Theor Appl Genet (2010) 120:1485–1500
DOI 10.1007/s00122-010-1270-4

ORIGINAL PAPER

# Genomic microstructure and differential expression of the genes encoding UDP-glucose:sinapate glucosyltransferase (UGT84A9) in oilseed rape (*Brassica napus*)

**Juliane Mittasch · Sabine Mikolajewski · Frank Breuer · Dieter Strack · Carsten Milkowski**

Received: 12 August 2009 / Accepted: 12 December 2009 / Published online: 20 January 2010


**Abstract** In oilseed rape (*Brassica napus*), the glucosyltransferase UGT84A9 catalyzes the formation of 1-*O*-sinapoyl-*β*-glucose, which feeds as acyl donor into a broad range of accumulating sinapate esters, including the major antinutritive seed component sinapoylcholine (sinapine). Since down-regulation of UGT84A9 was highly efficient in decreasing the sinapate ester content, the genes encoding this enzyme were considered as potential targets for molecular breeding of low sinapine oilseed rape. *B. napus* harbors two distinguishable sequence types of the *UGT84A9* gene designated as *UGT84A9-1* and *UGT84A9-2*. *UGT84A9-1* is the predominantly expressed variant, which is significantly up-regulated during the seed filling phase, when sinapate ester biosynthesis exhibits strongest activity. In the allotetraploid genome of *B. napus*, *UGT84A9-1* is represented by two loci, one derived from the Brassica C-genome (*UGT84A9a*) and one from the Brassica A-genome (*UGT84A9b*). Likewise, for *UGT84A9-2* two loci were identified in *B. napus* originating from both diploid ancestor genomes (*UGT84A9c*, Brassica C-genome; *UGT84A9d*, Brassica A-genome). The distinct *UGT84A9* loci were genetically mapped to linkage groups N15 (*UGT84A9a*), N05 (*UGT84A9b*), N11 (*UGT84A9c*) and N01 (*UGT84A9d*). All four *UGT84A9* genomic loci from *B. napus* display a remarkably low micro-collinearity with the homologous genomic region of *Arabidopsis thaliana* chromosome III, but exhibit a high density of transposon-derived sequence elements. Expression patterns indicate that the orthologous genes *UGT84A9a* and *UGT84A9b* should be considered for mutagenesis inactivation to introduce the low sinapine trait into oilseed rape.

Communicated by C. Quiros.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-010-1270-4) contains supplementary material, which is available to authorized users.

J. Mittasch · D. Strack · C. Milkowski
Department of Secondary Metabolism,
Leibniz Institute of Plant Biochemistry,
Weinberg 3, 06120 Halle (Saale), Germany

S. Mikolajewski
Institute for Crop Science and Plant Breeding,
Bavarian State Research Center for Agriculture,
Am Gereuth 8, 85354 Freising, Germany

F. Breuer
Molecular Breeding of Oilseed Rape, KWS Saat AG,
Grimsehlstraße 31, 37555 Einbeck, Germany

*Present Address:*
C. Milkowski (✉)
Institute of General Botany and Plant Physiology,
Friedrich Schiller University Jena,
Am Planetarium 1, 07743 Jena, Germany
e-mail: carsten.milkowski@uni-jena.de

## Introduction

Brassicaceae plants display a remarkable metabolic flux toward soluble sinapate esters, which accumulate in organ- and tissue-specific patterns. In oilseed rape (*Brassica napus*, Canola), sinapoylcholine (sinapine) constitutes the predominating phenolic seed compound ranging from 0.7 to 4% (w/w) with about 90% localized to the embryo (Blair and Reichert 1984; Velasco and Möllers 1998; Wang et al. 1998). As minor seed components, a set of at least 14 different sinapate esters was identified (Baumert et al. 2005), including sinapoylated glucose, gentiobiose and kaempferol glycosides.

The biological role of sinapine accumulation in seeds has not been definitely elucidated, although it has been suggested that highly concentrated sinapine serves as a ready-stored supply of the highly hygroscopic choline for the biosynthesis of phosphatidylcholine in young seedlings (Strack 1981). As a major antinutritive compound, however, sinapine spoils considerably the quality of rapeseed meal, which is enriched in valuable proteins of a well-balanced composition (Ohlson 1978). Sinapate esters may form complexes with proteins, thus causing reduced palatability of the meal and contribute to the bitter taste and astringency of rapeseed products (Nazck et al. 1998; Rawel et al. 2000). Together with other undesirable components like phytate, seed coat imposed fibers and complex proanthocyanidins, the high sinapate ester content of the embryo converts the valuable rapeseed meal into a by-product of the *B. napus* seed discounted about 35% relative to the meal from soybean (*Glycine max*) that currently dominates the market of food supplements. Accordingly, a strong decrease of the amount of seed-specific sinapate esters, commonly regarded as "low sinapine trait", has become a major aim in modern oilseed rape breeding.

While fibers and proanthocyanidines were substantially decreased by breeding of yellow seeded Canola lines (Rahman 2001), the pronounced sinapine accumulation of the embryo remains difficult to challenge because of the low level of genetic variability in oilseed rape. Currently, there is no germplasm available that meets the low sinapine requirement of less than 0.5% (w/w) (zum Felde et al. 2007). This hampers substantially the development of powerful breeding strategies. Hence, transgenic approaches based on metabolic engineering of the complex sinapate ester biosynthetic network became an alternative.

Sinapate (4-hydroxy-3,5-dimethoxycinnamate) is produced via the phenylpropanoid pathway from the aromatic amino acid phenylalanine and feeds via 1-*O*-sinapoyl-*β*-glucose into various sinapate esters (Fig. 1). Given its pivotal role as acyl donor, the formation of 1-*O*-sinapoyl-*β*-glucose from sinapate and UDP-glucose by the enzyme UDP-glucose: sinapate glucosyltransferase (SGT; EC 2.4.1.120) was hypothesized as critical metabolic step whose suppression might provoke a decrease of the sinapate ester content. From seeds of *B. napus*, a cDNA was cloned encoding SGT and classified as *UGT84A9* (Milkowski et al. 2000). Transgenic lines of oilseed rape that were homozygous for a single copy insertion of a dsRNAi cassette designed to silence *UGT84A9*, displayed a strong decrease (ca. 80%) in total sinapate ester content. In these lines, the remaining sinapine was about 30% of control plants (Hüsken et al. 2005). Interestingly, in the silenced plants there was no indication of a possible metabolic redirection of sinapate into other conjugates of the soluble seed metabolite fraction. This net loss of sinapoyl moieties approved inactivation of

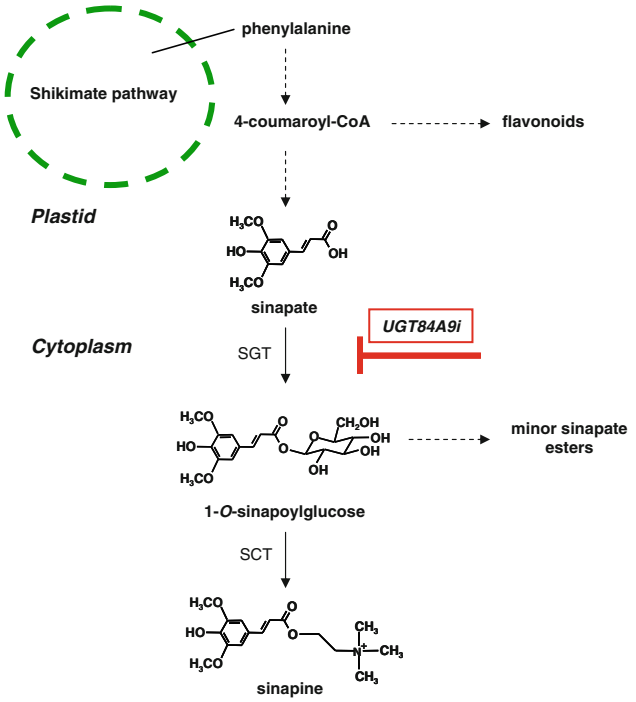


**Fig. 1** Sinapine biosynthesis. Phenylalanine provided by the plastid-localized shikimate pathway enters the general phenylpropanoid pathway. The activated 4-coumaroyl-CoA serves as precursor for a wide array of phenolic compounds, of which flavonoids are ubiquitous to plants. Sinapate is fed via the *β*-acetal ester 1-*O*-sinapoyl-*β*-glucose into various sinapate esters with sinapine (sinapoylcholine) as major seed component. Dashed arrows denote multiple biosynthetic steps. The dsRNAi strategy designed to reduce sinapine accumulation in *B. napus* seeds is indicated. SGT, UDP-glucose:sinapate glucosyltransferase (EC 2.4.1.120); SCT, 1-*O*-sinapoyl-*β*-glucose:choline sinapoyltransferase (EC 2.3.1.91)

*UGT84A9* genes as promising strategy to generate low sinapine oilseed rape lines.

To exploit this information for conventional breeding would require the selection of mutants impaired in *UGT84A9* genomic loci. This could be done by targeting induced local lesions in genomes (TILLING), a reverse genetics technique to identify mutations in any target gene through heteroduplex analysis (Till et al. 2003). For this approach, knowledge on gene copy number, expression profiles and the microstructure of the according loci in the allotetraploid genome of *B. napus* is an indispensable requirement.

Based on extensive analyses of cDNAs and genomic BAC clones, the present work was aimed at characterizing the *UGT84A9* loci present in the genome of *B. napus*. The genomic microstructure of the *UGT84A9* loci was elucidated and used to assess collinearity with the homologous region of the *Arabidopsis thaliana* genome. The differential expression patterns of individual *UGT84A9* gene copies were used to define target genes for breeding of low sinapine oilseed rape.

