# Hydroxycinnamate glucosyltransferase genes in *Brassica napus*, encoding key enzymes in sinapate ester metabolism.

**Dissertation**

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Naturwissenschaftlichen Fakultät I – Biowissenschaften
der Martin-Luther-Universität Halle-Wittenberg

von Juliane Mittasch

geb. am 03.12.1976 in Lauchhammer

Gutachter:

1. Prof. D. Strack
2. Prof. B. L. Møller
3. Prof. K. Humbeck

verteidigt am 04.12.2008 in Halle (Saale)

# Table of contents

# Table of figures

# List of tables

# Abbreviations

| | |
|---|---|
| A | adenine |
| AFLP | amplified restriction fragment length polymorphism |
| AP | alkaline phosphatase |
| APS | ammonium persulfate |
| At | *Arabidopsis thaliana* |
| BAC | bacterial artifical chromosome |
| BCIP | 5-bromo-4-chloro-3-indolyl phosphate |
| BLAST | basic local alignment search tool |
| Bn | *Brassica napus* |
| BnHCA-GT | *Brassica napus* hydroxycinnamate glucosyltransferase |
| BnSCT | *Brassica napus* 1-*O*-sinapoylglucose:choline sinapoyltransferase |
| Bo | *Brassica oleracea* |
| bp | base pairs |
| Br | *Brassica rapa* |
| BrHCA-GT | *Brassica rapa* hydroxycinnamate glucosyltransferase |
| BSA | bovine serum albumin |
| cDNA | complementary DNA |
| CoA | coenzyme A |
| CSPD | disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[1.3.1.1.$^{3,7}$]decan}-4-yl)pheryl phosphate |
| cv | cultivar |
| Da | dalton |
| dATP | deoxyadenosine triphosphate |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| dsRNAi | double-stranded ribonucleic acid interference |
| EDTA | ethylenediaminetetraacetic acid |
| EST | expressed sequence tag |
| g | gram |
| gDNA | genomic DNA |

| | |
|---|---|
| GT | glucosyltransferase |
| h | hour |
| HCA | hydroxycinnamate |
| HCA-GT | hydroxycinnamate glucosyltransferase |
| HPLC | high performance liquid chromatography |
| IgG | immunoglobulin G |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| kb | kilo base pairs |
| KBr | genomic *Brassica rapa* libraries |
| kDa | kilodalton |
| LB | Luria-Bertani medium |
| LTR | long terminal repeat |
| M | molar |
| MCS | multiple cloning site |
| MES | 2-(N-morpholino)ethanesulfonic acid |
| mg | milligram |
| min | minute |
| ml | milliliter |
| mM | millimolar |
| MOPS | 4-morpholinepropanesulfonic acid |
| mRNA | messenger ribonucleic acid |
| µg | microgram |
| µl | microliter |
| NBT | nitroblue tetrazolium |
| OD | optical density |
| ORF | open reading frame |
| PAGE | polyacrylamid gel electrophoresis |
| PCR | polymerase chain reaction |
| pI | isoelectric point |
| pkat | picokatal |
| put | putative |
| QTL | quantitative trait locus |

| | |
|---|---|
| RACE | rapid amplification of cDNA ends |
| RNA | ribonucleic acid |
| rpm | rotations per minute |
| RT-PCR | reverse transcriptase PCR |
| SAGE | serial analysis of gene expression |
| SAP | shrimp alkaline phosphatase |
| SCE | sinapine esterase |
| SCT | 1-*O*-sinapoylglucose:choline sinapoyltransferase |
| SDS | sodium dodecyl sulfate |
| sec | seconds |
| SGT | UDP-glucose:sinapate glucosyltransferase |
| siRNA | single stranded ribonucleic acid |
| SMT | 1-*O*-sinapoylglucose:malate sinapoyltransferase |
| T | thymine |
| TDF | transcript derived fragment |
| T-DNA | transferred DNA |
| TEMED | N,N,N',N'-tetramethylethylenediamine |
| TILLING | targeting induced local lesions in genomes |
| Tris | trishydroxymethylaminomethane |
| U | unit |
| UDP | uridine diphosphate |
| UGT | UDP-glycosyltransferase |
| UTR | untranslated region |
| v/v | volume per volume |
| w/v | weight per volume |

