Generation and application of genomic tools as important prerequisites for sugar beet genome analyses

Dissertation zur Erlangung des akademischen Grades des

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Cornelia Lange
aus Rheinhausen j. Duisburg

April 2010

Diese Arbeit wurde am Max-Planck-Institut für Molekulare Genetik, Berlin unter Anleitung von Dr. habil. Heinz Himmelbauer angefertigt.

1. Gutachter: Dr. habil. Heinz Himmelbauer
Centre de Regulacio Genomica (CRG)
C/ Dr. Aiguader, 88, E-08003 Barcelona, Spanien

2. Gutachter: Prof. Dr. Wolfgang Schuster
Institut für Angewandte Genetik, Freie Universität Berlin,
Albrecht-Thaer-Weg 6, D-14195 Berlin, Deutschland

Disputation am: 31.05.2010

Contents

1 List of publications enclosed in this thesis

Publication I

High-throughput identification of genetic markers using representational oligonucleotide microarray analysis.

Lange C*, Mittermayr L*, Dohm JC, Holtgräwe D, Weisshaar B, Himmelbauer H. Theor Appl Genet. 2010 Apr 9. [Epub ahead of print]

DOI: 10.1007/s00122-010-1329-2

Publication II

Haplotype divergence in *Beta vulgaris* and microsynteny with sequenced plant genomes.

Dohm JC, Lange C, Reinhardt R, Himmelbauer H. Plant J. 2009; 57 (1): 14-26.

DOI: 10.1111/j.1365-313X.2008.03665.x

Publication III

Construction and characterization of a sugar beet (*Beta vulgaris*) fosmid library.

Lange C, Holtgräwe D, Schulz B, Weisshaar B, Himmelbauer H. Genome. 2008; 51 (11): 948-51

DOI:10.1139/G08-071

Publication IV

Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in Beta vulgaris L.

Menzel G, Dechyeva D, Keller H, Lange C, Himmelbauer H, Schmidt T. Chromosome Res. 2006; 14 (8): 831-44

DOI: 10.1007/s10577-006-1090-1

2 Summary

Genetic and physical maps of a genome are essential tools for structural, functional and applied genomics. Genetic maps allow the detection of quantitative trait loci (QTLs), the characterisation of QTL effects and facilitate marker-assisted selection (MAS). The characterisation of genome structure and analysis of evolution is augmented by physical maps. Whole genome physical maps or ultimately complete genomic sequences, respectively, of a species display frameworks that provide essential information for understanding processes in respect to physiology, morphology, development and genetics. However, comprehensive annotation underpins the values a genome sequence or physical map represents. An important task of genome annotation is the linkage of genetic traits to the genome sequence, which is facilitated by integrated genetic and physical maps.

In the context of this study several sugar beet (*Beta vulgaris* L.) genomic tools were developed and applied for evolutionary studies and linkage analysis. A new technique allowing high-throughput identification and genotyping of genetic markers was developed, utilising representational oligonucleotide microarray analysis (ROMA). We tested the performance of the method in sugar beet as a model for crop plants with little sequence information available. Genomic representations of both parents of a mapping population were hybridised on microarrays containing custom oligonucleotides based on sugar beet bacterial artificial chromosome (BAC) end sequences (BESs) and expressed sequence tags (ESTs). Subsequent analysis identified potential polymorphic oligonucleotides, which were placed on new microarrays used for screening of 184 F_2 individuals. Exploiting known co-dominant anchor markers, we obtained 511 new dominant markers distributed over all nine sugar beet linkage groups and calculated genetic maps. Besides the method´s transferability to other species, the obtained genetic markers will be an asset for ordering of sequence contigs in the context of the ongoing sugar beet genome sequencing project. In addition, possible linkage of physical and genetic maps was provided, since genetic markers were based on source sequences, which were also used for construction of a BAC based physical map utilising a hybridisation approach. An example of the hybridisation based approach for physical map construction and its application for synteny studies was demonstrated. Since little is known about synteny between rosids and *Caryophyllales* so far, we analysed the extent of

synteny between the genomic sequences of two BAC clones derived from two different *Beta vulgaris* haplotypes and rosid genomes. For selection of the two BAC clones we hybridised 30 oligonucleotide probes based on ESTs corresponding to *Arabidopsis* orthologs on chromosomes 1 and 4 that were presumably co-localised in the reconstructed *Arabidopsis* pseudo ancestral genome (Blanc et al. 2003) on sugar beet BAC macroarrays comprising two different sugar beet libraries. A total of 27,648 clones were screened per sugar beet library, corresponding to 4.4-fold and 5.5-fold, respectively, sugar beet genome coverage. We obtained four and five positive clones for the probes on average. Two clones, one from each haplotype that were positive with the same five EST probes, were selected and their genomic sequences were determined, annotated and exploited for synteny studies.

Furthermore, I constructed and characterised a sugar beet fosmid library from the doubled haploid accession KWS2320 encompassing 115,200 independent clones. The insert size of the fosmid library was determined by pulsed field gel electrophoresis to be 39 kbp on average, thus representing 5.9-fold coverage of the sugar beet genome. Fosmids bear the advantage of narrowly defined size of the clone inserts, thus fosmid end sequences will essentially contribute to the future assembly and ordering of sequence contigs. Since repeats are a major obstacle for successful assembly of plant genome sequences, frequently causing gaps and misassembled contigs, I generated a genomic short-insert library. The short-insert library facilitated repeat identification within the sugar beet genome, which was exemplarily shown for three miniature inverted-repeat transposable element (MITE) families.

Altogether this work contributed substantially to a deeper understanding of the genome structure of sugar beet and provided the basis for successful sequencing of the sugar beet genome.

3 Zusammenfassung

Genetische und physikalische Karten eines Genoms sind essentielle Werkzeuge für strukturelle, funktionelle und angewandte Genomik. Genetische Karten erlauben die Identifizierung von „Quantitative Trait Loci“ (QTLs), die Charakterisierung von QTL-Effekten, und sie ermöglichen Marker gestützte Selektion in der Züchtung. Die Charakterisierung der Genomstruktur, sowie Evolutionsanalysen werden durch physikalische Karten ermöglicht. Umfassende physikalische Karten, bzw. letztendlich eine vollständige Sequenz des Genoms einer Spezies, stellen Gerüste dar, die essentielle Informationen zum Verständnis von Prozessen die Physiologie, Morphologie, Entwicklung und Genetik betreffend, enthalten. Der Nutzen einer physikalischen Karte bzw. einer Genomsequenz hängt jedoch von einer guten, umfangreichen Annotation ab. Ein wichtiger Bestandteil der Genomannotation ist die Verknüpfung von genetischen Merkmalen mit der Genomsequenz, die durch die Integration von genetischen und physikalischen Karten erreicht wird.

Im Kontext dieser Arbeit wurden verschiedene Werkzeuge für Genomik-Studien in der Zuckerrübe (*Beta vulgaris* L.) entwickelt und angewandt, um Evolutionsstudien und Kopplungsanalysen durchzuführen. Es wurde eine neue Technik entwickelt, die basierend auf „Representational Oligonucleotide Microarray Analysis” (ROMA), Hochdurchsatz-Identifikation und -Genotypisierung von genetischen Markern ermöglicht. Die Methode wurde in der Zuckerrübe als Model für Kulturpflanzen mit wenig verfügbarer Sequenzinformation getestet. Hierzu wurden genomische Repräsentationen beider Elternlinien einer Kartierungspopulation auf Microarrays mit benutzerdefinierten Oligonukleotiden, basierend auf Zuckerrüben „Bacterial Artificial Chromosome” (BAC) -Endsequenzen (BESs) und „Expressed Sequence Tags“ (ESTs), hybridisiert. Folgende Analysen führten zur Identifizierung von potentiell polymorphen Oligonukleotiden. Diese wurden zum Design weiterer Microarrays verwendet, die zum Screening von 184 F_2-Individuen dienten. Unter Zuhilfenahme bekannter kodominanter Marker konnten 511 neue, über alle neun Kopplungsgruppen der Zuckerrübe verteilte, dominante Marker gewonnen und genetische Karten berechnet werden. Zusätzlich zu der Möglichkeit die Technik auch auf andere Spezies anzuwenden, stellen die neu gewonnenen genetischen Marker ebenfalls einen Zugewinn für die Anordnung von Sequenz-Contigs im Rahmen des zurzeit laufenden Zuckerrübengenom-Sequenzierungsprojekts dar. Zudem wurde eine Verknüpfung der genetischen Karte mit der physikalischen Karte ermöglicht, da die Sequenzen auf denen die geneti-

schen Marker beruhen ebenfalls für die Konstruktion einer BAC basierten physikalischen Karte genutzt wurden. Ein Bespiel des Hybridisierungs-Ansatzes, der zur Generierung der physikalischen Karte genutzt wurde, und seine mögliche Anwendung für Syntänie-Studien wurde im Weiteren demonstriert. Da bisher wenig bekannt ist über Syntänie zwischen Rosiden und *Caryophyllales*, untersuchten wir den Syntäniegrad zwischen den genomischen Sequenzen zweier BAC-Klone, gewonnen aus zwei verschiedenen Zuckerrüben-Haplotypen und Rosiden-Genomen. Für die Auswahl der zwei BAC-Klone hybridisierten wir 30 Oligonukleotid-Sonden auf Zuckerrüben BAC-Makroarrays, die zwei verschiedene Zuckerrüben-Banken umfassten. Die Sonden basierten auf ESTs, welche *Arabidopsis* Orthologen auf den Chromosomen 1 und 4 entsprachen, die ursprünglich im rekonstruierten pseudo-anzestralen *Arabidopsis*-Genom kolokalisiert waren (Blanc et al. 2003). Insgesamt wurden 27.648 Klone pro Zuckerrüben Bank untersucht. Dies entspricht einer 4,4-fachen bzw. 5,5-fachen Abdeckung des Zuckerrübengenoms. Im Durchschnitt erhielten wir vier bzw. fünf positive Klone pro Sonde. Zwei Klone jedes Haplotyps, die positiv für die gleichen EST-Sonden waren, wurden ausgewählt und ihre genomischen Sequenzen wurden ermittelt, annotiert und für Syntänie-Studien verwendet.

Desweiteren konstruierte und charakterisierte ich eine Zuckerrüben-Fosmidbank mit 115.200 vereinzelten Klonen aus der doppelt-haploiden Linie KWS2320. Die Größe der Fosmid-Inserts wurde mittels Pulsfeldgelelektrophorese bestimmt und betrug 39 kbp im Durchschnitt, d.h. die Fosmidbank deckte das Zuckerrübengenom 5,9-fach ab. Fosmide weisen den Vorteil auf, dass ihre Insertgröße kaum variiert und daher Fosmid-Endsequenzen die zukünftige Assemblierung und Orientierung von Sequenz-Contigs ebenfalls entscheidend unterstützen werden. Da repetitive Elemente ein großes Hindernis für die erfolgreiche Assemblierung von Pflanzengenomsequenzen sind und häufig große Lücken in Genomsequenzen oder falsch assemblierte Contigs verursachen, habe ich zusätzlich eine genomische „Shotgun“-Klonbank mit kleinen Insert hergestellt. Diese ermöglichte die Identifikation von repetitiven Elementen im Zuckerrübengenom, wie es beispielhaft für drei „Miniature Inverted-Repeat Transposable Element” (MITE) Familien gezeigt wurde.

Insgesamt trug diese Arbeit erheblich zu einem tieferen Verständnis der Struktur des Zuckerrübengenoms bei und stellte die Basis für die erfolgreiche zukünftige Sequenzierung des Zuckerrübengenoms dar.

4 Introduction

4.1 Sugar beet (*Beta vulgaris*)

Sugar beet (*Beta vulgaris*) is a crop plant with high economic importance, since it is the only sucrose storing crop of moderate climates. It accounts for about 25% of the world-wide sugar production (Draycott 2006), provides the basis for a vast number of food products and can be utilised for bio-fuel production. Sugar beet is a relatively young crop, domesticated in the late 18th century (Fischer 1989). Before sugar beet was used for sugar production, sugar isolated from sugar cane had to be imported to Europe from tropical regions. After Marggraf had proven in 1784 that the sweet substance contained in *Beta* forms was chemically identical with cane sugar, beets were grown in central Europe and industrial sugar production started in the beginning of the 19th century. Systematic breeding of sugar beet led to continuous improvement of sugar beet, especially with regard to its sugar content. In the middle of the 19th century, the sucrose content of sugar beet amounted to about 8%, whilst today sucrose can account for around 17% of the plant´s dry weight (www.kws.de). Later and current breeding approaches do not in the first place aim for increased sucrose content in sugar beet roots, but rather for pest and disease resistances and minimizing the breeder´s effort. One early prominent example is the generation of monogerm sugar beet lines (Savitsky 1950). In contrast to old multigerm varieties, monogerm seeds produce single seedlings on germination, which obviated the need for performing laborious singling.

4.2 Taxonomy and angiosperm evolution

Taxonomically, within the angiosperms (flowering plants) sugar beet (*Beta vulgaris* L. ssp. *vulgaris* var. *altissima* Döll) is a member of the core eudicotyledons and belongs to the order of *Caryophyllales* (Figure 1).

Angiosperms are the most diverse extant plant group on earth occupying a large spectrum of habitats. Angiosperm diversification took place in the early Cretaceous period (145-99 million years ago). Sugar beet and *Arabidopsis thaliana* diverged from a common ancestor an estimated 120-130 million years ago (mya). The last common ancestor

of sugar beet and potato (*Solanum tuberosum*) is dated 110 mya (Wilkstrom et al. 2001). Both follow closely the divergence of dicot from monocot families about 140-150 mya (Moore et al. 2007). Genome analyses have revealed that a possible explanation for the fast diversification of the flowering plants might lie in gene or whole-genome duplications (WGD). Recent and ancient WGDs are also termed polyploidy and paleopolyploidy, respectively. It is generally accepted that *Arabidopsis* has undergone at least three WGDs, called α, β and γ events (Figure 1). Jaillon et al. (2007) suggested that the common ancestor of *V. vinifera*, *P. trichocarpa* and *A. thaliana* was an ancient hexaploid, resulting from a paleohexaploidy event (γ) after the divergence of monocots and eudicots early in angiosperm evolution, followed by gene loss leading to pseudodiploid genomes. The exact timing of the γ event is still controversial, ranging from before the split of the monocots and eudicots to the last common ancestor of all extant rosids (Soltis et al. 2009). However, comparative studies between tomato (*Solanum lycopersicum*) and grapevine revealing collinearity between tomato and triplicate regions in grapevine, led to the conclusion that the γ event took place before the split of asterids and rosids (Tang et al. 2008). The *Arabidopsis* genome underwent at least two other WGD events (β and α). The complete genome sequence of papaya has shown, that the β event probably took place in *Brassicales* after the split of papaya and *Arabidopsis* ~72 mya (Ming et al. 2008) and thus being much younger than previously proposed by Bowers et al. (2003) who placed the β event before the divergence of monocots and eudicots. The α event likely occurred within the *Brassicaceae* around 40 mya (Fawcett et al. 2009). Several investigated plants have undergone further lineage or species specific WGDs (Figure 1). Since all recently sequenced eudicot plant genomes represent rosids or asterids, the complete genome sequence of sugar beet will be of great value to gain new insights into angiosperm genome evolution.

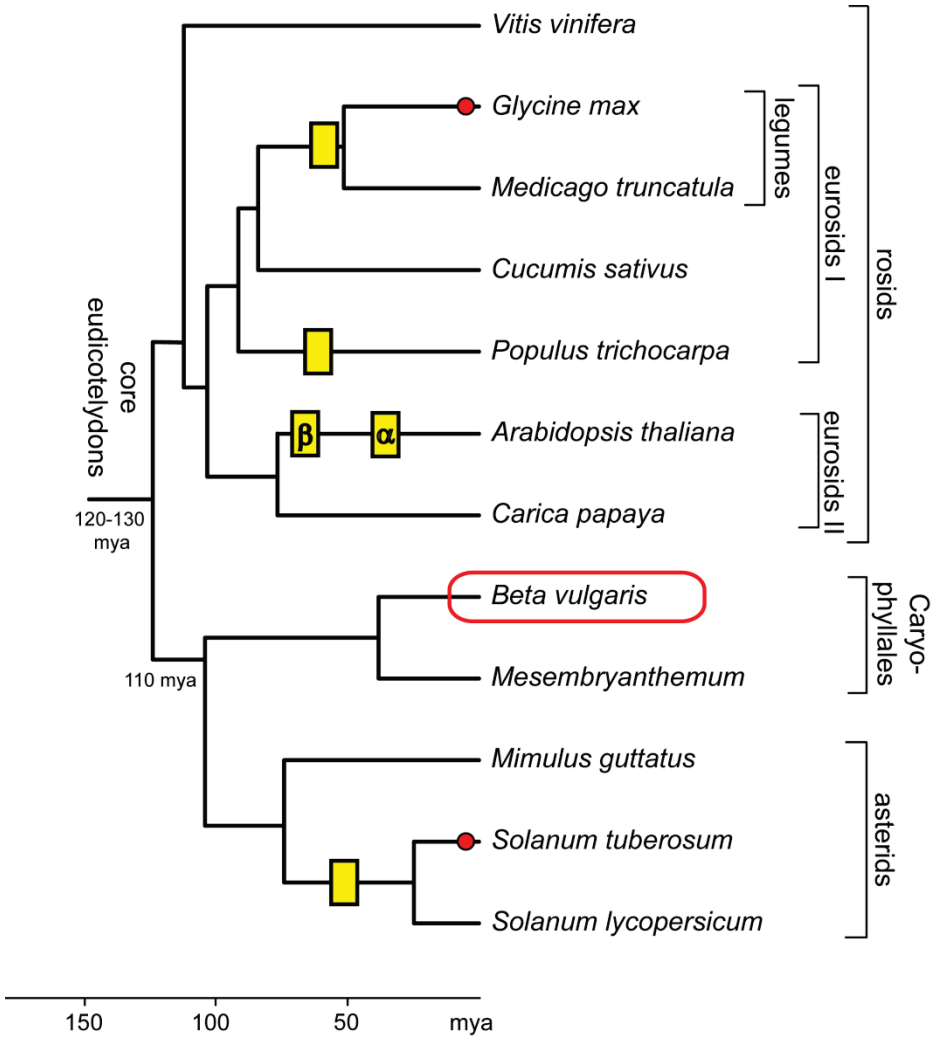


Figure 1. Phylogenetic tree of selected core eudicotelydons. Whole genome duplications (WGDs) in their approximate temporal context are indicated as yellow rectangles. WGDs were inferred from the studies of Bertioli et al. (2009) for legumes, Tuskan et al. (2006) for poplar (*Populus trichocarpa*), Fawcett et al. (2009) for the generally accepted α and β duplication events in *Arabidopsis* and Schlueter et al. (2004) for *Solanaceae*. Red circles denote recent polyploidy events that took place ~13 million years ago (mya) in soybean (*Glycine max*) (Schmutz et al. 2010) and in potato (*Solanum tuberosum*) (Schlueter et al. 2004). Figure modified from Fawcett et al. (2009).

4.3 Plant genome sizes and repeats

Plant nuclear genomes vary greatly in their sizes and structures. Genome sizes across land plants can range over several orders of magnitude. *Genlisea margaretae* is considered to have the smallest genome (63 Mbp) found so far within the angiosperms (Greilhuber et al. 2006), whereas members of the *Liliaceae* possess genomes with sizes of more than 120,000 Mbp (Bennett and Smith 1991).

Sugar beet is a diploid species encompassing n=9 chromosomes. The haploid genome size is estimated to be 758 Mbp (Arumuganathan and Earle 1991), which is larger than the genome sizes of all plants sequenced so far, except for the very recently sequenced soybean (*Glycine max*) genome with a size of 950 Mbp (Schmutz et al. 2010) and the maize (*Zea mays*) genome with 2,300 Mbp (Schnable et al. 2009) (Table 1).

Table 1. Characteristics of sequenced plant genomes.

Plant species	**Hapl. genome size [Mbp]**	**No. of chromosomes**	**No. of predicted protein-coding genes**	**Transposon content [%]**	**Reference**
Arabidopsis thaliana	125	n = 5	27,379*	14	(The Arabidopsis Genome Initiative 2000)
Oryza sativa (ssp. *japonica*)	389	n = 12	37,544	35	(International Rice Genome Sequencing Project 2005)
Populus trichocarpa	485	n = 19	45,555	42	(Tuskan et al. 2006)
Vitis vinifera	487	n = 19	30,434	41	(Jaillon et al. 2007)
Carica papaya	372	n = 9	24,746	52	(Ming et al. 2008)
Cucumis sativus	244	n = 7	26,682	43	(Huang et al. 2009)
Zea mays	2,300	n = 10	32,540	84	(Schnable et al. 2009)
Glycine max	1,115	n = 20	46,430	59	(Schmutz et al. 2010)
Brachypodium distachyon	272	n = 5	25,532	28	(International Brachypodium Initiative 2010)

*Updated with data from The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org)

Repetitive elements comprise a large fraction of plant genomes (Heslop-Harrison 2000). In sugar beet the repeat content was estimated to be around 63%, exploiting reannealing kinetic experiments (Flavell et al. 1974). In general, repetitive elements can be categorised into tandemly arranged and dispersed sequences. Transposable elements (TEs) are dispersed sequences, which comprise a large fraction of repetitive DNA in eukaryotes.

Tandem repeats include satellite DNA consisting of numerous tandemly arranged repeats that are non-coding and mostly located in heterochromatic regions, micro- and minisatellites, telomeric repeats and ribosomal genes. Several studies have shown that TEs together with satellites play a major role in plant genome evolution; in particular accumulation and proliferation of TEs are responsible for the different sizes of plant genomes (Sanmiguel et al. 1996; Heslop-Harrison 2000; Hawkins et al. 2006; Vitte and Bennetzen 2006).

TEs can be divided into two classes according to their mechanisms of transposition. Class I elements (retrotransposons) transpose by reverse transcription of an RNA intermediate performed by a multi-enzyme, while class II elements (DNA transposons) transpose directly from DNA to DNA mediated by an element-encoded transposase (Finnegan 1989). Retrotransposons can be further divided into highly repetitive long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. LTR retrotransposons, including *copia*-like and *gypsy*-like retrotransposons, make up the majority of the transposable element classes in most plants (see references in Table 1). Non-LTR retrotransposons include long interspersed elements (LINEs) and short interspersed elements (SINEs). They are grouped into superfamilies such as Tc1/*mariner*, *hAT* and *Mutator* (reviewed in: Feschotte et al. 2002). Additionally, miniature inverted-repeat transposable elements (MITEs) have been found in many plant genomes (Bureau and Wessler 1994; Tarchini et al. 2000; Jiang and Wessler 2001). MITEs are the predominant transposable element associated with the non-coding regions of the genes of flowering plants (Bureau et al. 1996; The Arabidopsis Genome Initiative 2000). Like DNA transposons, MITEs possess terminal inverted repeats and short target site duplication but they are non-autonomous, i.e. they do not encode a transposase.

In *Beta vulgaris* several studies have been performed analysing the abundance, genomic organization and evolution of tandemly repeated and dispersed repetitive DNA elements, including the characterisation of the physical distribution of microsatellites on chromosomes of sugar beet (Schmidt and Heslop-Harrison 1996). Dechyeva and Schmidt (2006) examined the structure and species-specific diversification of subtelomeric satellite DNA families of the genus *Beta* and related species, applying Southern blotting, fluorescent *in-situ* hybridisation (FISH) and multi-colour FISH on extended DNA fibres. Furthermore, sugar beet repetitive sequence belonging to various classes were identified and characterised, among them LINEs (Kubis et al. 1998), the Tc1/*mariner* DNA transposon Vulmar1 (Jacobs et al. 2004), dispersed repeats belong-

ing to the families pDvul1 and pDvul2, the pRv1 satellite repeat (Menzel et al. 2008), MITEs (Menzel et al. 2006), Ty1-*copia* retrotransposons (Schmidt et al. 1995) and a novel type of plant non-LTR retrotransposons, identified as the BNR family (Heitkam and Schmidt 2009).