## Materials and methods

### Plant material and cultivation

Winter oilseed rape (*B. napus* L. var. *napus* cv. Express), forage kale (*B. oleracea* L. var. *medullosa* cv. Markola) and turnip (*B. rapa* L. var. *campestris* cv. Rex) were obtained from Norddeutsche Pflanzenzucht (Holtsee, Germany). Seeds were sown on soil and germinated under constant light at 20°C for 5 days. Plants were grown in the greenhouse at 12–18°C under a 16 h light regiment. Winter cultivars were vernalized at 4°C for 8 weeks after reaching the five to seven leaves stage.

### Nucleic acids techniques

Purification of DNA and RNA was generally done by selective adsorption onto silica using appropriate commercial preparation kits (Qiagen, Hilden, Germany). Total RNA was extracted from different plant tissues. To prevent DNA contamination, on-column incubation with DNase I was included in the protocol. Poly($A^+$) RNA was purified from total RNA by selective binding to oligo-dT-Oligotex beads (Qiagen). Molecular cloning experiments and restriction analyses were performed according to the standard protocols (Sambrook et al. 1989). Routine sequencing of DNA was done by a commercial supplier (Eurofins MWG Operon, Ebersberg, Germany). PCR-based techniques were carried out according to standard protocols given by the commercial suppliers of reaction mixtures. The primers used for the different approaches are given in Table 1.

### RT–PCR

Total RNA (1 µg) was reversely transcribed in a 20 µl reaction mixture using the Omniscript RT kit (Qiagen) and an oligo-$(dT)_{15}$ primer (Promega, Madison, WI, USA). For amplification, the reverse transcription reaction mix was diluted 25-fold and included as template into a 20 µl PCR (Taq PCR Master Mix Kit, Qiagen) with primers 1–12 (Table 1). PCR was run for 30 cycles of denaturation (95°C, 30 s), annealing (5°C below primer melting temperature, 30 s) and elongation (72°C, 1 min $kb^{-1}$). PCR products were examined by agarose gel electrophoresis.

### Cloning of *UGT84A9* genes

Full-length sequences of *UGT84A9* genes were obtained by PCR with primers 13 and 14 (Table 1) and genomic DNA or cDNA as template. PCR was conducted for 35 cycles of denaturation (95°C, 10 s), annealing (60°C, 30 s) and elongation (68°C, 1 min $kb^{-1}$) with SuperMix High Fidelity (Invitrogen, Karlsruhe, Germany). PCR products were cloned into pGEM-T® Easy (Promega, Madison, WI, USA) and subsequently sequenced.

### cDNA-AFLP

Poly($A^+$) RNA from four developmental stages (flowers, seeds at cotyledon stage, seeds at well-developed mature embryo stage and 2-day-old seedlings) was used with the SMART™ PCR cDNA Synthesis Kit (BD Clontech, Heidelberg, Germany) to synthesize double-stranded cDNA by the template switch method, which produced first strand cDNA labeled on both ends. In a second step, first strand cDNA was amplified by PCR. The obtained double-stranded cDNA was purified by proteinase K treatment and nucleotide overhangs were removed by T4 DNA polymerase. The cDNA-AFLP (cDNA amplified restriction fragment length polymorphism) was carried out using the standard AFLP protocol (Vos et al. 1995) on a double-stranded cDNA template (Bachem et al. 1996). Primers used (19–34; Table 1) and AFLP adapters were derived according to the Keygene protocol (Keygene, Wageningen, The Netherlands). For the restriction of double-stranded cDNA the enzymes *Hin*dIII and *Mse*I were used. Identities of eluted transcript-derived fragments (TDFs) were confirmed by PCR product sequencing using the primers with three specific nucleotides at the 3′ end (25–27, 31–34; Table 1).

### Heterologous expression of *UGT84A9* genes

*UGT84A9a* and *UGT84A9c* sequences were amplified from the respective BAC clones by PCR using ProofStart DNA Polymerase (Qiagen). Restriction sites required for cloning were introduced by primers (15–18; Table 1). The produced amplicons were first cloned into pCR®-BluntII-TOPO® (Invitrogen) and after sequence evaluation transferred to the expression vector pET28a(+) (Novagen, La Jolla, CA, USA). For heterologous expression the *E. coli* strain BL21 Codon Plus (DE3)-PR (Stratagene, Darmstadt, Germany) was used. Cells were grown to the exponential phase ($OD_{600}$ 0.6), before heterologous protein expression was induced by adding IPTG to a final concentration of 1 mM to the culture. The bacteria were incubated in the presence of IPTG for 4 h at 30°C under gentle shaking. Cells were harvested by centrifugation and resuspended in a buffer containing 0.1 M MES [2-(*N*-morpholino)ethanesulfonic acid] pH 6.0, 10% (v/v) glycerol, 0.01% (v/v) $\beta$-mercaptoethanol, 0.5 mM EDTA. Cells were disrupted by sonication and the debris was sedimented by centrifugation at 4°C and 14,000 rpm for 20 min. The supernatants containing the soluble proteins were used as crude protein extracts. Protein concentrations were determined by the

**Table 1** Oligonucleotides used as PCR primers or adapters in cDNA-AFLP

| Oligonucleotide | Identifier | Sequence |
|---|---|---|
| 1 | UGT84A9fw7 | CAC GAC GAG ATC CCT TCT TTC |
| 2 | UGT84A9rev7 | GTC ACG AAA CAA ACC ACA GAA G |
| 3 | UGT84A9-1fw | TGC AAG CAG AAC CAA CTT AAC CAT CCT |
| 4 | UGT84A9-1rev | CTC CAT GCA GTC ATC CCT CGT CTC A |
| 5 | UGT84A9-2fw | CGT AAG AAG ACA CGA TTT CAC CAT CTA |
| 6 | UGT84A9-2rev | TTC CAT ACA ATG GTC CGT GGT CTC G |
| 7 | Napin fw | GCA CAA CAC CTA AGA GCT TG |
| 8 | Napin fw | GCA GCT GCT GTC CCT GCT GT |
| 9 | Ubiquitin fw | AGG CCA AGA TCC AGG ACA AAG |
| 10 | Ubiquitin rv | CGA GCC AAA GCC ATC AAA GAC |
| 11 | Genomic DNA control fw | TTC TAA GGT CTC GTG GGC TCA GTC AAC |
| 12 | Genomic DNA control rv | TAA CGT GGC CTT GTC CTG GGA AGG ATA C |
| 13 | FL_UGT84A9fw | ATG GAA CTA TCA TCT TCT CCT TTA CCT CCT CAT G |
| 14 | FL_UGT84A9rev | TTA TGA CTT TTC CAA TAA AAG TTC TTG AAC ACT TCC GTT TT |
| 15 | *Nco*I-UGT84A9_Ifw | GGC CCA TGG AAC TAT CAT CTT CTC CTT TAC |
| 16 | *Eco*RI-UGT84A9_Irv | GGC GAA TTC TGA CTT TTG CAA TAA AAG TTT TTG AAT ACT TCC |
| 17 | *Nco*I-UGT84A9_IIfw | GGC CCA TGG AAC TAG AAT CTT CTC CTT TAC |
| 18 | *Eco*RI-UGT84A9_IIrv | GGC GAA TTC TGA CTT TTC CAA TAA AAG TTC TTG AAC AGT TCC |
| 19[a] | *Hin*dIII-Adapter | CTC GTA GAC TGC GTA CC |
| 20[a] | *Hin*dIII-Adapter (second strand) | AGC TGG TAC GCA GTC TAC |
| 21[a] | *Mse*I-Adapter | GAC GAT GAG TCC TGA G |
| 22[a] | *Mse*I-Adapter (second strand) | TAC TCA GGA CTC AT |
| 23 | H03 | GAC TGC GTA CTA GCT TG |
| 24 | H04 | GAC TGC GTA CTA GCT TT |
| 25[b] | H22 | GAC TGC GTA CTA GCT TGT |
| 26[b] | H76 | GAC TGC GTA CTA GCT TGT C |
| 27[b] | H90 | GAC TGC GTA CTA GCT TTG T |
| 28 | M01 | GAT GAG TCC TGA GTA AA |
| 29 | M02 | GAT GAG TCC TGA GTA AC |
| 30 | M03 | GAT GAG TCC TGA GTA AG |
| 31 | M36 | GAT GAG TCC TGA GTA AAC C |
| 32 | M40 | GAT GAG TCC TGA GTA AAG C |
| 33 | M47 | GAT GAG TCC TGA GTA ACA A |
| 34 | M64 | GAT GAG TCC TGA GTA AGA C |

[a] Adapter in cDNA-AFLP

[b] 5′ fluorescence labeled

Bradford method (Bradford 1976) with bovine serum albumin as standard.

## SGT activity assays

Crude protein extracts were incubated in MES buffer (0.1 M MES at pH 6.0, 10% (v/v) glycerol, 0.01% (v/v) $\beta$-mercaptoethanol, 0.5 mM EDTA) with 4 mM UDP-glucose and 2 mM sinapate in a final volume of 150 µl at 30°C for 30 min. Reactions were terminated by adding 10 µl of trifluoroacetic acid (TFA). Reaction products were analyzed by HPLC as described previously (Milkowski et al. 2004).

## Screening of a genomic BAC library

A genomic BAC library from *B. napus* cv. Express providing an eightfold genome coverage was obtained from the University of Giessen. Screening was performed by hybridization of $^{33}$P-labeled cDNAs to high-density colony macroarrays according to the protocol given by the provider (German Resource Center for Genome

Research (RZPD), Berlin, Germany). The probe was labeled by random priming using the MegaPrime-DNA Labeling System (GE Healthcare Life Science). Non-incorporated nucleotides were removed by gel filtration chromatography on ProbeQuant G-50 Micro Columns (GE Healthcare Life Science). Signals were detected by exposing the filters to X-ray films for 2 days at −80°C and assigned to BAC clones according to evaluation schemes provided by RZPD.