# 1 Introduction

## 1.1 Phenylpropanoid metabolism in Brassicaceae

The phenylpropanoid pathway is one of the major pathways in plant secondary metabolism, providing a wide variety of secondary products (Hahlbrock and Scheel, 1989; Dixon, 2001; Boerjan et al., 2003). Quantitatively the most important, lignin is crucial for the structural integrity of land plants. It also waterproofs cell walls and thus enables the transport of water and solutes through the vascular system. Furthermore, it plays a role in pathogen defense by providing a physical barrier (Boerjan et al. 2003). Flavonoids and other phenylpropanoids play important roles in plant defense or UV-protection, as signaling molecules or by ensuring reproduction as flower pigments or fragrances (Hahlbrock and Scheel, 1989; Bharti and Khurana, 1997; Dixon, 2001). The phenylpropanoid metabolism starts with aromatic amino acids, mainly phenylalanine, derived from the shikimate pathway (Ehlting et al., 2006). In Brassicaceae one well known end product of the phenylpropanoid metabolism in seeds is sinapine (Shahidi and Naczk, 1992), first detected in black mustard seeds (Henry et al., 1825). Beside their occurrence in Brassicaceae, sinapate esters were found e.g. in *Daucus carota* (Halaweish and Dougall, 1990), *Ananas comosus* (Fernandez de Simon et al., 1992), *Capsicum annuum*, and *Beta vulgaris* (Winter and Herrmann, 1986). The biosynthesis of these components was studied in detail in the model plant *Arabidopsis thaliana* (Milkowski et al., 2000b; Lim et al., 2001; Shirley et al., 2001; Humphreys and Chapple, 2002), a close relative to the crop species of the genus Brassica.

Chemically, the phenylpropanoid metabolism is defined by several hydroxylation and *O*-methylation reactions at the aromatic ring and by side chain alterations, in which *trans*-cinnamic acid derived from phenylalanine is converted into a variety of hydroxycinnamates. In the scheme presented in Figure 1-1 the emphasis lies on the formation of sinapine in *A. thaliana*, but major branching points like flavonoid or lignin production are indicated. Due to the relatively close relationship between Arabidopsis and Brassica, it is assumed that the phenylpropanoid metabolism of both genera is comparable. Cinnamate is first hydroxylated to 4-coumarate which is then activated as 4-coumaroyl-CoA (Hahlbrock and Scheel, 1989). This thioester is further converted to 4-coumaroylshikimate or 4-coumaroylchinate, followed by the hydroxylation of the shikimate or quinate ester to the corresponding caffeic acid conjugates (Schoch et al., 2001; Franke et al., 2002). Caffeoylshikimate/-quinate is then transacylated to caffeoyl-CoA, which is further methylated to feruloyl-CoA and then reduced to coniferylaldehyde (Lacombe et al., 1997). Subsequently, coniferylaldehyde is hydroxylated (Humphreys et al., 1999) and then methylated to sinapaldehyde. These three hydroxycinnamaldehydes are precursors for lignin biosynthesis, for which they are reduced to the corresponding alcohols and assembled to the different types of lignin (review: Humphreys and Chapple, 2002). Alternatively, sinapaldehyde and coniferylaldehyde can be oxidized to their corresponding hydroxycinnamates, sinapate or ferulate (Nair et al., 2004; review: Ehlting et al., 2006).

Sinapate is the starting point of the sinapate ester metabolism. In Brassica species sinapate is esterified to sinapoylglucose by the enzyme UDP-glucose:sinapate glucosyltransferase (EC 2.4.1.120) (Nurmann and Strack, 1980; Milkowski et al., 2000a). Thereby, an energetic bond is formed, which drives subsequent transesterfication reactions (Mock and Strack, 1993). Thus, sinapoylglucose is a precursor for many sinapate esters (Baumert et al., 2005), but in seeds it is mainly converted to sinapoylcholine (sinapine) by 1-*O*-sinapoylglucose:choline sinapoyltransferase (SCT, EC 2.3.1.91) (Strack et al., 1983; Shirley et al., 2001). During germination sinapine is cleaved to sinapate and choline by sinapine esterase (EC 3.1.1.49; Tzagoloff, 1963; Strack et al., 1980), which turned out to be a GDSL lipase-like protein (Clauß et al., 2008). Sinapate is re-esterified to sinapoylglucose which is converted to sinapoylmalate (Tkotz and Strack, 1980; Lorenzen et al., 1996).