4.4 Genome Mapping

The term "genome mapping" describes different approaches to resolve the organisation of genes and other sequences within genomes. There are two major groups of genome maps: genetic maps and physical maps. Genetic maps reflect the relative positions of markers on a chromosome based on probability calculations, assuming that the more often two markers co-segregate, i.e. no crossing-over event has occurred between them during meiosis, the closer they are located within the genome. Physical maps, in contrast, display true physical distances measured in base pairs (bp).
Both map types are very useful tools for genomic studies. Genetic maps allow the detection of quantitative trait loci (QTLs), the characterisation of QTL effects and facilitate marker-assisted selection (MAS). The characterisation of genome structure and analysis of evolution is augmented by physical maps. However, when anchored to each other, providing an integrated map, they have the highest potential for advanced genomics research.
In the following chapters of the introduction, different techniques for construction of genetic and physical maps will be presented.

4.4.1 Genetic mapping and molecular marker systems

In order to construct a genetic map, the segregation pattern of markers in a mapping population is determined and the relative positions and distances, usually measured in centimorgan (cM), between markers are calculated. One cM is equal to a 1% chance that two markers on a chromosome will be separated in a single generation due to crossing-over during meiosis. Hence, genetic maps do not display physical distances but probabilities. cM units can be transferred into corresponding physical distances, i.e. base pairs (bp), but these are just estimated, averaged values, since recombination frequency is not equally distributed over a genome. Mapping of genetic markers is only

possible, if the markers are polymorphic, i.e. there is variation between the parental lines at the marker loci. Thus, the first task of genetic mapping is to find polymorphic markers, followed by scoring of their segregation pattern in the mapping population and finally ordering and calculation of genetic distances. Important factors influencing the robustness and value of a genetic map are the number of markers and individuals that are used for map calculation, its possible application for high-throughput and the character of the markers (dominant/codominant; anonymous/known). Over the past years several molecular marker technologies for genetic mapping have been developed, all addressing different tasks, depending on the desired needs and future usage. In practice several marker systems are often combined to produce genetic maps, in order to achieve higher marker densities.
The following sections describe the most prominent techniques and their advantages and disadvantages, respectively.

4.4.1.1 Restriction fragment length polymorphism (RFLP) scoring

One of the earliest technologies used on the DNA level was restriction fragment length polymorphism (RFLP) scoring (Botstein et al. 1980). RFLP methodology is based on the ability of DNA restriction enzymes to recognise and cleave specific DNA motives, yielding fragments of defined lengths. Base composition alterations within a recognition site can create or destroy a restriction site leading to variation in the number of sites. Alternatively, insertion or deletion of blocks of DNA within a fragment could alter its size. Fragments encoding specific sequences from two individuals can be analyzed by Southern hybridization (Southern 1975), allowing the identification of different sized fragments and thus visualising a polymorphic marker. RFLP scoring bears the advantages of being highly reproducible, providing co-dominant markers which are transferable between different populations. Restrictions in the number of loci that can be analysed simultaneously, laborious assay steps and the need for substantial amounts of genomic DNA are major drawbacks of RFLP scoring. The invention of PCR (Mullis et al. 1986) led to combination of both techniques and the development of PCR-RFLP (Deng 1988), also called cleavage amplification polymorphism (CAP). PCR-RFLP includes PCR amplification of a fragment with a sequence-specific primer pair followed by digestion with a restriction enzyme. Digested fragments can be directly analysed on an agarose gel without need of a subsequent hybridization step. RFLPs have been used for

construction of linkage maps for various plants, such as maize (Helentjaris et al. 1986), potato (Bonierbale et al. 1988) and sugar beet (Barzen et al. 1992; Hallden et al. 1996). RFLP scoring is still used, e.g. to test genetic diversity between different accessions of a species (Kojima et al. 2005), but PCR-RFLP is the prevalently applied method, in plants especially utilised for analyses of agronomically important loci (Kaundun and Matsumoto 2003; Avila et al. 2006; Asakura et al. 2009; Upadhyay et al. 2009) and in combination with different marker systems to construct linkage maps (Schneider et al. 1999; Cuevas et al. 2008; Orr and Molnar 2008).

4.4.1.2 Simple sequence repeats (SSRs) and inter-simple sequence repeats (ISSRs)

Simple sequence repeats (SSRs) or microsatellites are tandemly repeated short DNA stretches (unit size < 6 bp), which are scattered throughout eukaryotic genomes at many different locations. Since they have been shown to be highly polymorphic in the number of repeats within a block of tandemly repeated DNA within a species, they can be used as genetic markers (Weber and May 1989). Isolation of SSRs comprises screening of a small insert genomic library with microsatellite probes. Subsequently, inserts containing SSRs are sequenced and PCR primers flanking the particular loci are designed. Typing of the SSRs in a mapping population involves only two steps, PCR amplification and electrophoresis to generate DNA banding patterns on a gel and to reveal repeat number polymorphisms. Advantages of this marker system are the co-dominant character of the markers, its high reproducibility, straight-forward experimental performance and the transferability of SSR markers across different populations. However, they are quite costly to produce and cannot be multiplexed to a high extent. SSR markers have been widely utilised for studying genetic variation (Vigouroux et al. 2005; Malysheva-Otto et al. 2006; Blair et al. 2009) and linkage mapping (Taramino and Tingey 1996; Mccouch et al. 1997; Laurent et al. 2007; King et al. 2008) in plants.

A distinct marker system also based on microsatellites and PCR amplification, are inter-simple sequence repeats (ISSRs). Primers complementary to microsatellites are used as primers to amplify the regions between the microsatellite loci, resulting in a mixture of amplified fragments (Zietkiewicz et al. 1994). Length variations of amplifiable fragments between different individuals can be detected on gels. Unlike exploiting SSRs, for ISSRs applications no prior sequence information is necessary for primer design, but on the other hand a drawback of ISSR markers is their dominant character. The main

application of ISSR markers in plant is the exploration of genetic variation between populations and closely related species (Fang and Roose 1997; Joshi et al. 2000; Rout et al. 2009), but they have also contributed to the construction of genetic maps (Kojima et al. 1998; Casasoli et al. 2001; Gupta et al. 2008).

4.4.1.3 Random amplified polymorphic DNA (RAPD)

Another marker system exploiting PCR amplification is random amplified polymorphic DNA (RAPD), which is based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence (Williams et al. 1990). Polymorphisms are inherited in a Mendelian fashion and can be detected on agarose gels as DNA segments, which can be amplified from one parent but not the other. Since they are either present or absent on the detection gel, RAPD markers are dominant. The major benefits of RAPD assays are their independence of target DNA sequence information for the design of amplification primers and the very simple and cheap experimental design. Their very limited capability of being transferable between populations and species and their problems with reproducibility are critical disadvantages. Besides phylogenetic and diversity studies (Vierling and Nguyen 1992; Singh et al. 2009; Szczepaniak et al. 2009), RAPD markers have been exploited for genetic map construction (Bradshaw et al. 1994; Kesseli et al. 1994; Haque et al. 2008).

4.4.1.4 Amplified fragment polymorphism (AFLP)

Amplified fragment polymorphism (AFLP) technique combines PCR amplification and fragmentation with restriction enzymes (Vos et al. 1995). However, in contrast to PCR-RFPLs, AFLPs are generated by first performing restriction digestion of DNA, ligation of the resulting fragments to oligonucleotide adapters, serving as binding sites for the PCR primers, followed by selective PCR amplification of the fragments and analysis of the amplified fragments. Usually, the restriction digestion is performed with two restriction endonucleases. The choice of the utilised restriction enzymes determines the size range of the produced DNA fragments. Selective PCR amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotides flanking the restriction sites (Vos et al. 1995), resulting in a unique, reproducible profile. AFLP does not require any

prior knowledge of nucleotide sequence of the target DNA and the method allows co-amplification of high numbers of restriction fragments. Modern detection systems such as capillary electrophoresis of fluorescently labelled AFLP products have replaced conventional denaturing polyacrylamide gels, and thus allow to simultaneously evaluate a large number of loci and to produce high-resolution genetic maps with dominant markers (Schondelmaier et al. 1996; Bert et al. 1999; Kamisugi et al. 2008).

Sequence-specific amplification polymorphism (S-SAP) is a marker system derived from the AFLP technique. S-SAP exploits the high degree of sequence heterogeneity and insertional polymorphisms, both within and between species, of retrotransposons (Waugh et al. 1997). Only one AFLP primer and a second primer, complementary to the retrotransposon or another sequence of interest, are used for selective amplification. Waugh et al. (1997) demonstrated the usefulness of the system by detecting DNA polymorphisms based on position of LTR retrotransposon sequences in relation to adjacent restriction endonuclease sites in barley (*Hordeum vulgare*). The advantage of S-SAP is its usually higher degree of polymorphisms compared to AFLP. In principle, S-SAP is applicable to any transposable element in any organism (Syed and Flavell 2007) and has been successfully performed to construct linkage maps in a number of plant species, such as wheat (Queen et al. 2004), lettuce (Syed et al. 2006) and artichoke (Portis et al. 2009).

4.4.1.5 Single nucleotide polymorphisms (SNPs)

With ongoing progress in the development of sequencing technologies and the increase in available DNA sequence resources, single nucleotide polymorphisms (SNPs) have gained importance as molecular markers. SNPs are the most abundant variants within the genomes of eukaryotes. SNP rates vary between different species, but in general they are higher in non-coding regions, than in protein coding regions. SNP rates (average of coding and non-coding regions) of 1 per 247 bp in rape seed (*Brassica napus*) (Westermeier et al. 2009), 1 per 104 bp in maize (Tenaillon et al. 2001) and 1 per 65 bp (Schneider et al. 2007) or 1 per 29 bp (Dohm et al. 2009), respectively, in sugar beet have been estimated. Generally, SNPs are detected *in silico* by analysis of aligned sequences obtained from databases, through sequencing or re-sequencing of candidate genes, PCR products or whole genomes or transcriptomes of several genotypes. The ongoing improvements of next generation sequencing (NGS) technologies

(comprehensively reviewed in: Metzker 2009), such as 454 pyrosequencing (Margulies et al. 2005), Illumina/Solexa sequencing (Bentley et al. 2008) and SOLiD (Valouev et al. 2008), allow to re-sequence genomes or sequence transcriptomes for SNP detection in a high-throughput compatible manner. In this way, in the presence of a reference genome sequence, SNP detection for genes from secondary metabolite biosynthetic pathways utilizing NGS technologies was carried out in *Eucalyptus* species (Kuelheim et al. 2009). Furthermore, SNPs were detected in *Brassica napus* by Solexa transcriptome sequencing using a publicly available set of *Brassica* species unigenes as a reference sequence (Trick et al. 2009) and in maize performing transcriptome sequencing of shoot apical meristems with 454 sequencing technology (Barbazuk et al. 2007). Genotyping of previously detected SNPs, e.g. for linkage map construction, can be achieved in various ways. Several high-throughput SNP genotyping platforms have been developed, such as the molecular inversion probe (MIP) (Hardenbol et al. 2003) and Illumina GoldenGate (Fan et al. 2003) assays, both combining multiplex PCR with array hybridization and genotyping by hybridization of genomic representations on microarrays (Matsuzaki et al. 2004) (high-density genotyping platforms are extensively reviewed in: Fan et al. 2006). High-density genetic linkage maps of soybean, barley and cowpea (*Vigna unguiculata*) were generated by sequencing amplicons and ESTs, followed by genotyping with Illumina GoldenGate assays (Hyten et al. 2008; Close et al. 2009; Muchero et al. 2009).

4.4.2 Physical mapping techniques

In contrast to genetic maps, physical maps represent real physical distances of markers along chromosomes or DNA stretches. When integrated with genetic maps, physical maps allow linking of genetically mapped markers to actual physical locations. Different types of physical maps have been developed that vary in their degree of resolution. The physical map with the highest possible resolution would be the complete genome sequence of a given species. Complete genome sequences are available only for a very limited number of species so far. Most strategies for sequencing a whole genome require a physical map as an essential prerequisite, since it provides a scaffold for sequence or contig, i.e. a set of clones that are related to one another by overlap of their sequences, respectively, assembly.

The following parts of the introduction explain different physical map types and marker systems and point out their benefits and drawbacks.

4.4.2.1 Cytogenetic mapping

The lowest-resolution physical map is a cytogenetic map, which is based on the visual appearance of a chromosome when stained and examined under a microscope. Obtaining cytogenetic maps by hybridization with labelled probes was introduced with the development of the DNA *in situ* hybridisation technique (Gall and Pardue 1969; John et al. 1969). Initially used radiation based methods for probe labelling and signal detection were soon replaced by fluorescence-based techniques (Langer-Safer et al. 1982). The resolving power of fluorescence in situ hybridization (FISH) varies between 2 Mbp and 10 Mbp and depends on the cytological targets, encompassing interphase nuclei, mitotic prometaphase and metaphase chromosomes, super-stretched mitotic metaphase chromosomes, meiotic pachytene chromosomes, and extended DNA fibers (reviewed in Jiang and Gill 2006). The advantage of FISH rests mainly in the ability to directly determine and visualise the chromosomal location of DNA clones. FISH based physical maps covering entire chromosomes have been constructed for various plants including amongst others *Brassica oleracea* (Howell et al. 2005), maize (Koumbaris and Bass 2003), soybean (Walling et al. 2006) and potato (Iovene et al. 2008). Besides physical mapping different variants of FISH are applied in plants for chromosome identification (Pedersen and Langridge 1997; Dong et al. 2001; Kim et al. 2002; Lengerova et al. 2004; Szinay et al. 2008), karyotyping (Badaeva et al. 2002; Han et al. 2008; Falistocco 2009), repeat analyses (Dechyeva and Schmidt 2006; Menzel et al. 2006; Han et al. 2008; Macas et al. 2009) and chromosome-specific painting (Lysak et al. 2001; Tang et al. 2008).

4.4.2.2 Mapping with large insert clone libraries

Physical maps constructed by ordering a collection of overlapping cloned DNA fragments utilizing large insert clone libraries provide a higher resolution than cytogenetic maps. Several vector systems having different features can be selected for inserting the DNA fragments of interest (Table 2). Important factors to consider when choosing a vector system are the insert size, stability and ease of manipulation of the source library.

4.4.2.2.1 Large insert vector systems

Cosmids were among the earliest developed vectors capable of carrying large inserts with a size of approximately 40 kbp (Collins and Hohn 1978). They contain the cohesive - end site (*cos*) sequences of λ phage, enabling to serve as vectors in conjunction with the λ phage *in vitro* packaging system (Hohn and Murray 1977). The cosmid vector containing the particular DNA insert is packaged into λ phage particles and resulting phages are subsequently used for transfection of a bacterial host strain for propagation. Cosmids bear the advantages of high transfection rates and narrowly defined sizes of clone inserts. The major disadvantage of cosmids is their instability *in vivo*, i.e. they undergo drastic rearrangements and deletions, due to their high copy number. To overcome this drawback modified cosmids, termed fosmids, were developed by replacing the high copy origin of replication with an *Escherichia coli* F-factor single-copy origin of replication, which permits insert stability (Kim et al. 1992). Modern fosmid vectors, like pCC1FOS (Epicentre Biotechnologies, Madison,WI), usually also contain an additional inducible high-copy origin of replication, which facilitates high DNA yields along with insert stability.

The yeast artificial chromosomes (YACs) systems permits cloning of exogenous DNA fragments up to 3000 kbp into linear artificial chromosomes that are maintained in yeast (*Saccharomyces cerevisiae*) (Murray and Szostak 1983; Burke et al. 1987). YACs permit cloning of the largest DNA inserts compared to other vector system, however they posses several disadvantages. On the one hand the cloning procedure of YACs is complicated, leading to laborious production of a representative YAC library. Also preparation of YAC-DNA is difficult, since the artificial chromosome has to be separated from the background of yeast chromosomes. Moreover, chimeric YAC clones are frequently observed and recombination can occur within a YAC clone resulting in rearrangements or interstitial deletions.

The most widely used vector systems for large insert libraries nowadays are the bacterial artificial chromosome (BAC) system based on *E. coli* and its single-copy plasmid F-factor (Shizuya et al. 1992) and the P1 artificial chromosome (PAC) system, combining the features of bacteriophage P1-derived and F-factor based approaches (Ioannou et al. 1994). Both systems include the ability to maintain inserts up to 300 kbp along with high degree of structural stability in the bacterial host.

Besides these prevalent vector systems, there also exist several further vectors with specialised features for certain needs, e.g. the transformation-competent artificial chromosome (TAC) vector that can accept and maintain large genomic DNA fragments stably in both *E. coli* and *Agrobacterium tumefaciens* and it has the *cis* sequences required for *Agrobacterium*-mediated gene transfer into plants.

Table 2 : Vector systems for large insert library production

Vector	Insert size	Advantages (+)/ disadvantages (-)
Cosmid	35 – 45 kbp	+ high transformation efficiency + narrowly defined sizes of clone inserts - instability of clones - comparatively small insert size
Fosmid	35 – 45 kbp	+ high transformation efficiency + narrowly defined sizes of clone inserts + stable maintenance of insert DNA + high DNA yield with inducible origin of replication - comparatively small insert size
YAC	90 – 2000 kbp	+ largest insert size of all systems - difficult cloning and handling - instability of clones
BAC/PAC	70 – 300 kbp	+ relatively large insert size + stable maintenance of insert DNA - low DNA yield due to low copy origin of replication

4.4.2.2.2 Techniques for ordering of large insert clones

No matter which large insert vector library is utilised, the task of constructing a whole-genome physical map includes ordering of the clones by applying either fingerprinting methods or techniques based on screening for marker contents.
Fingerprinting methods include the digestion of large insert clones with restriction enzymes and subsequent analysis of the DNA fragments. Clones containing overlapping

DNA inserts, i.e. originating from the same genomic location, produce shared banding patterns on gels. The degree of overlap is indicated by the proportion of shared bands, thus the analysis of the overlap of numerous clones allows to built contigs (Staden 1980).

Classical fingerprinting techniques provide the basis for modern advanced fingerprinting applications. For construction of a physical map of *Caenorhabditis elegans*, whole-genome fingerprinting was utilised for the first time (Coulson et al. 1986). In this study, radioactively labelled fragments of cosmid clones were size separated on denaturing polyacrylamide gels. The fragments were produced by first performing restriction digestion with a rare cutter (i.e. 6-bp specificity), followed by radioactively labelling and subsequent digestion with a frequent cutter (i.e. 4-bp specificity) producing fragments appropriate for separation on polyacrylamide gels. Only a subset of fragments with a labelled rare-cutter-end can be detected. Different fingerprinting methods were developed, such as agarose fingerprinting (Olson et al. 1986), involving digestion of large-insert clones with a rare cutter and analysis of the fragments on agarose gels. Thus, almost all fragments are taken into consideration, allowing direct size estimation of the overlapping region and a reliable detection of rearranged clones. Several modifications adapted for increased throughput, accuracy and information content have been proposed. An important step was the adaptation of fingerprinting methods for use on automated sequencers and multiplexing (Gregory et al. 1997; Ding et al. 2001) by labelling fragments from different clones with different fluorescent dyes allowing simultaneous analyses in a single lane. Further on high-throughput pipelines combining the use of several restriction enzymes and capillary sequencers (Luo et al. 2003) were built, usually referred to as high information content fingerprinting (HICF). Until now, fingerprinted BAC based physical maps of whole genomes have been generated for many plant species, including rice (Zhang and Wing 1997; Tao et al. 2001), sorghum (Klein et al. 2000), soybean (Wu et al. 2004), maize (Nelson et al. 2005) , poplar (Kelleher et al. 2007) and recently papaya (Yu et al. 2009). The fingerprinting approach is well suited for relatively unexplored genomes and includes many possibilities for its adaptation to high-throughput. However, repetitive elements are a major source for producing false overlap using fingerprinting methods, since they result in similar banding patterns for clones originating from different genomic locations. Further major drawbacks of fingerprinting techniques are the anonymous character of produced clone contigs and the impossibility to detect small overlaps, in order to prevent a high rate of falsely as-

sembled contigs. Furthermore, maps can only be constructed based on a single, preferably homozygous accession, because haplotypic variation may result in the construction of two separate contigs for a given genomic region (see e.g. Kelleher et al. 2007).

"Screening for content" methods include PCR amplification of sequence tagged sites (STSs) and hybridisation based approaches. STSs are short stretches of DNA that can be specifically detected for establishing physical maps. If two STSs are located close to each other in the genome, there is a high chance of finding them together on the same DNA fragment or clone, respectively, when screening a genomic library. The further apart the positions of two STSs in the genome, the lower are the chances of finding them on the same fragment. Thus, the obtained screening data can be used to order clones according to their genomic location. Most commonly used sequence sources for STSs include expressed sequence tags (ESTs), simple sequence length polymorphisms (SSLPs) and random genomic sequences. STSs have first been employed for establishing physical maps of human chromosomes using PCR (Olson et al. 1989; Green et al. 1991). The development of systematic pooling strategies, where clone libraries are condensed in pools, allowed the screening of large libraries (Green and Olson 1990; Bruno et al. 1995). These approaches usually require a large number of PCRs to address one target sequence. Hybridisation based methods follow a different strategy. In general, all different variations of hybridisation based methods involve hybridization of labelled DNA probes to high density arrays of clones or DNA. Hybridisation probes may be labelled PCR amplicons, cDNA clones, overgo probes or short oligonucleotides. DNA sequence information is not necessarily required to develop probes. Systematic pooling of the probes before hybridisation to the arrayed clones or DNA, leads to a manifold reduction of required hybridisation experiments. Subsequently, individual probes can be assigned to clones by performing a deconvolution step. In comparison to PCR based approaches, hybridisation screening is faster and cheaper, because the entire library may be screened in one step, probe construction is less costly and multiple probes can be addressed in one step. Using PCR amplicons or cDNA as probes often bears the disadvantages of uneven labelling of all probes in one pool and the presence of repetitive elements in the labelled probes. When sequence data is available, hybridisation can be carried out using overgo probes. Overgo probes are 40-mer probes produced by designing a pair of 24-mers with an overlap of 8 bp from the target sequence. Radioactive nucleotides are then incorporated at the resulting 16-bp overhangs using Klenow fragment (Cai et al. 1998; Ross 1999). Due to their increased specific activity, sequence specific-

ity and small size, they are likely to hybridise in a locus specific manner and the chances of including repetitive sequences are reduced.

Rather than construction of physical maps for whole genomes solely based on overgo probe hybridisation, overgo probes have often been used in combination with maps obtained by fingerprinting techniques to increase the genome coverage (Cai et al. 2001), anchor fingerprinting derived contigs to genetic maps (Chen et al. 2002; Xu et al. 2008) or to perform comparative genomics by linking overgo probes conserved between closely related species to a physical map obtained by fingerprinting approaches (Hass-Jacobus et al. 2006; Yu et al. 2009).

An alternative but similar hybridisation approach is the use of labelled 35mer-oligonucletides as probes, which also allows conversion of any STS, genetically mapped marker or BAC end-fragment, for which sequence information exists, into a marker (Khorasani et al. 2004). Pooling and 5'-end labelling with polynucleotide kinase of the 35-mer oligonucleotides allow uniform activity of each probe in a pool and performance of hybridisation assays in a very straightforward high-throughput manner.

As described for overgo probes, a combination of fingerprinting maps and hybridisation based maps is of great value, since advantages and disadvantages of each method can substitute each other.

4.5 Aim of this work

Whole genome physical maps or genomic sequences, respectively, of a species display frameworks that provide essential information for understanding processes in respect to physiology, morphology, development and genetics. However, comprehensive annotation underpins the values a genome sequence represents. An important task of genome annotation is the linkage of genetic traits to the genome sequence, which is facilitated by integrated genetic and physical maps. An integrated map allows map based cloning of important genes, analyses of different varieties or species, QTL dissection and cloning, MAS and studies on genome structure. Sugar beet is an agronomically important plant, since it is the only sucrose storing crop of moderate climates. However, genomic resources and knowledge about its genome structure and evolution are limited. Prior to the studies included in this work, only medium dense genetic maps of sugar beet existed, including EST and RFLP- derived single SNP markers as well as microsatellite markers (Schumacher et al. 1997; Laurent et al. 2007; Schneider et al. 2007). In addition, several publicly accessible BAC libraries (Gindullis et al. 2001; Hohmann et al. 2003; Mcgrath et al. 2004; Hagihara et al. 2005; Jacobs et al. 2009), about 29,000 EST sequences, most of them originating from the study of Herwig et al. (2002) and about 25,000 BES in public databases were available. Yet, no high-density genetic map or a comprehensive physical map was available. This work aimed at generating and applying new tools, representing essential prerequisites for sugar beet genome analysis and providing new insight into evolution and genome structure. In addition, its goal was to facilitate successful sequencing, assembly and annotation of the sugar beet genome in the near future. Since it would be the first genome sequence originating from a member of the order *Caryopyllales*, it would be highly beneficial for comparative and evolutionary studies. A new method for generation of genetic markers in sugar beet was establishing with the potential to be linked to the physical map, constructed utilising a hybridisation based approach introduced in this work. In addition, fosmid and small insert libraries were produced and characterised supporting sequence assembly and facilitating repeat identification.