### Isolation and characterization of BAC-DNA

For small scale BAC-DNA preparation, the BACMAX DNA purification kit (Epicentre, Madison, WI, USA) was used, large scale preparation was done according to Birnboim (1983). BACs were initially characterized by PCR using the *Taq* PCR Master Mix Kit (Qiagen). PCR cycles included denaturation (95°C, 30 s), annealing (5°C below the primer melting temperature, 30 s) and elongation (72°C, 1 min $kb^{-1}$). Specific primers for *UGT84A9* sequence type evaluation (3–6) and primers for full-length amplification of *UGT84A9* (13, 14) are given in Table 1. BAC subfragments generated by restriction were cloned into pGEM-T® Easy (Promega) by standard protocols. Recombinant plasmids were selected by colony hybridization. Colonies were transferred to Hybond $N^+$ nylon membranes (Amersham Biosciences). After cell lysis and DNA denaturation by placing the membrane on denaturation buffer (0.5 M NaOH, 1.5 M NaCl), the membranes were neutralized with 1 M Tris–HCl, 2 M NaCl pH 7.5, and the DNA was fixed by UV-crosslinking. Hybridization and signal detection were done according to the Southern blot protocol.

### Southern blot

BAC-DNA (100 ng) was subjected to restriction with the appropriate endonuclease overnight. Fragments were separated by electrophoresis in TAE buffer (0.04 M Tris–acetate, 0.001 M EDTA pH 8.0) using a 0.75% agarose gel. For size determination, DIG-labeled DNA Molecular Weight Marker II (Roche) was applied. Blotting on a nylon membrane (Hybond NX, Amersham Biosciences, Piscataway, NJ, USA) was carried out by capillary blot technique using 20×SSC (3 M NaCl, 0.3 M sodium citrate pH 7.0) as transfer buffer (Sambrook et al. 1989). The probe was labeled with the PCR DIG Probe Synthesis kit (Roche, Mannheim, Germany), hybridization was conducted in DIG Easy Hyb buffer (Roche) at 42°C overnight. Signal detection was carried out using the DIG Luminescent Detection kit (Roche) according to the manufacturer's protocol.

### Sequence analysis and phylogenetic tree construction

DNA and protein sequences were analyzed by the software package Clone Manager (Scientific & Educational Software, Cary, NC, USA). For sequence comparison, the BLAST algorithm (Altschul et al. 1990) was employed. Online databases used were GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) and TAIR (http://www.arabidopsis.org/). For distance tree construction, DNA sequences of full-length open reading frames were aligned by the CLUSTALW algorithm. For distance calculations, the algorithm of Jukes and Cantor (1969) was used taking into account insertions and deletions. Tree construction was done with the program Treecon (Van de Peer and de Wachter 1994). Bootstrap analysis consisted of 100 replicates.

### Development of molecular markers and genetic mapping

The BAC sequence information raised for the four *UGT84A9* gene copies was used to develop locus specific primers and to establish three SNP derived pyrosequencing markers (PSQ) for *UGT84A9a*, *-b* and *-c* as well as one simple sequence repeat (SSR) marker for capillary electrophoresis (KPE) of *UGT84A9d*. Using the KWS internal PSQ- and capillary electrophoresis devices (ABI3730xl) the developed markers were used for genotyping polymorphisms identified between mapping parents. In order to map the candidate gene loci genetically the software package JoinMap3.0 (Kyazma, The Netherlands) with Haldane's mapping function and a linkage threshold of 3.0 was used.

### Mapping populations and genetic maps

Three double haploid (DH) mapping populations were available for genetic mapping of candidate gene loci in *B. napus*. These populations included the cross between cultivar Express and the resynthetic genotype RS239 with 93 DH lines and an SSR map (not published) containing anchor markers from the Celera AgGen Brassica Consortia (Piquemal et al. 2005), a population derived from a cross between the variety Samourai and the old landrace Mansholts Hamburger Raps with 139 genotypes and a genetic map based on RFLP data (Uzunova et al. 1995) as well as a third population from a cross between Express and the yellow seeded parent 1012-98 consisting of 166 genotypes with an AFLP map plus some SSR markers published (Badani et al. 2006). In the RFLP- and AFLP-based populations further SSR markers from the Celera AgGen Brassica Consortia (Piquemal et al. 2005) were mapped to connect the linkage groups to the public nomenclature (Parkin et al. 2005).

Accession numbers

The isolated sequences described in this paper have been submitted to GenBank and have been assigned the following accession numbers. Gene loci from *B. napus*, *UGT84A9a*, FM872284; *UGT84A9b*, FM872285; *UGT84A9c*, FM872286; *UGT84A9d*, FM872287. Genes from *B. oleracea*, *BoUGT84A9a*, FM872280; *BoUGT84A9c*, FM872281. Genes from *B. rapa*, *BrUGT84A9b*, FM872282; *BrUGT84A9d*, FM872283.

## Results

### *UGT84A9* is represented by two differently expressed sequence types

From cDNA of developing *B. napus* seeds and flowers 12 individual *UGT84A9* full-length clones were sequenced. The multiple sequence alignment revealed two distinct consensus sequences, designated as sequence types *UGT84A9-1* and *UGT84A9-2* that could be distinguished by PCR with selective primer pairs (3–6; Table 1; Supplement S1). *UGT84A9-1* was identical with the *UGT84A9* variant previously isolated from cDNA libraries of *B. napus* developing seeds and young seedlings (GenBank Accession AF287143; Milkowski et al. 2000), and shared a nucleotide sequence identity of 89% with *UGT84A9-2*. RT–PCR with selective primers revealed that *UGT84A9-1* was the predominantly expressed *UGT84A9* variant in *B. napus* with increasing RNA abundance during seed maturation and the maximum level in young seedlings (Fig. 2a). With cDNA from flowers, selective primers produced abundant signals for both *UGT84A9-1* and *UGT84A9-2*. For all other tissues analyzed, the contribution of sequence type *UGT84A9-2* to *UGT84A9* expression was negligible. To prove the results of RT–PCR, a cDNA-AFLP analysis was performed with developing seeds, young seedlings and flowers (Fig. 2b). The signal intensities produced by transcript-derived fragments confirmed *UGT84A9-1* as the predominantly expressed sequence type during seed development and early seedling growth of *B. napus*. In contrast to the results obtained by RT–PCR, cDNA-AFLP revealed low transcript levels for both *UGT84A9* sequence types in flowers with a slightly higher abundance of *UGT84A9-1*. To test functionality, *UGT84A9-1* and *UGT84A9-2* were expressed in *E. coli* and the protein extracts were assayed for enzymatic UGT activity toward cinnamate, 4-coumarate, ferulate and sinapate. As exemplified in Fig. 3 for sinapate, formation of the respective 1-*O*-*β*-glucose esters was observed after incubation of soluble proteins from *E. coli* expressing *UGT84A9-1* and *UGT84A9-2* with the specified hydroxycinnamates in the presence of UDP-glucose. As previously

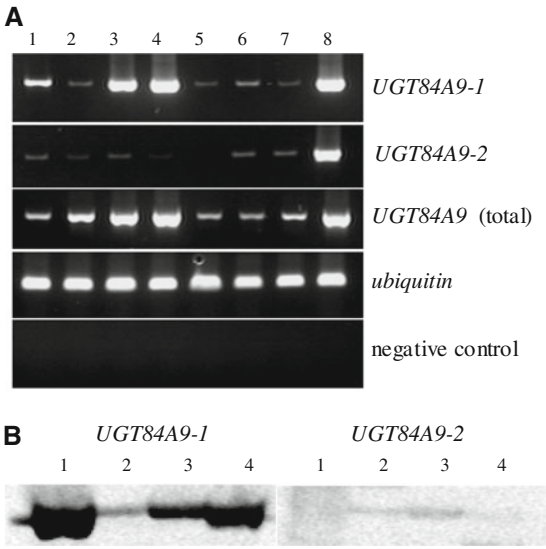


**Fig. 2** Expression of individual *UGT84A9* sequence types in *B. napus*. **a** RT–PCR with sequence type-specific primers for *UGT84A9-1* (primers 3, 4; Table 1) and *UGT84A9-2* (primers 5, 6; Table 1) as well as with a primer pair that recognizes both *UGT84A9* sequence types (primers 1, 2; Table 1) to evaluate the contribution of *UGT84A9-1* and *UGT84A9-2* transcription to the mRNA abundance of *UGT84A9* (*UGT84A9-total*). Template RNA was isolated from seeds at early globular to torpedo stage (*lane 1*), cotyledonary stage (*2*), well-developed mature embryo stage (*3*), from seedlings at 2 days after sowing (DAS; *4*), 7 DAS (*5*), from rosette leaves (*6*), flower buds (*7*) and flowers (*8*). Amplification of ubiquitin cDNA was used as positive control (primers 9, 10; Table 1). To exclude a possible contamination of RNA templates with genomic DNA, a PCR with selective primers for genomic sequences upstream from *UGT84A9a* was run (gDNA control; primers 11, 12; Table 1). **b** *UGT84A9* sequence type-specific cDNA-AFLP. Transcript derived fragments were amplified from seedlings at 2 DAS (*lane 1*), flowers (*2*), seeds at cotyledonary stage (*3*) and well-developed mature embryo stage (*4*)

reported for *UGT84A9-1* (Milkowski et al. 2000) the enzyme encoded by *UGT84A9-2* displayed substrate preference for sinapate as glucose acceptor.

### The genome of *B. napus* harbors four *UGT84A9* loci

A genomic BAC library from *B. napus* was screened by hybridization with *UGT84A9-1* as the probe. This led to the identification of 57 BAC clones supposed to carry *UGT84A9* sequences, which were designated as BAC1 to BAC57. For further characterization, these pre-selected BAC clones were subjected to Southern blot analyses as well as PCR-derived assignments of full-length ORFs and *UGT84A9* sequence types. Clones that produced ambiguous signals were subjected to BAC end sequencing. The results obtained for the individual genomic BAC clones are listed in Table 2.