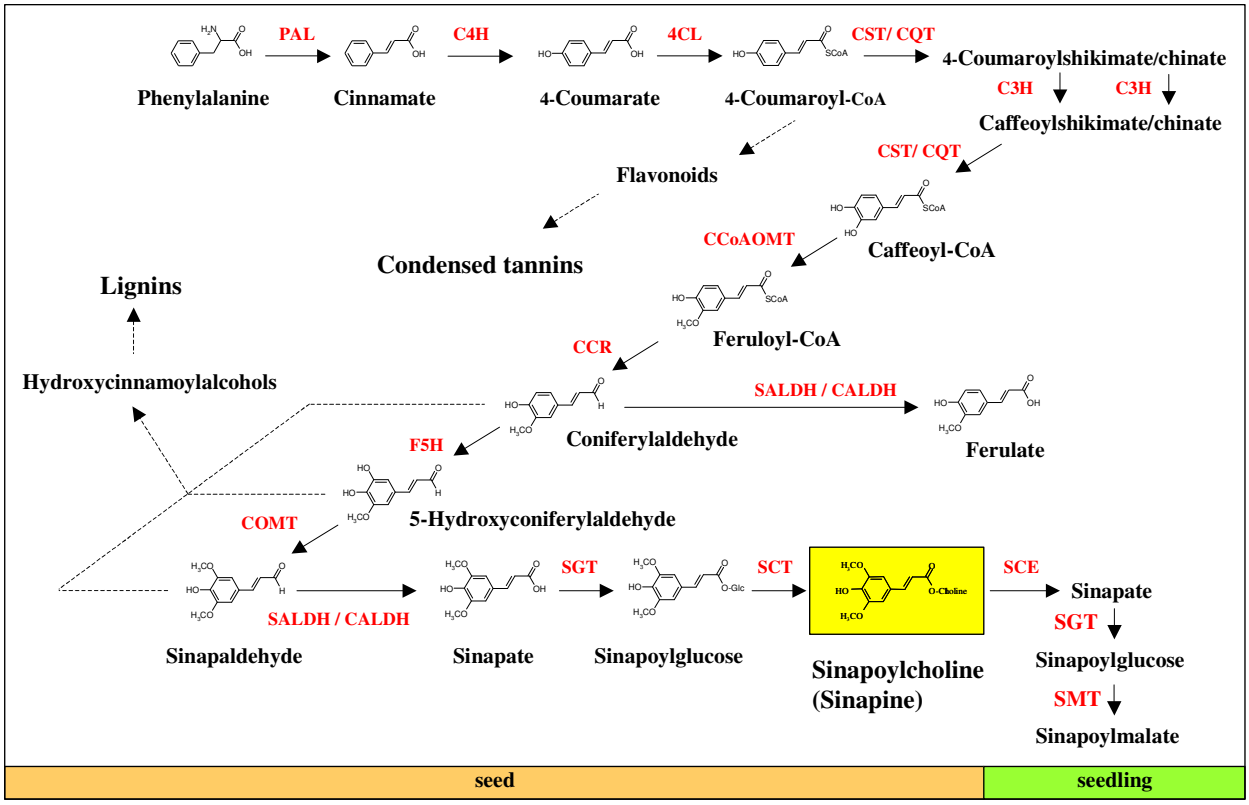



**Figure 1-1. Schematic overview of phenylpropanoid metabolism in *A. thaliana* emphasizing sinapine formation (Strack and Mock, 1993; Humphreys and Chapple, 2002, Nair et al., 2004; Ehlting et al., 2006). Enzymes are printed in red. PAL, phenylalanine ammonia lyase (EC 4.3.1.5); C4H, cinnamate 4-hydroxylase (EC 1.14.13.11); 4CL, 4-coumarate:CoA ligase (EC 6.2.1.12); CST, coumarate-CoA:shikimate coumaroyltransferase (EC 2.3.1.133); CQT, coumarate-CoA:quinate coumaroyltransferase (EC 2.3.1.99); C3H, coumaroyl shikimate/-quinate 3-hydroxylase (EC 1.14.13.36); CCoAOMT, caffeoyl-CoA *O*-methyltransferase (EC 2.1.1.104); CCR cinnamoyl-CoA reductase (EC 1.2.1.44); F5H ferulate 5-hydroxylase (monooxygenase, EC 1.14.13.x); COMT, 5-hydroxyconiferaldehyde *O*-methyltransferase (EC 2.1.1.68); SALDH/CALDH, sinapaldehyde/coniferaldehyde dehydrogenase (EC 1.2.1.68); SGT, UDP-glucose:sinapate glucosyltransferase (EC 2.4.1.120); SCT, 1-*O*-sinapoylglucose:choline sinapoyltransferase (EC 2.3.1.91); SCE, sinapine esterase (EC 3.1.1.49); SMT, 1-*O*-sinapoylglucose:malate sinapoyltransferase (EC 2.3.1.92)**

Sinapine is the most abundant phenolic ester in *B. napus* seeds (Bouchereau et al., 1991; Naczk et al., 1998). It contributes approximatly 1 to 2% to the seed dry mass (Fenwick, 1982; Shahidi and Naczk, 1992; Naczk et al., 1998), but the biological role of sinapine remains to be

elucidated. In *Raphanus sativa* it could be shown that the choline provided by sinapine hydrolysis during germination was consumed in the biosynthesis of phosphatidylcholine, which was later integrated into membranes (Strack, 1981). However, analysis of *A. thaliana* mutants impaired in sinapine biosynthesis revealed that the lack of sinapine was not a limiting factor in membrane biosynthesis (Chapple et al., 1992; Shirley et al., 2001). During early seedling development, the hydrolyzed sinapine provides sinapate for the production of the UV-protectant sinapoylmalate (Li et al., 1993; Landry et al., 1995; Sheahan, 1996).

