class respectively the genotype, the mean rank, the arithmetic mean and the number of individuals in the class. B73 and Mo17 will be referred to as 'origin' of a selected marker allele, when the AFLP marker has been identified as a B73 or Mo17 marker, respectively, mapped on the B73 × Mo17 RI linkage map (Chapter 3).
In order to keep the overall false positive rate low, a stringent significance level of 0.001 and 0.005 was used in the selection of markers significantly associated with

HP and SCA, respectively, across the 53 F_1's. Only those loci at the 0.001 and 0.005 significance level, respectively, for which all individuals are informative and the three genotypic classes are represented with at least one individual, were retained for further analysis.

Model for the prediction of HP

For each selected marker, the additive (a) and dominance (*d*) effects are estimated from the means of the three genotypic classes. Consider M as being the marker allele represented by an AFLP fragment, while m encodes the absence of that marker allele (i.e. m encodes one or different other alleles at the same marker locus). Then,

$$\hat{a} = (\bar{x}_{MM} - \bar{x}_{mm}) / 2$$

and

$$\hat{d} = \bar{x}_{Mm} - ((\bar{x}_{MM} - \bar{x}_{mm}) / 2)$$

where $\bar{x}_{MM}$, $\bar{x}_{mm}$ and $\bar{x}_{Mm}$ denote the arithmetic mean of the genotypic classes B, A and H, respectively. Using the notation of Hayman (1954), the genotype of individual *i* at locus *l* is represented by the variable θ_l^i, which takes the value -1, 0 and +1 for genotypes mm, Mm and MM, respectively. The genotypic value of individual *i* for a single locus *l* is written as:

$$Y_i = c_l + a_l \theta_l^i + d_l(1-(\theta_l^i)^2)$$

where c_l is the average value of homozygotes mm and MM, and *a* and *d* are the additive and dominance effects. Subsequently, if the trait is controlled by *nl* loci acting independently (no epistasis), the genotypic value of individual *i* is written as:

$$Y_i = C + \sum_{l=1}^{nl} a_l \theta_l^i + d_l(1-(\theta_l^i)^2)$$

with $C = \sum_{l=1}^{nl} c_l$, the mean of all homozygotes over all loci.

The genotypic value of the F_1 from the cross $i \times j$ can be written as:

$$Y_{ij} = C + \sum_{l=1}^{nl} a_l (\theta_l^i + \theta_l^j)/2 + d_l(1-\theta_l^i \theta_l^j)/2$$

Finally, a hybrid value $TCSM_{ij}$ can be calculated for any hybrid as $TCSM_{ij} = Y_{ij} - C$, representing the total contribution of the selected markers (TCSM) in terms of their a_l and d_l estimates. Different TCSMs can be calculated as a function of the significance level used in selecting markers, resulting in a $TCSM_{0.001}$, $TCSM_{0.0005}$ and $TCSM_{0.0001}$. Linear regression of the HP on the TCSM results in a model for the prediction of the HP.

Note that, since a parental line is supposed to be either homozygous for the absence or homozygous for the presence of the marker alleles showing significant association with QTL of the trait of interest, its TCSM *per se* reduces to

$$TCSM_{ii} = \sum_{l=1}^{nl} a_l \theta_l^i .$$

Model for the prediction of GCA

Analogous to the partitioning of HP into GCA and SCA of parents, the $TCSM_{ij}$ of the hybrid Y_{ij} between inbreds i and j can be written as:

$$TCSM_{ij} = \mu + GCSM_i + GCSM_j + SCSM_{ij}$$

with the analogous interpretation of general and specific contributions of selected markers. The GCSM of a line i is calculated as the deviation of its mean from the

overall mean μ, adjusted for the contribution of the other lines to the mean of the line in question (Falconer 1989):

$$GCSM_i = (\frac{n-1}{n-2})\,(\bar{i}-\mu)$$

where n = the number of lines crossed with inbred i, $\bar{i}$ is the mean of parental line i, and μ is the mean of the TCSMs, calculated in a analogous way as for GCA (Figure 5.1). Linear regression of the GCAs of the parental lines calculated for the trait of interest on the GCSMs results in an additive model for the prediction of the GCA.

An 'expected' TCSM of the hybrid from a cross between the *i*th female and the *j*th male can now be calculated as,

$$ETCSM_{ij} = \mu + GCSM_i + GCSM_j$$

Models for the prediction of SCA

There are two alternative models for the prediction of SCA based on selected markers: 1) The difference between the calculated and 'expected' TCSM for a hybrid results in an estimation of the SCSM of the two parental lines in combination. Linear regression of the SCAs of the hybrids calculated for the trait of interest on the SCSMs results in a **first model for the prediction of the SCA.**

2) In a way similar to finding markers significantly associated with HP, markers associated with SCA can be selected. The estimates of a_l and d_l of the marker alleles selected as being significantly associated with SCA are used to calculate a TCSM value of any hybrid. Linear regression of the SCAs on the TCSMs results in a **second model for the prediction of the SCA.**

Allelic divergence among groups

Allelic divergence (*ald*) among groups of inbreds at the marker loci and the QTL produces linkage disequilibrium between marker loci and QTL involved in SCA

(Charcosset and Essioux 1994). Since specific heterotic groups like BSSS and LSC have been classified on the basis of intra- and interheterotic heterosis, these groups should differ for their allelic frequencies at the QTL that exhibit dominance effects. Since we were able to determine group membership of the parental inbreds, allelic divergence among the two major groups for the markers showing sginificant association with SCA has been calculated as follows:

$$ald = (\mid f_1 - f_2 \mid)$$

where f_1 and f_2 are the allelic frequencies at the marker locus in group 1 (BSSS) and group 2 (LSC), respectively. High *ald* values must 1) provide evidence for the correlation between SCA and heterozygosity at marker loci, and 2) support the linkage association between the selected marker loci and QTL exhibiting dominance effects. The allelic frequency at the marker locus in the third group containing the two parental lines of miscellaneous origin, was left out of consideration.

Evaluation of the model for the prediction of HP

A first type of cross-validation performed to evaluate the additive model for prediction of HP is by a jackknife sampling procedure. The jackknife sampling procedure requires the partition of the initial set of N hybrids into 1) a set of N - 1 predictor hybrids used for parameter estimation and 2) one 'removed' hybrid used to compare predicted HP with observed HP. Evaluation of prediction efficiency was made by examining plots of observed vs. predicted values and two synthetic statistics: 1) the standard error (SE) estimated as:

$$SE = \sqrt{\ } s^2 \times \left(1 + \frac{1}{n} + (x_0 - \bar{x})^2 / \sum (x_i - \bar{x})^2\right)$$

where s^2 is the sample variance, n is the number of observations, x_0 is the predicted value, $\bar{x}$ is the mean of the observed values and x_i is the observed value

i; 2) the coefficient of determination (r^2, squared correlation coefficient) between observed and predicted values. The minimum value SE can reach is equal to σ, since a new hybrid will show some variation around the regression, equal to at least the residual variance σ^2. A second source of variation to be taken into account is the inaccuracy of the regression line: the estimates of the regression coefficients are stochastic, since they are based on a limited set of observations.

A second type of cross-validation to evaluate the additive models for prediction of HP is by linear regression of the HP of additional single crosses on their corresponding TCSM. In this study, sixteen parental testcrosses, obtained by testing eight BSSS inbreds (Lo999, N28, A1, A2, A3, A4, A5 and A8 (Table 4.1)) against Lo881 and C103 (LSC testers), were chosen. Evaluation of prediction efficiencies were made by examining plots of observed vs. predicted values and the corresponding r^2 values.

Results

Relationship of genetic distance to HP and SCA

The estimates of linear correlations (*r*) of GD and SGD calculated from different marker data sets with HP and SCA for GY, respectively, are presented in Table 5.1. It must be emphasized that the results obtained from the BSSS × BSSS and LSC × LSC groups of crosses, although these groups are of minor interest, should be interpreted with caution because of their small number of observations. The *r* values of GD with HP for the entire set of 78 hybrids were highly significant ($P<0.001$) but of moderate size (0.36-0.52); the highest *r* value (0.52) was obtained using the P/M marker data set. By contrast, a lack of relationship was noted between GD and HP in the three subset of crosses. Estimates of *r* values of SGD with the SCA effect were for the entire set of crosses and the BSSS×BSSS subset highly positive (0.80-0.84 and 0.83-0.86, respectively) and highly significant ($P<0.001$); the highest *r* values (0.84 and 0.86, respectively) were obtained using the "P/M marker data set. In addition, highly significant ($P<0.001$)

and of a high magnitude were the correlations found in the subset of unrelated lines (0.61-0.68). Again the highest *r* value (0.68) was obtained using the ^{m}P/M marker data set. The *r* values of GD and SGD calculated from different marker data sets with HP and SCA for GY calculated from the 78 F_1 data from the half-diallel only, were similar to those reported by Ajmone Marsan et al. (1998) (data not shown).

Table 5.1 Linear correlations of genetic distance (GD) and specific genetic distance (SGD) based on *EcoRI/MseI* (E/M), *PstI/MseI* (P/M) and ^{m}*PstI/MseI* (^{m}P/M) data, respectively, with hybrid performance (HP) and specific combining ability (SCA) for grain yield, for the total set of 78 single crosses and for different subsets of single crosses.

Variables	Crosses (n)			
	All (78)	BSSSxBSSS (15)	LSCxLSC (10)	Unrelated Lines (53)
		HP		
GD - E/M	0.47***	0.22	0.38	0.16
GD - P/M	0.52***	0.33	0.24	0.25
GD - ^{m}P/M	0.36***	0.41	0.11	0.08
GD – Tot	0.48***	0.30	0.28	0.19
		SCA		
SGD – E/M	0.80***	0.85***	0.58	0.61***
SGD – P/M	0.80***	0.83***	0.45	0.63***
SGD – ^{m}P/M	0.84***	0.86***	0.68*	0.68***
SGD – Tot	0.81***	0.86***	0.59	0.64***

***, *: Significant at the 0.001 and 0.05 probability level, respectively

Prediction of Hybrid Performance for GY

Table 5.2 gives an output list of the 20 marker alleles selected as being significantly ($p < 0.001$) associated with QTL alleles contributing to GY, as well as their corresponding map position, 'origin', Kruskal-Wallis test statistic, the arithmetic means of the quantitative trait for the three genotypic classes and the *a* and *d* values. It is clear from the *a* en *d* values that the marker alleles selected for

HP for GY fit single gene models with additive and partial dominance effects ($0 \leq |d| < |a|$).

The selected markers are clearly confined to particular regions of chromosomes, rather than being evenly distributed across the entire maize genome (residing on 8 of the 10 chromosomes). Only one putative QTL of GY was revealed on 1/94 where Mo17 contributed the superior allele. In contrast, B73 contributed the superior allele at the putative QTL on 4/56.2-58.0, 5/20.1, 6/10.3-10.8, 6/64.7-68.4, 8/124.6 and 9/54.1.