As a first step, BAC-DNAs were digested by *Nco*I restriction and re-hybridized with *UGT84A9-1* as the probe (data not shown). Since *Nco*I does not cleave within the

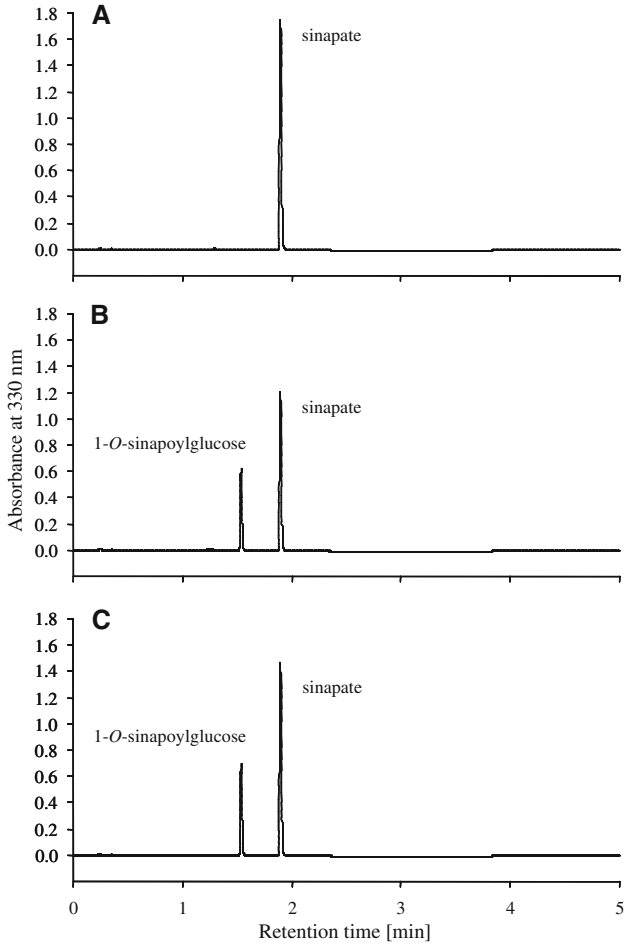


**Fig. 3** Traces of HPLC analyses of UDP-glucose:sinapate glucosyltransferase assays run with protein extracts of *E. coli* harboring the empty vector pET28a(+) (**a**) or the same vector containing cDNAs from *UGT84A9a* (*UGT84A9-1*) (**b**) or *UGT84A9c* (*UGT84A9-2*) (**c**)

intron-less *UGT84A9-1* sequence, single hybridization signals were expected for positive clones. As a result, 53 of the pre-selected BAC clones produced a hybridization signal. The BAC clones 13, 27, 31 and 48 failed to produce any hybridization signal and were excluded from further analyses (Table 2). Of the 53 positive clones, 49 displayed a single hybridization band whereas four (BACs 4, 22, 25, 54) developed two signals indicating an additional *Nco*I site within their putative *UGT84A9* open reading frames.

The 53 hybridization-positive BAC clones were evaluated by PCR for the presence of full-length *UGT84A9* reading frames (primers 13, 14; Table 1) and for the assignment to sequence types *UGT84A9-1* (primers 3, 4; Table 1) or *UGT84A9-2* (primers 5, 6; Table 1). We identified 27 BAC clones harboring sequence type *UGT84A9-1* of which 23 carried the full-length reading frame. 17 BAC clones were identified to contain *UGT84A9-2*, including 15 with the full-length coding sequence (Table 2). Since the *B. napus* BAC library was constructed from genomic fragments raised by partial *Hin*dIII restriction, the occurrence of BAC clones with partial reading frames was expected as it reflects the presence of two *Hin*dIII recognition sites in *UGT84A9*. For nine of the pre-selected BAC clones, the assignment to a distinct *UGT84A9* sequence type by PCR failed.

To determine the number of *UGT84A9* loci in the genome of *B. napus*, the 53 pre-selected BAC clones were subjected to a Southern blot analysis designed to produce locus-specific hybridization patterns (Fig. 4). BAC-DNAs were digested with the restriction endonucleases *Hin*dIII, *Xho*I, *Sal*I and *Bgl*II and hybridized with *UGT84A9-1* cDNA as the probe. Among the BAC clones carrying full-length ORFs of defined sequence types, four different hybridization patterns were generated with each restriction approach. Integration of PCR-derived sequence type determination revealed two distinct hybridization patterns for *UGT84A9-1* as well as for *UGT84A9-2*. This indicated the presence of two different loci for each sequence type in the *B. napus* genome. The resulting four genomic loci were designated as *UGT84A9a-d*, respectively. *UGT84A9a* and *UGT84A9b* belong to sequence type *UGT84A9-1*, whereas the loci *UGT84A9c* and *UGT84A9d* harbor *UGT84A9-2*. For analysis of genomic sequences, DNA sub-fragments were cloned from BAC clones representative for each of the defined *UGT84A9* loci—BAC 1 (representing locus *UGT84A9a*), BAC 4 (*UGT84A9b*), BAC 6 (*UGT84A9c*) and BAC 14 (*UGT84A9d*). Sequence analysis of full-length ORFs confirmed that BAC 1 and BAC 4 carried sequence type *UGT84A9-1* whereas BAC 6 and BAC 14 exhibited sequence type *UGT84A9-2*.

Besides loci determination, Southern blot analyses proved to be helpful in the assignment of BAC clones, which could not be classified by PCR approaches (Table 2; Fig. 4). Thus, for three BACs (7, 29 and 38), previously thought to carry an incomplete ORF, the hybridization patterns revealed a complete specific *UGT84A9* locus. BAC 7 and BAC 38 were found to harbor the *UGT84A9a* gene, whereas BAC 29 carries *UGT84A9d*. Two previously non-defined BACs (18 and 26) were assigned to *UGT84A9a*. The ORF carried by BAC 18 is incomplete, whereas BAC 26 harbors a full-length ORF. BAC 37, which most likely contains a partial *UGT84A9* ORF, showed restriction patterns partly compatible with both *UGT84A9a* and *UGT84A9d*. Due to the incompleteness of the ORF, characteristic bands are missing in the Southern blot, thus making a distinction impossible at this point. For seven BAC clones the Southern hybridization produced signal patterns that differed from those developed for the *UGT84A9* sequence types. Accordingly, these loci were classified as Novel 1 (BACs 2, 3, 21, 30) and Novel 2 (BACs 8, 20, 33).

Clones with an incomplete putative ORF as well as those not yet assigned to one of the four *UGT84A9* loci were

**Table 2** Characterization of genomic BAC clones from *B. napus* carrying *UGT84A9*-related sequences

| BAC no. | Hybridization with *UGT84A9-1* | Sequence type/length | Locus definition[a] | BAC end sequence | PCR product sequence[b] |
|---|---|---|---|---|---|
| 1 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 2 | + | nd | Novel 1 | | |
| 3 | + | 1 p | Novel 1 | KBrB089H07 | |
| 4 | + | 1 f | *UGT84A9b* | | *UGT84A9b* |
| 5 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 6 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 7 | + | 1 p[c] | *UGT84A9a* | | *UGT84A9a* |
| 8 | + | nd | Novel 2 | KBrB089H07 | |
| 9 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 10 | + | 1 p | *UGT84A9a/d* | *UGT84A9a* | |
| 11 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 12 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 13 | – | | | | |
| 14 | + | 2 f | *UGT84A9d* | | *UGT84A9d* |
| 15 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 16 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 17 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 18 | + | nd | *UGT84A9a* | *UGT84A9a* | |
| 19 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 20 | + | nd | Novel 2 | KBrB089H07 | |
| 21 | + | nd | Novel 1 | KBrB089H07, *At4g15475* | |
| 22 | + | 1 f | *UGT84A9b* | | *UGT84A9b* |
| 23 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 24 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 25 | + | 1 f | *UGT84A9b* | | *UGT84A9b* |
| 26 | + | nd[c] | *UGT84A9a* | | *UGT84A9a* |
| 27 | – | | | | |
| 28 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 29 | + | 2 p[c] | *UGT84A9d* | | *UGT84A9d* |
| 30 | + | nd | Novel 1 | KBrB089H07, *At4g15470* | |
| 31 | – | | | | |
| 32 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 33 | + | nd | Novel 2 | KBrB089H07, *At4g15890* | |
| 34 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 35 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 36 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 37 | + | nd | *UGT84A9a/d* | *UGT84A9a* | |
| 38 | + | 1 p[c] | *UGT84A9a* | | *UGT84A9a* |
| 39 | + | 2 f | *UGT84A9d* | | *UGT84A9d* |
| 40 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 41 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 42 | + | 2 p | *UGT84A9c* | *UGT84A9c* | |
| 43 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 44 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 45 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 46 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 47 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |

**Table 2** continued

| BAC no. | Hybridization with *UGT84A9-1* | Sequence type/length | Locus definition[a] | BAC end sequence | PCR product sequence[b] |
|---|---|---|---|---|---|
| 48 | – | | | | |
| 49 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 50 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 51 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 52 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 53 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |
| 54 | + | 1 f | *UGT84A9b* | | *UGT84A9b* |
| 55 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 56 | + | 1 f | *UGT84A9a* | | *UGT84A9a* |
| 57 | + | 2 f | *UGT84A9c* | | *UGT84A9c* |

*f* Full-length ORF, *p* partial ORF, *nd* not defined

[a] Hybridization patterns of individual *UGT84A9* genomic loci are shown in Fig. 3

[b] PCR products were generated covering the *UGT84A9* ORF from nucleotide 562 to 1094 and sequenced. Locus assignment was done by sequence comparison to BAC clones 1 (*UGT84A9a*), 4 (*UGT84A9b*), 6 (*UGT84A9c*) and 14 (*UGT84A9d*) for which the full-length ORFs had been sequenced