An important part of the Brassica sinapate ester metabolism takes place in the seeds and seedlings. Seed development in angiosperms can be divided into three phases: embryogenesis, seed maturation and desiccation. During embryogenesis, the basic architecture of the embryo is established and the embryonic longitudinal axis is defined. The root apical meristem and the cotelydons are formed. Furthermore, the embryonic organ system is generated. Seed maturation is characterized by a dramatic increase in volume and mass of the developing seed. The seed accumulates storage proteins and lipids and therefore this phase is considered as seed filling phase. The only differentiation event is the formation of the shoot apical meristem from cell layers of the upper axis region between the two cotyledons. During the desiccation phase the cells of the embryo and the surrounding layers become dehydrated and the metabolic activity ceases. This ensures seed dormancy (Goldberg et al., 1989; Thomas, 1993; Goldberg et al., 1994).

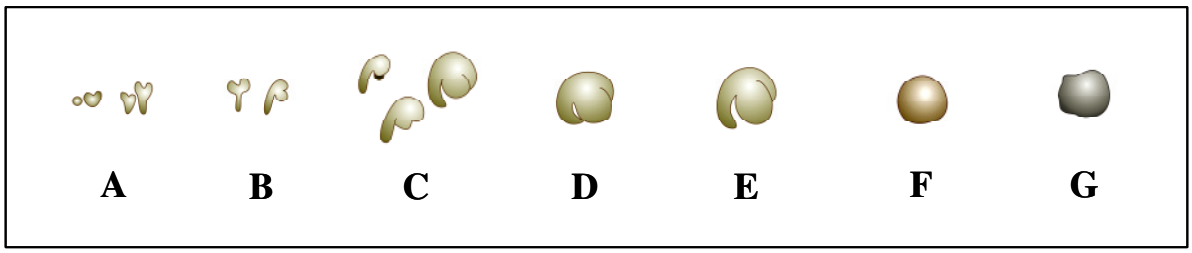



**Figure 1-2. Seed stages of *B. napus*. A-C: early embryogenesis, D and E: maturation, F and G: desiccation**

In *B. napus* seed development seven different stages can be distinguished morphologically (Figure 1-2; Milkowski et al., 2004). The developmental stages A to C describe the early embryogenesis, from globular, torpedo and late torpedo stages (stage A) to early cotyledon to "walking stick" stages (stage B) and the mid cotyledon stage (stage C), where the cotyledons expand in size and the maturation phase begins. The following phase is depicted as developmental stages D and E. In stage D, the cotyledons are well developed and the green embryo reaches its maximal size, but the seed coat is soft, whereas in stage E the desiccation process starts and the seed coat has become hard, but the embryo is still green. The desiccation process covers the two final developmental stages, F and G. Stage F includes seeds with brown seed coats and yellow embryos, while stage G defines the fully ripened, desiccated seeds with black seed coats.

The content of phenolic compounds in *B. napus* seeds increases during maturation (Naczk et al., 1998, Milkowski et al., 2004). Transcript levels and activities of the enzymes involved in sinapate ester metabolism are shown in Figure 1-3. In the early phases of seed development the transcript level of *BnSGT1* (*UGT84A9*) raises until it reaches a plateau at developmental stage C. The activity of the encoded enzyme increases to a maximum at developmental stage E and then decreases to base level activity for seeds. *BnSCT* is almost exclusively expressed

during the seed maturation phase. Transcript is first detectable at developmental stage C, the last stage of embryogenesis, rises to a maximum level in well developed embryos and vanishes thereafter. Enzyme activity follows this pattern. After germination a steep increase of SCE activity is followed immediately by an increase in transcript level and enzyme activity of SGT. This corresponds to the cleavage of sinapine after germination and the subsequent activation of sinapate by glucose ester formation. Starting at day two after sowing, SMT activity rises continuously until day 12 (Figure 1-3).

