Theor Appl Genet (2010) 121:549–565
DOI 10.1007/s00122-010-1329-2

ORIGINAL PAPER

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

Cornelia Lange · Lukas Mittermayr · Juliane C. Dohm · Daniela Holtgräwe · Bernd Weisshaar · Heinz Himmelbauer

Received: 2 December 2009 / Accepted: 22 March 2010 / Published online: 9 April 2010

Abstract We describe a novel approach for high-throughput development of genetic markers using representational oligonucleotide microarray analysis. 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 sequences, 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.

Communicated by T. Luebberstedt.

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

C. Lange and L. Mittermayr contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1329-2) contains supplementary material, which is available to authorized users.

C. Lange · L. Mittermayr · J. C. Dohm · H. Himmelbauer
Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany

J. C. Dohm · H. Himmelbauer (✉)
Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra, 08003 Barcelona, Spain
e-mail: heinz.himmelbauer@crg.es

D. Holtgräwe · B. Weisshaar
Center for Biotechnology (CeBiTec), University of Bielefeld, 33594 Bielefeld, Germany

Present Address:
L. Mittermayr
Department of Biology, Ludwig-Maximilians-University Munich, Großhaderner Strasse 2, 82152 Planegg-Martinsried, Germany

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*) (http://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*), 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 3,000 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 (http://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 invention 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 advantage 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). While Lucito et al. (2000) and Lezar 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 s at 30 Hz. Subsequently, 1.3 ml hot (65°C) extraction buffer (0.1 M Tris HCl; 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 min was performed, followed by 25 cycles consisting of 94°C for 30 s, 65°C for 30 s and 72°C for 3 min, and a final elongation step at 72°C for 10 min. 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 44 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 cut-off) = 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 *Arabidopsis 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 9,764 *B. vulgaris* sequences had matches with *A. thaliana*, 9,634 with *P. trichocarpa*, and 9,244 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 × 44 K (Agilent, Santa Clara, CA, USA) and 2 × 105 K (Agilent) for screening of the parental genotypes and 8 × 15 K (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 × 44 K and 2 × 105 K 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 15 K 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 15 K 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 290–6,319 bp were selected using the Agilent eArray platform and placed onto the 15 K array in fivefold replicates.

Amplicon labeling and array hybridization

In case of the F_2 samples for the 15 K 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 h followed by heat inactivation at 65°C for 10 min. The recommended amounts of DNA template for labeling varied between the different array formats and were 500 ng for the 15 K array, 1 μg for the 44 K array and 1.5 μg for the 105 K 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 15 K array, 44 μl for the 44 K array and 104 μl for the 105 K array). Specific labeling activity (pmol dye/μg DNA) of the samples was examined using a NanoDrop ND-1000 UV-VIS Spectrophotometer (NanoDrop 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 h.

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 44 and 105 K arrays; version 9.5.3.1 for 15 K arrays) applied on the individual grid file for each array format and the Agilent GE1-v5_91_0806 protocol (44 and 105 K arrays) and GE1-v5_95_Feb07 protocol with enabled "Local background method" (15 K arrays). For 44 and 105 K 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 15 K array as polymorphic marker candidates. For normalization of the 15 K 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 44 and 105 K 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 loci, 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 et al. 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 2,000 bp (Saiki et al. 1988). Within a mixture of differently sized templates, PCR preferentially generates products in the size range of 200–1,200 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 200–1,200 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

