the plant journal

SEB Society for Experimental Biology

The Plant Journal (2012) **70**, 528–540 doi: 10.1111/j.1365-313X.2011.04898.x

Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome (*Beta vulgaris*)

Juliane C. Dohm[1,2], Cornelia Lange[1], Daniela Holtgräwe[3], Thomas Rosleff Sörensen[3], Dietrich Borchardt[4], Britta Schulz[4], Hans Lehrach[1], Bernd Weisshaar[3,*] and Heinz Himmelbauer[1,2,*]

[1]*Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany,*

[2]*Centre for Genomic Regulation (CRG) and UPF, C. Dr Aiguader, 88, 08003 Barcelona, Spain,*

[3]*Bielefeld University, CeBiTec, Universitätsstr. 27, 33615 Bielefeld, Germany, and*

[4]*KWS SAAT AG, Grimsehlstr. 31, 37555 Einbeck, Germany*

Received 13 October 2011; revised 20 December 2011; accepted 23 December 2011; published online 14 February 2012.
*For correspondence (e-mail heinz.himmelbauer@crg.es or weisshaa@cebitec.uni-bielefeld.de).

SUMMARY

Sugar beet (*Beta vulgaris*) is an important crop plant that accounts for 30% of the world's sugar production annually. The genus *Beta* is a distant relative of currently sequenced taxa within the core eudicotyledons; the genomic characterization of sugar beet is essential to make its genome accessible to molecular dissection. Here, we present comprehensive genomic information in genetic and physical maps that cover all nine chromosomes. Based on this information we identified the proposed ancestral linkage groups of rosids and asterids within the sugar beet genome. We generated an extended genetic map that comprises 1127 single nucleotide polymorphism markers prepared from expressed sequence tags and bacterial artificial chromosome (BAC) end sequences. To construct a genome-wide physical map, we hybridized gene-derived oligomer probes against two BAC libraries with 9.5-fold cumulative coverage of the 758 Mbp genome. More than 2500 probes and clones were integrated both in genetic maps and the physical data. The final physical map encompasses 535 chromosomally anchored contigs that contains 8361 probes and 22 815 BAC clones. By using the gene order established with the physical map, we detected regions of synteny between sugar beet (order Caryophyllales) and rosid species that involves 1400–2700 genes in the sequenced genomes of Arabidopsis, poplar, grapevine, and cacao. The data suggest that Caryophyllales share the palaeohexaploid ancestor proposed for rosids and asterids. Taken together, we here provide extensive molecular resources for sugar beet and enable future high-resolution trait mapping, gene identification, and cross-referencing to regions sequenced in other plant species.

Keywords: *Beta vulgaris*, genome mapping, synteny, genome evolution, map integration.

INTRODUCTION

Beta vulgaris is a highly variable taxon that comprises the cultivated forms chard, beetroot, fodder beet, and sugar beet (*B. vulgaris* ssp. *vulgaris* var. *altissima* Döll). Among angiosperms, sugar beet is classified into the order of Caryophyllales. The beginning sugar beet cultivation dates back 250 years with the discovery by A. S. Marggraf in 1747 that sugar isolated from beets is the same substance as the sugarcane extract. Breeding experiments carried out by F. C. Achard in the late 18th century using fodder beet and chard led to the first sugar beet line, the white Silesian beet (Fischer, 1989). Traits of interest to breeders include sugar content, beet yield, quality traits such as the content of α-amino-nitrogen, potassium, and sodium, the control of bolting, the seed quality, as well as resistance against fungal pathogens, nematodes, viruses, and insects. Molecular resources are required for the identification of the genes underlying such traits: genetic markers are used to chromosomally assign and fine-map the trait and a physical map is needed to access the genomic region between the markers that flank the trait. Candidate gene discovery might be supported by integration with genome sequences of other species via map anchors of known sequence.

In recent years, sugar beet genetic maps that integrate different marker types have been constructed (Schneider

et al., 2007; Lange *et al.*, 2010). Transcript sequencing projects have led to a collection of about 30 000 sugar beet expressed sequence tags (ESTs) in the NCBI dbEST database. In the study of Herwig *et al.* (2002) many different genes were identified from relatively few ESTs by sequencing a cDNA clone set normalized via oligonucleotide fingerprinting. The Gene Index Project (http://www.compbio.dfci.harvard.edu; formerly TIGR Gene index) has prepared a BvGI Gene Index that consists of 17 186 unique entries (March 2011). Several genomic libraries in bacterial artificial chromosome (BAC) vectors (Hohmann *et al.*, 2003; McGrath *et al.*, 2004; Hagihara *et al.*, 2005; Jacobs *et al.*, 2009) and one sugar beet fosmid library (Lange *et al.*, 2008) have become available.

Physical maps are of value for many types of application (Meyers *et al.*, 2004; Dolezel *et al.*, 2007; Varshney *et al.*, 2010): in map-based cloning, physical maps allow immediate access to genomic regions to identify genes of interest. In the context of genomic sequencing projects, BAC maps are invaluable resources for the independent confirmation of an assembly produced from whole genome shotgun sequencing data. BACs facilitate targeted high coverage sequencing of regions that contain gaps in the sequence draft. Finally, the genomic clones are of value for transgenic approaches, e.g. for complementing a mutation by a functional gene copy contained in the BAC.

Different strategies have been developed to construct physical maps. Clone fingerprinting (Coulson *et al.*, 1986) relies on restriction digestion of clone DNA followed by size fractionation of the resulting fragments. Clones that share fragments are considered as overlapping and are grouped into contigs. This approach has been used successfully to prepare clone-based maps from many plant and animal species (Meyers *et al.*, 2004). While straightforward to implement, the fingerprinting approach has two disadvantages. First, the established contigs are not linked to genes or chromosomes and integration with genome resources such as linkage maps or gene indices needs to be addressed in separate projects. Second, fingerprinting is sensitive to haplotype variation and is best applied on genomic clones that have been prepared from homozygous genotypes. The combination of clone resources across genotypes will lead to highly fragmented maps. The latter also applies to the whole genome profiling method (van Oeveren *et al.*, 2011), which is based on the generation of short sequence tags from pooled clones after restriction enzyme digestion.

In this work, we apply a hybridization-based method adapted from the medaka physical map project (Khorasani *et al.*, 2004). Short 35mer oligonucleotide probes are designed from sequences of interest, e.g. genetic map anchors or ESTs, and used to hybridize arrayed BACs. The hybridization signals identify probe matches on clones that in turn are used for ordering clones into contigs. The information available for the probes adds substantial value to the resulting contig map and provides integration points to linkage maps and gene location assignments. Different genotypes can be included in the same map if probes are derived from genes, as coding regions are mostly conserved between genotypes. Conceptionally, the strategy is related to the 'overgo' approach (Cai *et al.*, 1998), which uses pairs of 24mer oligonucleotides with an eight-base overlap for labeling, but more straightforward as only a single oligonucleotide per locus needs to be designed.

Although dicot genomes have undergone extensive rearrangements and gene shuffling, regional collinearity of genes can be detected (Tuskan *et al.*, 2006; Jaillon *et al.*, 2007; Tang *et al.*, 2008a; Salse *et al.*, 2009; Argout *et al.*, 2011). Such comparative approaches are motivated by the interest in reconstruction of the genome of a common angiosperm ancestor and by the potential transfer of information about candidate genes from one species to another. Based on the genome analysis of grapevine, a hexaploid ancestor that encompassed seven different linkage groups before triplication was proposed for rosid species (Jaillon *et al.*, 2007). Evidence for the ancestral role of a hexaploid genome was also found for asterid species (Tang *et al.*, 2008b; Cenci *et al.*, 2010), a situation that implies the question if a palaeohexaploid scenario applies to the origin of further groups within the core eudicotyledons. Synteny between Caryophyllales and rosids was reported for a 153-kbp segment of the sugar beet genome (Dohm *et al.*, 2009).