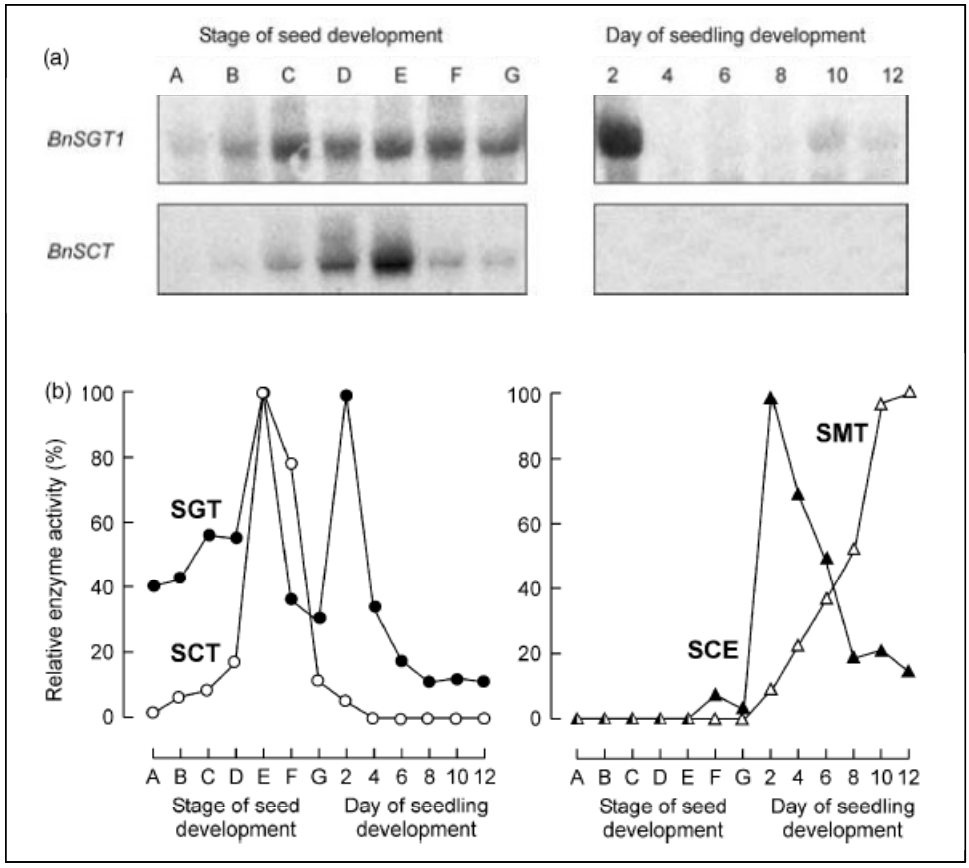



**Figure 1-3. Transcript levels (a) and enzyme activities (b) of enzymes of the sinapate ester metabolism in *B. napus* seeds and seedlings (Milkowski et al., 2004, modified)**

In *A. thaliana* the conversion of sinapoylglucose to sinapoylcholine or sinapoylmalate is catalyzed by unique enzymes, SCT and SMT, respectively, whereas the formation of sinapoylglucose can be catalyzed by four different glucosyltransferases (Milkowski et al., 2000b; Lim et al., 2001). So far however, in *B. napus* only *BnSGT1* (*UGT84A9*) was found (Milkowksi et al., 2000a). Given the genome complexity of *B. napus*, the presence of other hydroxycinnamate GTs that contribute to the formation of sinapoylglucose was hypothesized (Hüsken et al., 2005). Sinapoylglucose synthesis is a critical step in sinapate ester metabolism, and suppression of this step leads to the reduction of the majority of other sinapate ester compounds in seeds (Baumert et al., 2005; Hüsken et al., 2005). The nomenclature of HCA-GTs described from *A. thaliana* and *B. napus* is summarized in Table 1-1.

**Table 1-1. UDP-glucose:hydroxycinnamate glycosyltransferases in *Arabidopsis thaliana* and *Brassica napus*.**

| UGT nomenclature | Alternative nomenclature |
|---|---|
| *UGT84A1* | *At4g15480* |
| *UGT84A2* | *At3g21560, AtSGT1* |
| *UGT84A3* | *At4g15490* |
| *UGT84A4* | *At4g15500* |
| *UGT84A9* | *BnSGT1* |

(UGT nomenclature was assigned according to Mackenzie et al., 1997; Arabidopsis nomenclature according to TAIR database, SGT nomenclature according to Milkowski et al., 2000a and Milkowski et al., 2000b.)

## 1.2 Glycosyltransferases in plants

Glycosyltransferases form a highly divergent and polyphyletic multigene family. They catalyze glycosylation reactions, i.e. the transfer of an activated sugar moiety to a receptor molecule. This type of reaction is wide-spread among all living organisms and results in a great variety of products which serve many different functions (Mackenzie et al., 1997; Gachon et al., 2005). In 1997, a new classification of this enzyme superfamily was introduced. Based on sequences similarities of the proteins, 26 subfamilies were assigned (Campbell et al., 1997). Today 90 subfamilies are known and several sequences are still unaccounted for (www.cazy.org/fam/acc_GT.html). Two main catalytic mechanisms are conceivable for glycosyltransferases: the retention of the anomeric configuration of the sugar moiety – resulting in α-glucosides, or the inversion resulting in β-glucosides, catalyzed by β-UGT glycosyltransferases. The allocation into subfamilies consistently differentiates retaining from inverting enzymes (Campbell et al., 1997). The glycosyltranserases investigated in this work belong to family 1, that includes UDP-glucose-accepting glycosyltansferases of the inverting type: short UGTs (Mackenzie et al., 1997, Li et al., 2001, Paquette et al., 2003).

Plant secondary metabolism UGTs share a conserved amino acid sequence: the plant secondary product glycosyltransferase (PSPG) motif (Bairoch, 1991; Hughes and Hughes, 1994; Figure 1 4). Two peptide sequences of this motif, WAPQV and HCGWNS, can be found in 95% of all β-type glycosyltransferases (Vogt and Jones, 2000). Crystallographic data from bacterial enzymes indicate that the PSPG box could be the binding site of the nucleotide-activated sugar. A three-dimensional model of UGT73A5, a UGT from *Dorotheanthus bellidiformis* (Aizoaceae), suggests a direct interaction between UDP-glucose and the highly conserved HCGWNS motif of the PSPG box (Hans et al., 2004).

