Theor Appl Genet (2003) 107:168–180
DOI 10.1007/s00122-003-1236-x

G. Li · M. Gao · B. Yang · C. F. Quiros

Gene for gene alignment between the Brassica and Arabidopsis genomes by direct transcriptome mapping

Received: 20 November 2002 / Accepted: 10 January 2003 / Published online: 21 March 2003

Abstract We report a global gene for gene alignment of the genomes of *Brassica oleracea* and *Arabidopsis thaliana* by construction of a transcriptome map based on *B. oleracea* cDNAs obtained from leaf tissue. cDNAs were synthesized from total RNA extracted from individual F2s of a mapping population resulting from crossing double-haploids of broccoli and cauliflower. The map consisted of 247 cDNA markers obtained by the SRAP technique. After sequencing 190 of the polymorphic cDNA bands, FASTA detected 169 sequences with similarity to genes reported in *Arabidopsis*. There was extensive colinearity between the two genomes for chromosomal segments rather than for whole chromosomes, often showing inversions and deletions/insertions. Large-scale duplications were observed in the *B. oleracea* genome, but were unevenly distributed, arguing against ancient triplication of the entire genome. The most duplicated segments corresponded to those found on *Arabidopsis* chromosomes 1 and 5, whereas chromosomes 2 and 4 were the least represented in *Brassica*. Clear differences in the similarity score value of related sequences allowed the identification of orthologs. Transcriptome mapping is an efficient approach that allows gene-for-gene alignment between a fully sequenced and a poorly characterized genome.

Keywords Comparative genomics · Transcript profiling · cDNA

Communicated by H.F. Linskens

G. Li · M. Gao · B. Yang · C. F. Quiros (✉)
Department of Vegetable Crops,
University of California,
Davis, CA 95616, USA
e-mail: cfquiros@ucdavis.edu
Tel.: (530)-752-1734
Fax: (530)-752-9659

G. Li
Department of Biochemistry,
University of Saskatchewan,
Saskatoon, SK S7N 5E5, Canada

Introduction

The availability of the *Arabidopsis* and rice genome sequences provides the opportunity to analyze the similarities and differences between and within crop plants at a global genomic level (*Arabidopsis* Genome Initiative 2000; Lan et al. 2000; Paterson et al. 2000; Chen et al. 2002). It is clear that we are far from sequencing the genomes of most major crops; however, the conservation of gene sequences and gene order among taxa during their evolution in spite of million of years of divergence can be exploited through comparative genomics. In the past two decades great progress has been made in this area of research for several major crops. This trend will continue, and will certainly play a major role in current and future research activities. For comparative genomics, several tools based on DNA hybridization are commonly used, such as RFLP (restriction fragment length polymorphism), EST (expressed sequence tag) and physical mapping combined with genetic mapping (Cavell et al. 1998; O'Neill and Bancroft 2000; Draye et al. 2001; Fulton et al. 2002; Parkin et al. 2002). The cultivated species of the Brassicaceae in particular have benefited from this activity mainly due to the availability of the *Arabidopsis* genome sequences. However, the sequence information from this species reveals large-scale genome duplications, not only in *Arabidopsis* but also in most species in the plant kingdom, posing a challenge to the researcher. Often these are ancestral duplications involving the entire genome, making it difficult to pinpoint candidate genes for important traits (*Arabidopsis* Genome Initiative 2000). For example, 60% of the *Arabidopsis* genome is duplicated, which complicates comparative genetic mapping due to cross-hybridization of duplicated regions. In order to alleviate this problem, the development of new tools to improve comparative mapping is essential for transferring information from a sequenced genome to a non-sequenced one.

As EST data rapidly accumulate in many crops, these sequences become a useful resource to study large-scale gene expression with microarrays (Schena et al.1995;

Lockhart et al. 1996; Baldwin et al. 1999) or serial analysis of gene expression (SAGE) (Matsumura et al. 1999), which is based on the detection of transcripts from different tissues or produced under different environmental conditions. However, these techniques can be applied only to a few individuals, otherwise they become impractical, expensive and imprecise, especially when dealing with duplicated genomes. Furthermore, they become prohibitively expensive for co-segregation analysis involving large populations such as those used for crop-breeding programs. A more sensible approach in this situation is to develop transcriptome maps based on direct mapping of transcript polymorphisms. Brugmans et al. (2002) demonstrated the feasibility of this approach using cDNA-AFLP polymorphisms in segregating populations of diploid potato and *Arabidopsis*. In this report, we developed a transcriptome map in *Brassica oleracea* by visualizing transcript polymorphism with cDNA-SRAP, a PCR-based method designed to detect coding sequence polymorphisms of greater simplicity than cDNA-AFLPs (Li and Quiros 2001). Sequencing of the markers generated by this approach allowed us to align gene-for-gene chromosomal segments of the *Arabidopsis* and the *B. oleracea* genomes. This method is another important tool of comparative genomics allowing alignment of genes of a well-characterized model species, such as *Arabidopsis*, with those of a crop plant, such as *B. oleracea*.

Materials and methods

Plant material

An F2 population was developed from a cross of double-haploid lines of cauliflower, 'An-Nan Early', and broccoli 'Early Big'. Eighty eight F2 plants and their parental lines were used to construct the transcriptional map. This mapping population was used to carry out genetic analysis of three major genes involved in aliphatic glucosinolate synthesis (Li et al. 2001).

RNA extraction

We extracted RNA with phenol-chloroform (Sambrook et al. 1989). Two grams of young leaves were ground in liquid nitrogen, then adding 5 ml of extraction buffer (100 mM of Tris-HCl, pH 8.0, 100 mM of NaCl, 20 mM of EDTA and 1% sodium N-lauroyl sarcosine) and 5 ml of phenol:chloroform (3:1, Tris balanced phenol, pH 8.0, from Life Science Technologies, Calif.). The tissue was homogenized with a polytron at 3,000 rpm for 1 min and then centrifuged at 2,700 rpm for 5 min. The supernatant was washed once with chloroform. RNA was precipitated by adding 1/10 3 M sodium acetate and 2 vol of ethanol, and washed with 70% ethanol. The pellet was dissolved in 2 ml of deionized water, and then we added 2 ml of 4 M LiCl. The tubes were kept on ice for 4 h followed by centrifugation at 10,000 rpm. The pellet was washed briefly and dissolved in de-ionized water. DNA was removed with RNA-free DNAse I (Amersham Pharmacia Biotech, Calif.) following the manufacturer's protocol. The DNAse I was removed with phenol:chloroform (3:1). The RNA concentration was determined with the aid of a spectrophotometer.

Table 1 List of primers used in the present study

Primer name	Primer sequence 5'–3'
ME2	TGAGTCCAAACCGGAGC
ME8	TGAGTCCTTTCCGGTGC
EM1	GACTGCGTACGAATTCAAT
EM2	GACTGCGTACGAATTCTGC
DC1	TAAACAATGGCTACTCAAG
OD3	CCAAAACCTAAAACCAGGA
OD8	CACAAGTCGCTGAGAAGG
OD10	AGGAGGGAAAGGTCTGGT
OD12	TTGAATATCCAGTGTAAGGTT
OD13	AACAGCGAAACGATCCAGA
OD15	GCGAGGATGCTACTGGTT
OD17	GTTAGTATCAAGGTTAGAGTT
OD22	TACACCAGCCAAGGATGC
OD24	GATGCTTCTCGTCCACAA
OD26	CTATCTCTCGGGACCAAAC
OD30	GCGATCACAGAAGGAAGGT
OD32	ACTGTGATGTCGTTACTGAT
OD34	CAATCAGGGCGTAGCAGT
SA4	TTCTTCTTCCTGGACACAAA
SA7	CGCAAGACCCACCACAA
SA8	GGATGAAGCGACAAGTC
SA9	GTTGAGAGTGTTGATTGGT
SA12	TTCTAGGTAATCCAACAACA
SA14	TTACCTTGGTCATACAACATT
SA17	ATAAGAATCAGCAGACGCAT
SA18	ACGAGTTGCGGAAGTGG
SA21	GAATGCAGGAGAACACGTT
GA2	TTGAACTGGCAGAAAGGGT
GA3	TCATCTCAAACCATCTACAC
GA5	GGAACCAAACACATGAAGA
GA11	CATTGTGGTGGTTATTGTCA
GA12	CACCACCATCATCATATCTT
GA13	GTACCTGCAAGTGCTTCA
GA18	GGCTTGAACGAGTGACTGA
GA19	TTAAGGGCATAAAACATGGAT
GA25	TACTCCAGCCCAAATACAC
GA27	GAACGAAGCAAAGGATGAGA
GA28	GGTGATACACTTCAGATG
GA30	CTCTCCACCGCACATATC
GA33	GTTATGGGAAATTAGGTGAG
GA34	CCAAATGGAACAAAATGATG
GA38	CCTCTTCTTTAGCCGTTGA
GA45	AGTGGTATTTTTGCAGTTCTA
PM8	CTGGTGAATGCCGCTCT
PM18	AAGCGATCAAAGCGGGTG

