Die vorliegende Arbeit wurde am 21.01.2003 von der Fakultät III – Agrarwissenschaften der Universität Hohenheim als Dissertation zur Erlangung des Grades eines Doktors der Agrarwissenschaften (Dr. Sc. Agr.) angenommen.

Tag der mündlichen Prüfung: 31.01.2003

Prodekan:	Prof. Dr. K. Stahr
Berichterstatter, 1. Prüfer:	Prof. Dr. A.E. Melchinger
Mitberichterstatter, 2. Prüfer:	Prof. Dr. A. Pfitzner
Mitberichterstatter, 3. Prüfer:	Prof. Dr. R. Blaich

Aus dem Institut für
Pflanzenzüchtung, Saatgutforschung und Populationsgenetik
der Universität Hohenheim
Fachgebiet Angewandte Genetik und Pflanzenzüchtung
Prof. Dr. A.E. Melchinger

Resistance gene analogues as a tool for basic and applied resistance genetics exemplified by sugarcane mosaic virus resistance in maize (*Zea mays* L.)

Dissertation
zur Erlangung des Grades eines
Doktors der Agrarwissenschaften
der Fakultät III – Agrarwissenschaften
der Universität Hohenheim

von Dipl.-Ing. sc. agr.
Marcel Quint
aus
Celle

Stuttgart-Hohenheim

2002

Contents

[1]Quint M, Melchinger AE, Dußle CM, Lübberstedt T (2000) Breeding for virus resistance in maize. Genetika 32:529-545

[2]Quint M, Mihaljevic R, Dußle CM, Xu ML, Melchinger AE, Lübberstedt T (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. Theor. Appl. Genet. 105: 355-363

[3]Dußle CM, Quint M, Xu ML, Melchinger AE, Lübberstedt T (2002) Conversion of fragments tightly linked to SCMV resistance genes *Scmv1* and *Scmv2* into simple PCR-based markers. Theor. Appl. Genet. 105: 1190-1195

[4]Quint M, Dußle CM, Melchinger AE, Lübberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping, and MAS. Theor. Appl. Genet. 106: 1171-1177

[5]Frisch M, Quint M, Lübberstedt T, Melchinger AE (2004) Duplicate marker loci can result in incorrect locus orders on linkage maps. Theor. Appl. Genet. (in press)

Abbreviations

AFLP	amplified fragment length polymorphism
Avr gene	avirulence gene
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BSA	bulked segregant analysis
CAPS	cleaved amplified polymorphic sequence
DNA	deoxyribonucleic acid
GTP	guanosine triphosphate
EST	expressed sequence tag
indel	insertion/deletion
JGMV	Johnson grass mosaic virus
LRR	leucine-rich repeat
MAS	marker-assisted selection
MDMV	maize dwarf mosaic virus
MHC	major histocompatibility complex
NBS	nucleotide bindinge site
ORF	open reading frame
PCR	polymerase chain reaction
QTL	quantitative trait locus/loci
RGA	resistance gene analogue
R gene	resistance gene
SCMV	sugarcane mosaic virus
SSR	simple sequence repeat
SNP	single nucleotide polymorphism
SrMV	Sorghum mosaic virus
STS	sequence tagged site
TIR	*Toll* and interleukin receptor-like
WSMV	wheat streak mosaic virus

1 General introduction

Plant disease resistance

Plants are continually exposed to pathogen attack, but diseases are rare. There are basically three reasons for the missing success of pathogen infection and reproduction. (1) The plant does not supply the essential living requirements for a potential pathogen and is therefore a non-host. (2) Preformed plant defense compounds like structural barriers or pathotoxins restrict successful pathogen infection. (3) Plants are capable of defending themselves by means of a combination of constitutive and induced defenses. The latter resistance mechanism depends on recognition of the attacking pathogen. Knowledge about the genetic and biochemical basis of plant disease resistance has accumulated since the turn of the previous century, when plant breeders first recognised that disease was often controlled by Mendelian genes (Biffen 1905). The plant kingdom contains thousands of resistance genes (R genes) with specificities for particular viral, bacterial, fungal, or nematode pathogens. Despite the differences in defense responses induced by different plant-microbe interactions, some common characteristics are apparent during R gene mediated defenses. Therefore, it is becoming evident that plant genomes contain a large number of genes that are involved in the detection and discrimination of potential pathogens. Usually, the function of a certain R gene is limited to one or few genotypes of the respective pathogen (Keen 1992, de Wit 1992).

Generally, plant disease resistances can be inherited in a monogenic, oligogenic, or polygenic manner. Therefore, qualitative and quantitative resistances have to be distinguished concerning their resistance mechanism.

Qualitative resistance

The resistance mechanism of plants conferring qualitative or monogenic inherited resistance is comparable to the mammalian immune system with production of antigens by mammalian pathogens. Plant pathogens also produce a variety of potential signals. Some of these signals are detectable by plants. Genes expressing these signals in the pathogen are

designated avirulence (Avr) genes. Equivalent matching R and Avr gene pairs enable recognition of the pathogen and induce defense responses. Therefore, R gene products can be described as receptors for Avr-coded ligands in a gene-for-gene relationship (Flor 1956, 1971). R-Avr gene pairs resulting in resistance are epistatic over gene pairs that would otherwise result in susceptibility (Crute and Pink 1996). Gene pairs conferring higher degrees of resistance are generally epistatic over gene pairs associated with lower degrees of resistance, although phenotypic variation indicative of genetic additivity has also been reported, when more than one gene pair conferring resistance is effective. Following pathogen recognition, the resistance protein is presumed to activate signalling cascades that coordinate the initial plant defense response to impair pathogen ingress. Early signalling events following recognition are for example activation of protein kinases, induction of ion fluxes across the cellular membrane, and the release of reactive oxygen species probably triggering the transcriptional activation of defense responses. This signalling cascade results in the production of salicylic acid, cell wall fortification, and the expression of pathogenesis related proteins (reviewed in Hammond-Kosack and Jones 1996). Therefore, disease resistance triggered by genotype-specific pathogen recognition also became a model for signal transduction in plants.

Quantitative resistance

Polygenic or oligogenic inherited resistances are generally thought to be more durable and stable than monogenic resistances due to race or isolate unspecificity. The resulting resistance is caused by mechanisms different from the classical gene-for-gene concept and the subsequent hypersensitive response. The diversity of traits by which quantitative resistance can be expressed can be attributed to the variability of resistance mechanisms targeted at the mode of infection and the postinfectious effects of the virus. The defense response may be expressed in reduced rates of infection. Postinfectious mechanisms may cause extended incubation times, reduced virus concentration, incomplete virus spread, or reduction of the virus caused growth and yield losses.

The peculiarity of the resistance symptoms is also subject to several factors. While on the part of the virus its virulence and the dose of infection contribute to symptom expression,

on the part of the host these are mainly age, genetic background, and environmental effects like temperature and light intensity.

Structure of R genes

In absence of a known biochemical function of R gene products, R gene cloning strategies predominantly relied on defining the gene's chromosomal location using segregating populations, or transposon insertion to destroy biological activity, both followed by complementation to restore the resistance phenotype (Hammond-Kosack and Jones 1997). Since the first plant disease R gene had been cloned (Johal and Briggs 1992), R genes from several species have been isolated. Sequence comparisons among these genes have revealed remarkable similarities in general structure and conservation of specific domains that participate in protein-protein interactions and signal transduction. R genes contain similar sequence motifs, although they encode resistance to very different pathogens, including viruses, bacteria, fungi, and nematodes.

The majority of these genes encode a putative nucleotide binding site (NBS) and a leucine-rich repeat (LRR) region. The NBS domain is a common protein element that is required for ATP- and GTP-binding (Saraste et al. 1990). The presence of the highly conserved NBS domain suggests that nucleotide triphosphate binding is essential for the function of these proteins. The mechanistic role of NBS domains in the activation of plant defense remains unknown. Nucleotide triphosphate binding may alter the interactions between R gene products and other members of the defense signal transduction cascade (Bent 1996). However, only a few R genes have been functionally analysed and the origin and evolution of plant disease R genes remain obscure. The N-terminal region of some R genes contains a short sequence called TIR with homology to the animal innate immunity factors and cytoplasmic signalling domains, *Toll* and interleukin receptor-like genes. Pan et al. (2000) demonstrated that NBS domains faithfully predict whole gene structure and can be divided into two major groups. Group I NBS domains contain group specific motifs that are always linked with the TIR sequence at the N terminus. Group I NBS domains and their associated TIR domains are widely distributed in dicot species but were not detected in cereal databases nor could they be amplified by PCR. Group II domains are always associated with putative

coiled-coil domains in their N terminus and appear to be present throughout the angiosperms suggesting divergent evolution of the two main groups of R genes and possibly also diverged downstream signalling pathways.

LRRs have been defined as multiple, serial repeats of a motif of ~24 amino acids in length (Kobe and Deisenhofer 1994). They contain leucine or other hydrophobic amino acids at regular intervals. In animals, LRR domains mediate protein-protein interactions. Thus, the LRR domain of plant R gene products is considered as the site of pathogen recognition, providing specifities to interact with pathogen-derived elicitors, Avr gene products (Bent 1996; Baker et al. 1997).

Other motifs identified in R genes are the transcription factor Leucine zipper with a proposed role in homodimerisation of R gene products, and serine-threonine kinases modulating the phosphorylation state of proteins (reviewed in Bent 1996).

Resistance gene analogues

The majority of plant disease R genes cloned to date encode a predicted NBS domain attached to a C-terminal LRR of variable length (reviewed in Baker et al. 1997). In the case of these R genes, nucleotide binding has been predicted on the basis of sequence similarity only. Its biochemical function has yet to be demonstrated. Depending on the presence or absence of a N-terminal TIR, R genes can be categorised into two major phylogenetical types. NBS sequences of R genes have been recognised by the presence of at least five conserved domains including a P-loop, indicating that they are related to the ATP-/GTP-binding superfamily of proteins (Meyers et al. 1999). The overall sequence homology among members of the NBS-LRR genes is low – too low to be detectable by cross-hybridisation. However, the existence of conserved motifs provides opportunities for the design of degenerate primers and the isolation of RGAs by PCR from eukaryotic genomes. This approach has facilitated the amplification of RGAs from diverse plant genomes (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996).

In maize, Collins et al. (1998) amplified eleven classes of non-cross-hybridising sequences with amino acid identity to known NBS-LRR resistance proteins. Most of these RGAs were mapped to genomic regions known to contain R genes. Four of them mapped to

the potential R gene clusters on chromosomes 6S and 3L also known to contain sugarcane mosaic virus R genes *Scmv1* and *Scmv2*.

Sugarcane mosaic virus resistance

Sugarcane mosaic virus (SCMV) is one of the most important virus diseases of maize and causes serious yield losses in susceptible cultivars (Fuchs and Grüntzig 1995) (Fig. 1). Since the 1980s, SCMV and the closely related maize dwarf mosaic virus (MDMV) have been found in Germany (Fuchs and Kozelska 1984). In Germany, SCMV is more prevalent than MDMV and causes increasing damage to maize (Fuchs et al. 1996), while MDMV is a widespread viral disease in the southern US Corn Belt (Louie et al. 1991).

Figure 1: SCMV infection in European maize. The susceptible plants in the foreground show reduction of plant height, compared to the resistant plants in the background.

SCMV (Fig. 2) was formerly denoted as a MDMV isolate, MDMV-B (Shukla et al. 1989). Together with wheat streak mosaic virus (WSMV), Johnson grass mosaic virus (JGMV), Sorghum mosaic virus (SrMV), and MDMV it belongs to the same taxonomic group of related pathogenic potyviruses in maize. Owing to the non-persistent transmission, control of aphid vectors by chemical means is not effective and therefore, cultivation of resistant maize varieties is the most efficient method of virus control.

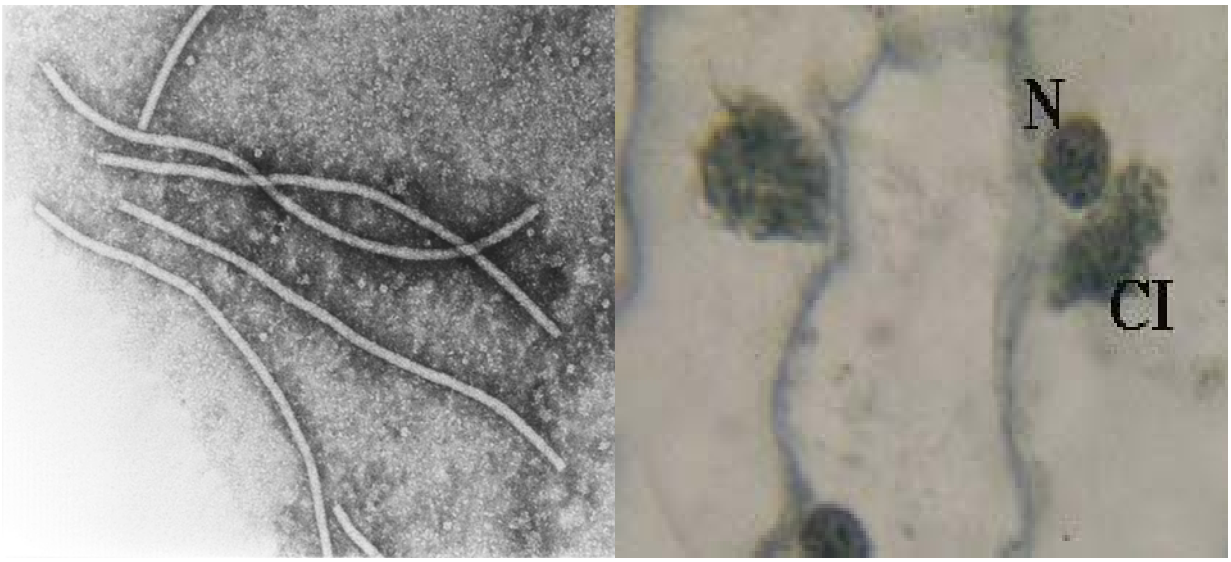

Figure 2: Electron microscopy picture of a potyvirus (http://www.virology.net/Big_ Virology/BVRNApoty.html) (left); light microscopy picture of SCMV with nucleus (N) and cylindrical inclusions (CI) (http://plantpath.ifas.ufl.edu/ pdc/Inclusionpage/Poty/poty.html) (right).

In a study with 122 early-maturing European maize inbreds, three lines (FAP1360A, D21, and D32) displayed complete resistance and four lines displayed partial resistance (FAP1396A, D06, D09, and R2306) against SCMV and maize dwarf mosaic virus (MDMV) (Kuntze et al. 1997). In field trials, resistance of all three European lines D21, D32, and FAP1360A seemed to be controlled by one to three genes (Melchinger et al. 1998). QTL and 'bulked segregant analysis' (BSA) mapped two major genes, *Scmv1* on the short arm of chromosome 6 and *Scmv2* near the centromere to chromosome 3 (Melchinger et al. 1998; Xia et al. 1999; Xu et al. 1999; Dußle et al. 2000). Minor QTL affecting SCMV resistance were identified on chromosomes 1, 5, and 10 (Xia et al. 1999). For complete resistance to SCMV, presence of both *Scmv1* and *Scmv2* is essential. *Scmv1* suppresses symptom expression

throughout all developmental growth stages at a high level, whereas *Scmv2* was mainly expressed at later stages of infection (Xia et al. 1999; Dußle et al. 2000).

In addition to the results of QTL analyses, SCMV resistance also shows physiological attributes of quantitative resistance. Lei and Agrios (1986) observed that SCMV (MDMV-B) replicated to high titres and spread locally in all genotypes, susceptible as well as resistant. Systemic spread was only evident for susceptible genotypes. In the inoculated leaves of susceptible plants the virus spread faster towards the proximal than the distal end of the leaf, whereas in the inoculated leaves of resistant plants the virus spread slowly and in similar rates towards both ends. These postinfectious mechanisms like the observed incomplete virus spread are typical for such quantitative resistances. The pattern of virus spread in the inoculated leaves of resistant plants suggests that the plant inhibits the virus from spreading through the leaf vascular system.

However, the two major genes, *Scmv1* and *Scmv2*, confer resistance independently of the isolate. They map to known R gene and RGA clusters on chromosomes 6 and 3, and therefore, the use of RGAs seems promising and provides a strong tool to identify candidate genes from these two target regions to accelerate the isolation of *Scmv1* and *Scmv2*.

Objectives

In this thesis I chose the example of SCMV resistance to evaluate the use of RGAs as tool for molecular breeding for disease resistance. The objectives of this thesis were to (i) give a review on the current status of virus resistance breeding in maize, (ii) identify RGAs, convert them to PCR-based markers, and map them to the potential R gene clusters on chromosomes 6 and 3, (iii) make use of putative sequence homologies of gene family members from the same genomic region to specifically identify further candidates in the target regions related to SCMV resistance, and (iv) investigate the implications of duplicated sequences like RGAs on the construction of genetic maps, MAS, and map-based cloning.

References

Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP (1997) Signalling in plant-microbe interactions. Science 276: 726-733

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell 8: 1757-1771

Biffen RH (1905) Mendel's laws of inheritance and wheat breeding. J. Agric. Sci. 1: 4-12

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogues in maize. Mol. Plant-Microbe Interact. 11: 968-78

Crute IR, Pink DAC (1996) Genetics and utilization of pathogen resistance in plants. Plant Cell 8: 1747-1755

de Wit PJGM (1992) Molecular characterization of gene-for-gene systems in plant-fungus interactions and the application of avirulence gene in control of plant pathogens. Annu. Rev. Phytopathol. 30: 391-418

Dußle CM, Melchinger AE, Kuntze L, Stork A, Lübberstedt T (2000) Molecular mapping and gene action of *Scm1* and *Scm2*, two major QTL contributing to SCMV resistance in maize. Plant Breed. 119: 299-303

Flor HH (1956) The complementary genetic systems in flax and flax rust. Adv. Genet. 8: 29-54

Flor HH (1971) The current status of the gene for gene concept. Annu. Rev. Phytopathol. 9: 275-296

Fuchs E, Kozelska S (1984) Zum Auftreten und zur Schadwirkung des Europäischen Mais Mosaik Virus (EMMV) in der DDR sowie Ergebnisse der Prüfung der Anfälligkeit von Inzuchtlinien und Hybriden. Nachrichtenblatt Pflanzenschutz DDR 38: 162-164

Fuchs E, Grüntzig M (1995) Influence of sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV) on the growth and yield of two maize varieties. J. Plant. Dis. Prot. 102: 44-50

Fuchs E, Grüntzig M, Kuntze L, Hohmann F (1996) Zur Epidemiologie der Potyviren des Maises in Deutschland. In Bericht über die 46. Arbeitstagung der Arbeitsgemeinschaft der Saatzuchtleiter im Rahmen der "Vereinigung österreichischer Pflanzenzüchter" 46: 43-49

Hammond-Kosack KE, Jones JDG (1996) Resistance gene-dependent plant defence responses. Plant Cell 8: 1773-1791

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48: 575-607

Johal GS, Briggs SP (1992) Reductase activity encoded by the *Hm1* disease resistance gene in maize. Science 258: 985-987

Kanazin V, Marek LF, Shoemaker RC (1996). Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. 93: 11746-11750

Keen NT (1992) Gene-for-gene complementarity in plant-pathogen interactions. Annu. Rev. Genet. 24: 447-463

Kobe B, Deisenhofer J (1994) The leucine-rich repeat: a versatile binding motif. Trends Biochem. Sci. 19: 415-421

Kuntze L, Fuchs E, Grüntzig M, Schulz B, Klein D, Melchinger AE (1997) Resistance of early maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breed. 116: 499-501

Lei JD, Agrios GN (1986) Mechanisms of resistance to maize dwarf mosaic virus. Phytopath. 76: 1034-1040

Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat. Genet. 14: 421-429

Louie R, Findley WR, Knoke JK, McMullen MD (1991) Genetic basis of resistance in maize to five maize dwarf mosaic virus strains. Crop. Sci. 31: 14-18

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

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. 20: 317-332

Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J. Mol. Evol. 50: 203-213

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop: a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15: 430-434

Shukla DD, Tosic M, Jilka J, Ford RE, Toler RW, Langham MAC (1989) Taxonomy of potyviruses infecting maize, sorghum and sugarcane in Australia and the United States as determined by reactivities of polyclonal antibodies directed towards virus-specific N-termini of coat proteins. Phytopath. 79: 223-229

Xia XC, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopath. 89: 660-667

Xu ML, Melchinger AE, Xia XC, Lübberstedt T (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. Mol. Gen. Genet. 261: 574-581

Yu YG, Buss GR, Saghaimaroof M (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding motif. Proc. Natl. Acad. Sci. 93: 11751-11756

UDC 575.827; 578.85
Review article

BREEDING FOR VIRUS RESISTANCE IN MAIZE

Marcel QUINT, Albrecht E. MELCHINGER, Christina M. DUßLE,
and Thomas LÜBBERSTEDT

Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

Marcel Quint, Albrecht E. Melchinger, Christina M. Dußle, and Thomas Lübberstedt (2000): *Breeding for virus resistance in maize*. – Genetika, Vol. 32, No. 3, 529-545.

Sugarcane mosaic virus (SCMV) is an important disease in maize, which is emerging in Germany since 1983. Using this pest as a model for the inheritance of oligogenic traits, we clarified the genetic basis for resistance in early maturing European maize germplasm. Screening of 122 adapted European inbred lines identified three completely resistant lines, which were used for further analyses. The genetics of SCMV resistance was investigated by allelism tests in field experiments combined with QTL and bulked segregant analyses (BSA) on the marker level. QTL analyses revealed the presence of two major genes Scm1 and Scm2 plus three minor QTL. Involvement of Scm1 and Scm2 in the inheritance of SCMV resistance could be confirmed by BSA in a second cross. Breeders can make use of tightly linked STS markers for marker-assisted selection (MAS) as well as our SCMV resistant flint lines to improve their elite germplasm. Currently, recurrent backcrossing with phenotypic selection is the most appropriate and cost effective breeding method. With decreasing costs of DNA chip technology, MAS can be competitive with phenotypic selection in the near future. Further objectives of our research are the isolation and cloning of Scm1 and Scm2. To achieve this goal we follow two different approaches. (1) Positional cloning based on more

Corresponding author: Marcel Quint,
Institute of Plant Breeding, Seed Science, and Population Genetics,
University of Hohenheim, Fruwirthstraße 21, 70593 Stuttgart, Germany
e-mail: quint@pz.uni-hohenheim.de

than 500 AFLP primer combinations resulted in Scm1/Scm2 specific markers with a resolution of approximately 0.2 cM in the respective regions. (2) Resistance gene analogues (RGAs), cosegregating with the target genes are used to identify further candidate genes for transformation experiments.

Key words: clustering, MAS, resistance genes, virus resistance, *Zea mays* L.

INTRODUCTION

There are more than 700 known plant pathogenic viruses of which more than 50 are pathogenic in maize (http://image.fs.uidaho.edu/vide). Particularly tropic, sub-tropic, and warm areas provide ideal conditions for the life cycle of viruses and their transmitting vectors. Viral diseases of maize are responsible for enormous economic losses. Intensive agricultural utilization of soils by monoculture and restriction of genetic diversity by the extensive cultivation of few varieties, have further aggravated the problem of viral diseases. Diverse approaches have been employed to minimize the losses caused by maize pathogenic viruses, such as chemical or integrated control of vectors, cultivation of resistant materials bred by conventional methods or marker-assisted selection (MAS) and transgenic crop cultivation. Plant protection by insecticides is ecologically and economically questionable. Transgenic approaches are still hampered by the missing acceptance of consumers. Regarding these aspects, breeding and cultivation of naturally resistant cultivars is currently the most promising way for the control of viral diseases in maize. Crucial items for effective resistance breeding are: (1) methods for fast and reliable evaluation of a large number of genotypes; (2) existance of natural resistance and sufficient genetic variation; (3) correct identification of the causative viruses, because symptoms alone can be misleading, (4) a good understanding of the genetic and physiological basis of resistance; and (5) choice of an efficient breeding method (KUNTZE *et al*. 1995).

This article reviews maize viral disease resistances, which have been intensively studied during the past 20 years. Starting with general information on maize virus diseases, we summarize genetic studies on virus resistance and discuss the prospects of conventional compared to new breeding methods using transgenes or molecular markers for virus resistance breeding in maize.

IMPORTANT MAIZE VIRUSES AND INHERITANCE OF RESISTANCE

Most maize virus diseases are of no economical importance. Hence, only some have been studied in greater detail. Tables 1 and 2 give basic information on the biology of those viruses and the inheritance of resistance. Symptom pictures can be obtained from the internet (http://www.agron.missouri.edu/Coop/images-/pesky.html). Starting from segregation analyses and tests of allelism, several models of inheritance have been proposed. Due to the complex nature of virus disease

Table 1. Choice of maize infecting viruses discussed in this review with general information

Virus	Taxonomy	Symptoms	Vector	Transmission	Geographical distribution
MDMV	*Potyviridae*	light green mosaic, stunting, casually reddening	insects, *Aphidiidae*	non-persistent, by mechanical inoculation, by seed	China, South Africa, USA, southern Europe
SCMV	*Potyviridae*	mosaic and/or ringspots, stunting	insects, *Aphidiidae*	non-persistent, by mechanical inoculation, by grafting, by seed	probably worldwide
WSMV	*Potyviridae*	light green streaky mosaic	mite, *Eriophydae*	by mechanical inoculation, by seed	Canada, Jordan, Romania, USA
HPV	not solved*	stunting, general necrosis mixed with mosaic, flecking, or streaking	mite, *Eriophydae*	by mechanical inoculaion	USA (central and west)
MCDV	*Sequiviridae*	stunting, reddening and yellowing, tertiary vein banding	insects, *Cicadellidae*	semi-persistent, does not multiply in the vector	USA (south)
MMV	*Rhabdoviridae*	green-yellow spotting, striping and vein-banding, some apical bending	insects, *Delphacidae*	persistent, multiplies in the vector	Australia, Colombia, Costa Rica, Fiji, India, Mauritius, Mexico, Peru, Spain, Tanzania, USA, Caribbean Islands
MSV	*Geminiviridae*	chlorotic or white streaking or lesions	insects, *Cicadellidae*	persistent, does not multiply in vector	African region, India, Madagascar, Reunion, Yemen

* HPV is morphologically similar to wheat spot chlorosis pathogen and wheat spot mosaic virus, but the issue of taxonomy is not solved.

resistances in maize, segregation analyses often did not give sufficient evidence for the number of genes involved in the inheritance of resistance. Gene number estimates were strongly affected by (1) the use of different resistant donor lines, (2) different susceptible parent lines in crosses for segregation analyses, and (3) environmental effects.

Two of the most intensively studied virus diseases are maize dwarf mosaic virus (MDMV) and sugarcane mosaic virus (SCMV). Both are closely related members of the *Potyviridae* family and can only be distinguished serologically and by capsid proteins (HOHMANN *et al*., 1996). SHUKLA *et al*. (1989) classified MDMV-B as a strain of sugarcane mosaic virus, apart from the other four MDMV strains. MDMV and SCMV were devastating in the late 1960s and still belong to the most widespread viral diseases in maize (GORDON *et al*., 1981) with the potential to reduce grain yield by as much as 45% (ROSENKRANTZ and SCOTT, 1978). In the maize-MDMV/SCMV pathosystem, complete qualitative and incomplete quantitative resistances have been identified. The qualitative resistance is characterized by extreme resistance, whereas quantitative resistance imparts decreased infection ratio, increased incubation time, decreased virus concentration, delayed virus movement, and decreased expression of symptoms (STEINBIß, 1993). LEI and AGRIOS (1986) investigated the mechanisms of resistance of SCMV (MDMV-B) in maize. In inoculated leaves of susceptible plants, spread of the virus was faster compared to resistant plants. This suggests that resistant plants inhibit the virus from moving through the leaf vascular system. Many efforts have been made to determine the genetic basis of resistance to MDMV using diallel crosses, backcrosses to resistant and susceptible parents, and reciprocal chromosomal translocations (LOUIE *et al*., 1991). Complete natural resistance to MDMV, SCMV as well as wheat streak mosaic virus (WSMV), was strongest in the inbred line Pa405. Segregation analyses indicated the presence of one to five genes conferring resistance to MDMV in Pa405 (FINDLEY *et al*., 1984; MIKEL *et al*., 1984; ROSENKRANTZ and SCOTT, 1987). By different approaches based on molecular markers, the major gene *Mdm1*, involved in inheritance of resistance against MDMV, was mapped to the short arm of chromosome 6 (Table 2). It perfectly cosegregates with the *nor* (SIMCOX *et al*., 1995). Similar to the situation with rust resistance genes in maize (SAXENA and HOOKER, 1974; SANZ-ALFEREZ *et al*., 1995), it cannot be ruled out that the *Mdm1* region involves a cluster of tightly linked R-genes.

In a study with 122 elite European maize inbreds, KUNTZE *et al*. (1995) identified three resistant lines (D21, D32, FAP1360A), conferring extreme resistance to MDMV as well as to SCMV. In field trials with segregating populations, resistance in these lines was apparently controlled by one to three genes, depending on the cross investigated. QTL analyses and 'bulked segregant analyses' (BSA) identified two major genes: *Scmv1* on the short arm of chromosome 6 and *Scmv2* near the centromere on chromosome 3 (MELCHINGER *et al*., 1998; XIA *et al*., 1999; XU *et al*., 1999; DUßLE *et al*., 2000). Minor QTL affecting SCMV resistance were

Table 2. Survey of studies on the inheritance of virus resistance by genetic mapping

Virus	Mapping approach	Number of genes	Map location (chromosome)	Mapping population	Reference	Resistant genotypes
MDMV	**a) Chromosomal translocation**	3	1, 4, 6	a) translocational lines	a) LOUIE *et al.* (1991)	Pa405, Oh1EP, Oh7B,
	b) cosegregation analysis	1	6	b) 20 BC1	b) LOUIE *et al.* (1991)	PB3187, A73Ohio, A239,
	c) BSA	1	6	c) 236 BC1	c) MCMULLEN and LOUIE (1989)	Va53, Va85, VaOm73,
	d) cosegregation analysis	1	6	d) 958 BC1, 7650 F2	d) SIMCOX *et al.* (1995)	ILM6161a, 6222a, 6222b, 6222c, 6222d, 6222e, 6223a, 6223b, Dekalb-Pfizer DK689, Jacques 840, McCurdy 85-60, D21, D32, FAP1360A
SCMV	**a) cosegregation analysis**	2	3, 6	a) 4 F2 pop. + 2 F2:3 lines	a) MELCHINGER *et al.* (1998)	Pa405, D21, D32, FAP1360A
	b) QTL	5	1, 3, 5, 6, 10	b) 220 F2:3 lines	b) XIA *et al.* (1999)	"
	c) QTL	2	3, 6	c) 121 F2:3 lines	c) DUßLE *et al.* (2000)	"
	d) BSA	2	3, 6	d) 40 BC5	d) XU *et al.* (1999)	"
WSMV	**a) cosegregation analysis**	1	6	a) BC1 + F2	a) MCMULLEN and LOUIE (1991)	most of the maize germplasm
	b) BSA	3	3, 6, 10	b) 61 F2	b) MCMULLEN *et al.* (1994)	
	c) cosegregation analysis	1	6	c) 2 F2 pops (60)	c) MARCON *et al.* (1997)	
	d) QTL	3	3, 6, 10	d) 129 RILs	d) MARCON *et al.* (1999)	
HPV	a) cosegregation analysis	1	6	a) 2 F2 pops (60)	a) MARCON *et al.* (1997)	B73, Mo17 (moderately
	b) QTL	2	3, 6	b) 129 RILs	b) MARCON *et al.* (1999)	susceptible)
MCDV	a) Chromosomal translocation	3	1, 3, 4	a) translocation lines	a) SCOTT and ROSENKRANTZ (1977)	Mp412, Mp444, Ky122, Tx29A
	b) QTL	2	3, 10	b) 87 F2	b) LOUIE *et al.* (1997)	
MMV	a) BSA	1	3	a) 91 RILs	a) MING *et al.* (1997)	Hi31
MSV	a) QTL	4	1, 2, 3, 4	a) 256 F2:3 lines	a) WELZ *et al.* (1998)	CML202, Tzi4, IB32, D211,
	b) QTL	1	1	b) 87 RILs	b) KYETERE *et al.* (1999)	CIRAD390
	c) QTL	5	1, 2, 3, 3, 10	c)165 F2:3 lines	c) PERNET *et al.* (1999a)	
	d) QTL	8	1, 2, 3, 5, 5	d) 191 F2:3 lines	d) PERNET *et al.* (1999b)	

identified on chromosomes 1, 5, and 10 (XIA *et al*., 1999). For extreme resistance, presence of *Scmv1* and *Scmv2* is essential. *Scmv1* suppresses symptom expression throughout all developmental growth stages at a high level, whereas *Scmv2* was expressed in later stages of infection (XIA *et al*., 1999; DUßLE *et al*., 2000). Thus, *Scmv1* is most valuable for SCMV control in hybrid breeding of maize. A high level of resistance can be obtained by selection for both major genes.

WSMV, another member of the *Potyviridae*, causes a serious disease in bread wheat. Most maize inbreds are resistant to WSMV, which appears to be only secondarily adapted to maize (MCMULLEN *et al*., 1994). Presence of WSMV in maize fields neighboring winter wheat may provide the opportunity that maize serves as an oversummering host for both WSMV and its mite vector (MCMULLEN and LOUIE, 1991). Because there are no immune varieties of wheat, considerable efforts have been made to introduce WSMV resistance from resistant cereal relatives into bread wheat (MCMULLEN and LOUIE, 1991). Segregation ratios suggested presence of multiple genes for resistance to WSMV in maize (MCMULLEN and LOUIE, 1991). Mapping with RFLPs identified three genes for resistance in Pa405: *Wsm1* on chromosome 6, *Wsm2* on chromosome 3, and *Wsm3* on chromosome 10. For resistance, each of the genes *Wsm1*, *Wsm2*, or *Wsm3* seems to be sufficient to reduce symptom expression in maize. Presence of multiple resistance genes might explain the high degree of WSMV resistance in maize (MCMULLEN *et al*., 1994).

High Plains Virus (HPV), infecting both maize and wheat, is the causal agent of the 'High Plains Disease'. Even though HPV and WSMV are unrelated viruses, they generally occur together as a mixed infection in the field due to the same transmitting vector (wheat curl mite) (JENSEN *et al*., 1994). So far, it has not been possible to isolate a pure culture of HPV or inoculate efficiently with HPV alone. Therefore, it is necessary to work with HPV/WSMV double-infected wheat curl mites for infection trials. MARCON *et al*. (1999) reported that the same genomic regions as for WSMV resistance on chromosomes 3 and 6 are involved in the inheritance of resistance to HPV. It is not known whether these major loci are identical or tightly linked to each other. Presence of HPV resistance on chromosomes 6 and 3 increases the number of disease resistance loci mapping to these genome regions.