Nterm < **WAPQVEVLAHPAVGCFVTHCGWNSTLESISAGVPMVAWPFFADQ** > Cterm

**Figure 1-4. PSPG box sequence of plant secondary product glycosyltransferases. Highly conserved amino acids are shaded in blue (identity > 50%) or red (identity > 80%) (Vogt and Jones, 2000; modified).**

In the *A. thaliana* genome 120 putative UDP-glycosyltransferase-encoding genes were identified, of which 112 contain a full open reading frame (ORF) and eight were classified as

apparent pseudogenes (Ross et al., 2001; Paquette et al., 2003). These glycosyltransferases were grouped into 14 subfamilies, based on amino acid sequences (Li et al., 2001; Ross et al., 2001). One of these subfamilies, group L, contains enzymes that catalyze the formation of glucose esters (Li et al., 2001). Paquette and co-workers (2003) confirmed this classification but suggested to combine several closely related groups. They further proposed two major groups for PSPG-containing UDP-glycosyltransferase genes: one with a conserved intron A and a second that includes genes without primary introns. They speculated that both of these groups are monophyletic (Paquette et al., 2003).

## 1.3 Genomic organisation of the tribe Brassiceae

The Brassicacea family consists of 25 tribes, 338 genera and 3,709 species (Al-Shehbaz et al. 2006, Beilstein et al. 2006). In the tribe Brassiceae, three major species have diverged ca. 7.9 – 3.7 million years ago (MYA): *Brassica nigra*, *Brassica rapa* and *Brassica oleracea*. The first divergence occurred between the *rapa-oleracea* clade and *B. nigra* ca. 7.9 MYA, later *B. rapa* and *B. oleracea* separated ca. 3.7 MYA (Truco et al. 1996, Rana et al. 2004, Al-Shehbaz et al. 2006, Beilstein et al. 2006). Spontaneous polyploidizations of these diploid species resulted in three allotetraploid species: *Brassica juncea*, *Brassica carinata* and *Brassica napus*. These three species all exhibit diploid genetics (Figure 1-5; U, 1935; Parkin et al., 1995).

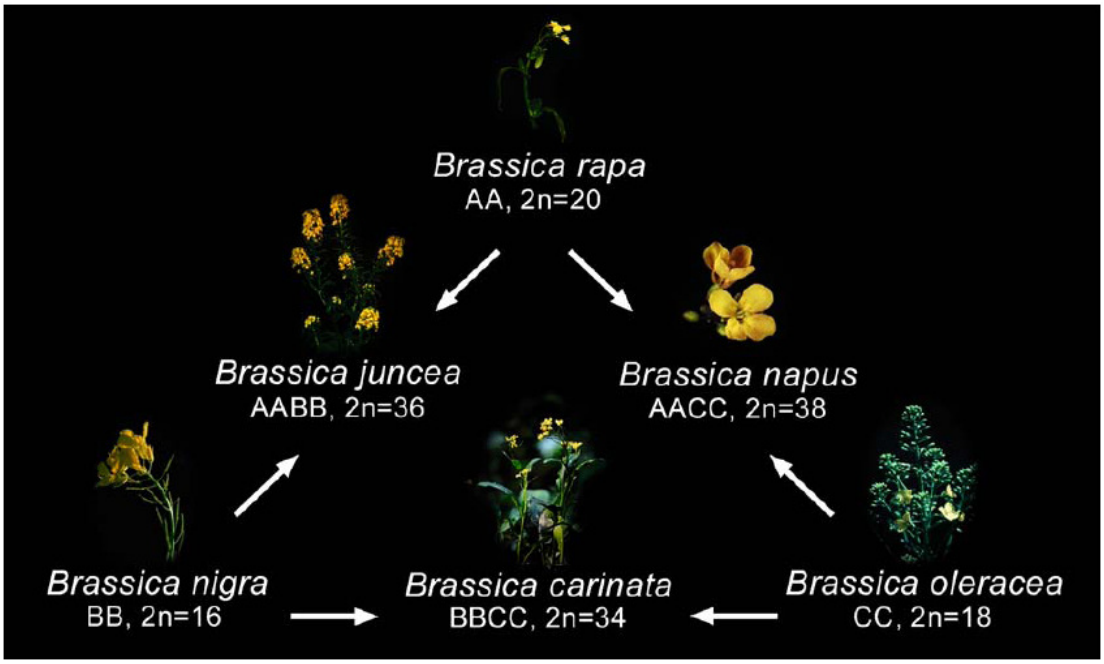



**Figure 1-5. Genomic composition of diploid and allotetraploid Brassica species (U, 1935, modified; Snowdon 2007).**

Oilseed rape (*B. napus*) is a relative young species. The oldest archeological records date to the 16th century and were documented for Northern Germany (Kroll, 1994). There are earlier references of "raepsaet" in the Netherlands (1360, 1421) and the lower Rhine region (1570), but it is not possible to differentiate between *B. napus* and *B. rapa* with certainty. Both

progenitor species are used since pre-Christian times in Europe and China (Körber-Grohne, 1995). Genetic studies estimate that *B. napus* emerged 0.01 MYA (Rana et al., 2004)