5 Publication I

5.1 High-throughput identification of genetic markers using representational oligonucleotide microarray analysis

Lange C*, Mittermayr L*, Dohm JC, Holtgräwe D, Weisshaar B, Himmelbauer H. Theor Appl Genet. 2010 Apr 9. [Epub ahead of print]
DOI: 10.1007/s00122-010-1329-2
Originalveröffentlichung erschienen auf www.springerlink.com
The original article is online available at:
http://www.springerlink.com/content/x111g83371335444

High-throughput identification of genetic markers using representational oligonucleotide microarray analysis

Cornelia Lange*[1], Lukas Mittermayr*[1, #], Juliane C. Dohm[1, 2], Daniela Holtgräwe[3], Bernd Weisshaar[3] and Heinz Himmelbauer[1, 2, †]

(*) these authors contributed equally to this work

(1) Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany

(2) Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra, 08003 Barcelona, Spain

(3) University of Bielefeld, Center for Biotechnology (CeBiTec), 33594 Bielefeld, Germany

(#) Present address: Department of Biology, Ludwig-Maximilians-University Munich, Großhaderner Strasse 2, 82152 Planegg-Martinsried, Germany

(†) To whom correspondence should be addressed to

++34 - 933 160 243

heinz.himmelbauer@crg.es

Keywords: Genetic linkage maps, DNA polymorphism, genomic representations, oligonucleotide microarray, genetic markers, sugar beet (*Beta vulgaris* L.)

Sequence data have been submitted to GenBank (accession numbers GS923105-GS923388)

Abstract

We describe a novel approach for high-throughput development of genetic markers using representational oligonucleotide microarray analysis (ROMA). We test the performance of the method in sugar beet (*Beta vulgaris* L.) as a model for crop plants with little sequence information available. Genomic representations of both parents of a mapping population were hybridized on microarrays containing in total 146,554 custom made oligonucleotides based on sugar beet bacterial artificial chromosome (BAC) end sequences and expressed sequence tags (ESTs). Oligonucleotides showing a signal with one parental line only, were selected as potential marker candidates and placed onto an array, designed for genotyping of 184 F_2 individuals from the mapping population. Utilizing known co-dominant anchor markers we obtained 511 new dominant markers (392 derived from BAC-end or BAC sequences, respectively and 119 from ESTs) distributed over all nine sugar beet linkage groups and calculated genetic maps. Further improvements for large-scale application of the approach are discussed and its feasibility for the cost-effective and flexible generation of genetic markers is presented.

Introduction

High-density genetic maps are essential tools for crop plant improvements. They facilitate the detection of quantitative trait loci (QTLs), the characterization of QTL effects and, when integrated with physical maps, enable the map based cloning of genes underlying QTLs. For precise transfer of QTLs between different genetic backgrounds, high density of genetic markers is crucial due to the need of polymorphic markers immediately flanking QTLs (Somers et al. 2004). Also linkage disequilibrium (LD) maps and association mapping require dense genetic maps (Bernardo et al. 2009). Genetic markers linked to genes and QTLs provide the framework for marker assisted selection (MAS), which is a very promising approach to accelerate line development in breeding programs (reviewed in: Collard et al. 2005; Collard and Mackill 2008; Ribaut and Hoisington 1998). Increasing availability of sequence resources for several crop plants has led to great advances in marker assisted breeding approaches. However, complete or draft, respectively, genome sequences exist only for a few crops, such as rice (*Oryza sativa*) (International Rice Genome Sequencing Project 2005), grapevine (*Vitis vinifera*) (Jaillon et al. 2007), papaya (*Carica papaya*) (Ming et al. 2008), sorghum (*Sorghum bicolor*) (Paterson et al. 2009), potato (*Solanum tuberosum*) (www.potatogenome.net), soybean (*Glycine max*) (Schmutz et al. 2010)) and cucumber (*Cucumis sativus*) (Huang et al. 2009). Thus, there is high demand for high-throughput, cost-effective marker technologies for crops with little sequence information available. Here, we focus on sugar beet (*Beta vulgaris* L.), a diploid species encompassing n=9 chromosomes and a haploid genome size of 758 Mbp (Arumuganathan and Earle 1991). Taxonomically, *B. vulgaris* is a member of the core eudicot plants and belongs to the order of Caryophyllales (APG 2009). As is the case for many crop plants, it is of high economic importance, but publicly available sequence resources are limited. At present, the GSS database of GenBank holds approximately 3000 end-sequences from sugar beet fosmid clones (Lange et al., 2008), and about 28,000 end-sequences from sugar beet BAC library USH20 (McGrath et al., 2004). Roughly 30,000 sugar beet EST sequences have been deposited in GenBank. Genomic sequencing of the sugar beet genome is under way in a collaborative effort by the authors of this paper (www.gabi.de).

Over the past years several molecular marker technologies for genetic mapping have been developed. One of the earliest technologies used on the DNA level was restriction fragment length polymorphism (RFLP) scoring (Botstein et al. 1980). With the inven-

tion of PCR, marker systems such as simple sequence repeat (SSR) (Weber and May 1989), random amplified polymorphic DNA (RAPD) (Williams et al. 1990) and amplified fragment polymorphism (AFLP) (Vos et al. 1995) followed. Subsequently, modifications of these mapping systems were developed in order to obtain performance improvements in terms of efficiency and reliability. One approach for identification and mapping of polymorphic markers in mouse established by Himmelbauer et al. (1998) included complexity reduction of genomic samples by performing AFLP prior to hybridization against a reference BAC library gridded at macroarrays. The concept of reducing the complexity of a genomic sample by producing genomic representations was originally introduced by Lisitsyn et al. (1993). They presented a method termed representational difference analysis (RDA) built upon subtractive hybridization techniques for identifying sequence differences between two DNA populations. RDA includes digestion of genomic DNA with restriction endonucleases, ligation of the resulting fragments to oligonucleotide adapters, followed by PCR amplification. Shorter restriction endonuclease fragments are preferentially amplified by *Taq* polymerase during PCR, resulting in genomic representations with reduced nucleotide complexity. The decreased complexity of the representations allows to achieve greater completeness during subtractive enrichment and, hence, a more effective kinetic enrichment. With ongoing progress in miniaturization of arrays, approaches using microarrays in combination with genomic representations were developed for analysis of copy number variations in the context of cancer (Lucito et al. 2000). A similar approach was used by Lezar et al. (2004) for fingerprinting in *Eucalyptus grandis*. For hybridization based methods the advantages of complexity reduction rests mainly in the lower noise to signal ratio, since opportunities for cross-hybridization are reduced, thus obtaining greater intensities for specific signals on the arrays (Kennedy et al. 2003). In addition, low amount of input material is needed per experiment. A technique that evolved from RDA is representational oligonucleotide microarray analysis (ROMA) that was established for the detection of structural variation in cancer and healthy tissue in a high-throughput profiling manner (Lucito et al. 2003). Whilst Lucito et al. (2000) and Lezart et al. (2004) initially applied microarrays of fragments from representations as probes to analyze genomic representations, microarrays of oligonucleotides were adopted for ROMA, thus representing a very flexible and reproducible method compatible with high-throughput applications. ROMA was further utilized in several studies for genome wide analysis of copy number variants in humans (Sebat et al. 2004) and analysis of copy number variants in

cancer tissue (Grubor et al. 2009; Hicks et al. 2006; Lakshmi et al. 2006; Stanczak et al. 2008).

Existing *Beta vulgaris* genetic maps covering all nine chromosomes include expressed sequence tag (EST)- and RFLP- derived single nucleotide polymorphism (SNP) markers as well as microsatellite markers (Laurent et al. 2007; Schneider et al. 2007; Schumacher et al. 1997). However, due to the limited sequence resources, no high-density genetic map is available for sugar beet so far.

In this study we explore and demonstrate the potential of ROMA for high-throughput, cost-effective and flexible development of genetic markers in crop plants. We apply ROMA for the identification of polymorphisms between two accessions of sugar beet (*Beta vulgaris* L.) and discuss further improvements. The information gained in this study will facilitate the production of similar platforms for other species.

Materials and methods

Plant material and DNA isolation

For array based genotyping we chose 196 F_2 individuals and both parents of the "K1" mapping population (kindly provided by B. Schulz, KWS SAAT AG, Einbeck, Germany). One parent of this mapping population was K1P1 (KWS2320), a German double haploid monogerm breeding line and the other parent was K1P2, a partly selfed line. The F_2 genotypes were generated by selfing of F_1 individuals (K1F1). A subset of the K1 mapping population was also used in the studies of Mohring et al. (2004) and Schneider et al. (2007). Genomic DNA was isolated from plant material cultivated *in vitro*. Briefly, young plants 3 – 5 cm in size were harvested, flash frozen in liquid nitrogen and stored at -80°C before DNA isolation. 100 – 200 mg frozen plant material was ground with 5 mm stainless steel beads (Qiagen, Hilden, Germany) using the TissueLyser (Qiagen) for 45 sec at 30 Hz. Subsequently, 1.3 ml hot (65°C) extraction buffer (0.1 M TrisHCl; 0.7 M NaCl; 0.05 M EDTA; pH 8) was added to the ground material, followed by incubation at 65°C for 15 min with repeated shaking. Genomic DNA was then purified from the lysate by extraction with phenol-chloroform. Remaining RNA was digested using 10 µl RNAse A (10 µg/µl) for 10 min at 37°C and the DNA was precipitated with isopropanol, followed by a wash-step with 70% ethanol. Finally, the dried pellet was dissolved in 100 µl TE-buffer (10 mM Tris-HCl; 1 mM EDTA; pH 8).

Amplicon generation

Amplicons were generated as described by Lucito and Wigler (2003) with slight modifications. Restriction digests were carried out in a reaction volume of 30 µl with 120 ng genomic sugar beet DNA, 20 U *Bam*HI (New England Biolabs, Ipswich, MA), 20 U *Bgl*II (New England Biolabs), 1× digestion buffer (New England Biolabs) and 1× BSA (New England Biolabs) followed by incubation overnight at 37°C. Completeness of the digestion was monitored by gel electrophoresis. In order to enable amplification of the fragments, adaptors were ligated to the protruding 5'-termini of the digested DNA. The adaptors consisted of a 24-mer oligonucleotide (5'-AGCACTCTCCAGCCTCTCACCGCT-3') and a partly complementary 12-mer (5'-

GATCAGCGGTGA-3‘), of which 7.5 µl each (62 µM) were added to 10 µl (40 ng) of digested DNA, 3 µl 10× T4 DNA ligase reaction buffer (New England Biolabs) and ddH_2O in a reaction volume of 29.5 µl. After heating to 55°C and slow cooling of the mixture to room temperature, 0.5 µl of T4 DNA ligase (400 U/µl, New England Biolabs) was added, followed by incubation at 16°C over night. Next, 3 µl (4 ng DNA) of the ligation reaction were used as PCR template with 0.4 µl dNTPs (100 mM), 3 µl 24-mer adaptor (62 µM) acting as primer, 2 µl *Taq* polymerase (5 U/µl), 6 µl 10× PCR buffer (481 mM KCl; 0.96% Tween 20; 14 mM $MgCl_2$; 337 mM Tris-base; 144 mM Tris-HCl; 1.44% cresol red) and 45.6 µl ddH_2O. PCR was performed with an initial elongation step at 72°C for 5 min to replace the 12-mer adaptor and fill in the recessive 3’-termini. Afterwards a denaturation step at 94°C for 4 minutes was performed, followed by 25 cycles consisting of 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 3 minutes, and a final elongation step at 72°C for 10 minutes. Finally the PCR products were purified using QIAquick PCR Purification Kit 50 (Qiagen) and QIAquick 96 Purification Kit (Qiagen) according to the manufacturer’s instructions.

Oligonucleotide design for microarrays

Custom oligonucleotides for the 44K and 105 K arrays were generated from two genomic BAC end data sets and one EST data set. 29,320 end sequences (Weisshaar et al., unpublished) from the sugar beet BAC clone library “ZR/KIEL” (genotype: KWS2320; Hohmann et al. 2003) and 25,850 end sequences from the sugar beet BAC clone library USH20 (McGrath et al. 2004) (NCBI database of Genome Survey Sequences; http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucgss) were searched for *Bam*HI and *Bgl*II restriction sites using the “restrict” program of the EMBOSS suite (Rice et al. 2000). In 6486 ZR BAC end sequences 1-12 restriction sites were found, and in 6493 USH20 BAC ends 1-9 restriction sites were found (either *Bam*HI or *Bgl*II). Perl scripts and the EMBOSS programs “seqret” and “extractseq” were used to extract the subsequences between restriction sites and to select fragments such that sequences of a length below 80 bp were discarded, sequences of lengths 80 – 200 bp were kept, and sequences of length above 200 bp were split into two parts. The resulting 20,759 fragments from the ZR BAC end data set, 21,882 fragments from the USH20 BAC end data set and 22,834 BAC sequences from the ZR BAC end data set containing no *Bam*HI or *Bgl*II restriction sites were repeat masked applying RepeatMasker (Smit et al. 1996-2004)

with a repeat library containing sugar beet specific repeats (from the NCBI nucleotide database http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore) and other known plant repeats (RepeatMasker inherent or downloaded from NCBI nucleotide database). The masked sequences were sent to the Agilent web page (https://earray.chem.agilent.com/earray) for the design of 60-mer oligonucleotides. In order to exclude repetitive oligonucleotides we searched against the BAC end data sets using BLASTN (Altschul et al. 1990) (-e 1e-5, -F F) and discarded oligonucleotides which matched more than two times. In addition, oligonucleotides based on two sugar beet BAC clone sequences, SBI-153H13 and ZR-47B15 (GenBank FJ752586 and FJ752587) (Dohm et al. 2009), were constructed in the same way.

For the design of oligonucleotides from ESTs we downloaded 22,209 publicly available nuclear sugar beet EST sequences from the NCBI nucleotide database. The sequences were repeat masked and clustered with the cap3 algorithm (Huang and Madan 1999) with parameters P (overlap percent identity) = 95, o (overlap length cutoff) = 50 and h (max. overhang percent length) = 100 resulting in a non-redundant data set of 14,517 EST sequences. This data set was compared with genomic sequences of *A. thaliana* (downloaded from the NCBI database), *Populus trichocarpa* (downloaded from http://genome.jgi-psf.org/Poptr1/Poptr1.download.ftp.html) and *O. sativa* (downloaded from http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml) on the protein level using TBLASTX (-S 1 and -e 1e-5). The mRNA-to-genomic alignment program spidey (Wheelan et al. 2001) with parameters -s T and -r p was applied for every pair of homologous sequences between *B. vulgaris* and *A. thaliana*, *P. trichocarpa*, or *O. sativa*, respectively, controlled and parsed by Perl scripts. In total 9764 *B. vulgaris* sequences had matches with *A. thaliana*, 9634 with *P. trichocarpa*, and 9244 with *O. sativa*. According to the match positions we extracted the subsequences from the ESTs using Perl scripts and the Emboss "exractseq" program. We removed subsequences shorter than 80 bp, reverse matching sequences, and single-exon sequences in cases where another homologous gene sequence with more than one exon for the same *B. vulgaris* EST sequence existed. Overlapping matches with different genes for one *B. vulgaris* EST sequence were combined. *Bam*HI and *Bgl*II restrictions sites were masked with "N"s, and 60-mer oligonucleotides were designed for each exon sequence at the Agilent web site. Oligonucleotides with a length of 30 bp were built from the central part of each 60-mer.

Microarray design

We used three different custom gene expression microarray formats: 4×44K (Agilent, Santa Clara, CA, USA) and 2×105K (Agilent) for screening of the parental genotypes and 8×15K (Agilent) for genotyping of the F_2 individuals. Gene expression arrays were preferred to comparative genome hybridization (CGH) arrays, since expression arrays contained more positions for custom made features. Previously designed custom oligonucleotides were placed onto the 4×44K and 2×105K arrays using the Agilent eArray platform. Features identified as being polymorphic between both parental lines in the course of this study were selected for the design of a 15K array. In addition to the polymorphic features, control features complementary to *Bgl*II/*Bam*HI double digest fragments of mouse BAC clone RP24-571N6 (GenBank: AC102017) were placed onto the 15K array. For the design of these control features, we performed *in silico* restriction digestion of the BAC clone sequence using NEBcutter V2.0 (Vincze et al. 2003) with *Bgl*II and *Bam*HI prior to repeat masking of the BAC sequences using RepeatMasker. Thereafter, appropriate feature sequences for each fragment in the size range of 290 – 6319 bp were selected using the Agilent eArray platform and placed onto the 15K array in fivefold replicates.

Amplicon labeling and array hybridization

In case of the F_2 samples for the 15K arrays, 247 ng of BAC clone RP24-571N6, double digested with *Bam*HI/*Bgl*II and amplified as described above, was spiked-into each sample before labeling as hybridization control. Labeling and hybridization were performed according to Agilent protocols. Briefly, amplicon samples were labeled with Cyanine 3-dUTP by random priming (Agilent Genomic DNA Labeling Kit Plus) at 37°C for 2 hr followed by heat inactivation at 65°C for 10 minutes. The recommended amounts of DNA template for labeling varied between the different array formats and were 500 ng for the 15K array, 1 µg for the 44K array and 1.5 µg for the 105K array. Labeled products were purified using Microcon YM-30 filters (Millipore, Billerica, MA) and if necessary 1×TE-buffer (10 mM Tris-HCl; 1 mM EDTA; pH 8) was added to the final hybridization volume (18 µl for the 15K array, 44 µl for the 44K array and 104 µl for the 105K array). Specific labeling activity (pmol dye / µg DNA) of the samples was examined using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Nano-

Drop Technologies, Rockland, DE). The recommended specific activity after labeling and clean-up was 25 – 40 pmol/µg. The Agilent Oligo aCGH Hybridization Kit was used for hybridization. Samples were prepared according to the manufacturer's protocol and hybridized with 10 rpm at 65°C for 24 hrs.

After two washing steps with Oligo aCGH wash buffers 1 and 2 (Agilent) the arrays were immediately scanned with the DNA microarray scanner G2505B (Agilent) at a wavelength of 532 nm and with 5 µm resolution.

Data analysis

We analyzed the scanned microarray images (.tif) using the Agilent Feature Extraction software (version 9.1.3.1 for 44K and 105K arrays; version 9.5.3.1 for 15K arrays) applied on the individual grid file for each array format and the Agilent GE1-v5_91_0806 protocol (44K and 105K arrays) and GE1-v5_95_ Feb07 protocol with enabled "Local background method" (15K arrays). For 44K and 105K arrays, signal thresholds separating positive signals from negative ones valid for all features on one array were determined manually by setting a threshold at which a weak optical signal was visible. Each feature signal on the particular arrays was divided by the determined threshold intensity and resulting signals above one were scored as present, and signals below one were scored as absent. Features having a signal in K1P1 and no signal in K1P2 or vice versa were placed onto the 15K array as polymorphic marker candidates. For normalization of the 15K arrays, signals of control feature groups, i.e. oligonucleotides complementary to one fragment of BAC clone RP24-571N6 present in five replicates were utilized. The average signal values of each control feature group on one array were summed up representing the normalization value. Subsequently, all feature values on one array were divided by the related normalization value.

Seventy-eight features on the array were based on 50 source sequences that had previously been used by Schneider et al. (2007) for marker development. By comparison of these features' scoring results with different thresholds to their known scoring results from Schneider et al. (2007), criteria for scoring the signal as absent or present for each feature were determined individually. The conclusive criteria were: (1) only features with a normalized signal value > 10 were scored; (2) a signal value was scored as positive when larger than 2.5 times the lower quartile and scored as negative when smaller than the lower quartile minus 10% of the lower quartile, signal values between these

thresholds were considered as missing genotypes ; (3) the number of genotypes scored as positive or negative (not missing) had to be more than 133 (72%); (4) only features with no significant deviation ($\chi^2 \leq \chi^2_{\alpha=0.05}$) from the expected 3:1 (signal: no signal) ratio, were included into further analysis. Before map calculation, an additional masking step was performed using RepeatMasker with all *Virdiplantae* specific repeats, *B. vulgaris* chloroplast (GenBank EF534108) and mitochondrial (GenBank NC_002511) sequences. Furthermore, features with more than one BLASTN hit (-e 1e-09) against the "nr" database or more than two hits (-e 1e-09) against the "gss" database were excluded. Based on the signal scores of the parental lines K1P1 and K1P2 on the 44K and 105K arrays, respectively, the marker scores were translated into A (homozygous K1P1); B (homozygous K1P2); C (known to be not homozygous A) and D (known to be not homozygous B). Due to the dominant character of the markers, heterozygous individuals could not be determined.

Map calculation

We calculated genetic maps using AntMap version 1.1 (Iwata and Ninomiya 2006) and performed grouping with the nearest neighboring locus option. We chose a LOD score of 12 or greater in order to minimize the number of falsely grouped markers. Groups known to be located on one linkage group based on previously mapped co-dominant markers were joined. Recombination percentage was converted to genetic distance by the Kosambi map function (Kosambi 1944) with optimization of locus ordering by minimizing the sum of adjacent recombination fractions (SARF) (Falk 1989) and with default parameters of AntMap Ant Colony Optimization. Thirty runs of locus ordering were performed. Linkage maps were plotted using the software MapChart 2.2 (Voorrips 2002) with post processing, i.e. adjustment of lines connecting homolog loci between linkage groups, applying an image processing software.

Results

Genomic representations

Genomic representations are reproducible subpopulations of genomic DNA in which the resulting sample has a new format, or reduced complexity, or both (Lisitsyn and Wigler 1993; Lucito et al. 1998). The complexity reduction leads to improved hybridization kinetics compared to that of the complete genome. In order to achieve a complexity reduction we digested the genomic DNA with endonucleases, ligated primers to the resulting fragments and amplified these by PCR, thus producing amplicons. *Taq* polymerase can amplify fragments up to approximately 2000 bp (Saiki et al. 1988). Within a mixture of differently sized templates, PCR preferentially generates products in the size range of 200 – 1200 bp. Hence, larger fragments will not be effectively amplified and produce no signals by hybridization on an array containing oligonucleotides complementary to subparts of the fragments. By scoring of presence or absence of fragments in representations from both parents of a mapping population, polymorphic marker candidates can be determined. Subsequent hybridization of representations from F_2 individuals of the mapping population on arrays containing the polymorphic marker candidates allows genotyping of the F_2 individuals (strategy outline: Fig. 1).

The complexity reduction rate depends predominantly on the choice of restriction enzymes and their cutting frequency, respectively. We wanted to achieve a complexity reduction to approximately 10% of the sugar beet genome. For determining suitable restriction endonucleases we performed *in silico* restriction enzyme double-digestion with different enzymes utilizing two genomic sugar beet sequences (BAC clones SBI-153H13 and ZR-47B15) and calculated the percentage of fragments in the range of 200 – 1200 bp. The restriction with both *Bgl*II and *Bam*HI led to a predicted amplifiable proportion of 7 – 11% of the DNA. In order to experimentally evaluate the preferred size range of the *Taq* polymerase, we digested the DNA of the two sugar beet BAC clones SBI-153H13 and ZR-47B15 with *Bgl*II and *Bam*HI and analyzed the amplified fragments by gel electrophoresis (Fig. 2). Fragments within an approximate size range of 250 – 1500 bp were preferentially amplified, suggesting that genomic amplicons generated by digestion with *Bgl*II and *Bam*HI represent 14 – 15% of the sugar beet genome.

Array design for identification of polymorphic markers

To identify polymorphic markers, labeled genomic representations of the P1 and P2 parental lines were hybridized on Agilent 44K and 105K custom microarrays.
The Agilent 44K array contained 45,220 oligonucleotide positions, named features, of which 1428 were structural controls, thus 43,792 custom features could be placed onto the array. We designed 21,720 oligonucleotides based on BAC-end sequences (BES), 21,720 based on ESTs and 352 based on two BAC sugar beet sequences (SBI-153H13 and ZR-47B15). Apart from structural controls, the 105K array provides space for 102,762 custom features. Of these, 79,506 were designed based on BES (45,980 with *Bam*HI or *Bgl*II restriction site and 33,526 without), 23,116 based on ESTs and 140 based on the sugar beet BAC sequence FJ752587. In total we designed 146,554 custom features based on sugar beet ESTs and BAC sequences, respectively, that were available in GenBank and the GABI beet physical map consortium. The standard oligonucleotide length for Agilent arrays is 60 nt. In order to test the hypothesis of previous studies (Castle et al. 2003) showing 30-mers to be more sensitive, we designed 50% of the 146,554 oligonucleotides as 60-mers and 50% as 30-mers representing sub-fragments of each 60-mer. The distribution of the origins of oligonucleotide sequences on the 44K and 105K arrays is shown in Fig. 3a.