Table 5.2 Map position, 'origin', Kruskal-Wallis test statistic (K), the means for the three genotypic classes and the *a* and *d* effects for the 20 marker alleles selected as being significantly ($p < 0.001$) associated with QTL alleles contributing to the hybrid performance for grain yield, across the 53 interheterotic crosses.

nr.	map position	marker	origin[a]	K	mm	Mm	MM	*a*	*d*
1	1/94.5	E42M4750-	Mo17	15.24 ****	10.39	11.80	12.48	1.04	0.37
2	2/24.5	P13M60-59.0	Mo17	16.06 ****	12.94	11.54	10.35	-1.3	-0.1
3	2/30.4	P12M50-146.4	Mo17	15.52 ****	12.23	10.93	10.14	-1.05	-0.26
4	2/53.4	P18M60-419.7	Mo17	13.88 ***	11.72	10.46	9.27	-1.23	-0.04
5	3/49.5	mP12M48-587.1	Mo17	17.21 ****	11.91	10.47	10.14	-0.88	-0.56
6	4/56.2	E35M5154-82.0	B73	19.28 *****	10.03	11.43	12.99	1.48	-0.08
7	4/58.0	E38M51-139.7	Mo17	19.33 *****	12.16	10.87	9.10	-1.53	0.24
8	4/58.0	E39M54-314.1	Mo17	19.33 *****	12.16	10.87	9.10	-1.53	0.24
9	4/58.0	E35M50-415.8	B73	17.04 ****	10.17	11.03	12.45	1.14	-0.28
10	5/20.1	P12M61-93.5	B73	18.19 ****	10.26	11.02	12.48	1.11	-0.35
11	6/10.3	E39M62-168.6	B73	15.22 ****	10.23	11.06	12.15	0.96	-0.13
12	6/10.8	E39M47-413.1	B73	23.81 *****	10.21	10.73	12.40	1.09	-0.58
13	6/10.8	E45M60-582.5	B73	14.22 ***	10.50	11.11	12.47	0.99	-0.38
14	6/64.7	P13M62-473.2	B73	21.21 *****	10.42	11.78	13.37	1.47	-0.12
15	6/68.4	E42M6061-	B73	16.89 ****	10.37	11.17	12.61	1.12	-0.33
16	8/124.6	E33M50-148.9	B73	13.99 ***	10.27	10.69	11.96	0.85	-0.65
17	8/124.7	E33M50-148.1	Mo17	17.21 ****	11.91	10.47	10.14	-0.88	-0.56
18	9/0.0	E39M50-174.0	Mo17	18.86 *****	12.62	11.16	10.27	-1.17	-0.28
19	9/54.1	E38M51-71.6	B73	20.22 *****	9.74	11.36	12.35	1.31	0.32
20	9/58.6	E35M50-228.1	Mo17	13.22 ***	11.94	10.90	8.90	-1.52	0.48

*****, ****, ***: Significant at the 0.0001, 0.0005 and 0.001 probability level, respectively

[a] To B73 and Mo17 will be referred as 'origin' of a selected marker allele, when the AFLP marker has been identified as a B73 or Mo17 marker, respectively, mapped on the B73 × Mo17 RI linkage map (Chapter 3).

The *r* values of TCSM with HP for GY for the 53 interheterotic crosses are reported in Table 5.3.

Table 5.3 Linear correlations of hybrid performance (HP), general combining ability (GCA) and specific combining ability (SCA) with the total, general and specific contribution of selected markers (TCSM, GCSM, SCSM), respectively, for interheterotic crosses only. Different TCSMs, GCSMs and SCSMs are calculated corresponding to the significance level used in selecting markers.

	# markers	HP/TCSM	GCA/GCSM	SCA/SCSM
P<0.001	20	0.79***	0.88***	0.49***
P<0.0005	16	0.78***	0.88***	0.47***
P<0.0001	7	0.77***	0.87***	0.48***

***: Significant at the 0.001 probability level

These *r* values calculated for different numbers of selected markers, are very highly significant (P<0.001) and of a much higher magnitude than the *r* values of HP with GD (Table 5.1). Simultaneous fit of all putative QTL accounted for 62.4% of the phenotypic variance among the 53 hybrids (Table 5.3; Figure 5.2), while still 59.3% of the phenotypic variance among the 53 hybrids is explained by only 7 markers.

The cross Lo881 × Lo951, of which GY is amongst the highest (12.94 t ha^{-1} at 15.5% moisture), has the highest $TCSM_{0.001}$ value (23.65) for GY among the 53 hybrids (data not shown). The maximal $TCSM_{0.001}$ value that can be reached, based on the maximal contribution of each selected marker listed in Table 5.2 equals 23.65 ($\sum_{i=1}^{20}|a_i|$). This means there is no additional gain in GY possible using the QTL detected in the germplasm under consideration.

Figure 5.2 Observed hybrid performance (t ha^{-1} at 15.5% moisture) vs. the total contribution of the markers selected at a significance level of 0.001 ($TCSM_{0.001}$). The 53 interheterotic crosses are considered. The straight line represents the linear regression of hybrid performance on $TCSM_{0.001}$.

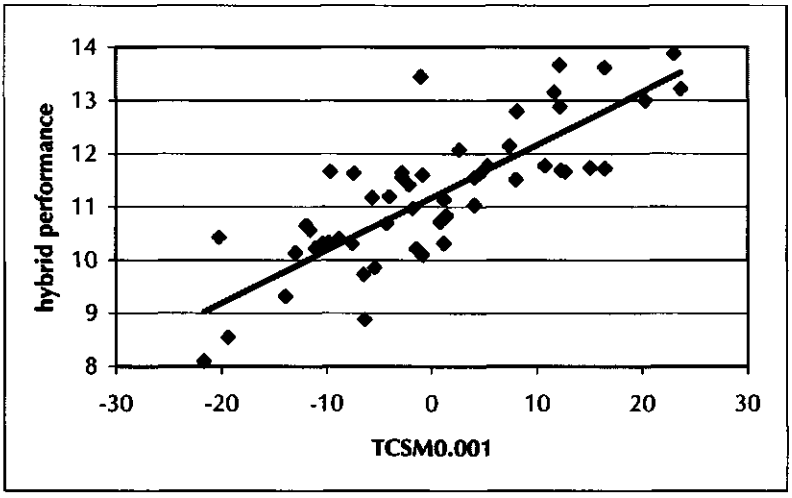
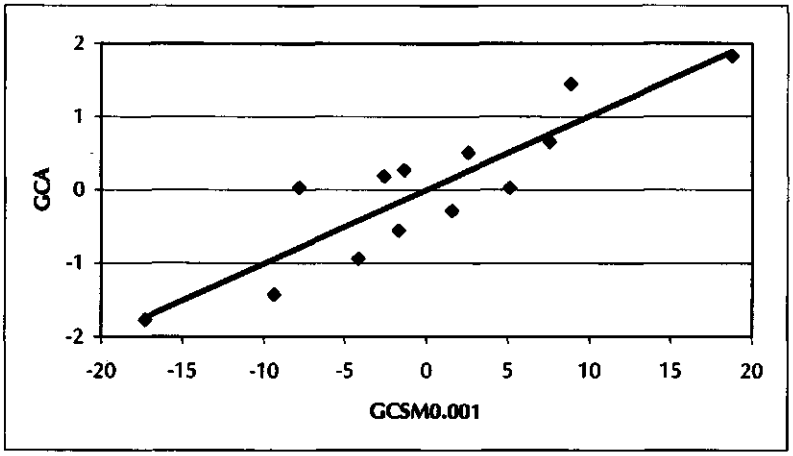


Figure 5.3 General combining ability (GCA) (t ha^{-1} at 15.5% moisture) vs. the general contribution of the markers selected at a significance level of 0.001 ($GCSM_{0.001}$). The 13 inbred lines are considered. The straight line represents the linear regression of GCA on $GCSM_{0.001}$.



Prediction of GCA for GY

The *r* values of GCSM with GCA for GY for the 13 inbred lines, calculated for different numbers of selected markers, are very highly significant ($P<0.001$) and are reported in Table 5.3. Simultaneous fit of all putative QTL accounted for 77.4% of the GCA variance among the 13 parental lines (Table 5.3; Figure 5.3), while still 75.7% of the GCA variance is explained by the 7 markers with the highest significance level.

Prediction of SCA for GY

The *r* values of SCSM with SCA for GY for the 53 interheterotic crosses, calculated for different numbers of selected markers, are very highly significant ($P<0.001$) but of moderate size (0.47-0.48) (Table 5.3) and of a lower magnitude than the *r* values of SGD with SCA (Table 5.1). Table 5.4 gives an output list of the 25 marker alleles selected as being significantly ($P<0.005$) associated with QTL alleles contributing to SCA for GY as well as their corresponding map locus, 'origin', Kruskal-Wallis test statistic, the arithmetic means of the quantitative trait for the three genotypic classes and the *a* and *d* values. It is clear from the *a* and *d* values that the marker alleles selected for SCA for GY fit single gene models with only overdominance effects ($|d| > |a|$). Again, the selected markers are confined to particular regions of the chromosomes, rather than being evenly distributed across the entire maize genome (they reside on 7 of the 10 chromosomes). The selected markers are showing positive overdominance (e.g. 1/74.7), as well as negative overdominance (e.g. 1/53.1). Where positive overdominance rules, the superior allele originates evenly from B73 and Mo17. Simultaneous fit of the 25 selected marker alleles accounted for 36.8% of the SCA variance among the 53 hybrids, which is of a higher extent than explained by SCSM, but less than explained by SGD.

All of the selected markers but two show a high *ald* value (either 80 or 100). These high *ald* values are in good agreement with the hypothesis that heterotic groups should differ for their allelic frequencies at the QTL that exhibit dominance effects, or at the marker loci tightly linked to these QTL (Charcosset and Essioux

1994). It is clear that divergence of allelic frequencies among groups for the QTL or marker loci in linkage disequilibrium with these QTL, produces correlation between heterosis and heterozygosity at marker loci.

Table 5.4 Map position, 'origin', Kruskal-Wallis test statistic (K), the means for the three genotypic classes, the *a* and *d* effects and the allelic divergence (*ald*) for the 25 marker alleles selected as being significantly ($p < 0.005$) associated with QTL alleles contributing to the specific combining ability for grain yield across the 53 interheterotic crosses.

nr.	map position	marker	origin*	K		mm	Mm	MM	*a*	*d*	*ald*
1	1/53.1	E33M50-169.1	B73	11.420	**	0.193	-0.177	0.443	0.125	-0.495	80
2	1/69.0	P12M50-278.2	Mo17	11.420	**	0.443	-0.177	0.193	-0.125	-0.495	80
3	1/74.7	P18M49-117.7	B73	11.093	**	-1.041	0.147	-0.135	0.453	0.734	100
4	2/70.4	E45M60-137.3	Mo17	11.093	**	-0.135	0.147	-1.041	-0.453	0.734	100
5	3/12.2	P12M61-154.0	Mo17	11.206	**	1.019	-0.156	0.073	-0.473	-0.702	100
6	3/34.6	E35M49-58.9	B73	11.093	**	-1.041	0.147	-0.135	0.453	0.734	100
7	3/70.4	E45M48-257.8	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
8	3/91.4	P12M61-159.3	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
9	3/96.6	P18M49-261.7	Mo17	11.093	**	-0.135	0.147	-1.041	-0.453	0.734	100
10	6/0.0	P18M48-136.8	B73	10.965	**	0.178	-0.144	0.881	0.351	-0.673	100
11	6/1.7	P18M48-142.1	Mo17	11.420	**	0.443	-0.177	0.193	-0.125	-0.495	80
12	6/11.4	"P12M49-148.9	B73	11.093	**	-1.041	0.147	-0.135	0.453	0.734	100
13	8/38.8	E39M60-394.1	Mo17	12.304	**	-0.164	0.236	-1.217	-0.527	0.926	60
14	8/55.7	"P12M61-507.8	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
15	8/57.1	E33M47-92.5	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
16	8/67.4	P18M48-196.9	B73	12.304	**	-1.217	0.236	-0.164	0.527	0.926	60
17	8/72.2	P13M49-226.9	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
18	8/74.6	P12M47-181.6	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
19	8/74.6	P12M47-178.4	B73	10.965	**	0.178	-0.144	0.881	0.351	-0.673	100
20	9/20.9	E38M51-153.0	Mo17	11.093	**	-0.135	0.147	-1.041	-0.453	0.734	100
21	9/58.6	E33M50-69.4	B73	10.965	**	0.178	-0.144	0.881	0.351	-0.673	100
22	9/58.6	E35M50-228.1	Mo17	12.304	**	-0.164	0.236	-1.217	-0.527	0.926	60
23	10/25.4	E35M58-103.0	Mo17	10.965	**	0.881	-0.144	0.178	-0.351	-0.673	100
24	10/37.7	P12M61-185.6	Mo17	11.420	**	0.443	-0.177	0.193	-0.125	-0.495	80
25	10/54.0	P13M60-96.1	Mo17	11.420	**	0.443	-0.177	0.193	-0.125	-0.495	80

**: Significant at the 0.005 probability level

*To B73 and Mo17 will be referred as 'origin' of a selected marker allele, when the AFLP marker has been identified as a B73 or Mo17 marker, respectively, mapped on the B73 × Mo17 RI linkage map (Chapter 3).