[c] Hybridization pattern reveals a full-length *UGT84A9* ORF

subjected to BAC end sequencing (Table 2). As a result, end sequences were produced for 11 BACs using T7 and SP6m primers. For classification, individual BAC end sequences were aligned with all four *UGT84A9* sequences described above or used for a BLAST search against the non-redundant nucleotide data base of GenBank. For the BAC clones 10 and 42 end sequences confirmed the pre-assignment by PCR and Southern hybridization to *UGT84A9a* and *UGT84A9c*, respectively. For BAC 18 pre-characterized as *UGT84A9a* by Southern hybridization, BAC end sequences confirmed the presence of a partial *UGT84A9a* ORF as well as for BAC 37 whose hybridization pattern was partly compatible to both *UGT84A9a* and *-d*. In the case of BAC 3, which produced a novel hybridization pattern (Novel 1), the BAC ends did not contain *UGT84A9* sequences. As closest homologue of the according SP6m BAC end sequence we found the genomic *Brassica rapa* BAC KBrB089H07 published by the Korea *Brassica* Genome Project (Yang et al. 2005). Likewise, the remaining BAC clones with novel hybridization patterns (BACs 2, 8, 20, 21, 30, 33) shared identity of end sequences with KBrB089H07, except BAC 2. Moreover, homologues of *A. thaliana* genes located near the hydroxycinnamate UGTs (HCA-GTs) *UGT84A1* (*At4g15480*), *UGT84A3* (*At4g15490*) and *UGT84A4* (*At4g15500*) were detected as closely related to these BAC ends. Sequence analysis of KBrB089H07 revealed the presence of a putative *B. rapa* hydroxycinnamate UGT (BrHCA-GT) by sequence identity of 85% on nucleotide level and 89% on amino acid level with *UGT84A1*. Since KBrB089H07 was found as closest homologue to end sequences of BACs producing either hybridization pattern Novel 1 (BACs 3, 21, 30) or Novel 2 (BACs 8, 20, 33), most likely the *B. napus* homologues of the putative BrHCA-GT from both the Brassica A- and C-genome were present in our BAC collection isolated by hybridization with *UGT84A9-1*.

To confirm the finding that the *B. napus* genome harbors four distinct *UGT84A9* loci, PCR products representing the ORF sub-sequence between nucleotides 562 and 1,094 were generated from each of the *UGT84A9*-related BAC clones. From the 42 BACs subjected to PCR, 39 amplicons could be derived. The BACs 10, 37 and 42 failed to produce a PCR product due to the presence of partial *UGT84A9* ORFs (Fig. 4; Table 2). The sequences of the 39 PCR products obtained were compared to the known sequences of BAC 1 (*UGT84A9a*), BAC 4 (*UGT84A9b*), BAC 6 (*UGT84A9c*) and BAC 14 (*UGT84A9d*). For each of the analyzed BACs, the previous classification by PCR and Southern blot analysis could be confirmed. No additional polymorphism was detected in this sequence set.

## Both *UGT84A9* sequence types are present in the diploid progenitors of *B. napus*

To trace back evolution of the four *UGT84A9* genomic loci, we analyzed homologous sequences from *B. rapa* and *B. oleracea*, the diploid progenitors of *B. napus*. PCR products covering the full-length coding region were amplified from both genomes and sequenced. Results from ten *B. rapa* sequences and nine *B. oleracea* sequences revealed two variants for each genome resembling the sequence types *UGT84A9-1* and *UGT84A9-2*. To elucidate evolutionary relationships, *UGT84A9*-related sequences from Brassica and Arabidopsis were used to construct an un-rooted

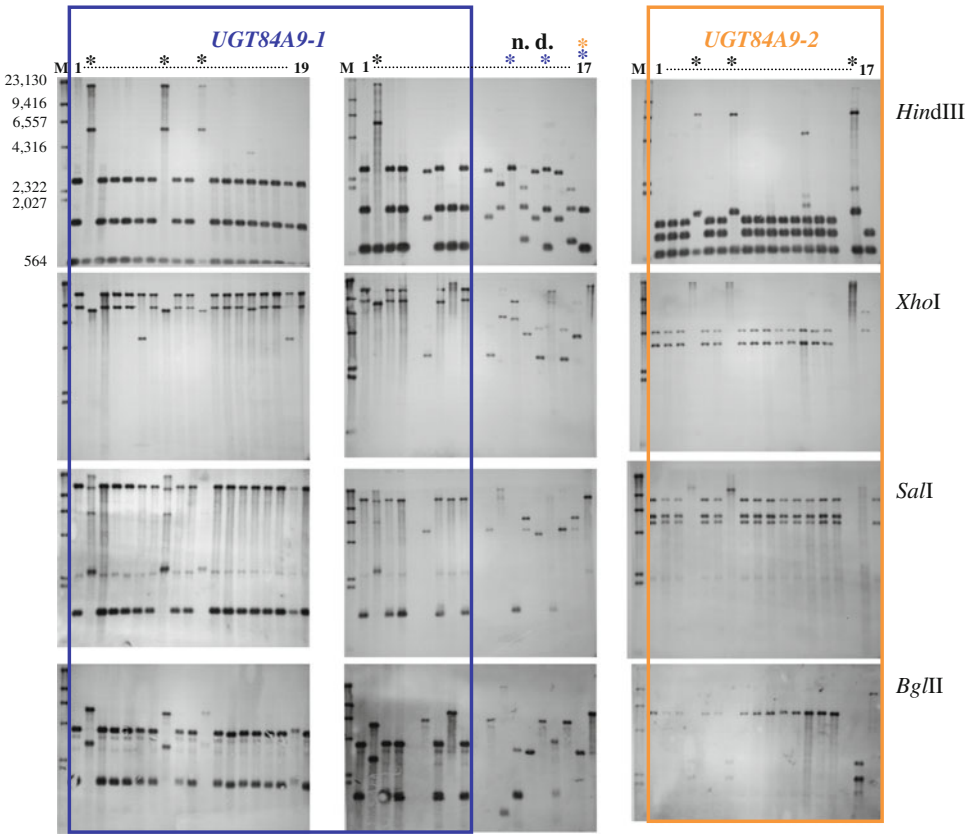


**Fig. 4** Southern blot analysis of 53 genomic BAC clones from *B. napus* carrying *UGT84A9*-related sequences. BAC-DNA was digested with restriction enzymes *Hin*dIII, *Xho*I, *Sal*I and *Bgl*II, and the fragments separated by agarose gel electrophoresis. Blotted membranes were hybridized with *UGT84A9-1* cDNA as the probe. Gels were loaded according to the sequence type characterization by PCR (Table 2). BACs carrying sequence type *UGT84A9-1* are marked by a *blue frame*; BACs related to sequence type *UGT84A9-2* are indicated by an *orange frame*. Within both sequence types, distinct gene loci are defined by different hybridization patterns; the less abundant patterns are labeled by *black asterisks*. Gene locus assignment was done as follows. *Lane 1* for sequence type *UGT84A9-1*: gene locus *UGT84A9a* (BAC 1); *lane 2* gene locus *UGT84A9b* (BAC 4). *Lane 1* for sequence type *UGT84A9-2*: gene locus *UGT84A9c* (BAC 6); *lane 4* gene locus *UGT84A9d* (BAC 14). BACs that escaped sequence type assignment by selective PCR are designated as non-defined (*n.d.*). Among the non-defined BACs, those revealing one of the distinct gene loci by restriction patterns are labeled with *asterisks* in the appropriate color. *Left panel lanes 1–19*, BACs 1; 4; 5; 9; 15; 17; 19; 22; 23; 24; 25; 28; 34; 35; 36; 41; 43; 47; 49. *Middle panel lanes 1–17*, BACs 51; 54; 55; 56; 3; 7; 10; 38; 2; 8; 18; 20; 21; 26; 30; 33; 37. *Right panel lanes 1–17*, BACs 6; 11; 12; 14; 16; 32; 39; 40; 44; 45; 46; 50; 52; 53; 57; 29; 42. For DIG-labeled DNA molecular weight marker (*M*), the fragment sizes are given in base pairs (bp)

neighbor-joining tree (Fig. 5). Tree topology confirmed that both *B. napus UGT84A9* sequence types, *UGT84A9-1* and *UGT84A9-2*, were present in *B. rapa* and *B. oleracea*. Within sequence type *UGT84A9-1*, the closest homologue of *UGT84A9a* was a *B. oleracea* sequence designated as *BoUGT84A9a*, whereas the *B. napus UGT84A9b* is closely related to the *B. rapa* sequence named *BrUGT84A9b*. For sequence type *UGT84A9-2* we could show that *B. napus UGT84A9c* clustered together with the *B. oleracea* variant *BoUGT84A9c*, whereas *B. napus UGT84A9d* grouped with the *B. rapa* sequence denoted as *BrUGT84A9d*. This clustering reveals that *UGT84A9a* and *-c* were both derived from the *Brassica* C-genome of *B. oleracea*, whereas *UGT84A9b* and *-d* originated from the Brassica A-genome of *B. rapa*. The not characterized putative hydroxycinnamate UGT from *B. rapa* BAC KBrB089H07 (BrHCA-GT) clustered with the functionally proven *UGT84A1* from Arabidopsis and with *UGT84A11*, which was isolated in a previous screen from *B. napus* cDNA (Mittasch et al. 2007). The *B. napus* gene *UGT84A10* identified by the same previous screen clustered as the homologue of Arabidopsis *UGT84A4*.