Scoring of K1P1 and K1P2: Selection of polymorphic markers for 15K oligonucleotide array

After hybridization of the labeled K1P1 and K1P2 amplicons, respectively, on the 105K and 44K arrays, the signals were scored as positive or negative. At this stage of our study, thresholds for signal scoring were determined by visual evaluation, and one signal value was defined for all features on one array.
Oligonucleotides giving a positive signal for K1P1 but no signal for K1P2 and vice versa were selected as potential polymorphic markers and were used for the design of a 15K array allowing the placement of 15,160 features onto the array. We used 245 positions for hybridization controls (described below), thus 14,915 oligonucleotides identified as potentially polymorphic before could be selected for the 15K array. Of these polymorphic marker candidates 83% showed a positive signal in K1P1 and 17% a positive signal in K1P2 This bias towards positive signals is probably due to the initially

used simple method for discrimination between signals or no signal, i.e. setting the same thresholds for all features on one array based on the visual impression of a weak optical signal. Criteria for scoring the signals as absent or present for each feature were optimized for scoring of the 15K arrays later on (see material and methods section). The distribution of 30-mers and 60-mers and their source sequence origin is shown in Fig. 3b. Even though the numbers of 30-mer and 60-mer probes were equal on the 44K and 105K arrays, 72% of the potential polymorphic markers were 60-mers, indicating that 60-mers are better suited for the detection of polymorphisms using ROMA.

Analysis of 15K oligonucleotide arrays with internal control features

For genotyping, genomic representations of 196 F_2 individuals from the K1 mapping population were hybridized on the 15K arrays containing the polymorphic marker candidates. Results were obtained only for 184 F_2genotypes, which were further analyzed. Oligonucleotides, one for each fragment of the mouse BAC clone RP24-571N6 produced by *Bgl*II and *Bam*HI digestion, served as an internal control. Five replicates of each of these control oligonucleotides were scattered across the 15K array, thus comparing the signal intensities allowed verification of even hybridization throughout an array. We selected 184 arrays with uniform hybridization results for further analysis. Since BAC clone RP24-571N6 was digested and amplified in the same way as the genomic sugar beet DNA and the resulting PCR products were spiked-into the genomic-representation samples of each genotype before labeling, the mouse BAC control oligonucleotides also provided verification of the performance of our method. Fig. 4 shows the distribution of the control oligonucleotides' signals from all 184 F_2 genotypes. As expected, large fragments did not produce signals on the array, since they could not be amplified by *Taq* polymerase. This observation confirms the feasibility of our method. However, we also observed control fragments within the amplifiable size range (C11: 290 bp and C30: 512 bp; Fig 4), which did not show hybridization signals. This may indicate some biases in the generation of amplicons other than size exclusion, for instance base composition of fragments. Another obvious conclusion from the results in Fig. 4 was the need for setting individual score thresholds for each feature, since the range of signal intensities varied largely between distinct features, presumably due to nucleotide composition.

Scoring of signals on the 15K array and selection of markers for map construction

In order to find general criteria for setting individual scoring thresholds for each feature, we utilized scoring data from markers mapped in a subset of the same mapping population K1 (Schneider et al. 2007). Seventy-eight features on the array were based on 50 source sequences that had previously been used by Schneider et al. for marker development. We determined the optimal scoring parameters by comparing the array-based scoring results for these features to the scoring results from Schneider et al. (2007). Based on signal intensities and the deviation from the expected ratio of signal to no signal (see Materials and Methods) individual thresholds were selected resulting in the least possible number of false positives and negatives. Applying these criteria, 1204 features were selected from the 15K array (Fig. 3c), of which the ratio of 60-mers (76%) to 30-mers (14%) was almost the same as it was on the whole 15K array. Within the 60-mers the proportion of oligonucleotides derived from BAC sequences increased from 52% to 62%. An additional masking step led to the removal of 30 features. After merging of features derived from the same locus, i.e. originating from the same BAC sequence or EST, we obtained 873 final marker candidates for genetic map construction. Six hundred-eighty-nine (79%) of these were derived from BAC sequences (621 with *Bam*HI or *Bgl*II restriction site and 68 without), and 184 (21%) from ESTs. The merging step provided an important verification step, since features derived from the same locus with discordant scoring results of more than 3% were discarded from the data set. The fraction of polymorphic markers having a positive signal in K1P1, constituted 82% (716), which reflects the fraction of K1P1 positive features being present on the whole 15K array. This suggests that the bias towards K1P1- positive features occurred due to the initially used strategy of applying one single threshold for all features per array to score the signals of the parental lines on the 44K- and 105K-arrays.

Proof of concept: Integration of markers into an existing genetic map and evaluation of marker orders

In order to test if the new markers could be integrated into an existing sugar beet genetic map, we obtained scoring data for 280 co-dominant RFLP- and EST-derived SNP markers. These markers were also mapped in a subset of the K1 mapping population in the study of Schneider et al. (2007). We combined the co-dominant scoring data of 80

F_2 individuals with the corresponding 873 dominant marker scores and grouped them. A stringent LOD score of 12 was chosen, to minimize the number of falsely grouped markers. In total 315 new dominant markers distributed over all nine *B. vulgaris* chromosomes (Table 1) could be assigned to linkage groups (LGs) and allowed the construction of a genetic map containing 595 markers, both co-dominant and dominant (Fig. 5, Table S1). This genetic map has a theoretical average density of one marker per 1.27 Mbp, assuming 758 Mbp as the size of the sugar beet genome (Arumuganathan and Earle 1991). For comparison of marker orders, a map containing only the 280 available co-dominant markers from Schneider et al. (2007) was constructed using the same parameters as described above for marker positioning (Table 1, Fig 5, Table S1). The marker order along the chromosomes was well preserved in LGs I – VIII, except for some local marker substitutions and rearrangements. These effects might be explained by the relatively small number (80) of K1F2 individuals used for map calculation and by the lesser information content on linkage of the dominant markers compared to co-dominant markers (Knapp et al. 1995; Sall and Nilsson 1994). LG IX showed more extensive shifting of the marker group containing anchor markers TG_E0246, MP_R0119, MP_R0002, MP_R0018 and MP_sc from one end of the linkage group to the opposite end. The size of the LGs varied between 147.3 cM (LG I) and 201.0 cM (LG III) and showed inflation for all LGs from 911.1 cM (sum of all LGs, only co-dominant markers) to 1589.5 cM (sum of all LGs, co-dominant and dominant markers), which could result from missing data points of some markers and from problematic markers, resulting in artificial inflation of the map size. We performed a second round of marker grouping and ordering using only the 873 dominant markers with their scores for all 184 K1F2 individuals. The 315 dominant markers mapped to LGs before served as anchor markers. This strategy led to the assignment of 196 additional dominant markers to LGs (Table 1, Fig. 5, Table S1), resulting in a total of 511 dominant markers, translating into an average marker density of one marker per 1.48 Mbp of the sugar beet genome. Of these 511 dominant markers 392 originated from BAC-end or BAC sequences and 119 from ESTs (Table S2).The overall genetic map size increased from 1589.5 cM to 1668.6 cM compared to the map with dominant and co-dominant markers. Except for LGs IV and IX, whose sizes decreased from 169.9 cM to 136.4 and from 184.7 cM to 152.5, respectively, the sizes of all LGs increased. This artificial map inflation was probably again resulting from missing scoring results and the dominant character of the markers. When comparing the marker order of the map with only dominant markers to

the one with both dominant and co-dominant markers, the need for carefully evaluating the dominant marker order becomes obvious. In LGs IV, V and VII, all of which containing only markers linked in coupling phase, the marker orders were well preserved. However, in the other LGs, containing also dominant markers in repulsion phase, there were severe marker rearrangements. Because of the stringent LOD score used within the grouping process, the assignment of markers to LGs was certainly very reliable; the marker order within LGs, however, was probably imperfect, originating mainly from the dominant character of the markers which is unfavorable in an F_2 intercross population. Especially for double heterozygotes from the F_2 population, the repulsion phase provides much less information about linkage than the coupling phase when considering two markers at a time (Liu 1998).

Discussion

In this study we showed that representational oligonucleotide microarray analysis can be successfully applied for high-throughput identification of genetic markers in species with limited sequence information. The marker yield could be drastically increased by optimizing the custom made arrays in several ways. On the one hand only 60-mer oligonucleotides should be placed onto the array, since they proved to perform better than 30-mers. Of initially 50% 60-mers used on the arrays for screening the parental genotypes (Fig. 3a), the fraction of 60-mers among the selected polymorphic marker candidates was 72% (Fig. 3b) and even slightly increased among the finally used markers for map construction (Fig. 3c). On the other hand, BAC-end derived oligonucleotides seem to be favorable compared to oligonucleotides designed based on EST sequences. If marker development is to take place for a genome that has not been sequenced, information on exon borders within ESTs needs to be determined by cross-species alignment. In the present work we aligned sugar beet ESTs against the genomes of *A. thaliana*, *P. trichocarpa* and *O. sativa*. However, such alignments may be erroneous, resulting in the design of some oligonucleotides that perform poorly in hybridization with amplicons prepared from genomic DNA in cases where exon-exon borders within the EST source sequences were missed. We also suggest placing each oligonucleotide in multiple replicates onto the array, to achieve more robust scoring results and thereby to reduce the number of missing data points. The dominant character of our markers provides less information on linkage compared to co-dominant markers (Liu 1998). Especially when the F_2 progeny is used and the markers are in repulsion phase, the quality of marker ordering within a multilocus map decreases drastically (Knapp et al. 1995; Mester et al. 2003). In practice, about half of the markers are expected in each coupling phase, since their identification should be random. Due to a bias in our initial approach for scoring the parental lines, i.e. setting the same signal threshold for all features on one array, the distribution of linkage phases in our experiment is deviating from the expected 1:1 ratio. The fact of having dominant markers in coupling and repulsion phase often leads to mapping the dominant markers from either parent separately to create two different maps in practice (Knapp et al. 1995; Mester et al. 2003; Peng et al. 2000; Sall and Nilsson 1994). We constructed phase separated maps containing co-dominant markers and dominant markers from one coupling group exemplarily for LG III and obtained indeed well preserved order of the

co-dominant markers (Figure S4). One approach to subsequently integrate the two maps into one final map applied before was using pairs of co-dominant and dominant markers, which have higher linkage information than pairs of dominant markers in the coupling phase (Mester et al. 2003). However, since this strategy requires every dominant marker to be paired with a co-dominant marker, it is extremely demanding. Tan and Fu (2007) proposed another method for estimating the recombination fraction between markers that improved the accuracy of estimation through distinction between the coupling phase and the repulsion phase of the linked loci. This method or other specialized algorithms as presented by Jansen (2009) could be utilized for map construction using a dataset of dominant markers like the ones presented in this work. In any case, the disadvantage of the dominant character of the markers could be reduced by using backcross progeny for genotyping. The amount of relative information per individual in an F_2 population drops drastically with higher recombination fraction. Only if dominant markers are in coupling phase and linked tightly, the information content of a F_2 population reaches the one of a backcross population (Allard 1956). Backcross populations map dominant and codominant markers with equal efficiency if the recurrent parent is recessive for the dominant loci, since in that case mapping is not affected by linkage phase. However, only half of the markers are expected to be informative in a backcross population when recessive and dominant loci are randomly distributed between both parents, contrary to F_2 populations where all markers are informative. This effect could be compensated by doubling the number of marker used for map construction.

Applied in an optimized fashion, our approach offers a straight-forward, cost-effective alternative for high-throughput identification and utilizing of genetic markers, when compared to existing methods: While the polymorphism that allows mapping the ROMA based marker is not known, some sequence information at the marker locus is available. The source sequences typically are 500-1000 bases in length (EST sequences and end sequences from genomic clones). The available sequence information is an advantage compared to other platforms such as AFLP or RAPD, because a ROMA based marker can be located on the genome sequence (once available). Also, the sequence information can be used to design a marker assay suitable for typing on sequence-based platforms, and for transfer of markers to other accessions. ROMA is easier to implement than the diversity arrays technology (DArT) (Jaccoud et al. 2001). DArT produces whole-genome fingerprints by scoring the presence versus absence of

DNA fragments in genomic representations and offers the possibility to develop genetic markers without any prior sequence information, but it includes a cloning step, which can be omitted using the ROMA approach. Another widely used method to identify single feature polymorphisms (SFP) in crop plants utilizes Affymetrix microarrays, which have a higher density (> 500,000 oligonucleotides per array) than the arrays used in this study (Bernardo et al. 2009; Das et al. 2008; Deleu et al. 2009; Kim et al. 2009; Rostoks et al. 2005), but depends on the availability of a comprehensive transcriptome catalogue and an Affymetrix GeneChip of the desired species or of a very closely related species, respectively. Our approach provides great flexibility, since arrays design can be adjusted to existing sequence resources that are available for the species of interest. Recently, also approaches combining next generation sequencing with complexity reduction methods, like AFLP or using transcriptome sequences for SFP markers in species without whole genome sequence information have been emerging (Barbazuk et al. 2007; Novaes et al. 2008; van Orsouw et al. 2007). The drawback of such methods might be a relatively high false positive rate in the absence of comprehensive genomic information, due to biased occurrences of sequencing errors (Dohm et al. 2008).

In summary, this study demonstrates the feasibility of ROMA to generate genetic markers in a cost-effective way with the potential for high-throughput analysis. The markers developed in this study will be an asset for the ongoing projects to map and sequence the sugar beet genome. Since the source sequence of each of the developed markers is known (Table S2), the new markers can be easily transferred onto other genotyping platforms.

Acknowledgements

We thank Ruben Rosenkranz and Ines Müller for technical instructions on Agilent array hybridization, Britta Schulz for providing plant material, Dietrich Borchardt for advices on strategies for genetic map construction and Thomas Rosleff-Sörensen for processing and submitting sequence data. This project was supported by the Federal Ministry of Education and Research (BMBF) with grants to H.H. and B.W. ("A physical map of the sugar beet genome to integrate genetics and genomics", Förderkennzeichen 0313127B and 0313127D; "BeetSeq – a reference genome sequence for sugar beet (*Beta vulgaris*)", Förderkennzeichen 0315069A and 0315069B).

References

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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Angiosperm Phylogeny Group (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc 161:105-121

Arumuganathan K, Earle ED (1991) Nuclear DNA Content of Some Important Plant Species. Plant Mol Biol Rep 9:208-218

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910-918

Bernardo AN, Bradbury PJ, Ma H, Hu S, Bowden RL, Buckler ES, Bai G (2009) Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. BMC Genomics 10:251

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a Genetic-Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am J Hum Genet 32:314-331

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557-572

Das S, Bhat PR, Sudhakar C, Ehlers JD, Wanamaker S, Roberts PA, Cui X, Close TJ (2008) Detection and validation of single feature polymorphisms in cowpea (Vigna unguiculata L. Walp) using a soybean genome array. BMC Genomics 9:107

Deleu W, Esteras C, Roig C, Gonzalez-To M, Fernandez-Silva I, Gonzalez-Ibeas D, Blanca J, Aranda MA, Arus P, Nuez F, Monforte AJ, Pico MB, Garcia-Mas J (2009) A set of EST-SNPs for map saturation and cultivar identification in melon. BMC Plant Biol 9:90

Dohm JC, Lange C, Reinhardt R, Himmelbauer H (2009) Haplotype divergence in Beta vulgaris and microsynteny with sequenced plant genomes. Plant J 57:14-26

Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 36:e105

Falk CT (1989) A simple scheme for preliminary ordering of multiple loci: application to 45 CF families. Prog Clin Biol Res 329:17-22

Grubor V, Krasnitz A, Troge JE, Meth JL, Lakshmi B, Kendall JT, Yamrom B, Alex G, Pai D, Navin N, Hufnagel LA, Lee YH, Cook K, Allen SL, Rai KR, Damle RN, Calissano C, Chiorazzi N, Wigler M, Esposito D (2009) Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). Blood 113:1294-1303

Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, Esposito D, Alexander J, Troge J, Grubor V, Yoon S, Wigler M, Ye K, Borresen-Dale AL, Naume B, Schlicting E, Norton L, Hagerstrom T, Skoog L, Auer G, Maner S, Lundin P, Zetterberg A (2006) Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res 16:1465-1479

Himmelbauer H, Dunkel I, Otto GW, Burgtorf C, Schalkwyk LC, Lehrach H (1998) Complex probes for high-throughput parallel genetic mapping of genomic mouse BAC clones. Mamm Genome 9:611-616

Hohmann U, Jacobs G, Telgmann A, Gaafar RM, Alam S, Jung C (2003) A bacterial artificial chromosome (BAC) library of sugar beet and a physical map of the region encompassing the bolting gene B. Mol Genet Genomics 269:126-136

Huang S, Li R, Zhang Z et al. (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275-81

Huang XQ, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868-877

Iwata H, Ninomiya S (2006) AntMap: Constructing genetic linkage maps using an ant colony optimization algorithm. Breed Sci 56:371-377

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:E25

Jaillon O, Aury JM, Noel B et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467

Jansen J (2009) Ordering dominant markers in F-2 populations. Euphytica 165:401-417

Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW (2003) Large-scale genotyping of complex DNA. Nat Biotechnol 21:1233-1237

Kim SH, Bhat PR, Cui X, Walia H, Xu J, Wanamaker S, Ismail AM, Wilson C, Close TJ (2009) Detection and validation of single feature polymorphisms using RNA expression data from a rice genome array. BMC Plant Biol 9:65

Knapp SJ, Holloway JL, Bridges WC, Liu BH (1995) Mapping Dominant Markers Using F2 Matings. Theor Appl Genet 91:74-81

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

Lakshmi B, Hall IM, Egan C, Alexander J, Leotta A, Healy J, Zender L, Spector MS, Xue W, Lowe SW, Wigler M, Lucito R (2006) Mouse genomic representational oligonucleotide microarray analysis: detection of copy number variations in normal and tumor specimens. Proc Natl Acad Sci U S A 103:11234-11239

Lange C, Holtgräwe D, Schulz B, Weisshaar B, Himmelbauer H (2008). Construction and characterization of a sugar beet fosmid library. Genome 51:948-951.Laurent V, Devaux P, Thiel T, Viard F, Mielordt S, Touzet P, Quillet MC (2007) Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. Theor Appl Genet 115:793-805

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD (2004) Development and assessment of microarray-based DNA fingerprinting in Eucalyptus grandis. Theor Appl Genet 109:1329-1336

Lisitsyn N, Lisitsyn N, Wigler M (1993) Cloning the differences between two complex genomes. Science 259:946-951

Liu B-H (1998) Statistical genomics : linkage, mapping, and QTL analysis. CRC Press, Boca Raton

Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 13:2291-2305

Lucito R, Nakimura M, West JA, Han Y, Chin K, Jensen K, McCombie R, Gray JW, Wigler M (1998) Genetic analysis using genomic representations. Proc Natl Acad Sci U S A 95:4487-4492

Lucito R, West J, Reiner A, Alexander J, Esposito D, Mishra B, Powers S, Norton L, Wigler M (2000) Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. Genome Res 10:1726-1736

Lucito R, Wigler M (2003) Microarray-based Representational Analysis of DNA Copy Number: Preparation of Target DNA In: Bowtell D, Sambrook J (eds) DNA Microarrays - A molecular cloning manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 386-391

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793-800

McGrath JM, Shaw RS, de los Reyes BG, Weiland JJ (2004) Construction of a sugar beet BAC library from a hybrid with diverse traits. Plant Mol Biol Rep 22:23-28

Mester DI, Ronin YI, Hu Y, Peng J, Nevo E, Korol AB (2003) Efficient multipoint mapping: making use of dominant repulsion-phase markers. Theor Appl Genet 107:1102-1112

Ming R, Hou S, Feng Y et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452:991-996

Mohring S, Salamini F, Schneider K (2004) Multiplexed, linkage group-specific SNP marker sets for rapid genetic mapping and fingerprinting of sugar beet (Beta vulgaris L.). Mol Breeding 14:475-488

Novaes E, Drost DR, Farmerie WG, Pappas GJ, Jr., Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. BMC Genomics 9:312

Paterson AH, Bowers JE, Bruggmann R et al. (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551-556

Peng J, Korol AB, Fahima T, Roder MS, Ronin YI, Li YC, Nevo E (2000) Molecular genetic maps in wild emmer wheat, Triticum dicoccoides: genome-wide coverage, massive negative interference, and putative quasi-linkage. Genome Res 10:1509-1531

Ribaut JM, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3:236-239

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends Genet 16:276-277

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol 6:R54

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491

Sall T, Nilsson NO (1994) The robustness of recombination frequency estimates in intercrosses with dominant markers. Genetics 137:589-596

Schmutz J, Cannon SB, Schlueter, et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463:178-183

Schneider K, Kulosa D, Soerensen TR, Mohring S, Heine M, Durstewitz G, Polley A, Weber E, Jamsari, Lein J, Hohmann U, Tahiro E, Weisshaar B, Schulz B, Koch G, Jung C, Ganal M (2007) Analysis of DNA polymorphisms in sugar beet (Beta vulgaris L.) and development of an SNP-based map of expressed genes. Theor Appl Genet 115:601-615

Schumacher K, Schondelmaier J, Barzen E, Steinrucken G, Borchardt D, Weber WE, Salamini CJF (1997) Combining different linkage maps in sugar beet (Beta vulgaris L.) to make one map. Plant Breed 116:23-38

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-528

Smit AFA, Hubley R, Green P (1996-2004) RepeatMasker Open-3.0 http://www.repeatmasker.org.