Evaluation of the additive model for prediction of HP

Performing the cross-validation of the additive models for prediction of HP by a jackknife sampling procedure, shows that the highest efficiency for prediction of HP of GY was reached when considering the selected markers at a significance

level of $P<0.001$ only (Table 5.5; Figure 5.4). In this situation, HP predicted by the model explained 45.1% of the variation observed. The 13 hybrids with a predicted HP value ≥ 12.0 t ha^{-1} at 15.5% moisture, had a observed value ≥ 11.5 t ha^{-1} at 15.5% moisture. The corresponding mean SE of predicted vs. observed values was 0.88 t ha^{-1} at 15.5% moisture. The ratio between SE and the total range of variation that was observed (7.5 to 13.88 t ha^{-1} at 15.5% moisture), and the fact that most of the best single crosses are identified, suggests that prediction based on the TCSM model is highly efficient for a preliminary screening of test hybrids before field evaluation. Note that cross-validation of the prediction model by a jackknife procedure involves only those hybrids forming part of the 13 by 13 half-diallel.

Cross-validating the additive model for prediction of HP by linear regression of the HP of the 16 parental testcrosses on their corresponding TCSM, the highest efficiency (r^2 = 33.0%; Table 5.5; Figure 5.5) was reached by simultaneous fit of the 20 selected marker alleles, given in Table 5.2. Note that here, in contrast with the cross-validation by jackknifing, the prediction model is evaluated using hybrids of which only one parent (the LSC tester) is forming part of the 13 by 13 half-diallel. Despite a moderate r^2 value, Figure 5.5 shows that the best single crosses are identified. This suggests that prediction, based on the a_i and d_i estimates of the 20 markers selected at $P<0.001$ (Table 5.2) is efficient as a preliminary screening of related test hybrids before field evaluation.

Table 5.5 Coefficient of determination (r^2) between observed and predicted hybrid performance (HP) of a) the 53 hybrids forming part of the 13 by 13 half-diallel and b) 16 hybrids partly related to the 13 by 13 half-diallel; the corresponding mean standard error (SE) and the empirical standard deviation of SE over the 53, resp. 16 cross-validations within brackets.

	HP			
	53 hybrids		16 hybrids	
	r^2(%)	SE	r^2(%)	SE
P<0.001	45.1	0.88 (0.03)	33.0	0.842(0.002)
P<0.0005	42.1	0.90 (0.03)	15.7	0.868(0.008)
P<0.0001	35.3	0.92 (0.02)	11.2	0.883(0.009)

Figure 5.4 Observed vs. predicted performance (t ha^{-1} at 15.5% moisture) of maize hybrids based on 52 predictor hybrids used for parameter estimation (jackknife sampling procedure), considering the selected markers at a significance level of $P<0.001$ (Table 5.2). The 53 interheterotic hybrids are considered. The straight line represents the predicted = observed equation.

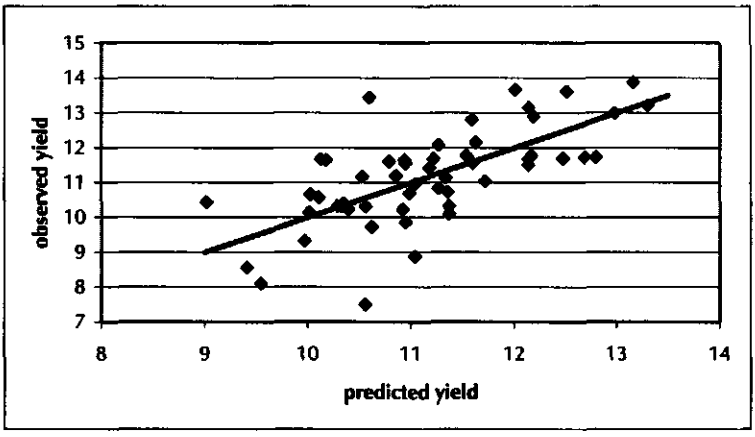
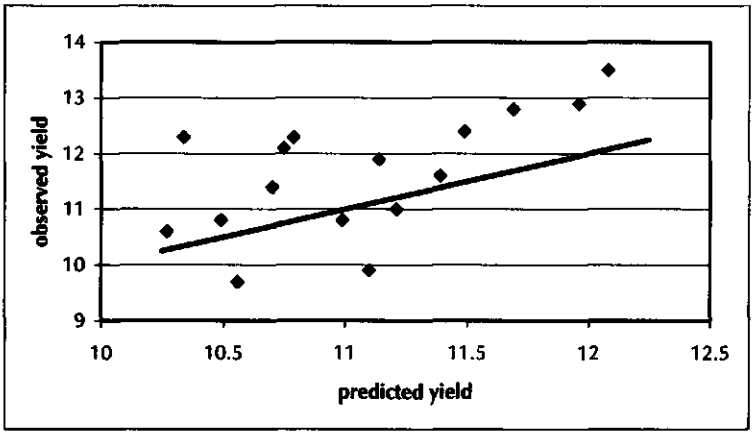


Figure 5.5 Observed vs. predicted performance (t ha^{-1} at 15.5% moisture) of the 16 hybrids partly related to the 13 by 13 half-diallel, based on the selected markers at a significance level of $P<0.001$ (Table 5.2). The straight line represents the predicted = observed equation.



Discussion

In good accordance with published results (Frei et al. 1986; Lee et al. 1989; Godshalk et al. 1990; Melchinger et al. 1990 a, b; Smith et al. 1990; Dudley et al. 1991; Melchinger et al. 1992; Boppenmaier et al. 1993; Burstin et al. 1995; Ajmone Marsan et al. 1998), estimates of the GD between parents did not consistently identify the best crosses, particularly not when the two parents are non-related lines. On the contrary, correlations between SGD and SCA found in the subset of unrelated lines were highly significant ($P<0.001$) and of a high magnitude suggesting practical utility in predicting SCA effects. Differences between our results and those reported by Ajmone Marsan et al. (1998), are solely due to differences in field data. Linear correlation values of SGDs calculated from different marker data sets with SCA calculated from the 78 F_1 data from the half-diallel only, were similar to those reported by Ajmone Marsan et al. (1998). These results stress the need for field data of hybrids available from multiple trials carried out in several different environments and years to reduce the error variance.

The established method for identification of QTL is the selection of two parents, representing a fraction of the germplasm used by the breeder, that differ markedly in one or more (related) particular quantitative traits of interest, followed by the determination of associations between markers and that character in a segregating population. However, the segregational resolution of conventional segregating populations (e.g. F_2, BC) is too low to distinguish tightly from less tightly QTL-marker combinations. The approach to QTL identification followed in the present study has the following potential advantages over QTL detection in a segregating population. Firstly, marker-trait associations are only expected to be found in case a marker is tightly linked to a QTL. This is because across a set of lines, associations between QTL and loosely linked markers will be non-existent due to accumulated recombination events during the establishment of the lines. Basically, the type of associations we have identified are due to identity by descent of QTL and marker alleles across lines. Secondly, only a limited number of lines representing the gene pool used by the breeder need to be genotyped. And thirdly, it may allow the detection of QTL that vary across a wide spectrum of the

germplasm used. The possible advantages are not easily generalized since especially the joint identity by descent of alleles of linked loci depends on factors that are largely unknown for most germplasm collections, i.e. the number of generations since the descent from a common ancestor and the amount of exchange between lines of descents by crossing in the past.

The *a* en *d* values of the markers selected for HP for GY indicate that QTL with additive to partial dominance effects are prevalent. Although the magnitude of the genetic effects for any single QTL contributing to GY can vary considerably, the joint added contribution of single QTL involved in GY explains 59.3 to 62.4% of the HP variance. Cross-validation of the prediction efficiency of the TCSM model for HP showed that the best crosses were identified, suggesting that the TCSM-approach is efficient as a preliminary screening of test hybrids before field evaluation. The higher prediction efficiency of the TCSM model in comparison with the 'distance' model can be explained as follows: rather than converting molecular polymorphism between inbred lines, having direct or no direct effect on the trait of interest, into the metric GD, only specific markers, supposed to be linked to loci that affect the quantitative trait of interest, were selected. Hybrids heterozygous for marker loci significantly associated with SCA, show often higher GY than hybrids homozygous for those marker loci. This pattern may be due to either true overdominance (i.e. particular single loci at which the heterozygote phenotype exceeds that of either homozygote), epistasis or pseudo-overdominance (i.e. closely linked loci at which alleles having dominant or partially dominant advantageous effects are in repulsion phase). With more than one QTL linked to the marker, epistatic effects modify the additive and dominance effects or pseudo-overdominance results. Although all QTL were detected at marker loci and deliver in this way the maximal genetic information, and an extensive dissociation of alleles at linked loci is most likely represented in the inbred lines, still our results can not distinguish these possibilities.

Supposing that the joint effect of multiple QTL involved in the heterotic response of GY is additive, 36.8% of the SCA variance among the 53 hybrids can be explained, which is of a higher extent than explained by SCSM (22.1% - 24.0%),

and almost equal to what is explained by SGD (37.2 – 46.2%). The high *ald* value of the marker alleles showing significant association with the SCA effects is consistent with the fact that 1) allelic divergence among groups at the marker loci and the QTL involved in the trait of interest produces linkage disequilibrium between marker loci and QTL involved in SCA (Charcosset et Essioux 1994), or 2) the process of inbreeding and selection by which lines are commonly developed generates linkage disequilibria by favoring linkage phases and by fixing them in homozygous condition (Weir et al. 1972).

Since many metabolic processes in the maize plant ultimately affect GY, it seems very likely that multiple genetic factors are involved in the inheritance of GY. Finding a large number of QTL is not surprising in view of the complex nature of the trait/phenotype. However, the aim of the breeder is to accumulate in the same genotype the maximum number of favorable genes. Since all putative QTL mentioned in Table 5.2 show additive to partial dominance gene effects, fixing one or more of the favorable QTL alleles in the inbred line is desirable. More than GCSM, the TCSM *per se* value of an inbred line is suited to monitor the improvement of an inbred line by fixation of favorable alleles and, subsequently, marker-assisted selection, since the TCSM *per se* value can be calculated directly from the genotype of the inbred line.

Although direct comparisons of QTL are complicated by differences in parental lines, design of the cross, number of progeny and the environments in which the progeny was assessed, as well as by different marker loci and QTL detection methods, other reports have identified some of the same regions detected in the present study to be associated with GY. Austin and Lee (1996; 1998) detected GY QTL on *5S/umc72*, *6L/bnl5.47-npi280* and *8L/umc7* that were also associated with GY in the present study (*5S/20.1*; *6L/64.7-68.4*; *8L/124.6-124.7*). Also Zehr et al. (1992) reported the GY QTL on *6L*, showing marker association with *umc38a*. Another GY QTL reported by Zehr et al. (1992) was associated with *umc44* on *2S*, likely to coincide with *2S/53.4* in our study. Ajmone Marsan et al. (1995) reported a major GY QTL associated with *umc051* on chromosome 6, which is in the vicinity of the putative QTL on *6C/10.3-10.8* found in our study. Another GY QTL,

found on *9C/54.1-58.6*, is likely to co-incide with the GY QTL found on 9C by Stuber et al. (1992), Zehr et al. (1992) and Ajmone Marsan et al. (1995). Finally, Ajmone-Marsan et al. (1995) also detected a GY QTL in the interval *4L/umc42-umc19*, associated with GY in the present study (*4C/56.2-58.0*). Beside the agreement in chromosomal location for a few GY QTL, agreement in origin of the superior and inferior allele is present only for the QTL on *4C*, *9C*, and on *2S*, respectively. Of the chromosomal regions selected as being significantly associated with QTL contributing to the SCA for GY, one (*9/58.6:E35/M50-228.1*) was also selected as being significantly ($P<0.001$) associated with a QTL contributing to the HP for GY in this study, and four (*1/53.1*, *1/69.0*, *2/70.4* and *10/37.7*) were associated with heterosis for GY in Stuber et al. (1992).