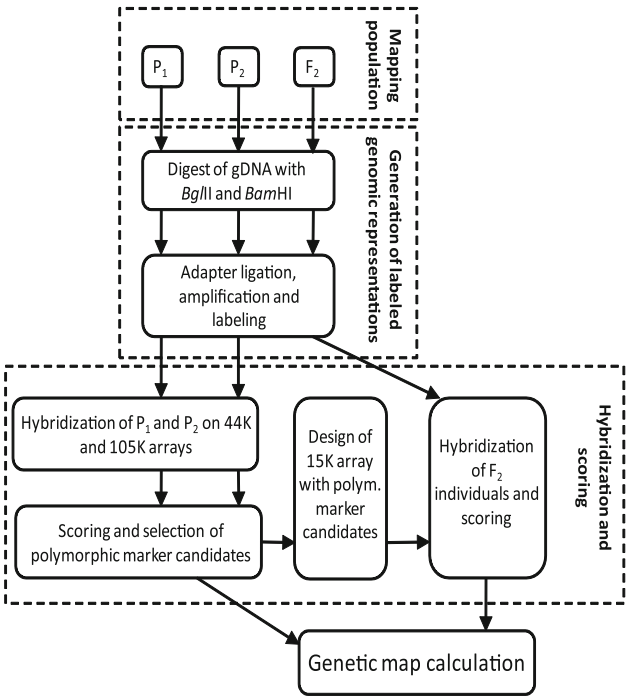

Fig. 1 Flow diagram to illustrate the mapping strategy

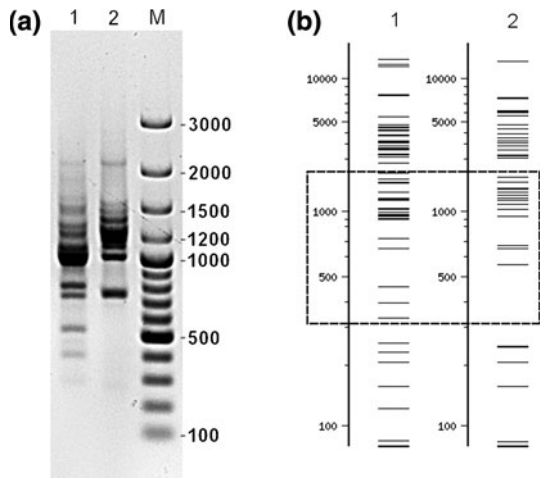

Fig. 2 Verification of size dependent preferential amplification of restriction fragments by *Taq* polymerase. **a** Purified amplicons of BAC clones SBI-153H13 (*lane 1*) and ZR-47B15 (*lane 2*). Sizes of marker bands (*lane M*) are indicated in base pairs. Only fragments in the size range of approximately 250–1,500 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*

with *Bgl*II and *Bam*HI and analyzed the amplified fragments by gel electrophoresis (Fig. 2). Fragments within an approximate size range of 250–1,500 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 44 and 105 K custom microarrays.

The Agilent 44 K array contained 45,220 oligonucleotide positions, named features, of which 1,428 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 105 K 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 44 and 105 K arrays is shown in Fig. 3a.

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

After hybridization of the labeled K1P1 and K1P2 amplicons, respectively, on the 105 and 44 K 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 15 K 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 15 K 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 15 K arrays later

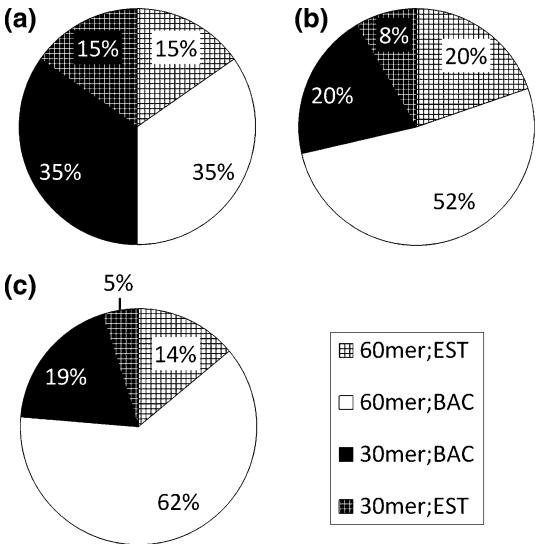

Fig. 3 Size distribution and source sequence origin of oligonucleotide arrays. **a** Sugar beet oligonucleotides on 44 and 105 K arrays used for identification of marker candidates. The total number of oligonucleotides on both arrays comprised 146,554, **b** 14,915 oligonucleotides selected from 44 and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

on (see "Material and methods"). 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 44 and 105 K 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 15 K oligonucleotide arrays with internal control features