Many of our crop plants are polyploids (review: Wendel, 2000; Osborn et al., 2003). Autopolyploids like alfalfa and potato are plants whose diploid genome has been multiplied, whereas plants like wheat, cotton, coffee and oilseed rape are allopolyploids. Their genomes consist of at least two different ancestral genomes. Other species like maize or soybean show features of polyploidization in their ancestry but the actual evidence for this is lost due to genomic rearrangements. These plants are called paleopolyploids. It is thought that genome doubling or multiplication played a role in the evolution of most plant species. Polyploids often show traits that are not present in their progenitors or show a stronger occurrence of them. Some of the observed advantages are: increased drought tolerance, pest resistance, organ size and biomass, or earlier flowering time. Various changes in the number and expression of genes could be the cause of these novel traits. Allele-dosage effects are known in diploids as intermediate gene expression. In polyploids, because of the increased copy number of genes, the potential variations are higher and a finer adjustment of gene expression is possible. In other cases the expression level of genes is positively correlated with their copy number, which is enhanced in polyploids. Because of the doubling or merging of genomes the number and the type of transcription factors also changes. This could cause alterations in the regulatory network of gene expression. Furthermore, there are changes on the epigenetic level. Altered DNA methylation patterns, histone modifications or changes in DNA packing influence gene expression as well. Polyploidization also triggers rapid non-Mendelian genetic changes, like homeologous recombination or gain and loss of DNA fragments in individual genomes. Therefore, polyploidy is a driving force in plant evolution. Duplicated genes may undergo a functional diversification, because the selective pressure of one gene copy is relaxed. Alternatively, they can decay through mutations which lead to the silencing of one copy, or both copies can retain their function, thus allowing the previously described allele-dosage effect and contributing to the redundancy of the plant genome.

In the genus Brassica, the genomes of the three diploid species *B. nigra*, *B. oleracea* and *B. rapa* contain triplicated copies of an ancestral genome (Lagercrantz and Lydiate, 1996, Rana et al., 2004 Parkin et al., 2005, Town et al., 2006). Each of this triplicated genomes is structurally similar to the *A. thaliana* genome (Lagercrantz, 1998). Comparative mapping (Parkin et al., 2005) and analysis of the genomic fine structure (Rana et al., 2004; Town et al., 2006) of *B. napus* and *A. thaliana* revealed large regions of collinearity between both genomes, disrupted by several chromosome rearrangements. The observed collinearity can be used to navigate between *A. thaliana* and Brassica genomes, to annotate Brassica sequences to the corresponding *A. thaliana* regions and thereby to find new molecular markers for fine mapping approaches or marker-assisted plant breeding. Also, it is possible to infer the homeologous regions of the Brassica genomes from a candidate gene in *A. thaliana*, investigating beside the genes also their genomic surroundings. Furthermore, the frequent chromosome rearrangements, especially in resynthesized rapeseed, can be utilized for the introduction of novel genes from the progenitor species into the narrow gene pool of modern oilseed rape (Snowdon, 2007).

## 1.4 Oilseed rape as a crop plant

Oilseed rape (*B. napus*) is the second most important oil plant in the world. It accounts for approximately 10% of the global oilseed production, surpassed only by soybean which contributes 60%. In the last 15 years the cultivation area in Germany has doubled to 1.43 million hectares in 2005. Main cultivation areas are in Mecklenburg-Western Pomerania and Saxony-Anhalt. In 2005, Germany produced 40% of the rape yield in the EU (www.gmo-safety.eu).

A reason for the success of oilseed rape is the development of double zero lines (Roebbelen, 1999) that are low in erucic acid and glucosinolates. Erucic acid (C22:1) is synthesized from oleic acid via chain elongation. In the 1950s, feeding experiments with laboratory animals suggested, that the high erucic acid content of the rape seed oil leads to poor animal performance like reduced growth or undesireable physiological changes in organs. By 1963 a Canadian plant breeding program produced the first low erucic acid lines in spring cultivars of oilseed rape (Stefansson et al., 1961; Downey, 1964). The second problem were glucosinolates, which were found in rape seed meal and led to unfavorable cleavage products like isothiocyanate, thiocyanate and nitrile. This prevented the use of rapeseed protein for human nutrition and feeding of non-ruminant animals. In 1967, a Polish *B. napus* spring cultivar, “Bronowski”, was identified which has very low levels of glucosinolates (Josefsson and Appelqvist, 1968; Downey et al., 1969). Crossing between low erucic acid lines and Bronowski resulted in the first double low spring rape lines, “Tower” in Canada and “Erglu” in Germany (review: Downey, 1976).