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.). Theor Appl Genet 109:1105-1114

Stanczak CM, Chen ZG, Nelson SE, Suchard M, McCabe ERB, McGhee S (2008) Representational oligonucleotide microarray analysis (ROMA) and comparison of binning and change-point methods of analysis: Application to detection of de122q11.2 (DiGeorge) syndrome. Hum Mutat 29:176-181

Tan YD, Fu YX (2007) A new strategy for estimating recombination fractions between dominant markers from an F2 population. Genetics 175:923-931

van Orsouw NJ, Hogers RC, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H, van Eijk MJ (2007) Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS One 2:e1172

Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: A program to cleave DNA with restriction enzymes. Nucleic Acids Res 31:3688-3691

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee 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

Weber JL, May PE (1989) Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain-Reaction. Am J Hum Genet 44:388-396

Wheelan SJ, Church DM, Ostell JM (2001) Spidey: A tool for mRNA-to-genomic alignments. Genome Res 11:1952-1957

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA Polymorphisms Amplified by Arbitrary Primers Are Useful as Genetic-Markers. Nucleic Acids Res 18:6531-6535

Tables

Table 1 Summary of marker numbers and sizes of linkage groups (LGs) of the constructed genetic maps with co-dominant and dominant markers (C + D), only co-dominant markers (C) and only dominant markers (D)

LG	C + D		C		D	
	No.	Size (cM)	No.	Size (cM)	No.	Size (cM)
I	35	147.3	23	82.2	23	171.8
II	63	153.9	31	82.8	59	216.2
III	64	201.5	33	116.0	48	206.6
IV	76	169.9	24	106.7	61	136.4
V	95	178.3	37	115.4	90	229.2
VI	51	177.9	35	104.0	51	183.9
VII	96	198.0	39	131.9	80	190.0
VIII	45	178.0	27	71.2	51	182.3
IX	70	184.7	31	100.9	48	152.2
∑	595	1589.5	280	911.1	511	1668.6

Figures

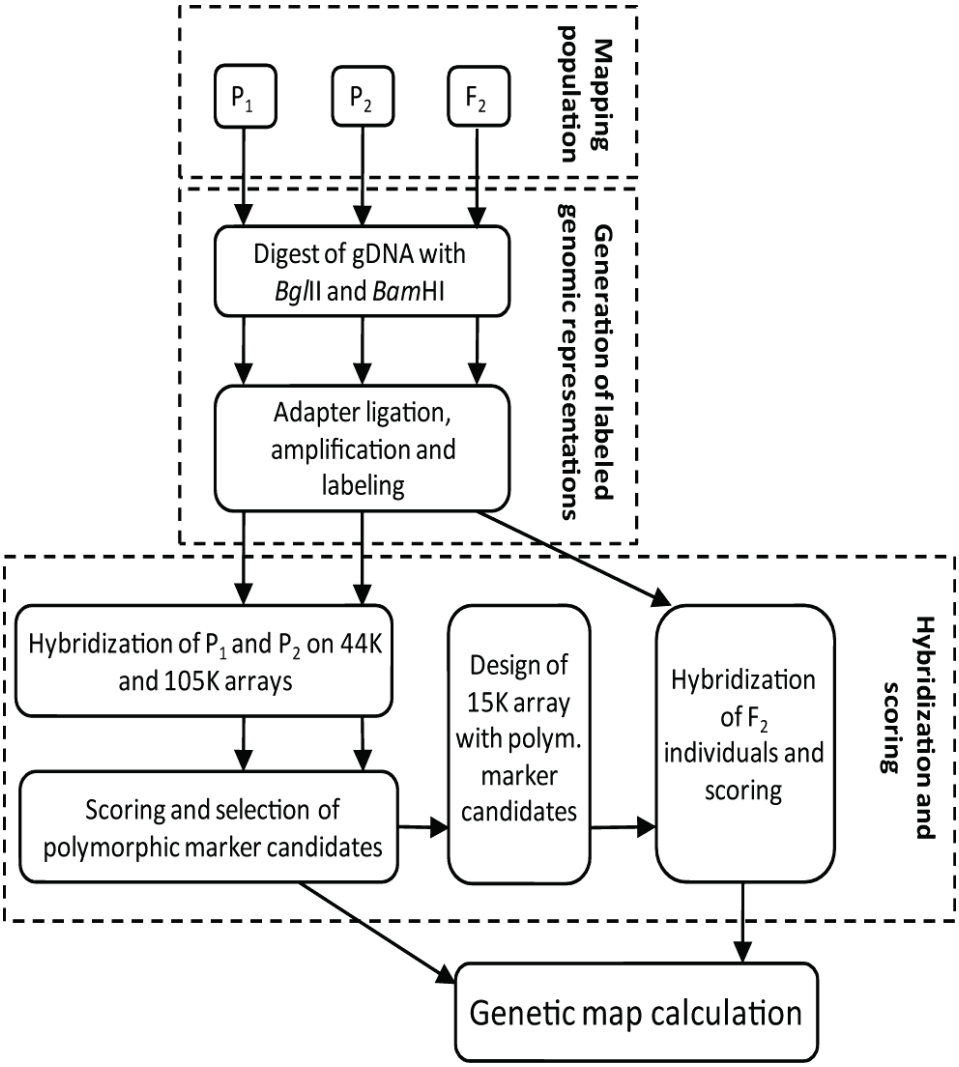


Fig. 1 Flow diagram to illustrate the mapping strategy

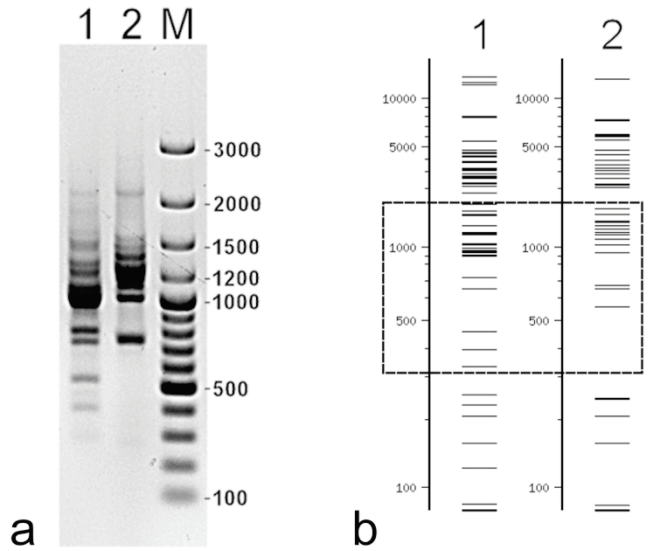


Fig. 2 Verification of size dependent preferential amplification of restriction fragments by *Taq*-polymerase. **a** Purified amplicons of BAC clones SBI-153H13 (lane 1) and ZR-47B15 (lane 2). Sizes of marker bands (lane M) are indicated in base pairs. Only fragments in the size range of approximately 250 – 1500 bp were amplified, **b** Virtual digest of BAC clones SBI-153H13 (lane 1) and ZR-47B15 (lane 2) with *Bam*HI and *Bgl*II. The size range of fragments that is amplified by PCR is indicated by a dashed box

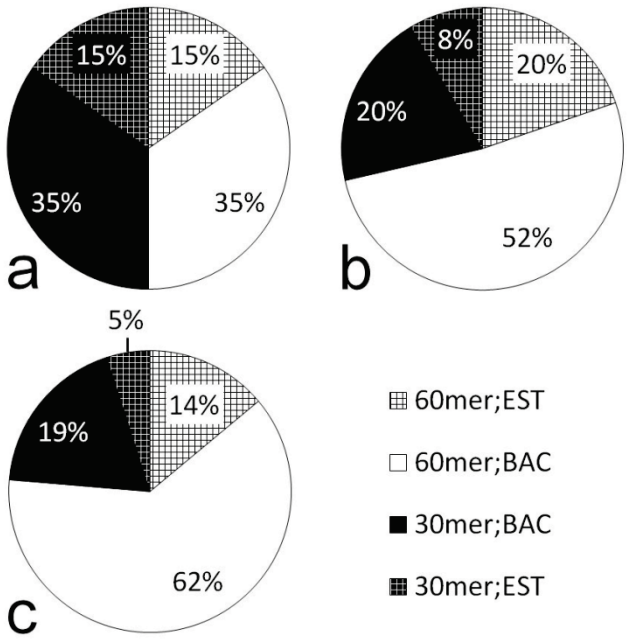


Fig. 3 Size distribution and source sequence origin of oligonucleotide arrays. **a** Sugar beet oligonucleotides on 44K and 105K arrays used for identification of marker candidates. The total number of oligonucleotides on both arrays comprised 146,554, **b** 14,915 oligonucleotides selected from 44K and 105K arrays and placed onto the 15K array for screening of the F_2 genotypes, **c** Features on the 15K array fulfilling the defined criteria for selection and scoring of individual features; used for map calculation

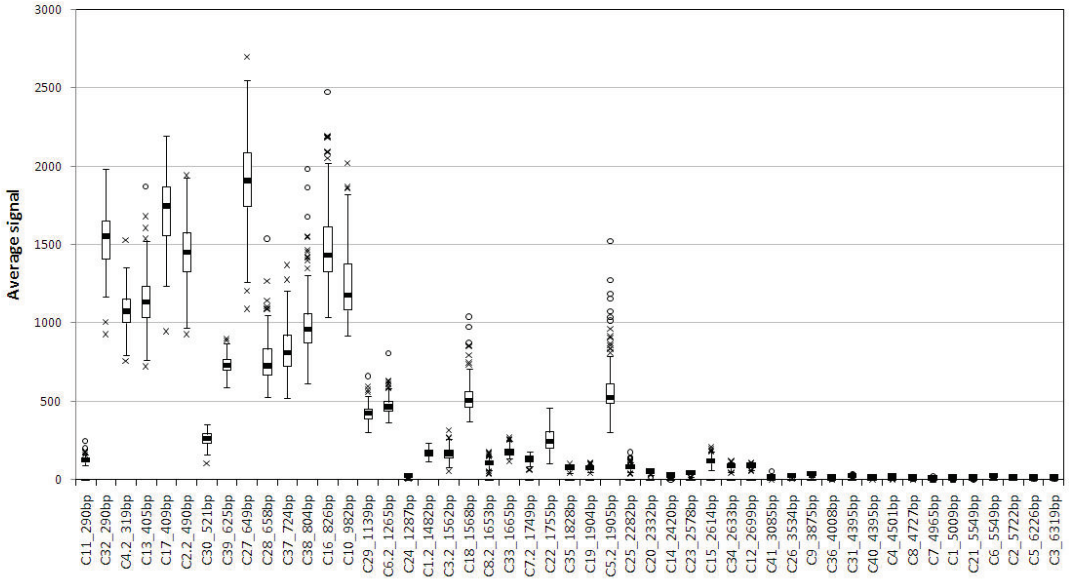
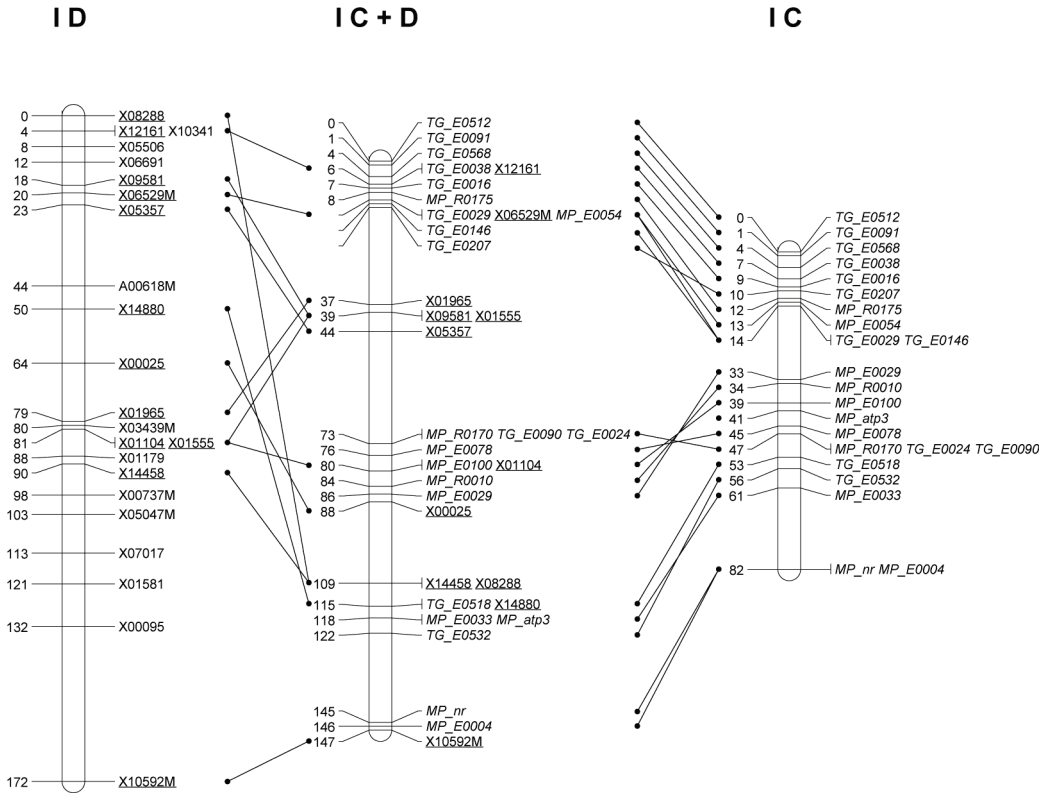
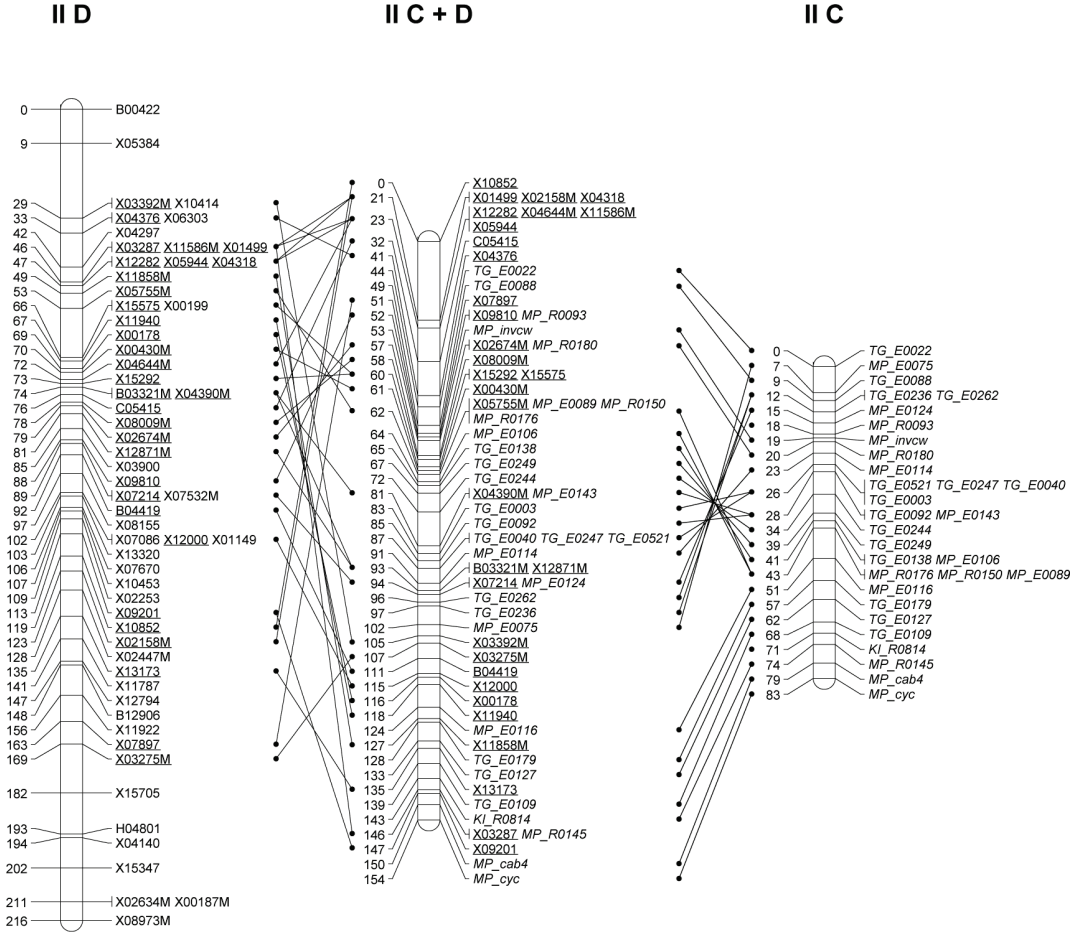
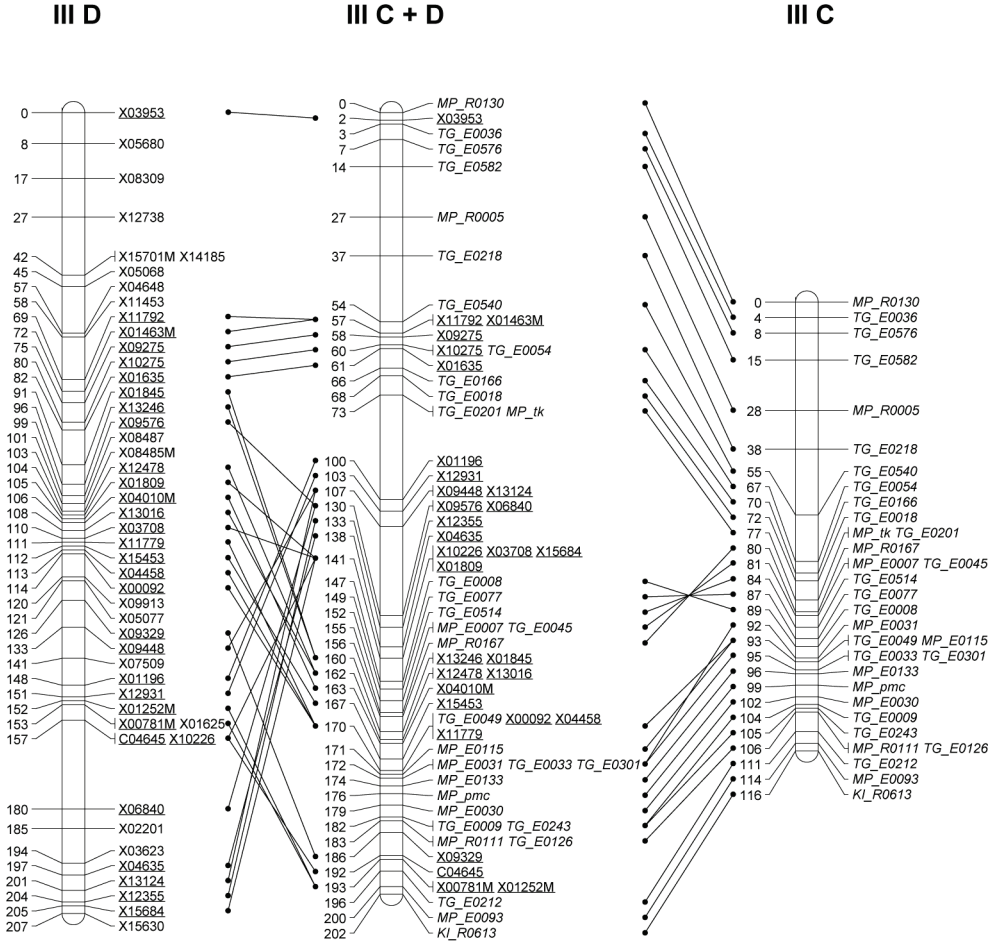
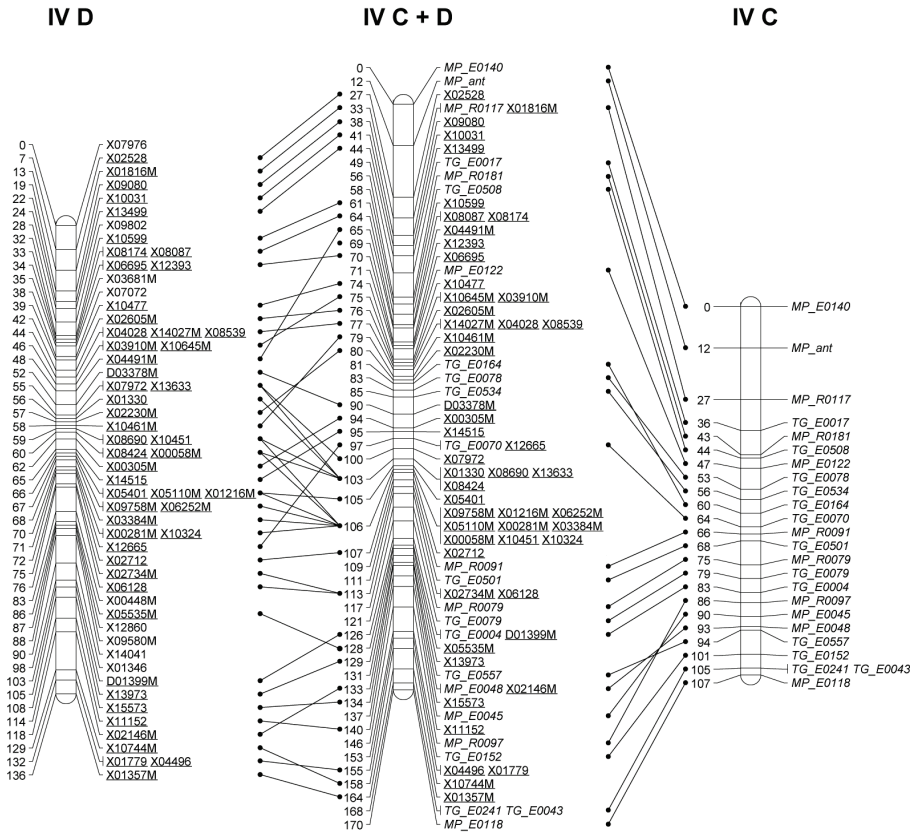


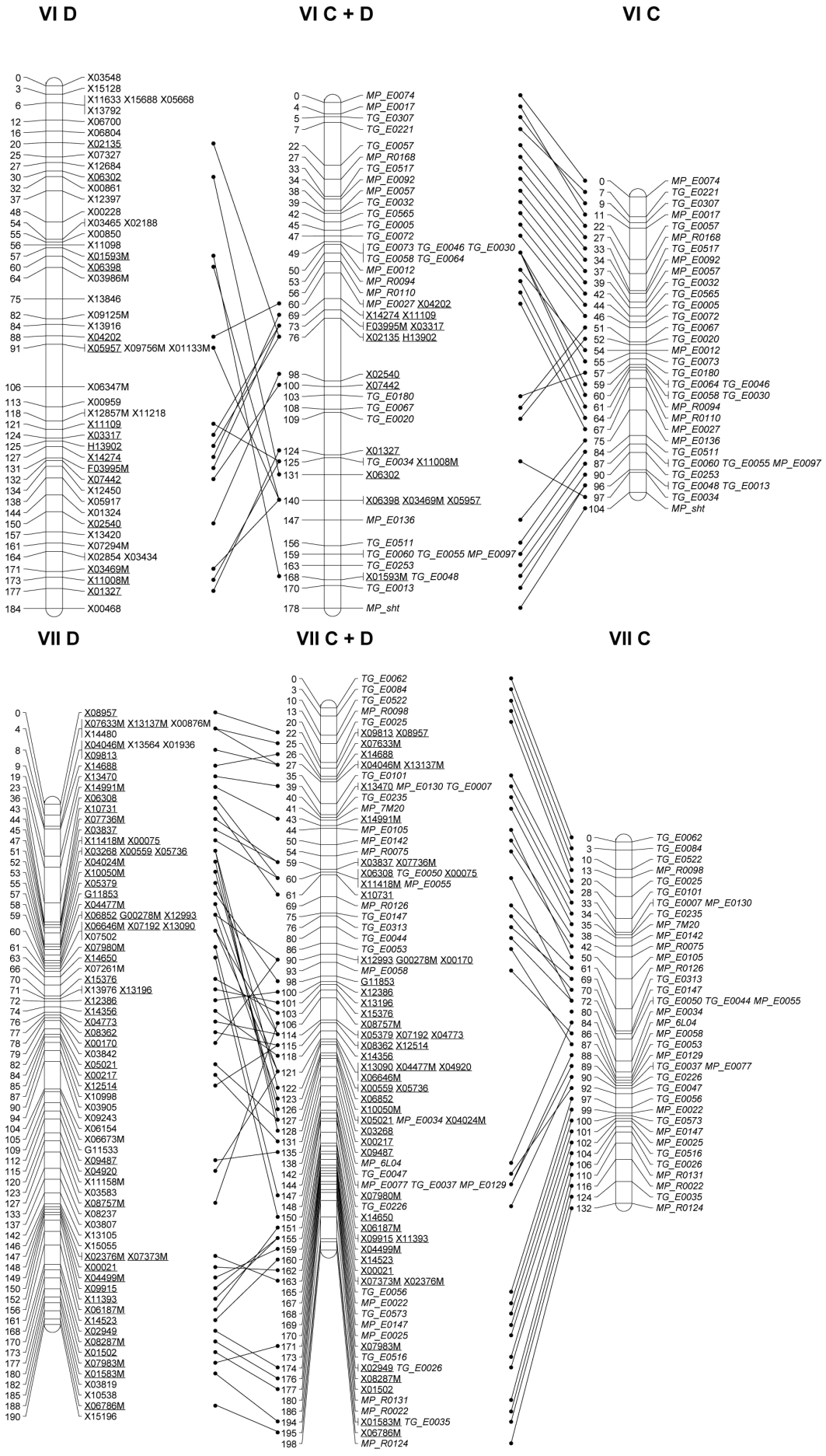
Fig. 4 Box plot of hybridization signals of mouse BAC control oligonucleotides from all genotypes. The signals were normalized and the average values of the five replicates were plotted. Oligonucleotides were ordered along the x-axis in ascending order according the size of the restriction fragments they are complementary to. The interquartile range (IQR) including the median and the inner fences (upper quartile plus 1.5× IQR and lower quartile minus 1.5× IQR, respectively) are shown. Mild outliers (points beyond the inner fences) are displayed as crosses, extreme outliers (points beyond the outer fences, i.e. larger than the upper quartile plus 3× IQR or smaller than the lower quartile minus 3× IQR) as circles