To provide sequence information for genetic mapping we sequenced several sub-fragments of representative BAC clones covering the genomic context upstream and downstream from the distinct ORFs (BAC 1, *UGT84A9a*; BAC 4, *UGT84A9b*; BAC 6, *UGT84A9c*; BAC 14, *UGT84A9d*). BAC sequences were used to develop molecular markers for each of the distinct *UGT84A9* gene loci. The gene loci *UGT84A9a* and *UGT84A9d* were both mapped in the population Express*1012-98. *UGT84A9a* was placed on linkage group C05 (N15) between the AFLP markers E33M62_188 and E32M48_288 (Badani et al. 2006), whereas *UGT84A19d* was mapped close to the SSR marker

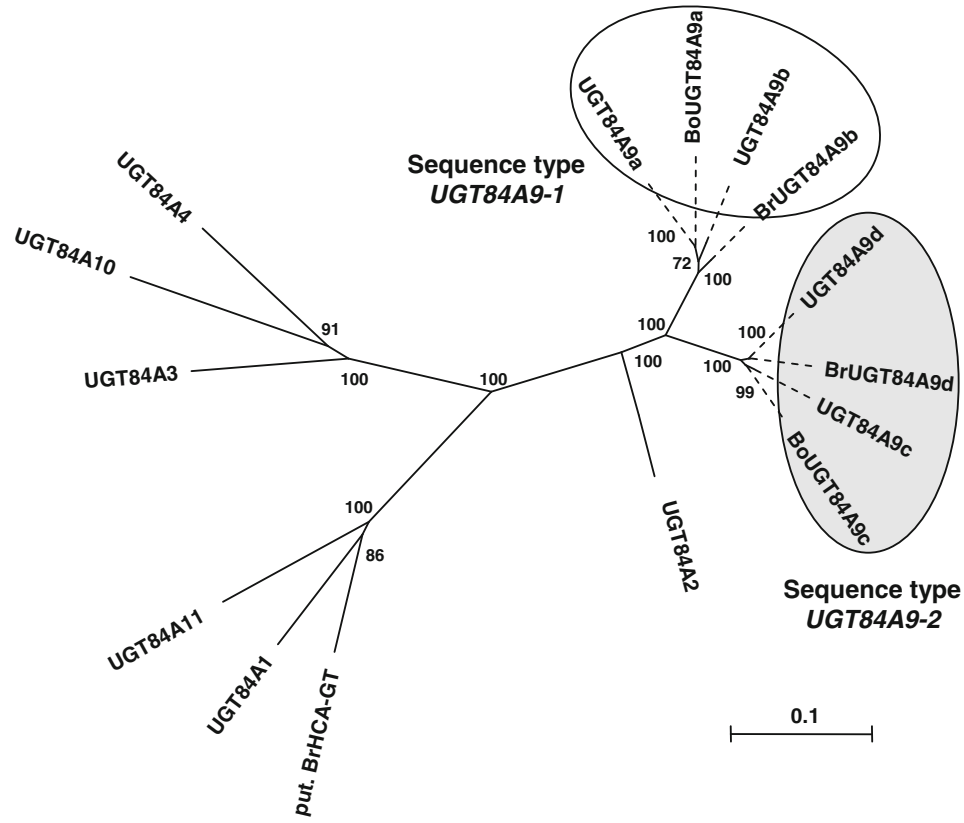


**Fig. 5** Phylogenetic relationships of *UGT84A* genes. The phylogenetic tree was derived by neighbor-joining algorithm from a multiple alignment of nucleotide sequences. The scale represents 0.1 fixed mutations per site. Clades comprising the different *UGT84A9* sequence types are highlighted. For clarity, dashed lines were added after tree construction. From Arabidopsis, the genes encoding the UDP-glucose:sinapate glucosyltransferase (SGT, UGT84A2, *At3g21560*) and unspecific ester forming HCA-GTs (UGT84A1, *At4g15480*; UGT84A3, *At4g15490*; UGT84A4, *At4g15500*) are shown. Sequences from *B. napus* include the genes identified in this work that encode SGT (*UGT84A9a*, GenBank Accession FM872276; *UGT84A9b*, FM872277; *UGT84A9c*, FM872278; *UGT84A9d*, FM872279) and the orthologs from *B. oleracea* (*BoUGT84A9a*, FM872280; *BoUGT84A9c*, FM872281) and from *B. rapa* (*BrUGT84A9b*, FM872282; *BrUGT84A9d*, FM872283). *UGT84A10* (AM231594) encodes an unspecific HCA-GT from *B. napus* whereas for the gene product of *B. napus UGT84A11* (AM231595) an enzymatic function could not be detected (Mittasch et al. 2007). The ORF designated as putative BrHCA-GT was found on the genomic BAC KBrB089H07 (AC189511) from *B. rapa*. Bo, *Brassica oleracea*; Br, *Brassica rapa*; HCA-GT, hydroxycinnamate glucosyltransferase; put., putative

CB10572 (Piquemal et al. 2005) on linkage group A01 (N01). In the population Mansholt*Samourai the gene locus *UGT84A9b* mapped on linkage group A05 (N05) close to the SSR marker CB10080 (Piquemal et al. 2005), whereas *UGT84A9c* could be mapped in the population Express*RS239 on linkage group C01(N11) close to the markers BRAS123 and CB10258 (Piquemal et al. 2005) (Fig. 6). The map positions confirmed the sequence-based assignments of *UGT84A9a* and *-c* to the Brassica C-genome of *B. oleracea* and of *UGT84A9b* and *-d* to the Brassica A-genome of *B. rapa*.

## The *UGT84A9* loci are localized to dynamic regions of the *B. napus* genome

To gain insight into the genomic fine structure of the four distinct *UGT84A9* loci from *B. napus* we used the sequence information of genomic BACs 1, 4, 6 and 14 for homology-based annotation approaches. Figure 7 shows schematic representations of a 7.7 kb fragment from BAC 1 harboring *UGT84A9a*, of 9.4 kb from BAC 4 for the *UGT84A9b* locus, of 7.5 kb from BAC 6 for *UGT84A9c* and of 20.3 kb from BAC 14 representing *UGT84A9d*. Detailed analyses revealed only remnants of micro-collinearity of the four *UGT84A9* loci with the homologous genomic region of Arabidopsis chromosome III, which carries the gene encoding the SGT homologue UGT84A2 (*At3g21560*). With exception of the *At3g21570* homologue located close to *UGT84A9d*, for none of the genes flanking *UGT84A2* in Arabidopsis the homologue was found in the genomic context of the distinct *UGT84A9* genes in *B. napus*. Instead, the genomic microstructures revealed the presence of several transposon-derived sequence elements at all four *UGT84A9* loci. On the *UGT84A9a*-related fragment, we detected a 1.2 kb sequence with homology to a portion of the non-coding A-region of the *B. oleracea* self incompatibility locus S7

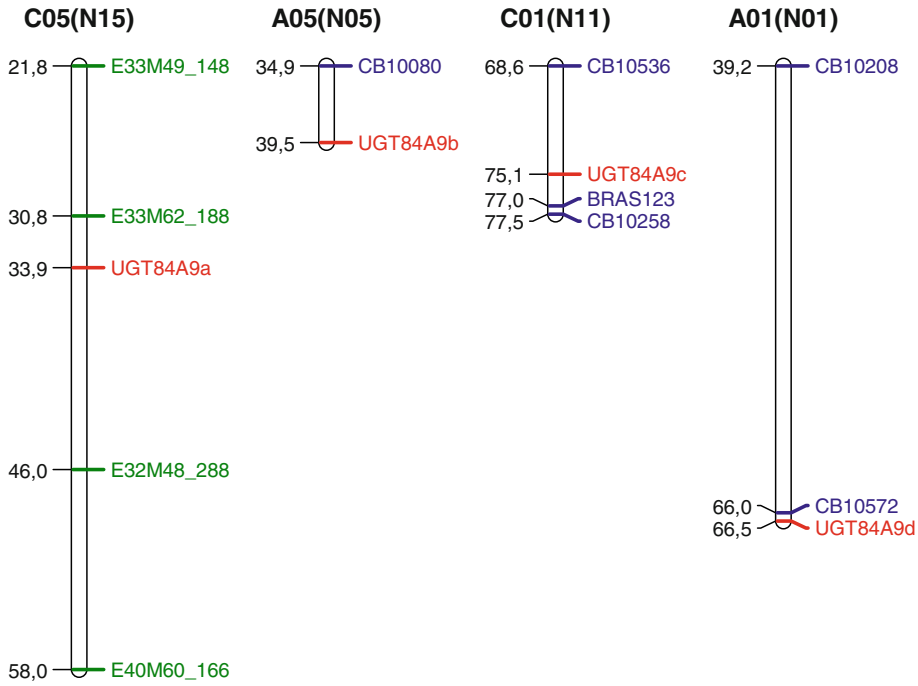


**Fig. 6** Genetic mapping positions of the four *UGT84A9* gene loci in *B. napus*. Distinct *UGT84A9* loci are shown in *red*. Closely linked public markers are highlighted in blue (Celera AgGen Brassica Consortia; Piquemal et al. 2005) and green (Badani et al. 2006). Linkage groups are given according to the A/C nomenclature as well as to the public N denomination (Parkin et al. 2005)

(Fujimoto et al. 2006). The non-coding B-region of this locus was shown to contain transposable elements. The sequence element shared about 60% of nucleotides with *At5g38383* from Arabidopsis, annotated as a member of the *gypsy*-like retrotransposon family Athila. *Gypsy*-like elements are the main type of plant long terminal repeat (LTR) retrotransposons, which belong to the class I of transposable elements and display high similarity to retroviruses (Sabot and Schulmann 2006). Upstream of *UGT84A9b*, we found an incomplete ORF homologous to *At3g15310* from Arabidopsis, annotated as transposable element (http://www.arabidopsis.org/). In the upstream region of *UGT84A9c*, an insertion sequence (IS) element was detected by overall sequence identity with five database entries from Arabidopsis (*At1g76035*, *At3g25185*, *At1g47875*, *At1g35666*, *At1g36930*). Downstream of *UGT84A9c*, three genomic regions with sequence identities to the Arabidopsis element ATLINE 1_4 (*At3g30745*) were identified. Long interspersed nucleotide elements (LINEs) are among the most abundant types of non-LTR retrotransposons and could be found in most eukaryotic genomes. The ATLINE 1_4 family belongs to the L1 familiy of LINEs, which is present in plants, vertebrates, slime molds and algal genomes (Malik et al. 1999). Close to *UGT84A9d*, two genomic regions with similarities to transposon-like sequences were identified. A sequence element upstream of *UGT84A9d* displayed identity to *At3g30418* described as member of the *gypsy*-like retrotransposon family Athila. Downstream of *UGT84A9d*, a second region was identified by similarity to four *CACTA*-like transposase genes from Arabidopsis. The 3′-ends of these genes (*At2g12990*, *At1g36270*, *At3g32226*, *At3g43126*) exhibit 60% sequence identity to the appropriate region of BAC 14. *CACTA* is a class II transposon located mainly in centromeric and peri-centromeric regions of the plant genome (Miura et al. 2004; Kwon et al. 2005), although some subfamily members were found to be spread over the whole genome and were used as segregating markers for genome mapping in maize and rice (Lee et al. 2006; Kwon et al. 2006). The insertion of transposon-derived elements could have expanded the genetic regions of the individual *UGT84A9* loci in *B. napus*. Therefore, to address the question of collinearity with the homologous *UGT84A2* locus of Arabidopsis deserves sequence analyses of larger genomic fragments from *B. napus*.