The efficiency of the TCSM method as a prediction method and/or as a QTL screening method may be increased by fulfilling the following requirements: 1) a higher marker density on the map will allow to pick up more marker alleles tightly linked to specific QTL, consolidating already identified QTL or identifying new putative QTL. Where a higher marker density can be obtained by intensifying the mapping efforts, a higher number of specific alleles per locus can be obtained by integrating linkage maps covering different genomes; 2) a more reliable and easier evaluation of the effect of a QTL allele will be obtained when the three genotypic classes are more equally represented. This balance can be obtained by enlarging the half-diallel; and 3) yield data of hybrids, available from multiple trials carried out across different environments and years are highly desirable in order to reduce the phenotypic variance, representing a gain in accuracy.

Acknowledgements

We would like to thank Dr. F. Salamini and Dr. M. Zabeau for the initial idea for this work, and Dr. M. Motto for providing us seeds of the inbred lines and the grain yield data of the hybrids. This work was financed by Keygene N.V. and 5 Dutch breeding companies: Cebeco Zaden, De Ruiter, ENZA Zaden, Rijk Zwaan and Van der Have.

Note: AFLP® is a registered trademark of Keygene N.V.

Note: The methylation AFLP® method is subject to a patent application filed by Keygene N.V.

6

Summarizing conclusion

In the 1980s, intensive research started using molecular markers to characterize and utilize genetic diversity for major crop plants such as maize (*Zea mays* L.). Nowadays, molecular markers are a routine component of 1) the genetic mapping of traits of high heritability and/or economic worth, and 2) backcrossing monogenic traits, thereby helping to more efficiently and effectively organize, combine and select new genotypic combinations. For purposes of varietal identification and the protection of Plant Breeders' Rights, molecular marker technology plays a preeminent role in the provision of data. So, molecular markers already are indispensable in some applications in maize breeding, but for other applications there is hesitancy to apply them. Nonetheless, molecular markers provide additional ancillary and important support for agriculture.
It is important to determine the obstacles that need to be overcome to facilitate the application of molecular markers in some applications in maize breeding.

Rapid map construction

In rapid (trait) mapping and marker based selection applications, emphasis must be placed upon attaining reduced time costs of data acquisition. Throughput using the time and labor intensive restriction fragment length polymorphism (RFLP; Botstein et al. 1980) assay is limiting, especially during progeny selection with time constraints imposed prior to pollination or harvest. For that reason, plant breeders are increasingly using PCR-based methods such as randomly amplified polymorphic DNA (RAPD, Williams et al. 1990), simple sequence repeats (SSR, Tautz et al. 1989) and the AFLP® technique (Vos et al. 1995), since these can provide faster and more time and cost effective (trait) mapping and marker assisted progeny selection.

In Chapter 3, it is shown that the high multiplex ratio of the AFLP technique offers the potential to improve the efficiency of genetic map construction in maize and to generate high-density maps around loci that control important traits in maize. The use of the AFLP method has led to the generation of two "second-generation" high-density AFLP linkage maps of maize based on: 1) a B73 × Mo17 Recombinant Inbred (RI) population and 2) a D32 × D145 Immortalized F_2 (IF_2) population. The IF_2 map spans 1376 cM and the lengths of most of its linkage groups are systematically longer than in the RI map (1178 cM). The number of mapped polymorphic AFLP fragments in the IF_2 and RI map are 1355 and 1539, respectively.

AFLP markers were generated using the enzyme combinations *Eco*RI/*Mse*I (E/M) and *Pst*I/*Mse*I (P/M). Among the mapped P/M AFLP markers, ^{m}P/M markers, fragments bounded by a methylated *Pst*I site, are included. ^{m}P/M markers were generated using the methylation AFLP® technique (Chapter 2), a novel PCR-based method to detect methylation of restriction sites randomly over the genome. The methylation AFLP assay allows the exploitation of two forms of DNA polymophisms: 1) variation in the primary nucleotide sequence either in the methylated restriction site or in the fragment size (^{m}AFLP markers) and 2) allele-specific methylation polymorphism (^{asm}AFLP markers).

By monitoring the segregation of both forms of DNA polymorphism in the RI and IF_2 populations, it was shown that DNA methylation inherits in a Mendelian fashion by the offspring and co-segregates in perfect accordance with the primary target site. In addition, simultaneous mapping of P/M and ^{m}P/M markers, obtained with complementary P/M PCs, allowed the identification of epi-alleles, showing allelic variation in the CpNpG methylation only. The percentage of methylation polymorphism in both mapping populations was low. Only three pairs of markers of a total of 673 and 595 mapped $^{(m)}$P/M markers (obtained with the 14 complementary P/M PCs), behaved like epi-alleles in the RI and IF_2 map, respectively. This level of DNA methylation variation is low in comparison with the high degree of sequence polymorphism detected by the AFLP method, and can be explained as follows: a point mutation in any of the 15 nucleotides targeted by

a *Pst*I+2/*Mse*I+3 primer combination (PC) will result in AFLP marker, while DNA methylation variation is restricted to the two cytosine residues of the *Pst*I restriction site only. Thus, using a *Pst*I+2/*Mse*I+3 PC, the probability of detecting DNA methylation variation is about eight times lower than detecting a point mutation. However, taking this into account, DNA methylation variation occurs still at a significant lower rate than DNA sequence variation. This would 1) indicate that methylation polymorphism at the CpNpG target site is not common in maize, and 2) show the high accuracy of the transmission of a given pattern of ^{5m}C, accounting for the specificity of the pattern that can be highly conserved between individuals.

AFLP markers segregating in the IF_2 mapping population were scored as co-dominant markers, using proprietary software developed specifically for AFLP analysis, at Keygene N.V. Quantitative analysis of AFLP marker bands in order to differentiate heterozygotes from both homozygous classes makes co-dominant scoring of AFLP markers feasible.

Taking the present studies as an example, about 26,000 data points (one data point = presence or absence of a band/lane) can be produced by one person, in one working day. With a mean absolute effective mapped multiplex ratio (EMM) (defining the number of mapped polymorphic fragments simultaneously analyzed in a single assay) of 19 and 22 for the IF_2 and RI map, respectively, 21% and 24% of the total amount of data points produced, respectively, are valuable. So, the marker data for a 1,500 AFLP marker linkage map can be acquired within 6 weeks by one person. By choosing P/M PCs only, with a mean EMM of 27 and 26 for the RI and the IF_2 map, respectively, the mapping efforts will be further reduced by 3-4 days.

In addition, investigation of the genetic location on the two linkage maps of the AFLP markers revealed that P/M markers are more uniformly distributed over the genetic map than E/M and ^{m}P/M. Clusters of E/M and ^{m}P/M markers co-localized well with the putative centromeric regions of maize. This is in accordance with the AT-richness of centromeres and the enrichment of methyl groups in the maize heterochromatin, consisting of particular regions that are not transcribed, like

centromeric regions, the nucleolus organizer region, telomeres and knobs. So, by using P/M PCs, the mapping process will not only be accelerated, but also a better coverage of the genome will be attained. Equal representation of the genome by genetic markers is of special interest in marker-assisted backcrossing: markers uniformly spread on the regions unlinked to the introgressed region are necessary to ensure that the genetic material from the recurrent parent is carried forward.

The AFLP technique also has been used to improve the efficiency of genetic map construction of other crops like barley (Becker et al. 1995; Qi et al. 1998), potato (van Eck et al. 1995), rice (Maheswaran et al. 1997; Zhu et al. 1998), soybean (Keim et al. 1997) and sugar beet (Schondelmaier et al. 1996). Even when no inbred lines of homozygous individuals are available, like for forest trees and potato, where standard analyses of backcross and inbred F_2 families for genetic studies are not feasible, the AFLP analysis has proven to be a rapid and efficient technique (van Eck et al. 1995; Marques et al. 1998).

The usefulness of a genetic map in top-down anchoring of physical maps and mapping commercially important traits like grain yield depends on the accurate determination of locus order. With the advent of AFLP markers, the bottleneck of low and inefficient production of segregation data in quickly building high-resolution genetic maps has been overcome. Because of the large number of possible locus orders (n!/2 for n loci) on a linkage map, the major problem is shifted to the 'locus ordering' with increasing numbers of markers. Therefore, a tool to order large numbers of markers in an accurate and efficient way is becoming the key factor in generating high-density linkage maps.

Finally, in the near future, with the availability of a huge amount of DNA sequences, co-ordination between the genetic and physical maps and DNA sequences may provide more efficient ways to identify and isolate genes controlling complex traits. So, genetic and physical maps will be bridges between complex traits and DNA sequences.

DNA-methylation

An extension of the AFLP assay is the methylation AFLP, a novel PCR-based method to detect DNA methylation. It shares the characteristics of the AFLP method such as genome-wide detection, no need for prior sequence knowledge, availability of only a limited set of generic primers, and a high multiplex ratio. However, the power of the methylation AFLP technique resides especially in its positive display of the native methylated sites, avoiding difficulties in interpretation.

In Chapter 2, we demonstrated that most moderately to highly repetitive DNA sequences are strongly methylated. Methylation of repetitive sequences, including transposable elements, is thought to serve as a genome-defense mechanism that guards against the deleterious effects of multicopy transposable elements (Bender 1998). So, in this respect, mapping ^{m}AFLP markers might localize insertions of cryptic transposons, which might contribute to the identification of genes (Martienssen 1998).

In addition, the methylation AFLP technique has the potential to shed light on the distribution of DNA methylation on the genetic map as well as on the physical map. Furthermore, the assay can furnish sets of epi-alleles near or in any locus of interest. DNA methylation is emerging as an important component of cell memory or genomic imprinting, the process by which dividing cells inherit states of gene activity, exhibiting a parent-of-origin effect. Therefore, detection of epi-alleles may become a prerequisite to identify genes involved in the violation of Mendel's law of inheritance.

Varietal identification

Apart from variety identification *sensu stricto* the use of molecular markers for identification purposes encompasses several other goals such as determining genetic distances between inbred lines, typing germplasm for patent protection, and assigning inbred lines to heterotic groups. The emphasis in varietal identification is on genotyping a large number of individuals with relatively small set of defined markers a in order to meet a given level of precision in the

estimation of genetic distances among the individuals. Information content and multiplex ratio of the marker assay, as well as an equal distribution of the markers on the genetic map, therefore become an issue for this application.

For purposes of varietal identification, the RFLP technology has played a pre-eminent role in providing data. The RFLP assay first revealed an abundance of genetic markers among elite inbred lines and varieties such that genetic distances reflective of pedigree could be obtained. PCR-based methods such as SSR and AFLP, however, offer significant practical improvements over RFLP. These methods are increasingly popular because of their overall information content, and may replace RFLPs for varietal identification.