For genotyping, genomic representations of 196 F_2 individuals from the K1 mapping population were hybridized on the 15 K arrays containing the polymorphic marker candidates. Results were obtained only for 184 F_2 genotypes, 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 15 K 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. Figure 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 15 K 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, 1,204 features were selected from the 15 K array (Fig. 3c), of which the ratio of 60-mers (76%) to 30-mers (24%) was almost the same as it was on the whole 15 K 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 15 K array. This suggests that the bias towards K1P1-positive features

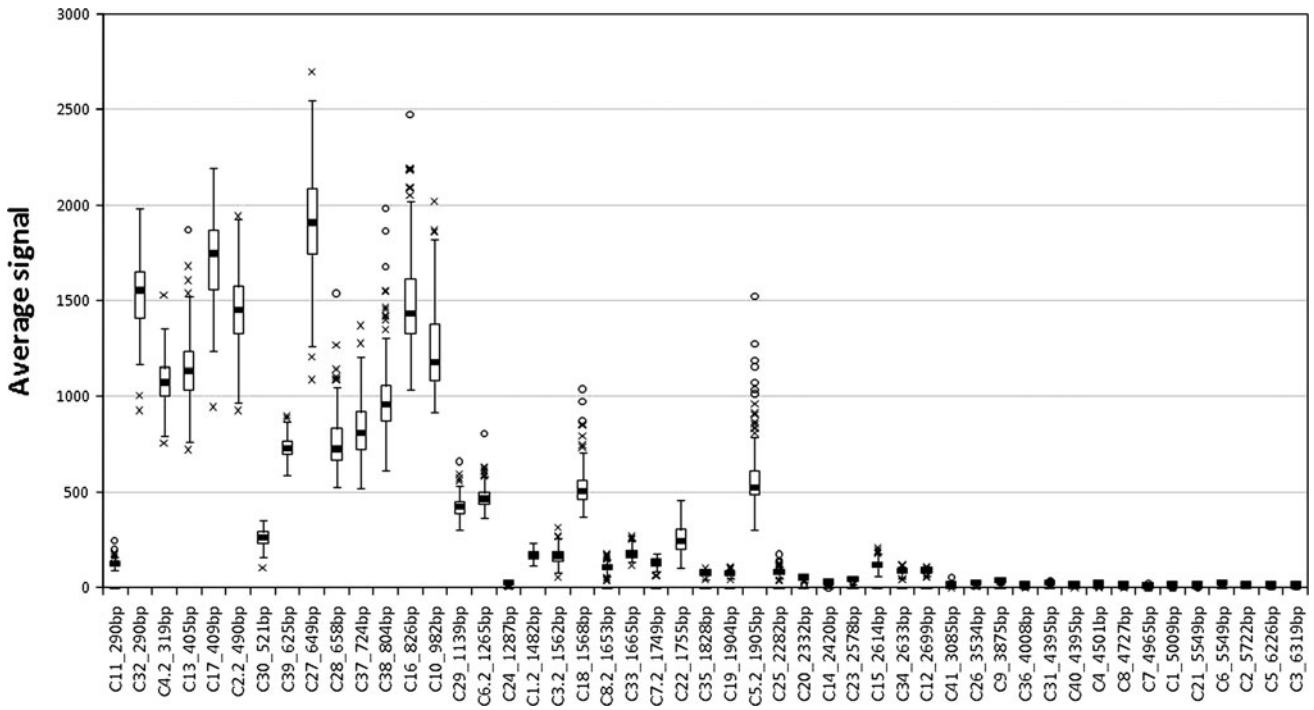

Fig. 4 Box plot of hybridization signals of mouse BAC control oligonucleotides from all genotypes. The signals were normalized and the average values of the five replicates were plotted. Oligonucleotides were ordered along the *x* axis in ascending order according the size of the restriction fragments 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