II D
II C + D
II C
0 B00422
9 X05384
29 X03392M X10414
33 X04376 X06303
42 X04297
46 X03287 X11586M X01499
47 X12282 X05944 X04318
49 X11858M
53 X05755M
66 X15575 X00199
67 X11940
69 X00178
70 X00430M
72 X04644M
73 X15292
74 B03321M X04390M
76 C05415
78 X08009M
79 X02674M
81 X12871M
85 X03900
88 X09810
89 X07214 X07532M
92 B04419
97 X08155
102 X07086 X12000 X01149
103 X13320
106 X07670
107 X10453
109 X02253
113 X09201
119 X10852
123 X02158M
128 X02447M
135 X13173
141 X11787
147 X12794
148 B12906
156 X11922
163 X07897
169 X03275M
182 X15705
193 H04801
194 X04140
202 X15347
211 X02634M X00187M
216 X08973M
0 X10852
21 X01499 X02158M X04318
23 X12282 X04644M X11586M
32 X05944
41 C05415
44 X04376
49 TG_E0022
51 TG_E0088
52 X07897
53 X09810 MP_R0093
57 MP_invcw
58 X02674M MP_R0180
60 X08009M
61 X15292 X15575
62 X00430M
64 X05755M MP_E0089 MP_R0150
65 MP_R0176
67 MP_E0106
72 TG_E0138
81 TG_E0249
83 TG_E0244
85 X04390M MP_E0143
87 TG_E0003
91 TG_E0092
93 TG_E0040 TG_E0247 TG_E0521
94 MP_E0114
96 B03321M X12871M
97 X07214 MP_E0124
102 TG_E0262
105 TG_E0236
107 MP_E0075
111 X03392M
115 X03275M
116 B04419
118 X12000
124 X00178
127 X11940
128 MP_E0116
133 X11858M
135 TG_E0179
139 TG_E0127
143 X13173
146 TG_E0109
147 KI_R0814
150 X03287 MP_R0145
154 X09201
MP_cab4
MP_cyc
0 TG_E0022
7 MP_E0075
9 TG_E0088
12 TG_E0236 TG_E0262
15 MP_E0124
18 MP_R0093
19 MP_invcw
20 MP_R0180
23 MP_E0114
26 TG_E0521 TG_E0247 TG_E0040
28 TG_E0003
34 TG_E0092 MP_E0143
39 TG_E0244
41 TG_E0249
43 TG_E0138 MP_E0106
51 MP_R0176 MP_R0150 MP_E0089
57 MP_E0116
62 TG_E0179
68 TG_E0127
71 TG_E0109
74 KI_R0814
79 MP_R0145
83 MP_cab4
MP_cyc


III D
III C + D
III C
0 X03953
8 X05680
17 X08309
27 X12738
42 X15701M X14185
45 X05068
57 X04648
58 X11453
69 X11792
72 X01463M
75 X09275
80 X10275
82 X01635
91 X01845
96 X13246
99 X09576
101 X08487
103 X08485M
104 X12478
105 X01809
106 X04010M
108 X13016
110 X03708
111 X11779
112 X15453
113 X04458
114 X00092
120 X09913
121 X05077
126 X09329
133 X09448
141 X07509
148 X01196
151 X12931
152 X01252M
153 X00781M X01625
157 C04645 X10226
180 X06840
185 X02201
194 X03623
197 X04635
201 X13124
204 X12355
205 X15684
207 X15630
0 MP_R0130
2 X03953
3 TG_E0036
7 TG_E0576
14 TG_E0582
27 MP_R0005
37 TG_E0218
54 TG_E0540
57 X11792 X01463M
58 X09275
60 X10275 TG_E0054
61 X01635
66 TG_E0166
68 TG_E0018
73 TG_E0201 MP_tk
100 X01196
103 X12931
107 X09448 X13124
130 X09576 X06840
133 X12355
138 X04635
141 X10226 X03708 X15684
147 X01809
149 TG_E0008
152 TG_E0077
155 TG_E0514
156 MP_E0007 TG_E0045
160 MP_R0167
162 X13246 X01845
163 X12478 X13016
167 X04010M
170 X15453
TG_E0049 X00092 X04458
X11779
171 MP_E0115
172 MP_E0031 TG_E0033 TG_E0301
174 MP_E0133
176 MP_pmc
179 MP_E0030
182 TG_E0009 TG_E0243
183 MP_R0111 TG_E0126
186 X09329
192 C04645
193 X00781M X01252M
196 TG_E0212
200 MP_E0093
202 KI_R0613
0 MP_R0130
4 TG_E0036
8 TG_E0576
15 TG_E0582
28 MP_R0005
38 TG_E0218
55 TG_E0540
67 TG_E0054
70 TG_E0166
72 TG_E0018
77 MP_tk TG_E0201
80 MP_R0167
81 MP_E0007 TG_E0045
84 TG_E0514
87 TG_E0077
89 TG_E0008
92 MP_E0031
93 TG_E0049 MP_E0115
95 TG_E0033 TG_E0301
96 MP_E0133
99 MP_pmc
102 MP_E0030
104 TG_E0009
105 TG_E0243
106 MP_R0111 TG_E0126
111 TG_E0212
114 MP_E0093
116 KI_R0613


IV D
IV C + D
IV C
0 X07976
7 X02528
13 X01816M
19 X09080
22 X10031
24 X13499
28 X09802
32 X10599
33 X08174 X08087
34 X06695 X12393
35 X03681M
38 X07072
39 X10477
42 X02605M
44 X04028 X14027M X08539
46 X03910M X10645M
48 X04491M
52 D03378M
55 X07972 X13633
56 X01330
57 X02230M
58 X10461M
59 X08690 X10451
60 X08424 X00058M
62 X00305M
65 X14515
66 X05401 X05110M X01216M
67 X09758M X06252M
68 X03384M
70 X00281M X10324
71 X12665
72 X02712
75 X02734M
76 X06128
83 X00448M
86 X05535M
87 X12860
88 X09580M
90 X14041
98 X01346
103 D01399M
105 X13973
108 X15573
114 X11152
118 X02146M
129 X10744M
132 X01779 X04496
136 X01357M
0 MP_E0140
12 MP_ant
27 X02528
33 MP_R0117 X01816M
38 X09080
41 X10031
44 X13499
49 TG_E0017
56 MP_R0181
58 TG_E0508
61 X10599
64 X08087 X08174
65 X04491M
69 X12393
70 X06695
71 MP_E0122
74 X10477
75 X10645M X03910M
76 X02605M
77 X14027M X04028 X08539
79 X10461M
80 X02230M
81 TG_E0164
83 TG_E0078
85 TG_E0534
90 D03378M
94 X00305M
95 X14515
97 TG_E0070 X12665
100 X07972
103 X01330 X08690 X13633
105 X08424
106 X05401
X09758M X01216M X06252M
X05110M X00281M X03384M
X00058M X10451 X10324
107 X02712
109 MP_R0091
111 TG_E0501
113 X02734M X06128
117 MP_R0079
121 TG_E0079
126 TG_E0004 D01399M
128 X05535M
129 X13973
131 TG_E0557
133 MP_E0048 X02146M
134 X15573
137 MP_E0045
140 X11152
146 MP_R0097
153 TG_E0152
155 X04496 X01779
158 X10744M
164 X01357M
168 TG_E0241 TG_E0043
170 MP_E0118
0 MP_E0140
12 MP_ant
27 MP_R0117
36 TG_E0017
43 MP_R0181
44 TG_E0508
47 MP_E0122
53 TG_E0078
56 TG_E0534
60 TG_E0164
64 TG_E0070
66 MP_R0091
68 TG_E0501
75 MP_R0079
79 TG_E0079
83 TG_E0004
86 MP_R0097
90 MP_E0045
93 MP_E0048
94 TG_E0557
101 TG_E0152
105 TG_E0241 TG_E0043
107 MP_E0118
V D
V C + D
V C
0 X10698
11 X13528
12 X13594
15 X11526
20 X00306M
21 X07596M
24 X15388
33 X03523 X15600
39 X08531
40 X14181
41 X01110 X05751 X02026
42 X12846M
45 X03886M X02546
46 X05783 X01629
47 X01833
48 X13308
50 X09686
51 X03489M
52 X08193M
53 X07180 X04483
56 X04486M
58 X13371 X02827
60 X10554
63 X14454
64 X02266 X14114M
66 X00963M X01903
68 X14567
70 X03141M
71 X02120 X05061
73 X00961
74 X15108
77 X03325
82 X09815
84 X02265 X03169
85 X00652 X11247
87 X09268 X07916
88 X05621
89 X07208 X15476
90 X00458 X14360
91 X10005
92 X04623M
96 X05177M
98 X13342
100 X02529M
102 X06865
105 X11981
110 X09856
114 X00427
121 X04640
128 X12254
138 X14015
146 X09236
151 X13418
156 X09466
163 X08925
166 X02911
168 X08792
170 X09273
175 X03727M X10220
177 X03650
185 X01352
188 X07742
191 X05436
192 X09445M
195 X09024M
198 X11361M
204 X02027M X03016M
207 X14243
212 X01775M
224 X07961
225 E01966
229 X12880 X12879
0 MP_E0128
16 MP_E0035
21 X07596M
23 TG_E0019
24 X00306M
31 MP_E0073
38 MP_R0038
42 MP_gdh
47 MP_fbpcp
50 TG_E0085 X14114M
51 X07916
53 X09268 X05621
54 X00963M
55 X03141M
59 X15476 X14454 X14360
61 X07208 X14567 X02266
62 TG_E0135 TG_E0255 X04623M
65 X04486M
69 MP_E0041 MP_R0115 TG_E0039
X13371 X02827 X08193M
70 TG_E0041
73 TG_E0075
75 TG_E0042
76 TG_E0554
78 X12846M
79 TG_E0513
80 TG_E0023
81 X02026 X08531 X14181
82 X01110 X05751
84 TG_E0081 X02546 X03886M
X03489M
85 X13308
88 X02265
93 X04483
97 X03169
100 X05783 X07180 X03325
101 X01629
104 X02911
X10005 TG_E0065 TG_E0071
TG_E0507 X08792 X00458
109 X09273
111 TG_E0010 X03727M
113 MP_R0058 X03650 X10220
114 MP_E0020
118 MP_cab5
124 MP_R0040
130 TG_E0569
131 X09024M
135 X11361M
136 TG_E0086
137 TG_E0027
140 X01352
141 X07742
145 X05436 X09445M TG_E0520
149 X02027M X03016M
150 X14243
154 X01775M
159 MP_E0132
162 MP_E0121
165 KI_R0954
174 X07961 TG_E0011 TG_E0550
176 E01966
178 X12880 X12879
0 MP_E0128
16 MP_E0035
20 TG_E0019
26 MP_E0073
30 TG_E0023
31 TG_E0513
33 TG_E0554
34 TG_E0042
36 TG_E0075
39 TG_E0081
40 TG_E0041
41 MP_E0041 MP_R0115 TG_E0039
42 TG_E0065 TG_E0071 TG_E0085
TG_E0507
46 TG_E0255
48 TG_E0135
51 MP_fbpcp
55 MP_gdh
59 MP_R0038
68 MP_cab5
72 MP_R0058
73 MP_E0020
74 TG_E0010
79 MP_R0040
86 TG_E0086 TG_E0027
88 TG_E0569
93 TG_E0520
101 MP_E0132
104 MP_E0121
107 KI_R0954
115 TG_E0011 TG_E0550


VI D
VI C + D
VI C
VII D
VII C + D
VII C

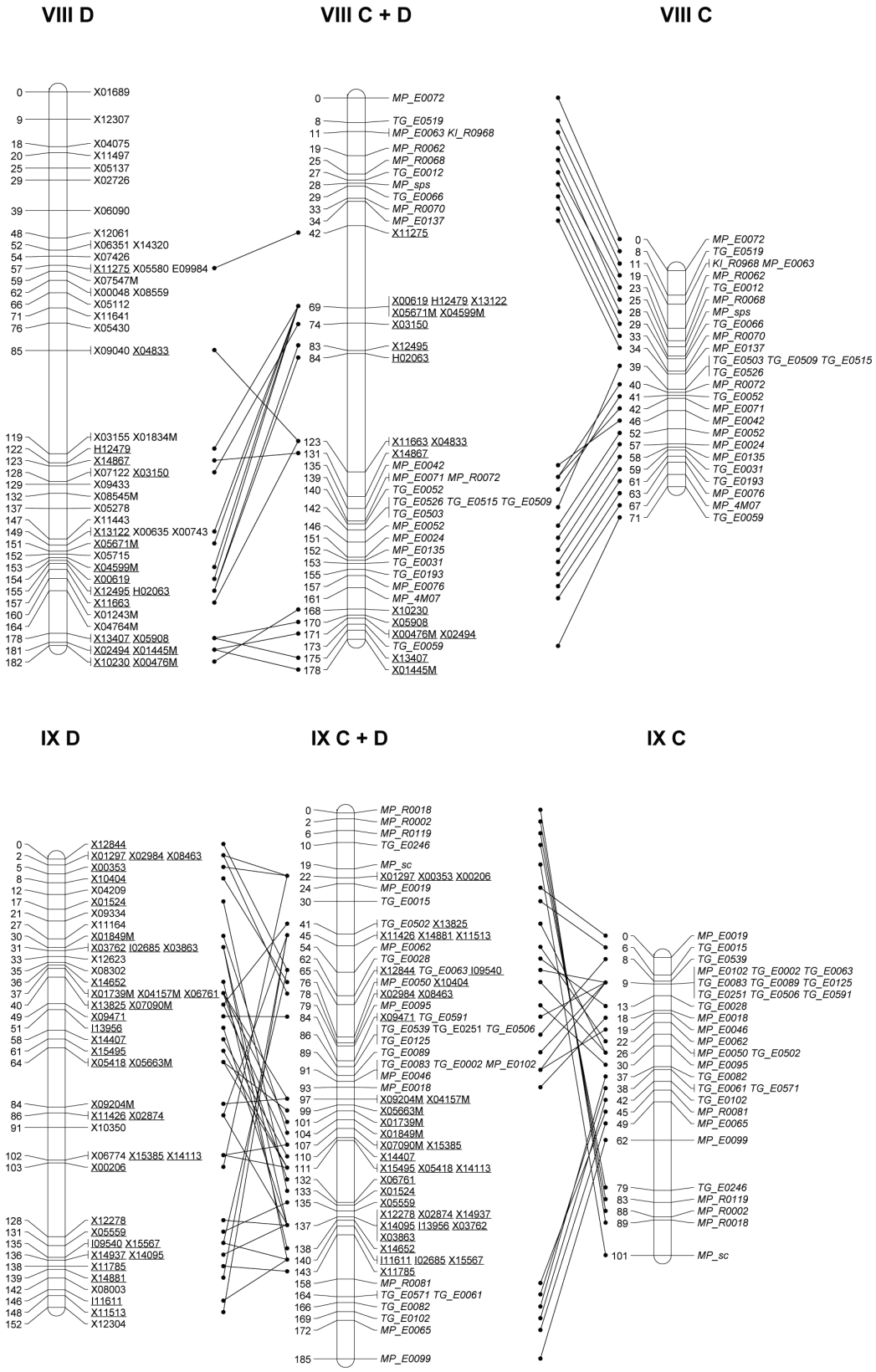


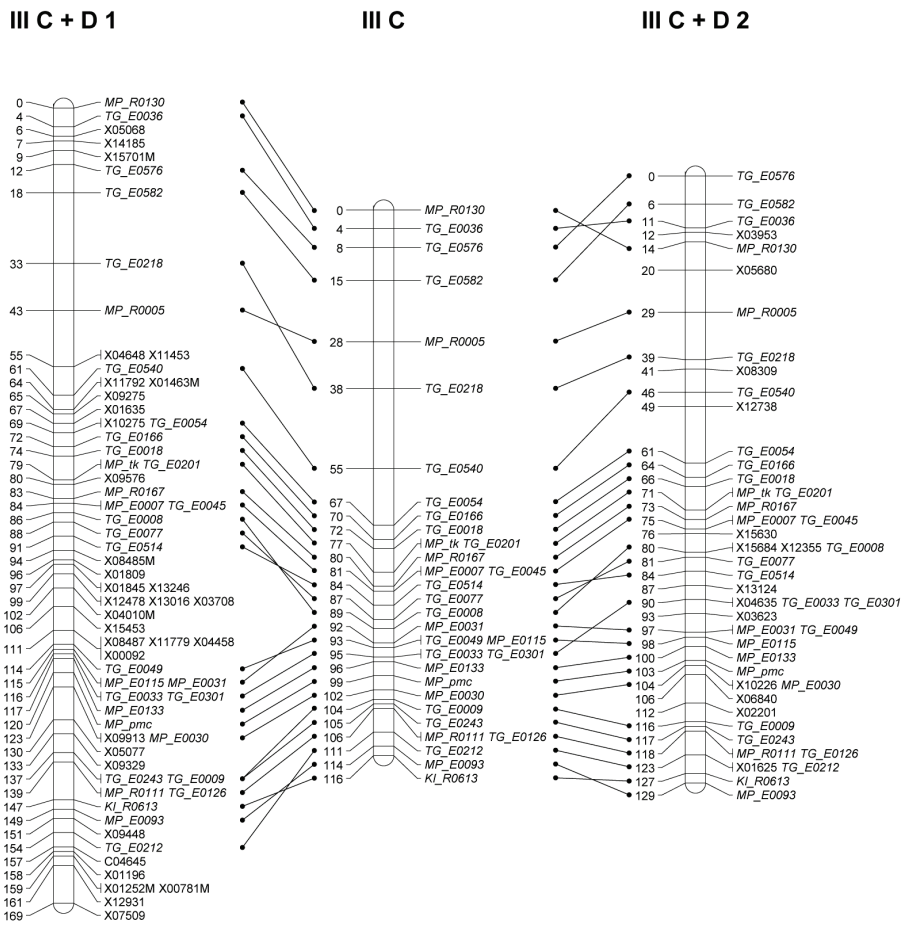
Fig. 5 Sugar beet linkage map constructed with co-dominant markers from Schneider et al. (2007) combined with dominant markers (C + D), only co-dominant markers (C) and only dominant markers (D). Marker names and the cumulative genetic distances in cM are indicated. Corresponding markers between the maps are connected with a line

5.2 Supplementary information

Provided on supplementary CD (see section 11.5):

Supplementary material S1 a: Genetic markers of the map containing only codominant markers (C) b: codominant and dominant markers (C+D) c: only dominant markers (D) and their cumulative genetic distances

Supplementary material S2: Dominant genetic markers with corresponding GenBank Accessions



Supplementary Figure S3: Sugar beet linkage map of linkage group III constructed with co-dominant markers from Schneider et al. (2007) (C) and combined with phase separated dominant markers (C +D 1 and 2). Marker names and the cumulative genetic distances in cM are indicated. Corresponding markers between the maps are connected with a line.

5.3 Contributions

I analysed the obtained data and calculated the genetic maps. In addition, together with HH I developed the concept of the manuscript, wrote the manuscript draft and included the suggestions from the co-authors.

Contributions of co-authors:

LM: Performed all wet lab experiments

JCD: Designed the oligonucleotides for the custom made arrays and wrote the corresponding part of the materials and methods section

DH and BW: Provided the BAC end-sequences

HH: Had the initial idea for the experiments, contributed to the concept of the manuscript

6 Publication II

6.1 Haplotype divergence in *Beta vulgaris* and microsynteny with sequenced plant genomes.

Dohm JC (JCD), Lange C (CL), Reinhardt R (RR), Himmelbauer H (HH). Plant J. 2009; 57 (1): 14-26.
DOI: 10.1111/j.1365-313X.2008.03665.x
The original article is online available at:
http://www3.interscience.wiley.com/journal/121391360/abstract

6.3 Contributions

I performed the wet lab experiments, i.e. production and hybridisation of the macroarrays, generation of the BAC shot-gun library for sequencing and performance of the Southern hybridisation experiments. Furthermore, I analysed the hybridisation results, created Figure 1 (phylogenetic tree), Figure 2 (part of the reconstructed gene map of the *Arabidopsis* ancestor) and Figures S1 (Probe-clone hit statistics), S2/S3 (Southern hybridisations) and S4 (comparison of 35mer hybridisation and Southern results). Together with HH and JCD I developed the concept of the manuscript, wrote parts of the manuscript and took part in critical discussion and optimising of the manuscript.

Contributions of co-authors:

JCD: Performed all bioinformatic analyses (sequence annotation, repeat identification and synteny analysis), took part in development of the manuscript concept, wrote major parts of the manuscript, included the suggestions from the co-authors and took part in critical discussion of the manuscript

RR: Sequenced the BAC shot-gun libraries and assembled the BAC sequences

HH: Had the initial idea for the experiments, took part in development of the manuscript concept, wrote parts of the manuscript and took part in critical discussion and optimisation of the manuscript

7 Publication III

7.1 Construction and characterization of a sugar beet (*Beta vulgaris*) fosmid library

Lange C (CL), Holtgräwe D (DH), Schulz B (BS), Weisshaar B (BW), Himmelbauer H (HH). Genome. 2008; 51 (11): 948-51.
DOI:10.1139/G08-071
The original article is online available at:
http://article.pubs.nrc-cnrc.gc.ca/ppv/RPViewDoc?issn=0831-2796&volume=51&issue=11&startPage=948

7.2 Contributions

I performed all wet lab experiments, i.e. construction of the fosmid library and screening of the library. In addition, I performed the BLAST analyses, together with HH I developed the manuscript concept, I wrote the manuscript draft and included suggestions of the co-authors.

Contributions of co-authors:

DH and BW: Generated the fosmid end-sequences and provided the corresponding parts of the manuscript

BS: Provided the plant material

HH: Had the initial idea for the experiments, contributed to the concept of the manuscript

8 Publication IV

8.1 Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in Beta vulgaris L.

Menzel G, Dechyeva D, Keller H, Lange C, Himmelbauer H, Schmidt T. Chromosome Res. 2006; 14 (8): 831-44
DOI: 10.1007/s10577-006-1090-1
http://www.springerlink.com/content/707841719561t8263

8.2 Contributions

I generated the *Beta vulgaris* small insert library and produced the macroarrays for isolation of the genomic *VulMITE I* clone. In addition, I wrote minor parts of the manuscript (material and methods).

9 Discussion

This study describes the generation of several sugar beet genomic tools and their application for evolutionary studies and linkage analysis. A new technique allowing high-throughput identification and genotyping of genetic markers in sugar beet was developed (Publication I). Besides its transferability to other species, the obtained genetic markers will be an asset for ordering of sequence contigs on the genetic map. In addition, possible linkage of physical and genetic maps is provided, since genetic markers are based on source sequences, which are also used for construction of a BAC based physical map utilising a hybridisation approach. An example of the hybridisation based approach for physical map construction and its relevance for synteny studies is demonstrated (Publication II). Furthermore, we constructed and characterised a sugar beet fosmid library (Publication III) supporting the assembly and orienting of sequence contigs and generated a short insert library facilitating repeat identification within the sugar beet genome (Publication IV).

9.1 Genome Mapping

Genetic and physical maps are essential tools for structural, functional and applied genomics. The method presented in Publication I demonstrated that representational oligonucleotide microarray analysis (ROMA) can be successfully applied for high-throughput identification of genetic markers in species with limited sequence information. Genomic representations of both parents of a mapping population were hybridised on 105K and 44K microarrays containing in total 146,554 custom made oligonucleotides based on sugar beet BAC-end sequences (BESs) and ESTs. Subsequent analyses resulted in selection of 14,915 oligonucleotides identified as potentially polymorphic, which were placed on a new 15K microarray used for screening of 184 F_2 individuals. Finally, 511 new dominant genetic markers could be placed onto a genetic map utilising co-dominant anchor markers. This low marker yield was due to control and test steps performed in order to establish the new technique and could be increases by several optimisations. On the one hand, 60mer oligonucleotides proved to perform superior in comparison to 30mer oligonucleotides. However, the crucial point to be improved is the scoring of the marker signals as present or absent. One possibility to obtain more robust

results is to place each oligonucleotide in multiple replicates, at least triplicates, onto the arrays. Oligonucleotides showing high variation within one group would be discarded, diminishing falsely scored markers. In addition, the approach utilised for determining polymorphisms between the parental lines P1 and P2 on the 105K and 44K arrays could be ameliorated by considering each feature signal individually instead of setting one signal threshold for the whole array. A promising strategy would be to examine the ratio of signal intensities between P1 and P2 specifically for each feature on the array. An individual threshold based on the level of signal differences between P1 and P2 for one feature could be set and features with P1/P2 or P2/P1 ratio, respectively, exceeding this threshold would be selected as polymorphic marker candidates. In order to test the feasibility of this approach, I determined the influence of the signal ratio between P1 and P2 on marker yield and the abundance of features exemplarily for the 44K array (Figure 2). The yield for each ratio category was calculated by dividing the number of markers falling into the particular ratio category and that could be placed onto the genetic map by the total number of features belonging to the particular category on the 15K array. For instance, 260 features whose signal intensities varied more than 10fold between P1 and P2 on the 44K array were successfully utilised for genetic map calculation, 976 features falling into this category (ratio between the signals of P1 and P2 on the 44k array > 10) were present on the whole 15K array, thus the marker yield for this category was 26.7% (260/976). In total, 5182 features on the 15K array originated from the 44K array (signal ratio >1), 339 of these were placed onto the genetic map. Hence, the overall marker yield for the 44K array in the performed experiments was 6.5%. A strong correlation between the extent of signal difference between P1 and P2 and the marker yield becomes obvious. However, increasing the signal ratio threshold above which marker candidates are selected, leads to drastic decrease of the number of features on the 44k array falling into the particular category (Figure 2). Features that show a more than 10fold higher signal with one parent in comparison to the other parent had a high marker yield (26.7%), but only 1429 features on the 44K array were in this range. The choice of the threshold should be considered carefully and adjusted to the available resources. Nevertheless, selection of about 15,000 features from the 44K with a P1/P2 or P2/P1 signal ratio >2 and subsequent genotyping of the F_2 progeny would have probably yielded about 1400 markers with the ability to be anchored onto the genetic map. This number could apparently be exceeded by using just 60mer oligonucleotides and replicates.