In this study, we present a high-resolution and high-density genetic map (and incorporate and greatly extend previous sugar beet genetic maps), a genome-wide physical map of sugar beet in BACs, the integration of the two types of maps, and the detection of synteny with other plant genomes leading to conclusions about the common ancestry of Caryophyllales, rosids, and asterids.

RESULTS

Resources and strategy for gene-based physical mapping

The genomic resources for genome-wide physical mapping of the sugar beet genome are two BAC clone libraries comprised of 27 648 clones each. One library is a sublibrary of the SBI library (72 size-selected 384-well plates that contained clones with an average insert size of 150 kbp) from a heterozygous sugar beet line, the other library is a sublibrary of the ZR library (72 randomly selected 384-well plates, average insert size 110 kbp) from the doubled haploid line KWS2320 (Hohmann *et al.*, 2003). Based on the estimated sugar beet genome size of 758 Mbp (Arumuganathan and Earle, 1991) these two sublibraries represent 5.5-fold (SBI) and fourfold (ZR) coverage of the genome, respectively.

We aimed to link 35mer oligonucleotide probes with BAC clones by hybridization experiments (Figure 1). As the clone libraries cover the genome several times over, each probe is expected to hybridize with several clones. By using a large

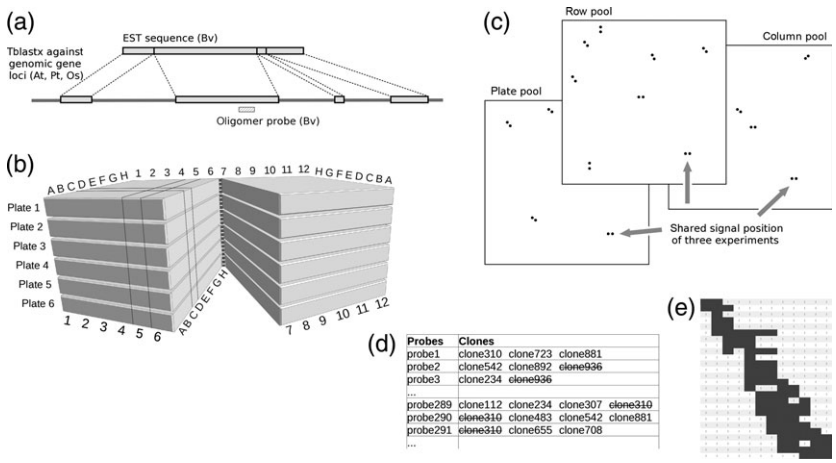

Figure 1. Overview on strategy and experimental settings for the generation of physical map data and map construction.
(a) Exon junctions within sugar beet expressed sequence tags (ESTs) (Bv: *Beta vulgaris*) were determined by sequence comparison to gene loci of *Arabidopsis thaliana* (At), *Populus trichocarpa* (Pt), *Oryza sativa* (Os). Oligonucleotides (35-mers) were selected from exonic regions as oligomers spanning exon-exon borders cannot hybridize over their entire length with a genomic clone insert.
(b) Oligomers were pooled in accordance with their spatial position within a block of six 96-well plates (virtually cut into halves) before being used as probes in hybridization experiments. Each half-block contained 288 probes that were combined into 20 pools: eight row pools of 36 probes each, six column pools of 48 probes each, six half-plate pools of 48 probes each. Names of rows (A–H) and columns (1–12) are indicated. As an example, row pool D and column pool 5 are framed by black lines. A single probe can be addressed via three coordinates: plate number, row letter, and column number.
(c) Each pool of probes was hybridized against two BAC macroarrays that contained clones of the SBI and ZR libraries, respectively. Three macroarrays of one library and a few hybridization signals are shown. Signals at the same positions (marked by arrows) of three experiments represent a clone that hybridized with the probe addressed by the three corresponding coordinates within the probe block.
(d) After resolution of pooled hybridization results to single probe results (deconvolution) a list of probe-clone assignments was generated. One probe (column 1) can be assigned to several clones (column 2). To avoid false-positive results we discarded clones that occurred more than once for probes within the same block half (crossed-out clones).
(e) Probes (columns) are brought into an order in accordance with shared clones (rows), and contigs are built by clone overlaps; connections between probes and clones are shown as dark squares (see also Figure 4).

number of randomly distributed probes it is expected that there will be clones that hybridize with more than one probe, and that single probe–clone connections can be assembled into rows of overlapping clones (contigs).

The following data resources to generate oligonucleotide probes for the hybridization experiments were used: (i) ESTs of *Beta vulgaris* transcripts from Herwig *et al.* (2002), Schneider *et al.* (2007), and published with this work (GenBank accession numbers in Data S1). (ii) Sequences based on markers for genetic map construction that were developed previously (Schneider *et al.*, 2007, as well as mapping data provided by KWS SAAT AG), both ESTs and genomic sequences. (iii) Genomic sequences from transcribed regions of the sugar beet genome (genotype KWS2320) published with this work. (iv) A subset of sequences from a sugar beet Roche-454 cDNA sequencing project (Data S2).

Optimal source sequences for probe design are characterized as representing single-copy loci in the genome and as being haplotype independent. The former point is important for the unambiguous assembly of contigs, the latter one facilitates the joint mapping of different haplotypes (here, the ones represented by the SBI and ZR BAC libraries). Although it is not possible to predict unique and haplotype-independent segments without additional information, transcribed loci are more likely to have these properties than other genomic regions. The divergence between *Beta vulgaris* haplotypes is low in coding regions (Dohm *et al.*, 2009), and multiple gene loci are not generally expected (except for particular cases like gene clusters, gene families, pseudogenes). There is no evidence for recent genome duplication in *Beta vulgaris* (Pillen *et al.*, 1992; Schondelmaier *et al.*, 1996; Hallden *et al.* 1998). Thus, we considered sequences of transcribed regions of the genome to be an appropriate data source for the generation of a large number of probe sequences.

We clustered publicly available ESTs and got a non-redundant set of 14 517 sequences. Source sequences from Roche-454 cDNA sequencing and ESTs used for genetic mapping by Schneider *et al.* (2007) were added separately (565 sequences) and we accept that some of these share sequence information with the clustered non-redundant set. We determined the exon structure of the ESTs by sequence homology searches with genomic gene loci of *Arabidopsis thaliana*, *Populus trichocarpa*, and *Oryza sativa.* We considered only ESTs that did not suggest high gene copy numbers in the target genomes. After masking known repetitive regions, we designed 35mer oligonucleotide probes for one exon per EST and for each of the genomic sequences. In total, probes for 13 748 source sequences were designed (10 330 ESTs, 3323 genomic sequences from gene loci, 95 genomic sequences not linked to gene loci).