Maize chlorotic dwarf virus (MCDV) was very likely the causing pathogen of 'corn stunting disease' in the 1960s. MCDV can be controlled by reducing leafhopper vectors with insecticides or by eliminating the infected overwintering host Johnson grass with herbicides (FOY and WITT, 1990). Although currently no maize germplasm displays complete resistance to MCDV, some inbreds are tolerant to MCDV infection (JOSEPHSON and SCOTT, 1981). Studies based on chromosomal translocations identified three genes for resistance on chromosomes 1, 3, and 4 (SCOTT and ROSENKRANTZ, 1977). Since technical problems due to leafhopper behaviour are solved (multiple-inoculation method) (LOUIE and ANDERSON, 1993), LOUIE *et al*. (1997) identified two major QTL on chromosomes 3 and 10 affecting

resistance/tolerance to MCDV using an inbred line derived from a Virgin Island population.

The maize mosaic nucleorhabdovirus (MMV) causes a major disease of maize in many tropical and sub-tropical regions (BREWBAKER, 1981). In bad years, entire fields can be dwarfed below 50 cm in height with no kernels produced under severe epiphytotics (BREWBAKER, 1979). MMV is infectious in almost all temperate-zone hybrids. The first resistance source was recognized in Cuban flint materials imported to Hawaii (MING *et al*., 1997). Using a BSA approach based on RFLPs, MING *et al*. (1997) mapped the single major gene *Mv1* near the centromere on chromosome 3, in the same region as *Wsm2*, *Scmv2*, as well as HPV, and MCDV resistance.

Maize streak virus (MSV), is a major pathogen throughout the sub-Saharan Africa. MSV outbreaks have been erratic over years and seasons depending on the insect vector population, which is favored by low rainfall and high temperatures, with devastating epidemics and complete crop failure in bad years (GUTHRIE, 1978; KIM *et al*., 1989). The first resistance sources for MSV were already observed in the 1930s (FIELDING, 1933). Several groups detected a single resistance gene on chromosome 1, designated *Msv1*, in different inbred lines (KYETERE, 1995; WELZ *et al*., 1998). Three adjacent marker loci from chromosome 9, displaying minor QTL, might be involved in a second mechanism of MSV resistance. KYETERE *et al*. (1999) hypothesized existance of two different systems for resistance. One, including major genes like *Msv1* controlling complete resistance and a second one with minor genes, controlling partial resistance. Additional QTL were detected by PERNET *et al*. (1999a, b).

The dominant nature of most R-genes is favorable for hybrid seed production. Resistance present in one parent is sufficient and the breeder is more flexible in the choice of parents.

BSA turned out to be the quickest method for the identification of markers linked to major genes. For BSA certain assumptions are necessary: (a) reliable phenotypic scoring of resistance, and (b) oligogenic or ideally monogenic inheritance. Minor genes can only be identified by QTL analyses and are often influenced by environmental factors. Therefore, QTL need to be investigated in trials across multiple environments.

CLUSTERING OF VIRUS RESISTANCE GENES?

The number of mapped qualitative or quantitative disease resistance loci has considerably increased over the past ten years. With each gene mapped, it became evident that disease resistance genes (R-genes), including virus R-genes, are not randomly distributed over the maize genome (MCMULLEN and SIMCOX, 1995). There are certain regions in the genome, where virus R-genes seem to be tightly linked or clustered. Figure 1 shows the genome positions of the virus R-genes mapped hitherto. The majority of these loci are located in chromosomal bins which are also containing other disease R-genes. Two major clusters containing virus R-

genes are on the short arm of chromosome 6 (bin 6.00-6.01) and on the short arm of chromosome 3 near the centromere (bin 3.04-3.05).

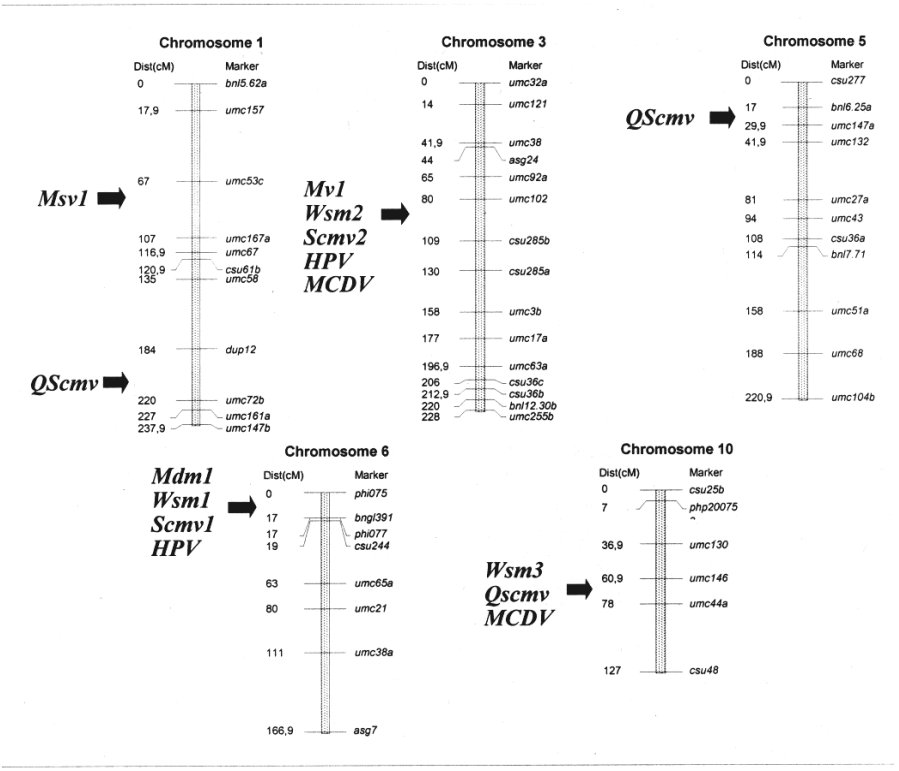

Fig. 1. Map locations of maize virus resistance genes by molecular markers

The cluster on chromosome 6 harbors besides a recessive gene, *Rhm1*, conferring resistance to the fungal pathogen *Cochliobolus heterostrophus* (Drechs.) Drechs. Race O (ZAITLIN *et al*., 1993), and three virus R-genes: *Mdm1* (MCMULLEN and LOUIE, 1989; LOUIE *et al*., 1991; SIMCOX *et al*., 1995), *Scmv1* (MELCHINGER *et al*., 1998), and *Wsm1* (MCMULLEN and LOUIE 1991; MCMULLEN *et al*. 1994), which confer resistance to the potyviruses MDMV, SCMV, and WSMV. Another member of the cluster might be a gene conferring resistance to the HPV. Because the WSMV resistant line B73 is highly susceptible to MDMV and SCMV (MCMULLEN and LOUIE, 1991), *Mdm1* and *Scmv1* are presumably not identical to *Wsm1*. However, a resistance allele for one gene, i. e. *Mdm1*, might be a susceptibility allele for the other virus, e. g. WSMV, and vice versa. MDMV and SCMV are closely related members of the *Potyviridae* family and identity between the respective R-genes cannot be ruled out. However, there is evidence suggesting that *Mdm1* and *Scmv1* are not identical. SIMCOX *et al*. (1995) found no recombination between *Mdm1* and the *nor* in a very large population of 7650 F_2 individuals,

whereas XU *et al*. (1999) identified 4 plants with recombination to the *nor* in a BC5 population consisting of only 40 SCMV resistant individuals suggesting that *Mdm1* and *Scmv1* are different. XU *et al*. (2000) performed AFLP-based pedigree analyses to clarify the origin of *Scmv1* and *Scmv2* in the ancestors of the three resistant European inbreds and Pa405. The results suggest identity of *Scmv1* in all three European inbreds, while the AFLP banding pattern of Pa405 in this region was completely different.

Another resistance gene against HPV was mapped to the same region in a HPV/WSMV-combined investigation (MARCON *et al*., 1997; 1999). The fact that HPV is not related to MDMV and SCMV suggests distinct, tightly linked genes controlling resistance to HPV and the two potyviruses (MARCON *et al*., 1997).

The resistance gene cluster on chromosome 3S near the centromere also includes virus R-genes. It contains the virus R-genes *Scmv2* (MELCHINGER *et al*., 1998), *Wsm2* (MCMULLEN *et al*., 1994) and *Mv1* (MING *et al*., 1997) and further major genes conferring resistance to HPV and MCDV (MARCON *et al*., 1999). As for chromosome 6, the question of a single gene with pleiotropic action or tightly linked genes still needs to be resolved. Additional nonviral linked R-genes are mapping to the same region, such as the *Rp3* locus, conferring resistance to *Puccina sorghi* (Schwein.) (SANZ-ALFEREZ *et al*., 1995), and QTL for resistance to *Fusarium* stalk rot, and the European corn borer (PE *et al*., 1993; SCHÖN *et al*., 1993).

Even if there is strong evidence for clustering of resistance genes, there is so far no firm proof for the existance of gene clusters. However, cloning and complementation of the respective R-genes may resolve this issue. Besides the clustering of R-genes in maize (BENNETZEN *et al*., 1991; SANZ-ALFEREZ *et al*., 1995; MCMULLEN and SIMCOX, 1995), clusters of disease R-genes have already been reported in other plant genera, e.g., the *MRC* loci in *Arabidopsis* (KUNKEL, 1996; HOLUB, 1997), the *N* locus in flax (ELLIS *et al*., 1997), and the *Cf* loci in tomato (KALOSHIAN *et al*., 1995; JONES *et al*., 1993). Furthermore, YU *et al*. (1996) observed colinearity between maize, barley and oat in genome regions containing clusters of R-genes.

NEW METHODS IN RESISTANCE BREEDING

Our knowledge of the genomic positions of R-genes and the rapid development of molecular techniques opened new avenues such as direct transfer of R-genes into susceptible elite germplasm for resistance breeding. Cloning of R-genes and pathogen-derived resistance offer new approaches for breeding of resistant lines and varieties. There are basically two methods of gene cloning: a) positional cloning such as chromosome walking or chromosome landing, and b) transposon tagging.

Positional cloning. - Cloning a gene by chromosome walking or chromosome landing (TANKSLEY *et al*., 1995) requires a high-resolution linkage map around the gene of interest. Furthermore, recovery of individuals with crossovers between the target locus and flanking markers is a prerequisite for determining

direction and physical distance during a chromosome walk (SIMCOX *et al*., 1995). These high-resolution maps can be used as a starting point for map-based cloning of a number of disease resistance traits. The high amount of repetitive DNA can make positional cloning in maize a difficult process. The existing colinearity between the genomes of maize and other related members of the grasses can be used to overcome the difficulties of positional cloning in species with large genomes like maize. By using BAC- or YAC-libraries of related but smaller genomes, establishing colinear relationships of R-genes across species could accelerate isolation of genes and their effects on resistance (MCMULLEN and SIMCOX, 1995). On the other hand, LEISTER *et al*. (1998) did not find evidence for a monocot-specific signature of R-genes. Interspecific analyses of RGAs frequently revealed non-syntenic map locations between the cereal species rice, barley, and foxtail millet although tight colinear gene order is a hallmark of monocot genomes.

Transposon tagging. - Maize contains several active transposable element systems which have been successfully used for cloning genes by transposon tagging. The first cloned disease R-gene in plants was *Hm1* in maize by means of transposon induced mutagenesis (JOHAL and BRIGGS, 1992). Hitherto, several disease R-genes in other plant species have been cloned using the maize *Ac* transposable element (JONES *et al*., 1994; WHITHAM *et al*., 1994). DINESH-KUMAR *et al*. (1995) cloned the first virus R-gene *N*, which confers resistance to tobacco mosaic virus (TMV) in tobacco. In association with the previously mentioned clustering of R-genes, a very promising attribute of transposable elements is their tendency to insert more frequently into physically adjacent linked sites than to more distant chromosome regions (DELLAPORTA and MORENO, 1994). Using this feature, genome regions containing clusters of R-genes could be more efficiently screened for transposon tagging by preselecting active transposable element populations with transposons linked to the region of interest (CHANG and PETERSON; 1994; MCMULLEN and SIMCOX, 1995).

Table 3. Survey of cloned virus resistance genes in general and cloned resistance genes originating from maize

Pathogen	Host	Gene	Approach	Reference
TMV	tobacco	*N*	transposon	DINESH-KUMAR *et al.* (1995)
PVX	potato	*Rx1*	Positional cloning	BENDAHMANE *et al.* (1999)
		Rx2	Candidate gene	BENDAHMANE *et al.* (2000)
Cochliobolus carbonum	maize	*Hm1*	transposon	JOHAL and BRIGGS (1992)
Puccinia sorghi	maize	*Rp1-D*	Positional cloning	COLLINS *et al.* (1999)

While TMV resistance gene *N* was cloned by transposon tagging, BENDAHMANE *et al*. (1999) employed map-based/positional cloning for the isolation of the second virus R-gene *Rx*, conferring resistance to potato virus X (PVX). Cloned virus R-genes and other R-genes originating from maize are summarized in Table 3.

Resistance gene analogues. - Although only about twenty disease R-genes in plants have been cloned, it is already obvious that they can be grouped on the basis of their deduced amino acid sequence. The sequence of R-gene products seems to be conserved in regions that display protein-protein interaction functions, which have already been successfully used by several groups creating PCR primers to generate 'resistance gene analogues' (RGAs) or 'resistance gene homologues'. (KANAZIN *et al*., 1996; YU *et al*., 1996). COLLINS *et al*. (1998) identified 20 RGA loci in maize, mapping mostly to regions known to carry R-genes like the previously mentioned R-gene clusters on chromosomes 3 and 6. These RGAs can be used in a 'candidate gene approach' to isolate target genes or as genetic markers for marker-assisted selection (MAS). This approach has been successfully used in maize to clone the maize *Rp1-D* gene, a member of the complex locus composed of approximately nine gene homologues, conferring resistance to common rust (*Puccinia sorghi*) (COLLINS *et al*., 1999) and the *Rx2* virus resistance gene in potato (BENDAHMANE *et al*., 2000). Conservation of R-genes and the colinearity of grass genomes can be used to transfer sequence information among grass species making it an important tool for the isolation of R-genes in related species. Cloned R-genes can also serve as perfect molecular markers.

Pathogen-derived resistance. - A totally different idea is the concept of pathogen-derived resistance (SANFORD and JOHNSTON, 1985). Viral RNA or the coat protein are usually regarded as the main resistance triggering factors. Various authors (DE ZOETEN and FULTON, 1975; SHERWOOD and FULTON, 1987) hypothesized that the coat protein of the first infecting viruses prevents replication of the following penetrating virus particles by coating the viroids before they are able to replicate. One of the first hypotheses in this context was the interaction of the coat protein with the *replicase* binding site on the minus strand of the replicating virion. There have been numerous attempts to generate virus resistance in transgenic plants through the expression of virus-derived genes (BEACHY, 1993; WILSON *et al*., 1993). Many of these attempts have been successful, and some have led to the development of virus-resistant potato cultivars for commercial use (BAULCOMBE, 1996). Besides coat protein-mediated resistance there are several other methods of pathogen-derived resistance, such as gene silencing, antisense RNA, RNA/DNA-mediated resistance or replicase-mediated resistance (BAULCOMBE, 1996). MURRY *et al*. (1993) reported the first example of coat protein-mediated protection in transformed maize using the coat protein of SCMV. They regenerated fertile plants, which expressed high levels of SCMV coat protein and were resistant to SCMV as well as MDMV and maize chlorotic mottle virus (MCMV). Furthermore, they demonstrated the expected segregation of the introduced gene in F_1 progeny.

Coat protein-mediated resistance is only one of several methods to generate pathogen-induced resistance. In contrast to the complexity of genetic resistance, expression of a single, dominant transgene can confer resistance to several virus strains (MURRY *et al*., 1993). By this approach it is possible to circumvent the incorporation of many R-genes in extensive backcrossing, which are needed in conventional breeding programs. With increasing public acceptance of genetically

engineered organisms transgenic approaches offer great possibilities in maize virus resistance breeding.

Generally, the routine and efficient application of transformation is still not yet available. Problems in control of expression, stability and inheritance of transgenes have led to the fact that presently only two transgenic events (Bt) are commercially used in maize. These problems, patent restrictions and limited public acceptance of transgenic varieties might aggravate a widespread use of transgenic resistances.

SELECTION METHODS

We can distinguish basically quantitative and qualitative resistance. For each mechanism of inheritance the breeder has to heed various parameters resulting in different optimal breeding methods. With the development of molecular marker technology and its indisputable advantages (Table 4), breeders have to figure out the most appropriate and cost effective breeding method.

Table 4. Comparison of marker-assisted selection (MAS) and phenotypic selection for breeding against virus resistance in maize

Selection criteria	MAS	Phenotypic selection
Artificial inoculation	not necessary	necessary when no natural infection pressure present
Date of scoring	seedlinge stage	later stages of plant development
Equipment	laboratory	inoculation supply
Scoring	time consuming (DNA isolation, PCR assay)	fast visual or serological scoring of large numbers
Required population size	small	large
Introgression of QTL	major and minor QTL	only major QTL
Double crossover observable	yes	no
Linkage drag	small	big
Escapes	no (excluding double crossovers)	yes ✍ progeny testing in multiple environments
Costs per 100 plants in $	600	70

Quantitative resistance. - For quantitatively inherited resistances the breeder can employ phenotypic and marker-assisted selection (MAS). Prerequisite for phenotypic selection is either a reliable scoring method or serological methods like ELISA and tissue immuno blot prints (TIBP). Application of MAS requires the availability of an adequate number of molecular markers, which is given for maize. In the case of SCMV, the costs per 100 plants are $70 for phenotypic selection and $600 for MAS (Table 4). However, in contrast to phenotypic selection, MAS helps to select not only for major QTL, but also for minor QTL. The main

part of the high costs for MAS is due to expensive and laborious DNA extraction. Automatisation of DNA extraction, the application of multiplex PCR methods and new developments in DNA-chip technology might decrease the high costs for MAS compared to phenotypic selection significantly in the near future.

Qualitative resistance. - The features of MAS and phenotypic selection of course remain the same for application in mono- or oligogenic inherited resistances. For monogenic resistances MAS is an important tool to increase the selection intensity and minimize linkage drag in resistance breeding against maize virus diseases. The number of backcross generations can be decreased significantly compared to phenotypic selection. Such a scheme could be easily integrated into a marker-assisted backcross program for the rapid recovery of the donor genome (FRISCH *et al*., 1999). It is also promising to combine several R-genes for the same pathogen in a single genotype. Such marker-assisted pyramidisation of virus R-genes for the same pathogen, enables the generation of more stable and durable resistances. This approach is suitable especially for virus resistances.

Viruses are the pathogens with the smallest genome, resulting in lower chances to break down qualitative resistances by the evolution of virulence genes. For geographical regions with high natural infection pressure of more than one maize infecting virus, besides incorporation of virus R-genes for the same virus, the generation of multiple-resistant lines with resistance to several viruses is another possibility for the breeder to prevent viral diseases in maize. In this context, the clustering of R-genes increases the challenge for breeders attempting to produce virus resistant lines. Large populations have to be evaluated to identify rare recombination events that enable the detection of genotypes conferring resistance to each pathogen of the cluster. Therefore, extensive crossing experiments are necessary, which could be significantly supported by MAS.

Acknowledgements. - This work was supported by the Deutsche Forschungsgemeinschaft (DFG), grant LU601/2-3.

Received October 20th, 2000
Accepted November 7th, 2000

REFERENCES

BAULCOMBE D.C. (1996): Mechanisms of pathogen-derived resistance to viruses in transgenic plants. Plant Cell *8*, 1833-1844.

BEACHY R.N. (1993): Transgenic resistance to plant viruses. Semin. Virol. *4*, 327-416.

BENDAHMANE A., K. KANYUKA, and D.C. BAULCOMBE (1999): The *Rx* gene from potato controls seperate virus resistance and cell death responses. Plant Cell *11*, 781-791.

BENDAHMANE A., M. QUERCI, K. KANYUKA, and D.C. BAULCOMBE, 2000. Agrobacterium transient expression system as a tool for the isolation of disease resistance genes: application to the *Rx2* locus. Plant Journal *21*(1), 73-81.

BENNETZEN J.L., S.H. HULBERT, and P.C. LYONS (1991): Genetic fine structure analysis of a maize disease resistance gene. *In*: Molecular strategies of pathogens and host plants. Springer-Verlag, Berlin, Germany. pp.: 177-188.

BREWBAKER J.L. (1981): Resistance to maize mosaic virus. *In*: "Virus and viruslike diseases of maize in the United States". Gordon D.T., J.K. Knoke, and G.E. Scott (Eds), So. Coop. Ser. Bull., *247*, 145-151.

CHANG R.Y. and P.A. PETERSON (1994): Chromosome labeling with transposable elements in maize. Theor. Appl. Gen., *87*, 650-656.

COLLINS N., J. DRAKE, M. AYLIFFE, Q. SUN, J. ELLIS, S. HULBERT, and T. PRYOR (1999): Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell, *7*, 1356-1366.

COLLINS N., C.A. WEBB, S. SEAH, E.G. ELLIS, S.H. HULBERT, and A. PRYOR (1998): The isolation and mapping of disease resistance gene analogs in maize. Molecular Plant-Microbe Interactions, *11* (10), 968-978.

DELLAPORTA S.L. and M. MORENO (1994): Gene-tagging with Ac-Ds elements in maize. *In*: "The Maize Handbook", M. Freeling and V. Walbot (Eds), Springer Verlag New York. pp: 219-233.

DEZOETEN G.A. and R.W. FULTON (1975): Understanding generates possibilities. Phytopathology *65*, 221-222.

DINESH-KUMAR S.P., S. WHITHAM, D. CHOI, R. HEHL, C. CORR, and B. BAKER (1995): Transposon tagging of tobacco mosaic virus resistance gene *N*: Its possible role in TMV-N-mediated signal transduction pathway. Proc. Natl. Acad. Sci. USA, *92*, 4175-4180.

DUßLE C.M., A.E. MELCHINGER, L. KUNTZE, A. STORCK, and T. LÜBBERSTEDT (2000): Molecular mapping and gene action of *Scm1* and *Scm2*, two major QTL contributing to SCMV resistance in maize. Plant Breeding *119*(4), 299-304.

ELLIS J., G. LAWRENCE, M. AYLIFFE, P. ANDERSON, N. COLLINS, J. FINNEGAN, D. FROST, J. LUCK, and T. PRYOR (1997) Advances in the molecular genetic analysis of the flax-flax rust interaction. Annu. Rev. Phytopathol., *35*, 271-291.

FIELDING W.L. (1933): Field experimental work on rotation crops. Empire Cotton Growing Association Progress Report 1931-1932. pp.: 10-14.

FINDLEY W.R., R. LOUIE, and J.K. KNOKE (1984): Breeding corn for resistance to corn viruses in Ohio. Proc. 41st Annu. Corn Sorg. Res. Conf., pp: 52-67.

FOY C.L. and H.L. WITT (1990): Johnsongrass control with DPX-V9360 and CGA-136872 in corn (*Zea mays*) in Virginia. Weed Technol., *4*: 615-619.

FRISCH M., M. BOHN, and A.E. MELCHINGER (1999): Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci., *39*, 967-975.

GORDON D.T., O.E. BRADFUTE, R.E. GINGERY, R.E. KNOKE, R. LOUIE, L.R. NAULT, and G.E. SCOTT (1981): Introduction: History geographical distribution, pathogen characteristics, and economic importance. In: "Virus and Viruslike Diseases of Maize in the United States". D.T. Gordon, J. K. Knoke, and G. E. Scott. (eds), South Coop Ser Bull *247*: 1-12.

GUTHRIE E.J. (1978): Measurement of yield losses caused by maize streak disease. Plant Dis. Rep., *62*: 839-840.

HOLUB E.B. (1997): Organization of disease resistance genes in *Arabidopsis*. *In*: "The gene-for-gene relationship in host-parasite interactions". Crute I., E. Holub, and J. Burden (Eds), CAB International, Wallingford, UK. pp. 5-26.

HOHMANN F., E. FUCHS, M. GRUNTZIG, and L. KUNTZE (1996): Development of a routine test for the detection of sugarcane mosaic virus (SCMV) and the maize dwarf mosaic virus (MDMV) using tissue print-immunoblotting (TPIB). Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz *103*(5), 464-471.

JENSEN S.G. (1994): The high plains virus - a new threat to corn and wheat production in the west. Proc. 49th Annu. Corn Sorg. Res. Conf., pp.: 156-164.

JOHAL G.S. and S.P. BRIGGS (1992): Reductase activity encoded by the *Hm1* disease resistance gene in maize. Science *258*, 985-987

JONES D.A., M.J. DICKINSON, P.J. BALINT-KURTI, M.S. DIXON, and J.D.G. JONES (1993), Two complex resistance loci revealed in tomato by classical and RFLP mapping of the *Cf-2*, *Cf-4*, *Cf-5*, and *Cf-9* genes for resistance to *Cladiosporum fulvum*. Molecular Plant-Microbe Interactions, *6*, 348-357.

JONES D.A., C.M. THOMAS, K.E. HAMMOND-KOSACK, P.J. BALINT-KURTI, and J.D.G. JONES (1994): Isolation of the tomato *Cf-9* gene for resistance to *Cladiosporum fulvum* by transposon tagging. Science *266*, 789-793.

JOSEPHSON L.M. and G.E. SCOTT (1981): Sources of disease resistance in corn. *In*: "Virus and viruslike diseases of maize in the United States". D.T. Gordon, J.K. Knoke and G.E. Scott (Eds), South Coop Ser Bull *247*, 132-136.

KALOSHIAN I., W.H. LANGE, and V.M. WILLIAMSON (1995): An aphid-resistance locus is tightly linked to the nematode resistance gene *Mi* in tomato. Proc. Natl. Acad. Sci. USA, *92*, 622-625.

KANAZIN V., L.F. MAREK, and R.C. SHOEMAKER (1996): Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. USA, *93*, 11746-11750.

KIM S.K., Y. EFRON, J.M. FAJEMISIN, and I.W. BUDDENHAGEN (1989): Mode of gene action for resistance in maize streak virus. Crop Sci., *29*, 890-894.

KUNKEL B.N. (1996): A useful weed put to work - genetic analysis of disease resistance in *Arabidopsis thaliana*. Trends in Genetics, *12*, 63-69.

KUNTZE L., E. FUCHS, M. GRÜNTZIG, B. SCHULZ, U. HENNING, F. HOHMANN, and A.E. MELCHINGER (1995): Evaluation of maize inbred lines for resistance to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Agronomie *15*, 463-467.

KYETERE D.T. (1995): Genetic basis of tolerance in maize to maize streak virus using molecular markers. Ph.D. dissertation, Ohio State University

KYETERE D.T., R. MING, M.D. MCMULLEN, R.C. PRATT, J. BREWBAKER, and T. MUSKET (1999): Genetic analysis of tolerance to maize streak virus in maize. Genome, *42*, 20-26.

LEI J.D. and G.N. AGRIOS (1986): Mechanisms of resistance in corn to maize dwarf mosaic virus. Phytopathology *76*(10), 1034-1040.

LEISTER D., J. KURTH, D.A. LAURIE, M. YANO, T. SASAKI, K. DEVOS, A. GRANER and P. SCHULZE-LEFERT (1998): Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA, *95*, 370-375.

LOUIE R. and R.J. ANDERSON (1993): Evaluation of maize chlorotic dwarf virus resistance in maize with multiple inoculations by *Graminella nigrifons* (Homoptera: Cicadellidae). Jour. Econ. Entom. *86*(5). 1579-1583.

LOUIE R., W.R. FINDLEY, J.K. KNOKE and M.D. MCMULLEN (1991): Genetic basis of resistance in maize to five maize dwarf mosaic virus strains. Crop Sci., *31*, 14-18.

LOUIE R., M. JONES, and R. ANDERSON (1997): Symptomology of maize chlorotic dwarf (MCD) and identification of loci for MCD tolerance in carribean maize. Maize Genetics Conference Abstracts 39.

MARCON A., S.M. KAEPPLER, and S.G. JENSEN (1997): Resistance to systemic spread of high plains virus and wheat streak mosaic virus cosegregates in two F_2 maize populations inoculated with both pathogens. Crop Sci., *37*, 1923-1927.

MARCON A., S.M. KAEPPLER, S.G. JENSEN, L. SENIOR, and C. STUBER (1999): Loci controlling resistance to high plains virus and wheat streak mosaic virus in a B73xMo17 population of maize. Crop Sci., *39*, 1171-1177.

MCMULLEN M.D., M.W. JONES, K.D. SIMCOX, and R. LOUIE (1994): Three genetic loci control resistance to wheat streak mosaic virus in the maize inbred Pa405. Molecular Plant-Microbe Interactions, *7*(6), 708-712.

MCMULLEN M.D. and R. LOUIE (1989): The linkage of molecular markers to a gene controlling the symptom response in maize to maize dwarf mosaic virus. Molecular Plant-Microbe Interactions *2*(6), 309-314.

MCMULLEN M.D. and R. LOUIE (1991): Identification of a gene for resistance to wheat streak mosaic virus in maize. Phytopathology, *81*, 624-627.

MCMULLEN M.D. and K.D. SIMCOX (1995): Genomic organization and insect resistance genes in maize. Molecular Plant-Microbe Interactions *8*(5), 811-815.

MELCHINGER A.E., L. KUNTZE, R.K. GUMBER, T. LÜBBERSTEDT, and E. FUCHS (1998): Genetic basis of resistance to sugarcane mosaic virus in european maize germplasm. Theor. Appl. Gen., *96*, 1151-1161.

MIKEL M.A., C.J. D'ARCY, A.M. RHODES, and R.E. FORD (1984): Genetics of resistance of two dent corn inbreds to maize dwarf mosaic virus and transfer of resistance into sweet corn. Phytopathology *74*(4), 467-473.

MING R., J.L. BREWBAKER, R.C. PRATT, T.A. MUSKET, and M.D. MCMULLEN (1997): Molecular mapping of a major gene conferring resistance to maize mosaic virus. Theor. Appl. Gen., *95*, 271-275.

MURRY L.E., L.G. ELLIOTT, S.A. CAPITANT, J.A. WEST, K.K. HANSON, L. SCARAFIA, S. JOHNSTON, C. DELUCA-FLAHERTY, S. NICHOLS, D. CUANNAN, P.S. DIETRICH, I.J. METTLER, S. DEWALD, D.A. WARNICK, C. RHODES, R. SINIBALDI, and K.J. BRUNKE (1993): Transgenic corn plants exrpressing MDMV strain B coat protein are resistant to mixed infections of maize dwarf mosaic virus and maize chlorotic mottle virus. Bio/Technology *11*, 1559-1564.

PE M.E., L. GIANFRANCESCHI, G. TARAMINO, R. TARCHINI, P. ANGELINI, M. DANI, and G. BINELLI (1993): Mapping quantitative trait loci (QTL) for resistance to *Gibberella zeae* infection in maize. Mol. Gen. Genet., *241*, 11-16.

PERNET A., D. HOISINGTON, J. FRANCO, M. ISNARD, D. JEWELL, C. JIANG, J.-L. MARCHAND, B. REYNARD, J.-C. GLASZMANN, and D. GONZALEZ DE LEON (1999a): Genetic mapping of maize streak virus resistance from the Mascarene source. I. Resistance in line D211 and stability against different virus clones. Theor. Appl. Genet., *99*, 524-539

PERNET A., D. HOISINGTON, J. DINTINGER, D. JEWELL, C. JIANG, M. KHAIRALLAH, P. LETOURMY, J.-L. MARCHAND, J.-C. GLASZMANN, and D. GONZALEZ DE LEON (1999b): Genetic mapping of maize streak virus resistance from the Mascarene source. II. Resistance in line CIRAD390 and stability across germplasm. Theor. Appl. Genet., *99*, 540-553.

ROSENKRANTZ E. and G.E. SCOTT (1978): Effect of plant age at time of inoculation with maize dwarf mosaic virus on disease development and yield in corn. Phytopathology *68*, 1688-1692.

ROSENKRANTZ E. and G.E. SCOTT (1987): Type of gene action in the resistance to maize chlorotic dwarf virus in corn. Phytopathology *77*(9), 1293-1296.

SANFORD J.C. and S.A. JOHNSTON (1985): The concept of parasite-derived resistance - Deriving resistance genes from the parasite's own genome. J Theor. Biol., *113*, 395-405.

SANZ-ALFEREZ S., T.E. RICHTER, S.H. HULBERT, and J.L. Bennetzen (1995): The *Rp3* disease resistance locus of maize: Mapping and characterization of introgressed alleles. Theor. Appl. Genet., *91*, 25-32.

SAXENA K.M.S. and A.L. HOOKER (1974): A study on the structure of gene *Rp3* for rust resistance in *Zea mays*. Can. J. Genet. Cytol., *16*, 857-860.

SCHÖN C.C., M. LEE, A.E. MELCHINGER, W.D. GUTHRIE, and W.L. WOODMAN (1993): Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. Heredity *70*, 648-659.

SCOTT G.E. and E. ROSENKRANTZ (1977): Location of genes conditioning resistance to the corn stunting disease complex in maize. Crop Sci., *17*, 923-925.

SHERWOOD J.L. and R.W. FULTON (1987): Demonstration of the specific involvement of coat protein in tobacco mosaic virus (TMV) cross protection using a TMV coat protein mutant. J. Phytopathology, *118*, 358-362.

SHUKLA D.D., M. TOSIC, J. JILKA, E. FORD, R.W. TOLER, and M.A.C. LANGHAM (1989): Taxonomy of potyviruses infecting maize, sorghum, and sugarcane in Australia and the United States as determined by reactivities of polyclonal antibodies directed towards virus-specific N-Termini of coat proteins. Phytopathology *7*(2), 223-229.

SIMCOX K.D., M.D. McMULLEN, and R. LOUIE (1995): Co-segregation of the maize dwarf mosaic virus resistance gene, *Mdm1*, with the nucleolus organizer region in maize. Theor. Appl. Genet., *90*, 341-346.

STEINBIß H.-H. (1993): Gentechnik zur Erzeugung von Virusresistenz. *In*: "Resistenz von Kulturpflanzen gegen pflanzenpathogene Viren". Kegler, H. and W. Friedt (Eds), Gustav Fischer Verlag, pp.: 214-231.

TANKSLEY S.D., M.W. GANAL, and G.B. MARTIN (1995): Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends in Genetics *11*(2), 63-68.

WELZ H.G., A. SCHECHERT, A. PERNET, K.V. PIXLEY, and H.H. GEIGER (1998): A gene for resistance to the maize streak virus in the African CIMMYT maize inbred line CML202. Molecular Breeding, *4*, 147-154.