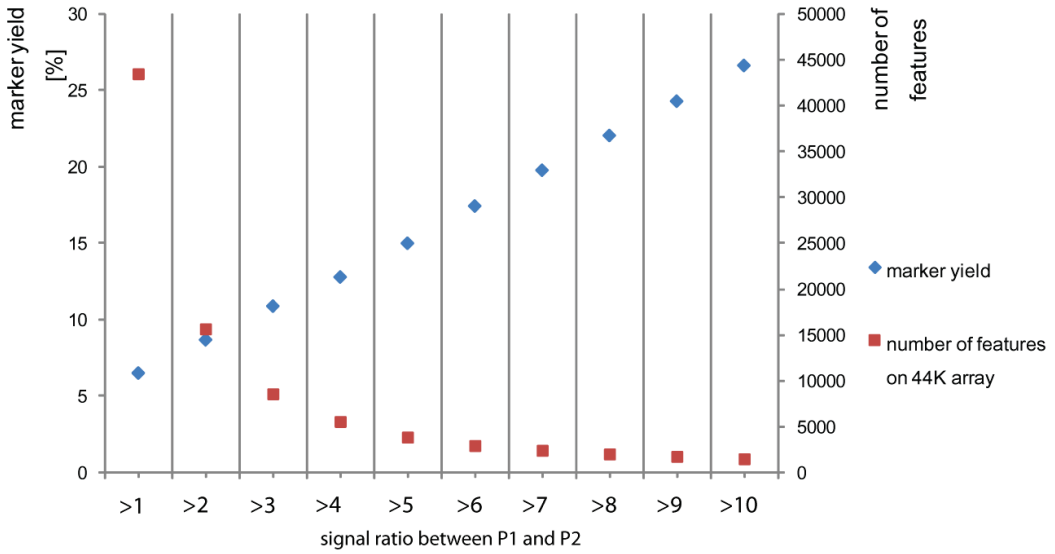


Figure 2: Influence of the signal ratio between P1 and P2 (i.e. signals of P1/P2 or P2/P1, respectively) of all features on the 44K array on marker yield and abundance. The yield for each ratio category was calculated by dividing the number of features falling into the particular category that could be anchored to the genetic map by the total number of features belonging to this category on the 15K array. Each feature that contributed to a marker was counted, although features representing one BES or EST were merged before map calculation (339 features originating from the 44K array contributed to 264 markers anchored to the genetic map).

The dominant character of the genetic markers developed in Publication I is a vital obstacle, since it provides less information on linkage compared to co-dominant markers (Liu 1998). Especially when an F_2 progeny is used and the markers are in repulsion phase, the quality of marker ordering within a multilocus map decreases drastically (Knapp et al. 1995; Mester et al. 2003). Different strategies can be exploited in order to alleviate this effect. In addition to the approach tested in Publication 1, i.e. construction of phase separated maps containing co-dominant markers and dominant markers from one coupling group, specialised algorithms can be applied for map calculation in order to improve the accuracy of estimations (Tan and Fu 2007; Jansen 2009). Furthermore, the choice of the mapping population has great influence on mapping efficiency. Different studies investigated the ability of different population types to detect recombinants by using either co-dominant or dominant markers (Allard 1956; Reiter et al. 1992). The amount of information about the recombination fraction provided by any data set depends on the completeness of classification and the closeness of linkage. Mapping efficiency of co-dominant markers in an F_2 population is high, but dominant markers map

less efficiently, since mixed linkage phases will negatively influence the information content. In contrast, if the recurrent parent is recessive for the dominant loci, backcross populations map dominant and co-dominant markers with equal efficiency, as mapping is not affected by the linkage phase. However, only half of the markers are expected to be informative in a backcross population when recessive and dominant loci are randomly distributed between both parents, contrary to F_2 populations where all markers are informative. This effect could be compensated by doubling the number of marker used for map construction.

The presented genetic mapping technique exploiting ROMA has considerable advantages compared to other existing high-throughput genetic mapping assays, such as diversity arrays technology (DArT) (Jaccoud et al. 2001) and detection of single feature polymorphisms (SFPs) utilising Affymetrix microarrays (Das et al. 2008; Bernardo et al. 2009; Deleu et al. 2009). DArT includes a laborious cloning step and the obtained markers are anonymous. The detection of SFPs on Affymetrix microarrays depends on the availability of a comprehensive transcriptome catalogue and an Affymetrix GeneChip of the desired species or of a very closely related species, respectively. However, since the ROMA approach is based on absence or presence of amplicons, but the underlying DNA polymorphism is not detected, the genetic markers might be difficult to transfer to high-throughput genotyping platforms such as Illumina`s GoldenGate assay, which has been deployed for SNP genotyping in soybean (Hyten et al. 2008), wheat (Akhunov et al. 2009) and loblolly pine (*Pinus taeda*) (Eckert et al. 2009).

In general, a genetic map can reveal if a marker is linked to a trait. Yet, the real physical distance remains unknown and the marker might be physically located quite far away from the gene of interest. Hence, linkage of genetic markers to a physical map is essential to exploit their full potential. Integrated genetic and physical maps are crucial for isolation of any gene of interest, e.g. underlying an important trait, by positional cloning. As introduced above, different approaches can be exploited for construction of genetic and physical maps all having individual advantages and disadvantages. The integration of genetic and physical maps is accomplished by obtaining markers that can be placed on both maps and thus establish a connection between different maps. For example, in papaya and grapevine, BAC clones were fingerprinted in order to construct physical maps (Lamoureux et al. 2006; Yu et al. 2009). Several previously genetically mapped markers (mainly SSRs) were anchored to these fingerprinted contig (FPC)-based physical maps by performing PCR screens of BAC clones and *in silico* analyses,

electronic PCR (Schuler 1997) and BLAST analyses (Altschul et al. 1990) utilising BESs. In addition, markers derived from BESs also used in physical map construction could directly be anchored. The BAC-based marker-content physical mapping technique introduced in Publication II allows direct linkage of genetic and physical maps. Although the presented genetic markers do not contain information about the molecular alteration underlying the polymorphisms per se, their approximate location is linked to the ESTs and BESs the oligonucleotides on the microarrays are based on. Using these source sequences for oligonucleotide design in the course of physical map construction facilitates direct linkage of the genetic and physical maps. Additional major advantages of the marker-content approach are its potential to use different genotypes (Publication II) and the small risk of detecting paralogous loci. Several problems can hamper anchoring of physical map contigs to the genetic map. An ideal situation would be to use single-copy probes derived from all genetically mapped loci to screen the BAC library. In theory the physical mapping technique presented in Publication II has the potential to accomplish this condition, but since this requires a high density genetic map at the same time as physical mapping is carried out, the number of available genetic probes is usually the bottleneck. In addition, ambiguous hybridisation results, e.g. due to the exploited pooling strategies or cross-contaminations, might lead to false contig assembly. Another problem might be that a probe from one genetic location hybridises to BACs in different contigs. Likely explanations for this phenomenon are either that the probe is complementary to a duplicated sequence in the genome or that physical contigs truly have to be merged. The latter explanation might presumably be true especially at the end of contigs. Thus, physical contig assembly and anchoring to the genetic map has to be carried out thoroughly and often includes manual post-editing of automated assembly steps.

9.2 Evolution and genome structure

Plant genomes differ substantially among species in their sizes and their number of genes. It is generally accepted that key factors for shaping of flowering plant genomes are large-scale DNA duplication events. However, the number, extent and timing of these duplications are still controversial. Whole genome physical maps or ultimately complete genome sequences of flowering plants from different plant lineages are essential prerequisites for studies on genome structure and evolution of angiosperms. Since

little is known about synteny between rosids and *Caryophyllales* so far, we determined and annotated the genomic sequences of two BAC clones derived from two different *Beta vulgaris* haplotypes and analysed the extent of synteny between them and rosid genomes (Publication II). In addition, we studied the intraspecific variation between the two sugar beet haplotypes. Initially, we used a hybridisation based approach for comparison of the gene order in *Beta vulgaris* and an ancient *Arabidopsis* genome before the recent α – WGD in *Arabidopsis,* reconstructed by Blanc (2003). They determined the approximate gene order of the ancestral genome by merging genes lying in sister regions, resulting in resulting in 20,187 genes arranged in a linear array. We hybridised 30 oligonucleotide probes based on ESTs corresponding to *Arabidopsis* orthologs on chromosome 1 and 4 that were co-localised in the reconstructed *Arabidosis* pseudo ancestral genome, on sugar beet BAC macroarrays comprising two different sugar beet libraries. One clone from each sugar beet library was chosen that hybridised with the same 5 probes, i.e. they span the same genomic region. The genomic region was identified to be located on sugar beet chromosome 1. The orthologs in *Arabidopsis* were assigned to chromosomes 1 and 4. Sequencing of the two clones and comprehensive annotation followed by collinearity analyses, revealed synteny between sugar beet, *Arabidopsis*, poplar, *Medicago* and grapevine. Although *Arabidopsis* has undergone at least one more WGD (β) after the split from sugar beet, the gene order of the *Arabidosis* pseudo ancestral genome showed broad collinearity to the sugar beet. The most comprehensive matches were found in grapevine and poplar (paralogous regions on two chromosomes). Grapevine has not undergone a lineage specific WGD. Its genome has a triplicate structure, resulting from a paleohexaploidy event (Jaillon et al. 2007), also shared with papaya (Ming et al. 2008) and probably tomato (Tang et al. 2008). Poplar has undergone at least one lineage specific WGD, resulting in a sextuplicate genome structure (Tuskan et al. 2006; Tang et al. 2008). These findings suggest that the paleohexaploidy event (γ) pre-dates the split between asterids and rosids, thus traces of the hexaploidy event should also be found in sugar beet. However, since we were analysing just a small, randomly chosen segment from the sugar beet genome, we could not address this question. Indications for the existence of many duplicated genes in the sugar beet genome have already been found (Mcgrath et al. 2004). The number of genes of a species is influenced by WGD events to some extent. Papaya has a reduced gene number (about 10% fewer genes than *Arabidopsis*), which may be accounted for by a paucity of genome duplications relative to other sequenced angiosperms (Soltis et al. 2009).

On the other hand, for soybean and poplar (both experienced lineage-specific WGD events) much higher numbers of genes were predicted (Table 1). For sugar beet we estimated a number of 29,000 protein coding genes (Publication II). This might be a hint for the absence of a lineage-specific WGD in sugar beet. But since *Arabidopsis* has a relatively small number of predicted protein coding genes, although it has undergone several rounds of WGDs, there must be additional reasons that cause the different number of genes. Differences in generation time and speed of evolution are proposed causes likely influencing the number of genes of plant species (Tuskan et al. 2006; Fawcett et al. 2009). A whole genome physical map and a complete genome sequence of sugar beet will facilitate essential insights into the genome structure of sugar beet and angiosperm evolution. Still, the oldest genome duplications will still be difficult to detect and time precisely, since more recent polyploidy events in concert with gene loss, chromosomal inversions and translocations can conceal the early events in angiosperm genome evolution.

9.3 Outlook: Whole genome physical map and genome sequencing

At the basis of all sequenced plant genomes lay physical, genetic and integrated maps, respectively. Until recently, sequencing of complete plant genomes was performed using traditional Sanger sequencing technology exploiting a clone-by-clone strategy (The Arabidopsis Genome Initiative 2000; International Rice Genome Sequencing Project 2005; Schnable et al. 2009) or whole genome shotgun (WGS) strategy (Tuskan et al. 2006; Jaillon et al. 2007; Ming et al. 2008; Schmutz et al. 2010). Both strategies rely on a comprehensive physical map either for the identification of a minimum tiling path, i.e. selection of large-insert clones that span a genomic region with minimal overlaps, or for orienting and ordering of sequence contigs. For sequencing of the cucumber genome, a novel *de novo* sequencing strategy was carried out taking advantage of the long reads of the Sanger technology and the high sequencing depth and low unit cost of NGS (Huang et al. 2009). For the ongoing sequencing of the sugar beet genome, a similar strategy is utilised, combining Sanger sequencing of BESs and fosmid end sequences (FESs) (Publication III) with several genomic NGS resources (454 whole genome shotgun single read data, 2.5 kbp and 4.5 kbp Illumina paired-end data and 20 kbp 454 paired-end data). Especially when exploiting the relatively short NGS reads, a comprehensive physical map and a combination of several resources with different defined sequence

lengths providing a robust scaffold are required for successful ordering and orienting of sequence contigs. The FESs bear the advantage of narrowly defined size of the clone inserts (Publication III), thus contributing essentially to the process of assembling scaffolds. As mentioned above, the BESs facilitate anchor points to the genetic map. Repeats are a major obstacle for successful assembly of plant genome sequences, frequently causing gaps and misassembled contigs. In addition, TEs might result in overestimation of the plant´s gene content, since they often acquire portions of genes, leading to the amplification of truncated gene fragments that have open reading frames (ORFs), which are easily mistaken for standard plant protein coding genes (Bennetzen et al. 2004). The best way to deal with repetitive elements is extensive identification and comprehensive annotation, as has been shown in Publication IV exemplarily for three MITE families in *Beta vulgaris*.

All new tools and findings, presented in this work contribute substantially to a deeper understanding of the genome structure of sugar beet and provide the basis for successful sequencing of the sugar beet genome.

10 References

Akhunov E, Nicolet C and Dvorak J (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theoretical and Applied Genetics 119(3): 507-517.

Allard RW (1956). Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24(10): 235-278.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic local alignment search tool. Journal of Molecular Biology 215(3): 403-410.

Arumuganathan K and Earle ED (1991). Nuclear DNA Content of Some Important Plant Species. Plant Molecular Biology Reporter 9(3): 208-218.

Asakura N, Mori N, Nakamura C and Ohtsuka I (2009). Genotyping of the Q locus in wheat by a simple PCR-RFLP method. Genes & Genetic Systems 84(3): 233-237.

Avila CM, Nadal S, Moreno MT and Torres AM (2006). Development of a simple PCR-based marker for the determination of growth habit in Vicia faba L. using a candidate gene approach. Molecular Breeding 17(3): 185-190.

Badaeva ED, Amosova AV, Muravenko OV, Samatadze TE, Chikida NN, Zelenin AV, Friebe B and Gill BS (2002). Genome differentiation in Aegilops. 3. Evolution of the D-genome cluster. Plant Systematics and Evolution 231(1-4): 163-190.

Barbazuk WB, Emrich SJ, Chen HD, Li L and Schnable PS (2007). SNP discovery via 454 transcriptome sequencing. Plant Journal 51(5): 910-918.

Barzen E, Mechelke W, Ritter E, Seitzer JF and Salamini F (1992). RFLP markers for sugar beet breeding - chromosomal linkage maps and location of major genes for rhizomania resistance, monogermy and hypocotyls color. Plant Journal 2(4): 601-611.

Bennett MD and Smith JB (1991). Nuclear-DNA amounts in angiosperms. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 334(1271): 309-345.

Bennetzen JL, Coleman C, Liu RY, Ma JX and Ramakrishna W (2004). Consistent over-estimation of gene number in complex plant genomes. Current Opinion in Plant Biology 7(6): 732-736.

Bentley DR, Balasubramanian S, et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218): 53-59.

Bernardo AN, Bradbury PJ, Ma H, Hu S, Bowden RL, Buckler ES and Bai G (2009). Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. BMC Genomics 10: 251.

Bert PF, Charmet G, Sourdille P, Hayward MD and Balfourier F (1999). A high-density molecular map for ryegrass (Lolium perenne) using AFLP markers. Theoretical and Applied Genetics 99(3-4): 445-452.

Bertioli DJ, Moretzsohn MC, et al. (2009). An analysis of synteny of Arachis with Lotus and Medicago sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics 10.

Blair MW, Diaz LM, Buendia HF and Duque MC (2009). Genetic diversity, seed size associations and population structure of a core collection of common beans (Phaseolus vulgaris L.). Theoretical and Applied Genetics 119(6): 955-972.

Blanc G, Hokamp K and Wolfe KH (2003). A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Research 13(2): 137-144.

Bonierbale MW, Plaisted RL and Tanksley SD (1988). Rflp Maps Based on a Common Set of Clones Reveal Modes of Chromosomal Evolution in Potato and Tomato. Genetics 120(4): 1095-1103.

Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a Genetic-Linkage Map in Man Using Restriction Fragment Length Polymorphisms. American Journal of Human Genetics 32(3): 314-331.

Bowers JE, Chapman BA, Rong JK and Paterson AH (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422(6930): 433-438.

Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S and Stettler RF (1994). Molecular-genetics of growth and development in Populus. 3. A genetic-linkage map of a hybrid poplar composed of RFLP,STS, and RAPD markers. Theoretical and Applied Genetics 89(2-3): 167-178.

Bruno WJ, Knill E, Balding DJ, Bruce DC, Doggett NA, Sawhill WW, Stallings RL, Whittaker CC and Torney DC (1995). Efficient Pooling Designs for Library Screening. Genomics 26(1): 21-30.

Bureau TE, Ronald PC and Wessler SR (1996). A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proceedings of the National Academy of Sciences of the United States of America 93(16): 8524-8529.

Bureau TE and Wessler SR (1994). Mobile Inverted-Repeat Elements of the Tourist Family Are Associated with the Genes of Many Cereal Grasses. Proceedings of the National Academy of Sciences of the United States of America 91(4): 1411-1415.

Burke DT, Carle GF and Olson MV (1987). Cloning of large segments of exogenous DNA into Yeast by means of artificial chromosome vectors. Science 236(4803): 806-812.

Cai WW, Chow CW, Damani S, Gregory SG, Marra M and Bradley A (2001). An SSLP marker-anchored BAC framework map of the mouse genome. Nature Genetics 29(2): 133-134.

Cai WW, Reneker J, Chow CW, Vaishnav M and Bradley A (1998). An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54(3): 387-397.

Casasoli M, Mattioni C, Cherubini M and Villani F (2001). A genetic linkage map of European chestnut (Castanea sativa Mill.) based on RAPD, ISSR and isozyme markers. Theoretical and Applied Genetics 102(8): 1190-1199.

Chen MS, Presting G, et al. (2002). An integrated physical and genetic map of the rice genome. Plant Cell 14(3): 537-545.

Close TJ, Bhat PR, et al. (2009). Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10: 13.

Collins J and Hohn B (1978). Cosmids - type of plasmid gene-cloning vector that is packable invitro in bacteriophage Lambda-heads. Proceedings of the National Academy of Sciences of the United States of America 75(9): 4242-4246.

Coulson A, Sulston J, Brenner S and Karn J (1986). Toward a physical map of the genome of the nematode Caenorhabditis-elegans. Proceedings of the National Academy of Sciences of the United States of America 83(20): 7821-7825.

Cuevas HE, Staub JE, Simon PW, Zalapa JE and McCreight JD (2008). Mapping of genetic loci that regulate quantity of beta-carotene in fruit of US Western Shipping melon (Cucumis melo L.). Theoretical and Applied Genetics 117(8): 1345-1359.

Das S, Bhat PR, Sudhakar C, Ehlers JD, Wanamaker S, Roberts PA, Cui X and Close TJ (2008). Detection and validation of single feature polymorphisms in cowpea (Vigna unguiculata L. Walp) using a soybean genome array. BMC Genomics 9: 107.

Dechyeva D and Schmidt T (2006). Molecular organization of terminal repetitive DNA in Beta species. Chromosome Research 14(8): 881-897.

Deleu W, Esteras C, et al. (2009). A set of EST-SNPs for map saturation and cultivar identification in melon. BMC Plant Biology 9: 90.

Deng GR (1988). A sensitive non-radioactive PCR-RFLP analysis for detecting point mutations at 12th codon of oncogene c-Ha-ras in DNAs of gastric cancer. Nucleic Acids Research 16(13): 6231.

Ding Y, Johnson MD, Chen WQ, Wong D, Chen YJ, Benson SC, Lam JY, Kim YM and Shizuya H (2001). Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using type IIS restriction endonucleases. Genomics 74(2): 142-154.

Dohm JC, Lange C, Reinhardt R and Himmelbauer H (2009). Haplotype divergence in Beta vulgaris and microsynteny with sequenced plant genomes. Plant Journal 57(1): 14-26.

Dong F, McGrath JM, Helgeson JP and Jiang J (2001). The genetic identity of alien chromosomes in potato breeding lines revealed by sequential GISH and FISH analyses using chromosome-specific cytogenetic DNA markers. Genome 44(4): 729-734.

Draycott AP (2006). Sugar beet. Oxford ; Ames, Iowa, Blackwell Pub.

Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM and Neale DB (2009). High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (Pinus taeda L.). Tree Genetics & Genomes 5(1): 225-234.

Falistocco E (2009). Presence of triploid cytotypes in the common fig (Ficus carica L.). Genome 52(11): 919-925.

Fan JB, Chee MS and Gunderson KL (2006). Highly parallel genomic assays. Nature Reviews Genetics 7(8): 632-644.

Fan JB, Oliphant A, et al. (2003). Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology 68: 69-78.

Fang DQ and Roose ML (1997). Identification of closely related citrus cultivars with inter-simple sequence repeat markers. Theoretical and Applied Genetics 95(3): 408-417.

Fawcett JA, Maere S and Van de Peer Y (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proceedings of the National Academy of Sciences of the United States of America 106(14): 5737-5742.

Feschotte C, Jiang N and Wessler SR (2002). Plant transposable elements: Where genetics meets genomics. Nature Reviews Genetics 3(5): 329-341.

Finnegan DJ (1989). Eukaryotic Transposable Elements and Genome Evolution. Trends in Genetics 5(4): 103-107.

Fischer HE (1989). Origin of the Weisse-Schesische Rube (White Silesian Beet) and resynthesis of sugar-beet. Euphytica 41(1-2): 75-80.

Flavell RB, Bennett MD, Smith JB and Smith DB (1974). Genome Size and Proportion of Repeated Nucleotide-Sequence DNA in Plants. Biochemical Genetics 12(4): 257-269.

Gall JG and Pardue ML (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proceedings of the National Academy of Sciences of the United States of America 63(2): 378-383.

Gindullis F, Dechyeva D and Schmidt T (2001). Construction and characterization of a BAC library for the molecular dissection of a single wild beet centromere and sugar beet (Beta vulgaris) genome analysis. Genome 44(5): 846-855.

Green ED, Mohr RM, Idol JR, Jones M, Buckingham JM, Deaven LL, Moyzis RK and Olson MV (1991). Systematic generation of sequence-tagged sites for physical mapping of human-chromosomes - Application to the mapping of human chromosome-7 using yeast artificial chromosomes. Genomics 11(3): 548-564.

Green ED and Olson MV (1990). Systematic Screening of Yeast Artificial-Chromosome Libraries by Use of the Polymerase Chain-Reaction. Proceedings of the National Academy of Sciences of the United States of America 87(3): 1213-1217.

Gregory SG, Howell GR and Bentley DR (1997). Genome mapping by fluorescent fingerprinting. Genome Research 7(12): 1162-1168.

Greilhuber J, Borsch T, Muller K, Worberg A, Porembski S and Barthlott W (2006). Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biology 8(6): 770-777.

Gupta SK, Souframanien J and Gopalakrishna T (2008). Construction of a genetic linkage map of black gram, Vigna mungo (L.) Hepper, based on molecular markers and comparative studies. Genome 51(8): 628-637.