Each probe was hybridized against the two clone libraries. To reduce the experimental effort we combined probes into pools (Figure 1). Pooling strategies were first established for the screening of clone libraries (Kusumi *et al.*, 1993; Hunter *et al.*, 1994); the application to probes is described by Cai *et al.* (1998) and Khorasani *et al.* (2004). Here, we extended the probe pooling to a three-dimensional scheme that processed each unit of 288 probes in only 20 pooled hybridization experiments (see Data S1 for details). Intersections of three hybridization results on the data level ('deconvolution') led to the information which clones are matched by a single probe. The reduction of experiments implicates certain ambiguities for clone–probe assignments in cases in which a clone has a positive signal with more than one probe within the same experimental unit. In order to test the probability of such events, we simulated 27 648 clone identifiers and 288 probe identifiers and linked these probes and clones randomly with an average of five clones per probe. On average (10 repetitions) 2.6% of all assigned clones were assigned to more than one probe and 24% of the probes were affected. After building virtual probe pools and performing the deconvolution to get back to single probes, clones that occurred only once were assigned correctly but clones that occurred more than once were assigned to 83% of probes. Thus, the combination of simulated pools resulted in a large proportion of probes that contained at least one false-positive clone in their assignment result.

We decided to clean the data by removal of all clones that occurred more than once within the clone–probe assignments of a unit of 288 probes. By doing so, 24% of probes loose at least one correct clone but at the same time we safely eliminated all false-positive clones introduced by the deconvolution step. Later tests during the contig construction phase confirmed that the use of unfiltered data leads to incorrect contigs with arbitrary connections across all chromosomes (which is avoided by our cleaning step).

From all probes hybridized we selected those being positive with at least one clone and with a maximum of 12 SBI clones and a maximum of 10 ZR clones (about two times the number of expected clones per probe) in order to exclude most multi-copy regions of the genome. We kept probe–clone assignments for 12 088 single probe experiments and found 7.4 clones per probe on average. For the two clone libraries separately, the average result was 4.8 SBI clones per probe and 3.7 ZR clones per probe (excluding probes with zero clones of the library under consideration). Of 27 648 clones per library 16 891 (61%) SBI clones and 14 973 (54%) ZR clones were assigned to at least one probe.

The 12 088 probe hybridizations that included controls and confirmations contained probes derived from 11 358 different source sequences (probes from redundant source sequences may be composed of the same 35mer sequence; however, each hybridization experiment effectively has a unique result due to slight variation in experimental conditions and because the removal of clones during the data cleaning step depends on the set of probes jointly contained in an experimental unit). As 99.5% of these sequences are gene-associated, i.e. either ESTs or genomic sequences from gene loci, the gene-rich regions of the genome are expected to be well covered compared with non-genic regions. Some of the source ESTs were integrated in a consensus sequence of an EST cluster and additionally used as a separate EST sequence and resulted in two different probes (141 ESTs affected). After subtraction of such cases, the number of involved genes as inferred from source sequences was 11 217. As larger genes might not be covered by a continuous EST or cluster sequence, the number of different gene loci covered by our data is assumed to be a bit lower.

Extended genetic map and BAC end sequencing

The resources used in the hybridization experiments for the physical mapping were also the basis for marker development to build an extended genetic map for sugar beet. A comprehensive genetic map is of great value for physical contig assembly as the genetic markers, once reliably linked to probes or clones of the physical data, can assist the process of contig assembly and provides information on orientation and order. The K1 mapping population was prepared by selfing a single F1 clonal plant derived from the cross of a doubled haploid line (KWS2320: genotype origin of the ZR BAC library) and a P2 parent (partly selfed line, cloned plant, referred to as K1P2 in Schneider *et al.* (2007)). A previously constructed genetic map based on a K1 mapping population of 97 F2 individuals contained 305 markers (Schneider *et al.*, 2007). We extended the mapping population to a size of 183 F2 individuals. Source sequences for marker development were existing sugar beet ESTs as well as newly generated BAC end sequences from the ZR library. We generated 32 753 unique BAC end sequences singularized in accordance with clone addresses. BACs composed of highly repetitive regions were identified by sequence comparisons (see Data S1) in order to minimize the use of repeat containing BAC ends as marker sources. Previous estimations on the repeat content of sugar beet (Flavell *et al.*, 1974) could be confirmed: about 60% of the BAC end sequence data was found to be repetitive. Based on 1148 selected ESTs (selection in accordance with previously used markers (Schneider *et al.*, 2007) and ESTs used for probe design for physical mapping) and 11 180 BAC end sequences we generated primers for comparative sequencing of the parent plants resulting in 548 (EST) and 4810 (BAC) high quality alignments, respectively. In these alignments we searched for single nucleotide polymorphisms (SNPs) between the KWS2320 and P2 parental accessions. The Sequenom platform for genotyping could be applied on 139 (EST) and 1239 (BAC) alignments, respectively. SNP genotyping of the K1

mapping population resulted in 1127 markers for map construction. We observed a SNP-frequency of 1/324 bp.

We mapped 983 markers (283 EST-based, 700 BAC-based) in the extended K1 population with a total map length of 886 centi-Morgan (cM; Table 1, Figures 2 and S1). The nine linkage groups were assigned to chromosome numbers in accordance with the information of Schondelmaier and Jung (1997). The map contains 670 distinct loci with an average distance of 1.32 cM. Of distinct marker pairs 61% have a distance of ≤1 cM and 4% have a distance of ≥5 cM. Clusters of co-segregating markers are located in the center of linkage groups rather than at the top or bottom and reflected the expected reduced recombination rate in centromere regions. Successive map positions of markers related to their cM assignments are shown in Figure 3.

To illustrate gene densities in relation to recombination rates we determined the positions of EST-derived markers and BAC-derived markers separately (Figure S2). EST-derived markers represent gene densities, and BACs for marker development were chosen at random from the genome. If a high gene density was correlated with high recombination rates, we would expect steep slopes of the graphs to be EST-enriched and flat slopes to show more BAC-derived than EST-derived markers. Most of the chromosomes show at least one region of low recombination rate in which BAC markers accumulate rather than EST markers. However, there are also flat parts of the curves that show EST markers (e.g. in chromosome 7) as well as steep parts with little or no EST marker accumulation (e.g. in chromosome 5). The latter may be attributed to the fact that only a subset of genes is represented by the EST data. The former suggests that gene-rich regions of low recombination rates exist.

Physical contig construction and integration with genetic maps

The genetic positioning of markers in the extended genetic map ('Beetmap') was used to guide the physical contig construction and ordering. Additionally, we used information of genetically mapped markers from maps based on other mapping populations (Schneider *et al.*, 2007; KWS SAAT AG) to assign the corresponding probes to chromosomes.