Chapter 4 has focused on the use of AFLP markers for varietal identification. AFLP markers generated by CNG methylation sensitive (P/M) and CNG methylation insensitive (E/M) ECs and AFLP markers collected from hypomethylated (P/M) and hypermethylated (^{m}P/M) regions have been compared for their polymorphism information content (PIC), marker utility index (MI) and patterns of genetic diversity among a representative sample of 33 maize inbred lines. The PIC value provides an estimate of the discriminatory power of a marker by taking into account not only the number of alleles at a locus, but also the relative frequencies of those alleles in the population under study. An AFLP marker is assumed to correspond to a locus having only two alleles, defined by presence and absence of the band, respectively. Therefore, the maximum PIC value of an AFLP marker equals 0.5. Sets of P/M or ^{m}P/M PCs generate significantly higher PIC_{av} (an arithmetic mean of PIC values of *N* AFLP markers generated by a PC) (0.38 and 0.38, respectively) than sets of E/M PCs (0.32), demonstrating that the efficiency of polymorphism detection by the AFLP method can be adjusted by altering the ECs. In contrast, similar studies have shown that mean PIC values for multi-allelic SSR markers are higher, ranging from 0.59 to 0.62 (Senior et al. 1998; Smith et al. 1997), similar to a reported mean PIC value for RFLP markers (0.60) by Dubreuil et al. (1996).

The MI is defined as the product of the PIC value and the effective multiplex ratio (EM), the extent to which an assay can identify multiple polymorphisms

simultaneously (Powell et al. 1996). Sets of P/M PCs generate, for a pair of genotypes being compared, an average of 22.5 polymorphic fragments per assay. This is four to five polymorphic fragments more per assay than sets of E/M (18.7) or ᵐP/M (17.2) PCs. So, not only the efficiency of polymorphism detection but also the marker utility index of the AFLP method can be adjusted by altering the ECs. In contrast, the mean MI of the SSR marker equals 0.6. Technical developments to multiplex SSRs in maize (Smith et al. 1997; Senior et al. 1998) are underway. Multiplexing SSRs will provide a boost in the information yielded per assay. However, it will add to the initial development costs since the primers must often be redesigned and extensively tested to amplify all of the targets efficiently in the same reaction vessel (Bates et al. 1996).

So, although AFLP markers have a lower polymorphism rate than SSR, the numbers of AFLP markers generated per assay are sufficiently large to offset their lower polymorphism rates.

The numerical values of PIC, EM and, consequently, MI of the AFLP assay will vary across species and, in a species, with the degree of genetic relatedness among the individuals analyzed. Assuming an equal representation of the alleles in the individuals analyzed, a decreasing number of alleles per locus (with a minimum of two alleles per locus) will result in a mean AFLP PIC value asymptotically increasing to 0.5, while the mean SSR PIC value asymptotically decreases to 0.5. Simultaneously, a lower number of alleles per locus narrows down the heterogeneity among the individuals, resulting in a lower EM. So, the difference in efficiency of polymorphism detection between the AFLP and the SSR assay is positively correlated with the degree of allele diversity present in a species.

Several factors may affect the estimate of genetic relationships between individuals, like number of markers used and the distribution of markers over the genome. In order to estimate the sample size of AFLP markers required to provide a given level of precision in the estimation of genetic distances among a set of inbred lines, bootstrap sampling variances were determined. Furthermore, the overrepresentation of certain genomic regions and, subsequently, the sampling variance of the estimates of genetic relationships were minimized by using subsets

of AFLP markers uniformly distributed over the linkage groups. Over-sampling of certain regions of the genome introduces some redundancy of information carried by the markers, causing higher sampling variance of the estimates of genetic relationships. Reducing marker information redundancy was achieved by sampling or weighing markers conditionally on their map position in the B73 × Mo17 genetic map, whether or not in combination with further selection or weighing relative to their PIC value. The sample size of E/M markers had to be substantially larger (173) than those of either P/M (135) or "P/M (129) in order to reduce the mean standard deviation (*sd*) in the estimation of genetic distances among the 33 inbred lines to 5%. Accordingly, the average number of P/M, "P/M and E/M PCs required to obtain a 5% precision in the GD estimate are 2-3, 3-4 and 3-4, respectively. In contrast, investigating a set of 33 maize inbred lines, a similar precision (5%) was attained with 30-40 clone-enzyme RFLP combinations, 40-50 RAPD primers and 20-30 SSR primers, respectively (Pejic et al. 1998). So, it is clear from this comparison that the AFLP method will produce in a faster and more cost effective way than other methods a set of random markers providing a given level of precision in the estimation of genetic distances among a set of inbred lines. The speed with which the AFLP method generates marker data is due not only to its high multiplex ratio, but also to its inherent ease of genotyping. In the case of a inbred analysis, AFLP genotyping requires only a plus/minus assay rather than distinguishing several length-based alleles such as in the SSR genotyping process, permitting easier automation.

Reducing marker information redundancy by selecting markers evenly distributed over each chromosome is even more effective in minimizing the sampling variance: when evenly spread over the genome, only 106 AFLP markers are required for a mean *sd* of 5%.

This result shows the clear advantages of performing an analysis based on 'map-based' AFLP markers in order to minimize the number of markers required for a given level of precision in the estimation of genetic distances among individuals, or alternatively in order to estimate genetic relationships among individuals more precisely for a given number of markers. However, in order to generate this subset

of 106 AFLP markers, the number of assays required tremendously increases. Therefore, deriving from a larger data set the minimal set of PCs generating AFLP markers which meets best the criteria of being uniformly distributed over the linkage groups and of giving the least sampling variance in the estimation of genetic relationships, in practice will offer a more time- and cost-effective alternative.

Marker-assisted breeding

Although AFLP markers are extremely useful for varietal identification, with regard to agronomic performance, genetic distance (GD) data *per se* are not always practically useful predictors of heterosis. The term heterosis is usually described in terms of the increase in size or rate of growth of offspring over inbred parents. In maize, the main strategy that has been followed towards hybrid prediction during the last decade is based on the 'distance' model: heterosis in a hybrid is related to the genetic divergence between its parental lines (Lee et al. 1989). In Chapter 5, the 'distance' model-strategy has been applied to 78 hybrids, evaluated for grain yield (GY). Estimates of the GD between parents did not consistently identify the best crosses in terms of hybrid performance (HP), particularly not when the two parents are non-related lines. On the contrary, linear correlations found between the specific genetic distance (SGD) and the specific combining ability (SCA) were highly significant ($P<0.001$) and of a fairly high magnitude (r = 0.61-0.68) suggesting some practical utility in predicting SCA effects. The overall results, therefore, were partly in good accordance with the general tendency found, namely that the prediction efficiency of the 'distance' model is high when 1) hybrids between related lines (intraheterotic crosses) and 2) hybrids between both related and unrelated lines (intra- and interheterotic crosses) are considered. However, correlations between genetic distances of unrelated lines only and their respective interheterotic crosses, were of a low practical predictive value.

Drawbacks of the 'distance' model are the following: 1) although the major part of the molecular polymorphism detected between inbred lines has no direct effect on the trait of interest (spatial separation), it is converted into the metric GD; and

2) especially in the case of maize, linkage disequilibria between markers and QTL may differ randomly from one heterotic group to the other. Therefore, a novel method for predicting HP and SCA, circumventing the drawbacks of the 'distance' model has been developed and presented in Chapter 5.

The method is based on: 1) AFLP markers which are, across a set of F_1's, significantly associated with the trait of interest; and 2) the joint added effect of the several selected marker loci. Linear regression of the trait of interest on a hybrid value representing the total contribution of the selected markers (TCSM) in terms of their additive and dominance effects estimates, results in a model for the prediction of the trait of interest in the hybrid. Since only the interheterotic crosses are of commercial importance and therefore of interest to the breeder, the practical value of the TCSM approach has been evaluated by applying this procedure to the 53 interheterotic hybrids out of the 78 hybrids, evaluated for GY. The *a* en *d* values of the markers selected for HP for GY indicate that QTL with additive ($d = 0$) to partial dominance ($0 < |d| < |a|$) effects are prevalent. In contrast, the *a* and *d* values of the markers selected for SCA for GY fit single gene models with only overdominance effects ($|d| > |a|$). Results suggested practical utility of the TCSM method especially in predicting the HP for GY: the joint added contribution of single QTL involved in GY explained 59.3% (based on 7 markers selected at $P<0.0001$) to 62.4% (based on 20 markers selected at $P<0.001$) of the HP variance, indicating that the TCSM model has a higher HP prediction efficiency than the 'distance' model (explained HP variance ranges from 28.3% to 43.6%). In addition, cross-validation procedures, evaluating the HP prediction efficiency of the TCSM method, identified the best single crosses, suggesting that the TCSM approach is efficient as a preliminary screening of test hybrids before field evaluation.

The novel method presented in Chapter 5 is attractive not only for its predictive value, but also for its use as a 'low cost and effort' QTL screening method. The established method for identification of QTL is the selection of two parents, representing a fraction of the germplasm used by the breeder, that differ markedly in one or more (related) particular quantitative traits of interest, followed by the

determination of associations between markers and that character in a segregating population. However, the segregational resolution of conventional segregating populations (e.g. F_2, BC) is too low to distinguish between tightly linked and less tightly linked QTL-marker combinations. The approach to QTL identification followed in Chapter 5 has the following potential advantages over QTL detection in a segregating population. Firstly, marker-trait associations are only expected to be found in case a marker is highly linked to a QTL. This is because across a set of lines, associations between QTL and loosely linked markers will be non-existent due to accumulated recombination events during the establishment of the lines. Basically, the type of associations we have identified are due to identity by descent of QTL and marker alleles across lines. Secondly, only a limited number of lines representing the gene pool used by the breeder need to be genotyped. Thirdly, it may allow the detection of QTL that vary across a wide spectrum of the germplasm used.

Although direct comparisons of QTL are complicated by differences in parental lines, design of the cross, number of progeny, the environments in which the progeny was assessed, and by different marker loci and QTL detection methods, other reports have identified some of the same regions detected by the TCSM method to be associated with GY.

Despite the detection of many QTL such as for grain yield, the impact of QTL mapping on the establishment of improved inbred lines has so far been modest in maize MAB programs. This may be due to: 1) the observed inconsistency of trait means of marker genotypes across experiments, causing difficulty in improving traits by selecting for desirable marker alleles in any given breeding population; 2) the lack of congruent QTL, best explained by recognizing that there is little power to identify QTL unless the trait exhibits a fairly high heritability (van Ooijen 1992); 3) the use of insufficient numbers of mapped markers, not optimizing the probability to find tight marker-trait associations; and 4) focusing on QTL studies that have been uncoupled from breeding activities or else having been performed in a non-elite breeding germplasm.

In order to find consistent and more applicable results, large segregating population sizes (>1000) and many inbred lines, being used in actual breeding programs, are needed. Furthermore, tools are needed to distinguish tightly linked from less tightly linked QTL-marker combinations, either by analyzing a set of inbred lines (Chapter 5) or by relaxing linkage disequilibrium through random sibmating (Beavis et al. 1992), requiring both a high saturation with genetic markers. Regardless of the mapping technology, other serious and fundamental challenges to be met include: 1) obtaining reliable performance data for quantitative traits; 2) an understanding of genotype × environment interaction; and 3) improved detection of QTL with epistatic effects.

As already mentioned, most QTL studies have been uncoupled from standard breeding practice. A recently proposed strategy for molecular breeding, referred to as advanced backcross QTL (AB-QTL; Tanksley and Nelson 1996) attempts to integrate QTL analysis and variety improvement efforts. This method identifies beneficial alleles in unadapted germplasm in order to transfer these into elite cultivars, thus exploiting the hidden value of exotic germplasm. The general strategy of AB-QTL analysis is comprised of the following experimental phases: 1) generation of an elite × unadapted donor hybrid; 2) backcross to the elite parent to produce BC_1 and BC_2 populations which are subjected to marker and/or phenotypic selection against undesirable donor alleles; 3) molecular marker characterization of the BC_2 or BC_3 population; 4) generation of BC_3 or BC_4 families which are evaluated for agronomic performance and analyzed for QTL; 5) selection of target genomic regions containing useful donor alleles for the production of near-isogenic lines (NILs) in the elite background using marker-assisted selection; and 6) evaluation of the agronomic performance of the NILs and elite parent controls in multi-location trials (Tanksley and Nelson 1996). Recently, published reports (Tanksley et al. 1996; Bernacchi et al. 1998a, b) represent the first full cycle of AB-QTL analysis in tomato from the production of the initial interspecific hybrid (*L. esculentum* × *L. hirsutum* and *L. esculentum* × *L. pimpinellifolium*) to the development of breeding lines improved for several important agronomic traits, including total yield, soluble solids, fruit color and fruit

firmness. These results indicate that unadapted germplasm has genetic potential for the improvement of quantitative traits of agronomic importance and that molecular markers are very instrumental to establish NILs.