WHITHAM S., S.P. DINESH-KUMAR, D. CHOI, R. HEHL, C. CORR, and B. BAKER (1994): The product of the tobacco mosaic virus gene *N*: Similarity to toll and the interleukin-1 receptor. Cell *78*, 1101-1115.

WILSON T.M.A. (1993): Strategies to protect crop plants against viruses: Pathogen-derived resistance blossoms. Proc. Natl. Acad. Sci. USA, *90*: 3134-3141.

XIA X.C., A.E. MELCHINGER, L. KUNTZE, and T. LÜBBERSTEDT (1999): Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology *8*, 660-667.

XU M.L., A.E. MELCHINGER, X.C. XIA, and T. LÜBBERSTEDT (1999): Fine mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR and AFLP markers. Mol. Gen. Genet. *261* (3), 574-581.

XU M.L., A.E. MELCHINGER, and T. LÜBBERSTEDT (2000): Origin of *Scm1* and *Scm2* - two loci conferring resistance to sugarcane mosaic virus (SCMV) in maize. Theor. Appl. Genet., *100*, 934-941.

YU Y.G., G.R. BUSS, and M.A. SAGHI MAROOF (1996): Isolation of a superfamily of candidate disease-resistance genes in soybean on a conserved nucleotide-binding site. Proc. Natl. Acad. Sci. USA, *93*: 11751-11756.

ZAITLIN D., S. DEMARS, and Y. MA (1993): Linkage of *rhm*, a recessive gene for resistance to southern corn leaf blight, to RFLP marker loci in maize (*Zea mays*) seedlings. Genome *36*: 555-564.

Theor Appl Genet (2002) 105:355–363
DOI 10.1007/s00122-002-0953-x

M. Quint · R. Mihaljevic · C. M. Dussle · M. L. Xu
A. E. Melchinger · T. Lübberstedt

Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize

Received: 23 September 2001 / Accepted: 12 December 2001 / Published online: 19 June 2002

Abstract Three previously published resistance gene analogues (RGAs), pic13, pic21 and pic19, were mapped in relation to sugarcane mosaic virus (SCMV) resistance genes (*Scmv1*, *Scmv2*) in maize. We cloned these RGAs from six inbreds including three SCMV-resistant lines (D21, D32, FAP1360A) and three SCMV-susceptible lines (D145, D408, F7). Pairwise sequence alignments among the six inbreds revealed a frequency of one single nucleotide polymorphism (SNP) per 33 bp for the three RGAs, indicating a high degree of polymorphism and a high probability of success in converting RGAs into co-dominant cleaved amplified polymorphic sequence (CAPS) markers compared to other sequences. SNPs were used to develop CAPS markers for mapping of the three RGAs in relation to *Scmv1* (chromosome 6) and *Scmv2* (chromosome 3), and for pedigree analyses of resistant inbred lines. By genetic mapping pic21 was shown to be different from *Scmv2*, whereas pic19 and pic13 are still candidates for *Scmv1* and *Scmv2*, respectively, due to genetic mapping and consistent restriction patterns of ancestral lines.

Keywords RGA · SNP · CAPS · SCMV · Maize

Communicated by C. Möllers

M. Quint · R. Mihaljevic · C.M. Dussle · M.L. Xu
A.E. Melchinger · T. Lübberstedt (✉)
Institute of Plant Breeding, Seed Science,
and Population Genetics, University of Hohenheim,
Fruwirthstraße 21, 70593 Stuttgart, Germany
e-mail: Thomas.Luebberstedt@agrsci.dk
Fax: +45-58-11-33-01

Present addresses:
M.L. Xu, College of Bioscience and Biotechnology,
Yangzhou University, Yangzhou, Jiangsu, 225009, PR China

T. Lübberstedt, Danish Institute of Agricultural Sciences,
4200 Flakkebjerg, Denmark

Introduction

Sugarcane mosaic virus (SCMV) is one of the most important virus diseases of maize and causes serious yield losses in susceptible cultivars (Fuchs and Grüntzig 1995). Owing to the non-persistent transmission, control of aphid vectors by chemical means is not effective and, therefore, cultivation of resistant maize varieties is the most efficient method of virus control. In a study with 122 early maturing European maize inbreds, three lines (FAP1360A, D21 and D32) displayed complete resistance and four lines displayed partial resistance (FAP1396A, D06, D09 and R2306) against SCMV and maize dwarf mosaic virus (MDMV) (Kuntze et al. 1997). In field trials, resistance of all three European lines D21, D32 and FAP1360A seemed to be controlled by one to three genes (Melchinger et al. 1998). Linkage mapping and 'bulked segregant analysis' (BSA) mapped two major genes, *Scmv1* on the short arm of chromosome 6 and *Scmv2* near the centromere of chromosome 3 (Melchinger et al. 1998; Xia et al. 1999; Xu et al. 1999; Dußle et al. 2000). Minor quantitative trait loci (QTLs) affecting SCMV resistance were identified on chromosomes 1, 5, and 10 (Xia et al. 1999). For full resistance to SCMV, the presence of both *Scmv1* and *Scmv2* is essential. *Scmv1* suppresses symptom expression throughout all developmental growth stages at a high level, whereas *Scmv2* was expressed at later stages of infection (Xia et al. 1999; Dußle et al. 2000).

With the cloning of a number of disease resistance genes (R genes) from several plant species, it became obvious that these R genes share homologies in protein domains such as the nucleotide-binding site (NBS) and leucine-rich repeats (LRRs) (reviewed in Bent 1996; reviewed in Hammond-Kosack and Jones 1996). Degenerate primers based on the amino-acid sequence of these domains have meanwhile allowed successful PCR amplification of several RGAs from various plant species with significant homology to known plant disease R genes. Collins et al. (1998) identified 20 RGA loci in maize, which mapped preferentially to chromosomal regions

known to carry R genes (McMullen and Simcox 1995). These RGAs can be further analysed for their potential use in marker-assisted selection (MAS) or even the cloning of target genes. The latter approach has been successfully used in maize to clone the *Rp1-D* gene, a member of the complex locus composed of approximately nine gene homologues, conferring resistance to common rust, *Puccinia sorghi* (Collins et al. 1999).

Most sequence variation is attributable to single nucleotide polymorphisms (SNPs), with the rest attributable to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements (The International SNP Map Working Group 2001). In maize, one SNP between two randomly sampled sequences occurs approximately every 104 bp (Tenaillon et al. 2001). Therefore, SNPs are present at sufficient density for comprehensive haplotype analysis as applied in this study.

The objectives of the present study were to: (1) clone three RGAs previously mapped to chromosomal regions known to harbour SCMV R genes from six inbred lines resistant or susceptible to SCMV, (2) measure the frequency of SNP and CAPS occurring within RGAs and evaluate their usefulness as a source for marker development, and (3) map RGA-CAPS markers in relation to *Scmv1* and *Scmv2* to examine their role as candidates for SCMV R genes.

Materials and methods

Plant materials

Seventeen maize inbred lines were analysed for CAPS restriction pattern including: (1) three resistant European dent lines D21, D32, and FAP1360A, (2) three highly susceptible European lines D408 (dent), D145 (flint), and F7 (flint), (3) ancestral lines of the three resistant lines, A632, V3, WD, W401, Co158, Co125, FAP493B, and FAP954A, and (4) three partially resistant lines D06, D09, and FAP1396A.

The mapping populations consisted of: (1) 87 $F_{2:3}$ families derived from a cross between D32 and D145 (PopA) (Vuylsteke et al. 1999; Xia et al. 1999), (2) 27 resistant BC5 individuals from the cross (F7 × FAP1360A) × F7 (PopB) (Xu et al. 1999), and (3) 30 resistant BC7 individuals from the cross (D408 × D21) × D408 (PopC). PopB originally consisted of 40 individuals, but 13 were excluded after progeny testing in field trials for resistance to SCMV. In addition, the map of PopB was saturated with SSR markers *phi075*, *umc1002*, *umc1018*, *bmc1600*, *bmc1433*, *bngl107*, *bmc1538*, and *bngl426*. Primer sequences were obtained from the maize database (http://www.agron.missouri.edu/ssr.html). PCR amplification was performed as described by Xu et al. (1999). pic13 was also mapped in oat-maize chromosome addition lines (Ananiev et al. 1997).

Cloning and sequencing of RGAs

Genomic DNA was extracted from 0.1 g of freeze-dried leaf tissue following the CTAB procedure as described by Hoisington et al. (1994). RGAs were isolated and cloned from the parental lines using PCR with specific primers (Table 1) based on the original sequences of pic13 (pic13L2/pic13R2), pic21 (pic21L1/pic21R1), and pic19 (pic19L/pic19R) described by Collins et al. (1998). DNA amplifications were performed in a standard reaction mix containing 100 ng of genomic DNA, 10 mM of Tris–HCl (pH 9), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.3 mM of dNTPs, 0.5 pMol of each primer and 0.625 U of *Taq* DNA polymerase (Amersham Pharmacia, Freiburg). After an initial denaturation step at 94 °C for 2 min, template DNAs were amplified using 35 cycles with the following conditions: 1 min at 94 °C, 2 min at 55 °C, and 2 min at 72 °C. The final extension step was conducted at 72 °C for 2 min. Amplification products were excised from the agarose gel, extracted with the Nucleospin Kit (Macherey and Nagel, Düren), and blunt-end cloned into the pBluescript vector.

Table 1 Primer sequences used for specific amplifications of pic13, pic19 and pic21

RGA	Primer	Sequences 5′–3′
pic13	pic13L2	TTGAAGCCATTGCTGGTGAC
	pic13R2	GCCATGAGCTATCCATTGAG
pic19	pic19L	TAGATGATGTCTGGACGGCT
	pic19R	AGCCAATGGCAAACCATCAC
	pic19X	GCAGTTCCTCTCTGCAACGTG
	ippic19L	CCAGAGTTACATCAGTGTGG
	ippic19R	ACATCAGCCGTCCAGACATC
pic21	pic21L1	GGAAGACCACGCTGCTCAAC
	pic21R1	CTCATCAGGTGGTCGCCAAC

Elongation of pic19 was performed by inverse PCR. Genomic DNA (500 ng) of the resistant parent FAP1360A was digested with *Rsa*I for 1 h at 37 °C, self-ligated overnight at 16 °C, and amplified with inverse PCR primers ippic19L and ippic19R. The elongated pic19 can be recreated after amplification by using primers pic19L and pic19X (Table 1). DNA sequencing was performed using the ALFExpress automated DNA sequencer (Amersham Pharmacia, Freiburg). Sequencing reaction conditions were chosen as suggested by the manufacturer (Amersham Pharmacia, Freiburg). For verification, pic19 was again sequenced by SEQLAB Sequence Laboratories Göttingen GmbH. The DNA sequences and the deduced amino-acid sequences were analysed using the ALIGN Plus 2.0 software package (http://www.scied.com/ses_alim.htm).

Generation and mapping of CAPS markers

Sequence regions displaying single nucleotide changes within restriction enzyme recognition sites between parental lines of mapping populations were used to identify RGA-CAPS using the CLONE Manager 3.11 software package (http://www.scied.com/ses_cm6.htm). PCR products were generated using the primers (Table 1) and the amplification conditions listed above. PCR products were digested with restriction endonucleases *Nla*III (pic19), *Rsa*I (pic13), and *Mbo*II (pic21) at 37 °C for 4 h according to the manufacturers' suggestions. RGAs were separated on 3% MetaPhor agarose gels in 0.5 × TBE buffer except for pic21, which was separated on denaturing polyacrylamide gels (SequaGel6, Biozym, Hessisch Oldendorf). Mapping of pic13, pic21, and pic19 was performed with PopA. In addition, pic13 was mapped in oat-maize addition lines and pic19 in PopB (Xu et al. 1999).

Linkage and statistical analyses

RGAs were mapped using data generated in previous studies (Xia et al. 1999; Xu et al. 1999). Marker orders and genetic distances for PopA were calculated with MAPMAKER 3.0b (Lander et al. 1987) using a LOD threshold of 3.0 and the Kosambi mapping function. Marker orders and genetic distances for PopB were calculated with Cri-Map (http://biobase.dk/Embnetut/Crimap).

Significant differences among the six inbred lines (resistant: D21, D32, FAP1360A; susceptible: D408, D145, F7) for SNP and CAPS frequencies per 100 bp of RGA sequence were tested by ANOVA using the software package PLABSTAT (http://www.uni-

Table 2 Genetic similarities (GS) of RGAs and frequency of SNPs per 100 bp between six maize inbred lines

Inbred lines	F7	FAP1360A	D21	D408	D32	D145	
pic19							
F7		6.38	5.10	6.54	5.58	6.34	GS
FAP1360A	0.34		4.78	6.22	5.90	2.55	
D21	0.50	0.53		2.55	1.91	5.10	
D408	0.34	0.34	0.74		1.59	5.58	
D32	0.45	0.41	0.84	0.86		4.15	
D145	0.38	0.81	0.55	0.47	0.57		
	SNP/100 bp					$\Sigma^a = 58$	
pic13							
F7		0.92	4.94	1.23	3.94	0.56	GS
FAP1360A	0.79		4.80	1.83	4.23	0.89	
D21	0.50	0.43		4.62	3.34	3.30	
D408	0.93	0.86	0.57		3.90	0.84	
D32	0.29	0.36	0.50	0.36		3.62	
D145	0.79	0.86	0.43	0.86	0.36		
	SNP/100 bp					$\Sigma^a = 14$	
pic21							
F7		1.17	0.70	0.70	0.70	3.29	GS
FAP1360A	1.00		0.47	0.94	0.70	3.99	
D21	1.00	1.00		0.47	0.23	3.29	
D408	0.93	0.93	0.93		0.47	3.52	
D32	0.93	0.93	0.93	0.87		3.52	
D145	0.13	0.13	0.13	0.07	0.07		
	SNP/100 bp					$\Sigma^a = 15$	

[a] Total number of SNPs between all six investigated inbred lines

hohenheim.de/~ipspwww/soft.html). Analyses of variance for randomised complete block designs were used to obtain a mean value for each RGA sequence. Least significant differences were calculated to test for differences between individual entry means. To determine SNP frequencies, pairwise comparisons between each two of the six inbred lines were performed. CAPS frequencies were calculated in pairwise comparisons of the three parental pairs of the mapping populations PopA, PopB, and PopC.

Genetic similarity of the RGA sequences between the six investigated inbred lines was measured on the basis of SNPs (Table 2) to test the six inbred lines for clustering into the heterotic flint and dent groups or resistant and susceptible lines using the software package NTSYSpc Version 2.02i (Rohlf 1989). The data set consisted of 58 detected SNPs between all six inbred lines for pic19, 14 for pic13, and 15 for pic21. Graphical genotypes of pic19 were displayed by software package GGT:GraphicalGenoTyping (http://www.spg.wau.nl/pv/pub/ggt).

Results

Isolation of pic13, pic19, and pic21 sequences from six inbred lines

All six RGA alleles of pic19 and pic21 were identical in size and similar to the sequences published by Collins et al. (1998). All pic19 sequences contained continuous open reading frames (ORFs), except those for D145 and D408. Likewise, pic21 sequences of all six inbreds showed continuous ORFs, except F7. The sequence of pic13 was incomplete, i.e. it was sequenced from both ends without identifying overlaps. pic13 showed a continuous ORF for both partial sequences of FAP1360A, D32, and F7. Lines D21, D408, and D145 displayed at least one stop codon.

Frequency of SNPs and probability of conversion into RGA-CAPS

The overall mean for the six genotypes over the three RGAs was 3.05 SNPs per 100 bp (= 1 SNP per 33 bp) of the RGA sequence. SNP frequencies of pic19 (4.70 SNPs per 100 bp) were significantly higher ($P < 0.01$) than those of pic13 (2.86 SNPs per 100 bp) and pic21 (1.61 SNPs per 100 bp). The number of CAPS sites per 100 bp of the RGA sequence was measured for the three parental pairs of the mapping populations FAP1360A/F7, D32/D145, and D21/D408. The number of CAPS varied between 0.00 and 4.15 per 100 bp, with an overall mean of 2.20 CAPS sites per 100 bp. CAPS frequencies of pic19 (3.72) were significantly higher ($P < 0.01$) than those of pic13 (1.69) and pic21 (1.18). PopA revealed the highest degree of polymorphism (3.13), followed by PopC (2.18), and PopB (1.82).

Genetic similarity analyses revealed no clustering into the heterotic flint and dent groups or resistant and susceptible lines (Fig. 1). The graphical haplotype for pic 19 was constructed on the basis of the same SNP data for all six inbred lines, revealing conserved sequence blocks.

Development of CAPS markers

CAPS polymorphisms were identified for all investigated RGAs in at least one of the three mapping populations. Sequence alignment between the cloned pic13 fragments revealed a SNP within a *Rsa*I recognition site resulting in polymorphism between D32 and D145

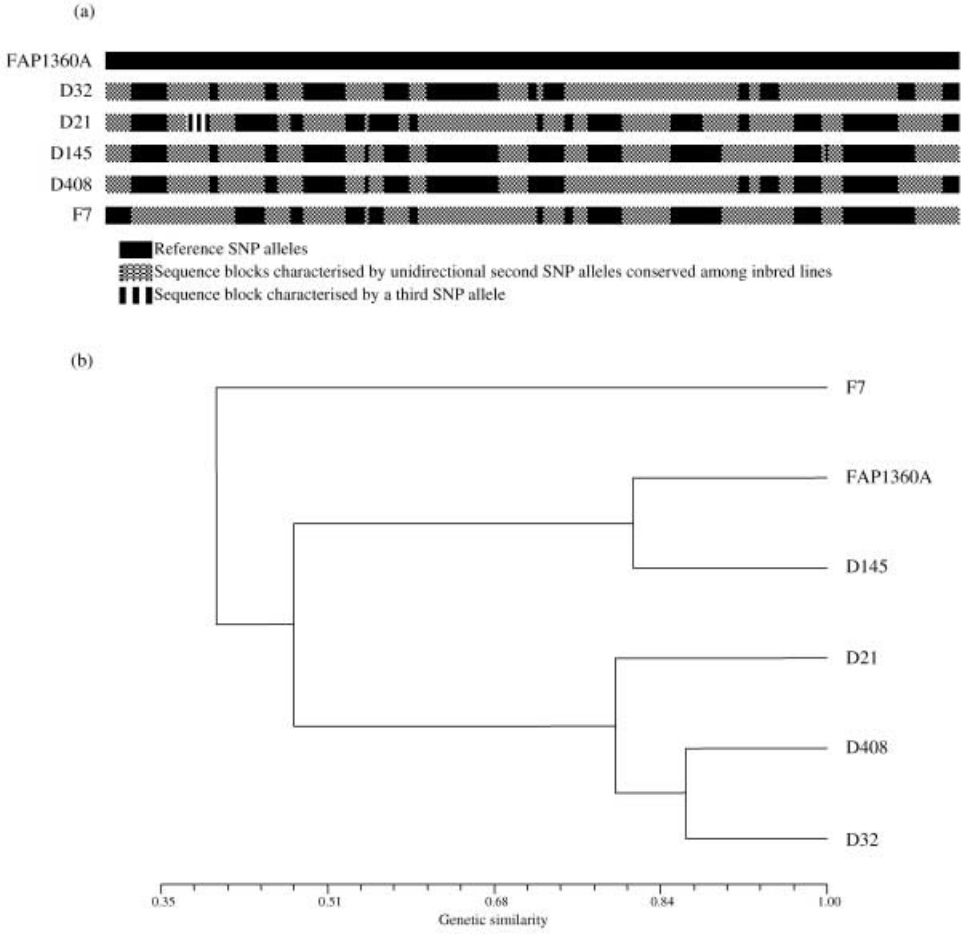

Fig. 1 (**a**) Graphical haplotyping for the example of pic19 (624 bp) in six European inbred lines based on 58 SNPs between the six maize inbred lines FAP1360A, D32, D21, D145, D408 and F7. (**b**) Genetic similarities among the six inbred lines based on pic19 SNPs

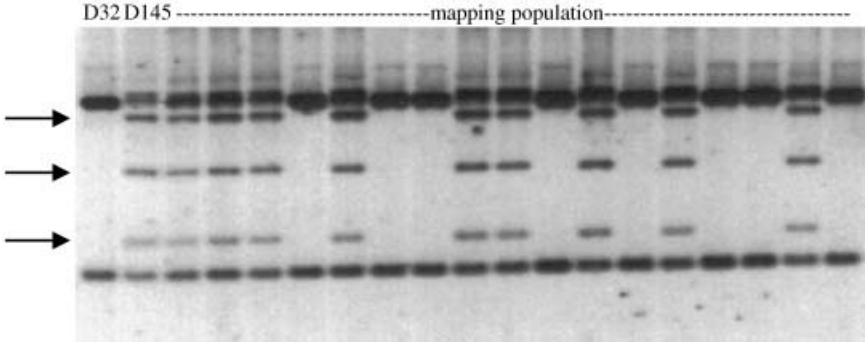

Fig. 2 *Rsa*I restriction pattern of pic13 PCR products from parental inbred lines D32 (SCMV resistant) and D145 (SCMV susceptible) and part of PopA

(PopA) (Fig. 2) and between D21 and D408 (PopC) (data not shown). The sequences of pic21 were polymorphic between D32 and D145 at a *Mbo*II recognition site. Digestion of elongated pic19 amplification products with *Nla*III generated polymorphisms in all three parental pairs (data not shown).

Genetic mapping of pic13, pic21, and pic19 relative to *Scmv1* and *Scmv2*

Amplification of pic13 from oat-maize addition lines showed a PCR product of the expected size only in the oat line carrying an additional maize chromosome 3. Amplification of pic19 and pic21 did not result in differential amplification of the oat line carrying the respective additional maize chromosome. For PopA, pic13 was mapped into the interval of *umc102/csu285b* on chromosome 3L (Fig. 3), which completely overlaps with the *Scmv2*-QTL region previously identified by Xia et al. (1999). However, pic13 could not be mapped in PopB because of identical banding patterns between heterozygotes and the recurrent parent F7 for all tested restriction enzymes.

Using PopA, pic21 was located on chromosome 3L but, contrary to pic13, it was located outside the *Scmv2*

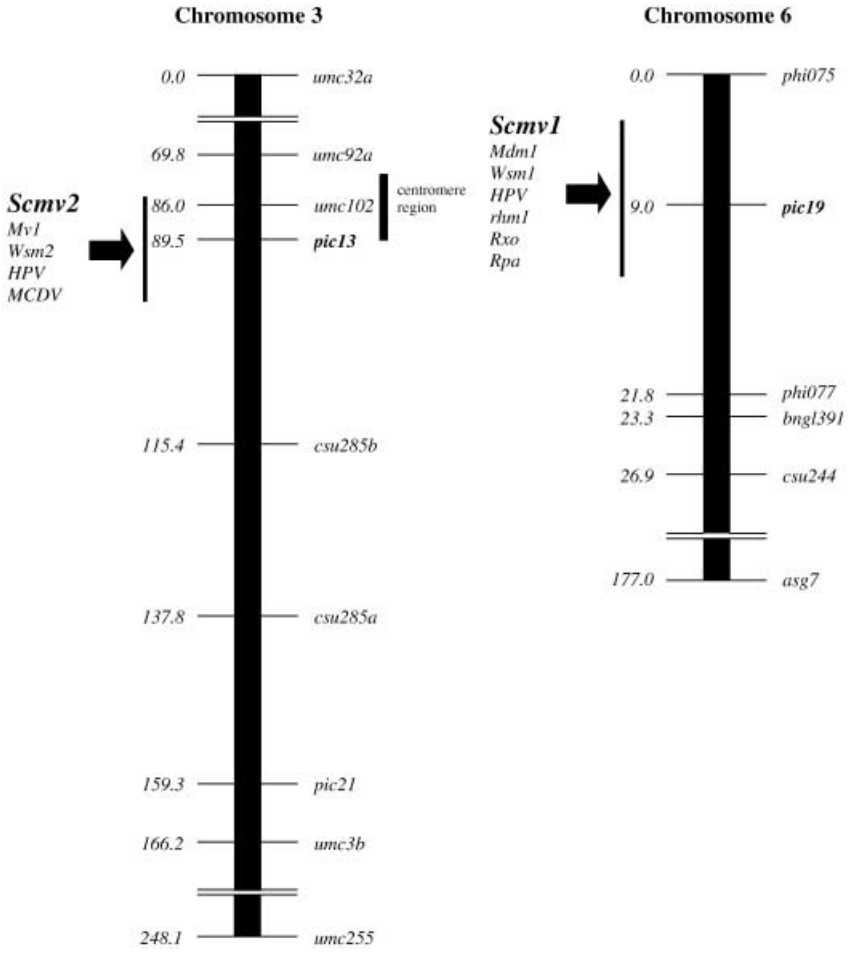

Fig. 3 Genetic map of genome regions conferring resistance to SCMV on maize chromosomes 3 and 6 derived from D32 × D145 $F_{2:3}$ mapping populations (PopA). Chromosome 3 containing RGAs pic13 and pic21, and genes for resistance to sugarcane mosaic virus (*Scmv2*), maize mosaic virus (*Mv1*), wheat streak mosaic virus (*Wsm2*), high plains virus (*HPV*) and maize chlorotic dwarf virus (*MCDV*). Chromosome 6 containing pic19 and genes for resistance to sugarcane mosaic virsu (*Scmv1*), wheat streak mosaic virus (*Wsm1*), high plains virus (HPV), southern corn leaf blight (*rhm1*), rice bacterial streak (*Rxo*) and sorghum bacterial stripe (*Rpa*)

Table 3 Evaluation of restriction pattern of pic19 and pic13 in resistant, susceptible, partially resistant, and ancestral inbred lines

Inbred lines		pic13		pic19		
		Genotype A	Genotype B	Genotype A	Genotype B	Genotype C
Resistant	FAP1360A		X	X		
	D21	X		X		
	D32	X		X		
Susceptible	F7		X		X	
	D408	X				X
	D145	X			X	
Partially resistant	D06	X		X		
	D09	X		X		
	FAP1396A	X		X		
Ancestors	A632	X		X		
	V3	X			X	
	WD	X			X	
	W401	X				X
	Co158	X			X	
	Co125		X		X	
	FAP493B	X			X	
	FAP954A	X		X		

region flanked by markers *csu285a* and *umc3b* (Fig. 3). In PopA, pic19 mapped to maize chromosome 6S between markers *phi075* and *phi077b* (Fig. 3). In PopB, pic19 mapped into the marker interval *umc1023/E2M7-1* with a map distance of 3.5 cM to *Scmv1*.

RGA analyses in ancestral lines and partially resistant inbreds

Two genotypes were observed when pic13 amplification products were digested with *Rsa*I. F7, FAP1360A, and Co125, the direct ancestor and potential donor of *Scmv2*,

showed the same restriction banding pattern (genotype B). The remaining inbred lines displayed a different restriction pattern (genotype A) (Table 3).

Three genotypes were observed for pic19 when digested with *Nla*III. Besides the three resistant lines FAP1360A, D21, and D32, the potential donors of *Scmv1*, A632 and FAP954A, and the three partial resistant inbreds D06, D09, and FAP1396A, as well as FAP1360A, showed the same restriction pattern (genotype A). The second restriction banding pattern (genotype B) was observed for the following six lines: V3, WD, Co158, Co125, FAP493B, and F7. The remaining two lines D408 and W401 lack one *Nla*III restriction site and thus revealed a third restriction pattern (genotype C) (Table 3). In contrast to pic13, the resistant lines revealed the same restriction pattern.

Discussion

One of the long-term aims of molecular marker technology in plant breeding is the selection of superior individuals directly at the level of DNA. RGAs provide an excellent source for the development of molecular markers, especially for resistance traits, because of their high level of polymorphism and their putatively functional character. Thereby, RGAs can be converted to single-copy PCR markers like CAPS (Konieczny and Ausubel 1993), while RGA-RFLPs frequently result in complex banding patterns because of sequence homology to related RGA sequences (Collins et al. 1998). Diagnostic markers like allele-specific RGA-CAPS will have wider applications in MAS strategies in the future.

SNP and CAPS frequency in three maize RGAs

Rafalski et al. (2001) analysed random cDNA clones in a collection of over 30 public and private maize lines representative for the North American corn germplasm. This analysis was restricted to coding regions. Sequence alignment revealed one single nucleotide change per 70 bp among the 30 lines. However, the authors pointed out that pairwise comparison between any two lines – such as in the results presented in this article – reveals a lower degree of polymorphism. Useche et al. (2001) performed similar experiments in maize and reported a SNP frequency of only one SNP per 138 bp in coding regions. Another study revealed that maize has an average of one SNP every 104 bp between two randomly sampled sequences (Tenaillon et al. 2001). Hence, the frequency of one SNP per 33 bp originating from single nucleotide changes in pairwise comparisons is significantly higher in the present study. This result confirms the highly polymorphic character of plant RGAs and possibly R genes, which is an important feature in evolutionary processes at R gene loci and especially R gene clusters (Parniske and Jones 1999).

Among the RGAs, pic19 displayed a significantly higher degree of polymorphism than pic13 and pic21 (Table 2); pic19 might be phylogenetically older than pic13 and pic21 and, therefore, displays a significantly higher degree of polymorphism. A second possibility would be the coding character of pic13 and pic21 in contrast to pic19. Since a SNP in the first two positions within a triplet generates more amino-acid changes, the mutations in coding regions should be mainly in the third position of the triplet. Hence, the significantly higher SNP frequency at the third position of triplets for pic19 suggests expression of the respective sequence, but was not observed in the current study.

The generally high degree of polymorphism of the three RGAs might also be influenced by their map position. Two of the three RGAs map to potential R gene clusters on chromosomes 3L close to the centromere and 6S close to the nucleolus organiser region (nor). In gene clusters the recombination frequency is expected to be very high because of the high density of coding sequences. In many plant species, recombination rates can vary up to an order of magnitude over relatively small intervals (reviewed by Schnable et al. 1998). Detailed physical and genetic mapping of grass genomes revealed the clustering of genes, and supports the hypothesis that much of the meiotic recombination occurs in genes and most recombination events are restricted to few chromosome regions containing gene clusters (Gill et al. 1996; Künzel et al. 2000). The maize genome exhibits a very striking gene distribution with almost all genes present in 10–20% of the genome (Carels et al. 1995). In fact, genes exhibit recombination rates 10 to 100-fold higher than the genome average (reviewed by Lichten and Goldman 1995). In maize, one-fifth of all recombination events in a 140-kb interval between the *anthocyaninless1* (*a1*) and the *shrunken2* (*sh2*) genes were resolved within a 377-bp region of the *a1* gene (Xu et al. 1995). It is not known whether the *Scmv* regions on chromosomes 6 and 3 contain R gene clusters or single pleiotropically acting genes. Zhao et al. (2001) mapped two maize R genes, *Rxo* and *Rpa*, to the same position on the short arm of maize chromosome 6. The authors identified 5–6 RGAs in the direct neighbourhood of *Rxo* and *Rpa*. These findings and the high degree of polymorphism for pic19 support the presence of a R gene cluster on chromosome 6S. However, one member of the putative R gene cluster on chromosome 6S, *Mdm1*, shows 100% linkage with nor (Simcox et al. 1995), whereas *Scmv1* and nor were not absolutely linked (Xu et al. 1999). Very low recombination rates are typically exhibited by regions surrounding the centromeres and the nor. Therefore, the putative R gene cluster on chromosome 6S is located in the direct neighbourhood of regions with suppressed recombination. Interestingly, the situation seems to be similar on chromosome 3 with *Scmv2* in the direct neighbourhood of the centromere.

The structure of plant materials (flint-dent, SCMV resistant-susceptible; relationship by descent of the three resistant lines) was not reflected by the pic19-based genetic similarity analysis (Fig. 1). Explanations for the lack of grouping of the six inbred lines are either a miss-

ing correlation of the RGA allele sequences with these factors or the above mentioned high degree of polymorphism within RGAs. Furthermore, Xu et al. (2000) identified a 7.2 cM interval containing eight molecular markers as well as pic19 flanking *Scmv1* conserved across the three resistant lines D21, D32, and FAP1360A, whereas the susceptible lines displayed a different segregation pattern. Therefore, this 7.2 cM chromosomal segment seemed to be identical in the three resistant lines. However, disagreement of RGA-based genetic similarity analysis with phenotypic, pedigree, and marker data question the utility of SNPs for association or disequilibrium mapping (reviewed in Lazzeroni 2001), at least for this genome region.

Single nucleotide changes in all three RGAs between all six inbred lines generated only two of the possible four SNP alleles except one. The formation of the second SNP alleles was not randomly distributed. Several consecutive SNPs were ordered in sequence blocks (haplotypes) with unidirectional mutations. Most of the sequence blocks contain more than one single SNP (Fig. 1a). Presuming a low number of RGA alleles in European founder materials, the driving force for the development of the high degree of polymorphism and variability within these RGAs seems to be recombination.

In conclusion, RGA-derived markers are especially interesting because of their genomic localisation within R gene clusters which putatively contain numerous genes relevant for plant breeding purposes. The frequencies of SNPs and CAPS are promising prerequisites for conversion of even short RGA sequences into molecular markers. This should further enhance the mapping of RGAs as potential candidates for genes conditioning resistance to pathogens in plant species.

Candidate gene evaluation

The identification of CAPS restriction sites between the parents of three mapping populations facilitated mapping of the three RGAs, pic13, pic19, and pic21, to genomic regions known to be involved in the inheritance of SCMV resistance on chromosomes 3 and 6. To address the question of whether any of the RGAs is a likely candidate for one of the target genes *Scmv1* or *Scmv2*, respectively, we evaluated the following criteria: (1) *Linkage with the target gene*. A potential role as a candidate gene can be ruled out in case of a large map distance to the target gene. Cosegregation and even tight linkage to the R gene indicates a functional role of the RGA because of: (i) incomplete penetrance of SCMV resistance and escapes resulting in mis-scoring, as well as (ii) the potential presence of more than one SCMV R gene within both regions. (2) *Comparison of the R alleles of the six inbred lines with those of the ancestor and partially resistant lines*. Sharing the same allele among resistant, partially resistant inbred lines and putative R gene donors on the one hand and different alleles in the other lines on the other hand, indicates a functional role in the inheritance of SCMV resistance. (3) *Continuous ORFs as a prerequisite for coding regions*. A continuous ORF spanning the complete RGA sequence strongly suggests a coding character on the basis of the statistical probability of the occurrence of stop codons.