Table 1 Number of genetically mapped markers in Beetmap and map lengths by chromosome

	Number of markers			
Chromosome	EST	BAC	Total	Map length (cM)
1	18	67	85	93.2
2	26	56	82	87.2
3	30	59	89	107.5
4	33	53	86	108
5	41	104	145	101.8
6	44	103	147	102.7
7	36	96	132	101.9
8	23	67	90	92.5
9	32	95	127	92
Sum	283	700	983	886.8

The physical data contain 493 clones and 364 probes that can be directly linked to Beetmap as BAC end sequences of these clones and the source sequences of these probes, respectively, were used for Beetmap marker development. Together with the chromosomal information for additional 492 probes linked to genetic maps of Schneider *et al.* (2007) (partially overlapping with the set of 364 probes contained in Beetmap) and 2069 probes linked to genetic maps provided by KWS SAAT AG, we extracted probe–clone connections from the physical data (i.e. hybridization results) for each of the nine sugar beet chromosomes. The contig assembly of the data subsets was performed chromosome-wise using the program called probeorder (Mott *et al.*, 1993) followed by manual inspection of the resulting contigs and taking the order of integrated genetic markers into account (see Data S1). The complex nature of the genome and the experimental setting can involve cross-hybridizations, false-positive hybridization signals, and missing data. We built as many connections as possible but used the data records conservatively in order to avoid incorrect connections. In total, we constructed: (i) 391 genetically localized contigs all of which contained at least one Beetmap marker probe or marker clone regardless of the contig size, (ii) additional 144 chromosomally assigned contigs that contained one or more markers from maps of other populations, and (iii) a pool of 1684 remaining contigs. The genetically anchored contigs contain 6010 probes and 16 671 clones; the chromosomally assigned contigs without Beetmap placement contained 2351 probes and 6144 clones. The remaining pool of 3888 probes and 14 364 clones were either in contigs without any genetically mapped probe or clone, in contigs with conflicting chromosomal assignment, or in very small contigs (<10 probes) that were not manually inspected.

Of the 8361 chromosomally located probes, 154 probes appeared in more than one contig in cases in which several possible connections of comparable reliability existed, and 1950 clones appear in more than one chromosome. Several biological and technical factors can lead to multiple inclusions into different contigs: (i) cross-hybridizing probes with clones from duplicated regions or genes in the genome, (ii) contamination of a probe with another probe, (iii) a mixed population of clones in the microtiter plate, (iv) incomplete removal of previous hybridization signals after prolonged exposure to X-ray film; or (v) hybridization artifacts. The variation of signal intensities (dependent upon clone spot size, probe activity, exposure time) can result in false-negative clone information that influences the connection strength between probes in the contigs and potentially leads to separate locations for the same clone. However, most probes and clones could be placed uniquely.

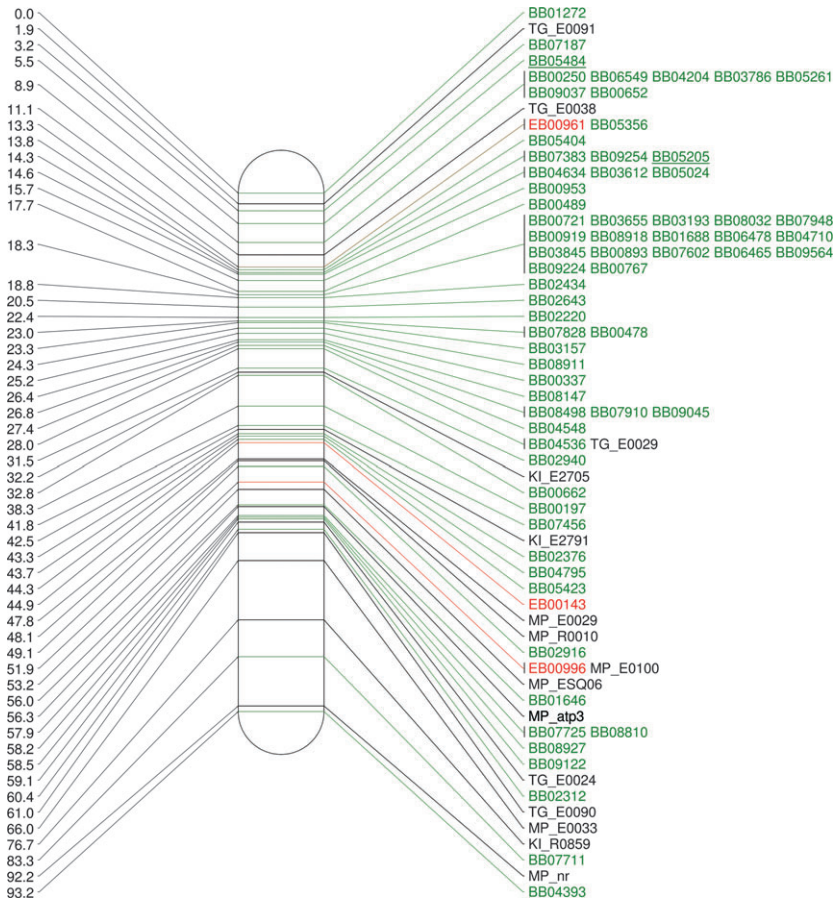

Figure 2. Genetic map of sugar beet chromosome 1 (Beetmap).
Numbers on the left side refer to cM positions. Markers derived from expressed sequence tags (ESTs) sequences are shown in red, markers derived from BAC end sequences are shown in green. Black names indicate markers that were also used for genetic mapping by Schneider *et al.* (2007) (see Figure S1 for figures of all nine Beetmap chromosomes).

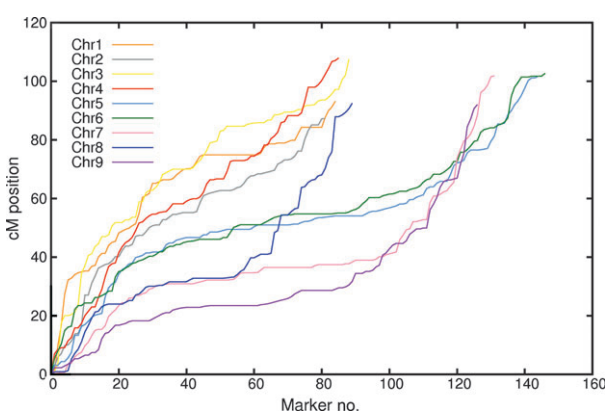

Figure 3. Increase of centi-Morgan (cM) units within the genetic map along with successive marker positions.
See also Figure S2.

The largest two BAC contigs were derived from chromosome 4 (contig01, 136 probes and 246 clones) and chromosome 7 (contig48, 127 probes and 261 clones), respectively, based on the genetic map location of integrated genetic markers. In this paper, we show a smaller contig as an example in Figure 4: this contig contains several genetic markers assigned to chromosome 8 at cM positions 0 and 1 in Beetmap or assigned consistently to chromosome 8 in other mapping populations. The seven Beetmap-linked probes and clones corresponded to six different Beetmap markers at three distinct cM positions. In the BAC contig these probes resolved to seven different positions. Two marker probes (030A05 and 180E08) of different sequence, but derived from the same source EST, resulted in overlapping sets of clones and placed the two probes next to each other within the contig. The remaining marker probes were distributed over the contig and showed an order slightly different from the one in the genetic map. Although not all of the 38 probes were derived from unique source sequences each probe had a unique clone assignment result. Of 98 different clones (65 SBI, 33 ZR) in this contig, 68 clones were hit by at least two probes. The remaining 30 clones (19 SBI, 11 ZR) were hit only once, thus they were integrated in the contig but did not contribute to connections between probes. BAC end sequences are available for 27 clones (ZR) of this contig.