As a result, the high frequency of transposon-derived sequence elements indicate that the four *UGT84A9* loci of *B. napus* are localized to dynamic genome regions, which underlie increased reorganization. By comparative analyses with a summer type Canola line we produced evidence that the presence of four distinct *UGT84A9* loci is a common feature of *B. napus* cultivars (data not shown). However, with regard to genomic fine structure and nucleotide sequence, the results from cultivar ‘Express’ could not be directly transferred to other cultivars.

## Discussion

Given the importance of the glucosyltransferase UGT84A9 as the key enzyme for sinapate ester biosynthesis in *B. napus*, we characterized the encoding gene with respect to copy number, expression and microstructure of individual

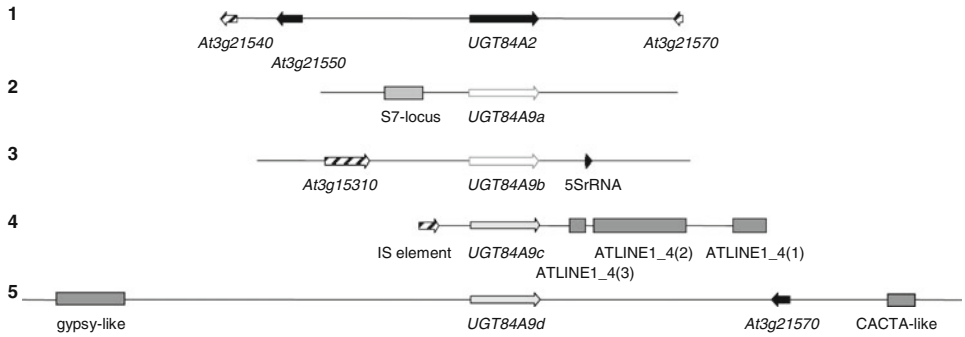


**Fig. 7** Microstructures of the *UGT84A9* genomic regions from *B. napus* compared to the homologous *UGT84A2* locus from Arabidopsis. Genes are shown as *arrows* pointing in the direction of transcription. *Hatched arrows* represent partial genes. The *UGT84A9* orthologs of sequence type 1 (*UGT84A9a*, *UGT84A9b*) are shown in *white*; those representing sequence type 2 (*UGT84A9c*, *UGT84A9d*) are marked in *light gray*. Transposon-derived elements are shown in *dark gray*. *1* 10 kb section of *Arabidopsis thaliana* chromosome III carrying *At3g21540* (nucleotides 1–362 complementary strand (C), incomplete ORF), *At3g21550* (1,214–1,768 C), *At3g21560* (*UGT84A2*, 5,390–6,880), *At3g21570* (10,000–9,809 C, incomplete ORF); *2* genomic locus *UGT84A9a* from *B. napus* BAC 1 (7.7 kb, GenBank Accession FM872284) carrying homologous region to S7-locus from *B. oleracea* (1,365–2,198, 91% identity), *UGT84A9a* (3,208–4,701); *3* genomic locus *UGT84A9b* from *B. napus* BAC4 (9.4 kb, FM872285), features incomplete ORF homolog to *At3g15310* (1,458–2,434, 72% identity), *UGT84A9b* (4,581–6,074), 5SrRNA homologous to *B. nigra* gene II 5SrRNA (7,099–7,225, 95% identity); *4* genomic locus *UGT84A9c* from *B. napus* BAC6 (7.5 kb, FM872286), features IS element (1–431, incomplete, 100%), *UGT84A9c* (1,118–2,611), ATLINE1_4(3) (3,258–3,601, C, 61% identity to *At3g30745*), ATLINE1_4(2) (3,777–5,779, C, 63% identity to *At3g30745*), ATLINE1_4(1) (6,792–7517, C, 58% identity to *At3g30745*); *5* genomic locus *UGT84A9d* from *B. napus* BAC14 (20.3 kb, FM872287) features *gypsy*-like region with homology to *At3g30418* annotated as *gypsy*-like retrotransposon (737–2,233, 53% identity), *UGT84A9d* (9,688–11,181), ORF homologous to *At3g21570* (16,225–16,641, C, 82% identity), CACTA-like region with homology to *At3g32226* (18,741–19,304, 60% identity)

genomic loci. As illustrated by a classical cytogenetic study (UN 1935), the major crop *B. napus* arose as a natural spontaneous allotetraploid (genome AACC, $2n = 38$) from the diploid species *Brassica rapa* (genome AA, $2n = 20$) and *Brassica oleracea* (genome CC, $2n = 18$). During evolution, after the split of the Arabidopsis and Brassica lineages about 20–24 million years ago (MYA), the latter exhibited a genome triplication (Jackson et al. 2000; Lagercrantz 1998; O'Neill and Bancroft 2000). The later process of diploidization was characterized by extensive chromosomal rearrangements and the loss of about 35% of the initial gene copies from the triplicated Brassica genome (Town et al. 2006). As a result, the Brassica genomes display a high rate of segmental chromosomal duplication, and the chromosomal collinearity between *Arabidopsis thaliana* and the closely related Brassica lineages is often obscured (Yang et al. 2006). The genomic organization of *UGT84A9* reflects the complex genome evolution in the Brassica lineage. From both diploid progenitor genomes of *B. napus* we identified a paralogous gene pair—*UGT84A9a/-c* from the *B. oleracea* C-genome and *UGT84A9b/-d* from the *B. rapa* A-genome. The paralogs shared about 88% sequence identity on nucleotide level, whereas the orthologs were nearly identical, thus forming *UGT84A9* sequence types 1 (*UGT84A9a/-b*) and 2 (*UGT84A9c/-d*). A phylogenetic analysis revealed that the two *UGT84A9* sequence types from Brassica were closer related to each other than to the Arabidopsis SGT homologue *UGT84A2*. This indicates that the *UGT84A9* sequence types appeared after separation of Brassica from the Arabidopsis branch.

Gene duplication is considered a major force in molecular evolution. Paralogs originating from duplication events may undergo different developments. They could gain new functional roles, contribute to gene redundancy or become silenced (Wendel 2000). During genomic reorganization duplicated genes might be lost without immediate negative effects on the organism. These developments were first described for animal and plant resistance genes, designated as birth-and-death mechanism (Nei et al. 1997; Michelmore and Meyers 1998). Genes evolving through birth-and-death mechanism exhibit higher sequence similarities between orthologs than between paralogs, as it was found in this work for the *UGT84A9* genes. Bioinformatics and gene expression studies on duplicated genes in Arabidopsis have shown that either one paralogous copy disappears or—in case of remaining—these copies most frequently acquire different functions (Blanc and Wolfe 2004). In some cases, the strongly decreased expression of one paralog in almost all organs or treatment conditions pointed to a "regulatory hypofunctionalization" (Duarte et al. 2006). For *UGT84A9*, our data on transcript abundance indicate such regulatory hypofunctionalization for the paralogous gene pairs *UGT84A9a/-c* and *UGT84A9b/-d* in *B. napus*. The two orthologs of sequence type 1 (*UGT84A9a/-b*) mainly contributed to *UGT84A9* gene expression in all observed tissues exhibiting SGT enzyme activity, whereas expression

of sequence type 2 representing *UGT84A9c*/*-d* could only be detected in flowers. Transcript quantification by sequence type-specific RT–PCR and cDNA-AFLP produced similar results, except for flowers. While RT–PCR showed a strong signal for both *UGT84A9* sequence types, cDNA-AFLP analysis revealed low transcript abundance for sequence type 2 orthologs in flowers. This discrepancy might be caused by unspecific primer binding in RT–PCR to a non-characterized abundant UGT-cDNA present in the flower sample. Since the cDNA-AFLP method was deliberately developed to provide independence from the amplification efficiency of individual primers (Bachem et al. 1996), the results of the latter should reflect more precisely the transcript abundance. A sequence comparison of *UGT84A9a* and *-b* with cDNA clones revealed that indeed both genes of sequence type 1 were expressed in developing seeds of *B. napus*. In accordance with our previous finding that SGT enzyme activity parallels transcript abundance (Milkowski et al. 2004), we conclude that in developing seeds of *B. napus* SGT activity is mainly provided by *UGT84A9* sequence type 1. Therefore, conventional breeding and TILLING approaches aimed at generating low-sinapine *B. napus* lines should focus on the genes *UGT84A9a* and *UGT84A9b*. From an evolutionary point of view, the maintenance of the weakly expressed sequence type 2 orthologs *UGT84A9c* and *-d* in the Brassica genomes might indicate that these gene copies are necessary for genetic robustness.