Chromosome 6

Mapping in PopA located pic19 within the QTL peak of *Scmv1*. Mapping in PopB showed two recombinant individuals. However, field experiments, BSA (Xu et al. 1999), and QTL analyses (Xia et al. 1999; Dußle et al. 2000) did not exclude the presence of more than one SCMV R gene in the *Scmv1* region. Furthermore, incomplete penetrance of virus resistance and escapes might result in mis-scorings during the phenotypic evaluation of the mapping populations. Therefore, a functional role of pic19 in the expression of SCMV resistance cannot be ruled out by genetic mapping. Moreover, the three resistant, the partially resistant lines, as well as the putative donors of *Scmv1*, A632, and FAP954A, displayed the same restriction patterns for pic19 (Table 3). Xu et al. (2000) investigated the chromosomal regions harbouring *Scmv1* and *Scmv2* by AFLP analyses in the same three resistant European inbreds, their ancestral lines and partially resistant inbred lines. The banding patterns indicated the identity of a 7.2 cM window harbouring the *Scmv1* locus in D21, D32, and FAP1360A. Therefore, it is most likely that these three European inbred lines share the same *Scmv1* allele at this Mega-locus. Absence of stop codons in all three resistant lines strengthens the possible functional involvement in the expression of SCMV resistance. However, a continuous ORF is not restricted to the resistant lines because the susceptible inbred F7 also showed a continuous ORF. In conclusion, pic19 remains a candidate for *Scmv1*, but our findings cannot distinguish between close linkage and identity of pic19 with *Scmv1*.

Chromosome 3

In PopA, pic21 mapped to chromosome 3, and was located approximately 50 cM outside the marker interval flanking *Scmv2*. Therefore, it was excluded as a candidate gene for *Scmv2*.

In contrast, pic13 mapped in PopA within the QTL peak of *Scmv2*. According to Xu et al. (2000), the *Scmv2* region in FAP1360A most likely originated from Co125, while all other ancestral lines showed AFLP patterns in this region different from FAP1360A. The authors proposed the following explanations for the origin of the *Scmv2* region: (1) *Scmv2* in FAP1360A originates from Co125, or (2) Co125 is also the *Scmv2* donor of D32 displaying a very short donor region not detectable by the employed flanking markers or another inbred line. The fact that restriction patterns of the potential donor of *Scmv2*, Co125, and FAP1360A, are identical and show a

different allele at this locus than D21 and D32, confirms the previously published hypothesis of different *Scmv2* genes in FAP1360A and D32 (Xu et al. 2000). The presence of *Scmv1* seems to be sufficient for conferring partial resistance. Hence, *Scmv2* is not necessarily expected to be present in the partially resistant lines, explaining a different genotype in all three partially resistant inbred lines. No continuous ORF was found for pic13 in D21, in agreement with the lack of evidence for *Scmv2* in this inbred line.

The mapping results and analyses of ancestor lines cannot distinguish between identity and close linkage of pic19 and *Scmv1*, and pic13 and *Scmv2*, respectively. The previously mentioned results of Zhao et al. (2001), as well as studies on the *Pto*, *Cf*, and *Dm* clusters in tomato and lettuce (Martin et al. 1994; Thomas et al. 1997; Meyers et al. 1998), revealed the existence of numerous RGAs in the direct physical neighbourhood of the R genes. At least seven apple *Vf* gene analogues have been identified from a BAC contig of 290 kb, encompassing the *Vf* locus (Xu 2001, personal communication). These RGAs showed very similar amino-acid domains except for the LRRs, where deletion of several LRR units and point mutations occurred frequently. Therefore, even in the case of non-identity of pic19 and pic13 with *Scmv1* and *Scmv2*, respectively, they should provide excellent starting points for a map-based approach for cloning of the target genes themselves as well as other members of these clusters, such as MDMV, wheat streak mosaic virus (WSMV), maize mosaic virus (MMV), High Plains virus (HPV), and maize chlorotic dwarf virus (MCDV) R genes. Large and often continuous ORFs of the RGAs investigated in this study indicate that they are part of coding sequences. This is in agreement with large-scale sequencing of the above mentioned tomato and lettuce R gene loci. The majority of RGAs seemed to be functional and expressed, with only a few of them clearly being pseudogenes. Screening of DNA libraries will extend the number of RGAs in these regions and, therefore, broaden the probability of identifying the target genes. In contrast to the simple procedures to clone homologues, establishing potential functions of the RGAs remains challenging. Since the proof of function of a given candidate for a certain resistance is highly dependent on pathogen-specific plant material, exchange of R gene candidates between research groups working on different resistances is essential and can greatly accelerate the cloning of R genes by the RGA approach.

Acknowledgements The present study was supported by a grant from the Deutsche Forschungsgemeinschaft, Grant No. LU601/3. The authors gratefully acknowledge the mapping of SSRs to PopB by Susanne Dreissigacker and the skilled technical assistance of Elisabeth Kokai-Kota.

References

Ananiev EV, Riera-Lizarazu O, Rines HW, Phillips RL (1997) Oat-maize chromosome addition lines: a new system for mapping the maize genome. Proc Natl Acad Sci USA 94:3524–3529

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell 8:1757–1771

Carels N, Barakat A, Bernardi G (1995) The gene distribution of the maize genome. Proc Natl Acad Sci USA 92:11,057–11,060

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogues in maize. Mol Plant–Microbe Interact 11:968–78

Collins NC, Drake J, Aycliffe M, Sun Q, Ellis J (1999) Molecular characterisation of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

Dußle CM, Melchinger AE, Kuntze L, Stork A, Lübberstedt T (2000) Molecular mapping and gene action of *Scm1* and *Scm2*, two major QTLs contributing to SCMV resistance in maize. Plant Breed 119:299–303

Fuchs E, Grüntzig M (1995) Influence of sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV) on the growth and yield of two maize varieties. J Plant Dis Protection 102:44–50

Gill KS, Gill BS, Endo TR, Taylor T (1996) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 144:1883–1891

Hammond-Kosack KE, Jones JDG (1996) Resistance gene-dependent plant defense responses. Plant Cell 8:1773–1791

Hoisington DA, Khairallah MM, Gonzales-de-Leon D (1994) Laboratory protocols. CIMMYT Applied Molecular Genetics Laboratory, Mexico, DF

Konieczny A, Ausubel FA (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Kuntze L, Fuchs E, Grüntzig M, Schulz B, Klein D, Melchinger AE (1997) Resistance of early maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breed 116:499–501

Künzel G, Korzun L, Meister A (2000) Cytogenetically integrated physical restriction fragment length polymorphism maps for barley genome based on translocation breakpoints. Genetics 154:397–412

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

Lazzeroni (2001) A chronology of fine-scale gene mapping by linkage disequilibrium. Stat Methods Med Res 10:57–76

Lichten M, Goldman AS (1995) Meiotic recombination hotspots. Annu Rev Genet 29:423–444

Martin GB, Frary A, Wu T, Brommonschenkel S, Chunwongse J, Earle ED, Tanksley SD (1994) A member of the tomato *Pto* gene family confers sensitivity to fenthion resulting in rapid cell death. Plant Cell 6:1543–1552

McMullen MD, Simcox KD (1995) Genomic organisation of disease and insect resistance genes in maize. Mol Plant–Microbe Interact 8:811–815

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

Meyers BC, Chin, DB, Shen, KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1831

Parniske M, Jones JD (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc Natl Acad Sci USA 96:5850–5

Rafalski JA, Ching A, Bhattramakki D, Morgante M, Dolan M, Register JC, Smith OS, Tingey S (2001) SNP markers in maize: discovery and applications. Plant and Animal Genome IXth Conference, p 44

Rohlf FJ (1989) NTSYS-pc numerical taxonomy and multivariate analysis system. Setauket, Exeter Publishing Company Limited

Schnable PS, Hsia AP, Nikolau BJ (1998) Genetic recombination in plants. Curr Opin Plant Biol 1:123–129

Simcox KD, McMullen MD, Louie R (1995) Co-segregation of the maize dwarf mosaic virus resistance gene, *Mdm1*, with the nucleolus organiser in maize. Theor Appl Genet 90:341–346

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JDG (1997) Characterisation of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9:2209–2224

Useche F, Morgante M, Hanafey M, Tingey S, Gao GR, Rafalski JA (2001) Computer detection of single nucleotide polymorphisms (SNPs) in maize ESTs. Plant and Animal Genome IXth Conference, p 144

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Xia XC, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology 89:660–667

Xu ML, Melchinger AE, Xia XC, Lübberstedt T (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. Mol Gen Genet 261:574–581

Xu ML, Melchinger AE, Lübberstedt T (2000) Origin of *Scm1* and *Scm2* – two loci conferring resistance to sugarcane mosaic virus (SCMV) in maize. Theor Appl Genet 100:934–941

Xu X, Hsia AP, Zhang L, Nikolau BJ, Schnable PS (1995) Meiotic recombination breakpoints resolve at high rates at the 5′end of a maize coding sequence. Plant Cell 12:2151–2161

Zhao B, Lin X, Leach JE, Hulbert S (2001) Molecular analysis of the maize gene conferring a non-host hypersensitive reaction to rice bacterial streak. Plant and Animal Genome IXth Conference, p 90

Theor Appl Genet (2002) 105:1190–1195
DOI 10.1007/s00122-002-0964-7

C. M. Dussle · M. Quint · M. L. Xu · A. E. Melchinger
T. Lübberstedt

Conversion of AFLP fragments tightly linked to SCMV resistance genes *Scmv1* and *Scmv2* into simple PCR-based markers

Received: 31 October 2001 / Accepted: 14 December 2001 / Published online: 22 October 2002

Abstract In a previous study, bulked segregant analysis with amplified fragment length polymorphisms (AFLPs) identified several markers closely linked to the sugarcane mosaic virus resistance genes *Scmv1* on chromosome 6 and *Scmv2* on chromosome 3. Six AFLP markers (E33M61-2, E33M52, E38M51, E82M57, E84M59 and E93M53) were located on chromosome 3 and two markers (E33M61-1 and E35M62-1) on chromosome 6. Our objective in the present study was to sequence the respective AFLP bands in order to convert these dominant markers into more simple and reliable polymerase chain reaction (PCR)-based sequence-tagged site markers. Six AFLP markers resulted either in complete identical sequences between the six inbreds investigated in this study or revealed single nucleotide polymorphisms within the inbred lines and were, therefore, not converted. One dominant AFLP marker (E35M62-1) was converted into an insertion/deletion (indel) marker and a second AFLP marker (E33M61-2) into a cleaved amplified polymorphic sequence marker. Mapping of both converted PCR-based markers confirmed their localization to the same chromosome region (E33M61-2 on chromosome 3; E35M62-1 on chromosome 6) as the original AFLP markers. Thus, these markers will be useful for marker-assisted selection and facilitate map-based cloning of SCMV resistance genes.

Communicated by G. Wenzel

C.M. Dussle · M. Quint · M.L. Xu · A.E. Melchinger
T. Lübberstedt (✉)
Institute of Plant Breeding, Seed Science, and Population Genetics,
University of Hohenheim, 70593 Stuttgart, Germany
e-mail: thomas.luebberstedt@agrsci.dk
Fax: +49-4558-113301

Present address:
M.L. Xu, College of Biosciences and Biotechnology,
Yangzhou University, 225009, Yangzhou, PRC

Present address:
T. Lübberstedt, Danish Institute of Agricultural Sciences,
Research Centre Flakkebjerg, 4200 Slagelse, Denmark

Keywords AFLP · CAPS · Indel marker · Maize · Marker conversion · Sugarcane mosaic virus (SCMV)

Introduction

Sugarcane mosaic virus (SCMV) is one of the most important virus diseases of maize (*Zea mays* L.) and causes serious yield losses in susceptible cultivars (Fuchs and Grüntzig 1995). It is naturally transmitted by aphids in a non-persistent manner, which makes control of SCMV vectors rather inefficient. Therefore, cultivation of resistant varieties is the most promising approach for controlling of SCMV.

Kuntze et al. (1997) screened 122 early-maturing European inbred lines for resistance to SCMV and MDMV (maize dwarf mosaic virus) and identified three dent inbreds (D21, D32 and FAP1360A) displaying complete resistance under both field and greenhouse conditions. Two major genes, *Scmv*1 and *Scmv*2 (previously named *Scm1* and *Scm2*), conferring resistance to SCMV were mapped to chromosome arms 6S and 3L, respectively, in cross D145 × D32 by quantitative trait loci (QTL) analysis (Xia et al. 1999) and in cross F7 × FAP1360A by bulked segregant analysis (BSA) (Xu et al. 1999) and QTL analysis (Dussle et al. 2000). As resistance against SCMV is strongly affected by environmental conditions (Melchinger et al. 1998), molecular markers turned out to be a good tool to determine the resistance genotype.

Identification of molecular markers closely linked to the SCMV resistance genes is an essential step towards both marker-assisted selection (MAS) and map-based cloning of these genes. Xu et al. (1999) identified 23 tightly linked amplified fragment length polymorphism (AFLP) markers for both major resistance genes by BSA: 11 markers linked to *Scmv2* on chromosome 3 and 12 linked to *Scmv1* on chromosome 6, including one AFLP marker cosegregating with *Scmv1*.

Although the AFLP technique is powerful and reliable in identifying markers closely linked to genes of in-

terest, it has some disadvantages for use in MAS and map-based cloning. Limitations to the large-scale, locus-specific application of AFLPs include their dominant type of inheritance, the intensity of labour involved, and the high costs. Hence, conversion of AFLP markers into sequence-specific polymerase chain reaction (PCR) markers is required for screening large breeding populations at low costs.

Sequence-specific PCR markers have been successfully developed by conversion of different marker types such as RFLPs, RAPDs and SSRs, (Bradshaw et al. 1994; Cheung et al. 1997; Jung et al. 1999). The conversion of AFLP markers into PCR-based markers has been accomplished for several species such as carrot (Bradeen and Simon 1998), brassica (Negi et al. 2000), asparagus (Reamon-Büttner et al. 2000), soybean (Meksem et al. 2001), apple (Xu et al. 2001), barley and wheat (Shan et al. 1999). However, the conversion of AFLP markers seems to be more difficult than the conversion of other marker types due to the loss of their sequence specificity after amplification of the AFLP-derived internal primers (Shan et al. 1999). Hence, AFLP polymorphisms related to *Eco*RI or *Mse*I restriction site differences will not be reflected in primers from an internal sequence (Shan et al. 1999).

The objective of the study reported here was to sequence the respective AFLP bands linked to SCMV resistance genes in order to convert these dominant markers into either indel (insertion/deletion) or cleaved amplified polymorphic sequence (CAPS) markers. These AFLP markers were previously identified by BSA to be closely linked with *Scmv1* on chromosome 6 (two markers) or *Scmv2* on chromosome 3 (six markers). Our goal was to obtain codominant, simple PCR-based markers as a tool for marker-assisted selection as well as for map-based cloning of *Scmv1* and *Scmv2*.

Materials and methods

Plant materials

Inbred lines used in this analysis were the SCMV-resistant European inbred lines FAP1360A, D21 and D32 and the highly susceptible lines F7, D408 and D145. The mapping population consisted of (1) a subset 87 $F_{2:3}$ families derived from a cross between D32 and D145 previously used by Vuylsteke et al. (1999) to develop a high-density AFLP map (1,355 markers), and (2) 27 resistant BC_5 individuals from the cross (F7 × FAP1360A) × F7.

Isolation and cloning of tightly linked AFLP markers

AFLP markers flanking *Scmv1* (E35M62-1, E33M61-1) and *Scmv2* (E33M61-2, E33M52, E38M51, E82M57, E84M59, E93M53) were identified in a BSA employing four DNA samples: both parental lines FAP1360A (resistant parent) and F7 (susceptible parent), as well as a resistant and a susceptible bulk (Xu et al. 1999). AFLP markers were named according to the standard list for AFLP primer nomenclature (Keygene, The Netherlands, http://wheat.pw.usda.gov/ggpages/keygene/AFLPs.html). AFLP primer pairs corresponding to the tightly linked AFLP markers were used to re-amplify the linked AFLP markers from resistant parent FAP1360A. The resulting bands were excised from the dried gel with a sharp, clean razor blade. The sliced DNA-containing gel was transferred into an Eppendorf tube, eluted twice with 200 µl TE (1 h each) and once with 200 µl ddH_2O (2 h). The eluted gel was then mixed with 50 µl ddH_2O and kept on boiling water for 5 min to release the DNA from the gel. After the gel debris was spun down, the DNA-containing supernatant was transferred into a new Eppendorf tube and used as template for the subsequent amplification.

For a given linked AFLP marker, the corresponding primer pair and the same reaction conditions as for the main amplification of AFLPs (Vos et al. 1995) were employed to re-amplify the isolated AFLP marker bands. Re-amplification products were excised from an agarose gel, extracted with Nucleospin Kit (Macherey & Nagel) and blunt-end cloned into the pBluescript vector.

Conversion of AFLP markers

After cloning, five white colonies from each transformation event were selected. Respective inserts were sequenced using the ALF-Express automated sequencer (Amersham Pharmacia, Freiburg). Sequencing reaction conditions were chosen following the manufacturer's (Amersham Pharmacia) suggestions, and the DNA sequences were analysed using the ALIGNPLUS 2.0 software package (http://www.scied.com/ses_alim.htm). If the sequencing of these first five clones showed identity for at least three of the five clones, new primers internal to the AFLP selective primers were designed using the PRIMER1.02 programme (http://www.scied.com/ses_pd4.htm) (Table 1). Otherwise, additional five white clones were sequenced to receive a majority of identical sequences for one genotype. The internal primers synthesized for fragments corresponding to the AFLP markers were employed to amplify fragments from the inbred lines F7 and FAP1360A, which represent the parent lines of the mapping population for BSA. Internal primers of the three AFLP markers E33M61-1, E33M61-2 and E35M62-1 were additionally employed on the four inbred lines D21, D32 (SCMV resistant), and D145, D408 (susceptible) in order to evaluate the relationship between polymorphisms and SCMV resistance. The extension "STS" was added to the names of the AFLP marker after synthesizing the internal primers in order to distinguish AFLP markers and converted markers.

Sequenced tagged site (STS) markers that differed in length after amplification were used immediately as indel (insertion/deletion) markers. In the case of an identical sequence length, enzyme recognition sites were identified using the CLONE manager software package (http://www.scied.com/ses_cm6.htm). Sequence regions displaying single nucleotide differences in restriction enzyme recognition sites between parent lines of mapping populations were used to identify CAPS markers, which were separated on a 3% MetaPhor agarose gel in 0.5× TBE buffer.

Linkage and statistical analyses

Based on the segregation data, the STS markers were mapped to previously constructed genetic linkage maps (Xia et al. 1999; Xu et al. 1999). Marker orders and map distances for population D32 × D145 were calculated with MAPMAKER 3.0B (Lander et al. 1987) using a LOD threshold of 3.0 and the mapping function of Kosambi (1944). Marker orders and genetic distances for population FAP1360A × F7 were calculated with CRIMAP 2.4 (Green et al. 1990) taking into account the meiotic interdependence of progenies and ancestors within a population of BC individuals from different generations.

Table 1 STS marker development in maize: detailed information on eight STS markers converted from AFLPs that are closely linked to resistance genes *Scmv1* (chromosome 6) and *Scmv2* (chromosome 3)

Marker locus	Chromosome	STS forward primer (5′ → 3′) STS reverse primer (5′ → 3′)	Sequence length after amplification with STS primers (bp)	Number of SNPs[a]	Marker type
		Primers analysed with inbred F7 and FAP1360A			
E33M52STS	3	CCATATCGTGTTGAGAAGGC CCACTCAATGCGGTGTCTAT	173	1	–
E38M51STS	3	CACCAAGAAGGTTTGGATCC GCGTACCAATTCACTAACCG	146	–	–
E82M57STS	3	AACCTCCTAGCGTCATGTAG AGTCCTGAGTAACGGATCC	166	–	–
E84M59STS	3	AACAACAGTTACCAGGCCAG CTTCAGATTCTCCCGAACCA	168	2	–
E93M53STS	3	GCTTGCCAATTCTGCATGCA	203	–	–
		Primers analysed with inbred lines F7, FAP1360A D21, D32, D145 and D408			
E35M62-1STS	6	GAGTCCTGAGTAACCGCCTA CTTCATGCCTCTCGTCG	152; 160	7	Indel
E33M61-1STS	6	ACTGCTTAGTCCTCGACAGA CGTACCAATTCAAGAGCGAC	195	–	–
E33M61-2STS	3	TCTTGTGCAACTACGACACC GATGATGGCATTGTCGAGGA	152	8	CAPS

[a] Identified between pairs of inbred lines

Results

The polymorphic markers linked to SCMV resistance, obtained after AFLP analysis, were in the range of 200 bp to 300 bp. All sequenced AFLP fragments contained the *Eco*RI adapter on the one end and the *Mse*I adapter on the other end. Five to ten clones obtained from inbred line FAP1360A were used to design primers for PCR amplification of genomic DNA (Table 1). Following amplification using these PCR primers, no difference in sequence length was identified for any of the markers between inbred lines F7 and FAP1360A, with sequence lengths ranging from 146 bp for marker E38M51STS to 203 bp for marker E93M53STS. For the additionally analysed inbred lines (D21, D32, D145 and D408), markers E33M61-1STS and E33M61-2STS showed identical sequence length for all six inbred lines (195 bp and 152 bp, respectively). A 152-bp sequence was identified with marker E35M62-1STS for all of the inbred lines D32, D21 and D408 except D145 (160 bp). This polymorphism of 8 bp between the parents of cross D32 × D145 could be easily detected on a 3% MetaPhor agarose gel (FMC) (Fig. 1). Genetic mapping was performed using 87 $F_{3:2}$ families of cross D32 × D145, which located E35M62-1STS on maize chromosome 6S between markers phi075 and phi077 within the *Scmv1* QTL region previously identified by Xia et al. (1999).

Six out of the eight markers showed identical sequences among the clones within each inbred line. In contrast, markers E33M61-1STS and E84M59STS resulted in single nucleotide differences between the clones within each inbred line. Single nucleotide polymorphisms (SNPs) between inbred lines F7 and FAP1360A were found for the four markers E33M61-2STS, E35M62-1STS, E33M52STS and E84M59STS. Recognition sites for restriction enzymes could be found only for marker E33M61-2STS, resulting in a different number of recognition sites for the restriction enzyme *Mnl*I (Fig. 2). For marker E33M61-2STS, *Mnl*I cuts the fragments of FAP1360A (Fig. 2) and D408 four times. The fragments of inbred lines F7, D21, D32 and D145 were cut only three times with *Mnl*I. Therefore, E33M61-2STS could be used as a CAPS marker (Fig. 3). Mapping of E33M61-2STS with the BC_5 mapping population (FAP1360A × F7) confirmed the same segregation pattern with its corresponding AFLP marker E33M61-2 and its location 7.3 cM above *Scmv*2.

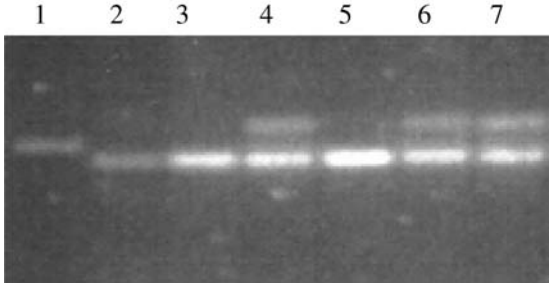

Fig. 1 Polymorphism in population D32 × D145 after conversion of AFLP primer E35M62-1 to the indel marker E35M62-1STS on a 1.5% agarose gel. *Lanes: 1* Susceptible parent D145, *2* resistant parent D32, *3–7* genotypes of the mapping population (*3*, *5* homozygous; *4*, *6*, *7* heterozygous band pattern)

Fig. 2 Recognition sites for restriction enzyme *Mnl*I in parents F7 (susceptible) and FAP1360A (resistant) after amplification with CAPS marker E33M61-2STS

```
F7          (   1)   GATGATGGCATTGTCGAGGATGAAGTCCCTGTGGCTCGATGAAGATGTCG
FAP1360A    (   1)   GATGATGGCATTGTCGAGGATGAAGTCCCTGTGGCTCGATGAAGAgGTCG

F7          (  51)   GAGAAGGATTTCCAGCAACGTGCCCACGAGCTTGTGGGACATGTCCATGG
FAP1360A    (  51)   GAGAAGGATTTCCAGCAACGTGCCCACGAGCTTGTGGGACATGTCCATGG

F7          ( 101)   AGGCAAGTTGTCCTACTATTCCACCCAGGTCGTTCCAGCGAACCAGGTGG
FAP1360A    ( 101)   AGGCAAGTTGTCCTgCTATTCCACCCAGGTCGcTCCAaCGAACCAGGTGG

F7          ( 151)   CCTTGCCGCCAGAAAGTCATGGTGAGGGTGTCGTAGTTGCACAA
FAP1360A    ( 151)   CCTTGCCGCCAGAAAGTCATGGTGAGGGTGTCGTAGTTcCACAA
```

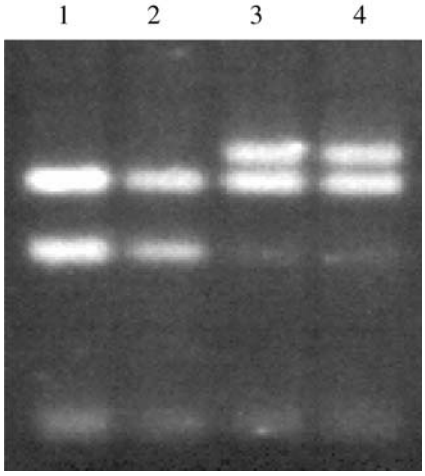

Fig. 3 CAPS marker E33M61-2STS digested with the restriction enzyme *Mnl*I corresponding to AFLP marker E33M61-2. *Lanes 1, 2* Resistant parent FAP1360A, *3*, *4* susceptible parent F7

Four out of the eight investigated markers revealed no SNPs after pairwise comparison of the inbreds. Although only half of the analysed markers showed between one and eight SNPs per inbred pair, 2.1 SNPs were found per inbred pair on average, resulting in one SNP per 71 bp. Analyses for stop codons identified continuous open reading frames for none of the sequences.

Discussion

In combination with BSA, AFLPs proved to be highly efficient for finding tightly linked molecular markers to the SCMV resistance genes *Scmv1* and *Scmv2* (Xu et al. 1999). However, AFLP markers are too costly and laborious for high-throughput monitoring of large numbers of genotypes. Hence, the conversion of AFLP markers closely linked to resistance genes is an important step to implement useful markers for MAS and map-based cloning, both of which require large population sizes of thousands of individuals. Marker conversion requires the characterization of the linked marker sequences and the design of locus-specific primers (Paran and Michelmore 1993). Reports on successful AFLP marker conversion are lacking so far in maize. With a size of 500–1500 bp, RAPD fragments are easier to convert to either sequence-characterised amplified region (SCAR) or CAPS markers than AFLP markers (Barret et al. 1998). Although there are doubts in converting short fragments like AFLPs (Negi et al. 2000), we were able to convert two short AFLP bands (150–300 bp) into PCR-based indel and CAPS markers without using methods like inverse PCR or chromosome walking. In contrast to DeJong et al. (1997) and Negi et al. (2000), who applied either inverse PCR or PCR walking to isolate the flanking regions for conversion of indel markers, we found with E35M62-1STS a polymorphism that could be used directly as an indel marker in populations generated from cross of D32 and D145.

After sequencing five to ten clones for each STS marker, we obtained six markers with identical sequences over most of the clones originating from one inbred. For markers E33M61-1STS and E84M53STS, SNPs were also found within all inbred lines. In total, 8 out of 28 inbreds investigated with the eight markers revealed single nucleotide changes within the inbred lines. The small number of published reports on AFLP marker conversion might be due to similar findings. Because identical sequences for the several clones of one inbred line were found technical problems in sequencing could be ruled out. The probability of residual heterozygosity for the inbred lines is below 0.025% because they were self-fertilized for more than 12 generations. Taking into account that the independent inbred lines revealed the same SNP within these inbreds in every case, it seems very unlikely that these polymorphisms were caused by residual heterozygosity. If we accept, the hypothesis that maize is an ancient tetraploid species (Gaut and Doebley 1997), the whole region harbouring the SCMV resistance gene might be duplicated, even though located at different regions of the genome. Under these conditions the segregation ratio would shift from 1r:3s (1 resistant to 3 susceptible) under a two dominant gene model for BC plants to 1r:7s under a three gene model. Although Xu et al. (1999) found a better fit with a three dominant gene model for the segregation within 20 $BC_{4:5}$ families of population F7 × FAP1360A segregating for SCMV resistance, the presence of additional SCMV resistance genes beside those on chromosomes 3 and 6 could not be confirmed for population F7 × FAP1360A. In contrast, mapping of the original AFLP markers did locate markers E33M61-1STS and E84M59STS exclusively to chromosomes 6 and 3, respectively.

Clusters of resistance genes originating from duplication during the evolution of maize may have led to slight sequence differences of paralogs, which may differ only in single nucleotides but not in total length. In mapping BAC clones of lettuce in order to analyze resistance gene clusters, Meyers et al. (1998) identified duplicates of AFLP markers in the same chromosome region. Genes conferring resistance to different pathogens are often clustered in the same chromosome region in the maize genome (McMullen and Simcox 1995). The fact that the AFLP markers corresponding to the converted STS markers mapped in the same regions previously reported to harbour clusters of resistance genes (McMullen et al. 1995) allows the assumption that the different marker sequences found within one inbred are linked to different resistance genes in the same chromosome region. The closer a marker is linked to a specific resistance gene, the higher might be the probability of being duplicated with the resistance gene during evolution. Hence, the occurrence of different sequences within one inbred line that map to the same chromosome region seems to be possible due to clustering. However, none of the sequenced AFLP fragments revealed any similarity to sequences known to be conserved within resistance genes.

The development of markers that can be easily handled is a prerequisite to the screening of large populations in order to clone the resistance genes *Scmv1* and *Scmv2*. The converted CAPS and indel markers will be useful to identify recombination events close to *Scmv1* and *Scmv2*. So far, it is unknown whether resistance genes cluster due to linkage or whether some of them are identical and display pleiotropy. In support of the existence of closely linked but different major resistance genes, Lübberstedt et al. (1999) found some susceptible plants in an allelism test between the three European dent inbreds D21, D32 and FAP1360A. The converted, closely linked markers identified in the present study could be used as probes for BAC screening in order to solve the question of whether the *Scmv1* and the *Scmv2* regions each harbour only a single locus or clusters of resistance loci.

Rafalski et al. (2001) analysed random cDNA clones in a collection of over 30 maize lines representative for the North American corn germplasm. Their analysis was restricted to coding regions. Sequence alignment revealed one SNP per 70 bp among the 30 lines. The authors emphasized that pairwise comparison between any two lines reveals a lower degree of polymorphism. In similar experiments, Useche et al. (2001) detected one SNP per 49 bp, although in non-coding regions. The low number of one SNP per 71 bp found in our study might be due to the pairwise sequence comparison in contrast to the sequence alignment of a large number of inbreds analysed by the previous authors. Taking into account that only four out of the eight converted AFLP markers showed polymorphism between inbred lines, it seems very likely that extending the fragment size by inverse PCR would increase the number of polymorphic STS-primers.

The CAPS marker E33M61-2STS turned out to be dominant in the mapping population. The polymorphism resulted in the presence of an additional band in the resistant parent FAP1360A that was absent in the susceptible parent F7 (Fig. 3). As in this mapping population the individuals were either homozygous for the susceptible parent allele or heterozygous, mapping with our mapping population of 27 resistant BC_5 individuals was not affected. However, even in this dominant case, CAPS markers are easier to apply than the original AFLP markers. In contrast to the AFLP markers, the converted markers do not require purified, high-molecular-weight DNA. Consequently, the application of simple STS markers enables a faster DNA isolation for a high number of individuals required for fine mapping. Additionally, the polymorphisms revealed by STS markers could be separated by an agarose gel, where no radioactivity is required to visualize the results. Compared to the analyses of AFLP markers in which polyacrylamide gels and radioactivity are used, the application of STS markers can reduce costs to about 20%.

The quality of a marker used for MAS depends on its predictive and/or diagnostic value (Borchardt and Weissleder 2000). Whereas the predictive value of a marker is determined by the inheritance of the marker and the linkage between marker and trait, the diagnostic value can be measured as the frequency of the desired linkage phase between marker and trait. Taking into account that resistant individuals of different populations harbour different resistance alleles of the same gene, cosegregation of these markers with the resistance trait in each population (F7 × FAP1360A, D32 × D145, D21 × D408) is not consequently preconditioned. By analysing inbred lines showing resistance, partial resistance and susceptibility to SCMV, Xu et al. (2000) suggested a single common ancestor for the resistance gene *Scmv1*. In the present study, no marker allele identical for all resistant or susceptible genotypes was identified. Therefore, the diagnostic value of these markers seems to be low. In the case of a low diagnostic value, the allelic phase of a marker has to be checked in each cross before it can be used in MAS (Borchardt and Weissleder 2000). A reason for the lack of resistance-allele-specific, cosegregating markers could be the presence of more than one SCMV resistance gene in the *Scmv1* region. Field experiments, BSA (Xu et al. 1999), and QTL analyses (Xia et al. 1999; Dussle et al. 2000) did not preclude the presence of more than one gene in the *Scmv1* region. Since different ancestors were expected for *Scmv2* (Dussle et al. 2000; Xu et al. 2000) and, therefore, different SCMV resistance genes within the *Scmv2* region, it was not possible to develop one single resistance-allele-specific marker for *Scmv2*.

Acknowledgements The present study was supported by a grant from the Deutsche Forschungsgemeinschaft, Grant No. LU601/2. The authors gratefully acknowledge the skilled technical assistance of Elisabeth Kokai-Kota.