Stuber and Sisco (1991) identified QTL in the inbred lines B73 and Mo17 that could be improved after replacement with alleles from the lines TX303 and Oh43, respectively. Subsequent crosses among the top performing enhanced lines resulted in 5% of the hybrids outperforming the original B73 × Mo17 hybrid.

The superiority of exotic alleles over adapted alleles was also demonstrated clearly at a few QTL for several traits, including grain yield and number of ears per plant, in maize F_2 individuals and their selfed F_3 progeny (Ragot et al. 1995). However, it remains to be seen what response these favorable alleles produce once introgressed and evaluated in hybrid combinations and if these can surpass the leading processing hybrids.

A partisan view on the future of the AFLP assay

To date, more than any other molecular marker system, the AFLP assay fulfils best the requirements of a high-throughput low-cost genotyping technology in order to genetically map traits of high heritability and backcross monogenic traits. The AFLP technology has a high amenability to automation, and at Keygene the AFLP-related software used for the scoring of AFLP gels has greatly streamlined the data production. Data for a 1500 AFLP marker linkage map, based on 100 segregating individuals (15×10^4 data points), can be acquired in 6 weeks by one person. In addition, AFLP markers are mutationally stable (while SSR exhibit some instability (Weber and Wong 1993)) making them more suitable for association studies in which linkage disequilibrium between markers and an unknown variant is used to map trait associations.

However, the greatest gains in efficiency from MAB would be for those traits that are polygenic and exhibit low heritabilities (Lande and Thompson 1990). In order to realize these gains in MAB efficiency, it is important to identify and fine-map as many QTL as possible. This can be achieved by evaluating large numbers of progeny (>1000) with large marker data sets (>10000). Thus, at least 10×10^6

datapoints are required. Using the AFLP method, this would take more than 400 weeks, making clear that the AFLP method in its present use is not appropriate for this task, since it is inherently limited for a higher sample throughput efficiency by its gel based nature. A major jump in throughput, and therefore lowering of the cost of data, can be obtained by the introduction of multichannel capillary electrophoresis. Currently, two instrument builders have or will have 96-capillary instruments available (Molecular Dynamics and Perkin-Elmer) that perform a largely automated sample analysis. Using these machines, an average turn-around time for a group of 96 AFLP reactions will be in the order of 2,5 hours. When proven to be robust, these platforms will be able to process samples largely unattended. It is expected, that the implementation of such machines will increase the throughput by at least a factor of 5, when both the data production process and data extraction process are considered. The introduction of multiple dyes to analyze 4 different product sets simultaneously in a single capillary run will add a further factor of 3–4 to this increase. Where capillary analysis of AFLP reactions may thus allow a 15-20 fold increase in throughput, the next generation of molecular markers clearly will be analyzed on a non-gel based platform, e.g., solid-phase DNA-arrays. Such non-gel based assays are expected to assay at least a 100-fold more quickly than a gel-based method. They are expected to enter the field of molecular plant breeding soon after they are borne out in human diagnostic applications. DNA polymorphisms that lend themselves very well to automated analysis on a very large scale, e.g., with DNA array technology, are single nucleotide polymorphisms (SNPs). SNPs are the most important component of genetic variation, widely distributed across the genome, mostly bi-allelic in nature and mutationally stable. Nowadays, SNPs are acquiring a strong momentum in human genetics, especially by the new U.S. Human Genome Project plan for 1998-2003 (Collins et al. 1998). This ambitious program includes goals for developing technologies for rapid, large-scale identification and scoring of SNPs and for the generation of an SNP map of at least 100000 markers. So, SNPs are expected to take the place of SSRs, just as the SSR markers rapidly replaced the RFLP markers, in human genetics.

For small sample size, the genome density of SNPs (dSNP) as they are detected by the AFLP method can be estimated as follows:

$$dSNP = \beta / k$$

where β is the the fraction of polymorphic bands and k is the number of targeted nucleotides (restriction sites + selective bases) by an AFLP PC, e.g., *Pst*I+2/*Mse*I+3 targets 15 nucleotides, while *EcoR*I+3/*Mse*I+3 targets 16 nucleotides. Note that the occurrence of two point mutations in this short motif, is neglected, leading to a small underestimation of the genomic dSNP. Calculated on the basis of experimental data obtained from the two inbred lines B73 and Mo17 (Table 3.2), the dSNP in the maize genome is estimated to be on average 1 per 50 base pairs. In comparison, when two human haploid genomes are compared, the dSNP is estimated to be on average 1 per 1000 base pairs (Collins et al. 1998). Based on the experimental data obtained from the thirty three maize inbred lines investigated in Chapter 4 (Table 4.2), the dSNP in the maize genome is even estimated to be on average 1 per 20 base pairs. Note that on average 10% of the polymorphic AFLP bands are the result of insertion/deletion (Chapter 3), which may lead to slight overestimation of the dSNP in maize.

As increasing amounts of sequence data become available in public and private databases of plant species, and the pressing need for far greater throughput in sample and marker analyses does not cease, biochips containing plant SNPs are expected to enter the field of molecular plant breeding soon. However, it remains to be seen whether a certain marker density on the biochip as well as a certain level of automation in the analysis can be obtained in order to assay at least a 100-fold more quickly than the AFLP method.

References

1. Ajmone Marsan P, Livini C, Messmer MM, Melchinger AE, Motto M (1992) Cluster analysis of RFLP data from related maize inbred lines of the BSSS and LSC heterotic groups and comparison with pedigree data. Euphytica 60: 139-148
2. Ajmone Marsan P, Monfredini G, Ludwig RF, Melchinger AE, Franceschini P, Pagnotto G, Motto M (1995) In an elite cross of maize a major quantitative trait locus controls one-fourth of the genetic variation for grain yield. Theor Appl Genet 90: 415-424
3. Ajmone Marsan P, Castiglioni P, Fusari F, Kuiper M, Motto M (1998) Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor Appl Genet 96: 219-227
4. Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J, Kuiper MTR (1998) Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant J 14: 259-271
5. Austin DF, Lee M (1996) Comparitve 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 2: 817-826
6. Austin DF, Lee M (1998) Detection of quantitative trait loci for grain yield and yield components in maize across generations in stress and nonstress environments. Crop Sci 38: 1296-1308
7. Barrett BA, Kidwell KK (1998) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci 38: 1261-1271
8. Bates SRE, Knorr DA, Weller JW, Ziegle JS (1996) Instrumentation for automated molecular marker acquisition and data analysis. In: Sobral BWS (ed) The impact of plant molecular genetics. Bikhäuser, Boston, pp 239-255
9. Beavis WD, Grant D (1991) A linkage map based on information from four F_2 populations of maize (*Zea mays* L.). Theor Appl Genet 82: 636-644

10. Beavis WD, Lee M, Grant D, Hallauer AR, Owens T, Katt M, Blair D (1992) The influence of random mating on recombiniation among RFLP loci. Maize Genet Coop Newsletter 66: 52-53
11. Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F_4 progeny from maize. Crop Sci 34: 882-896
12. Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249: 65-73
13. Beckett JB (1987) B-A translocations in maize. J Hered 69: 27-36
14. Bender J (1998) Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. Trends Biochem Sci 23: 252-256
15. 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 *Lycopersicum hirsutum*. Theor Appl Genet 97: 381-397
16. Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig J, Zamir D, Tanksley S (1998b) Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicum hirsutum* and L. *pimpinellifolium*. Theor Appl Genet 97: 170-180
17. Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34: 20-25
18. Bird AP, Southern EM (1978) Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol 118: 27-47
19. Boppenmaier J (1993) Genetic Diversity for RFLPs in European Inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. Plant Breeding 111: 217-226
20. Botstein D, White RL, Skolnick MH, Davis RW (1980) Construction of a genetic map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314-331

21. Brigneti G, Garciamas J, Baulcombe DC (1997) Molecular mapping of the potato virus Y resistance gene Ry (Sto) in potato. Theor Appl Genet 94: 198-203

22. Burr BM, Burr FA, Thompson KH, Albertson M, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118: 519-526

23. Burstin J, Charcosset A, Barrière Y, Hébert Y, De Vienne D, Damerval C (1995) Molecular markers and protein quantities as genetic descriptors in maize. II. Prediction of performance of hybrids for forage traits. Plant Breeding 114: 427-433

24. Canada K, Meeley R (1996) PCR-based selection of gene-specific Mutator insertion alleles. In: Abtracts of the 38th annual maize genetics conferences

25. Carlson WR (1988) The cytogenetics of corn. In: Corn and Corn development, 3rd edition. Sprague GF, Dudley JW (eds). American Society of Agronomy, Madison, Wisconsin. pp. 259-343

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

27. Causse M, Santoni S, Damerval C, Maurice A, Charcosset A, Deatrick J, de Vienne D (1996) A composite map of expressed sequences in maize. Genome 39: 418-432

28. Cervera MT, Gusmão J, Steenackers M, Peleman J, Storme V, Vanden Broeck A, Van Montagu M, Boerjan W (1996) Identification of AFLP molecular markers for resistance agains *Melampsora larici-populina* in *Populus*. Theor Appl Genet 93: 733-737

29. Cervera MT, Cabezas JA, Sancha JC, Martínez de Toda F, Martínez-Zapatar JM (1998) Application of AFLPs t the charactarization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theor Appl Genet 97: 51-59

30. Charcosset A, Essioux L (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet 81: 571-575

31. Charcosset A, Bonniseau B, Touchebeuf O, Burstin J, Dubreuil P, Barrière Y, Gallais A, Denis J-B (1998) Prediction of maize hybrid silage performance using marker data: Comparison of several models for specific combining ability. Crop Sci 38: 38-44

32. Cockerham CC, Zeng Z-B (1996) Design III with marker loci. Genetics 143: 1437-1456

33. Coe E, Hancock D, Kowaleski S, Polacco M (1995) Gene list and working maps. Maize Genet Coop Newsl 69: 191-267

34. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L, the members of the DOE and NIH planning groups (1998) New goals for the US Human Genome Project 1998-2003. Science 282: 682-689

35. Davis G, McMullen M, Polacco M, Grant D, Musket T, Baysdorfer C, Staebell M, Xu G, Koster L, Houchins K, Melia-Hancock S, Coe EH (1998) UMC 1998 molecular marker map of maize. Maize Genet Coop Newsl 72: 118-128

36. Demeke T, Sasikumar B, Hucl P, Chibbar RN (1997) Random amplified polymorphic DNA (RAPD) in cereal improvement. Maydica 42: 133-142

37. Dillmann C, Bar-Hen A, Guérin D, Charcosset A, Murigneux A (1997) Comparison of RFLP and morphological distances between maize *Zea mays* L. inbred lines. Consequences for germplasm protection purposes. Theor Appl Genet 95: 92-102

38. Dubreuil P, Dufour Pm Krejci E, Causse M, de Vienne D, Gallais A, Charcosset A (1996) Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. Crop Sci 36:790-799

39. Dudley JW, Saghai Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31: 718-723

40. Eathington SR, Dudley JW, Rufener GK (1997) Marker effects estimated from testcrosses of early and late generations of inbreeding in maize. Crop Sci 37: 1679-1685

41. Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116: 113-125

42. Emerson RA, Beadle GW, Fraser AC (1935) A summary of linkage studies in maize. Cornell Univ Agric Exp Stn Memoir 180: 1-83

43. Everson EH, Schaller CW (1955) The genetics of yield differences associated with awn barbing in the barley hybrid (Lion × Atlas^{10}) × Atlas. Agron J 47: 276-280

44. Falconer DS (1989) Inbreeding and crossbreeding: II. Changes of variance. In: Introduction to quantitative genetics, 3rd edition. Longman Scientific Technical, Essex, England. pp. 264-281

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

46. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci 89: 1827-1831

47. Gardiner JM, Coe EH, Melia-Hancock S, Hoisington DA, Chao S (1993) Development of a core RFLP map in maize using an immortalized F_2 population. Genetics 134: 917-930

48. GENSTAT-5-COMMITTEE 1993. Genstat 5: reference manual. Oxford University Press, New York.

49. Godshalk EB, Lee M, Lamkey KR (1990) Relationship of restriction fragment length polymorphisms to single cross hybrid performance of maize. Theor Appl Genet 80: 273-280

50. Gonzalgo ML, Jones PA (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res 25: 2529-2531

51. Griffing B (1956) Concept of general and specific combining ability in relation to the diallel crossing system. Aust J Biol Sci 9: 463-493

52. Gruenbaum Y, Naveh-Maney T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292: 860-862

53. Hayes HK (1952) Development of the heterosis concept. In: Gowen JW (ed) Heterosis. Iowa State College Press. Ames, IA.