The numbers of probes, clones, and markers for all chromosomally assigned contigs are listed as a summary of chromosomes in Table 2. The individual contigs can be

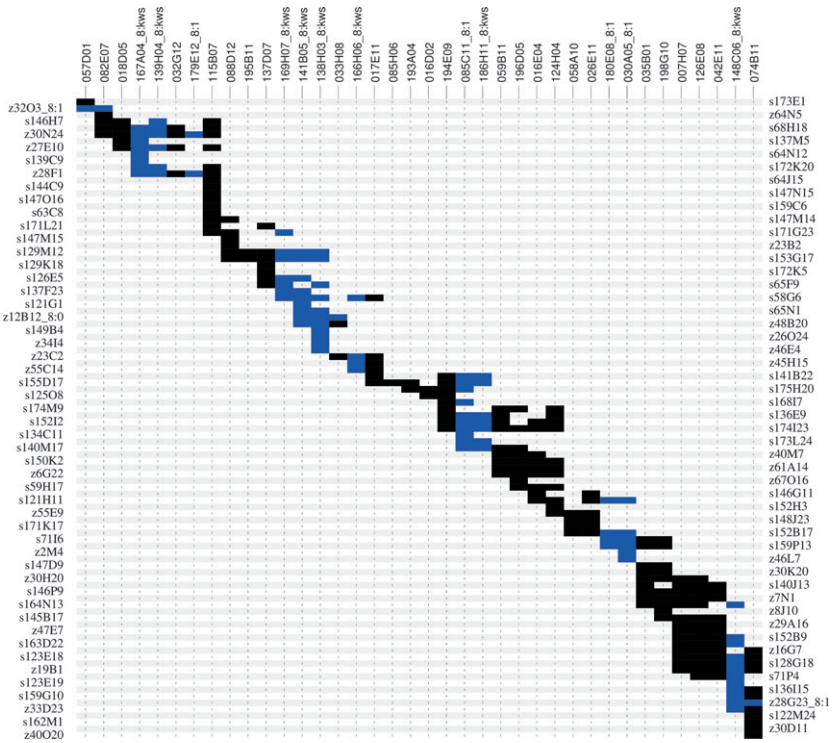

Figure 4. Display of contig 1 of chromosome 8. Probes (in columns) were sorted with the program wprobeorder in accordance with clone assignments (clones in rows, clone names alternating on both sides). Connections between probes and clones are shown as black squares. Genetically mapped probes are indicated by extended probe names or clone names and by blue boxes. The first part of an extension corresponds to the chromosome, the second part to the cM position within Beetmap. The extension 'kws' indicates that the probe was mapped in mapping populations different from Beetmap.

Table 2 Numbers of contigs, probes, and clones contained in the physical map

	Genetically anchored contigs			Chromosomally assigned contigs		
Chromosome	Contigs	Probes	Clones	Contigs	Probes	Clones
1	32	551	1337	19	317	746
2	41	564	1646	20	371	939
3	32	724	1751	14	189	537
4	28	693	1728	13	212	624
5	60	850	2578	17	282	883
6	67	802	2274	15	219	545
7	48	691	2019	15	249	726
8	36	448	1319	14	226	469
9	47	678	1912	17	286	658
Sum	391	6001	16 564	144	2351	6127

Genetically anchored contigs contain Beetmap markers, and chromosomally assigned contigs contain only markers from maps other than Beetmap.

viewed and searched via the public webpage http://bvseq.molgen.mpg.de.

Genetic and physical distances

We expected a physical distance of 855 kbp per 1 cM on average based on the total length of the genetic map of 886 cM and the estimated sugar beet genome size of 758 Mbp. In 26 cases, markers with different genetic positions were located on the same BAC clone in the physical map. As BAC clones have a known physical length of about 100–150 kbp we could use such cases to compare expected and actual physical distances. The smallest difference between two markers located on the same BAC clone was found to be 0.31 cM, a number that would correspond to a physical length of 265 kbp according to the average expectation. This figure corresponds to an elevated genetic distance, and thus to an elevated recombination rate, of about twofold when compared with the maximum physical distance of 100–150 kbp. The majority of marker pairs within one BAC clone differed by 0.4–2.9 cM and represented increased recombination rates of about threefold to 20-fold at these particular genomic regions. The largest genetic distance found between two markers that shared the same BAC clone was 7.5 cM (recombination rate elevated by about 50-fold) and was at the terminal region of chromosome 3.

Synteny analysis and assignment of ancestral linkage groups

The source sequences of the oligomer probes mapped in chromosomally localized physical map contigs were used for a sequence homology search against coding sequences of the genomes of *Arabidopsis thaliana*, *Populus trichocarpa*, *Theobroma cacao*, and *Vitis vinifera*. The matches were filtering by length and sequence identity (and excluded probes which matched more than 10 different genes of the target genome), then we used the information on gene positions within the sequenced genomes and the positions of probes

within the physical map of sugar beet to detect syntenic regions (see Data S1). The criteria to determine contiguous regions of links between sugar beet and the target genomes were tested by application on 913 (the average number of probes per chromosome) sugar beet probes that had been ordered randomly. For 10 000 repetitions, we could not find any 'syntenic' region by chance in poplar and grapevine genomes. As the gene order of *Arabidopsis* is shuffled by several duplications and re-arrangement events (Blanc *et al.*, 2003) there is a higher chance of finding 'synteny' by shuffling the sugar beet probes; we found such regions in 20% of the created random chromosomes. In cacao this rate was 1%.

Parts of all sugar beet chromosomes could be assigned to one or more chromosomes of all target species based on sequence similarity and gene order. In total, we found 26 syntenic regions in grapevine (1424 grapevine genes involved), 35 syntenic regions in cacao (2393 cacao genes involved), 64 syntenic regions in poplar (2718 poplar genes involved), and 73 syntenic regions in *Arabidopsis* (2752 *Arabidopsis* genes involved). In Figure S3, the syntenic regions between all sugar beet chromosomes and these species are visualized using circular plots generated with Circos (Krzywinski *et al.*, 2009). Figure 5 shows such a plot for sugar beet chromosome 1. In this example, we found syntenic regions in poplar chromosomes 2 and 14 (spanning large regions) as well as 5 and 13 (smaller regions), in grapevine chromosome 15 und 18, and in *Arabidopsis* chromosomes 1, 2, 3 and 4. In a previous study, the order of 15 genes annotated within two overlapping BAC clones was analyzed, and microsynteny with poplar chromosomes 2 and 14, grapevine chromosome 15, and *Arabidopsis* chromosomes 1, 2 and 4 was found (Dohm *et al.*, 2009). These two clones were contained in contig 2 of chromosome 1 in our physical map and this finding thus supported the synteny findings based on physical data.

We summarized the syntenic regions in a graphical overview from the perspective of the four target species (Figure 6). Their chromosomes were thoroughly analyzed in previous studies, and seven ancestral linkage groups have been proposed that underwent hexaploidization prior to diversification of the rosid species (Jaillon *et al.*, 2007). Using information on the assignments of the target species' chromosomes to these ancestral linkage groups (for grapevine, poplar, cacao: Jaillon *et al.*, 2007; Argout *et al.*, 2011) we transferred these assignments to the corresponding regions within sugar beet. As in the grapevine genome 16 chromosomes were assigned to only one ancestral linkage group and three (chromosomes 4, 7 and 14) were assigned to two of these linkage groups, we started with this genome