cDNA synthesis

M-MLV reverse transcriptase (Invitrogen, Calif.) was used to synthesize single-strand cDNA following the manufacturer's protocol, except that only 1/10 of the specified Oligo $(dT)_{15}$ primer concentration was used; 50 μg of total RNA was added to make 100 μl of the reaction mixture and incubated at 37 °C for 2 h. After incubation, 400 μl of water was added to bring the total volume to 0.5 ml. Then, we added 1/10 vol of 3 M sodium acetate and 0.7 vol of iso-propyl alcohol. The tubes were placed on ice for 3 min and then centrifuged at 14,000 rpm for 3 min. The cDNA pellet was washed with 70% ethanol once and then dissolved in 100 μl of de-ionized water.

Fingerprinting of cDNA with the sequence related amplified polymorphism (SRAP) protocol

We applied the SRAP protocol to fingerprint 88 cDNA samples using 47 primer combinations following the procedure of Li and Quiros (2001) (Table 1). In order to detect and isolate polymorphic

bands for sequencing, we used three steps. (1) For detecting polymorphism we ran all the 88 F2 and two parental samples in a LI-COR sequencer IR^2, model 4,200, after amplifying the cDNAs with two primers, one of which was labeled with IRDye 800 or IRDye 700, (LI-COR, Lincoln, Neb.). (2) For collecting DNA from the polymorphic bands we re-amplified only the DNA of the two parental cDNAs with the same primer combination used for the LI-COR assay, except that one of the primers was labeled with (γ^{33} P)-ATP. The amplicons were separated by denaturing acrylamide-gel electrophoresis and detected by autoradiography (Li and Quiros 2001). All bands showing polymorphism between these two parental lines were cut from the dried gel. The DNA was eluted from the gel with buffer (0.5 M of ammonium acetate, 10 mM of magnesium acetate, 1 mM of EDTA, pH 8.0, 0.1% SDS) by shaking at 300 rpm at 37 °C overnight, and precipitated with ethanol. (3) In order to align the polymorphic bands with the isolated bands from both gel systems the DNA from the isolated bands was re-amplified for 30 cycles as follows: 94 °C for 50 s, 55 °C for 50 s and 72 °C for 40 s. The PCR products from these bands were run side by side in the LI-COR system along with the amplified products of the two parental lines used previously for detecting polymorphism. This approach allowed us to match the corresponding bands whose sequences were used for comparative analysis to the *Arabidopsis* sequences. The specific marker for the *BoGSL-ELONG* gene (Genbank AF399834) was obtained by amplification of the cDNA samples with specific primers for this gene, PM8 and PM18 (Table 1) as follows: 94 °C for 50 s, 55 °C for 50 s and 72 °C for 60 s for 35 cycles. The amplicons were fractionated by agarose-gel eletrophoresis, which allowed us to detect polymorphism for this sequence (Li and Quiros, unpublished).

Phenotypic analysis

The phenotypes for genes *BoGSL-PRO* and *BoGSL-ELONG*, (the presence of 3-carbon and 4-carbon aliphatic glucosinolates, respectively) in *B. oleracea* were determined in the F2 mapping population as reported by Li et al. (2001).

Sequence analysis and map construction

Sequences were produced by the LI-COR IR^2 sequencing using the manufacturer's protocol. The sequences were analyzed with the FASTA searching program (Pearson and Lipman 1988) allowing to match the *Arabidopsis* homologs to the *Brassica* cDNA markers, including their map positions and gene products when known. E-values of less than 10^{-5}, showing over 70% identity in more than 100 nt, were considered as high confidence matching between two sequences. The transcriptome map in *B. oleracea* was constructed with Mapmaker version Mac 2.0 with a LOD value of 3.0.

Results

Using 48 primer combinations, we detected 281 polymorphic bands as markers in 88 cDNA pools from the same number of plants. Each primer combination gave 1–15 polymorphic bands with an average of 6.0 bands per primer set. Most (78.9%) of the polymorphic bands showed dominant expression, the rest of the markers were co-dominant (Table 2). Since the size of some bands was too small to be informative, we sequenced only 190 of the polymorphic bands, most of which had a size larger than 100 bp. The FASTA search allowed us to identify 169 sequences having similarity to the genes reported in *Arabidopsis*. Of these, 113 had high confidence matches, whereas 56 matched at a lower confidence level displaying E values higher than 10^{-5} due to their smaller size. There were 132 unique sequences, each of which represented a single expressed gene, if multiple amplifications of the same gene are not counted. Sequence analysis allowed unambiguous identification of multiple amplifications of the same gene. This event is illustrated by the fact that nine dominant markers, T9, T61, T63, T64, T66, T87, T88, T137 and T152, amplified by five different primer sets hit the same *Arabidopsis* gene, namely glycine SRC2-like (Genbank NM-100778) (Table 2, Fig. 1). In another case four primer sets amplified three co-dominant markers, T138, T146, T156, and one dominant marker, T120b. All these markers corresponded to a gene coding for a glycine-rich protein (Genbank NM-120087) in *Arabidopsis*. Similar cases were observed for genes similar to glutathione transferase, the putative ribosomal protein L17, the DAG-like protein and for other several unknown proteins in *Arabidopsis*.

The band intensity observed roughly represents the abundance of the template cDNAs, and presumably that of their corresponding RNAs in the pools. This interpretation is based on the fact that multiple markers displaying the same gene matches, but amplified with different primer combinations, showed the same band intensity (data not shown).

When we checked the sequences that appeared to be codominant in the gels, we found indeed that most had nearly identical sequences as expected for alleles at the same locus, except for insertions or deletions, which might correspond to splicing-site changes. However, there were two markers, T22, and T131, which in spite of appearing codominant in segregation pattern and mapping to the same region, displayed two totally different sequences indicating that they were not allelic.

Two other interesting results are worthwhile mentioning. One is that only 40% of the 132 unique *Arabidopsis* genes identified by the *Brassica* cDNA marker sequences had available ESTs or were supported by cDNA sequences in *Arabidopsis*. This finding indicates that SRAP might detect some genes with low levels of expression or detect gene expression more evenly than ESTs from cDNA libraries. Another surprise was that 3% of marker sequences displaying strong band intensity did not match any genes in *Arabidopsis*. These transcripts might originate from non-protein coding RNAs (MacIntosh et al. 2001) or genes that have been lost in *Arabidopsis*, but further studies are needed to pinpoint their origin and nature.