References

Barret P, Delourme R, Foisset N, Renard M (1998) Development of a SCAR (sequence characterized amplified region) marker for molecular tagging of the dwarf BREIZH (*Bzh*) gene in *Brassica napus* L. Theor Appl Genet 97:828–833

Borchardt DC, Weissleder KK (2000) Marker-assisted selection in sugar beet: a breeder's perspective. In: Gallais A, Dillmann C, Goldringer I (eds) EUCARPIA Quantitative genetics and breeding methods: the way ahead. Proc Meet EUCARAPIA Sect Biometrics Plant Breed. INRA, Yessailles, pp 213–220

Bradeen JM, Simon PW (1998) Conversion of an AFLP fragment linked to the carrot Y_2 locus to a simple, codominant PCR-based marker form. Theor Appl Genet 97:960–967

Bradshaw HD Jr, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89:167–178

Cheung WY, Champagne G, Hubert N, Landry BS (1997) Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. Theor Appl Genet 94:569–582

DeJong W, Forsyth A, Leister D, Gebhardt C, Baulcombe DC (1997) A potato hypersensitive resistance against potato virus X maps to a resistance gene cluster on chromosome 5. Theor Appl Genet 95:246–252

Dussle CM, Melchinger AE, Kuntze L, Stork A, Lübberstedt T (2000) Molecular mapping and gene action of *Scm1* and *Scm2*, two major QTL contributing to SCMV resistance in maize. Plant Breed 119:299–303

Fuchs E, Grüntzig M (1995) Influence of sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV) on the growth and yield of two maize varieties. J. Plant Dis Protect 102:44–50

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809–6814

Green P, Falls K, Crooks S (1990) CRI-MAP, version 2.4. http://biobase.dk/Embnetut/Crimap

Jung G, Skroch PW, Nienhuis J, Coyne DP, Arnaud-Santana E, Ariyarathne HM, Marita JM (1999) Confirmation of QTL associated with common bacterial blight resistance in four different genetic backgrounds in common bean. Crop Sci 39:1448–1455

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Kuntze L, Fuchs E, Grüntzig M, Schulz B, Klein D, Melchinger AE (1997) Resistance of early-maturing European maize germplasm to sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV). Plant Breed 116:499–501

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

Lübberstedt T, Xia XC, Xu ML, Kuntze L, Melchinger AE (1999) Inheritance of resistance to SCMV and MDMV in European maize. In: Mugnozza GTS (ed) Proc 15th EUCARPIA 1998 Gen Congr Genet Breed Crop Quality Resistance. Kluwer, Dordrecht, pp 241–250

McMullen MD, Simcox KD (1995) Genomic organization of disease and insect resistance genes in maize. Mol Plant Microbe Interact 8:811–815

Meksem K, Ruben E, Hyten D, Triwitayakorn K, Lightfoot (2001) Conversion of AFLP bands into high-throughput DNA markers. Mol Gen Genet 265:207–214

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

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Negi MS, Devic M, Delseny M, Lakshmikumaran M (2000) Identification of AFLP fragments linked to seed coat color in *Brassica juncea* and conversion to SCAR marker for rapid selection. Theor Appl Genet 101:146–152

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance in lettuce. Theor Appl Genet 85:985–993

Rafalski JA, Ching A, Bhattramakki D, Morgante M, Dolan M, Register JC, Smith OS, Tingey S (2001) SNP markers in maize: discovery and applications. In: Grant D, Lazo G, Carollo V, Marshall K (eds) Plant Anim Genome 9th Conf

Reamon-Büttner SM, Jung C (2000) AFLP-derived STS markers for the identification of sex in *Asparagus officinalis* L. Theor Appl Genet 100:432–438

Shan X, Blake TK, Talbert LE (1999) Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. Theor Appl Genet 98:1072–1078

Useche F, Morgante M, Hanafey M, Tingey S, Martins W, Gao GR, Rafalski A (2001) Computer detection of single nucleotide polymorphisms (SNPs) in maize ESTs. In: Grant D, Lazo G, Carollo V, Marshall K (eds) Plant Anim Genome 9th Conf. San Deigo, p 144

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

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Xia XC, Melchinger AE, Kuntze L, Lübberstedt T (1999) QTL mapping of resistance to sugarcane mosaic virus in maize. Phythopathology 89:660–667

Xu ML, Melchinger AE, Xia XC, Lübberstedt T (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR and AFLP markers. Mol Gen Genet 261:574–581

Xu ML, Melchinger AE, Lübberstedt T (2000) Origin of *Scm1* and *Scm2* – two loci conferring resistance to sugarcane mosaic virus (SCMV) in maize. Theor Appl Genet 100:934–941

Xu ML, Huaracha E, Korban SS (2001) Development of sequence-characterized amplified regions (SCARs) from amplified fragment length polymorphism (AFLP) markers tightly linked to the *Vf* gene in apple. Genome 44:63–70

Theor Appl Genet (2003) 106:1171–1177
DOI 10.1007/s00122-002-1105-z

M. Quint · C. M. Dußle · A. E. Melchinger
T. Lübberstedt

Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping and MAS

Received: 4 March 2002 / Accepted: 7 May 2002 / Published online: 25 January 2003

Abstract The resistance gene analogue (RGA) pic19 in maize, a candidate for sugarcane mosaic virus (SCMV) resistance gene (R gene) *Scmv1*, was used to screen a maize BAC library to identify homologous sequences in the maize genome and to investigate their genomic organisation. Fifteen positive BAC clones were identified and could be classified into five physically independent contigs consisting of overlapping clones. Genetic mapping clustered three contigs into the same genomic region as *Scmv1* on chromosome 6S. The two remaining contigs mapped to the same region as a QTL for SCMV resistance on chromosome 1. Thus, RGAs mapping to a target region can be successfully used to identify further-linked candidate sequences. The pic19 homologous sequences of these clones revealed a sequence similarity of 94–98% on the nucleotide level. The high sequence similarity reveals potential problems for the use of RGAs as molecular markers. Their application in marker-assisted selection (MAS) and the construction of high-density genetic maps is complicated by the existence of closely linked homologues resulting in 'ghost' marker loci analogous to 'ghost' QTLs. Therefore, implementation of genomic library screening, including genetic mapping of potential homologues, seems necessary for the safe application of RGA markers in MAS and gene isolation.

Keywords RGA · SCMV · Maize · pic19 · Ghost marker · BAC

Communicated by G. Wenzel

M. Quint · C. M. Dußle · A. E. Melchinger · T. Lübberstedt
Institute of Plant Breeding, Seed Science,
and Population Genetics, University of Hohenheim,
Fruwirthstraße 21, 70593 Stuttgart, Germany,

T. Lübberstedt (✉)
Danish Institute of Agricultural Sciences,
Research Centre Flakkebjerg, DK-4200 Slagelse, Denmark,
e-mail: Thomas.Luebberstedt@agrsci.dk
Fax: ++45-58-11-33 01

Introduction

Plant pathogens exhibit a high mutation rate from avirulence to virulence, and natural selection favours the spread of new virulent races. In response to this, there is a co-evolution in plants that generate novel resistance protein variants, which are capable of recognising either the modified avirulence determinant or other pathogen components (Hammond-Kosack and Jones 1997). One mechanism for this is the clustering of R genes within large complex loci. Recombination between clustered loci leads to new pathotype-specific resistances. In a classical study with maize, Saxena and Hooker (1968) identified genes conferring resistance to *Puccinia sorghi*. They found 16 linked, but separate, loci conferring resistance to 16 different pathogen isolates (*Rp1A* to *Rp1N, Rp5, Rp6*). Another mode of clustering of R genes is displayed by linkage groups containing genes for recognition of different pathogens. In lettuce, linkage group I contains eight genes for resistance to downy mildew (*Bremia lactucae*) and a gene for aphid resistance (*Pemphigus bursarius*), whereas in linkage group II two additional downy mildew R genes are clustered with a gene for turnip mosaic virus resistance and a gene for resistance to the root pathogen *Plasmopara lactucae-radicis* (Landry et al. 1987; Kesseli et al. 1993; Witsenboer et al. 1995; Meyers et al. 1998a).

The recent cloning of a number of R genes by transposon tagging and positional cloning uncovered sequence homology in conserved amino-acid domains. This has been exploited to isolate hundreds of sequences with homology to R genes, RGAs, based on degenerate PCR primers derived from the conserved amino-acid domains. RGAs belonging to the same gene cluster displayed an even higher degree of sequence homology than unlinked RGAs (Ellis et al. 1995; Parniske et al. 1997; Meyers et al. 1998a), which can be explained by the evolution of many R genes by duplication. Circumstancial evidence would suggest that all organisms have experienced at least one round of genome duplication in their phylogenetic past. Thus all eukaryotes probably are

ancient polyploids (Leipoldt and Schmidtke 1982). Therefore, in the plant kingdom a high degree of homology between unlinked RGAs is also very likely. This has been shown for Arabidopsis (Meyers et al. 1999) and may also be valid for maize with its potential allotetraploid evolution (Gaut et al. 1997). Therefore, the maize genome contains duplicated chromosome segments with colinear gene arrangements (Helentjaris et al. 1995).

On maize chromosomes 6 and 3, R gene clusters include genes conferring resistance to several virus and bacterial diseases. Chromosome 6S contains R genes for SCMV, maize dwarf mosaic virus, wheat streak mosaic virus, high plains virus, rice bacterial streak, sorghum bacterial stripe and Southern corn leaf blight, whereas the cluster on chromosome 3L contains genes conferring resistance to SCMV, wheat streak mosaic virus, high plains virus, maize mosaic virus and maize chlorotic dwarf virus (for a review see Quint et al. 2000). There is strong evidence for the clustering of R genes on chromosomes 6 and 3, but the issue of pleiotropy versus clustering still needs to be solved for both regions.

Collins et al. (1998) isolated and mapped several RGAs to the maize genome. One of them, pic19, was mapped as a single-copy RGA to the *Scmv1* region on chromosome 6, and more intense investigations in SCMV-specific materials suggested pic19 to be a candidate gene for *Scmv1* (Quint et al. 2002). Hence, RGAs cosegregating or linked to a target R gene may provide a tool for the isolation of further linked RGAs based on homology to members of the same gene cluster. R gene and RGA clusters have the potential to be highly duplicated and span large physical distances (Meyers et al. 1998a) which might even be interrupted by other sequences. Therefore, the isolation of overlapping BAC/YAC clones spanning flanking markers by chromosome walking may be time-consuming and ineffective. In contrast, the approach outlined in this study fosters isolation of further candidate genes.

The objectives of this study were to: (1) identify homologues of pic19, (2) investigate their genomic organisation, (3) determine their map position to clarify whether they map to (a) the direct neighbourhood within the same cluster, (b) independent loci or (c) duplicated genome regions in the maize genome, and (4) evaluate the use of RGAs for genetic mapping, MAS, and gene cloning on the basis of these results.

Materials and methods

BAC screening

Scmv1 candidate pic19 was previously elongated to 624 bp (Quint et al. 2002). It was used as a probe to screen a B73 BAC library, purchased from Texas A & M BAC Center (http://hbz.tamu.edu/bac.html), covering the genome approximately 4 x. Labelling was performed by α-32P-dATP/dCTP random priming. Hybridisation and washing procedures of BAC membranes were conducted at 65 °C as proposed by the supplier (http://hbz.tamu.edu/bac.html). Autoradiography was carried out for 1 to 2 days at –80 °C with an intensifying screen.

Table 1 Primer sequences used for genetic mapping of BAC contigs 1 to 4

BAC contig	Primer	5′–3′
1	71L1 71R1	GGA AGC ATA TTG TCG TTG T GCA TGC TCC GTC GTA TG
2	pic19L 230R1	TAG ATG ATG TCT GGA CGG CT GGC ACA ATA CAG GGA A
3	239L2 pic19X	AGC CCT TGT GCC AAT AA GCA GTT CCT CTC TGC AAC GTG
4	UBC860	$(TG)_8$ RA

Restriction analysis

For identification of overlapping BAC clones and classification of independent contigs, BAC DNA was isolated using the Macherey & Nagel (Düren) Nucleobond BAC 100 Kit. 2µg of BAC DNA was digested with 10 U of either *Bam*HI or *Hin*dIII in a 30 µl reaction for 4 h at 37 °C. Digested DNA was separated on a 0.8% agarose gel. The presence or absence of bands of the same size was converted into a 1/0 matrix to identify overlapping BAC clones by scoring for fragments of common size. This process was performed for both restriction enzymes independently. To identify false positives to be excluded from further experiments, gels were blotted and hybridised with pic19 following the same protocols as mentioned above.

Cloning of pic19 homologues

Specific primers for pic19 (pic19L + pic19X, Table 1) were used to amplify pic19 homologous sequences out of each BAC clone. DNA amplification was performed with a standard reaction mix containing 10 ng of BAC DNA, 10 mM of Tris–HCl (pH 9), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.3 mM of dNTPs, 0.5 pMol of each primer and 0.625 U of *Taq* DNA polymerase (Amersham Pharmacia, Freiburg). After an initial denaturation step at 94 °C for 2 min, template DNA was amplified using 35 cycles under the following conditions: 1 min at 94 °C, 2 min at 55 °C and 2 min at 72 °C. Final extension was conducted at 72 °C for 2 min. Amplification products were excised from the agarose gel, extracted with the Nucleospin Kit (Macherey & Nagel, Düren) and ligated into pGEM-T (Promega, Mannheim) plasmids. DNA sequencing was performed using the ALFExpress automated DNA sequencer (Amersham Pharmacia, Freiburg). Sequencing reaction conditions were chosen as suggested by the manufacturer (Amersham Pharmacia, Freiburg). Sequence alignment was performed using the ALIGN Plus 2.0 software package (http://www.scied.com/ses_alim.htm).

Mapping populations

Since the BAC library was derived from maize inbred line B73, all BAC contigs except one were mapped using the commercially available B73 × Mo17 RI (IBM) mapping population (PopA) consisting of a subset of 94 individuals. PopA contains more than 570 RFLP and SSR markers (Davis et al. 2000; http://www.cafnr.missouri.edu/mmp/ibmmaps.htm). The mapping population is derived from a cross of B73 × Mo17, which underwent random mating for four generations.

The last contig was mapped to a subset of 84 individuals of a D32 × D145 $F_{2:3}$ mapping population (PopB) containing 100 RFLP and SSR markers (Xia et al. 1999).

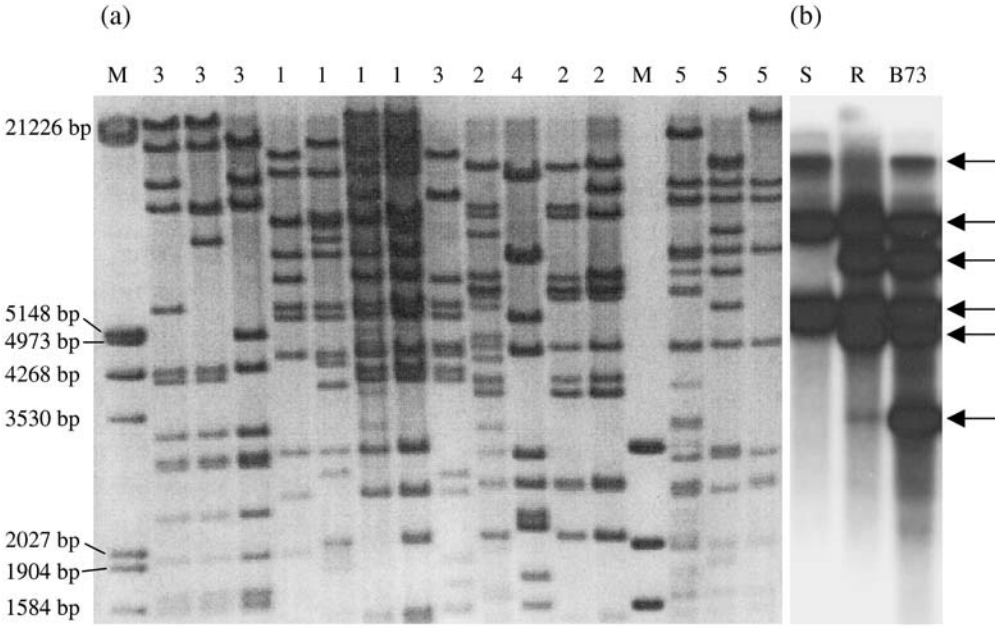

Fig. 1 (a) Identification of overlapping BAC clones and classification by *Hin*dIII restriction analysis into contigs 1–5; (b) Autoradiogram of *Hin*dIII-digested DNA of SCMV susceptible maize inbred line F7 (*S*), SCMV resistant maize inbred line FAP1360A (*R*) and B73 using pic19 as a probe

Mapping strategy and BAC mapping

The BAC mapping strategy included two steps: (1) the design of BAC contig-specific primers (Table 1) based on the DNA sequence of the pic19 homologues; (2) amplification of contig-specific PCR products discriminating both parents of the PopA, B73 and Mo17 (PCR conditions a, b). In the case of identical size of PCR fragments, pic19 homologues were converted to CAPS markers (c). Independent from the pic19 homologues, contig-specific amplification of inter-simple sequence repeat (ISSR) fragments in PopA (d) or contig-specific single-copy RFLPs of BAC subclones in PopB (e) were employed as alternative strategies for BAC mapping.

DNA amplification was performed in a standard reaction mix containing 50 ng of genomic DNA, 10 mM of Tris–HCl (pH 9), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.3 mM of dNTPs, 0.5 pMol of each primer and 0.625 U of *Taq* DNA polymerase (Amersham Pharmacia, Freiburg). After an initial denaturation step at 94 °C for 3 min, template DNA was amplified using 30 cycles with one of the three conditions: 45 s at 94 °C, 45 s at 62 °C (a), 48 °C (b) or 45 °C (c), and 1 min at 72 °C. The final extension step was conducted at 72 °C for 10 min. For CAPS analysis a *Mse*I recognition site was used to construct polymorphism between the parental DNAs. Therefore, the PCR product was digested with *Mse*I (10 µl of the PCR product with 5 U of *Mse*I in a 20-µl volume for 4 h at 37 °C) and separated on 1.5% agarose gels.

The ISSR PCR reaction mix contained 50 ng of genomic DNA, 10 mM of Tris–HCl (pH 9), 50 mM of KCl, 2.5 mM of $MgCl_2$, 0.5 mM of dNTPs, 1.0 pMol of the primer and 1.0 U of *Taq* DNA polymerase (Amersham Pharmacia, Freiburg). After initial denaturation at 94 °C for 1 min, template DNA was amplified using 35 cycles under the following conditions: 30 s at 94 °C, 30 s at 45 °C and 2 min at 72 °C. Final extension was conducted at 72 °C for 10 min.

Subcloning was performed by double-digesting BAC clones with 1 U of both *Bam*HI and *Hin*dIII in a 30-µl reaction using 200 ng of BAC DNA (Amersham Pharmacia Biotech, Freiburg). The restriction reaction was conducted at 37 °C for 4 h. After purification, 50 ng of double-digested BAC DNA was shotgun-cloned into pBluescript. For identification of single-copy clones, plasmid DNA was Southern blotted and hybridised using 100 ng of genomic maize DNA as a probe. Plasmids revealing no signals were tested for polymorphism between parents of PopB, because PopA was not available for Southern analysis. Genomic DNA (10 µg) of 84 individuals of PopB was *Hin*dIII-digested and Southern blotted to the Hybond N+ nylon membrane (Amersham Pharmacia Biotech, Freiburg). Hybridisation procedures were identical to those from BAC screening.

Linkage and statistical analyses

For PopA, linkage groups were constructed using MapMaker for UNIX version 3.0 (Whitehead Institute, Cambridge, Mass.). Initially, linkage groups were defined at a LOD of 6.0. The maps were constructed using the RI self option and the Haldane map function. All remaining markers were assigned at LOD 4.0. These computations were conducted at the University of Columbia, Missouri. The IBM population is an intermated recombinant inbred population. The random mating process increased the average number of recombination events per individual by approximately 3-fold compared with F_2 or RI-derived mapping lines. This is equivalent with a 3-fold expansion of map distances (Liu et al. 1996). Therefore, all map distances in the IBM map were divided by three.

For PopB, marker orders and genetic distances were calculated with MapMaker 3.0b (Lander et al. 1987) using a LOD threshold of 3.0 and the Kosambi mapping function.

Results

BAC analysis

Southern hybridisation of *Hin*dIII-digested B73 genomic DNA using pic19 as a probe identified six prominent bands (Fig. 1b). Screening of the BAC library identified 19 positive clones. Four clones were excluded that showed no positive signals when hybridised to pic19. Evaluation of the *Bam*HI or *Hin*dIII restriction gels of the remaining 15 BACs resulted in five groups of overlapping BAC clones (Table 2).

Table 2 Grouping of overlapping BAC clones based on *Bam*HI and *Hin*dIII restriction analysis

Contig 1 – pic19-1	Contig 2 – pic19-2	Contig 3 – pic19-3	Contig 4 – pic19-4	Contig 5 – pic19-5
71E12	44E24	191F2	181B13	62H1
103E15	230A6	198O20		192O11
1,75O11	267B13	212D24		241P7
2,05O18		239B24		

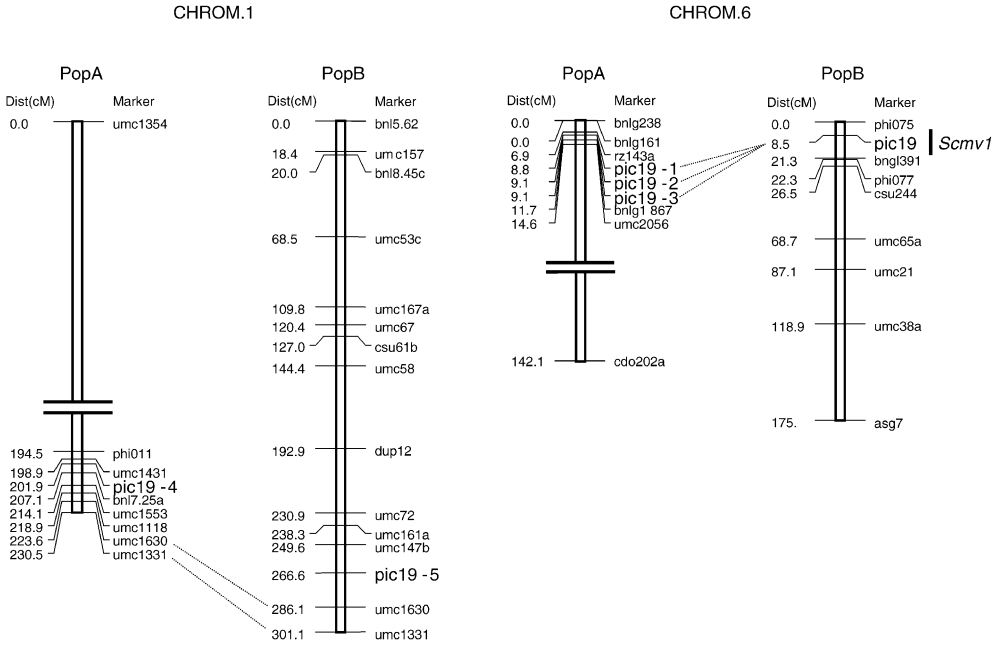

Fig. 2 Genetic mapping of pic19 homologues to maize chromosomes 1 and 6

Sequence alignments

RGA sequences homologous to pic19 could be amplified in four out of the five BAC contigs. Sequence comparisons revealed an identity of 94 to 98% among the pic19 RGA homologues of contigs 1 to 4. All differences were due to single nucleotide changes. Only contig 4 showed a continuous open reading frame (ORF). Contig 3 showed one stop codon. Contigs 1 and 2 revealed several frame shifts due to InDels.

Genetic mapping

PCR amplification of the pic19 homologues was not possible for contig 5. For contigs 1 and 2 [corresponding to mapping conditions (a) and (b) from the 'Materials and methods' section], SNP primers were designed resulting in differential amplification on the BAC as well as the genomic DNA level (Table 2). Genetic mapping placed both contig 1 and contig 2 into the chromosomal bins 6.00–6.01 within the marker interval *rz143a–bnlg1867* (Fig. 2) in PopA. Contig 1 and 2 displayed close linkage with a genetic distance of 0.37 cM.

For contig 3 (c), differential amplification was possible on the BAC level. Using the same primer pair, the parents of the mapping population did not show polymorphism. Cloning and sequencing of the regarding parental fragments revealed a SNP in a *Mse*I recognition site, which was used as a CAPS marker to place contig 3 in the same marker interval as contigs 1 and 2, cosegregating with contig 2. The evaluation of the marker scores revealed only two recombinants between contigs 1 and 2. Both individuals did not amplify in contig 3. For the remaining individuals, contig 3 scores were identical to both contig 1 and contig 2 scores.

For contig 4 (d), none of the previous approaches was successful. Screening of the five contigs with ISSR primers identified a contig 4-specific fragment using ISSR primer UBC860. The polymorphism was conserved between the two parental lines B73 and Mo17 on the genomic DNA level. Contig 4 was mapped to chromosomal bin 1.10 within the marker interval *umc1431–bnl7.25* (Fig. 2).

It was not possible to amplify pic19 homologues by PCR from contig 5 (e). Therefore, one member of the contig, BAC 192O11, was shotgun-subcloned. Subclone III-17 was used as a single-copy RFLP probe in PopB and this mapped contig 5 to the end of chromosome 1 into the marker interval *umc161a–umc147b* (Fig. 2). Since contigs 4 and 5 were mapped in different populations, their map distance was not determined.

Discussion

Genomic organisation of the pic19 RGA family

Southern hybridisation of pic19 to B73 genomic DNA revealed six prominent bands (Fig. 1b). Taking into account the size of the bands on the Southern blot and the size of the random primed pic19 probe, these six bands represent between three and six loci in the B73 genome. The five different contigs of overlapping BAC clones agree with this hypothesis. However, due to potentially absent sequences in the used maize BAC library (approximately 4 × genome coverage) and the high stringency of hybridisation conditions (65 °C), the existence of further homologues to pic19 cannot be ruled out.

The function of RGAs flanking R genes may be various: they might display: (1) R genes conferring isolate-specific resistance, (2) pseudogenes representing an important source for the evolution of new resistance-specificities by recombination or gene conversion with functional genes, or (3) rudiments of already overcome resistances. This study demonstrates, as others before, that homologous RGAs can be tightly linked. But is also demonstrates that they can be simultaneously located at different loci. There is no evidence for duplicated loci between the two map positions of pic19 homologues in chromosomal bins 1.10 and 6.00/01 according to the autotetraploidy hypothesis (Helentjaris et al. 1995; Gaut and Doebley 1997). However, initial duplication followed by single nucleotide changes is apparently the driving force in the evolution of these loci. NBS regions appear to be under purifying selection consistent with its proposed but unproven effector function (Michelmore and Meyers 1998).

Clustering of RGA homologues, such as observed in this study, for pic19 is consistent with reports in maize as well as other crops. Studies on the *Pto, Cf* and *Dm* clusters in tomato and lettuce (Martin et al. 1993; Thomas et al. 1997; Meyers et al. 1998b) revealed the existence of numerous RGAs in the direct physical neighbourhood of R genes. Data from the Arabidopsis genome sequencing project uncovered 160 R genes (The Arabidopsis Genome Intiative 2000). The NBS sequences currently in the databases are located in approximately 21 genomic clusters and 14 isolated loci (Meyers et al. 1999). Besides the RGA clusters in maize identified by Collins et al. (1998), Zhao et al. (2001) noted the mapping of two maize R genes, *Rxo* and *Rpa*, to the same position on the short arm of maize chromosome 6 (Fig. 2) in the *Scmv1* region. Furthermore, the authors identified 5–6 RGAs in the direct neighbourhood of *Rxo* and *Rpa*. Therefore, clustering of homologous sequences within short chromosomal stretches seems also to be a common feature in the organisation of R genes and RGAs in maize.

Restriction analysis identified five contigs which are not overlapping with each other. Since contigs 1–3 and 4–5 are mapping to different chromosomes, overlapping can also be excluded. Genetic mapping identified two recombinants between contigs 1 and 2. Furthermore, restriction analysis clearly separates contigs 1 to 3. Therefore, the minimum size of the region covered by the three contigs on chromosome 6 estimated by the length of restriction fragments is 450 kbp. Repeated PCR using contig 3-specific primers did not amplify any product of the two genotypes recombinant between contigs 1 and 2. Therefore, it seems likely that these individuals lost the corresponding sequence. Pryor et al. (1987) and Bennetzen et al. (1988) reported spontanous mutations to susceptibility in different *rp1* maize rust R genes with frequencies of up to 0.5%. It has been proposed that the instability of this region is due to gene conversion or unequal crossing-over events between mispaired sequence repeats during meiosis (Sudupak et al. 1993; Hu and Hulbert 1994). The same mechanisms could explain the loss of contig 3-specific sequences in these two recombinant genotypes.

Xia et al. (1999) identified five QTLs for SCMV resistance in PopB. One major QTL mapped near *nor* on the short arm of chromosome 6 (*Scmv1*) and another major QTL near the centromere on the long arm of chromosome 3 (*Scmv2*). Three minor QTLs were located on chromosomes 1, 5 and 10. Quint et al. (2002) mapped pic19 in PopB to the *Scmv1* region on chromosome 6.00/01. Therefore, it was expected that at least one pic19 homologue maps to this region. Linkage of three BAC contigs containing pic19 homologues supports the hypothesis of a R gene cluster rather than one single pleiotropic R gene in this region. Especially interesting is the map position of contigs 4 and 5, which map to the same region as the QTL previously detected for SCMV resistance on chromosome 1 (Fig. 2). Therefore, these contigs might as well contain candidate genes underlying the QTL in this region.

Sequence comparison of pic19 homologues

The pic19 homologues which were amplified by PCR in four out of the five contigs displayed a very high sequence similarity of 94–98% at the nucleotide level. A comparison of 11 genes of the tomato Cf-4/9 cluster (Parniske et al. 1997) exhibits an equally high degree of overall sequence homology (92–99%). Shen et al. (1998) sequenced part of the NBS motif of cross-hybridising lettuce RGAs and showed that linked members of the same RGA families had an identity in the range of 54–98% at the deduced amino-acid level. The authors defined RGAs displaying at least 50% sequence similarity on the amino-acid level as members of the same RGA family. The pic19 RGA family includes members clustering at two different genome regions in contrast to the other discussed RGA families. However, the relationship between sequence similarity and the physical position of NBS-encoding RGAs is very complex and cannot be generalised because some clusters also contain quite diverse sequences (Meyers et al. 1999).

Consequences for gene isolation and development of closely linked molecular markers

Our study shows that RGAs mapping to the target region can be successfully used to identify further-linked candidate sequences. This approach is useful with regard to map-based cloning, which has been successfully used in maize to clone the *Rp1-D* gene, a member of the complex locus composed of approximately nine gene homologues, conferring resistance to common rust (*Puccinia sorghi*) (Collins et al. 1999).

A second benefit is the possibility to develop closely linked molecular markers for application of MAS in practical plant breeding. Furthermore, RGAs were shown to be highly polymorphic and the rate of success of converting RGAs into codominant CAPS or dominant SNP markers is high compared to other genomic sequences (Quint et al. 2002).

However, the high sequence similarity of these linked RGA sequences reveals potential problems for the use of RGAs as molecular markers. Discrimination of homologous sequences by PCR might be problematic because of simultaneous amplification of several homologues (Fig. 3c). Therefore, these indistinguishable and genetically linked homologues might be mapped as a single-copy 'ghost' marker analogous to a 'ghost' QTL (Martinez and Curnow 1992) as illustrated by an example in Fig. 3. Suppose a RGA was sequenced from one end in two genotypes 1 and 2, and a SNP (SNP1) is identified between them (in Fig. 3a) and used for mapping in a segregating population. Suppose further that the two real homologous copies RGA1 and RGA2, which can be distinguished in both genotypes by a second SNP (SNP2), remain unrecognised. Hence, scoring the polymorphism between the parental genotypes 1 and 2 at SNP1 results in a single-copy 'ghost' marker representing RGA1 and RGA2. Using a codominant marker system in a segregating progeny of the parental genotypes 1 and 2 results in deviations between the marker genotypes of the 'ghost' RGA and those of the real RGAs (Fig. 3c, d). The expected genotype frequencies will be shifted towards heterozygotes. Instead of estimating the correct map positions of RGA1 and RGA2 relative to linked marker loci, the position of the fictive 'ghost' marker is estimated. As a consequence, the position of the 'ghost' marker may change the positions and even the order of other marker loci and target genes or QTLs (Fig. 3c, for details see Frisch et al. 2003). Map distances will be overestimated and the positions of the fictive 'ghost' marker might deviate greatly from the true marker loci RGA1 and RGA2. The divergence of the map positions of the 'ghost' marker and its real underlying loci grows with the map distance between the real loci of RGA1 and RGA2. Wrong estimates of map positions of 'ghost' marker loci might have fatal effects for map-based cloning or MAS. Relative to the position of this 'ghost' marker and other markers, one also arrives at completely wrong estimates of map positions of the hypothetical target genes.

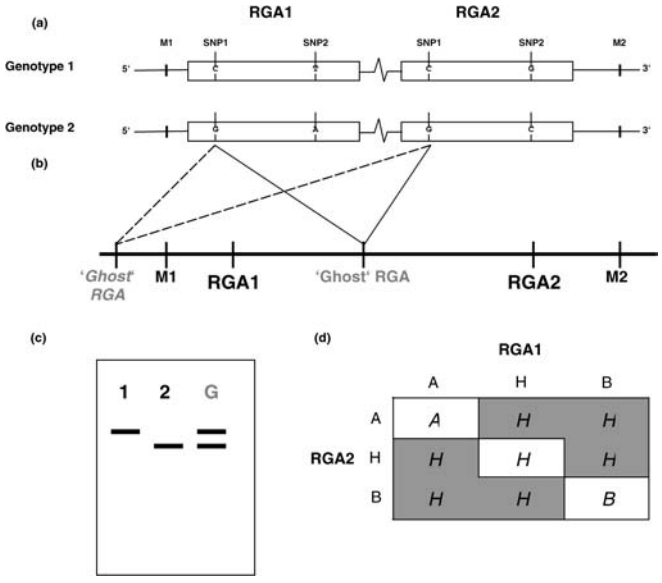

Fig. 3a–d Occurrence of 'ghost' markers. (**a**) Physical organisation of two homologous linked sequences RGA1 and RGA2 and two flanking markers M1 and M2; the RGAs can be distinguished between two genotypes by SNP1 and amongst each other by SNP2; (**b**) using SNP1 for scoring the RGA genotype in a segregating population results in the mapping of both homologous sequences to one single-copy 'ghost' RGA; the map position of this 'ghost' RGA may either be in between RGA1 and RGA2 but can also be outside this interval, and may even result in different orders of marker loci ('ghost' RGA in *italics*). (**c**) banding pattern of RGA1 (1), RGA2 (2) and the resulting 'ghost' marker (*G*), simultaneous amplification of 1 and 2 would result in banding pattern G; (**d**) the resulting scoring of RGA1 and RGA2 with a codominant marker system; scores for the 'ghost' marker are shown in *italics*, deviations of scores between the 'ghost' marker and one of the real RGAs are indicated by a *shaded background*

To avoid these scenarios, the implementation of a genomic library screening, including genetic mapping of potential homologues as reported in this paper, seems necessary for safe application of RGA markers for gene isolation and MAS. Especially for R genes, it has been demonstrated that clustering of homologous sequences is a common feature in the plant kingdom. However, these complications in the application of RGA sequences as molecular markers are not restricted to RGAs. They are existent for all DNA markers derived from potentially duplicated sequences, such as ESTs from large gene families, AFLPs or other markers located in recombination hotspots.

Acknowledgements The present study was supported by a grant from the Deutsche Forschungsgemeinschaft, Grant No. LU601/3. The authors gratefully acknowledge the skilled technical assistance of Elisabeth Kokai-Kota. We would also like to thank Michael D. McMullen for linkage analysis and Matthias Frisch for his helpful suggestions on the issue of "ghost" markers.