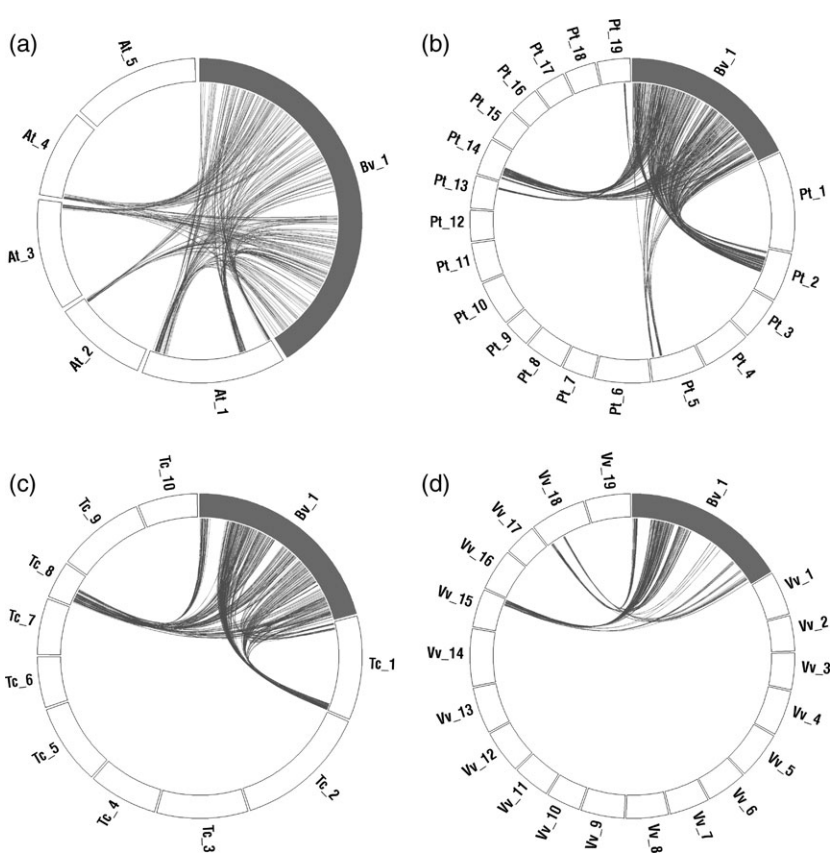

Figure 5. Syntenic regions of sugar beet chromosome 1 ('Bv_1') and the chromosomes of four rosid species.
(a) *Arabidopsis thaliana.*
(b) *Populus trichocarpa.*
(c) *Theobroma cacao.*
(d) *Vitis vinifera.* Connecting lines indicate corresponding genes based on sequence homology.

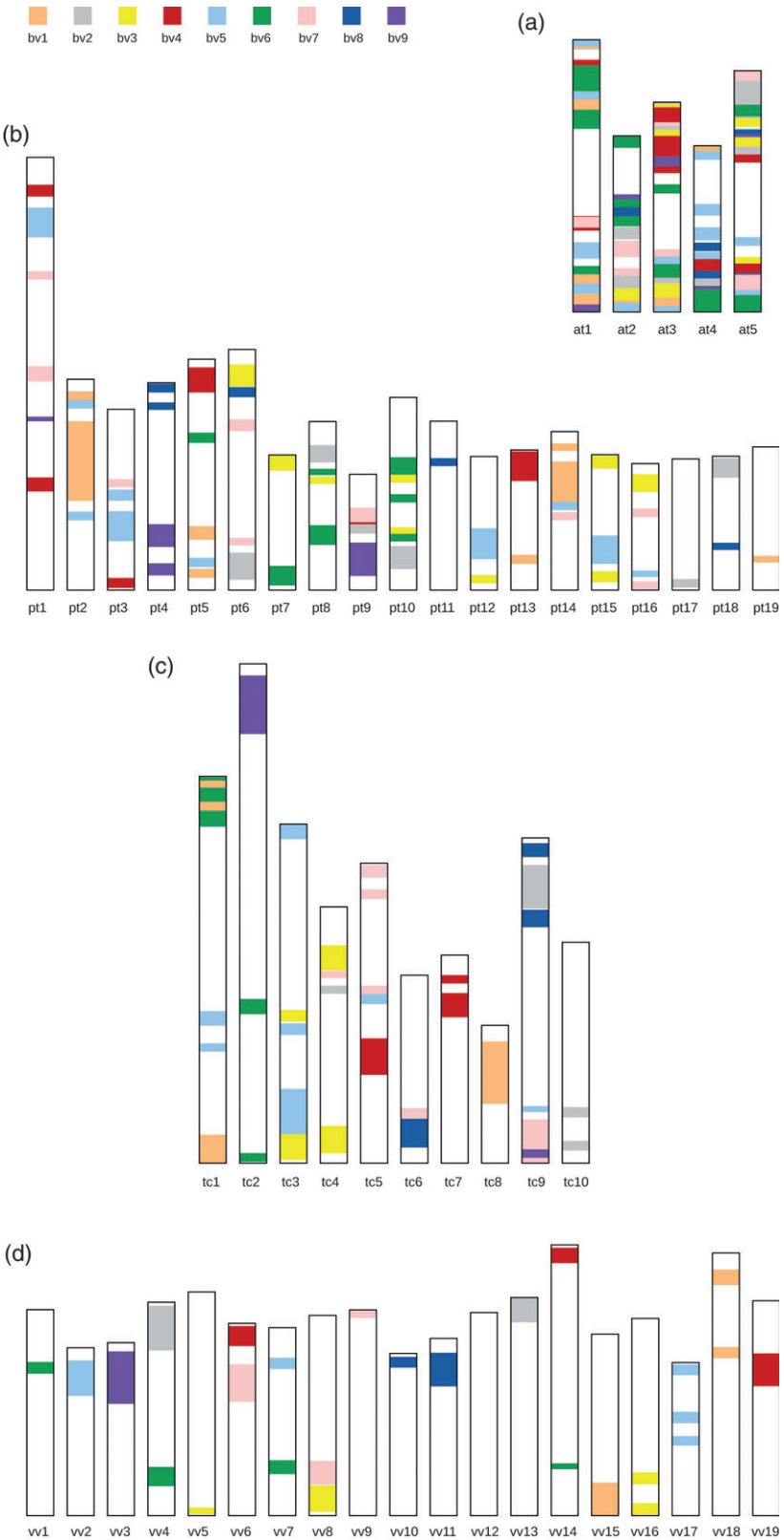

Figure 6. Chromosomes of the four target species and their syntenic regions with sugar beet. Syntenic regions were inferred from homology searches between probe source sequences (sugar beet) and coding sequences (target species). The colors for the nine sugar beet chromosomes are indicated at the top of the figure. Chromosome lengths and horizontal positions of syntenic regions are displayed to scale in accordance with genome annotations.
(a) *Arabidopsis thaliana.*
(b) *Populus trichocarpa.*
(c) *Theobroma cacao.*
(d) *Vitis vinifera.*