After assembling the 281 cDNA markers and the phenotypic marker, [presence of the 3-carbon side chain glucosinolates (*Bo GSL-PRO* gene) (Li et al. 2001)] with Mapmaker, we produced a transcriptome map consisting of 247 markers. This map also included two cDNA markers for two members of the isopropyl malate synthase-like gene (IPMS), which presumably determine carbon side-chain length in aliphatic glucosinolates. (Li et al. 2001; Li and Quiros 2002). Perfect co-segregation was observed for the presence of 3-carbon and the 4-carbon

Table 2 List of sequenced cDNA markers from *B. oleracea* including their properties, locations and their physical correspondence to the *Arabidopsis* genome

Marker code	Primer	Band intensity[a]	Type[b]	Accession #	E-value	EST or cDNA	Size (bp)[c]	*A. thaliana* chrom.	*A. thaliana* loc. (MB)	Link. grp	Gene product
T7	ME2+OD3	VS	D	NM-104927	8.00E-07	Yes	280	1	22.3		Expressed protein
T8	ME2+OD3	VS	D	NM-105621	4.10E-20	Yes	214	1	25.7	5	Expressed protein
T9	ME2+OD3	S	D	NM-100778	3.40E-22	Yes	326	1	2.9	4	Similar to glycine SRC2
T11	ME2+OD8	VS	D	NM-126366	1.80	Yes	110	2	1	1	Putative calmodulin
T13	ME2+OD8	VS	D	NM-100765	3.00E-44	Yes	272	1	2.8	1	Zinc finger protein ATZF1, putative
T14	ME2+OD8	VS	CO-D	NM-103993	7.70E-16	Yes	290	1	22	9	Expressed protein
T16	ME2+OD8	VS	CO-D	NM-101532	6.00E-12	No	353	1	5.7	1	Unknown protein
T18	ME2+OD15	S	D	NM-129304	1.00E-06	No	100	2	15.7	4	Unknown protein
T20	ME2+OD15	VS	D	NM-129304	1.00E-06	No	150	2	15.7	4	Unknown protein
T21	ME2+OD15	S	D	NM-129304	1.00E-06	No	250	2	15.7	4	Unknown protein
T22*	ME2+OD15	S	CO-D	NM-127156	5.00E-09	Yes	301	2	6.9	5	Similar to cold acclimation protein WCOR413
T24	ME2+OD17	S	D	NM-115590	7.00E-81	No	126	3	21.2	2	Helicase-like protein
T27	ME2+OD17	S	D	NM-100013	0.2	No	185	1	0.1	1	Pathogenesis related protein, putative
T28	ME2+OD17	W	D	NM-129278	5.00E-19	Yes	403	2	15.7	2	Putative RNA-binding protein
T29	ME2+OD26	VS	D	NM-112946	0.002	No	171	3	7.1	1	DNA polymerase, putative
T30	ME2+OD26	S	D	NM-111311	5.00E-07	Yes	138	3	1.2	9	Ribosomal protein L17, putative
T32	ME2+OD26	S	D	NM-129704	3.30E-06	Yes	180	2	17.2	2	Calmodulin-like protein
T33	ME2+OD26	VW	D	NM-129705	0.014	No	194	2	17.1	4	Unknown, protein
T34	ME2+OD26	VW	D	NM-129705	0.014	No	194	2	17.1	4	Unknown, protein
T35	ME2+OD26	VW	D	NM-129705	0.014	No	194	2	17.1	4	Unknown, protein
T36	ME2+OD32	S	D	NM-123083	0.58	No	139	5	14.5	4	Putative protein
T40	ME2+OD32	S	D	NM-123083	0.58	No	139	5	14.5	4	Putative protein
T43	ME2+OD32	W	D	NM-116659	5.90E-21	Yes	329	4	2	3	Putative membrane trafficking factor
T44	ME2+OD34	S	D	NM-116414	1.9	Yes	246	4	0.8	3	Putative potassium channel
T45	ME2+OD34	S	D	NM-116414	1.9	Yes	246	4	0.8	3	Putative potassium channel
T46	ME2+OD34	S	D	NM-116414	1.9	Yes	246	4	0.8	3	Putative potassium channel
T50	ME2+SA4	S	D	NM-123460	1.20E-19	Yes	180	5	16.1	2	50s ribosomal protein L47
T51	ME2+SA4	W	D	NM-116511	4.70E-08	Yes	140	4	1.2	1	Putative photosystem I reaction center subunit
T52	ME2+SA4	S	D	NM-102509	7.70E-18	Yes	289	1	9.5	3	Adenine phosphoribosyltransferase 1, APRT
T53	ME2+SA4	S	CO-D	NM-104981	1.50E-19	No	237	1	22.9	3	Transcription factor DREB1A, putative
T55	ME2+SA7	S	CO-D	NM-120596	2.90E-18	No	420	5	1.5	7	Similar to unknown protein.
T57	ME2+SA7	W	D	NM-103082	1.50E-05	Yes	149	1	12.2	4	Expressed protein
T58	ME2+SA7	S	CO-D	NM-120596	2.90E-18	No	420	5	1.5	7	Similar to unknown protein.
T59	ME2+SA7	W	D	NM-120596	2.90E-18	No	488	5	1.5	7	Similar to unknown protein.
T61	ME2+SA9	S	D	NM-100778	0.00016	Yes	208	1	2.8	4	Similar to glycine SRC2
T63	ME2+SA9	S	D	NM-100778	0.00016	Yes	208	1	2.8	4	Similar to glycine SRC2
T64	ME2+SA9	S	D	NM-100778	0.00016	Yes	208	1	2.8	4	Similar to glycine SRC2
T65	ME2+SA9	S	D	AC079288	1.20E-11	No	170	1	10.4	6	Similar to putative selenium binding protein
T66	ME2+SA9	W	D	NM-100778	0.00016	Yes	208	1	2.8	4	Similar to glycine SRC2
T68	ME2+SA9	W	D	NM-106107	5.40E-45	Yes	374	1	27.6	6	Geranylgeranyl reductase
T71	ME2+SA12	W	D	NM-100397	0.023	Yes	79	1	1.5	1	Putative chloroplast 50S ribosomal protein, L6
T72	ME2+SA12	VS	D	NM-129704	0.013	Yes	151	2	17.1	2	Calmodulin-like protein
T75	ME2+SA12	S	D	NM-101504	0.65	No	276	1	5.5	1	Putative Na/H antiporter
T77	ME2+SA14	S	D	NM-116358	1.7	No	189	4	0.5	4	Putative myb-related DNA binding protein
T79	ME2+SA14	S	D	NM-124583	0.0024	Yes	206	1	20.9	3	Ariginine/semine-rich splicing factor Rsp41 homolog
T80	ME2+SA14	W	D	NM-101993	5.40E-21	Yes	186	1	7.5	3	Expressed protein
T82	ME2+SA14	S	D	NM-128407	1.10E-39	No	284	2	12.2	3	Putative beta-galactosidase

Table 2 (continued)