Hagihara E, Matsuhira H, Ueda M, Mikami T and Kubo T (2005). Sugar beet BAC library construction and assembly of a contig spanning Rf1, a restorer-of-fertility gene for Owen cytoplasmic male sterility. Molecular Genetics and Genomics 274(3): 316-323.

Hallden C, Hjerdin A, Rading IM, Sall T, Fridlundh B, Johannisdottir G, Tuvesson S, Akesson C and Nilsson NO (1996). A high density RFLP linkage map of sugar beet. Genome 39(4): 634-645.

Han YH, Zhang ZH, Liu JH, Lu JY, Huang SW and Jin WW (2008). Distribution of the tandem repeat sequences and karyotyping in cucumber (Cucumis sativus L.) by fluorescence in situ hybridization. Cytogenetic and Genome Research 122(1): 80-88.

Haque S, Ashraf N, Begum S, Sarkar RH and Khan H (2008). Construction of Genetic Map of Jute (Corchorus olitorius L.) Based on RAPD Makers. Plant Tissue Culture & Biotechnology 18(2): 165-172.

Hardenbol P, Baner J, et al. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. Nature Biotechnology 21(6): 673-678.

Hass-Jacobus B, Futrell-Griggs M, et al. (2006). Integration of hybridization-based markers (overgos) into physical maps for comparative and evolutionary explorations in the genus Oryza and in Sorghum. BMC Genomics 7(1): 199.

Hawkins JS, Kim H, Nason JD, Wing RA and Wendel JF (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Research 16(10): 1252-1261.

Heitkam T and Schmidt T (2009). BNR - a LINE family from Beta vulgaris - contains a RRM domain in open reading frame 1 and defines a L1 sub-clade present in diverse plant genomes. Plant Journal 59(6): 872-882.

Helentjaris T, Weber DF and Wright S (1986). Use of Monosomics to Map Cloned DNA Fragments in Maize. Proceedings of the National Academy of Sciences of the United States of America 83(16): 6035-6039.

Herwig R, Schulz B, et al. (2002). Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes. Plant Journal 32(5): 845-857.

Heslop-Harrison JS (2000). Comparative genome organization in plants: From sequence and markers to chromatin and chromosomes. Plant Cell 12(5): 617-635.

Hohmann U, Jacobs G, Telgmann A, Gaafar RM, Alam S and Jung C (2003). A bacterial artificial chromosome (BAC) library of sugar beet and a physical map of the region encompassing the bolting gene B. Molecular Genetics and Genomics 269(1): 126-136.

Hohn B and Murray K (1977). Packaging recombinant DNA-molecules into bacteriophage particles invitro. Proceedings of the National Academy of Sciences of the United States of America 74(8): 3259-3263.

Howell EC, Armstrong SJ, Barker GC, Jones GH, King GJ, Ryder CD and Kearsey MJ (2005). Physical organization of the major duplication on Brassica oleracea chromosome O6 revealed through fluorescence in situ hybridization with Arabidopsis and Brassica BAC probes. Genome 48(6): 1093-1103.

Huang S, Li R, et al. (2009). The genome of the cucumber, Cucumis sativus L. Nature Genetics 41(12): 1275 - 1281.

Hyten DL, Song Q, et al. (2008). High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theoretical and Applied Genetics 116(7): 945-952.

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463(7282): 763-768.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. Nature 436(7052): 793-800.

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA and De Jong PJ (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genetics 6(1): 84-89.

Iovene M, Wielgus SM, Simon PW, Buell CR and Jiang JM (2008). Chromatin Structure and Physical Mapping of Chromosome 6 of Potato and Comparative Analyses With Tomato. Genetics 180(3): 1307-1317.

Jaccoud D, Peng K, Feinstein D and Kilian A (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29(4): E25.

Jacobs G, Dechyeva D, Menzel G, Dombrowski C and Schmidt T (2004). Molecular characterization of Vulmar1, a complete mariner transposon of sugar beet and diversity of mariner- and En/Spm-like sequences in the genus Beta. Genome 47(6): 1192-1201.

Jacobs G, Dechyeva D, Wenke T, Weber B and Schmidt T (2009). A BAC library of Beta vulgaris L. for the targeted isolation of centromeric DNA and molecular cytogenetics of Beta species. Genetica 135(2): 157-167.

Jaillon O, Aury JM, et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161): 463-467.

Jansen J (2009). Ordering dominant markers in F-2 populations. Euphytica 165(2): 401-417.

Jiang J and Gill BS (2006). Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome 49(9): 1057-1068.

Jiang N and Wessler SR (2001). Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. Plant Cell 13(11): 2553-2564.

John HA, Birnstie.Ml and Jones KW (1969). RNA-DNA hybrids at cytological level. Nature 223(5206): 582-587.

Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK and Brar DS (2000). Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus Oryza. Theoretical and Applied Genetics 100(8): 1311-1320.

Kamisugi Y, von Stackelberg M, Lang D, Care M, Reski R, Rensing SA and Cuming AC (2008). A sequence-anchored genetic linkage map for the moss, Physcomitrella patens. Plant Journal 56(5): 855-866.

Kaundun SS and Matsumoto S (2003). Development of CAPS markers based on three key genes of the phenylpropanoid pathway in Tea, Camellia sinensis (L.) O. Kuntze, and differentiation between assamica and sinensis varieties. Theoretical and Applied Genetics 106(3): 375-383.

Kelleher CT, Chiu R, et al. (2007). A physical map of the highly heterozygous Populus genome: integration with the genome sequence and genetic map and analysis of haplotype variation. Plant Journal 50(6): 1063-1078.

Kesseli RV, Paran I and Michelmore RW (1994). Analysis of a detailed genetic-linkage map of Lactuca-sativa (lettuce) constructed from RFLP and RAPD markers. Genetics 136(4): 1435-1446.

Khorasani MZ, Hennig S, et al. (2004). A first generation physical map of the medaka genome in BACs essential for positional cloning and clone-by-clone based genomic sequencing. Mechanisms of Development 121(7-8): 903-913.

Kim JS, Childs KL, Islam-Faridi MN, Menz MA, Klein RR, Klein PE, Price HJ, Mullet JE and Stelly DM (2002). Integrated karyotyping of sorghum by in situ hybridization of landed BACs. Genome 45(2): 402-412.

Kim UJ, Shizuya H, Dejong PJ, Birren B and Simon MI (1992). Stable propagation of cosmid sized human DNA inserts in an F-factor based vector. Nucleic Acids Research 20(5): 1083-1085.

King J, Thorogood D, Edwards KJ, Armstead IP, Roberts L, Skot K, Hanley Z and King IP (2008). Development of a genomic microsatellite library in perennial ryegrass (Lolium perenne) and its use in trait mapping. Annals of Botany 101(6): 845-853.

Klein PE, Klein RR, et al. (2000). A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Research 10(6): 789-807.

Knapp SJ, Holloway JL, Bridges WC and Liu BH (1995). Mapping Dominant Markers Using F2 Matings. Theoretical and Applied Genetics 91(1): 74-81.

Kojima T, Nagaoka T, Noda K and Ogihara Y (1998). Genetic linkage map of ISSR and RAPD markers in einkorn wheat in relation to that of RFLP markers. Theoretical and Applied Genetics 96(1): 37-45.

Kojima Y, Ebana K, Fukuoka S, Nagamine T and Kawase M (2005). Development of an RFLP-based rice diversity research set of germplasm. Breeding Science 55(4): 431-440.

Koumbaris GL and Bass HW (2003). A new single-locus cytogenetic mapping system for maize (Zea mays L.): overcoming FISH detection limits with marker-selected sorghum (S. propinquum L.) BAC clones. Plant Journal 35(5): 647-659.

Kubis SE, Heslop-Harrison JS, Desel C and Schmidt T (1998). The genomic organization of non-LTR retrotransposons (LINEs) from three Beta species and five other angiosperms. Plant Molecular Biology 36(6): 821-831.

Kuelheim C, Yeoh SH, Maintz J, Foley WJ and Moran GF (2009). Comparative SNP diversity among four Eucalyptus species for genes from secondary metabolite biosynthetic pathways. BMC Genomics 10: Article No.: 452.

Lamoureux D, Bernole A, et al. (2006). Anchoring of a large set of markers onto a BAC library for the development of a draft physical map of the grapevine genome. Theoretical and Applied Genetics 113(2): 344-356.

Langer-Safer PR, Levine M and Ward DC (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 79(14): 4381-4385.

Laurent V, Devaux P, Thiel T, Viard F, Mielordt S, Touzet P and Quillet MC (2007). Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. Theoretical and Applied Genetics 115(6): 793-805.

Lengerova M, Kejnovsky E, Hobza R, Macas J, Grant SR and Vyskot B (2004). Multicolor FISH mapping of the dioecious model plant, Silene latifolia. Theoretical and Applied Genetics 108(7): 1193-1199.

Liu B-H (1998). Statistical genomics : linkage, mapping, and QTL analysis. Boca Raton, CRC Press.

Luo MC, Thomas C, et al. (2003). High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82(3): 378-389.

Lysak MA, Fransz PF, Ali HBM and Schubert I (2001). Chromosome painting in Arabidopsis thaliana. Plant Journal 28(6): 689-697.

Macas J, Koblizkova A, Navratilova A and Neumann P (2009). Hypervariable 3 ' UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. Gene 448(2): 198-206.

Malysheva-Otto LV, Ganal MW and Roder MS (2006). Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (Hordeum vulgare L.). BMC Genetics 7: 6.

Margulies M, Egholm M, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057): 376-380.

Matsuzaki H, Dong SL, et al. (2004). Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nature Methods 1(2): 109-111.

McCouch SR, Chen XL, Panaud O, Temnykh S, Xu YB, Cho YG, Huang N, Ishii T and Blair M (1997). Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Molecular Biology 35(1-2): 89-99.

McGrath JM, Shaw RS, de los Reyes BG and Weiland JJ (2004). Construction of a sugar beet BAC library from a hybrid with diverse traits. Plant Molecular Biology Reporter 22(1): 23-28.

Menzel G, Dechyeva D, Keller H, Lange C, Himmelbauer H and Schmidt T (2006). Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in Beta vulgaris L. Chromosome Research 14(8): 831-844.

Menzel G, Dechyeva D, Wenke T, Holtgrawe D, Weisshaar B and Schmidt T (2008). Diversity of a complex centromeric satellite and molecular characterization of dispersed sequence families in sugar beet (Beta vulgaris). Annals of Botany 102(4): 521-530.

Mester DI, Ronin YI, Hu Y, Peng J, Nevo E and Korol AB (2003). Efficient multipoint mapping: making use of dominant repulsion-phase markers. Theoretical and Applied Genetics 107(6): 1102-1112.

Metzker ML (2009). Sequencing technologies - the next generation. Nature Reviews Genetics 11(1): 31-46.

Ming R, Hou S, et al. (2008). The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452(7190): 991-996.

Moore MJ, Bell CD, Soltis PS and Soltis DE (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences of the United States of America 104(49): 19363-19368.

Muchero W, Diop NN, et al. (2009). A consensus genetic map of cowpea [Vigna unguiculata (L) Walp.] and synteny based on EST-derived SNPs. Proceedings of the National Academy of Sciences of the United States of America 106(43): 18159-18164.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G and Erlich H (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51 Pt 1: 263-273.

Murray AW and Szostak JW (1983). Construction of artificial chromosomes in yeast. Nature 305(5931): 189-193.

Nelson WM, Bharti AK, Butler E, Wei FS, Fuks G, Kim H, Wing RA, Messing J and Soderlund C (2005). Whole-genome validation of high-information-content fingerprinting. Plant Physiology 139(1): 27-38.

Olson M, Hood L, Cantor C and Botstein D (1989). A common language for physical mapping of the human genome. Science 245(4925): 1434-1435.

Olson MV, Dutchik JE, Graham MY, Brodeur GM, Helms C, Frank M, Maccollin M, Scheinman R and Frank T (1986). Random-clone strategy for genomic restriction mapping in yeast. Proceedings of the National Academy of Sciences of the United States of America 83(20): 7826-7830.

Orr W and Molnar SJ (2008). Development of PCR-based SCAR and CAPS markers linked to beta-glucan and protein content QTL regions in oat. Genome 51(6): 421-425.

Pedersen C and Langridge P (1997). Identification of the entire chromosome complement of bread wheat by two-colour FISH. Genome 40(5): 589-593.

Portis E, Mauromicale G, Mauro R, Acquadro A, Scaglione D and Lanteri S (2009). Construction of a reference molecular linkage map of globe artichoke (Cynara cardunculus var. scolymus). Theoretical and Applied Genetics 120(1): 59-70.

Queen RA, Gribbon BM, James C, Jack P and Flavell AJ (2004). Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Molecular Genetics and Genomics 271(1): 91-97.

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV and Scolnik PA (1992). Global and Local Genome Mapping in Arabidopsis-Thaliana by Using Recombinant Inbred Lines and Random Amplified Polymorphic Dnas. Proceedings of the National Academy of Sciences of the United States of America 89(4): 1477-1481.

Ross (1999). Screening Large-Insert Libraries by Hybridization. Current Protocols in Human Genetics.

Rout GR, Sahoo DP and Aparajita S (2009). Studies on Inter and intra-population variability of Pongamia pinnata: a bioenergy legume tree. Crop Breeding and Applied Biotechnology 9(3): 268-273.

SanMiguel P, Tikhonov A, et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science 274(5288): 765-768.

Savitsky VF (1950). Monogerm sugar beets in the United States. Proceedings of the American Society of Sugar Beet Technologists 1950: 156-159.

Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ and Shoemaker RC (2004). Mining EST databases to resolve evolutionary events in major crop species. Genome 47(5): 868-876.

Schmidt T and Heslop-Harrison JS (1996). The physical and genomic organization of microsatellites in sugar beet. Proceedings of the National Academy of Sciences of the United States of America 93(16): 8761-8765.

Schmidt T, Kubis S and Heslopharrison JS (1995). Analysis and chromosomal localization of retrotransposons in sugar-beet (Beta vulgaris L.) - LINEs and TY1-copia-like elements as major components of the genome. Chromosome Research 3(6): 335-345.

Schmutz J, Cannon SB, et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463(7278): 178-183.

Schnable PS, Ware D, et al. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326(5956): 1112-1115.

Schneider K, Borchardt DC, Schafer-Pregl R, Nagl N, Glass C, Jeppsson A, Gebhardt C and Salamini F (1999). PCR-based cloning and segregation analysis of functional gene homologues in Beta vulgaris. Molecular and General Genetics 262(3): 515-524.

Schneider K, Kulosa D, et al. (2007). Analysis of DNA polymorphisms in sugar beet (Beta vulgaris L.) and development of an SNP-based map of expressed genes. Theoretical and Applied Genetics 115(5): 601-615.

Schondelmaier J, Steinrucken G and Jung C (1996). Integration of AFLP markers into a linkage map of sugar beet (Beta vulgaris L). Plant Breeding 115(4): 231-237.

Schuler GD (1997). Sequence mapping by electronic PCR. Genome Research 7(5): 541-550.

Schumacher K, Schondelmaier J, Barzen E, Steinrucken G, Borchardt D, Weber WE and Salamini CJF (1997). Combining different linkage maps in sugar beet (Beta vulgaris L.) to make one map. Plant Breeding 116(1): 23-38.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y and Simon M (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia-coli using an F-factor-based vector. Proceedings of the National Academy of Sciences of the United States of America 89(18): 8794-8797.

Singh N, Lal RK and Shasany AK (2009). Phenotypic and RAPD diversity among 80 germplasm accessions of the medicinal plant isabgol (Plantago ovata, Plantaginaceae). Genetics and Molecular Research 8(4): 1273-1284.

Soltis DE, Albert VA, et al. (2009). Polyploidy and Angiosperm Diversification. American Journal of Botany 96(1): 336-348.

Southern EM (1975). Detection of Specific Sequences among DNA Fragments Separated by Gel-Electrophoresis. Journal of Molecular Biology 98(3): 503-&.

Staden R (1980). A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Research 8(16): 3673-3694.

Syed NH and Flavell AJ (2007). Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. Nature Protocols 1(6): 2746-2752.

Syed NH, Sorensen AP, Antonise R, van de Wiel C, van der Linden CG, van't Westende W, Hooftman DAP, den Nijs HCM and Flavell AJ (2006). A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. Theoretical and Applied Genetics 112(3): 517-527.

Szczepaniak M, Bieniek W, Boron P, Szklarczyk M and Mizianty M (2009). A contribution to characterisation of genetic variation in some natural Polish populations of Elymus repens (L.) Gould and Elymus hispidus (Opiz) Melderis (Poaceae) as revealed by RAPD markers*. Plant Biology 11(5): 766-773.

Szinay D, Chang SB, et al. (2008). High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. Plant Journal 56(4): 627-637.

Tan YD and Fu YX (2007). A new strategy for estimating recombination fractions between dominant markers from an F2 population. Genetics 175(2): 923-931.

Tang HB, Wang XY, Bowers JE, Ming R, Alam M and Paterson AH (2008). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Research 18(12): 1944-1954.

Tang XM, Szinay D, et al. (2008). Cross-Species Bacterial Artificial Chromosome-Fluorescence in Situ Hybridization Painting of the Tomato and Potato Chromosome 6 Reveals Undescribed Chromosomal Rearrangements. Genetics 180(3): 1319-1328.

Tao QZ, Chang YL, Wang JZ, Chen HM, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM and Zhang HB (2001). Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158(4): 1711-1724.

Taramino G and Tingey S (1996). Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39(2): 277-287.

Tarchini R, Biddle P, Wineland R, Tingey S and 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(3): 381-391.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF and Gaut BS (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp mays L.). Proceedings of the National Academy of Sciences of the United States of America 98(16): 9161-9166.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408(6814): 796-815.

Trick M, Long Y, Meng JL and Bancroft I (2009). Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnology Journal 7(4): 334-346.

Tuskan GA, DiFazio S, et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313(5793): 1596-1604.

Upadhyay SK, Singh DP and Saikia R (2009). Genetic Diversity of Plant Growth Promoting Rhizobacteria Isolated from Rhizospheric Soil of Wheat Under Saline Condition. Current Microbiology 59(5): 489-496.

Valouev A, Ichikawa J, et al. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Research 18(7): 1051-1063.

Vierling RA and Nguyen HT (1992). Use of RAPD markers to determine the genetic diversity of diploid, wheat genotypes. Theoretical and Applied Genetics 84(7-8): 835-838.

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS and Doebley J (2005). An analysis of genetic diversity across the maize genome using microsatellites. Genetics 169(3): 1617-1630.

Vitte C and Bennetzen JL (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences of the United States of America 103(47): 17638-17643.

Vos P, Hogers R, et al. (1995). AFLP - a New Technique for DNA-Fingerprinting. Nucleic Acids Research 23(21): 4407-4414.

Walling JG, Shoemaker R, Young N, Mudge J and Jackson S (2006). Chromosome-level homeology in paleopolyploid soybean (Glycine max) revealed through integration of genetic and chromosome maps. Genetics 172(3): 1893-1900.

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT and Powell W (1997). Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Molecular & General Genetics 253(6): 687-694.

Weber JL and May PE (1989). Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain-Reaction. American Journal of Human Genetics 44(3): 388-396.

Westermeier P, Wenzel G and Mohler V (2009). Development and evaluation of single-nucleotide polymorphism markers in allotetraploid rapeseed (Brassica napus L.). Theoretical and Applied Genetics 119(7): 1301-1311.

Wilkstrom N, Savolainen V and Chase MW (2001). Evolution of the angiosperms: Calibrating the family tree. Proceedings of the Royal Society Biological Sciences Series B 268(1482): 2211-2220.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990). DNA Polymorphisms Amplified by Arbitrary Primers Are Useful as Genetic-Markers. Nucleic Acids Res 18(22): 6531-6535.

Wu CC, Sun SK, Nimmakayala P, Santos FA, Meksem K, Springman R, Ding K, Lightfoot DA and Zhang HB (2004). A BAC and BIBAC-based physical map of the soybean genome. Genome Research 14(2): 319-326.

Xu ZY, Kohel RJ, Song GL, Cho JM, Yu J, Yu SX, Tomkins J and Yu JZ (2008). An integrated genetic and physical map of homoeologous chromosomes 12 and 26 in Upland cotton (G. hirsutum L.). BMC Genomics 9.

Yu Q, Tong E, et al. (2009). A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genomics 10.

Zhang HB and Wing RA (1997). Physical mapping of the rice genome with BACs. Plant Molecular Biology 35(1-2): 115-127.

Zietkiewicz E, Rafalski A and Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain-reaction amplification. Genomics 20(2): 176-183.

11 Appendix

11.1 Abbreviations

AFLP	amplified fragment polymorphism
BAC	bacterial artificial chromosome
BES	bacterial artificial chromosome end sequence
bp	base pairs
CAP	cleavage amplification polymorphism
cM	centi Morgan
DArT	diversity arrays technology
DNA	deoxyribonucleic acid
e.g.	exempli gratia
EST	expressed sequence tag
FES	fosmid end sequence
FISH	fluorescence in situ hybridization
FPC	fingerprinted contig
HICF	high information content fingerprinting
i.e.	id est
ISSR	inter-simple sequence repeat
kbp	kilo base pairs
LINE	long interspersed element
LTR	long terminal repeat
MAS	marker-assisted selection
Mbp	mega base pairs
MIP	molecular inversion probe
MITE	miniature inverted-repeat transposable element
mya	million years ago
NGS	next generation sequencing
ORF	open reading frame
PAC	P1 artificial chromosome
PCR	polymerase chain reaction
RAPD	random amplified polymorphic DNA

RFLP	restriction fragment length polymorphism
ROMA	representational oligonucleotide microarray analysis
SFP	single feature polymorphism
SINE	short interspersed element
SNP	single nucleotide polymorphism
S-SAP	sequence-specific amplification polymorphism
SSLP	simple sequence length polymorphism
STS	sequence tagged site
TAC	transformation-competent artificial chromosome
TE	transposable element
WGD	whole genome duplication
WGS	whole genome shotgun
YAC	yeast artificial chromosome

11.2 Curriculum Vitae

For reasons of privacy protection, a complete CV is not included in the electronic version of the thesis.

.

11.3 Danksagung (Acknowledgements)

Ich möchte mich bei Herrn Dr. Himmelbauer für die Überlassung der Projekte, seine Betreuung und die Erstellung des Gutachtens bedanken.

Ebenfalls möchte ich Herrn Prof. Schuster für seine Bereitschaft danken, auch sehr kurzfristig die Betreuung meiner Arbeit am Fachbereich Biologie, Chemie, Pharmazie der Freien Universität Berlin zu übernehmen.

Herrn Prof. Lehrach danke ich für die Möglichkeit meine Arbeit in seiner Abteilung am MPI anfertigen zu können.

Ein ganz besonderer Dank geht an diejenigen Mitglieder der AG Himmelbauer, die für eine kollegiale Atmosphäre des Zusammenhalts gesorgt haben. Insbesondere danke ich Marion Klein und Stefanie Palczewski für hervorragende Unterstützung und Ausführung von Laborarbeiten und ihr Durchhaltevermögen auch in schwierigeren Phasen und Tobias Nolden, der bis zum Schluss mit mir die Stellung gehalten hat.

Desweiteren danke ich allen Mitgliedern der GABI PhysMap- und BeetSeq - Konsortien für die stimulierende, konstruktive Atmosphäre während der Meetings und für die fruchtbare Zusammenarbeit.

Während meiner Arbeit am MPI haben mich viele Kollegen begleitet und mir auf verschiedene Weise eine unvergessliche Zeit bereiteten (auch außerhalb des Instituts), deren einzelne Aufzählung den Rahmen sprengen würde. Hervorheben möchte ich Ute Nonhoff, mit der ich zu weitreichenden Themengebieten immer regen Gedankenaustausch betreiben konnte, die mit mir schlimme Ohrwürmer geteilt hat und mich bei der Überarbeitung meiner Dissertation ebenso wie Martin Kerick unterstützt hat.

Ebenso gilt mein herzlichster Dank meinen Eltern, die mich immer in meinen Vorhaben unterstütz haben und mir stets Vertrauen entgegenbrachten.

Meinen Freunden danke ich für die nötige Ablenkung und Rückhalt außerhalb der Forschung.

11.4 Selbständigkeitserklärung

Hiermit erkläre ich, dass ich diese Arbeit selbst verfasst habe sowie keine anderen als die angegebenen Quellen und Hilfsmittel in Anspruch genommen habe. Ich versichere, dass diese Arbeit in dieser oder anderer Form keiner anderen Prüfungsbehörde vorgelegt wurde.

Cornelia Lange Berlin,