to infer the partitioning of sugar beet chromosomes. The distribution of the ancestral linkage groups within the genomes of poplar and cacao were more fragmented, but in many cases results were sufficiently clear when taking into account the position of the syntenic sugar beet region within the chromosomes of these genomes. *Arabidopsis* was not analyzed due to its extensive fragmentation of ancestral traces. All assignments of the seven ancestral linkage groups to sugar beet inferred from grapevine were also inferred consistently from assignments based on the genomes of poplar and cacao. Some additional assignments based on poplar and cacao existed, which we only considered if the two genomes were confirming each other. We found each of the ancestral linkage groups I, II, III, V in three different sugar beet chromosomes. The ancestral linkage group VII was found to be present within two, linkage group VI within four, and linkage group IV within six sugar beet chromosomes (Table 3). The presence of predominantly three locations of the different ancestral linkage groups within sugar beet suggested that also Caryophyllales, like rosids and asterids, are evolved from a hexaploid ancestor. If this situation is the case, the corresponding regions within sugar beet would be expected to be paralogous. We performed an all-against-all comparison at the protein level with the source sequences of genetically anchored sugar beet physical map probes. After setting cutoffs for sequence identity and match length (see Data S1) we counted the matches between different sugar beet chromosomes. The top-four most frequently occuring matching chromosome pairs (excluding chromosome 8, which has the least number of genetic markers and the smallest physical contig size and thus is likely to be a less reliable data source) were in line with the location of the linkage groups assigned by comparison with the genomes of grapevine, poplar and cacao (Table 3). Alternative scenarios would be an independent genome triplication of Caryophyllales or several independent chromosomal triplications. However, in the case of independent triplication events, one would not expect to find each of the three resulting copies corresponding to exactly one copy within the other species. For example, linkage group I (LGI) segments were found in sugar beet chromosomes 1, 3, and 5 (Table 3) and within *Vitis* LGI was found to be located in chromosomes 2, 15, and 16 (Jaillon *et al.*, 2007). For each of the three sugar beet chromosomes (bvchr) there is only one of the three *Vitis* chromosomes (vvchr) that show synteny: LGI in bvchr1 corresponds to LGI in vvchr15, bvchr3 corresponds to vvchr16, and bvchr5 corresponds to vvchr2 (see Circos plots in Figure S3), and none of the other two LGI regions of *Vitis* was found to be syntenic in each of these cases. The same result is true for other LGs. If the triplication was independent we would expect to see synteny scattered between all three copies of other species for each of the three sugar beet copies. The one-to-one relationship of syntenic blocks that reflects triplicated ancestral linkage groups is most parsimoniously explained by a common triplication event before speciation.

DISCUSSION

In the context of this work we presented a dense linkage map of the sugar beet genome and a gene-based physical map in BAC clones. Both types of maps are linked tightly by reference points that they have in common, either because anchor probes for physical mapping and SNP detection employed the same set of EST source sequences or because genetic marker development was based on BAC end

Table 3 Assignment of sugar beet chromosomes to ancestral linkage groups inferred from comparisons with the genomes of grapevine, poplar, and cacao

	Ancestral linkage groups assigned							
Sugar beet chromosomes	I	II	III	IV	V	VI	VII	Bvchr matched by tblastx
bvchr1	x				x			9, 6, 5, 3
bvchr2		x		x		x		7, 9, 3, 5
bvchr3	x		x	x		x		5, 9, 2, 1
bvchr4			x	x			x	5, 3, 9, 7
bvchr5	x		x	x		x		9, 3, 1, 2
bvchr6					x	x		(8,) 1, 9, 5, 2
bvchr7		x		x				2, 9, 3, 5
bvchr8		x					x	(6, 9, 2, 5)
bvchr9				x	x			1, 2, 5, 3
Total[a]	3	3	3	6	3	4	2	

The seven proposed ancestral linkage groups are indicated by roman numerals (the numbering follows the order of the linkage groups in Figure 4(c) in Argout *et al.*, 2011). Assignments are based on comparisons with paralogy information of grapevine, poplar, and cacao genomes. The last column lists the top-four matching chromosomes in an all-against-all sequence homology search (protein level) of sugar beet sequences anchored in the physical map.

[a]The last row contains the number of different sugar beet chromosomes to which the ancestral linkage groups were assigned.

sequences and the underlying BAC clones are present in the map. The integrated nature of the maps is particularly rewarding and will allow rapid progress from linkage results to genomic regions that contain candidate genes.

A comparative inspection of genetic and physical maps suggested uneven distribution of recombination events in the sugar beet genome. The linkage map contained markers based on both gene sequences and genomic sequences; bins of unresolved markers on the genetic map contained markers from both types of sources. While co-segregation of markers from coding sequences might be explained by the uneven distribution of genes, markers from BACs should represent a random sampling of the genome. We concluded, therefore, that the distribution of crossovers is non-random. We also noticed enhanced recombination at chromosome ends and that linkage maps were compressed centrally. Such map distortion may be a feature of any large plant genome and has been reported previously for soybean (Schmutz *et al.*, 2010).

The construction of the physical map described herein relies on hybridization of gene-derived probes. The contigs will cover gene containing intervals well while gene-poor stretches in the genome will be sparsely covered due to the uneven distribution of coding regions in genomes. If a gene complement of about 29 000 sugar beet genes (Dohm *et al.*, 2009) is assumed, we prepared probes from roughly 30% of all sugar beet genes. Little information on the spacing of sugar beet genes is presently available. Dohm *et al.* (2009) sequenced and annotated overlapping BAC clones and estimated a density of one gene per 11 kbp of genomic sequence. As the average insert size of the two BAC libraries used for map construction was 110 and 150 kbp, respectively, distances of up to 150 kbp between genes could be spanned. Given that about one-third of all genes was represented by oligomer probes, we assume a maximum average distance of 30–40 kbp between genes included in the physical map. However, gene densities might vary greatly.

Gene duplicates, pseudogenes, or gene fragments located on different chromosomes may lead to map assignments of BACs to wrong positions in the map by means of cross-hybridization. However, the short oligo sequences are very specific in hybridizing with their genomic location. Thus, only highly similar (if not identical) sequence segments in the genome would result in cross-hybridizations. The hybridization data were cleaned for probes with high-copy matches, and the contig construction relied on the presence of several links between clones that supported each other consistently. A cross-hybridizing sequence in the target clone will presumably be embedded into an entirely different genomic context and probes from the vicinity will drive the correct placement during contig construction. However, in the case of sparsely distributed probes or missing data, clones might be linked incorrectly. The location of a particular clone should be read with caution taking into account the number of probes connecting this clone with its neighbors.

The taxonomic position of sugar beet (order Caryophyllales) within the core eudicotyledonous plants makes it a relevant target for filling a gap in comparative genomics data. The Caryophyllales represent a sister taxon to both rosids and asterids. While genome sequences that represent the rosid or asterid lineage are available, or are in progress, no genome project for Caryophyllales other than the one for sugar beet has been announced so far. The gene-derived mapping probes ordered within the physical map provide access to a genome-wide gene set and its spatial arrangement within a member of the Caryophyllales. Although the gene set is not comprehensive, it could be used successfully to identify syntenic blocks with sequenced rosid species.

It has been suggested previously that an angiosperm ancestor had existed that contained a haploid set of seven chromosomes and had undergone hexaploidization of its genome (Jaillon *et al.*, 2007). Evidence for the hexaploid origin of different rosid and asterid genomes has been reported based on complete or partial genomic sequence data (Jaillon *et al.*, 2007; Argout *et al.*, 2011; Tang *et al.*, 2008b; Cenci *et al.*, 2010). We compared the genome-wide syntenic regions of sugar beet in grapevine, poplar, and cacao with the ancestral paralogy data described for these species (Jaillon *et al.*, 2007; Argout *et al.*, 2011). Assuming a hexaploid ancestor for sugar beet we expected to find the ancestral linkage groups three times (e.g. in three different chromosomes, if there was little re-arrangement like in grapevine) within the haploid sugar beet genome. We indeed found three paralogous regions in sugar beet for four of the proposed ancestral linkage groups (and two, four, and six paralogous regions for the remaining three ancestral linkage groups). As the three target genomes under consideration confirmed each other and, consistently, no more than two to four different ancestral linkage groups were assigned per sugar beet chromosome (despite of many different syntenic blocks of sugar beet within the target genomes), we concluded that sugar beet shares the hexaploid ancestor with these species. The scenario of an independent triplication event in Caryophyllales, in contrast, seems to be unlikely. In many cases, the individual copies of triplicated ancestral linkage groups in sugar beet were found to correspond to distinct copies in the other species. Thus, the most likely scenario is that hexaploidization of the proposed ancestral angiosperm genome has occurred prior to lineage separation of Caryophyllales, asterids, and rosids. For the same reasons, we further concluded that inter-chromosomal rearrangements have been relatively rare during the evolution of Caryophyllales, a finding that contrasted with findings for other angiosperm lineages such as Brassicales (Lagercrantz, 1998; Kejnovsky *et al.*, 2009); we

consider sugar beet to be a rewarding target for further comparative analyses.