54. Hayman BI (1954) The theory and analysis of diallel crosses. Genetics 39: 789-809

55. Heiskanen M, Syvänen AC, Siitari H, Laine S, Palotie A (1994) A novel method to quantitate methylation of specific genomic regions. PCR Methods and Applic 4: 26-30

56. Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72: 761-769

57. Helentjaris T, Weber D, Wright S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphism. Genetics 118: 353-363

58. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci 93: 9821-9826

59. Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R (1996) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in Lactuca spp. Theor Appl Genet 93: 1202-1210

60. Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaudoise Sci Nat 44: 223-270

61. Jansen RC (1993a) Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. Biometrics 49: 227-231

62. Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding 3: 381-390

63. Karp A, Edwards K, Bruford M, Vosman B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D, Hewitt G (1997) Newer molecular technologies for biodiversity evaluation: opportunities and challenges. Nature Biotechnology 15: 625-628

64. Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997) A high-density soybean genetic map based on AFLP markers. Crop Science 37: 537-543

65. Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12: 172-175

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

67. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181

68. Lanza LLB (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor Appl Genet 94: 1023-1030

69. Lee M, Godshalk EB, Lamkey KR, Woodman WL (1989) Association of restriction length polymorphism among maize inbreds with agronomic performance of their crosses. Crop Sci 29: 1067-1071

70. Lincoln SE, Lander ES (1992) Systematic detection of errors in genetic linkage data. Genomics 14: 604-610

71. Livini C, Ajmone Marsan P, Melchinger AE, Messmer MM, Motto M (1992) Genetic diversity of maize inbred lines within and among heterotic groups revealed by RFLP. Theor Appl Genet 84: 17-25

72. Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N (1997) Polymorphism, distribution and segregation of AFLP markers in a doubled haploid rice population. Theor Appl Genet 94: 39-45

73. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27: 209-220

74. Marques CM, Araújo JA, Ferreira JG, Whetten R, O'Malley DM, Liu B-H, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. Tereticornis*. Theor Appl Genet 96: 727-737

75. Martienssen RA, Richards EJ (1995) DNA methylation in eukaryotes. Curr. Opin. Genet. and Dev. 5: 234-242

76. Martienssen R (1998) Transposons, DNA methylation and gene control. Trends in Genetics 14: 263-264

77. Matz EC, Burr FA, Burr B (1994) Molecular map based on TXCM and COXTX recombinant inbred families. Maize Genet Coop Newsl 68: 198-208

78. Maughan PJ, Saghai Maroof MA, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theor Appl Genet 93: 392-401

79. Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499-560

80. McClelland M, Nelson M, Raschke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res. 22: 3640-3659

81. Meksem K, Leister D, Peleman J, Zabeau M, Salamini F, Gebhardt C (1995) A high resolution map of the vicinity of the R1 locus on Chromosome V of potato based on RFLP and AFLP markers. Mol Gen Genet 249: 74-81

82. Melchinger AE, Lee M, Lamkey KR, Hallauer AR, Woodman WL (1990a) Genetic diversity for restriction length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor Appl Genet 80: 488-496

83. Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990b) Genetic diversity for restriction length polymorphisms: relation to estimate genetic effects in maize inbreds. Crop Sci 31: 1033-1040

84. Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among U.S. Maize inbreds revealed by Restriction Fragment Length Polymorphisms. Crop Sci 31: 669-678

85. Melchinger AE, Boppenmaier J, Dhillon BS, Pollmer WG, Herrmann RG (1992) Genetic diversity for RFLPs in European maize inbreds: II. Relation to performance of hybrids within versus between heterotic groups for forage traits. Theor Appl Genet 84: 672-681

86. Melchinger AE, Gumber RK, Leipert RB, Vuylsteke M, Kuiper M (1998a) Prediction of testcross means and variances among F_3 progenies of F_1 crosses from testcross means and genetic distances of their parents in maize. Theor Appl Genet 96: 503-512

87. Melchinger AE, Kuntze L, Gumber RK, Lübberstedt T, Fuchs E (1998b) Genetic basis of resistance to sugarcane mosaic virus in European maize germplasm. Theor Appl Genet 96: 1151-1161

88. Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Charactarization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. Plant Mol.Biol 16: 753-770

89. Messmer MM, Melchinger AE, Lee M, Woodman WL, Lee EA, Lamkey KR (1991) Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor Appl Genet 83: 97-107

90. Messmer MM, Melchinger AE, Hermann RG, Boppenmaier J (1993) Relationship among early European maize inbreds: II. Comparison of pedigree and RFLP data. Crop Sci 33: 944-950

91. Meyer P, Niedenhof I, ten Lohuis M (1994) Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*. The EMBO Journal 13: 2084-2088

92. Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Mol Breed 3: 127-136

93. Montero LM, Filipski J, Gil P, Capel J, Martínez-Zapater JM, Salinas J (1992) The distribution of 5-methylcytosine in the nuclear genome of plants. Nucleic Acids Res. 20: 3207-3210

94. Moore G (1995) Cereal genome evolution: Pastoral pursuits with 'Lego' genomes. Curr Opin Gen and Dev 5: 717-724

95. Mumm RH, Dudley JW (1994) A classification of 148 U.S. maize inbreds: I. Cluster analysis based on RFLPs. Crop Sci 34: 842-851

96. Paul S, Wachira FN, Powell W, Waugh R (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea (Camellia-Sinensis (L) Kuntze, O) revealed by AFLP markers. Theor Appl Genet 94: 255-263

97. Pejic I, Ajmone Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor Appl Genet 97: 1248-1255

98. Pfeiffer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. Science 246: 810-813

99. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2: 225-238

100. Qi X, Stam P, Lindhout P (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. Theor Appl Genet 96: 376-384

101. Ragot M, Sisco PH, Hoisington DA, Stuber CW (1995) Molecular-marker-mediated characterization of favorable exotic alleles at quantitative trait loci in maize. Crop Sci 35: 1306-1315

102. Rasmusson JM (1933) A contribution to the theory of quantitative character inheritance. Hereditas 18: 245-261

103. Rohlf FJ (1992) NTSYS-PC Numerical Taxonomy and Multivariate Analysis System version 1.7. Owner manual

104. Ross K, Fransz PF, Jones GH (1996) A light microscopic atlas of meiosis in *Arabidopsis thaliana*. Chromosome Res 4: 507-516

105. Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95: 714-722

106. Sadri R, Hornsby PJ (1996) Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. Nucleic Acids Res 24: 5058-5059

107. Saghai Maroof MA, Soliman KM, Jorgenson R, Allard RW (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal locations and population dynamics. Proc Natl Acad Sci USA 81: 8014-8018

108. Saluz HP, Jost JP (1993) Major techniques to study DNA methylation. In: Saluz HP, Jost HP (eds) DNA Methylation: Molecular Biology and Biological Significance. Birkhauser Verlag, Basel, Switzerland, pp 11-26

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

110. Schondelmaier J, Steinrucken G, Jung C (1996) Integration of AFLP markers into a linkage map of sugar beet (*Beta vulgaris* L.) J Plant Breed 115: 231-237

111. Schut JW, Qi X, Stam P (1997) Association between relationship measures based on AFLP-markers, pedigree data and morphological traits in barley. Theor Appl Genet 95: 1161-1168

112. Selker EU, Fritz DY, Singer MJ (1993) Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. Science 262: 1724-1728

113. Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1997) Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction. Crop Sci 36: 1676-1683

114. Senior ML, Murphy JP, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci 38: 1088-1098

115. Sharma SK, Knox MR, Ellis THN (1996) AFLP analysis of the diversity and phylogeny of Lens and its comparison with RAPD analysis. Theor Appl Genet 93: 751-758

116. Shoemaker J, Zaitlin D, Horn J, De Mars S, Kirshman J, Pitas J (1992) A comparison of three agrigenetics maize RFLP linkage maps. Maize Genet Coop Newsl 66: 65-69

117. Simons G, van der Lee T, Diergaarde P, van Daelen R, Groenendijk J, Frijters A, Büschges R, Hollricher K, Töpsch S, Schulze-Lefert P, Salamini F, Zabeau M, Vos P (1997) AFLP-based fine mapping of the *Mlo* gene to a 30-kb DNA segment of the barley genome. Genomics 44: 61-70

118. Singer-Sam J, LeBon JM, Tanguay RL, Riggs AD (1990) A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res 18: 687

119. Smith JSC, Smith OS (1989) The description and assessment of distance between inbred lines of maize. I. The use of morphological traits as descriptors. Maydica 34: 141-150

120. Smith JSC, Smith OS (1991) Restriction fragment length polymorphisms can differentiate among U.S. maize hybrids. Crop Sci 31: 893-899

121. Smith JSC, Beavis WD (1996) Molecular marker assisted breeding in a company environment. In: Sobral BWS (ed) The impact of plant molecular genetics. Bikhäuser, Boston, pp 259-272

122. Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (Zea mays L.): comparisons with data from RFLPs and pedigree. Theor Appl Genet 95: 163-173

123. Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, grain yield, heterosis and RFLPs. Theor Appl Genet 80: 833-840

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

125. Stam P, van Ooijen JW (1995) JoinMap™ version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen

126. Stewart CN, Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14: 748-750

127. Stuber CW, Sisco P (1991) Marker-facilitated transfer of QTL alleles between elite inbred lines and responses in hybrids. In: Proceedings of the 46th annual corn and sorghum industry research conference; Dec 11-12, 1991, Chicago, IL. Washington, DC: American Seed Trade Assocation

128. Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132: 823-839

129. Stuber CW (1995) Mapping and manipulating quantitative traits in maize. Trends in Genetics 11: 477 - 481

130. Sturtevant AH (1913) The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. J of Exp Zool 14: 43-59

131. Sumner AT (1990) Chromosome banding. Unwin Hyman Ltd pp.157-169

132. Tanksley SD, Young ND, Paterson AH, Bonier Bale MW (1989) RFLP mapping in plant breeding: new tools for an old science. BioTechnology 7: 257-264

133. 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

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

135. Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17: 6463-6471

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

137. Thomas CM, Vos P, Zabeau M, Jones DA, Norcott KA, Chadwick BP, Jones JDG (1995) Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato *Cf-9* gene for resistance to *Cladosporium-Fulvum*. Plant J 8: 785-794

138. Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor Appl Genet 89: 259-264

139. van Eck HJ, van der Voort JR, Draaistra JR, van Zandvoort P, van Enckevort E, Segers B, Peleman J, Jacobsen E, Helder J, Bakker J (1995) The inheritance of chromosomal localization of AFLP markers in a non-inbred potato offspring. Mol Breed 1: 397-410

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

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

142. Veldboom LR, Lee M, Woodman WL (1994) Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. Theor Appl Genet 88: 7-16