References

Bennetzen JL, Qin MM, Ingels S, Ellingboe AH (1988) Allele-specific and mutator-associated instability at the *Rp1* disease-resistance locus of maize. Nature 332:369–370

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogues in maize. Mol Plant Microbe Interact 11:968–78

Collins NC, Drake J, Aycliffe M, Sun Q, Ellis J (1999) Molecular characterisation of the maize Rp1-D rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

Davis G, Musket T, Melia-Hancock S, Duru N, Qu J, Sharopova N, Schultz L, McMullen MD, Woodman WL, Long MJ, Lee M, Vogel JM (2000) A high-resolution genetic map of the B73 × Mo17 population. Maize Genetics Conference Abstracts 42:P79

Ellis J, Lawrence GJ, Finnegan EJ, Anderson PA (1995) Contrasting complexity of two rust resistance loci in flax. Proc Natl Acad Sci USA 92:4185–4188

Frisch M, Quint M, Lübberstedt T, Melchinger AE (2003) Duplicate marker loci can result in incorrect locus order on linkage maps. (submitted)

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809–6814

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 48:575–607

Helentjaris T (1995) Atlas of duplicated sequences. Maize Genet Coop Newslett 69:67

Hu G, Hulbert S (1994) Evidence for the involvement of gene conversion in meiotic instability of the *Rp1* rust resistance genes in maize. Genome 37:742–746

Kesseli R, Witsenboer H, Stanghellini M, Vandermark G, Michelmore RW (1993) Recessive resistance to *Plasmopara lactucae-radicis* maps by bulked segregant analysis to a cluster of dominant resistance genes in lettuce. Mol Plant-Microbe Interact 6:722–728

Liu SS, Kowalsky SP, Lan TH, Feldmann KA, Peterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics 142:247–258

Lander ES, Green P, Abraramson J, Baslow 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 1:174–184

Landry BS, Kesseli R, Farrara B, Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331–337

Leipoldt M, Schmidtke J (1982) In: Dover G, Flavell R (eds) Genome evolution. Academic Press, New York, pp 219–236

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–6

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

Meyers BC, Chin, DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998a) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998b) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 11:1833–1846

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD (1997) Novel disease resistance specifities results from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell 91:821–832

Pryor T (1987) Stability of alleles of *Rp* (resistance to *Puccinia sorghi*). Maize Genet Newslett 61:37–38

Quint M, Melchinger AE, Dußle CM, Lübberstedt T (2000) Breeding for virus resistance in maize. Genetika 32:529–545

Quint M, Mihaljevic R, Dußle CM, Xu ML, Melchinger AE, Lübberstedt T (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for SCMV resistance in maize. Theor Appl Genet 105:355–363

Saxena KMS, Hooker AL (1968) On the structure of a gene for disease resistance in maize. Proc Natl Acad Sci USA 61:1300–1305

Shen KA, Meyers BC, Islani-Faridi MN, Chin DB, Stelly DM, Michelmore RW (1998) Resistance gene candidates indetified by PCR with degenerate oligonnucleotide primos map to clusters of resistance genes in lettuce. Mol Plant-Microbe Interact 8:815–823

Sudupak MA, Bennetzen JL, Hulbert SH (1993) Unequal exchange and meiotic instability of disease-resistance genes in the *Rp1* region on maize. Genetics 133:119–125

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JDG (1997) Characterisation of the tomato *Cf-4* gene for resistance to *Cladiosporum fulvum* identifies sequences that determine recognitial specifity in *Cf-4* and *Cf-9*. Plant Cell 9:2209–2224

Witsenboer H, Kesseli R, Fortin MG, Stanghellini M, Michelmore RW (1995) Sources and genetic fine structure of a cluster of genes for resistance to three pathogens of lettuce Theor Appl Genet 91:178–188

Xia XC, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. Phytopathology 89:660–667

Zhao B, Lin X, Leach JE, Hulbert S (2001) Molecular analysis of the maize gene conferring a non-host hypersensitive reaction to rice bacterial streak. Plant and Animal Genome IX Conference, p 90

Duplicate marker loci can result in incorrect locus orders on linkage maps

M. Frisch,[1] M. Quint,[1] T. Lübberstedt,[2] and A. E. Melchinger

Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

Abstract: Genetic linkage maps, constructed from multi-locus recombination data, are the basis for many applications of molecular markers. For the successful employment of a linkage map, it is essential that the linear order of loci on a chromosome is correct. The objectives of this theoretical study were to (i) investigate the occurrence of incorrect locus orders caused by duplicate marker loci, (ii) develop a statistical test for detection of duplicate markers, and (iii) discuss the implications for practical applications of linkage maps. We derived conditions, under which incorrect locus orders do occur or do not occur with duplicate marker loci for the general case of n markers on a chromosome in a BC_1 mapping population. We further illustrated these conditions numerically for the special case of four markers. On the basis of the extent of segregation distortion, an exact test for the presence of duplicate marker loci was suggested and its power was investigated numerically. Incorrect locus orders caused by duplicate marker loci can (a) negatively affect the assignment of target genes to chromosome regions in a map-based cloning experiment, (b) hinder indirect selection for a favorable allele at a quantitative trait locus, and (c) decrease the efficiency of reducing the length of the chromosome segment attached to a target gene in marker-assisted backcrossing.

GENETIC linkage maps are constructed from observed recombination frequencies between loci in experimental or natural populations with known pedigree. They are an essential tool for practical applications such as marker-assisted selection, marker-assisted backcrossing, and map-based cloning of target genes. For these applications a correct linear order of loci within linkage groups is essential. Errors in locus order can seriously hamper the ability to map, isolate, or select for simple Mendelian and complex traits.

Duplications of chromosome regions occur frequently and seem to be an important mechanism of genome evolution (Ohno 1970). More than one third of a typical eukaryotic genome consists of duplicated genes and gene families. Such gene families can arise from polyploidization events such as those presumed to have preceded the origin of many plant species (Wendel 2000).

The portion of genes in the different model organisms concerning their presence as singletons or duplicate members of gene families is very variable. Tandem gene duplications appear to be ubiquitous in plant genomes (Acarkan et al. 2000, Tarchini et al. 2000). The complete genome sequence of Arabidopsis has revealed that an estimated 17% of the 25 000 genes is arranged in tandemly repeated segments (The Arabidopsis Genome Initiative 2000). For the monocot model organism rice, this portion of locally duplicated genes accounts for an estimated 22% of the approximately 30 000 genes available from the draft sequence (Goff et al. 2002). For both species, it has been recognized that 60% of the genome is contained within large duplicated segments (Blanc et al. 2000, Goff et al. 2002), with almost half of the Arabidopsis genes within the duplicated segments being conserved.

If a duplicate chromosome region contains a DNA sequence, which can be used as a molecular marker, the marker alleles at the two duplicate marker loci cannot be distinguished. Equal fragment length results in an equal banding pattern, and consequently, the alleles of duplicate markers are scored in a mapping population as the alleles of one single marker. The recombination frequency between this non-existing 'ghost marker' and non-duplicated markers are different from those between the non-duplicated markers and the duplicate marker loci actually underlying the ghost marker. Since the locus order of linkage groups is determined on the basis of recombination frequencies between loci, incorrect recombination values for a linkage group can result in an incorrect locus order for the chromosome.

We encountered this phenomenon in a study with resistance gene analogues (RGAs, Quint et al. 2003). More generally, it can also be found for marker systems based on polymorphisms in short sequence stretches such as amplified fragment length polymorphism (AFLP) markers. Vuylsteke et al. (1999) conducted comprehensive AFLP mapping in two maize populations. From more than 1000 markers mapped in the individual populations, 353 AFLP markers were in common, i.e., a given AFLP primer combination resulted in polymorphic AFLP fragments of identical size in both populations. 327 of these 353 AFLP markers (>92%) were considered as colinear between both populations. However, the remaining 26 common AFLP markers (7.4%) mapped to different chromosomes in both populations. Thirteen of the respective AFLP fragments were sequenced. For three of these bands, sequences were (almost) identical, whereas for the other 10 bands sequence identity was restricted to restriction sites and selective nucleotides employed in the AFLP assay.

Corresponding author: Albrecht E. Melchinger, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany. Email: melchinger@uni-hohenheim.de. Received Nov. 11, 2003.

[1] Both authors contibuted equally to this article.

[2] Present address: T. Lübberstedt, Danish Institute of Agricultural Sciences, Research Centre Flakkebjerg, DK-4200 Slagelse, Denmark.

The term ghost marker was coined in analogy to the ghost QTL phenomenon (Martinez and Curnow 1992). The effect of such ghost markers on the construction of linkage maps and the consequences for marker-assisted selection as well as map-based cloning of target genes has not yet been investigated.

The objectives of our study were to (i) derive the recombination frequency between a marker and a ghost marker, (ii) derive conditions under which duplicate markers result in an incorrect locus order of the respective linkage group, (iii) investigate under which conditions the correct locus order is found, even if there are duplicate markers in the linkage group, (iv) develop a test for detection of duplicate markers, and (v) discuss the consequences of duplicate markers for applications of linkage maps such as map-based cloning, marker-assisted selection and marker-assisted backcrossing.

Theory

Definitions: Assuming (1) a diploid species and (2) two duplicate marker loci carrying alleles, which cannot be distinguished by the laboratory method used for the molecular marker analysis, the four alleles at the two duplicate markers are scored as the alleles of only one marker, which we call 'ghost marker'. Segregation ratios, recombination frequencies with other loci, and the map position of a ghost marker are in general not identical with the corresponding parameters for the underlying duplicate markers.

Non-duplicate markers, of which the alleles can be distinguished by the laboratory method used for the molecular marker analysis, are referred to as 'distinguishable markers'. The term marker is used in the sense of distinguishable marker, when there is no further specification as a ghost marker or a duplicate marker.

An incorrect locus order is defined as an order, which cannot be obtained by omitting loci from the correct locus order of all loci on the chromosome. In this study, we focus on incorrect locus orders for which a ghost marker maps to a chromosome interval, in which none of the underlying duplicate markers is located.

Notation: Consider a chromosome with n distinguishable marker loci at positions $k_1, \ldots, k_n$. The positions are measured in map distance from the beginning of the chromosome, and

$$k_u < k_{u+1} \text{ for } u \in \{1, \ldots, n-1\}.$$

In addition we define the telomere map positions as k_0 and k_{n+1}.

We consider two duplicate markers, located at positions $i_1 < i_2$. The indices of the map positions k_u, which are located next to the duplicate markers and have a smaller map position than these, are denoted with x and y, respectively:

$$x = \max(u | u \in \{0, \ldots, n\}, k_u < i_1)$$
$$y = \max(u | u \in \{0, \ldots, n\}, k_u < i_2)$$

The map position of the ghost marker, resulting from linkage analysis involving the two duplicate loci i_1 and i_2, is denoted by i. The index of the map position k_u located next to the ghost marker and having a smaller map position than it is denoted with z:

$$z = \max(u | u \in \{0, \ldots, n\}, k_u < i)$$

Without loss of generality we assume $z \leq n - z$.

In this notation, a correct locus order is characterized by

$$z = x \text{ or } z = y,$$

whereas an incorrect locus order is characterized by

$$z \neq x \text{ and } z \neq y.$$

Assumptions and basic results: For our derivations, we assume no interference in crossover formation (Stam 1979). Under this assumption, crossover formation in adjacent marker intervals is stochastically independent, and the recombination frequency ρ between two loci is related to the respective map distance d by Haldane's (1919) mapping function

$$\rho = (1 - e^{-2d})/2.$$

Linkage between two loci is measured by the linkage value (Schnell 1961)

$$\lambda = 1 - 2\rho = e^{-2d}. \quad (1)$$

Linkage values between distinguishable markers at positions k_u and k_v are denoted with $\lambda_{u,v}$, those between a distinguishable marker at position k_u and the duplicate markers with $\lambda_{u,i1}$ and $\lambda_{u,i2}$, respectively, and linkage between a distinguishable marker at position k_u and the ghost marker with $\lambda_{u,i}$. For sake of convenience in the subsequent derivations, in which linkage values are summed over marker intervals, we define the linkage between the telomere and the first locus next to the telomere to be zero:

$$\lambda_{0,1} = 0 \quad \text{if } x > 0 \qquad \lambda_{n,n+1} = 0 \quad \text{if } y < n$$
$$\lambda_{0,i1} = 0 \quad \text{if } x = 0 \qquad \lambda_{i2,n+1} = 0 \quad \text{if } y = n$$
$$\lambda_{0,i} = 0 \quad \text{if } z = 0 \qquad \lambda_{i,n+1} = 0 \quad \text{if } z = n.$$

Using the stochastic independence of crossover formation in adjacent marker intervals delimited by the loci at positions $k_u < k_v < k_w$, it can be shown that

$$\lambda_{u,w} = \lambda_{u,v}\lambda_{v,w} \quad (2)$$

and, because $\lambda < 1$,

$$\lambda_{u,w} < \lambda_{u,v}. \quad (3)$$

Another property used in the subsequent derivations is

$$\lambda_{u,v} = \lambda_{v,u}. \quad (4)$$

Linkage between a marker and a ghost marker: In this section we use results of Schnell (1961) and therefore adopt his notation. A haplotype is denoted by a sequence of digits, where each digit corresponds to the origin of the allele at a certain

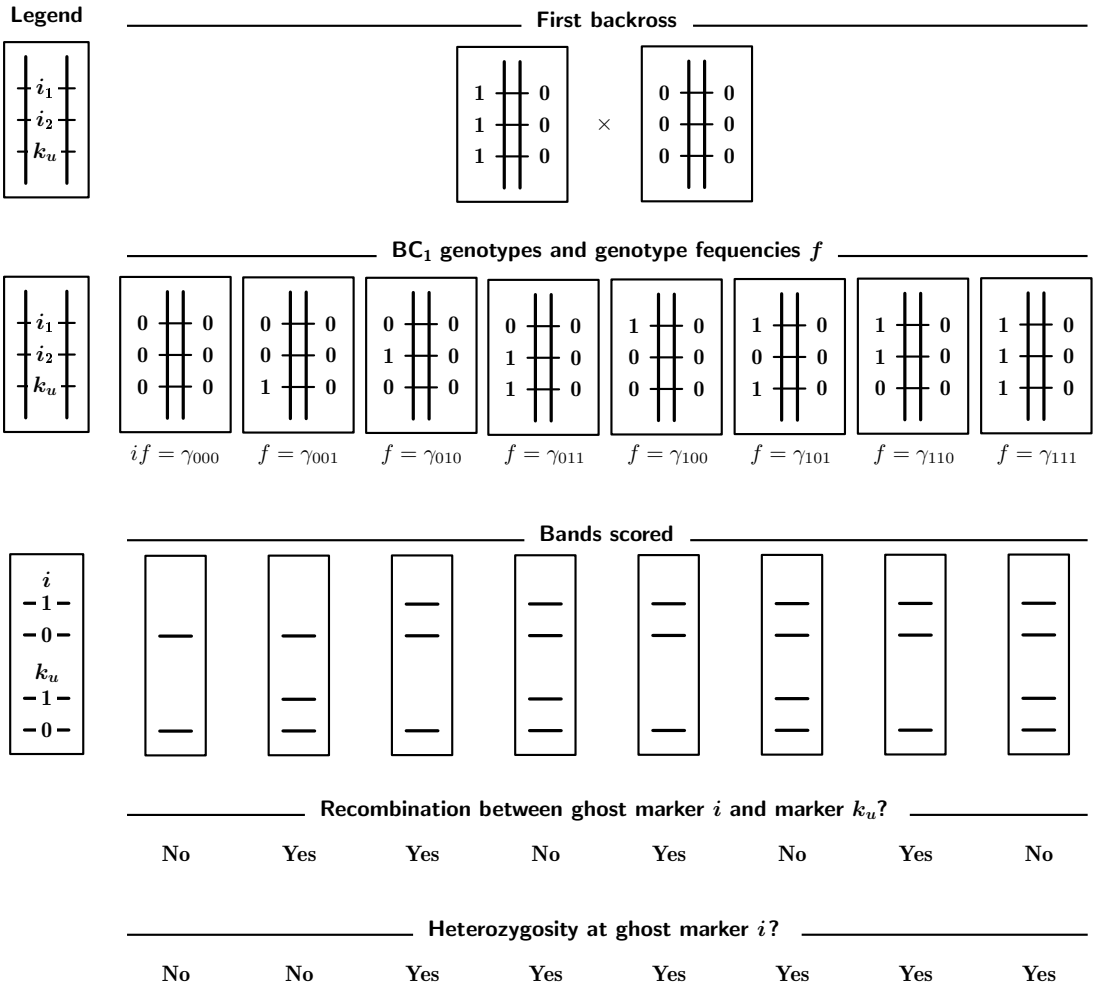

Figure 1 Genotypes of a BC$_1$ population with respect to two duplicate markers i_1, i_2 and a marker k_u. For each multi-locus genotype, its frequency f and the bands scored for ghost marker i and marker k_u are given (assuming codominant inheritance). It is listed for which genotypes recombination between the ghost marker i and marker k_u is observed, and which genotypes are scored as heterozygous with respect to the ghost marker i.

locus. In this sequence, the digits 0 and 1 are used to denote that an allele is of maternal or paternal origin, respectively. The probability that an individual transmits a certain gamete to its progenies is denoted by γ, which is indexed with a sequence of digits describing its haplotype.

We consider three linked loci at positions i_1, i_2, k_u on a chromosome, where the loci at positions i_1 and i_2 are duplicate marker loci. We further assume a BC$_1$ mapping population $\frac{111}{000} \times \frac{000}{000}$ of indefinite population size and codominant markers. Since the alleles at i_1 and i_2 cannot be distinguished, the BC$_1$ genotypes $\frac{001}{000}$, $\frac{010}{000}$, $\frac{100}{000}$, and $\frac{110}{000}$ with respect to i_1, i_2, k_u are scored as recombinant individuals with respect to the ghost marker at position i and the marker at position k_u (Figure 1).

Applying Equation (4) of Schnell (1961) yields the recombination frequency between the ghost marker at position i and the marker at position k_u as

$$\begin{aligned}\rho_{i,u} &= \gamma_{001} + \gamma_{010} + \gamma_{100} + \gamma_{110} \\ &= (1 + \lambda_{i1,i2} - \lambda_{i1,u} - \lambda_{i2,u} \\ &\quad + 1 - \lambda_{i1,i2} + \lambda_{i1,u} - \lambda_{i2,u} \\ &\quad + 1 - \lambda_{i1,i2} - \lambda_{i1,u} + \lambda_{i2,u} \\ &\quad + 1 + \lambda_{i1,i2} - \lambda_{i1,u} - \lambda_{i2,u})/8 \\ &= 1/2 - 1/4\lambda_{i1,u} - 1/4\lambda_{i2,u} \\ &= (\rho_{i1,u} + \rho_{i2,u})/2\end{aligned} \tag{5}$$

and consequently (Equation 1)

$$\lambda_{i,u} = (\lambda_{i1,u} + \lambda_{i2,u})/2. \tag{6}$$

Note that Equation 5 is valid for any linear order of the loci i_1, i_2, k_u on the chromosome, irrespective of the applied mapping function and additional loci on the chromosome (Schnell 1961).

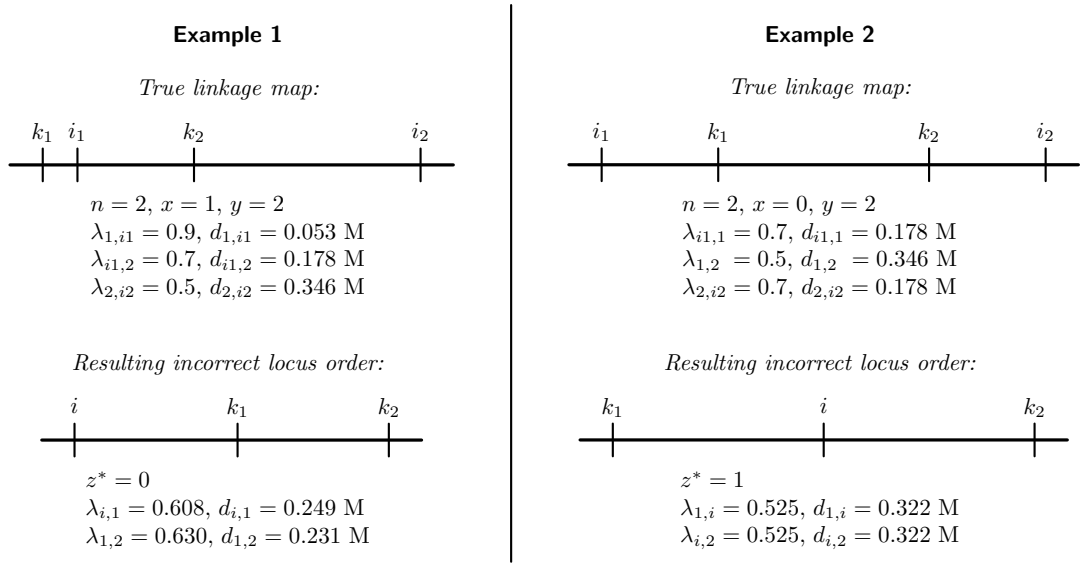

Figure 2 Two examples for incorrect locus orders resulting from duplicate marker loci i_1 and i_2. For description of variable names see text.

The SAR criterion: In the following, we use the 'sum of adjacent recombination frequencies' (SAR) criterion for locus ordering on multi-locus linkage maps and therefore briefly describe its properties. When applying the SAR criterion, ordering of a multi-locus linkage map is done in two steps. First, the pairwise recombination frequencies between all loci of the linkage group are calculated. Second, the locus order is searched, which minimizes the sum of recombination frequencies ρ between adjacent loci on the linkage map. According to Equation 1 this is mathematically equivalent to maximizing the sum of linkage values λ between adjacent loci, briefly referred to as sum of adjacent linkage values. This procedure is based on the proposition that only the correct locus order maximizes the sum of adjacent linkage values on a chromosome. We prove this proposition in the appendix to show that the SAR criterion is a valid method for constructing multi-locus linkage maps.

Because the SAR criterion is only minimized for the correct locus order, the locus order determined on the basis of the SAR criterion must be the same as the one found by any other valid method. Consequently, the results subsequently derived by using the SAR criterion also apply to any other valid locus ordering method.

Incorrect locus orders: The sum of adjacent linkage values for a locus order described by n, x, y, and z is

$$\begin{aligned} L(n,x,y,z) &= \sum_{0\le u<z} \lambda_{u,u+1} + \lambda_{z,i} + \lambda_{i,z+1} + \sum_{z<u\le n} \lambda_{u,u+1} \\ &= \sum_{0\le u<n} \lambda_{u,u+1} - \lambda_{z,z+1} + \lambda_{z,i} + \lambda_{i,z+1}. \end{aligned} \tag{7}$$

Comparing the sum L for two alternative values z' and z'' and omitting equal terms yields

$$\begin{aligned} &L(n,x,y,z') > L(n,x,y,z'') \\ \Leftrightarrow\ &\lambda_{z',i} + \lambda_{i,z'+1} + \lambda_{z'',z''+1} > \lambda_{z'',i} + \lambda_{i,z''+1} + \lambda_{z',z'+1}. \end{aligned} \tag{8}$$

All incorrect locus orders for a combination n, x, y can be described by their value of $z^* \in J = \{0,\dots,n\} \setminus \{x,y\}$. Mapping results in an incorrect locus order characterized by $z^* \in J$ if and only if

$$L(n,x,y,z^*) > L(n,x,y,x) \text{ and} \tag{9}$$

$$L(n,x,y,z^*) > L(n,x,y,y) \text{ and} \tag{10}$$

$$L(n,x,y,z^*) = \max_{z\in J}(L(n,x,y,z)) \tag{11}$$

Proposition (Case 1): In a BC_1 population of infinite size, locus ordering according to the SAR criterion results in an incorrect locus order of type

$$z^* < x < y \quad \text{or} \quad x < z^* < y \quad \text{or} \quad x < y < z^*$$

with $z^* \in J$ if and only if

$$\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{x,x+1} > \lambda_{x,i} + \lambda_{i,x+1} + \lambda_{z^*,z^*+1} \tag{12}$$

and

$$\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{y,y+1} > \lambda_{y,i} + \lambda_{i,y+1} + \lambda_{z^*,z^*+1} \tag{13}$$

and Equation 11 is true.

Proof (Case 1): From Equation 8 follows that

$$\begin{aligned} &L(n,x,y,z^*) > L(n,x,y,x) \\ \Leftrightarrow\ &\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{x,x+1} > \lambda_{x,i} + \lambda_{i,x+1} + \lambda_{z^*,z^*+1} \end{aligned}$$

and

$$\begin{aligned} &L(n,x,y,z^*) > L(n,x,y,y) \\ \Leftrightarrow\ &\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{y,y+1} > \lambda_{y,i} + \lambda_{i,y+1} + \lambda_{z^*,z^*+1} \end{aligned}$$

which completes the proof.

Examples for Case 1: Two scenarios resulting in incorrect locus orders of type $z^* < x$ and $x < z^* < y$ for $n = 2$ are shown in Figure 2.

Proposition (Case 2): In a BC_1 population of infinite size, locus ordering according to the SAR criterion results always in the correct locus order

$$z^* = x = y$$

if the two duplicate marker loci are located in the same marker interval, i.e., $x = y$.

Proof (Case 2): We assume without loss of generality $z^* < x$ and obtain a contradiction. Because of Equation 2

$$\lambda_{z^*,i} = \lambda_{z^*,z^*+1}\lambda_{z^*+1,i} < \lambda_{z^*,z^*+1}$$
$$\lambda_{i,z^*+1} = \lambda_{z^*+1,x}\lambda_{x,i} \leq \lambda_{x,i}$$

and because

$$\lambda_{x,x+1} = \lambda_{x,i1}\lambda_{i1,x+1} < \lambda_{i1,x+1} < \lambda_{i2,x+1}$$

we have

$$\lambda_{x,x+1} < (\lambda_{i1,x+1} + \lambda_{i2,x+1})/2 = \lambda_{i,x+1}.$$

In consequence

$$\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{x,x+1} < \lambda_{z^*,z^*+1} + \lambda_{x,i} + \lambda_{i,x+1}$$

or equivalently (using Equation 8)

$$L(n, x, x, z^*) < L(n, x, x, x),$$

which completes the proof.

Proposition (Case 3): In a BC_1 population of infinite size, locus ordering according to the SAR criterion does not result in an incorrect locus order of type

$$z^* < x \quad \text{or} \quad z^* > y$$

if all markers are equally spaced with linkage value c

$$\lambda_{u,u+1} = c \tag{14}$$

for $u \in \{1, n-1\}$ and $0 < c < 1$.

Proof (Case 3): We assume without loss of generality $z^* < x$ and obtain a contradiction. Because of Equation 2 we have

$$\lambda_{x-1,i1} = \lambda_{x-1,x}\lambda_{x,i1} < \lambda_{x-1,x} < \lambda_{i1,x+1}$$
$$\lambda_{x-1,i2} = \lambda_{x-1,x+1}\lambda_{x+1,i2} < \lambda_{i2,x+1},$$

from which follows (Equation 6)

$$\lambda_{x-1,i} < \lambda_{i,x+1}. \tag{15}$$

Moreover, for any $z < x$,

$$\lambda_{z-1,i} = \lambda_{z-1,z+1}\lambda_{z+1,i} < \lambda_{z+1,i}. \tag{16}$$

From equations 15 and 16 follows (using Equations 4 and 14)

$$\lambda_{x-1,i} + \lambda_{i,x} + \lambda_{x,x+1} < \lambda_{x-1,x} + \lambda_{x,i} + \lambda_{i,x+1} \quad \text{and}$$
$$\lambda_{z-1,i} + \lambda_{i,z} + \lambda_{z,z+1} < \lambda_{z-1,z} + \lambda_{z,i} + \lambda_{i,z+1}$$

or equivalently (using Equation 8)

$$L(n, x, y, x-1) < L(n, x, y, x) \quad \text{and}$$
$$L(n, x, y, z-1) < L(n, x, y, z)$$

from which follows

$$L(n, x, y, z^*) < L(n, x, y, x),$$

which completes the proof.

Proposition (Case 4): In a BC_1 population of infinite size, locus ordering according to the SAR criterion results in a correct locus order

$$z^* = x \quad \text{or} \quad z^* = y$$

if all markers are equally spaced with linkage value c (Equation 14) and the duplicate marker loci are located in the center between their flanking markers

$$\lambda_{x,i1} = \lambda_{i1,x+1} = \lambda_{y,i2} = \lambda_{i2,y+1} = \sqrt{c}. \tag{17}$$

Proof (Case 4): For $z^* < x$ and $y < z^*$ the proof corresponds to the proof for Case 3.

Because of Equation 7

$$\begin{aligned} L(n, x, y, z) - L(n, x, y, z+1) &= \lambda_{z,i} - \lambda_{z+2,i} \\ &= \lambda_{z,i} - \lambda_{z+1,i} + \lambda_{z+1,i} - \lambda_{z+2,i} \end{aligned} \tag{18}$$

and for any $x < z < (y + x)/2$

$$\lambda_{z,i} > \lambda_{z+1,i} \tag{19}$$

because (Equations 6, 17 and 14)

$$\begin{aligned} & \lambda_{z,i1} + \lambda_{z,i2} > \lambda_{z+1,i1} + \lambda_{z+1,i2} \\ \Leftrightarrow\ & \lambda_{i1,x+1}\lambda_{x+1,z} + \lambda_{z,y}\lambda_{y,i2} > \lambda_{i1,x+1}\lambda_{x+1,z+1} + \lambda_{z+1,y}\lambda_{y,i2} \\ \Leftrightarrow\ & c^{z-x-1} + c^{y-z} > c^{z-x} + c^{y-z-1} \\ \Leftrightarrow\ & c^{z-x-1} - c^{z-x} > c^{y-z-1} - c^{y-z} \\ \Leftrightarrow\ & c^{z-x} > c^{y-z} \\ \Leftrightarrow\ & z - x < y - z \\ \Leftrightarrow\ & z < (y + x)/2. \end{aligned}$$

For symmetry reasons, we have for $x + 1 + \delta \leq (x - y)/2$

$$L(n, x, y, x + 1 + \delta) = L(n, x, y, y - \delta), \tag{20}$$

and as a special case

$$L(n, x, y, x + 1) = L(n, x, y, y). \tag{21}$$

From Equations 18, 19 and 20 follows that for $x < z < (y + x)/2$

$$L(n,x,y,z) \geq L(n,x,y,z+1). \quad (22)$$

We now assume without loss of generality $x+1 < z^* < (y+x)/2$ and obtain a contradiction, because from Equations 21 and 22 follows that

$$L(n,x,y,y) = L(n,x,y,x+1) \geq L(n,x,y,z^*), \quad (23)$$

which completes the proof.

A test for detection of duplicate markers: Duplicate markers result in an excess of heterozygotes at the ghost locus in a BC_1 mapping population. Therefore, a statistical test of the null hypothesis that the frequency p of heterozygotes at the locus under consideration is $1/2$ can help to identify them. While approximate χ^2 tests for segregation distortion are common (Weir 1996), we propose to use an exact test based on the binomial distribution because of its superior statistical properties.

Given the null hypothesis H_0: $p = 1/2$ is true, the probability of obtaining more than m heterozygotes in a population of size s can be obtained from the probability function of the binomial distribution as

$$P_0(M > m) = 1 - \sum_{k=0}^{m} \binom{s}{k} 1/2^s.$$

We use the observed number m_b of heterozygotes in a population of size s as a test statistic, and determine the corresponding critical value m^* for testing H_0 for a given Type I error α by solving $P_0(M > m^*) \leq \alpha$. The null hypothesis is rejected if $m_b > m^*$.

The Type II error β of the test depends (a) on the recombination frequency $\rho_{i1,i2}$ between the duplicate loci, which determines the expected frequency of heterozygotes and (b) the size s of the mapping population. For two duplicate markers i_1 and i_2, the genotypes $\frac{10}{00}$, $\frac{01}{00}$, and $\frac{11}{00}$ are scored as heterozygous with respect to the ghost marker i (Figure 1). Hence, the frequency of heterozygotes at the ghost marker can be determined in analogy to Equation 5 from γ_{10}, γ_{01}, and γ_{11} as $p = (1 + \rho_{i1,i2})/2$. Consequently, under the alternative hypothesis

$$H_{A,\rho}\colon p = \frac{1+\rho_{i1,i2}}{2},$$

the probability of obtaining m heterozygotes is

$$P_{A,\rho}(M = m) = \binom{s}{m} \left(\frac{1+\rho_{i1,i2}}{2}\right)^m \left(\frac{1-\rho_{i1,i2}}{2}\right)^{s-m}$$

and the power $1-\beta$ of the test can be obtained from

$$\beta = \sum_{k=0}^{m^*} P_{A,\rho}(M = k).$$

Numerical results

Occurrence of incorrect locus orders: In the previous section, we derived for the general case of n markers the conditions under which incorrect locus orders occur, if there are duplicate marker loci on a chromosome. Here, we illustrate the typical properties of situations for which incorrect locus orders occur with numerical examples for the four locus case.

The locus order k_1, i_1, i_2, k_2 is characterized by $n = 2$ and $x = y$. As shown in the theory section (Case 2), no incorrect locus order can occur for this situation.

The locus order k_1, i_1, k_2, i_2 is characterized by $n = 2$, $x = 1$, $y = 2$. For these parameters the locus orders k_1, i, k_2 ($z^* = x = 1$) and k_1, k_2, i ($z^* = y = 2$) are correct. Simple combinatorical considerations show that the only incorrect locus order is i, k_1, k_2 ($z^* = x - 1 = 0$). Applying Equations 12 and 13 yields the conditions under which this incorrect locus order occurs:

$$\lambda_{1,i1}\lambda_{i1,2} > (\lambda_{i1,2} + \lambda_{2,i2})/2 \quad \text{and} \quad (24)$$

$$\lambda_{1,i1} + \lambda_{1,i1}\lambda_{i1,2}\lambda_{2,i2} > \lambda_{i1,2} + \lambda_{2,i2}. \quad (25)$$

Setting $\lambda_{1,i1} = 0.60, 0.80, 0.90, 0.99$ and solving both inequalities for $\lambda_{i1,2}$ results in the graphs in Figure 3, showing combinations of linkage values $\lambda_{1,i1}$, $\lambda_{i1,2}$, $\lambda_{2,i2}$ for which incorrect locus orders do occur. A prerequisite for an incorrect map order is that $\lambda_{i1,2}$ is greater than $\lambda_{2,i2}$. The set of parameter combinations, for which mapping results in an incorrect locus order, increases with increasing linkage between k_1 and i_1: For $\lambda_{1,i1} < 0.5$ only correct map orders are found, as can be seen from Equation 24 using simple arithmetics. In contrast, for very tight linkage ($\lambda_{1,i1} = 0.99$) incorrect map orders occur for a broad range of parameter settings, including $0.8 > \lambda_{i1,2} > \lambda_{2,i2}$. Summarizing, incorrect map orders occur if (a) k_1 and i_1 are tightly linked and (b) linkage between i_1 and k_2 is greater than linkage between k_2 and i_2 (Figure 3).