Marker code	Primer	Band intensity[a]	Type[b]	Accession #	E-value	EST or cDNA	Size (bp)[c]	*A. thaliana* chrom.	*A. thaliana* loc. (MB)	Link. grp	Gene product
T85	ME2+SA17	S	D	NM-114810	5.70E-10	Yes	200	3	18.4	4	RNA-directed RNA polymerase
T87	ME2+SA17	S	D	NM-100778	3.10E-11	Yes	174	1	2.8	4	Similar to glycine SRC2
T88	ME2+SA17	S	D	NM-100778	9.70E-12	Yes	271	1	2.8	4	Similar to glycine SRC2
T89	ME2+SA17	S	D	NM-100943	1.60E-34	Yes	298	1	3.5	8	ATP citrate-lyase, putative
T95	ME2+SA21	W	D	NM-115227	0.12	No	112	3	19.8	9	Putative protein
T101	ME2+GA3	W	D	NM-124089	4.50E+08	Yes	184	5	18.9	3	VAMP(vesicle associated membrane protein) associate protein like
T102	ME2+GA3	W	D	NM-129704	0.0051	Yes	173	2	17.2	2	Calmodulin-like protein
T103	ME2+GA3	S	D	NM-121805	5.10E+10	No	208	1	5.9	8	Similar to unknown protein.
T112	ME2+GA5	S	D	NM-118121	3.60E+09	No	261	4	9.8	6	DAG like protein
T114	ME2+GA5	W	D	NM-121352	0.0025	Yes	132	5	4.4	1	Adenosine nucleotide translocator
T115	ME2+GA5	W	D	NM-118121	3.60E+09	No	261	4	9.8	6	DAG like protein
T116	ME2+GA5	S	D	NM-118121	1.00E-14	No	290	4	9.8	6	DAG like protein
T120	ME2+GA11	S	D	NM-103993	0.65	Yes	143	1	18.5	1	Expressed protein
T121	ME2+GA11	S	D	NM-111311	1.60E-16	Yes	156	3	1.2	9	Ribosomal protein L17, putative
T122	ME2+GA11	W	D	NM-117130	1.3	Yes	219	4	5.5	1	Putative protein
T123	ME2+GA12	VS	D	NM-122380	0.92	No	86	5	8.4	1	Unknown protein
T125	ME2+GA12	VS	D	NM-124463	0.49	Yes	173	5	20.4	3	Pyrurvate dehydrogenase E1 component beta subunit mitochondrial
T128	ME2+GA12	W	CO-D	NM-123865	0.45	No	157	5	17.8		Putative protein
T131	ME2+GA13	S	CO-D	NM-103495	2.90E-21	Yes	394	1	15.9	1	Putative transcription factor
T133	ME2+GA18	S	D	NM-100778	1.2	Yes	186	1	2.8	4	Similar to glycine SRC2
T134	ME2+GA18	S	CO-D	AC009894	3.30E-08	No	142	1	20.5	3	Elongation factor EF-2
T137	ME2+GA18	S	D	NM-100778	1.2	Yes	186	1	2.8	4	Similar to glycine SRC2
T138	ME2+GA18	S	CO-D	NM-120087	1.10E-06	Yes	246	4	17.2	7	Glycine rich protein
T139	ME2+GA19	S	CO-D	NM-100765	3.00E-26	Yes	220	1	2.8	1	Zinc finger protein ATZF1, putative
T143	ME2+GA25	S	D	NM-111124	6.00E+13	Yes	349	3	0.5	9	Putative 40S ribosomal protein.
T146	ME2+GA28	S	CO-D	NM-120087	4.20E-18	Yes	240	4	17.7	7	Glycine rich protein
T147	ME2+GA28	S	D	NM-130246	3.30E-56	Yes	363	2	19.2	2	Putative zinc transporter
T149	ME2+GA30	S	CO-D	NM-105873	3.60E-13	Yes	422	1	26.8	7	Cytosolic factor putative
T151	ME2+GA30	S	D	NM-106555	4.80E-35	Yes	175	1	29.3	5	Photosystem II polypeptide, putative
T152	ME2+GA30	S	D	NM-100778	6.60E-37	Yes	388	1	2.8	4	Similar to glycine SRC2
T154	ME2+GA33	W	D	NM-121258	0.00039	Yes	227	5	3.9	6	Dehydropyrimidinase
T155	ME2+GA33	W	D	NM-117632	1.2	No	160	4	7	1	Hypothetical protein
T156	ME2+GA33	W	CO-D	NM-120087	2.70E-05	Yes	161	4	17.2	7	Glycine rich protein
T157	ME2+GA33	W	D	NM-100314	0.98	No	192	1	1.1		Unknown protein
T158	ME2+GA33	VW	D	NM-123221	0.94	Yes	184	5	15.6	1	Nitrilase 4
T159	ME2+GA33	VW	D	NM-121916	2.70E-20	No	208	5	6.4	3	Dermal glycoprotein like
T160	ME2+GA33	S	D	NM-106069	1.10E-27	Yes	231	1	27.4	8	Putative flavonol sulfotransferase
T161	ME2+GA33	S	D	NM-115539	1.50E-13	Yes	254	3	21	2	Calmodulin-3
T162	ME2+GA33	W	D	NM-104330	1.10E-17	Yes	395	1	19.8		Expressed protein
T163	ME2+GA33	W	D	AL162751	0.77	No	315	5	0.8	2	EIN2
T164	ME2+GA34	S	D	NM-129353	8.40E-34	Yes	297	2	15.8	2	Unknown protein
T165	ME2+GA34	S	D	NM-127596	6.80E-29	No	293	2	8.7		Unknown protein
T170	ME2+GA38	W	D	NM-125137	3.00E-88	Yes	457	5	23	1	TCH4 protein
T172	ME2+GA27	W	D	NM-103909	5.50E-06	No	220	1	18.2	5	Chloroplast FtsH protease
T173	ME2+GA27	S	D	NM-125366	5.00E-16	No	185	5	23.7	8	Protein serine/threonine kinase like protein

Table 2 (continued)

Marker code	Primer	Band intensity[a]	Type[b]	Accession #	E-value	EST or cDNA	Size (bp)[c]	*A. thaliana* chrom.	*A. thaliana* loc. (MB)	Link. grp	Gene product
T174	ME2+GA27	S	D	NM-123742	8.30E-06	Yes	402	5	17.3	2	Expressed protein
T175	ME2+GA45	W	D	NM-125779	8.50E-39	No	203	5	25.8	3	Glucosidase II alpha subunit
T176	ME2+GA45	S	D	NM-123387	7.10E-21	No	225	5	15.8	2	Unknown protein
T178	ME8+SA7	VS	D	NM-127531	0.00011	Yes	160	5	8.5	8	Putative ribosomal protein L28
T179	ME8+SA7	S	D	NM-124984	1.1	No	115	5	22.4	1	Unknown protein
T182	ME8+SA7	S	D	NM-127531	0.00011	Yes	160	5	8.5	8	Putative ribosomal protein L28
T187	ME8+SA7	W	D	NM-114260	1.2	No	269	3	15.7	1	Putative protein
T188	ME8+SA8	W	D	NM-102749	0.089	No	360	1	10.6	8	9-cis-epoxycarotenoid, putative
T189	ME8+SA8	W	D	NM-121206	7.30E-06	Yes	167	5	3.8	3	Putative protein
T190	ME8+SA8	W	D	NM-102871	2.70E-24	Yes	191	1	11.2	1	Photosystem I subunit III precursor, putative
T192	ME8+SA8	S	D	NM-102749	0.089	No	360	1	10.6	8	9-cis-epoxycarotenoid dioxygenase, putative
T194	ME8+SA18	S	D	NM-125029	0.15	No	170	5	22.5		Protein kinase like protein
T195	ME8+SA18	S	D	NM-122135	1.20E-07	Yes	186	5	7.3	1	Nitrilase 4
T196	ME8+SA18	VS	D	NM-127189	1.00E-19	Yes	281	2	7	3	Expressed protein
T197	ME8+SA18	S	D	NM-102732	2.10E-09	Yes	216	1	10.4	2	Photosystem II type I chlorophyla/b binding protein putative
T199	ME8+GA2	W	D	NM-122080	1.10E-05	Yes	252	5	7	8	Unknown protein
T200	ME8+GA2	S	D	NM-120366	1.00E-18	No	219	5	0.6	1	Similar to thyroid receptor interact in protein
T201	ME8+GA2	W	D	NM-100036	7.10E-11	Yes	298	1	0.2	1	Protein serine/threonine kinase, putative
T202	ME8+GA2	W	D	NM-106555	7.90E-23	Yes	267	1	29.3	5	Photosystem II polypeptide, putative
T203	ME8+GA2	S	D	NM-121983	2.80E-26	Yes	345	5	6.6	3	Tubulin alph-5 chain
T206	DC1+OD10	VS	D	NM-127265	1.10	Yes	149	2	7.3	4	Expressed protein
T210	DC1+OD10	S	CO-D	NM-125683	6.90E-14	Yes	249	5	24.9	8	Permease-like protein
T211	DC1+OD10	S	D	NM-128994	1.00E-54	Yes	374	2	14.4	2	Chlorophyl a/b-binding protein
T212	DC1+OD10	S	D	NM-106485	3.40E-39	Yes	360	1	29.1	1	Glutathione transferase, putative
T213	DC1+OD15	S	D	NM-125648	3.20E-11	Yes	187	5	24.8	1	Strong similar to ubiquitin conjugation enzyme
T214	DC1+OD15	W	D	NM-112356	1.80	Yes	264	3	5	5	Phosphoenolpyruvate carboxylase (PPC)
T216	DC1+OD26	S	D	NM-106485	3.60E-46	Yes	406	1	29	1	Glutathione transferase, putative
T218	DC1+OD30	S	D	NM-106485	3.60E-46	Yes	406	1	29	1	Glutathione transferase, putative
T219	DC1+OD34	S	D	NM-116415	0.14	No	190	4	0.8	7	S-adenosylmethionine synthase 2
T224	EMI+OD15	W	D	NM-113459	6.80E-11	No	120	3	9.3	6	ATPase II, putative
T225	EMI+OD15	W	D	NM-125477	0.0025	No	138	5	24.2	6	Unknown protein
T227	EMI+OD15	W	CO-D	NM-125477	0.47	No	180	5	24.2	6	Unknown protein
T228	EMI+OD15	S	D	NM-128100	5.10E-05	Yes	233	2	10.7	9	Expressed protein
T229	EMI+OD15	W	D	NM-100919	0.95	No	135	1	3.5	9	Unknown protein
T230	EMI+OD15	W	D	NM-106608	6.10E-19	Yes	282	1	29.6	1	Unknown protein
T231	EMI+OD15	W	D	NM-122501	1.20E-14	Yes	258	5	9	8	Myrosinase precursor
T232	EMI+OD15	W	D	NM-115631	2.50E-14	Yes	351	3	21.4	2	Protein kinase – like protein
T235	EMI+OD17	S	CO-D	NM-119373	1.10E-07	No	132	4	14.5	7	Putative protein
T236	EMI+OD17	W	D	AC073433	1.70E-04	No	169	1	15.2	5	Pseudogene
T237	EMI+OD22	S	CO-D	NM-127641	0.92	Yes	299	2	8.9	4	Expressed protein
T238	EM2+OD12	S	CO-D	NM-112317	1.7	No	145	3	4.9	9	Hypothetical protein
T239	EM2+OD12	W	D	NM-115463	2.50E-11	Yes	177	3	20.8	2	Putative protein kinase
T240	EM2+OD12	S	D	NM-101857	1.20E-09	Yes	187	1	6.9		Similar to ferredoxin-NADP+ reductase
T241	EM2+OD12	S	D	NM-112470	6.10E-38	No	344	3	5.4	4	Myosin heavy chain-like protein.
T243	EM2+OD13	S	D	NM-112170	1.90E-11	No	143	3	4.3	8	Hypothetical protein
T245	EM2+OD15	W	D	NM-112422	0.21	Yes	249	3	5.2	2	Early auxin-induced protein. IAA19