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 44 K- and 105 K-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. 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

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
$\sum$	595	1589.5	280	911.1	511	1668.6

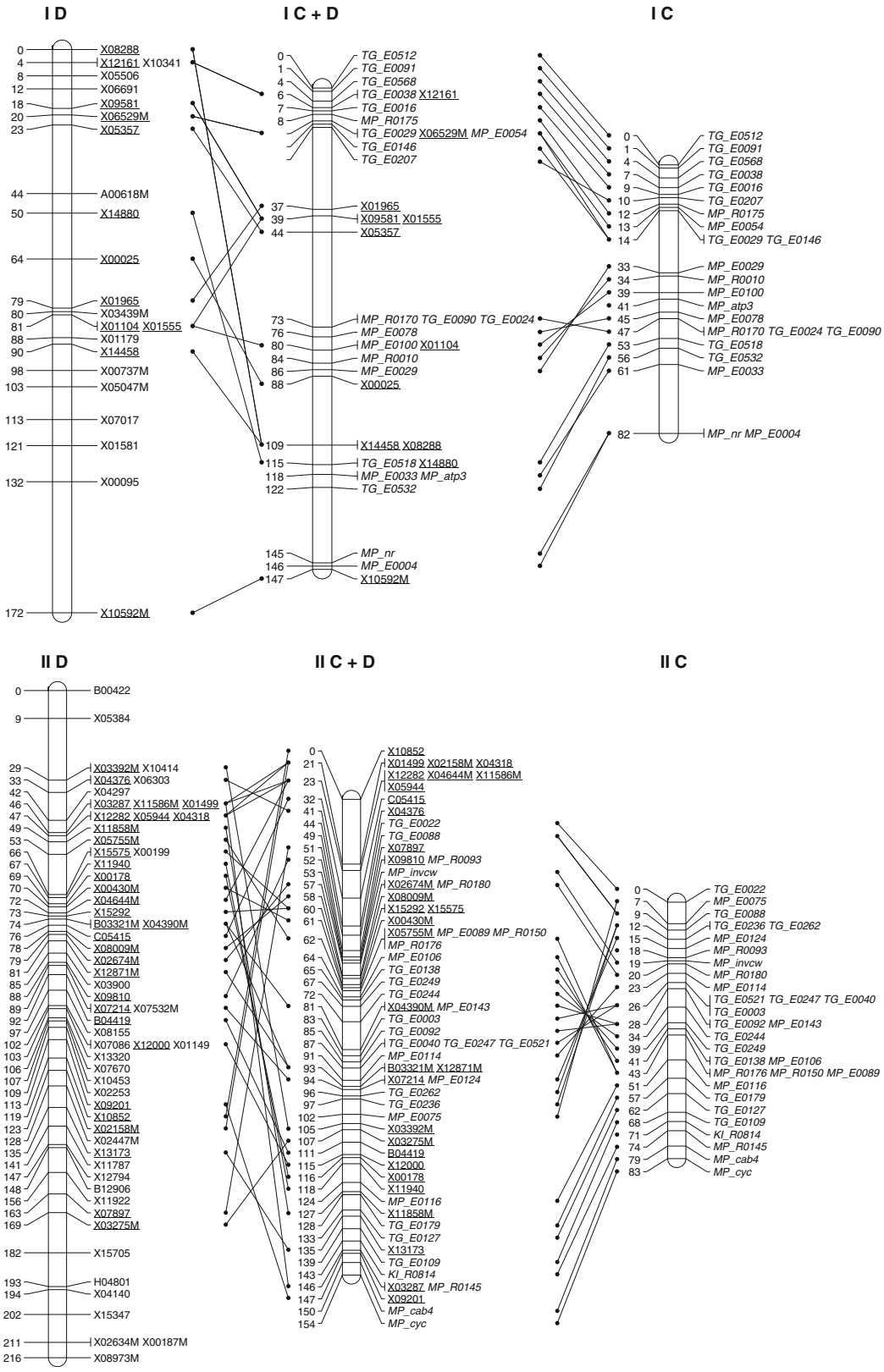

Fig. 5 Sugar beet linkage map constructed with co-dominant markers from Schneider et al. (2007) combined with dominant markers (C + D), only co-dominant markers (C) and only dominant markers (D). Marker names and the cumulative genetic distances in cM are indicated. Corresponding markers between the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