The locus order i_1, k_1, k_2, i_2 is characterized by $n = 2$, $x = 0$, and $y = 2$. For these parameters, correct locus orders are i, k_1, k_2 ($z^* = x = 0$) and k_1, k_2, i ($z^* = y = 2$). The only incorrect locus order is k_1, i, k_2 ($z^* = x + 1 = 1$). Applying Equations 12 and 13 yields the conditions under which this incorrect locus order occurs:

$$\left(\lambda_{i1,1}\lambda_{1,2} + \lambda_{i2,2}\right)/2 > \lambda_{1,2} \quad \text{and} \quad (26)$$

$$\left(\lambda_{i1,1} + \lambda_{1,2}\lambda_{i2,2}\right)/2 > \lambda_{1,2}. \quad (27)$$

Setting $\lambda_{1,2} = 0.1, 0.5$ and solving both inequalities for $\lambda_{2,i2}$ results in the graphs in Figure 4, showing combinations of linkage values $\lambda_{i1,1}$, $\lambda_{1,2}$, $\lambda_{2,i2}$, for which incorrect locus orders do occur. For loose linkage between k_1 and k_2 ($\lambda_{1,2} = 0.1$), the set of parameter combinations resulting in incorrect locus orders is quite large, a prerequisite is that neither $\lambda_{i1,1}$ nor $\lambda_{2,i2}$ is smaller than 0.2. With increasing linkage of k_1 and k_2, the set of parameter combinations decreases, for which incorrect locus orders do occur. However, even for very tight linkage, incorrect map orders do occur, if the linkage values $\lambda_{i1,1}$ and $\lambda_{2,i2}$ are large and have approximately the same value. Summarizing, incorrect locus orders occur for loose linkage of k_1 and k_2, when linkage between i_1 and k_1 as well as between k_2 and i_2 is almost equal and each value is at least twice as large as linkage between k_1 and k_2 (Figure 4).

Power of detecting duplicate loci: Testing for segregation distortion is important to detect duplicate marker loci and, hence, avoid inappropriate application of incorrect linkage

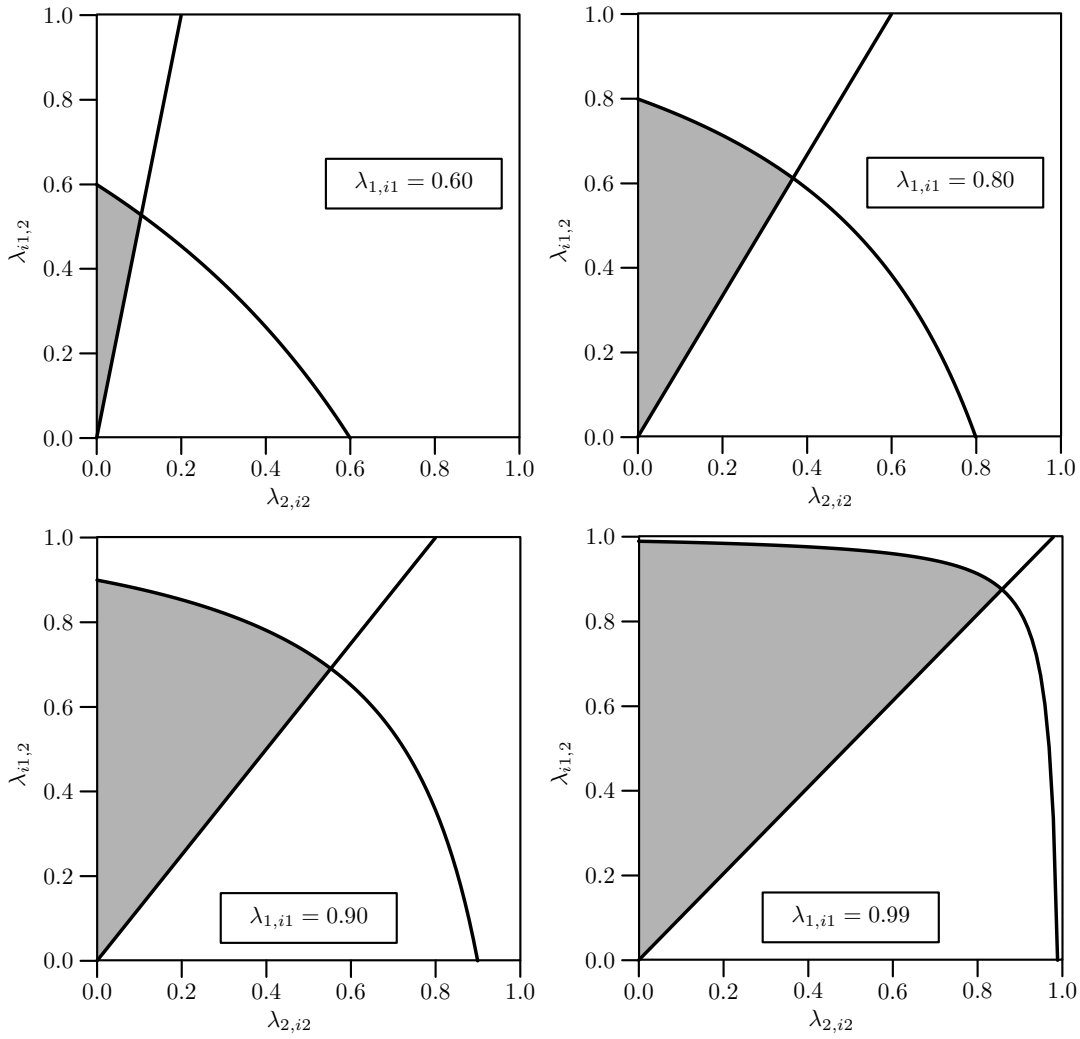

Figure 3 Occurrence of incorrect locus orders for the true locus order k_1, i_1, k_2, i_2. The lines are obtained by using $\lambda_{1,i1} = 0.60, 0.80, 0.90, 0.99$ in Equations 24 and 25 and solving for $\lambda_{2,i2}$. The shaded areas indicate parameter combinations of $\lambda_{1,i1}$, $\lambda_{i1,2}$, and $\lambda_{2,i2}$ for which incorrect locus orders do occur.

maps. Here, we investigate the power of the exact test for segregation distortion in a BC_1 population, depending on the size of the mapping population and the linkage value between the duplicate marker loci.

For a Type I error $\alpha = 0.05$ of incorrectly assuming the presence of a ghost marker, using a mapping population of size $s = 50$, the power of detecting a ghost marker is only greater than 0.9, if the linkage between the duplicate loci is greater than 0.2 (which corresponds to a map distance of approximately 50 cM) (Figure 5). Mapping populations with size $s = 500$ or even 1000 are required to detect with a high probability ghost loci, resulting from duplicate markers with linkage values between 0.8 and 1.0 (corresponding to map distances of about 10 and 0 cM).

For a smaller Type I error $\alpha = 0.001$, the minimum population size required to detect duplicate loci with a high probability is $s = 100$, if the linkage value is lower than 0.2 (Figure 5). For linkage values greater 0.8, populations larger than $s = 1000$ individuals are required.

Discussion

Genetic model: For our derivations we used the assumption of no interference (Stam 1979) underlying Haldane's (1919) mapping function. This is a simplified mathematical model and there exist more sophisticated models of crossover formation in meiosis, which fit experimental data better (McPeek and Speed 1995). Briefly, the assumption of no interference has (1) the advantage of mathematical simplicity, yielding equations, which can be easily evaluated and (2) the results can be applied without knowing the exact amount of interference in the chromosome region under consideration. For a more detailed discussion concerning the use of the assumption of no interference see Frisch and Melchinger (2001). Note that Equation 6, defining linkage between a ghost locus and a distinguishable marker, holds true for arbitrary mapping functions. However, the results for locus ordering may be affected when applying a mapping function different from Haldane's.

The definition of an incorrect locus order in the theory sec-

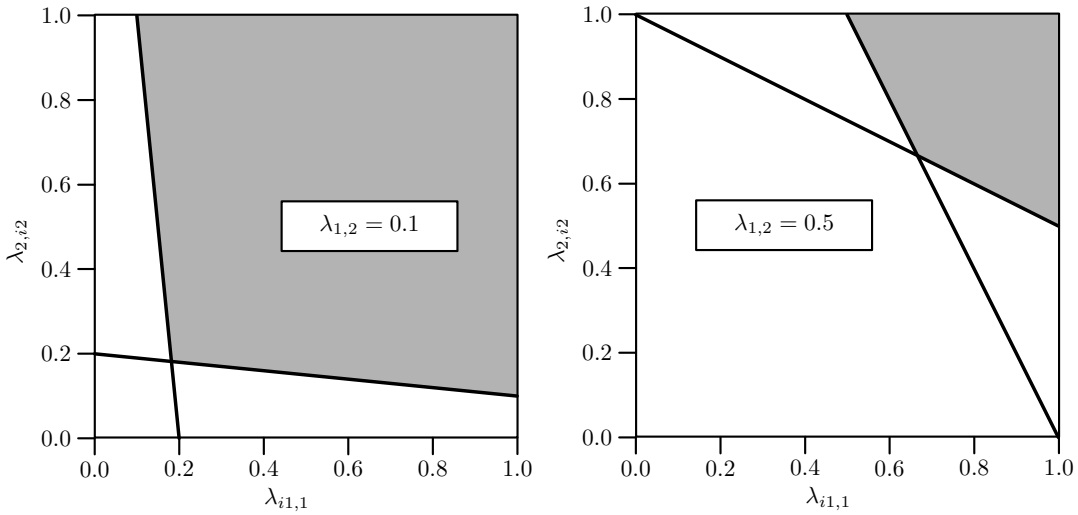

Figure 4 Occurence of incorrect locus orders for the true locus order i_1, k_1, k_2, i_2. The lines are obtained by using $\lambda_{1,2} = 0.1, 0.5$ in Equations 26 and 27 and solving for $\lambda_{2,i2}$. The shaded areas indicate parameter combinations of $\lambda_{i1,1}$, $\lambda_{1,2}$, and $\lambda_{2,i2}$ for which incorrect locus orders do occur.

tion considers a locus order as correct if the ghost marker maps to one of the two intervals in which the duplicate markers are located. This is appropriate for two situations: (a) The marker itself is part of the gene (e.g., RGA or EST markers), and the target gene is duplicate, or (b) the marker is tightly linked to a target gene and the complete region containing marker and target gene is duplicate.

In contrast, if only the marker is duplicate, but not the target gene, and only one of the two duplicate marker loci is tightly linked with the target gene, then this definition of incorrect locus orders is not appropriate. In such a case, linkage analysis does not identify the ghost marker as being tightly linked to the target gene, because recombination between ghost marker and a target is the mean recombination frequency between the target and the two duplicate loci (Equation 5). This situation may also negatively affect construction of linkage maps, but is not the subject of the present study.

Ghost QTL and ghost markers: Ghost QTL and ghost markers share the properties that (i) biometrical analysis maps a locus to an incorrect position on the linkage map, and (ii) this is caused by the fact that not a single locus but two indistinguishable loci are underlying the observed differences between individuals.

However, there are also fundamental differences between the two phenomena: (1) Ghost markers occur in the initial construction of a linkage map, whereas ghost QTL are detected in QTL analysis conducted after having a linkage map available. (2) Ghost markers can map outside the interval of the duplicated markers, whereas ghost QTL are located between the underlying QTL. (3) Ghost markers result from duplicated DNA sequences, whereas ghost QTL may occur from two loci having entirely different DNA sequences but affecting the same phenotypic trait. Summarizing, the ghost marker phenomenon has similarities to the ghost QTL phenomenon, but from the differences mentioned above, the implications for practical applications are different.

Segregation distortion caused by zygotic selection: If segregation distortion is detected at a marker locus, this may not only be due to duplicate markers, but also due to various other reasons, one of which is zygotic section. In this case, an excess of heterozygotes follows from a reduced fitness of homozygotes. To distinguish both situations, the following considerations can be made: For duplicate markers, segregation distortion occurs only at the ghost locus. In contrast, for zygotic selection, segregation distortion occurs not only at the locus which is affected by selection, but also at closely linked loci.

This is illustrated by a numerical example: Consider a 2 M chromosome, carrying 21 equally spaced markers and a BC_1 mapping population consisting of $s = 100$ individuals. If two duplicate marker loci are located at map positions 0.87 and 1.13 ($\lambda_{i1,i2} = 0.6$), then the test for segregation distortion ($\alpha = 0.05$) detects segregation distortion with a probability $1 - \beta \approx 0.7$ at map position 1.0 (Figure 6). At all other loci, segregation distortion is only detected with the probability of the Type I error $\alpha = 0.05$. If there is zygotic selection at the locus at position 1.0 such that from the homozygotes only 50% survive, then with a comparable probability of about 0.7 segregation distortion is detected by the test at map position 1.0. However, in this case also linked markers adjacent to the locus at map position 1.0 display segregation distortion with a high probability (Figure 6).

Consequently, if segregation distortion is detected only at one locus, chances are high that duplicate markers are the reason, whereas if segregation distortion is detected at several closely linked loci, this can be taken as an indicator for zygotic selection.

Effects of sampling and locus ordering method: The proofs in our theoretical investigation assume an indefinite population size resulting in exact linkage values λ. However, in a mapping study, linkage is estimated from a finite sample of a population, and a considerable estimation error may occur depending on the sample size. Therefore, we used a simulation study to investigate whether the theoretically expected results for indefinite populations, known linkage values, and the SAR criterion are obtained when applying different mapping programs to finite populations.

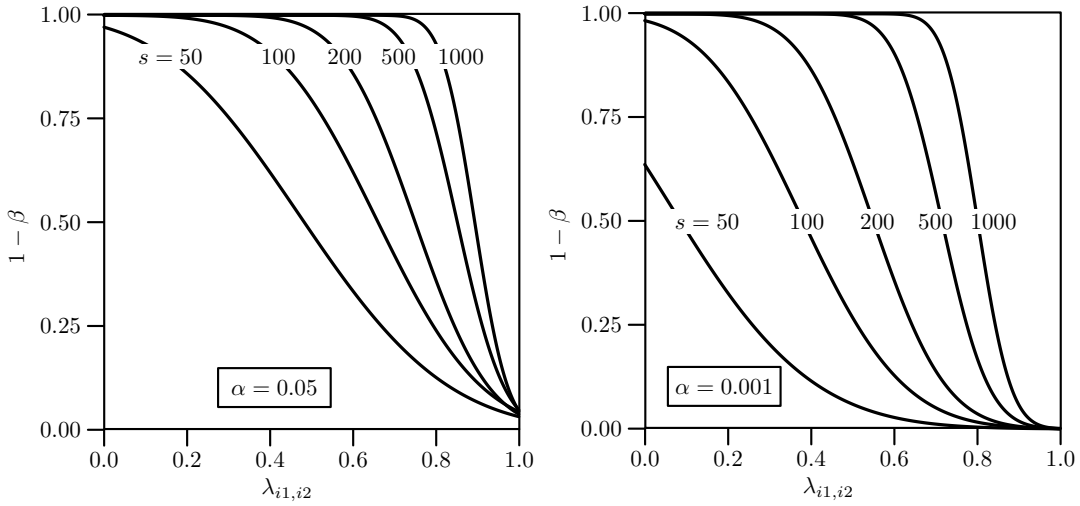

Figure 5 Power $1-\beta$ of the exact test for segregation distortion in a BC_1 population for Type I errors $\alpha = 0.05$ and 0.001, depending on population size s and linkage $\lambda_{i1,i2}$ between two duplicate marker loci.

The simulation program Plabsim (Frisch et al. 2000) was used to generate the datasets and the mapping programs GMendel (Liu and Knapp 1990, Holloway and Knapp 1993), Mapmaker (Lander et al. 1987), and Joinmap (Stam 1993) were applied for linkage analysis. The GMendel software performs locus ordering with the SAR criterion, while Joinmap uses a modification of the SAR criterion and Mapmaker applies a maximum likelihood approach.

We investigated a chromosome with $(k_1, i_1, k_2, i_2) = (0.0, 0.1, 0.3, 1.0)$ and generated with Plabsim 100 BC_1 populations $\left(\frac{111}{111} \times \frac{000}{000}\right) \times \frac{000}{000}$ for each population size $s = 50, 100, 250, 100, 5000$. The populations were evaluated for the genotype at loci k_1 and k_2. The genotypes $\frac{10}{00}$, $\frac{01}{00}$, and $\frac{11}{00}$ with respect to loci i_1 and i_2 were scored as $\frac{1}{0}$ with respect to the ghost marker i, $\frac{00}{00}$ was scored as $\frac{0}{0}$.

The resulting datasets were analyzed with GMendel. For a population size of $s = 50$, the incorrect locus order i, k_1, k_2 was found in 23% (Table 1). With increasing population size, the percentage of incorrect maps increased and reached 93% for $s = 5000$. In consequence, for small populations the estimation error of the recombination frequencies resulted in that a correct locus order or no linkage at all was found quite frequently. However, for large populations the incorrect locus order was observed in most cases, as expected from theory.

One population of size $n = 5000$ was analyzed with GMendel, Joinmap, and Mapmaker. All three programs yielded the incorrect locus order i, k_1, k_2. The programs GMendel and Mapmaker estimated the map distances $(\hat{d}_{i,1}, \hat{d}_{1,2}) = (0.372, 0.297)$, which were close to those expected from Equation 6 $(\hat{d}_{i,1}, \hat{d}_{1,2}) = (0.370, 0.300)$. However, Joinmap estimated $(\hat{d}_{i,1}, \hat{d}_{1,2}) = (0.289, 0.508)$, which is surprising because according to theory, the occurrence of duplicate markers should not influence the map distance $\hat{d}_{1,2}$ between distinguishable markers k_1 and k_2. Consequently, the incorrect locus order was observed irrespective of the locus ordering method implemented in these programs, as expected from theoretical considerations.

Type of the mapping population: Throughout our study we focused on BC_1 mapping populations, because determining the fractions of recombinant gametes and heterozygotes is simple in BC_1 (Figure 1). However, all results obtained for locus ordering using the SAR criterion depend only on the known recombination frequencies between loci. How and from which type of population they are obtained, is irrelevant for the derivations. For a different type of mapping population, e.g., an F_2 population, the procedure of obtaining recombination frequencies between loci differs, but the locus ordering procedure based on the SAR criterion does not. Therefore, the presented results are valid for any type of mapping population in which heterozygous individuals occur.

In recombinant inbred lines or doubled haploids, both homologues of a chromosome are identical copies and, therefore, all loci are homozygous. However, two duplicate marker loci may carry two different alleles and therefore will be scored as heterozygous with respect to the ghost marker. In consequence, markers which are scored as heterozygous these two types of mapping populations may be ghost markers.

Implications for applying linkage maps The incorrect map position of a ghost marker affects application of linkage maps, for which not only tight linkage of a target gene and adjacent markers is required, but the correct location of the target gene with respect to flanking markers is important. Examples are map-based cloning, marker-assisted backcrossing, and marker-assisted selection.

In map-based cloning, the chromosome region where a gene is located, is first determined with a low-density linkage map. Then, the region of the target gene is analyzed with a marker density higher than 1 marker per cM, in order to finemap the gene and to locate a marker interval, which will be used for genomic library screening.

In the fine mapping step, no problems from incorrect locus orders are expected. First, incorrect locus orders do not occur for equal marker spacing (Proposition 4); second, numerical evaluation of Equations 12 to 13 shows, that for marker distances smaller than 1 cM always the correct locus order is found. However, in the first stage with low density linkage maps, duplicate markers can result in mapping the target

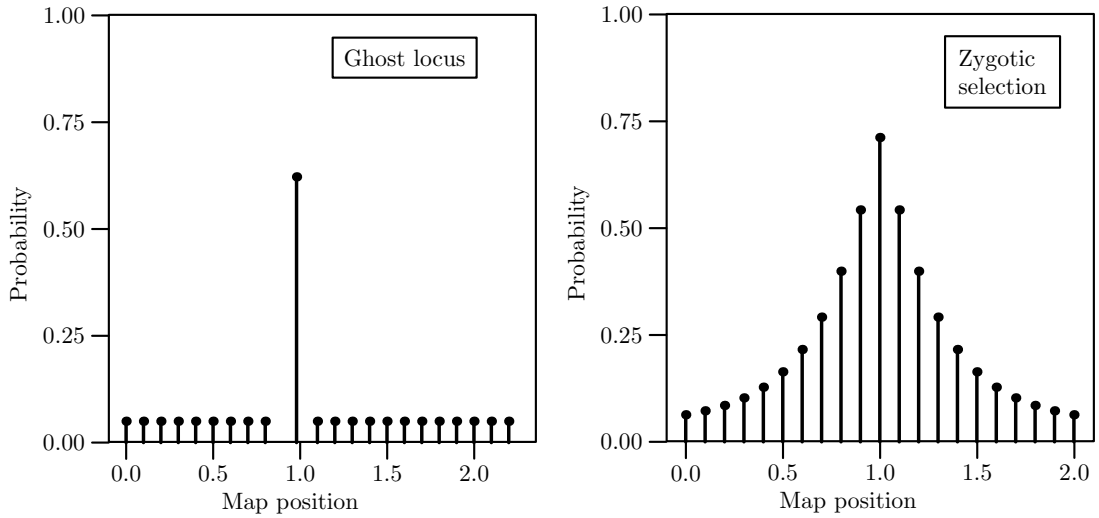

Figure 6 Probability of detecting segregation distortion ($\alpha = 0.05$) at loci equally distributed on a 2 M chromosome with a BC_1 population of size $s = 100$. Left diagram: Duplicate loci at map positions 0.87 and 1.13 ($\lambda_{i1,i2} = 0.6$). Right diagram: Zygotic selection with a survival rate of homozygotes of 0.5.

gene into an incorrect chromosome region, such that none of the high-resolution markers investigated in the second step are tightly linked to the target gene.

For marker-assisted selection, a QTL is mapped to a chromosome interval, subsequently the markers flanking the chromosome interval are used for indirect selection for the presence of the favorable allele at the QTL. Estimated locations of the QTL are usually not precise point estimates, but the QTL is assumed to be located in a so-called support interval, which often covers large chromosome segments up to 90 cM (Visscher et al. 1996).

If a chromosome region is duplicate, which contains a marker and a QTL, then tight linkage between marker and QTL is detected, irrespective of the duplication. Because of the large marker distances required to select for a QTL in a support interval, marker and QTL may map into an incorrect flanking marker interval. Selection for the markers incorrectly assumed to be flanking the target region may not be an indirect selection for the chromosome region which carries the QTL. This can greatly reduce the efficiency of marker-assisted selection.

In marker-assisted backcrossing for introgression of a target gene from a donor parent into the genetic background of a recipient parent, markers can be used for two purposes: (a) to select for the presence of a tightly linked target gene (foreground selection) and (b) to select against the genetic background of the recipient parent (background selection) (Tanksley et al. 1989). Marker-assisted backcrossing is routinely applied, e.g., in maize breeding to introgress transgenes in inbred lines used for production of commercial hybrids.

For foreground selection, linkage between marker and target gene needs to be very tight. If a chromosome region is duplicate, which contains the target gene and a tightly linked marker, no negative effects of the duplication with respect to foreground selection are expected.

In marker-assisted background selection, a primary goal is to reduce the length of the donor chromosome segment around the target gene (Stam and Zeven 1981, Young and Tanksley 1989, Frisch et al. 1999). This is achieved by selecting for the allele of the recipient at markers flanking the target gene. In backcross programs, population size is usually restricted by the reproduction coefficient of the species and practical constraints. In order to observe recombination between the target gene and the flanking markers with a high probability in a finite population, linkage between the marker and target gene should not be extremely tight (Frisch et al. 1999). This implies the use of more distant marker brackets for background selection, which can result in incorrect locus orders.

In consequence, if the target gene maps into an incorrect chromosome interval, selection for markers incorrectly assumed to flank the target gene does not reduce the donor chromosome segment attached to the target gene. This can greatly reduce the efficiency of fast recovery of the recurrent parent genome.

Conclusions and further research needs: Pointing out the extent of duplicated sequences as well as the evolutionary formation of large gene families by duplication events in eukaryotic genomes, the ghost marker phenomenon was presumably overlooked so far. The existence of ghost markers in linkage maps is very likely and many of them remain undetected because of very close linkage of the underlying duplicate loci and the insufficient size of mapping populations. Furthermore, the ghost marker phenomenon has the potential to provide new explanations for distorted segregation at numerous loci in existing and emerging linkage maps.

The application of duplicated sequences as molecular markers is not restricted to gene-derived markers like ESTs or RGAs. Single bands of several other molecular marker types like AFLPs and SSRs are also known to frequently represent multiple sequences resulting in the same complications for the construction of linkage maps. Because correct linkage maps are essential for important applications in genetics and breeding, many interesting questions concerning this subject warrant further research: What are the consequences, if only the marker but not the target gene is duplicate? How do duplicate markers affect map distances between observable markers in multipoint estimation of recombination frequencies? Can the presented

Table 1 Locus orders resulting from applying GMendel to simulated BC_1 datasets of size $s = 50, 100, 250, 1000, 5000$. The underlying linkage map was $(k_1, i_1, k_2, i_2) = (0, 0.1, 0.3, 1.0)$.

Estimated locus order	Population size s				
	50	100	250	1000	5000
	[%]				
no linkage	55	6	0	0	0
k_1, k_2, i	17	38	38	38	7
k_1, i, k_2	5	7	3	0	0
i, k_1, k_2	23	49	59	62	93

approach be extended to more sophisticated crossover formation models? How does the mode of inheritance (codominance vs. dominance) affect the ghost marker phenomenon? Furthermore, besides the implications of ghost markers on map-based cloning, MAS, and MAB, it is also necessary to investigate their influence on known discrepancies between genetic and physical maps regarding locus order and contig assembly.

References

Acarkan A, Rossberg M, Koch M, Schmidt R (2000) Comparative genome analysis revealed extensive conservation of genome organziation among Arabidopsis thaliana and Capsella rubella. Plant J 23:233–243

Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, van Staveren M, Stiekema W (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. Nature 391:485–488

Blanc G, Barakat A, Guyot R, Cook R, Delseny M (2000) Extensive duplication and reshuffling in the Arabidopsis genome. Plant Cell 12:1093–1101

Frisch M, Bohn M, Melchinger AE (1999) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975 (erratum: Crop Sci 39:1913)

Frisch M, Bohn M, Melchinger AE (2000) Plabsim: Software for simulation of marker-assisted backcrossing. J Heredity 91:86–87

Frisch M, Melchinger AE (2001) The length of the intact chromosome segment around a target gene in marker-assisted backcrossing. Genetics 157:1343–1356

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92–100

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linkage factors. J Genet 8:299–309

Holloway JL, Knapp SJ (1993) GMendel 3.0: Software for the analysis of genetic markers and maps. Oregon State University, Corvallis.

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

Liu BH, Knapp SJ (1990) GMendel: A program for Mendelian segregation and linkage analysis of individual or multiple progeny populations using likelihood ratios. J Hered 81:407

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

McPeek MS, Speed TP (1995) Modeling interference in genetic recombination. Genetics 139:1031–1044

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, Heidelberg

Quint M, Dußle CM, Melchinger AE, Lübberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping, and MAS. Theor Appl Genet (in press)

Schnell FW (1961) Some general formulations of linkage effects in inbreeding. Genetics 46:947–957

Stam P (1979) Interference in genetic crossing over and chromosome mapping. Genetics 92:573–594

Stam P (1993) Construction of integrated genetic linkage maps by means of na new computer package: Joinmap. Plant J 3:739–744

Stam P, Zeven AC (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227–238

Tanksley SD, Young ND, Patterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257-263

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A. (2000) The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lubbersted T, Xia XC, Stam P, Zabeau M, Kuiper M. (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

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

Weir B (1996) Genetic data analysis. Sinauer Associates, Sunderland, Massachussets, p 27.

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–49

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

Appendix

To show that the SAR criterion is suitable for locus ordering we consider a chromosome consisting of n loci at positions $k_1, \ldots, k_n$ and prove by mathematical induction:

Proposition (SAR): The SAR for the correct locus order is smaller than the SAR for any other order of the n loci on the chromosome.

Induction basis: From Equation 1 follows that the locus order with a minimum SAR has a maximum sum of adjacent linkage values. For a subset of three loci at positions $k_{u1} < k_{u2} < k_{u3}$ with $u_1, u_2, u_3 \in \{1, \ldots, n\}$ the correct locus order has the greatest sum of adjacent linkage values because

$$\begin{aligned} & L(k_{u1}, k_{u2}, k_{u3}) > L(k_{u2}, k_{u1}, k_{u3}) \\ \Leftrightarrow\ & \lambda_{u1,u2} + \lambda_{u2,u3} > \lambda_{u2,u1} + \lambda_{u1,u3} \\ \Leftrightarrow\ & \lambda_{u2,u3} > \lambda_{u1,u2}\lambda_{u2,u3} \end{aligned}$$

and

$$\begin{aligned} & L(k_{u1}, k_{u2}, k_{u3}) > L(k_{u2}, k_{u3}, k_{u1}) \\ \Leftrightarrow\ & \lambda_{u1,u2} + \lambda_{u2,u3} > \lambda_{u2,u3} + \lambda_{u3,u1} \\ \Leftrightarrow\ & \lambda_{u1,u2} > \lambda_{u1,u2}\lambda_{u2,u3} \end{aligned}$$

Induction hypothesis: Consider a subset of $3 < m < n$ loci at map positions $k_{u1} < \ldots < k_{um}$ with $u_1, \ldots, u_m \in \{1, \ldots, n\}$ and denote $U = \{u_1, \ldots, u_m\}$. We assume that for the subset of m loci, the correct locus order has the largest sum of adjacent linkage values.

Induction step: Consider an $m+1$th locus at map position k_v ($v \in \{1, \ldots, n\}$ and $v \notin U$). We further define

$$w = \max(s | s \in U, k_s < k_v).$$

The correct locus order for the $m+1$ loci has the sum of adjacent linkage values

$$L = \sum_{\mu \in U, \mu < w} \lambda_{\mu,\mu+1} + \lambda_{w,v} + \lambda_{v,w+1} + \sum_{\mu \in U, w < \mu} \lambda_{\mu,\mu+1},$$

whereas an incorrect locus order has the sum of adjacent linkage values

$$L^* = \sum_{\mu \in U, \mu < w-\varepsilon} \lambda_{\mu,\mu+1} + \lambda_{w-\varepsilon,v} + \lambda_{v,w-\varepsilon+1} + \sum_{\mu \in U, w-\varepsilon < \mu} \lambda_{\mu,\mu+1}$$

where $w - \varepsilon \in U$ and without loss of generality $0 < \varepsilon$.

Omitting equal terms in L and L^* yields

$$\begin{aligned} & L^* < L \\ \Leftrightarrow\ & \lambda_{w-\varepsilon,v} + \lambda_{v,w-\varepsilon+1} + \lambda_{w,w+1} < \lambda_{w-\varepsilon,w-\varepsilon+1} + \lambda_{w,v} + \lambda_{v,w+1}, \end{aligned}$$

which can be shown to be true using the induction assumption and Equation 3:

$$\begin{aligned} \lambda_{w-\varepsilon,v} &= \lambda_{w-\varepsilon,w-\varepsilon+1}\lambda_{w-\varepsilon+1,v} < \lambda_{w-\varepsilon,w-\varepsilon+1} \\ \lambda_{v,w-\varepsilon+1} &= \lambda_{w-\varepsilon+1,w}\lambda_{w,v} < \lambda_{w,v} \\ \lambda_{w,w+1} &= \lambda_{w,v}\lambda_{v,w+1} < \lambda_{v,w+1}. \end{aligned}$$

Hence, if the SAR criterion is valid for m loci, it is also valid for $m+1$ loci. The induction basis shows that the SAR criterion is valid for $m = 3$, which completes the proof.

Note that the SAR criterion is used e.g., in software GMendel (Liu and Knapp 1990, Holloway and Knapp 1993), which is widely used by plant geneticists. However, to our knowledge a proof of its optimum properties has not yet been published. Moreover, the above proof is constructive in that it describes an algorithm which finds the correct locus order for n loci with $n^2/2 - 4$ comparisions. To our knowledge, available mapping software applies either a search of the entire parameter space, which requires $n!/2$ comparisons to find the correct locus order, or it applies numerical optimization methods, which may converge only to local maxima of the SAR instead of the global maximum.

7 General discussion

Genomic organisation of R genes – R gene clusters

Our studies on pic19 as well as several other investigations on the model organism *Arabidopsis thaliana*, tomato, lettuce, and other crops (Martin et al. 1994; Thomas et al. 1997; Meyers et al. 1998; Meyers et al. 1999) demonstrated clustering of RGAs within short chromosomal stretches. Although the Arabidopsis Genome Initiative (2000) identified 46 singletons among the 150 Arabidopsis R genes, the majority of plant disease R genes seems to be organised in R gene clusters. Of the non-singletons, ~60% of R gene pairs are ordered in direct repeats, and ~40% in inverted repeats. The number of genes within known R gene clusters varies greatly from a minimum of two genes for the tomato *Cf2*/*Cf5* cluster (Jones et al. 1993; Dixon et al. 1996; Hammond-Kosack and Jones 1997) to >24 genes at the lettuce *Dm3* locus (Meyers et al. 1998). Analogously, the physical genome distance spanned by the clusters varies from several kbp to >4 Mbp for the *Dm3* R gene cluster.

Hybridisation of the B73 maize BAC library identified five contigs of independent pic19 gene family members. These were located in two clusters on maize chromosomes 1 (pic19-4,-5) and 6 (pic19-1,-2,-3). The minimum size of the region covered by the three contigs on chromosome 6 estimated by the length of restriction fragments was 450 kb making one RGA in 150 kb. However, since the three pic19 contigs are not overlapping with each other, these values are probably even larger. Other known clusters are mostly smaller like the 230 kb rice *Xa21* cluster with eight RGAs (Song et al. 1997), the tomato *Cf4* and *Cf9* clusters both with five homologues spread over 36 kb (Parniske et al. 1997), or nine *RPP5*-related sequences in *Arabidopsis* spanning a distance of 70 kb (The EU Arabidopsis Genome Project 1998). Generally, members of R gene clusters are spaced much closer together (7 to 70 kb apart). Therefore, the pic19 multilocus on chromosome 6S spans a comparable large physical distance. Only the complex multigene family at the lettuce *Dm3* region reveals a comparable spacing with an average space of 145 kb (Meyers et al. 1998). On the one hand, this increased spacing for the pic19 gene family could be a reflection of the large genome size of maize compared to these other species. On the other hand, there is so far no sequence information

about the pic19 BAC contigs available and there is a high probability for the existence of further RGAs which could be identified, if less stringent screening conditions would be employed.

Evolution of R genes

The picture emerging for the organisation and evolution of plant R genes and RGA sequences is similar to that of the vertebrate major histocompatibility complex (MHC), T-cell receptor, and immunoglobulin genes (Michelmore and Meyers 1998). Therefore, the evolutionary forces shaping the plant and vertebrate defence systems may be similar, although the specific types of genes involved are different.