Table 2 (continued)

Marker code	Primer	Band intensity[a]	Type[b]	Accession #	E-value	EST or cDNA	Size (bp)[c]	*A. thaliana* chrom.	*A. thaliana* loc. (MB)	Link. grp	Gene product
T22a*	ME2+OD15	S	D	NM-101964	9.00E-08	Yes	361	1	7.4		O-methyltransferase, putative
T22b	ME2+OD15	W	D	NM-115419	1.50E-52	No	366	3	20.5		Delta-1-pyrroline-5-carboxylate synthetase
T22e	ME2+OD15	S	D	NM-129304	5.00E-09	No	284	2	15.7	2	Unknown protein
T22f	ME2+OD15	S	D	NM-129304	1.00E-06	No	320	2	15.7	4	Unknown protein
T28b	ME2+OD17	W	D	NM-101131	9.90E-08	No	160	1	4.3	2	Transcriptional activator CBF1, putative
T28c	ME2+OD17	W	D	NM-129304	7.10E-15	No	254	2	15.7		Unknown protein
T44b	ME2+OD34	W	D	NM-111995	1.70E-05	Yes	162	3	3.7	8	Putative 2-cys peroxinedoxin BAS1 precursor
T44c	ME2+OD34	W	D	NM-100398	1.20E-06	Yes	129	1	1.5	4	Putative ligand-gated ion channel protein
T65b	ME2+SA9	S	D	NM-112264	5.80E-08	Yes	106	3	4.7	2	sm protein putative
T72b	ME2+SA12	W	D	NM-115453	5.80E-14	No	100	3	20.7		Receptor kinase - like protein
T76b	ME2+SA12	W	D	NM-128262	7.50E-07	Yes	265	2	11.5	7	Argonaute (AGO1)-like protein.
T83b	ME2+SA17	W	D	NM-103435	9.00E-07	No	108	1	15.4	1	Niemann-Pick C disease protein-like protein
T83c	ME2+SA17	W	D	NM-101131	1.80E-37	No	262	1	4.3		Transcriptional activator CBF1, putative
T106b	ME2+GA3	W	D	NM-128289	1.80E-33	No	463	2	18.8		Nam(no apical meristem) like protein.
T105b	ME2+GA3	W	D	NM-121531	4.00E-33	No	303	5	4.9	8	Putative protein
T108b	ME2+GA5	W	D	NM-102854	3.40E-12	No	120	1	11.1		Putative protein kinase C inhibitor
T116c	ME2+GA5	W	D	NM-102760	8.00E-09	Yes	311	1	10.6	4	Expressed protein
T120b	ME2+GA11	S	D	NM-120087	2.40E-06	Yes	139	4	17.2	7	Glycine rich protein
T122b	ME2+GA11	W	D	NM-111311	7.50E-14	Yes	268	3	2.5	9	Ribosomal protein L17,putative
T136b	ME2+GA18	S	D	NM-120884	7.30E-16	No	313	5	24.1		Replication factor A - like protein
T143b	ME2+GA25	W	D	NM-125459	3.80E-33	Yes	425	5	17.9		mip C protein. like (aquaporin)
T177b	ME2+GA45	W	D	NM-103861	9.50E-24	Yes	273	1	17.4	5	Expressed protein
T200b	ME8+GA2	S	D	NM-123772	2.80E-13	No	184	5	3		Putative protein
T213b	DC1+OD15	W	D	NM-121018	7.70E-08	Yes	215	5	25.9		ACTIN2/7
T214b	DC1+OD15	W	D	NM-105651	2.20E-31	Yes	229	1	11.5	5	Putative alpha-amylase
T216b	DC1+OD24	S	D	NM-118561	1.60E-23	No	232	4	15.7	6	Hsp 70 like protein.
T236b	EMI+OD17	S	D	NM-119676	2.00E-27	Yes	212	4	18.9		Plasma membrane intrinsic protein
T236c	EMI+OD17	W	D	NM-130171	3.90E-04	Yes	190	2	5.8		Expressed protein
T237b	EMI+OD22	W	D	NM-112570	6.30E-03	No	342	3	20.6		Calmodulin-binding protein, putative
T237c	EMI+OD22	W	D	NM-124520	2.90E-08	Yes	248	5	2.5	8	Arginine-aspartate-rich RNA binding protein-like
T237d	EMI+OD22	W	D	NM-100668	1.60E-24	Yes	263	1	1.5	4	Elongation factor 1 alpha
ELONG	PM8-PM18	S	CO-D	NM-122208	1.80E-40	No	400	5	7.6	1	2-isopropylmalate synthase-like

[a] W= weak band, VW= very weak, S= strong, VS= very strong
[b] D=dominant, CO-D, co-dominant
[c] for co-dominant markers the size of a single allele is shown
*T22 and T22a, co-dominant marker