In summary, this work provides well characterized resources for the study of sugar beet. The genome-wide comparative analysis based on the presented data shed light onto the evolutionary history of Caryophyllales, and a future genome sequencing project will benefit substantially from these resources. The maps provide a backbone for a sugar beet draft assembly based on whole genome shotgun sequencing, and will be an invaluable tool to order, orientate, verify, and connect assembled sequence scaffolds.

EXPERIMENTAL PROCEDURES

The experimental procedures used in this work are described in detail in the Data S1.

ACKNOWLEDGEMENTS

We are grateful to Stefanie Palczewski, Marion Klein, Anja Feldner, and Prisca Viehöver for excellent technical assistance, Andreas Menze, Steffen Hennig and Ralf Herwig for help with data, and Donald Buczek for software support. We thank Georg Koch and Thomas Schmidt for helpful discussions and André Minoche for comments on the manuscript. Thomas Illig and Henning Gohlke contributed to genotyping on the Sequenom platform. This work was supported by the BMBF grant 'Verbundprojekt GABI-Beet Physical map', FKZ 0313127A and 0313127D and by the 'Verbundprojekt GABI BeetSeq', FKZ 0315069A and 0315069B.

We dedicate this work to the memory of our dear colleague Dr Katharina Schneider who passed away September 5th 2007. She was deeply involved in the genetic mapping performed at the beginning of the project.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Genetic map of sugar beet for all nine chromosomes (Beetmap).

Figure S2. Increase of centi-Morgan (cM) units within the genetic map along with successive marker positions.

Figure S3. Syntenic regions of all sugar beet chromosomes and the chromosomes of four rosid species.

Data S1. Experimental procedures.

Data S2. Source sequences for probe design (112 sequences) derived from Roche-454 cDNA sequencing.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Argout, X., Salse, J., Aury, J.M. *et al.* (2011) The genome of *Theobroma cacao*. *Nat. Genet.* **43**, 101–108.

Arumuganathan, K. and Earle, E.D. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–211.

Blanc, G., Hokamp, K. and Wolfe, K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Res.* **13**, 137–144.

Cai, W.W., Reneker, J., Chow, C.W., Vaishnav, M. and Bradley, A. (1998) An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics*, **54**, 387–397.

Cenci, A., Combes, M.C. and Lashermes, P. (2010) Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. *Mol. Genet. Genomics*, **283**, 493–501.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA*, **83**, 7821–7825.

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

Dolezel, J., Kubaláková, M., Paux, E., Bartos, J. and Feuillet, C. (2007) Chromosome-based genomics in the cereals. *Chromosome Res.* **15**, 51–66.

Fischer, H.E. (1989) Origin of the 'Weisse Schlesische Rübe' (white Silesian beet) and resynthesis of sugar beet. *Euphytica*, **41**, 75–80.

Flavell, R.B., Bennett, M.D., Smith, J.B. and Smith, D.B. (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* **12**, 257–269.

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. *Mol. Genet. Genomics*, **274**, 316–323.

Hallden, C., Ahren, D., Hjerdin, A., Säll, T. and Nilsson, N.O. (1998) No conserved homoeologous regions found in the sugar beet genome. *J. Sugar Beet Res.* **35**, 1–13.

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

Hohmann, U., Jacobs, G., Telgmann, A. *et al.* (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.

Hunter, K.W., Ontiveros, S.D., Watson, M.L. *et al.* (1994) Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR), generation of a 5-cM, > 5 megabase contig on mouse chromosome 1. *Mamm. Genome*, **5**, 597–607.

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**, 157–167.

Jaillon, O., Aury, J.M., Noel, B. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Kejnovsky, E., Leitch, I.J. and Leitch, A.R. (2009) Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends Ecol. Evol.* **24**, 572–582.

Khorasani, M.Z., Hennig, S., Imre, G. *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. *Mech. Dev.* **121**, 903–913.

Krzywinski, M., Schein, J., Birol, I., Connors, J. and Gascoyne, R. (2009) Circos, an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.

Kusumi, K., Smith, J.S., Segre, J.A., Koos, D.S. and Lander, E.S. (1993) Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome*, **4**, 391–392.

Lagercrantz, U. (1998) Comparative mapping between Arabidopsis thaliana and Brassica nigra indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, **150**, 1217–1228.

Lange, C., Holtgräwe, D., Schulz, B., Weisshaar, B. and Himmelbauer, H. (2008) Construction and haracterization of a sugar beet (*Beta vulgaris*) fosmid library. *Genome*, **51**, 948–951.

Lange, C., Mittermayr, L., Dohm, J.C. *et al.* (2010) High-throughput identification of genetic markers using representational oligonucleotide microarray analysis. *Theor. Appl. Genet.* **121**, 549–565.

McGrath, J.M., Shaw, R.S., de los Reyes, B.G. and Weiland, J.J. (2004) Construction of a sugar beet BAC library from a hybrid with diverse traits. *Plant Mol. Biol. Rep.* **22**, 23–28.

Meyers, B.C., Scalabrin, S. and Morgante, M. (2004) Mapping and sequencing complex genomes, let's get physical! *Nat. Rev. Genet.* **5**, 578–588.

Mott, R., Grigoriev, A., Maier, E., Hoheisel, J. and Lehrach, H. (1993) Algorithms and software tools for ordering clone libraries, application to the

mapping of the genome of schizosaccharomyces pombe. *Nucleic Acids Res.* **21**, 1965–1974.

van Oeveren, J., de Ruiter, M., Jesse, T. *et al.* (2011) Sequence-based physical mapping of complex genomes by whole genome profiling. *Genome Res.* **21**, 618–625.

Pillen, K., Steinrücken, G., Wricke, G., Herrmann, R.G. and Jung, C. (1992) A linkage map of sugar beet (*Beta vulgaris* L.). *Theor. Appl. Genet.* **84**, 129–135.

Salse, J., Abrouk, M., Bolot, S. *et al.* (2009) Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. *Proc. Natl Acad. Sci. USA*, **106**, 14908–14913.

Schmutz, J., Cannon, S.B., Schlueter, J. *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Schneider, K., Kulosa, D., Soerensen, T.R. *et al.* (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.

Schondelmaier, J. and Jung, C. (1997) Chromosomal assignment of the nine linkage groups of sugar beet (*Beta vulgaris* L.) using primary trisomics. *Theor. Appl. Genet.* **95**, 590–596.

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

Tang, H., Bowers, J.E., Wang, X. *et al.* (2008a) Synteny and collinearity in plant genomes. *Science* **320**, 486–488.

Tang, H., Wang, X., Bowers, J.E. *et al.* (2008b) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954.

Tuskan, G.A., Difazio, S., Jansson, S. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.

Varshney, R.K., Glaszmann, J.C., Leung, H. and Ribaut, J.M. (2010) More genomic resources for less-studied crops. *Trends Biotechnol.* **28**, 452–460.