Duplication obviously is the initial step in creating such complex arrays of R gene loci. Duplications create new loci, alter the number of gene family members, or generate repeated sequences within a gene. In the evolution of pic19 homologues, duplication is likely involved in different ways. Since genomic regions containing large amounts of repeated DNA sequences enhance duplication through unequal crossing-over via mispairing between different genes, this mechanism is likely to be the driving force of recombination in the *Scmv1* region on chromosome 6, closely linked to the highly repetitive *nor*. Assuming either chromosome 6 or chromosome 1 to be the founder locus of pic19, duplication created a new locus on a different chromosome. Subsequently, duplication or unequal crossing-over altered the number of gene family members on both chromosomes. Alternatively, and analogous to the formation of the *Hcr9s* gene family in tomato containing *Cf-9* homologous genes (Parniske et al. 1999), the first step might as well have been a local duplication creating a cluster on one of the two chromosomes followed by duplication of the whole cluster resulting in a duplicated segment on the other chromosome. However, our data do not give indications for any of the two possibilities. In addition to shuffling of large genomic regions, single nucleotide changes can lead to fine structural changes within the genes. The high sequence similarity between the pic19 homologues indicates that the initial duplication event happened recently in an evolutionary perspective. Another explanation might be a high selection pressure resulting in conservation of similar gene copies. This high degree of sequence

similarity might result in frequent unequal crossing-over leading to further duplications and deletions. Therefore, the pic19 gene clusters are yet very likely to be genetically unstable.

The organisation of R genes in such large arrays of R gene clusters provides a number of opportunities for plants to face the challenge of fast pathogen evolution. Since pathogens are under selection pressure to prevent the expression of avirulence factors and gain access to the host plant, there are basically two different ways in that clustering of R genes supports evolution of novel plant disease resistances: (1) provision of building blocks for rapidly evolving R genes with new recognition capabilities by recombination, gene conversion, unequal crossing-over, etc. and (2) reservoirs of active and inactive genes with unique recognition specificities. Not all genes may be active. Function of inactive pseudogenes may be restored and recycled by infrequent recombination between active and inactive copies or gene conversion with functional genes (Meyers et al. 1998). Analogous to this, in the mammalian MHC, the class I genes contain approximately 20 genes with only three functional copies. The additional copies are most likely pseudogenes (Trowsdale 1993). Therefore, non-expressed pic19 homologues may serve as a reservoir for genetic diversity. For the pic19 homologous sequences from the BAC contigs only contig 4 showed a continuous open reading frame (ORF). Contig 3 showed one stop codon. Contigs 1 and 2 revealed several frame shifts due to SNPs. But since the BAC library originated from the SCMV susceptible inbred line B73, this does not say anything about the potential function of these copies in resistant genotypes.

The domain of R genes responsible for the recognition event and therefore, of highest interest for the development of new recognition specificities is the LRR. Mutations leading to amino acid replacements (nonsynonymous substitutions) in the recognition determining regions and intragenic recombination leading to expansion or contraction of the LRR number may facilitate rapid evolution of such novel recognition specificities. Michelmore and Meyers (1998) emphasise divergent selection acting on arrays of solvent-exposed amino acid residues in the LRR resulting in evolution of individual R genes within a haplotype. Comparisons between resistance haplotypes reveal that orthologs (genes seperated by a speciation event occupying allelic positions within the cluster) are more similar than paralogs (genes that have arisen by duplication events) implying a low rate of sequence homogenisation from unequal crossing-over and gene conversion. Therefore, intergenic unequal crossing-over and gene

conversion are important for the initial formation of R gene clusters and also generation of genetic variation, but not the primary mechanism generating novel recognition specificities.

Although the features of clustering and rapid evolution of R genes suggest that a gene specific for recognition of a particular pathogen species can evolve to recognise a different pathogen species, there is no evidence yet to support this hypothesis. Future cloning and sequencing of linked R genes conferring resistance to different pathogenes may accelerate understanding and eventually demonstrate such a common evolutionary origin.

RGAs and R gene isolation

The RGA approach has previously been successfully applied for the isolation of R genes. Namely two R genes, *Rp1-D* from maize (Collins et al. 1999) and *Rx* from potato (Bendahmane et al. 1999), have been cloned using RGAs.

A search of candidates for SCMV R genes showed that pic19 is a candidate for *Scmv1*. Genetic mapping identified few resistant individuals showing recombination between pic19 and *Scmv1*. However, because of the existence of escapes and incomplete penetration the recombination cannot rule out identity between pic19 and *Scmv1*. Xu et al. (2000) identified the maize inbred lines A632 and FAP954A to be the putative donors of *Scmv1*. Ancestor analysis of the resistant inbred lines revealed the same restriction pattern for pic19 as expected for *Scmv1*. Only the resistant, the partial resistant, and the putative *Scmv1* donor lines showed the same restriction pattern, whereas the rest of the ancestor lines and the susceptible inbreds showed a different pattern. Therefore, genetic mapping cannot rule out identity for pic19 and *Scmv1* and ancestor analysis suggest at least the same evolutionary origin. Furthermore, pic19 showed continuous ORFs in all three resistant inbred lines suggesting expression of the respective sequence. This result could be confirmed by screening of a cDNA library constructed from the SCMV resistant inbred line FAP1360A using pic19 as probe with three different kinds of positive cDNA clones identified (Liang and Quint, unpublished results). The presence of more than one SCMV R gene in the *Scmv1* region was suggested before. Latest results from a QTL analysis in a FAP1360A x F7 F_3 population underline these assumptions and identified another SCMV QTL in the *Scmv1* region (Yuan et al. 2002). Consequently, inheritance of SCMV resistance seems to be more complex and

isolation of SCMV R genes in this region is much more complicated than expected. This also gives more importance on the RGA approach since map-based cloning of SCMV R genes in this region requires a clear understanding of the mode of inheritance of SCMV resistance and knowledge of the number of genes involved in this process.

Mapping of the BAC contigs containing pic19 homologous sequences identified two clusters on chromosomes 6 and 1 in genome regions known to be involved in the inheritance of SCMV resistance. Therefore, pic19 seems to be a very interesting candidate regarding SCMV resistance. Furthermore, this study demonstrated that sequence homologies can be used to identify further candidate genes from the same target region on basis of high sequence similarities of clustered RGA paralogues. Prerequisite for this approach is the identification of at least one RGA in the target region. The stringency of screening conditions significantly affects the number and degree of sequence similarity of the identified homologous RGA sequences. Therefore, application of low-stringent screening conditions increases the number of homologues with this approach. This opens the possibility to detect a number of evolutionary more distant RGA sequences, and thus, increasing the number of candidates significantly. Therefore, RGAs provide an excellent tool for the identification of candidate genes in target regions such as the *Scmv1* region on maize chromosome 6S. To exploit this approach it is necessary to design PCR primers representing all known classes of plant disease R genes. NBS and LRR motifs are the most widespread conserved amino acid domains in plant disease R genes, but also making use of conserved sequences in other motifs like leucine zippers (coiled coil), TIR, and transmembrane domains could increase the number of potential candidates from target regions.

The initial identification of a RGA in the target region could be simplified by construction of a public database gathering all RGA information over varieties and species. RGAs could be ordered into species and genome position, thus, facilitating the initial step of identification of RGAs in target regions. Such a database also facilitates making use of synteny and colinear relationships between species. For the near future, BAC contigs spanning most of the genomes of major crops will become available. This opens completely new possibilities for RGA analysis making use of degenerate primers. If the genome position of the target gene is known, BACs spanning this region could be selected and used as templates for such a degenerate primer approach resulting in numerous candidate genes.

However, the RGA approach alone is insufficient for efficient cloning of plant disease R genes. There are different suggestions about the distribution of RGAs between active genes and pseudogenes. Meyers et al. (1998) suggested most RGAs to be expressed, while Kanazin et al. (1996) questioned the isolation of R genes by RGAs because pseudogenes accumulate more mutations compared to active genes as source of polymorphism between genotypes. Therefore, the fraction of RGAs representing inactive pseudogenes seems to be significant making the RGA approach not a straight forward method for the isolation of R genes. Genetic mapping of RGAs may be important for locating molecular markers in the immediate neighbourhood of R genes. Hence, RGAs are at present mostly applied for map-based cloning and MAS (Collins et al. 1998, Shen et al. 1998). If the RGA sequence identifies a pseudogene, it cannot be used as a candidate gene but for map-based cloning and MAS.

Another aspect concerning isolation of R genes making use of RGAs is the fact that there are also R genes containing neither NBS nor LRR domains. To date, these are namely two genes, *Hm1* in maize conferring resistance to *Cochliobolus carbonum* (Johal and Briggs 1992), and *Pto* in tomato conferring resistance to *Pseudomonas syringae* pv *tomato* (Martin et al. 1993). While *Hm1* - encoding for a toxin reductase - does not even contain an avr gene and is therefore, probably not involved in a signal tranduction pathway in agreement with the gene-for-gene concept (Flor 1971), *Pto* has an avr counterpart and encodes for a protein kinase. Existence of such types of R genes has to be taken into account and identification of those cannot be accomplished with the RGA approach even though the majority of R genes cloned to date indicates that most of plant disease R genes contain conserved amino acid domains like NBS or LRR. Regarding this aspect, it is also possible that SCMV resistance as a model for oligogenic inherited resistances is not triggered by a recognition event through an RGA type of R gene. SCMV resistance is not expressed by a hypersensitive response reaction. SCMV resistance is a threshold character and, therefore, possibly not underlying the gene-for-gene concept.

Summarising, map-based cloning using any marker type is the more targeted approach compared to RGA analysis and the method of choice, if genetic maps and material segregating for the trait/gene of interest are available. However, in certain genome regions like chromosomal stretches with suppressed recombination a map-based approach is unsuitable and alternative approaches like transposon tagging or RGA analysis provide better chances to isolate the target gene. RGA analysis can generally accompany a map-based

approach broadening the possibilities to identify closely linked markers or even candidate genes.

RGAs and AFLPs as molecular markers for plant disease resistance

Genetic and physical clustering of plant R genes and RGAs and the abundance of the NBS-LRR superfamily in plant genomes complicate the use of RGAs for direct isolation of disease resistance specificities. However, the existence of hundreds of RGAs in each genome of plant species provides a valuable source and ideal starting point for the development of molecular markers.

For efficient development and application of molecular markers, important features are occurrence in large numbers and a high degree of polymorphism between genotypes. As members of a large superfamily, RGAs occur in great numbers. Appr. 1% of maize and rice genes belong to the NBS-LRR class (Michelmore 2002). Furthermore, as shown for pic13, pic19, and pic21, RGAs display a high degree of polymorphism compared to other genome sequences, which is an important prerequisite for the conversion to simple PCR-based markers. Therefore, RGAs meet important criteria for efficient application in map-based cloning or MAS. Their preferential mapping to resistance related genome regions makes them an ideal tool for map-based isolation of R genes or MAS of R genes with respect to pyramidisation of a number of resistance specificities in the same genotype and thus, prevention of fast breakdown of qualitative resistances against plant diseases.

When the R gene does not belong to one of the gene classes containing LRR and/or NBS domains, the `gene-narrowing´ AFLPs are the method of choice while the `gene-landing´ RGAs must fail in this case. In the literature there are several examples for successful cloning of target genes in using map-based cloning based on the AFLP method (e.g. Büschges et al. 1997). Therefore, parallel to the RGA approach, an AFLP-based high-resolution mapping was conducted for map-based cloning of the target genes *Scmv1* and *Scmv2* using BSA based on AFLPs (Dußle et al. 2002). In BSA, AFLP markers specific for the resistant bulk are supposed to be closely linked to one of the target genes and need to be fine mapped in a large mapping population. Unfortunately, AFLP markers are too expensive and laborious for high-throughput monitoring of large mapping populations. Therefore,

analogue to RGAs, AFLPs need to be converted to simple PCR-based STS markers. These converted STS markers significantly reduce the costs of molecular marker analysis. AFLP fragments are usually shorter than RGA sequences and therefore, provide less possibilities to identify polymorphism between parental genotypes. In our study, the degree of polymorphism was significantly lower in the same six inbred lines compared to RGAs. Only two of eight AFLP markers could be converted into one CAPS and one indel marker, while all four (only three included in this thesis) RGAs could easily be converted to CAPS markers. Therefore, RGAs seem to be better suited to conversion into STS markers. One possibility to increase the success ratio of AFLP conversion is elongation of the respective fragments by inverse PCR, which means an additional laborious cloning step. Analogously to RGAs, our results from conversion of AFLP fragments revealed multiple sequences represented by single bands. Hence, AFLPs frequently do not display only a single locus making the map positions of unsequenced AFLP markers questionable comparable to the potentially duplicated RGAs.

Duplicated sequences as molecular markers

Identification of five different BAC contigs with the previously single-copy mapped RGA pic19 posed the question whether the locus in the *Scmv1* region on chromosome 6 is true or fictive. Sequence analysis of the respective pic19 homologous sequences of the BAC contigs showed extremely high sequence similarity of 94-98%. It became clear that the single-copy pic19 locus represents at least five duplicated sequences which can be distinguished only by single nucleotide changes. However, since no BAC library was available for the parents of the D145 x D408 F_3 mapping population, the original pic19 RGA and the homologous sequences from the BAC contigs were mapped in different populations and therefore, we can only be sure about the IBM population that pic19 displays a duplicate locus with at least five copies in the maize genome. These discrepancies and the high sequence similarities led to the conclusion that molecular markers derived from potentially duplicated sequences of similar fragment length might falsify genetic mapping caused by simultaneous scoring of multiple loci in the genome and thus, resulting in fictive ghost markers. For this phenomenon high sequence similarity is not even necessary. For the application of PCR markers sequence identity is only needed in the primer region. In the case of similar fragment

length, different sequences can only be distinguished by sequencing but not on the level of gel electrophoresis, which is the usual means of molecular marker analysis.

Excess of heterozygotes resulting in distorted segregation ratio is the major consequence of simultaneous scoring of two markers. Instead of estimating the correct map positions for the two duplicate loci relative to linked marker loci, the position of the fictive ghost marker is estimated. Simultaneous mapping of more than two homologous sequences would shift the genotype frequency even more in favour of the heterozygotes. However, existing mapping programs use maximum likelihood estimation for calculation of pairwise recombination frequencies, requiring the expected genotype frequencies at two markers. This results in an incorrect estimation of the pairwise recombination frequency in the case of segregation distortion. As a consequence, incorrect map positions may be derived by the discrepancy concerning the expected genotype frequencies within the mapping programs and the shifted genotype frequencies. Bailey (1961) suggested that heterozygosity and segregation distortion lead to biased multipoint linkage maps. But until today it was unknown if these phenomena affect locus order within linkage groups. The occurrence of pic19 homologous sequences followed by the idea of ghost markers caused by excess of heterozygotes connected Baileys suggestion with our experimental results. We were able to theoretically prove the existence of ghost markers, derive recombination frequencies to other marker loci, and derive conditions under which duplicate markers result in incorrect locus orders.

As a consequence, incorrect locus orders can negatively affect the assignment of target genes to chromosome regions in a map-based cloning experiment, hinder indirect selection for a favourable allele at a quantitative trait locus, and decrease the efficiency of reducing the chromosome segment attached to the target gene in marker-assisted backcrossing. The abundance of duplications in eukaryote genomes and the fact that several types of molecular markers are generated from potentially duplicated sequences underlines the importance of this phenomenon.

Conclusions

RGA analysis provided general information on SCMV resistance as well as on resistance genetics. Despite of indications for the presence of a R gene cluster in the *Scmv1*

region on maize chromosome 6S, the question of clustering versus a pleiotropically acting single gene could not be answered. Identification of the RGA clusters on chromosomes 6S and 1 provides strong evidence for clustering of R genes in these regions. Furthermore, this RGA analysis identified new candidates for SCMV QTL and successfully applied a new method on candidate gene identification in target regions. We evaluated the use of RGAs as molecular markers and investigated the implications of duplications for the construction of genetic maps. It became clear that duplicated genome sequences greatly influence genetic linkage mapping. This phenomenon was not recognised before and has consequences for genetic mapping in general, not restricted to the plant kingdom or a special marker type. Therefore, RGA analysis is of great importance for basic research in resistance genetics, but for targeted isolation of genes of interest, it can only be accompanying other approaches like map-based cloning or transposon tagging. For the future, the availability of BAC contigs representing whole genomes is very promising and will together with the identification of RGAs within defined regions greatly accelerate cloning of plant disease R genes.

For the SCMV project, a main task for the future will be the connection of the two different approaches - RGAs and AFLPs - to expand the collection of candidate genes and relate them to closely linked molecular markers looking forward to cloning SCMV R genes.

References

Bailey NTJ (1961) Introduction to the mathematical theory of genetic linkage. Clarendon press, Oxford

Bendahmane, A, Kanyuka K, Baulcombe DC (1999) The *Rx* gene from potato controls seperate virus resistance and cell death responses. Plant Cell 11: 781-791

Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Töpsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley Mlo gene: a novel control element of plant pathogen resistance. Cell 88: 695-705

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogues in maize. Mol. Plant-Microbe Interact. 11: 968-78

Collins NC, Drake J, Aycliffe M, Sun Q, Ellis J (1999) Molecular characterisation of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11: 1365-1376

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JDG (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84: 451-459

Dußle CM, Quint M, Xu ML, Melchinger AE, Lübberstedt T (2002) Saturation of two chromosome regions conferring resistance to SCMV with SSR and AFLP markers by targeted BSA. Theor. Appl. Genet. (in press)

Flor HH (1971) Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. 9: 275-296

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48: 575-607

Johal GS, Briggs SP (1992) Reductase activity encoded by the *Hm1* disease resistance gene in maize. Science 258: 985-987

Jones DA, Dickinson MJ, Balint-Kurti BJ, Dixon MS, Jones JDG (1993) Two complex resistance loci revealed in tomato by classical and RFLP mapping of the *Cf-2*, *Cf-4*, *Cf-5*, and *Cf-9* genes for resistance to *Cladiosporum fulvum*. Mol. Plant-Microbe Interact. 6: 348-357

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. 93: 11746-11750

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262: 1432-1436

Meyers BC, Chin, DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10: 1817-1832

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. 20:317-332

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8: 1113-1130

Michelmore RW (2002) Natural and artificial evolution of plant disease resistance genes. Plant, Animal & Microbe Genomes X Conference, p. 7

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91: 821-32

Parniske M, Wulff BHM, Bonnema G, Thomas CM, Jones DA, Jones JDG (1999) Homologues of the *Cf-9* disease resistance gene (*Hcr9s*) are present at multiple loci on the short arm of tomato chromosome 1. Mol. Plant-Microbe Interact. 12: 99-102

Shen KA, Meyers BC, Islam-Faridi N, Stelly DM, Michermore RW (1998) Resistance gene candidates identified using PCR with degenerate oligonucleotide primers map to resistance gene clusters in lettuce. Mol. Plant-Microbe Interact. 11: 815-823

Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC (1997) Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9: 1279-1287

The EU Arabidopsis Genome Project (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391: 485-488

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JDG (1997) Characterisation of the tomato *Cf-4* gene for resistance to *Cladiosporum fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9: 2209-2224

Trowsdale (1993) Genomic structure and function in the MHC. Trends Genet 9: 117-122

Xu ML, Melchinger AE, Lübberstedt T (2000) Origin of *Scm1* and *Scm2* – two loci conferring resistance to sugarcane mosaic virus (SCMV) in maize. Theor Appl Genet 100: 934-941

Yuan L, Dußle CM, Melchinger AE, Utz HF, Lübberstedt T (2002) Clustering of QTL conferring SCMV resistance in maize. Maydica (in press)

Summary

With the recent cloning of a number of plant disease resistance genes (R genes) it became apparent that R genes share certain homologies in conserved amino acid domains. PCR amplification of genomic DNA using degenerate primers on the basis of these conserved amino acid domains identified sequences with homologies to plant disease R genes - resistance gene analogues (RGAs). RGAs exist in large numbers in plant genomes and provide new possibilities for the investigation of resistance genetics in general and also for the analysis of certain plant disease resistances.

The overall objective of this thesis was to evaluate the use of RGAs for plant breeding for the example of sugarcane mosaic virus (SCMV) resistance in maize. SCMV is one of the most important virus diseases of maize and causes serious yield losses in susceptible cultivars. Owing to the non-persistent manner of transmission, control of aphid vectors by chemical means is not effective and therefore, cultivation of resistant maize varieties is the most efficient method of virus control. Previous studies on the inheritance of oligogenic SCMV resistance located two major quantitative trait loci (QTLs) - *Scmv1* and *Scmv2* - on chromosomes 6S and 3L, respectively. The objectives of this study were to (1) give an overview on the current status of breeding for virus resistance in maize, (2) identify and genetically map candidate genes for *Scmv1* and *Scmv2*, (3) use potential sequence homologies of linked RGAs for targeted increase of the number of candidate genes in the target regions, (4) convert closely linked amplified fragment length polymorphism (AFLP) markers into codominant, simple PCR-based markers as a tool for marker-assisted selection (MAS) and map-based cloning, (5) evaluate RGAs for the development of molecular markers, MAS, and map-based cloning, and (6) investigate the consequences of duplicate markers for the construction of linkage maps and their implications for MAS and map-based cloning.

Three previously published RGAs, pic13, pic21, and pic19 were cloned from six maize inbred lines, converted to cleaved amplified polymorphic sequence (CAPS) markers, and mapped in relation to SCMV R genes (*Scmv1*, *Scmv2*) in maize. Pairwise sequence alignments among the six inbreds revealed a frequency of one single nucleotide polymorphism (SNP) per 33 bp for the three RGAs, indicating a high degree of polymorphism and a high probability of success in converting RGAs into codominant CAPS markers compared to other sequences. Therefore, RGAs meet important requirements for the development of molecular markers, i.e., a high degree of polymorphism and availability in great numbers throughout the genome. In contrast to this, the degree of polymorphism for

AFLPs closely linked to *Scmv1* an *Scmv2* was significantly lower in the same six inbred lines compared to RGAs. Only two of eight AFLP markers could be converted into one CAPS and one indel (insertion/deletion) marker. By genetic mapping, pic21 was shown to be different from *Scmv2*, whereas pic19 and pic13 could be mapped as single-copy markers to the target regions and are candidates for *Scmv1* and *Scmv2*, respectively, due to genetic mapping and consistent restriction patterns of ancestral lines.

Subsequently, pic19 was used as candidate for *Scmv1* to screen a maize BAC library to identify homologous sequences in the maize genome and to investigate their genomic organisation. Fifteen positive BAC clones were identified and classified into five physically independent contigs consisting of overlapping clones. Genetic mapping clustered three contigs into the same genomic region as *Scmv1* on chromosome 6S. The two remaining contigs mapped to the same region as a QTL for SCMV resistance on chromosome 1. Thus, RGAs mapping to a target region can be successfully used to identify further linked candidate sequences. The pic19 homologous sequences of these clones revealed a sequence similarity of 94-98% at the nucleotide level. The high sequence similarity and the multi-locus character of the previously single-copy mapped RGA pic19 show potential problems for the use of RGAs as molecular markers. The existence of ghost markers analogous to ghost QTL was suggested to be a result of simultaneous mapping of several homologous gene family members which cannot be distinguished at the level of PCR.

The idea of ghost loci derived by potentially duplicated sequences such as expressed sequence tags (ESTs), AFLPs, or simple sequence repeats (SSRs) was the subject of a theoretical and computer simulation study. Simultaneous amplification of homologous sequences results in an excess of heterozygotes causing distorted segregation ratios. We were able to theoretically prove the existence of such ghost markers resulting in changes of the correct marker orders. If these fictive ghost markers are part of a genetic map which is the subject of MAS or map-based cloning this may have fatal effects like locating a target gene into an incorrect marker interval. This incorrect locus order caused by duplicate marker loci can negatively affect the assignment of target genes to chromosome regions in a map-based cloning experiment, hinder indirect selection for a favourable allele at a QTL, and decrease the efficiency of reducing the chromosome segment attached to the target gene in marker-assisted backcrossing.

In conclusion, this thesis demonstrates the use of RGAs for plant breeding and resistance genetics in general. RGAs provide a good source for the development of simple PCR-based markers. Furthermore, RGAs are an excellent tool for MAS, the identification of

candidate genes and effective increase of such candidates in target regions using sequence homologies between RGAs. The duplicate nature of RGAs revealed potential problems for genetic mapping of potentially duplicated sequences which are widespread in eukaryote genomes and existent for several types of molecular markers. For resistance genetics in general, investigation of RGAs is important for the understanding of R gene organisation and evolutionary genetics of plant disease resistance.

Zusammenfassung

Pflanzliche Krankheitsresistenzgene (R-Gene) weisen Homologien in konservierten Aminosäureregionen auf. Homologien zu R-Genen ausfweisende, sog. Resistenzgenanaloga (RGAs), existieren in großer Anzahl in pflanzlichen Genomen und eröffnen neue Möglichkeiten für die Erforschung der Resistenzgenetik allgemein sowie die Untersuchung spezifischer pflanzlicher Resistenzen.

Im Rahmen dieser Dissertation wurde am Beispiel der Zuckerrohrmosaikvirusresistenz (sugarcane mosaic virus, SCMV) bei Mais der Nutzen von RGAs für die Pflanzenzüchtung untersucht und bewertet. SCMV ist eine der wichtigsten Viruserkrankungen bei Mais und verursacht signifikante Ertragsverluste in anfälligen Sorten. Aufgrund der nicht-persistenten Übertragungweise des Virus durch Aphiden ist es nicht möglich, die SCMV-Vektoren auf chemischem Wege effizient zu bekämpfen. Daher ist der Anbau resistenter Sorten die einzig wirksame Bekämpfungsmaßnahme. Vorhergehende Studien zur genetischen Basis der oligogen vererbten SCMV-Resistenz identifizierten zwei Genomregionen (quantitative trait loci, QTLs), *Scmv1* und *Scmv2*, auf den Chromosomen 6S und 3L, die massgeblich an der Ausprägung der Resistenz beteiligt sind. Die Ziele dieser Arbeit waren (1) einen Überblick über den aktuellen Status der Virusresistenzzüchtung bei Mais zu geben, (2) die Identifizierung und genetische Kartierung von Kandidatengenen für die beiden Zielgene *Scmv1* und *Scmv2*, (3) die Nutzung potentieller Sequenzhomologien gekoppelter RGAs zur gezielten Erhöhung der Anzahl von Kandidatengenen in den Zielregionen, (4) die Konvertierung eng gekoppelter AFLP (amplified fragment length polymorphism) Marker in codominante, einfach handhabbare PCR Marker, (5) die Bewertung von RGAs für die Entwicklung molekularer Marker, markergestütze Selektion (marker-assisted selection, MAS) und kartengestützte Klonierung, sowie (6) die Untersuchung der Auswirkungen duplizierter Loci auf die Erstellung genetischer Kopplungskarten, MAS und kartengestützte Klonierung.

Die drei zuvor veröffentlichten RGAs pic13, pic21 und pic19 wurden aus sechs Maisinzuchtlinien kloniert und in CAPS (cleaved amplified polymorphic sequence) Marker umgewandelt. Paarweise Sequenzvergleiche ergaben eine Frequenz von einem SNP (single nucleotide polymorphism) alle 33 bp über die drei RGAs. Demnach haben RGAs verglichen mit anderen genomischen Sequenzen einen sehr hohen Polymorphiegrad und lassen sich daher leicht in codominante CAPS Marker umwandeln. RGAs erfüllen wichtige Anforderungen für die Entwicklung molekularer Marker bezüglich des Polymorphiegrades und des zahlreichen Vorkommens in pflanzlichen Genomen. Im Gegensatz dazu war der

Polymorphiegrad der eng koppelnden AFLP (amplified fragment length polymorphism) Marker wesentlich kleiner und nur zwei von acht AFLPs konnten in einen CAPS und einen Indel (Insertion/Deletion) Marker umgewandelt werden. Die genetische Kartierung der RGAs zeigte, daß *Scmv2* nicht mit pic21 identisch ist, während pic19 und pic13 als single-copy Marker in die Zielregionen kartierten und auch aufgrund der Vorfahrenanalysen als Kandidaten für *Scmv1* und *Scmv2* in Frage kommen.

Daraufhin wurde mit pic19 als Kandidat für *Scmv1* in einer Mais BAC-Bibliothek ein Screening durchgeführt, um pic19-homologe Sequenzen im Maisgenom zu identifizieren und deren genomische Organisation zu untersuchen. 15 positive BAC-Klone wurden identifiziert und konnten in fünf physikalisch unabhängige Contigs (Bereiche überlappender Klone), angeordnet werden. Drei Contigs kartierten eng gekoppelt in die *Scmv1*–Region auf Chromosom 6S, während zwei weitere in dieselbe Region auf Chromosom 1 kartierten wie ein QTL für SCMV-Resistenz. Diese Ergebnisse zeigen beispielhaft, dass RGAs aus Zielregionen unter Nutzung potentieller Sequenzhomologien gekoppelter RGAs erfolgreich genutzt werden können, um benachbarte Kandidatengene zu identifizieren. Paarweise Sequenzvergleiche der pic19-homologen Sequenzen der einzelnen Contigs ergaben eine 94-98%ige Sequenzidentität auf Nukleotidebene.

Die hohe Sequenzähnlichkeit und der Multilocuscharakter des ursprünglich single-copy kartierten RGAs pic19 zeigen potentielle Probleme für die Verwendung von RGAs als molekulare Marker auf. Daraufhin wurde als Folge von simultaner Amplifikation homologer Sequenzen, die auf PCR Ebene nicht unterschieden werden können, die Existenz von Ghost-Markern analog zu den bekannten Ghost-QTL vorgeschlagen.

Die Idee der auf potentiell duplizierten Sequenzen wie z.B. ESTs (expressed sequence tags), AFLPs oder SSRs (simple sequence repeats) beruhenden Ghost-Marker war Gegenstand einer detaillierten theoretischen Untersuchung und Computersimulationsstudie. Demnach resultiert die simultane Amplifikation homologer Sequenzen in einem Überschuß an Heterozygoten, wodurch es zu schiefen Spaltungsverhältnissen für den betreffenden Marker kommt. Die Existenz von Ghost-Markern und die daraus folgenden Verfälschungen der korrekten Markerfolge konnten theoretisch nachgewiesen werden. Werden diese fiktiven Ghost-Marker über eine genetische Kopplungskarte für MAS oder kartengestützte Klonierung herangezogen, so kann die Kartierung eines Zielgens in ein falsches Markerintervall aufgrund des Ghost-Markers fatale Auswirkungen haben. Die durch duplizierte Loci hervorgerufene inkorrekte Locusfolge kann die Kartierung von Zielgenen negativ beeinflussen und zur Feinkartierung falscher Markerintervalle bei der kartengestützen Klonierung führen. In der

MAS behindern inkorrekte Locusfolgen die indirekte Selektion auf vorteilhafte Allele in einer QTL-Region und verringern die Effizienz der Reduzierung des Genomanteils des rekurrenten Elters bei markergestützer Rückkreuzung.

Zusammenfassend konnte diese Studie den Nutzen von RGAs sowohl für die Pflanzenzüchtung als auch für die Erforschung der Resistenzgenetik deutlich machen. RGAs stellen aufgrund ihres zahlreichen Vorkommens und des hohen Polymorphiegrades einen hervorragenden Ausgangspunkt für die Entwicklung molekularer Marker dar. Darüber hinaus konnte gezeigt werden, daß RGAs von großem Nutzen für die MAS, die Identifizierung von Kandidatengenen, sowie die effektive Erhöhung der Anzahl der Kandidatengene in einer bestimmten Zielregion durch Nutzung von Sequenzhomologien zwischen RGAs sind. Die potentielle Duplikation von RGAs zeigt mögliche Probleme für die genetische Kartierung von duplizierten Sequenzen auf, die trotz des hohen Anteils von Duplikationen in eukaryotischen Genomen bislang unterschätzt wurden. Diese Problematik trifft nicht nur für RGAs zu, sondern ist über zahlreiche Markertypen hinweg existent. Für die Resistenzgenetik im Allgemeinen liefert die Erforschung von RGAs wichtige Erkenntnisse zum Verständnis der genomischen Organisation und der Evolution von R-Genen und R-Genclustern.

10 Acknowledgements

First of all I would like to thank my family, namely Rolf, Doris, and Nicola Quint, for their uncompromising support and for being the constant in my life.

I am very grateful to my doctoral advisor Prof. Dr. Melchinger for his advise and support during this work, but mainly for imparting me motivation and his concept of scientific research.

Sincere thanks to Dr. Thomas Lübberstedt for his excellent supervision, numerous fruitful discussions, and his scientific creativity.

Special thanks go to Christina Dußle for the time we spent together, her tolerance and helpfulness. In the same way I would like to thank Jasmina Muminovic and Matthias Frisch for their friendship and repeated accommodation.

I highly appreciated the technical assistance of Elisabeth Kokai-Kota, as well as the `Hiwis´ and diploma students Renata Mihaljevic, Eva Knötzinger, Frank Luz, and Nadine Burkarth. Many thanks to Mrs. Barbara Boesig and Beate Devezi-Savula for their friendliness and helpfulness in everyday life in the institute. Finally, I want to thank Franz Mauch and Astrid Karg for their support during the field experiments, my office mate Martin Heckenberger, and all members of our group who are not mentioned here for creating such a pleasant working environment.

This research was financially supported by grants from the Deutsche Forschungsgemeinschaft.

11 Curriculum vitae

Name		Marcel Quint
Date of Birth		21.03.1974 in Celle
Nationality		German
Marital Status		Single

School

80 - 84	Grundschule Vorwerk
84 - 86	Orientierungsstufe Vorwerk
86 - 93	Kaiserin-Auguste-Viktoria Gymnasium Celle
05/93	Graduation: Abitur

University

10/93 - 10/98	Horticulture (Gartenbau) at the University of Hannover
	Diploma thesis at the Institute of Applied Genetics, University of Hannover: 'Construction of a genetic linkage map for *Daucus carota* based on AFLP markers' (Prof. Dr. G. Wricke)
10/98	Graduation: Diplom-Gartenbauingenieur
11/98 - 01/03	Doctorate candidate in Plant Breeding and Applied Genetics (Prof. Dr. A.E. Melchinger), University of Hohenheim, Germany

Practical Experience

07/93 - 10/93	Internship at a landscape gardening company, Sandau & Heindorff in Celle-Nienhagen
10/95 - 03/96	Internship at a tree nursery, Baumschule Lochte in Celle
04/96 - 08/96	Research assistant at the Department of Vegetable Crops, University of California, Davis, USA
08/97 - 10/97	Research assistant at the Institute for Plant Genetics and Crop Plant Research (IPK), Gatersleben
04/95 - 10/98	Research assistant at the Institute of Applied Genetics, University of Hannover
12/02-01/03	Visiting scientist at the Danish Institute of Agricultural Sciences, Slagelse, Denmark