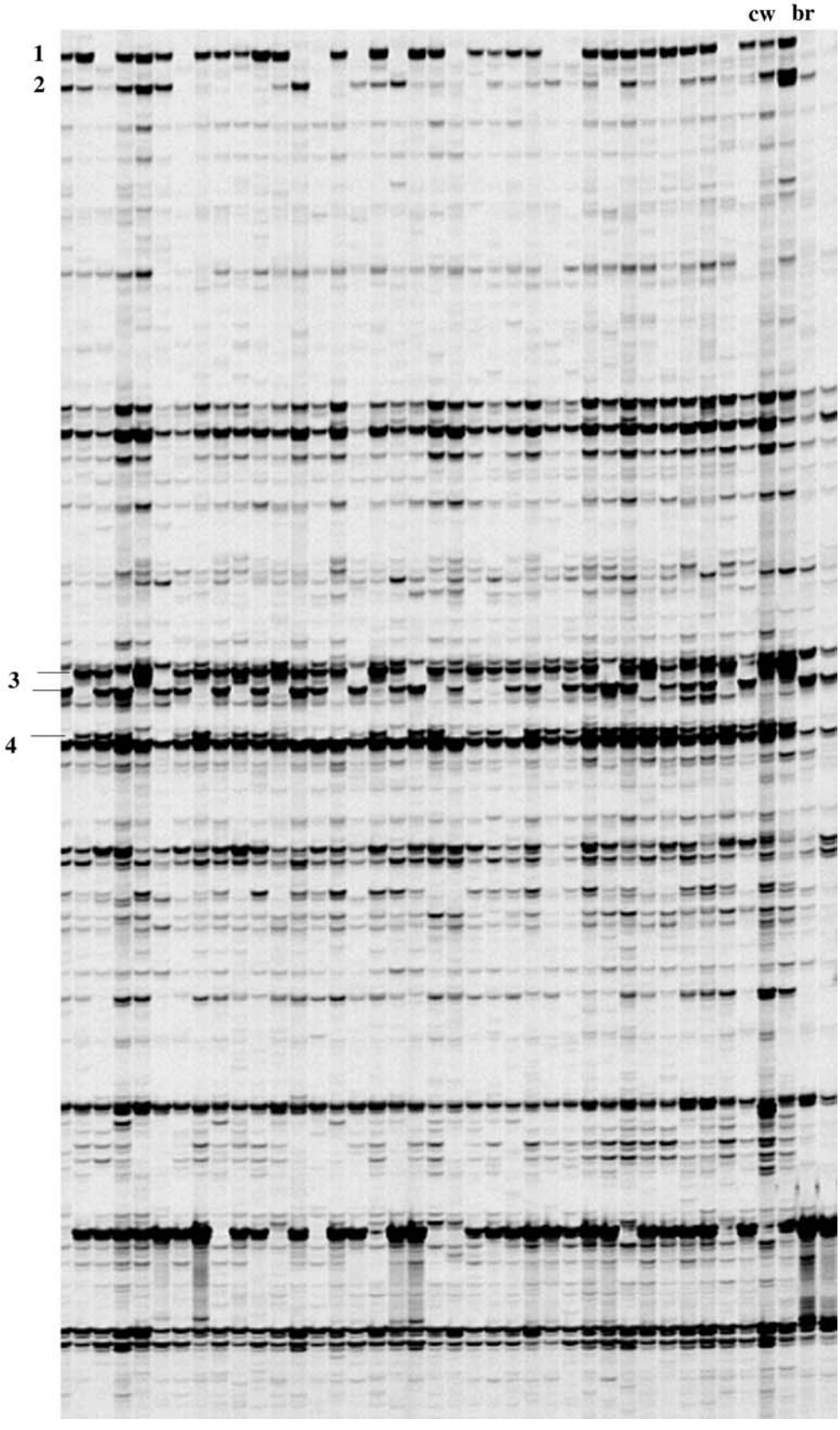


Fig. 1 cDNAs from F2 population and parental lines (br = broccoli, cw = cauliflower) amplified with primer combination DC1 + ODD10.
1 = dominant marker T212,
2 = dominant marker T211,
3 = co-dominant marker T210,
4 = dominant marker T209

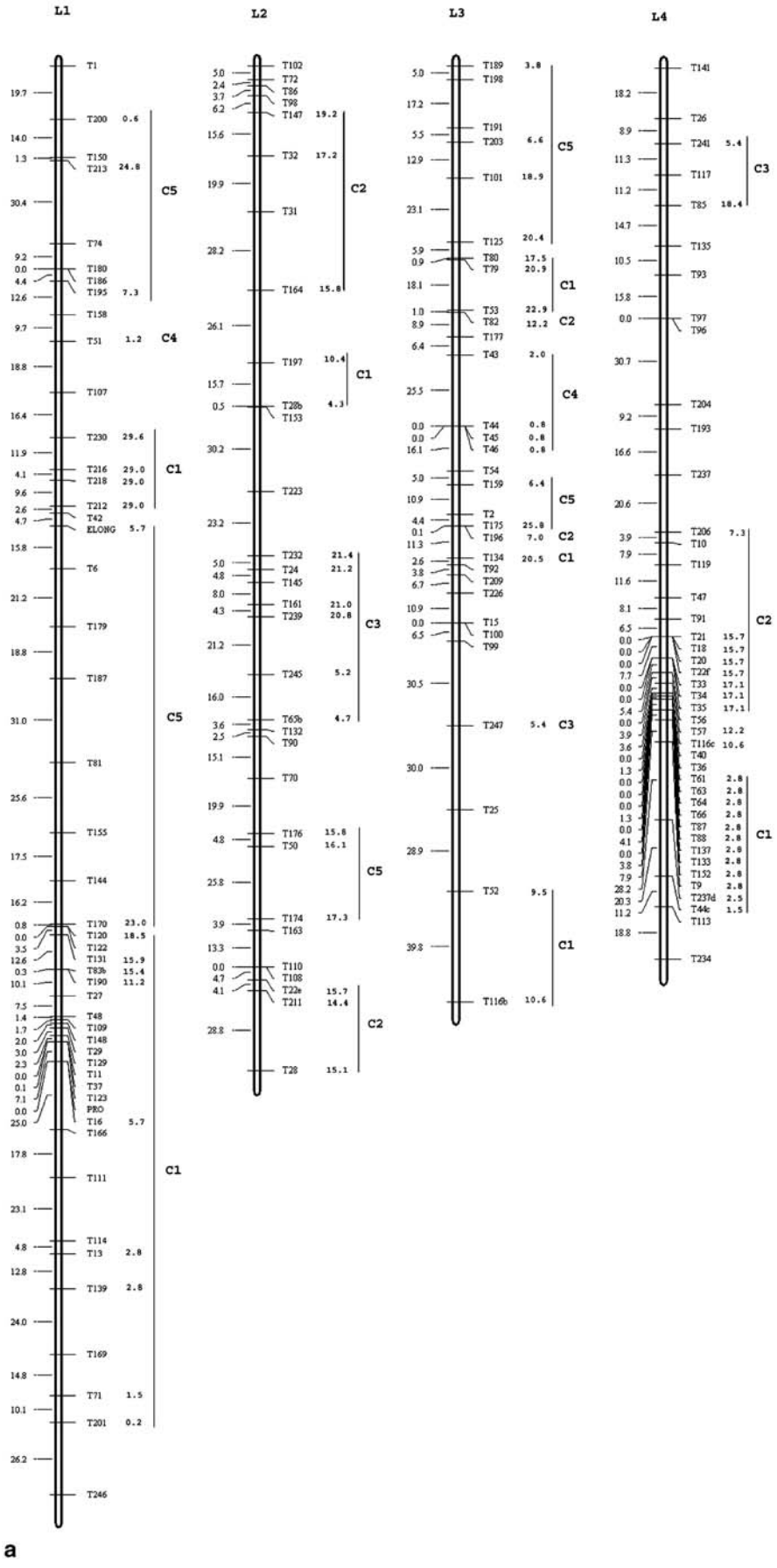
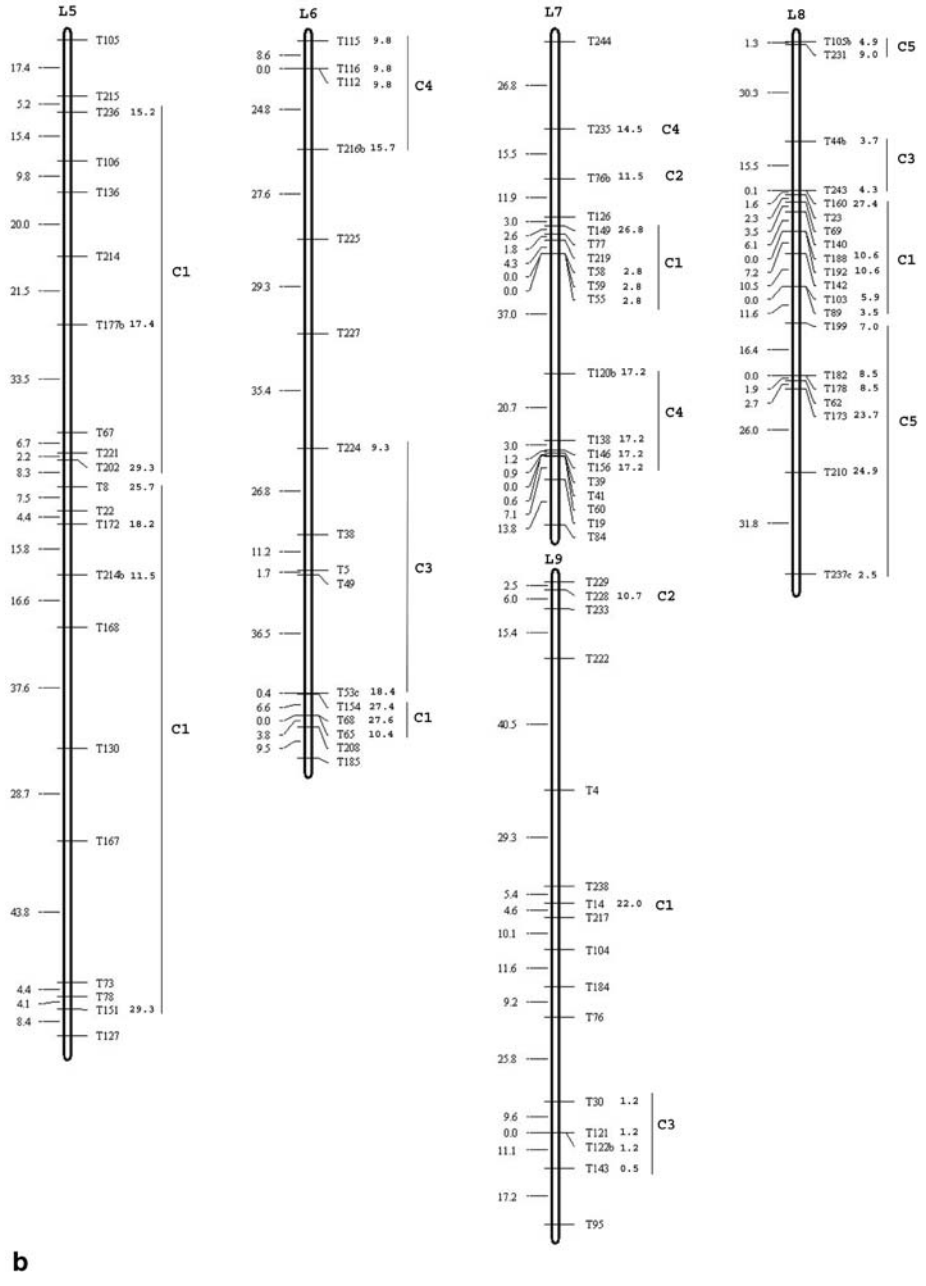