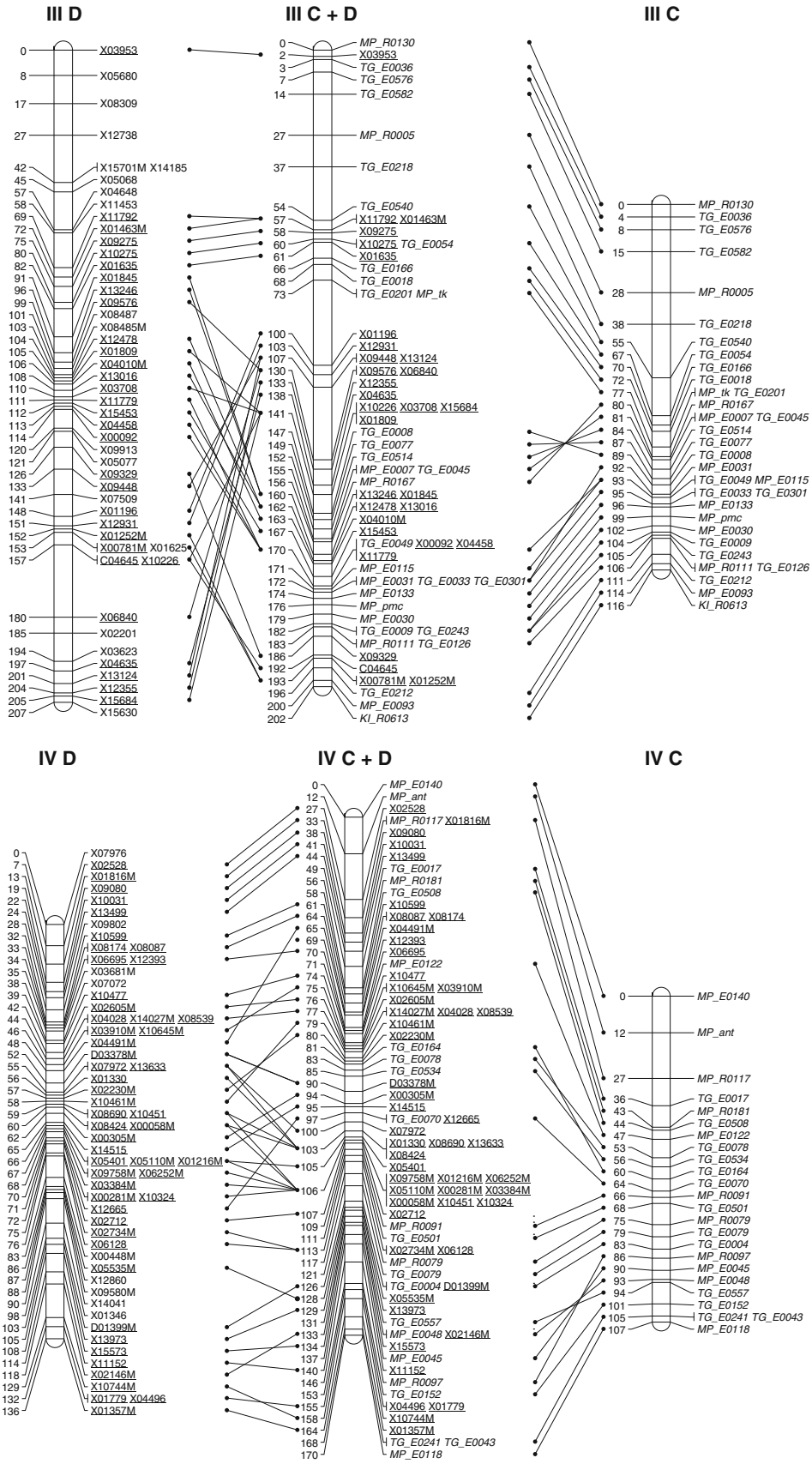

Fig. 5 continued

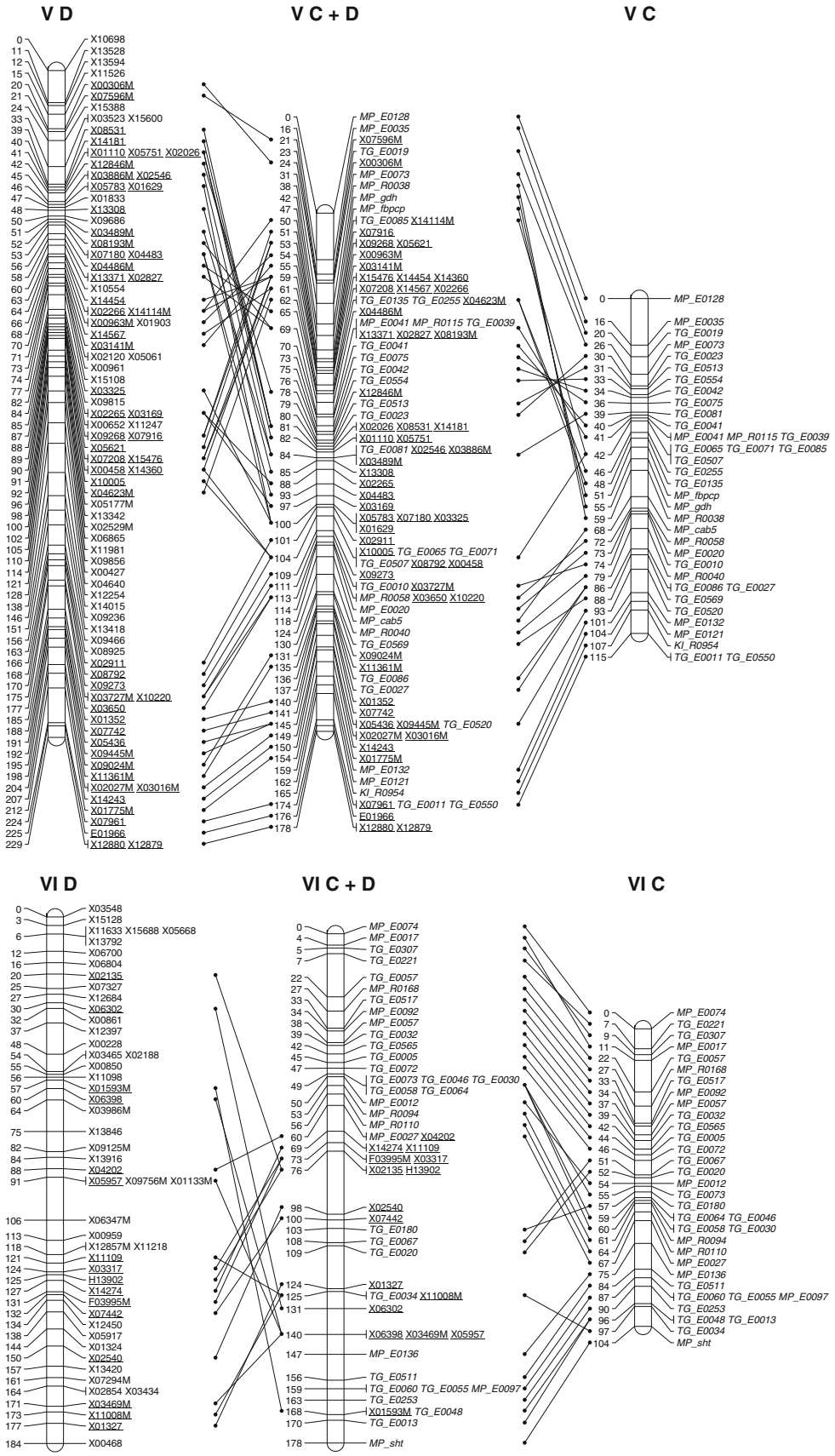

Fig. 5 continued

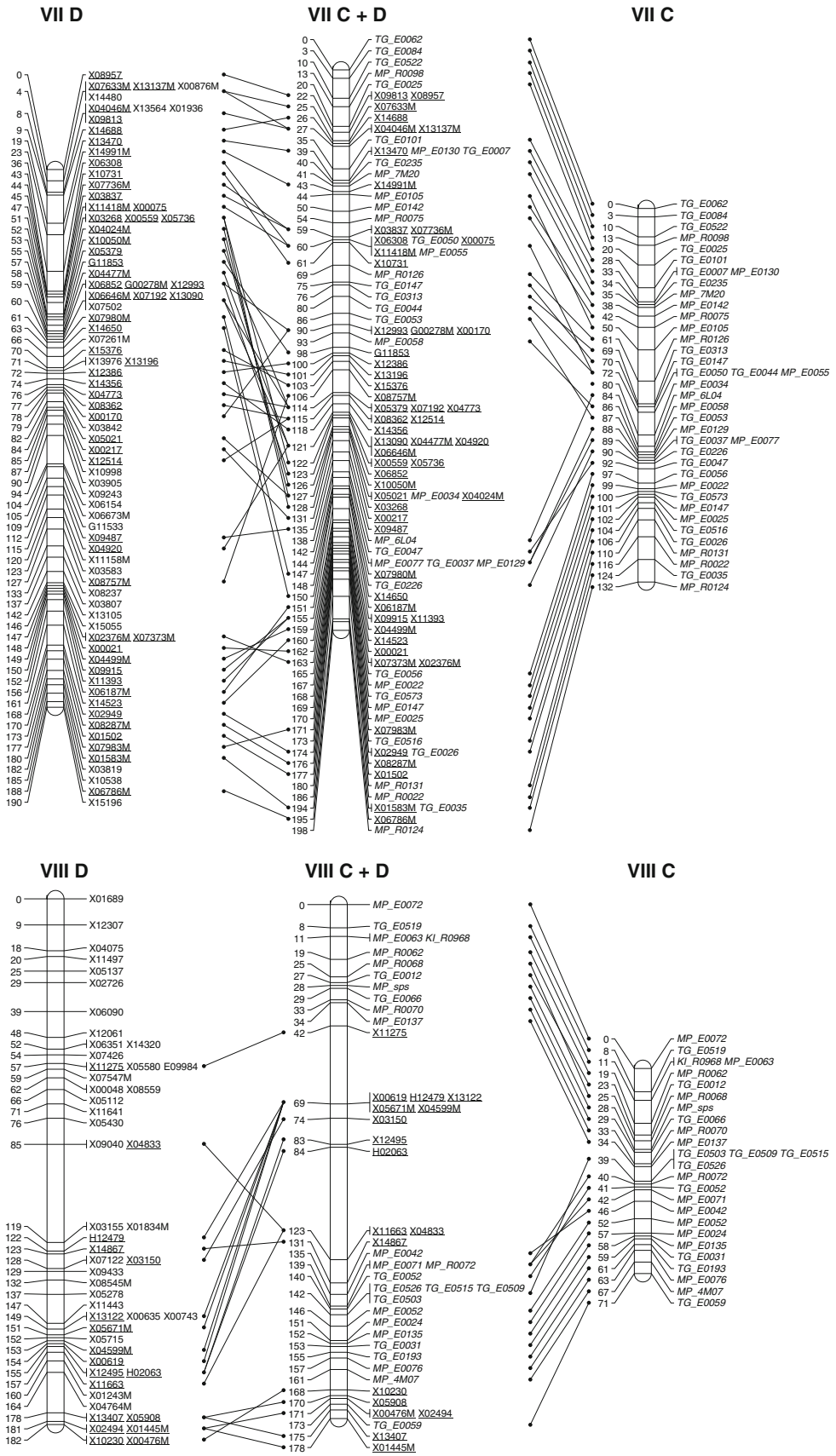

Fig. 5 continued

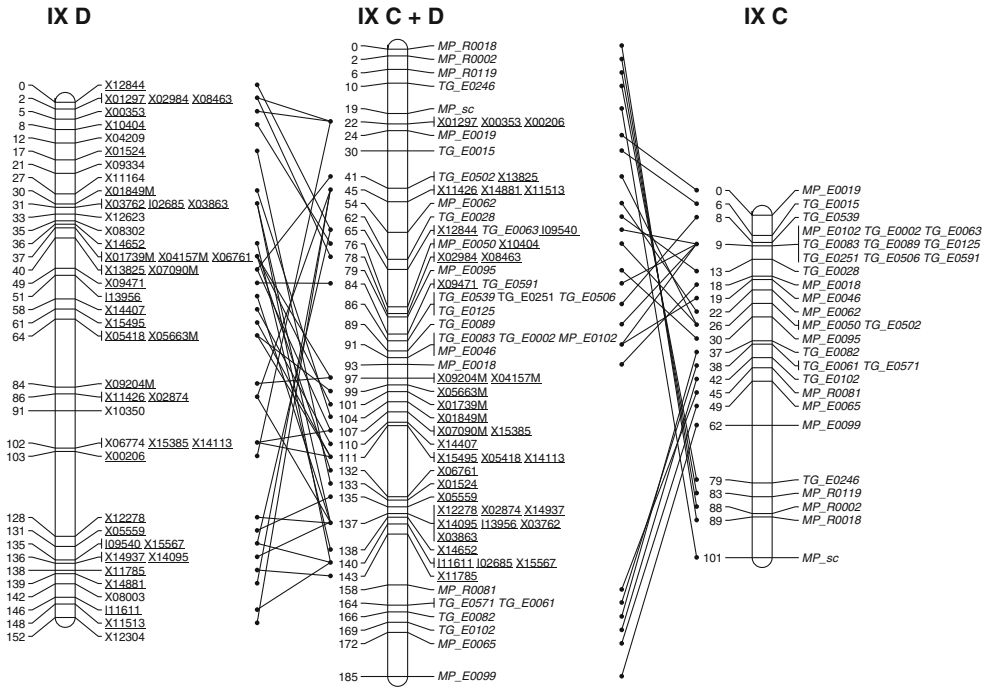

Fig. 5 continued

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 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 arrays, 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 (Fig. S3). 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 500–1,000 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 array 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.

Acknowledgments 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

Castle J, Garrett-Engele P, Armour CD, Duenwald SJ, Loerch PM, Meyer MR, Schadt EE, Stoughton R, Parrish ML, Shoemaker DD, Johnson JM (2003) Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. Genome Biol 4:R66

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

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

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, 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

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

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

Huang S, Li R, Zhang Z et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–1281

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

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 USA 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, 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

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 USA 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, 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

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 LeinJ, 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