Sequence analyses of representative BAC regions unraveled the microstructure of the four *UGT84A9* genomic loci from *B. napus*, which was in each case different from the homologous portion of Arabidopsis chromosome III encoding *UGT84A2*. This reflects the unusually high rate of chromosomal rearrangements calculated for the Brassica lineages (6.5–9.7 rearrangements per million years according to Koch and Kiefer 2005). In contrast to *UGT84A9*, the genomic loci of *B. napus* encoding the final enzyme in sinapine biosynthesis, sinapoylglucose:choline sinapoyltransferase (BnSCT; Weier et al. 2008) displayed a significantly higher degree of micro-collinearity with the homologous region of Arabidopsis chromosome V, which carries the gene *SNG2* (*sinapoylglucose accumulator 2*; Shirley et al. 2001) encoding Arabidopsis SCT. On all four genomic *UGT84A9* loci of *B. napus* we detected transposon-like sequences or incomplete transposon-like elements. Transposable elements (TEs) are assumed to be a major driver of genome and gene evolution (for reviews, see Bennetzen 2005; Casacuberta and Santiago 2003; Fedoroff 2000; Kazazian 2004). In plants, differential amplification of TEs is associated with genome size variability (Hawkins et al. 2006). Moreover, large-scale genome rearrangements are frequently found near regions enriched for TEs (Eichler and Sankoff 2003). Accordingly, the presence of TE-like sequences on each of the four *UGT84A9* loci indicates that these loci form part of dynamic regions of the *B. napus* genome coined by frequent rearrangements. This is in agreement with preliminary analyses of other *B. napus* cultivars (J. Mittasch, personal communication), which produced evidence that the presence of four genomic *UGT84A9* loci might be conserved in *B. napus* whereas the fine structure of these loci may differ. With regard to function, TEs were shown to play a role in determining heterochromatin regions, in epigenetic gene regulation (Lippman et al. 2004), direct gene regulation (Kobayashi et al. 2004) and diversifying evolution of duplicated genes (Akhunov et al. 2007). In the case of the differential expression found for the *UGT84A9* loci, however, a possible influence of the closely related TE-derived sequences remains to be elucidated.

**Acknowledgments** The authors thank Rod Snowdon (University of Giessen, Germany) for providing the genomic BAC library from *B. napus*, for help with the library screen and for supplying the identified positive clones. BAC end sequencing by Prisca Viehöver and Bernd Weisshaar (University of Bielefeld, Germany) is greatly acknowledged. Seeds of *B. napus*, *B. oleracea* and *B. rapa* were kindly provided by Norddeutsche Pflanzenzucht (Holtsee, Germany). Excellent technical assistance was given by Sylvia Vetter, Claudia Horn, Anja Henning and Alexandra Jestadt. This work was part of the research project "*YelLowSin* Rapeseed: Functional genomics approaches for the development of yellow-seeded, low sinapine ("YelLowSin") oilseed rape/canola (*Brassica napus*)", financially supported by the Bundesministerium für Bildung und Forschung.

## References

Akhunov ED, Akhunova AR, Dvorak J (2007) Mechanism and rates of birth and death of dispersed duplicated genes during the evolution of a multigene family in diploid and tetraploid wheats. Mol Biol Evol 24:539–550

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

Bachem CWB, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RGF (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9(5):745–753

Badani AG, Snowdon RJ, Wittkop B, Lipsa FD, Baetzel R, Horn R, De Haro A, Font R, Lühs W, Friedt W (2006) Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). Genome 49:1499–1509

Baumert A, Milkowski C, Schmidt J, Nimtz M, Wray V, Strack D (2005) Formation of a complex pattern of sinapate esters in *Brassica napus* seeds, catalyzed by enzymes of a serine carboxypeptidase-like acyltransferase family. Phytochemistry 66:1334–1345

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Birnboim HC (1983) A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol 100:243–255

Blair R, Reichert RD (1984) Carbohydrate and phenolic constituents in a comprehensive range of rapeseed and canola fractions: nutritional significance for animals. J Sci Food Agric 35(1):29–35

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16:1679–1691

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1–11

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altmann N, de Pamphilis CW (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol 23:469–478

Eichler EE, Sankoff D (2003) Structural dynamics of eukaryotic chromosome evolution. Science 301:793–797

Fedoroff N (2000) Transposons and genome evolution in plants. Proc Natl Acad Sci USA 97:7002–7007

Fujimoto R, Okazaki K, Fukai E, Kusaba M, Nishio T (2006) Comparison of the genome structure of the self-incompatibility (S) locus in interspecific pairs of S haplotypes. Genetics 173:1157–1167

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

Hüsken A, Baumert A, Strack D, Becker HC, Moellers C, Milkowski C (2005) Reduction of sinapate ester content in transgenic oilseed rape (*Brassica napus*) by dsRNAi-based suppression of BnSGT1 gene expression. Mol Breed 16:127–138

Jackson SA, Cheng Z, Wang ML, Goodman HM, Jiang J (2000) Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. Genetics 156:833–838

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kazazian HH (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982

Koch MA, Kiefer M (2005) Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species—*Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. Am J Bot 92:761–767

Kwon SJ, Park KC, Kim JH, Lee JK, Kim NS (2005) Rim 2/Hipa CACTA transposon display; a new genetic marker technique in Oryza species. BMC Genet 6. doi:10.1186/1471-2156-6-15

Kwon SJ, Hong SW, Son JH, Lee JK, Cha YS, Eun MY, Kim NS (2006) CACTA and MITE transposon distributions on a genetic map of rice using $F_{15}$ RILs derived from Milyang 23 and Gihobyeo hybrids. Mol Cells 21:360–366

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

Lee JK, Park JY, Kim JH, Kwon SJ, Shin JH, Hong SK, Min HK, Kim NS (2006) Genetic mapping of the Isaac-CACTA transposon in maize. Theor Appl Genet 113:16–22

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471–476

Malik HS, Burke WD, Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805

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

Milkowski C, Baumert A, Strack D (2000) Cloning and heterologous expression of a rape cDNA encoding UDP-glucose:sinapate glucosyltransferase. Planta 211:883–886

Milkowski C, Baumert A, Schmidt D, Nehlin L, Strack D (2004) Molecular regulation of sinapate ester metabolism in *Brassica napus*: expression of genes, properties of the encoded proteins and correlation of enzyme activities with metabolite accumulation. Plant J 38:80–92

Mittasch J, Strack D, Milkowski C (2007) Secondary product glycosyltransferases in seeds of *Brassica napus*. Planta 225:515–522

Miura A, Kato M, Watanabe K, Kawabe A, Kotani H, Kakutani T (2004) Genomic localization of endogenous mobile CACTA family transposons in natural variants of *Arabidopsis thaliana*. Mol Genet Genomics 270:524–532

Nazck M, Amarowicz R, Sullivan A, Shahidi F (1998) Current research developments on polyphenolics of rapeseed/canola: a review. Food Chem 62(4):489–502

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci 94:7799–7806

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

Ohlson R (1978) Functional properties of rapeseed oil and protein product. In: Proceedings of the 5th international rapeseed congress, Malmö, Sweden, pp 152–167

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, Doucet I, Perret D, Villeger M-J, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor Appl Genet 111:1514–1523

Rahman MH (2001) Production of yellow-seeded *Brassica napus* through interspecific crosses. Plant Breed 120:463–472

Rawel H, Rohn S, Kroll J (2000) Reaction of selected secondary plant metabolites (glucosinolates and phenols) with food proteins and enzymes—influence on physicochemical protein properties, enzyme activity and proteolytic degradation. Recent research developments. Phytochemistry 4:115–142

Sabot F, Schulmann AH (2006) Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97:381–388

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning—a laboratory manual. Cold Spring Harbor Press, New York

Shirley AM, McMichael CM, Chapple CCS (2001) The *sng2* mutant of Arabidopsis is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. Plant J 28:83–94

Strack D (1981) Sinapine as a supply of choline for the biosynthesis of phosphatidylcholine in *Raphanus sativus* seedlings. Z Naturforsch 36c:215–221

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13:524–530

Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortmann JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, Vigouroux M, Trick

M, Bancroft I (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell 18(6):1348–1359

U N (1935) Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn J Bot 7:389–452

Uzunova M, Ecke W, Weissleder K, Röbbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90:194–204

Van de Peer Y, de Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 10:569–570

Velasco L, Möllers C (1998) Nondestructive assessment of sinapic acid esters in Brassica species: II. Evaluation of germplasm and identification of phenotypes with reduced levels. Crop Sci 38:1650–1654

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acid Res 23(21):4407–4414

Wang S, Oomah BD, McGregor DI, Downey RK (1998) Genetic and seasonal variation in the sinapine content of seed from Brassica and Sinapis species. Can J Plant Sci 78:395–400

Weier D, Mittasch J, Strack D, Milkowski C (2008) The genes *BnSCT1* and *BnSCT2* from *Brassica napus* encoding the final enzyme of sinapine biosynthesis: molecular characterization and suppression. Planta 227:375–385

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Yang TJ, Kim JS, Lim KB, Kwon SJ, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Kim SH, Lim YP, Park BS (2005) The Korea Brassica Genome Project: a glimpse of the Brassica genome based on comparative genome analysis with Arabidopsis. Comp Funct Genomics 6:138–146

Yang TJ, Kim JS, Kwon SJ, Lim KB, Choi BS, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Lim YP, Kang JJ, Hong JH, Kim CB, Bhak J, Bancroft I, Park BS (2006) Sequence-level analysis of the diploidization process in the triplicated *FLOWERING LOCUS C* region of *Brassica rapa*. Plant Cell 18:1339–1347

zum Felde T, Baumert A, Strack D, Becker HC, Moellers C (2007) Genetic variation for sinapate ester content in winter rapeseed (*Brassica napus* L.) and development of NIRS calibration equations. Plant Breed 126(3):291–296