143. 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 Res 23: 4407-4414

144. Vos P, Simons G, Jesse T, Wijbrandi J, Heinen L, Hogers R, Frijters A, Groenendijk J, Diergaarde P, Reijans M, Fierens-Onstenk J, de Both M, Peleman J, Liharska T, Hontelez J, Zabeau M (1998) The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. Nature Biotechnology 16: 1365-1369

145. Waalwijk C, Flavell RA (1978) *Msp*I, an isoschizomer of *Hpa*II, which cleaves both unmethylated and methylated *Hpa*II sites. Nucleic Acids Res 5: 3231-3236

146. Waugh R, Powell W (1992) Using RAPD markers for crop improvement. Trends Bio/technol 10: 186-191

147. Weber JL, Wong C (1993) Mutation of short tandem repeats. Hum Mol Genet 2: 1123-1128

148. Weir BS, Allard RW, Kahler AL (1972) Analysis of complex allozyme polymorphisms in a barley population. Genetics 72: 505-523

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

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

151. Xia XC, Melchinger AE, Lübberstedt T (1998) Genetische Kartierung von SCMV-Resistenz in Europäischem Mais mittels RFLP- und SSR-Markern. Vortraege fuer Pflanzenzuechtung 42: 13-15

152. Xiong Z, Laird PW (1997) COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res 25: 2532-2534

153. Yang AS, Gonzalgo ML, Zingg J-M, Millar RP, Buckley JD, Jones PA (1996) The rate of CpG mutation in Alu repetitive elements within the p53 tumor-suppressor gene in the primate germline. J Mol Biol 258: 240-250

154. Zabeau M, Vos P (1993) Selective retriction fragment amplification: a general method for DNA fingerprinting. European Patent Application 92402629.7; Publ no. EP 0534858 A1

155. Zehr BE, Dudley JW, Chojecki J, Saghai-Maroof MA, Mowers RP (1992) Use of RFLP markers to search for alleles in a maize population for improvement of an elite hybrid. Theor Appl Genet 83: 903-911

156. Zhu J, Gale MD, Quarrie S, Jackson MT, Bryan GJ (1998) AFLP markers for the study of rice diversity. Theor Appl Genet 96: 602-611

Samenvatting

Deze samenvatting is geschreven voor diegenen die niet vertrouwd zijn met het onderwerp van dit proefschrift of met biologie in het algemeen.

Dit proefschrift beschrijft de toepassing van de AFLP® merkertechnologie in de genetische analyse van mais. De AFLP technologie is een DNA merkertechnologie die begin jaren negentig zijn opmars maakte. De AFLP techniek gaat als volgt te werk: het DNA, de streng erfelijk materiaal die elke levende cel bevat, wordt met behulp van restrictie-enzymen in kleine fragmenten geknipt. Restrictie-enzymen zijn eiwitten die het DNA op heel specifieke plaatsen, de restrictieplaatsen genoemd, knippen. Zo heeft nagenoeg ieder restrictie-enzym een eigen specifieke restrictieplaats. Hoe frequenter een bepaalde restrictieplaats aanwezig is in het DNA, hoe meer DNA fragmenten er worden gegenereerd door een bepaald restrictie-enzym. De AFLP techniek maakt gebruik van een paar restrictie-enzymen (een enzyme combinatie), die tegelijkertijd het DNA in miljoenen stukjes knippen.

Dankzij een elegante techniek, kan de onderzoeker een fractie van alle DNA stukjes ontelbare keren kopiëren. Vervolgens worden de gekopieerde DNA fragmenten op een gelkolom gezet. Hier geldt: hoe korter het stukje DNA, hoe sneller het over de kolom beweegt en hoe lager het in de kolom terecht komt. Met behulp van een radioactieve detectiemethode worden de verschillende DNA fragmenten geïdentificeerd als streepjes. Zo krijgt iedere plant zijn eigen 'vingerafdruk', zijn eigen verzameling streepjes. Wanneer streepjes in bepaalde individuen wel en in andere niet aanwezig zijn, dan spreekt men van een DNA merker. In het geval deze DNA merker wordt opgepikt met de AFLP techniek spreekt men van een AFLP merker.

Inherent aan de AFLP methode is het grote aantal DNA fragmenten dat tegelijkertijd op een gelkolom kan worden geplaatst, en dus het grote aantal DNA merkers dat tegelijkertijd kan worden opgevist en geanalyseerd. Deze hoge mate van efficiëntie heeft er toe geleid dat het gebruik van de DNA merkertechnologie,

meer bepaald de AFLP methode, voor toepassingen in bv. de plantenveredeling, commercieel aantrekkelijk is geworden.

Restrictie-enzymen kunnen worden belemmerd in het knippen van het DNA wanneer hun restrictieplaats 'geblokkeerd' is. Een kleine chemische verbinding, namelijk een methylgroep, bindt zich op de restrictieplaats en zorgt er voor dat het DNA op deze plaats niet kan worden geknipt. Het blokkeringsproces wordt 'DNA-methylatie' genoemd; geblokkeerde restrictieplaatsen zijn dus 'gemethyleerd'.

In hoofdstuk 2 wordt een variant van de AFLP methode, namelijk de 'methylatie AFLP®' gepresenteerd. De methylatie AFLP techniek demethyleert restrictieplaatsen vooraleer AFLP fragmenten te genereren. Deze AFLP fragmenten worden 'gemethyleerde' AFLP fragmenten genoemd, verwijzend naar het oorspronkelijk gemethyleerd zijn van de plaats op het DNA waar het fragment vandaan komt.

De methylatie AFLP methode kan dus worden ingezet in de studie van DNA methylatie. DNA methylatie geniet al geruime tijd veel aandacht in de biologische wetenschap. Er wordt vermoed dat het de moleculaire basis vormt van chromosoominprenting. Chromosoominprenting lijkt een rol te spelen in het ontstaan van kanker en in een aantal genetische ziekten bij de mens. In dit proefschrift is de studie van DNA methylatie beperkt gebleven tot het meten van de graad van DNA methylatie in mais, tomaat en koolzaad, en het lokaliseren van gemethyleerd DNA op de chromosomen van mais.

Het totale DNA, ook wel het genoom genoemd, is opgedeeld in een aantal pakketjes, chromosomen genoemd. Mais telt 10 chromosoomparen. Alle eigenschappen van de plant, bv. virusresistentie, bloemkleur, bloeitijd, korrelgrootte, hebben een welbepaalde plaats op de chromosomen. Niet alleen eigenschappen, maar vanzelfsprekend ook de DNA fragmenten vinden hun oorsprong op de chromosomen. DNA fragmenten en eigenschappen kunnen genetisch worden gekarteerd, d.w.z. dat de relatieve positie van meerdere DNA fragmenten en eigenschappen, liggend op één chromosoom of verspreid over

meerdere chromosomen, kan worden bepaald. Het resultaat noemt men een genetische kaart.

Tot voor kort was het toewijzen van een chromosoompositie aan DNA merkers een kostbare en tijdsintensieve operatie. Met behulp van de AFLP techniek kan het genetisch karteren van DNA fragmenten in een hogere versnelling plaatsvinden. In hoofdstuk 3 worden twee hoge dichtheidskaarten van mais gepresenteerd. De twee kaarten bevatten respectievelijk 1539 and 1355 AFLP merkers. Een fractie van deze gekarteerde AFLP merkers zijn gemethyleerde AFLP merkers. Ze geven weer waar het chromosomaal DNA precies gemethyleerd is.

Genetische kaarten zijn onmisbaar wil men DNA merkertechnologie implementeren in plantenveredeling. Wanneer een merker Y en een eigenschap X op eenzelfde plaats liggen op de genetische kaart, dan kan het voorkomen van merker Y in een bepaald individu worden geassocieerd met het voorkomen van de eigenschap X. Zo kunnen veredelaars met behulp van DNA merkertechnologie reeds in een vroeg groeistadium van de plant inzicht krijgen in de aanwezigheid van eigenschappen die zich pas maanden of jaren later manifesteren.

De meest simpele toepassing van de DNA merkertechnologie is het bepalen van genetische afstanden tussen individuen, m.a.w. verwantschapsanalyse. De genetische afstand tussen 2 individuen is gelijk aan de fractie DNA fragmenten die de 2 individuen niet gemeenschappelijk hebben. In hoofstuk 4 werden genetische afstanden tussen 33 mais inteeltlijnen bepaald aan de hand van AFLP merkers. Deze verwantschapsanalyse maakt duidelijk in welke mate bepaalde maislijnen genetisch verwant zijn met elkaar. Verwantschapsanalyses zijn voor de maisveredelaar van praktisch nut bij het selecteren van ouderlijnen voor het opzetten van hybriden: twee ouderlijnen met een quasi identieke fingerprint (sterk genetische verwant) zijn niet interessant om met elkaar te kruisen. Verder hebben de resultaten uit hoofdstuk 4 aangetoond dat de efficientie van het uitvoeren van een verwantschapsanalyse kan worden opgeschroefd (m.a.w. er zijn minder merkers vereist) door 1) de keuze van de enzym combinatie waarmee het

DNA wordt geknipt, en 2) een meer uniforme spreiding van de merkers over de genetische kaart.

Hoofdstuk 5 is gewijd aan het onderwerp 'heterosis'. Heterosis wordt gedefinieerd als de betere prestatie die de eerste generatie nakomelingen leveren in vergelijking met hun ouders. Heterosis komt zowel bij planten als bij dieren voor. Op zich is het geen eigenschap, maar een fenomeen dat eigenschappen positief beinvloedt. Begin deze eeuw werd voor het eerst gebruik gemaakt van heterosis in de maisveredeling, wat leidde tot de eerste hybride rassen. Tegenwoordig domineert het hybride-concept niet alleen de maisveredeling, maar ook de veredeling van heel wat andere land- en tuinbouwgewassen, zoals rijst, zonnebloem, en tomaat.
In hoofdstuk 5 werd de manifestatie van heterosis in mais gemeten aan de eigenschap korrelopbrengst. AFLP merkers geassocieerd met de eigenschap korrelopbrengst, alsook met heterosis in korrelopbrengst werden geïdentificeerd. Deze AFLP merkers werden op de genetische kaart geplaatst, wat leidde tot de identificatie van een aantal chromosomale gebieden die hoogstwaarschijnlijk de genen herbergen die een belangrijke rol spelen in korrelopbrengst.
Deze set AFLP merkers geassocieerd met korrelopbrengst kunnen heel waardevol zijn voor de maisveredelaar in zijn speurtocht naar goede oudercombinaties die superieure hybriden moeten opleveren. De veredelaar kan zijn collectie ouderlijnen screenen of de aanwezigheid van deze AFLP merkers: ouderlijnen die veel van deze AFLP merkers in hun fingerprints vertonen worden geselekteerd om mee verder te werken. Meer zelfs, twee ouderlijnen die samen veel van deze AFLP merkers in hun fingerprints vertonen (ouderlijnen zijn complementair), zijn geschikt om met elkaar te worden gekruist. Zo kan de veredelaar efficiënter te werk gaan in het selekteren van geschikte oudercombinaties.

Curriculum vitae

De auteur is geboren op 6 januari 1965 in Roeselare, België. In 1983 beëindigde hij het secundair onderwijs aan het St. Aloysiuscollege te Diksmuide en begon de opleiding voor landbouwkundig ingenieur aan de Rijksuniversiteit te Gent. In 1988 sloot hij deze opleiding, richting 'Voeding- en dieetleer', af met onderscheiding. Na een jaar militaire dienst kwam hij terecht op het Centrum voor Voeding- en Drankonderzoek in Ieper. In 1991 startte hij met de opleiding 'Bijzonder licentiaat in de Moleculaire Biologie en Biotechnologie' aan de Vrije Universiteit te Brussel. In 1993 wist hij deze opleiding af te ronden met onderscheiding. In hetzelfde jaar werd hem een baan aangeboden op Keygene N.V. te Wageningen, wat tegelijkertijd zijn promotieplaats werd en waar hij tot op heden nog steeds werkzaam is als junior projectleider Genomics.