Fig. 2 Nine linkage groups (L1 to L9) in the transcriptional map of *B. oleracea*. *Vertical bars* indicate corresponding sizes on *Arabidopsis* chromosomes (C1 = chromosome 1 to C5 = chromosome 5). On the left of each group genetic distance in cM is shown. On the right, next to the marker number, the physical location of the corresponding gene is shown in MB

Fig. 2 (continued)



glucosinolates and the presence of cDNA markers matching the IPMS *Arabidopsis* homologs on chromosomes 1 (corresponding to the *Brassica* gene *BoGSL-PRO*) and 5 (corresponding to the *Brassica* gene *BoGSL-ELONG)*. The markers in the map fell into nine linkage groups, which were named arbitrarily since we did not attempt at this time to align them to those of other existing *B. oleracea* maps (Hu et al. 1998). The largest group contained 49 markers, and the smallest one had 16 markers. Of the 247 mapped markers we produced, 149 sequences had homologs in *Arabidopsis*. These sequences allowed us to do a gene-for-gene global genome alignment between *B. oleracea* and *Arabidopsis* (Fig. 2, Table 2). Based on the conservation of gene order in these two species, we found broad colinearity between the two genomes for chromosomal segments rather than for whole chromosomes. In general, the *Brassica* linkage groups were composed of multiple syntenic *Arabidopsis* chromosome segments dispersed on all chromosomes and often showing inversions and deletions/insertions (Fig. 2). For example, in linkage group 3, there were syntenic segments corresponding to three *Arabidopsis* chromosomes, including two overlapping segments for chromosome C5, but positioned at different locations on the *Brassica* linkage group. Interestingly, nearly all markers on linkage group 5, except T22, hit genes on chromosome 1 in *Arabidopsis*, but covered two duplicate but inverted and overlapping regions, the first one ranging from genes at positions15–29 Mb and the second one from 11 to 29 Mb of chromosome 1 of *Arabidopsis* (Fig. 2).

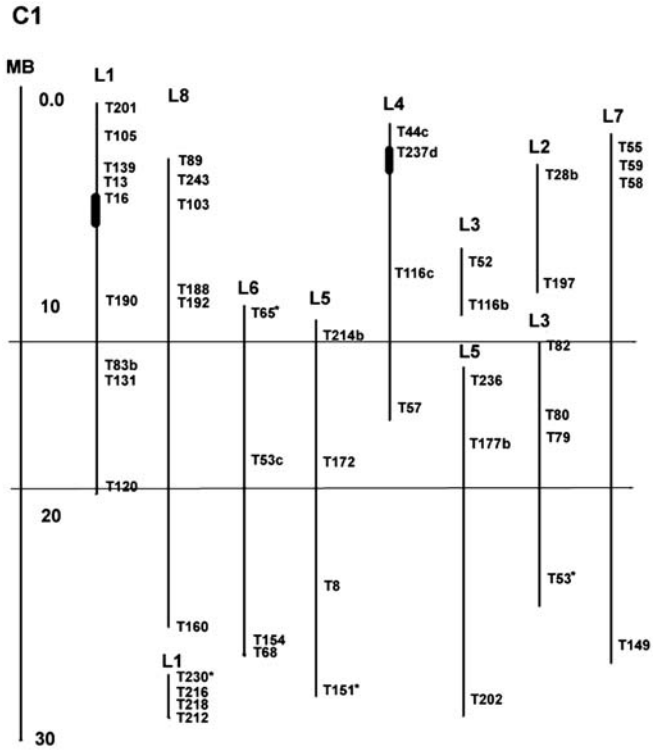


Fig. 3 Diagrammatic representation of *Arabidopsis* chromosome 1 (C1, in MB) showing, between the two horizontal lines, duplicated regions in 7 to 8 copies of *B. oleracea* linkage group (LG) segments. cDNA markers on linkage groups corresponding to coding sequences in C1 are shown at their approximate positions. These markers had high confidence scores with Arabidopsis, mostly maintaining gene order, with a few exceptions (asterisk). Refer to Table 2 for corresponding Arabidopsis sequences to these markers. The *black bubble* indicates a cluster of up to 20 markers (see Fig. 1 for details)

Inspection of the *B. oleracea* genome structure using the *Arabidopsis* genome as a reference, revealed extensive duplication in the *B. oleracea* genome, as reported before by many other studies (Quiros 2001). The distribution of the duplicated segments, however, was uneven. For example, *Arabidopsis* chromosome 1 aligned to 11 segments from eight *Brassica* linkage groups (Figs. 2, 3). Most of these display up to six overlapping duplicated regions. There were four segments located on linkage groups 1, 2, 4 and 8 with similar gene order and content as that observed at the top arm of chromosome 1 in *Arabidopsis*, covering the regions ranging from 0.2 to 18.5, 4.3 to 10.4, 2.5 to 12.2 and 3.5 to 10.6 Mb on the *Arabidopsis* physical map, respectively. Similarly, for the rest of chromosome 1, the region spanning from 10 to 30 Mb was also aligned to four segments of linkage groups 1, 5 and 6 in the *B. oleracea* transcriptome map. *Arabidopsis* chromosome 5 aligned with seven segments derived from four *Brassica* linkage groups displaying up to six overlapping duplications, although five of them included larger overlapping areas (Fig. 2). *Arabidopsis* chromosome 3 was mainly represented by a segment on linkage groups 2 and 7, containing homologs for genes at positions 4.7 to 21.4 MB and 0.5 to 1.2 MB, respectively. There were four other segments, but containing only two markers each, on linkage groups 4, 6 and 8. Three of these segments displayed short overlapping duplications. *Arabidopsis* chromosomes 2 and 4 were under represented in the *Brassica* genome. On linkage groups 2 and 4 there are two overlapping segments corresponding to *Arabidopsis* chromosome-2 regions at positions 15.8–19.2 Mb and 7.3 to 17.1, respectively. Additionally, linkage group 2 has a rearranged segment for genes included at 14.4 to 15.7 MB, and linkage group 3 included a short segment containing two markers matching genes at positions 12.2 and 22.0 MB. We could not find duplications in the *Brassica* linkage map for *Arabidopsis* chromosome 4. Two segments for this chromosome were present on linkage group 3 at positions 0.8–2.0 MB and group 6 at positions 9.8–15.7 MB.

Discussion

The observed polymorphism of the transcriptome markers observed comes from template differences due to SNPs (Brugmans et al. 2002), and splicing-site changes resulting in transcripts of different size (e.g., *BoGLS-ELONG*, Li and Quiros 2002). Although we did not score band intensity, the quantitative appraisal of the bands is expected to disclose many more polymorphisms, like those reported in budding yeast (Brem et al. 2002). For such an evaluation it would be ideal to work with gene circuits where gene members are known to be coordinately regulated.

Our report on the alignment of both genomes is based on the *Brassica* gene members displaying the highest level of similarity to their orthologs in *Arabidopsis*. It was possible to align the chromosomal segments of both species by their similarity values as well as their expected sequential order based on the *Arabidopsis* homologs on those segments. Not surprisingly, some of the cDNA marker sequences often hit more than one gene in *Arabidopsis* due to the extensive duplication in the genome of this species. However, there was a clear difference in their similarity score values for most of these genes, thus allowing the identification of orthologs and eliminating the ambiguity often observed by EST mapping, unless extensive computer algorithms are applied (Fulton et al. 2002). Therefore, it is clear that when dealing with duplications, a common situation in plant genomes, transcriptome mapping for cross-genome comparisons is superior to maps generated by DNA hybridization. Our approach makes comparative genomics straightforward and precise.

The large number of sizable duplications, as well as the uneven representation of these segments in *B. oleracea* observed in our study, is not unexpected considering the high level of duplication in *Arabidopsis thaliana*, which is nearly equivalent to that expected for a tetraploid (*Arabidopsis* Genome Initiative 2000). *B. oleracea* has almost twice the number of chromosomes

of *Arabidopsis*. Therefore, assuming that the *Brassica* species derive from an ancestral lineage undergoing a similar level of duplications as the lineage leading to *Arabidopsis*, one would expect to find mostly four copies of chromosomal segments, which was not the case. We observed instead that some segments, like those corresponding to *Arabidopsis* chromosome 1 are in seven to eight copies in the *Brassica* genome, whereas other segments, like those corresponding to *Arabidopsis* chromosomes 2 and 4 are poorly represented, with few or no copies. Incidentally, these two chromosomes are the ones reported to contain sizable duplicated segments in *Arabidopsis*, which indicate that the *Arabidopsis* and *Brassica* lineages are quite divergent from each other. This is consistent with their estimated time of separation of over 20 million years, and their placement in different tribes (Wroblewski et al. 2000). The lack of even representation of all five *Arabidopsis* chromosomes in the duplicated segments of the *Brassica* genome argues against ancient hexaploidy or the triplication of the whole genomes followed by gene loss in *Brassica*, as suggested by Cavell et al. (1998) and Parkin et al. (2002) among others, based mostly on DNA hybridization analysis. Our results certainly contradict the statement of Gale and Devos (1998) stating that the *Arabidopsis* genome is "essentially triplicated in the diploid *Brassica* species". If such was the case, one would expect to find six copies per segment for all five *Arabidopsis* chromosomes evenly distributed in the *Brassica* genome. Instead, the variable number of duplications and rearrangements we observed is rather consistent with events of higher complexity than simple polyploidization, leading to the synthesis of *Brassica* genomes including also aneuploidy and chromosomal rearrangement (Quiros 2001).

Transcriptome mapping not only places genes on the map accurately but also detects gene function directly based on their co-segregation with the trait they control. For example, a candidate gene for 4-C side-chain glucosinolates *BoGSL-ELONG* was identified by this approach after finding a cDNA marker for this gene that was completely linked to the presence of 4-carbon glucosinolates (Li and Quiros 2002). Another important advantage of transcriptome mapping worth stressing is the fact that repetitive DNA, introns and gene spacers are excluded from the sample. This greatly reduces the effective genome size, making it much easier to find a marker physically closely associated to a gene, such as the 3-carbon side-chain glucosinolate trait determined by the *BoGSL-PRO* candidate gene included in the map. The perfect co-segregation of a cDNA marker matching an IPMS *Arabidopsis* homolog and presence of 3-carbon glucosinolates is good evidence that it is the right candidate gene.

Multiple amplification of the same gene by different primer sets demonstrates that a single gene transcript can be found efficiently when there are differences in gene expression between the two alleles at a locus detected by segregation in a mapping population. Additionally, since genes with tissue-specific expression are the rule in eukaryotes, we could dissect the whole genome into different pools, where each pool is a genome subsample, by isolating RNA from different tissues. In the present study we used young leaves to extract RNA, and, not surprisingly, the analysis of the sequences indicate that most of the genes detected are related to plant growth, such as genes coding for photosystem proteins. This advantage could be very useful for species with large genomes.

In conclusion, transcriptome mapping is an efficient and relatively low-cost approach superior to DNA hybridization techniques allowing gene alignment between a fully sequenced and a poorly characterized genome. Furthermore, this procedure allows gene-expression studies and quick development of markers associated with genes of economic importance for cloning and marker-assisted selection.

Acknowledgements We are indebted to Vincent D'Antonio, Isha Guo and Nicolas Rios for technical assistance, and to Elizabeth Earle, Thomas Osborn, Sheila McCormick and Roger Chetelat for reading of the manuscript. Research was supported by USDA-IFAFS grant # 00-52100-9683, and the "Development of Genomic Tools and Resources for *B. oleracea*".

References

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

Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2:96–103

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755

Brugmans B, Carmen A, Bachem C, Os H, Eck H, Visser R (2002) A novel method for construction of genome-wide transcriptome maps. Plant J 31:211–222

Cavell A, Lydiate D, Parkin I, Dean C, Trick M (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome 41:62–69

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang FC, Kim H, Frisch D, Yu YS, Sun SH, et al. (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Draye X, Lin Y, Xian X, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren S, Wing R, Paterson AH (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol 125:1325–1341

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Gale M, Devos K (1998) Plant comparative genomics after 10 years. Science 23:656–659

Hu J, Sadowski J, Osborn TC, Landry BS, Quiros CF (1998) Linkage group alignment from four independent *Brassica oleracea* RFLP maps. Genome 41:226–235

Lan T, DelMonte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S, Paterson AH (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res 10:776–788

Li G, Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theor Appl Genet 103:455–461

Li G, Quiros CF (2002) Genetic analysis, expression and molecular characterization of *BoGSL-ELONG*, a major gene involved in the aliphatic glucosinolate pathway of *Brassica* species. Genetics (in press)

Li G, Riaz A, Goyal S, Abel S, Quiros CF (2001) Inheritance of three major genes involved in the synthesis of aliphatic glucosinolates in *Brassica oleracea*. J Am Soc Hort Sci 126:427–431

Lockhart DJ, Dong HL, Byrne MC, Follettie MF, Gallo MV, Chee MS, Mittmann M, Wang CW, Kobayashi M, et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnol 14:1675–1680

MacIntosh GC, Wilkerson C, Green PJ (2001) Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs. Plant Physiol 127:765–776

Matsumura H, Nirasawa S, Terauchi R (1999) Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant J 20:719–726

O'Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. Plant J 23:233–243

Parkin IAP, Lydiate DJ, Trick M (2002) Assessing the level of colinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. Genome 45:356–366

Pearson WR, Lipman D (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C, Katsar C, Lan T, Lin Y, et al. (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Quiros CF (2001) DNA-based marker maps of *Brassica*. In: Phillips RL, Vasil JK (eds) DNA-based markers in plants. Kluwer Academic Publishers, London, pp 20-1-238

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Wroblewski T, Coulibaly S, Sadowski J, Quiros CF (2000) Variation and phylogenetic utility of the *Arabidopsis thaliana* Rps2 homolog in various species of the tribe Brassiceae. Mol Phylog Evol 16:440–448