The seeds of oilseed rape contain approximately 40% oil, the seed meal consists of 40-50% proteins (Fenwick, 1982). Although rape seed is mainly an oil crop, its meal is rich in essential amino acids, B-vitamins and minerals (Ohlson, 1978; Bell, 1993). However, high fiber content and phenolic compounds, like tannins and sinapine, decrease the nutritional value of the seed proteins and cause a bitter taste (Fenwick, 1982; Shahidi and Naczk, 1992; Naczk et al., 1998). The content of phenolic acids in rapeseed meal (defatted whole rapeseed) is up to five times higher than in soybean meal (Naczk et al., 1998). In the flours (defatted and dehulled seeds) of both species the differences goes up to 30-fold (Shahidi and Naczk, 1992; Naczk et al., 1998). It was suggested that phenolic acids can form insoluble complexes with rapeseed proteins (Shahidi and Naczk, 1992). The most abundant phenolic compound is sinapine, the ester of sinapate and choline (Bouchereau et al., 1991; Shahidi and Naczk, 1992, zum Felde et al., 2007). Sinapine is also the main cause of the “tainted egg” problem. When rape seed meal is included in the diet for laying hens producing brown eggs, a fishy taint occurs in many eggs. This taint is caused by trimethylamine, a degradation product of sinapine. Certain varieties of brow egg laying hens have a genetic deficiency that reduces the amount of the hepatic enzyme trimethylamine oxidase and thereby their capacity to oxidize and excrete trimethylamine in a normal way. Moreover, trimethylamine oxidase is inhibited by tannins and degradation products of glucosinolates (review: Fenwick and Curtis, 1980; Fenwick, 1982, Shahidi and Naczk, 1992, Naczk et al., 1998). Great efforts have been undertaken to clean the rape seed meal from these unwanted compounds. After harvesting, different processing procedures have been applied to reduce the amounts of unwanted compounds in rapeseed meal. Dehulling before and after oil extraction have been tested to remove fiber and tannins of the seed coat from the meal, but no procedure has been found that

appeals to the industry (Downey and Bell, 1990). Rapeseed meal and protein concentrates have been extracted with several mixtures of solvents like water, ethanol, methanol, ammonia and hexane. Seven extractions with 70% ethanol were needed to reduce the phenolic compounds to trace amounts in rapeseed meal (review: Shahidi and Naczk, 1992). Although processing has no negative influence on protein quality, it is too expensive for broader use (Downey and Bell, 1990). Another approach to reduce phenolic compounds could be conventional breeding. It was shown that the low sinapine trait is heritable (Bouchereau et al., 1991), but it depends also very much on environmental conditions (zum Felde et al., 2006). Although several promising lines could be identified (zum Felde et al., 2007), there is no successful breeding program for low sinapine rapeseed until now. Since the biosynthesis of sinapate esters in rapeseed is known, biotechnological approaches can also be used. First experiments look promising, showing reductions of sinapine content by up to 40% (Nair et al., 2000). A first dsRNAi approach silencing *UGT84A9* (*BnSGT1*), a glucosyltransferase of the sinapate ester metabolism, reduced total sinapate content to 24% in the best performing lines (Hüsken et al., 2005).

## 1.5 Aim of this work

Previous studies suggest that the sinapate ester accumuluation in *B. napus* seeds can be controlled by engineering the formation of sinapoylglucose, the metabolic precursor of the sinapate esters (Hüsken et al., 2005; Baumert et al., 2005). This requires a detailed knowledge on the genes encoding enzymes involved in the formation of sinapoylglucose, regarding their existence, copy number and gene expression.

In *A. thaliana* four glucosyltransferases were found, which can convert sinapate into sinapoylglucose (Milkowski et al., 2000b; Lim et al., 2001), whereas in *B. napus* only *UGT84A9* (*BnSGT1*) was described (Milkowski et al., 2000a). Hence, additional HCA-GTs from *B. napus* should be cloned and characterized by transcript analysis and recombinant protein expression. Their contribution to the formation of sinapoylglucose in seeds should be evaluated.

A first dsRNAi approach to silence *UGT84A9* reduced the total sinapate ester content in *B. napus* seeds to up to 24 % (Hüsken et al., 2005). The T-DNA region of the used dsRNAi construct is depicted in Figure 1-6. However, questions remain with regard to the stability of the suppression in the following transgenic plant generations and the sinapate ester metabolism of the seedlings. To answer these questions, transcript levels of *UGT84A9* in transgenic *B. napus* plants (T5 generation) and transgenic *B. napus* seedlings (T6 generation) should be investigated. To further reduce sinapate ester levels in *B. napus* seeds, a dsRNAi vector targeting *UGT84A9* and *BnSCT1* simultaneously should be constructed and the resulting transgenic plants should be characterized.

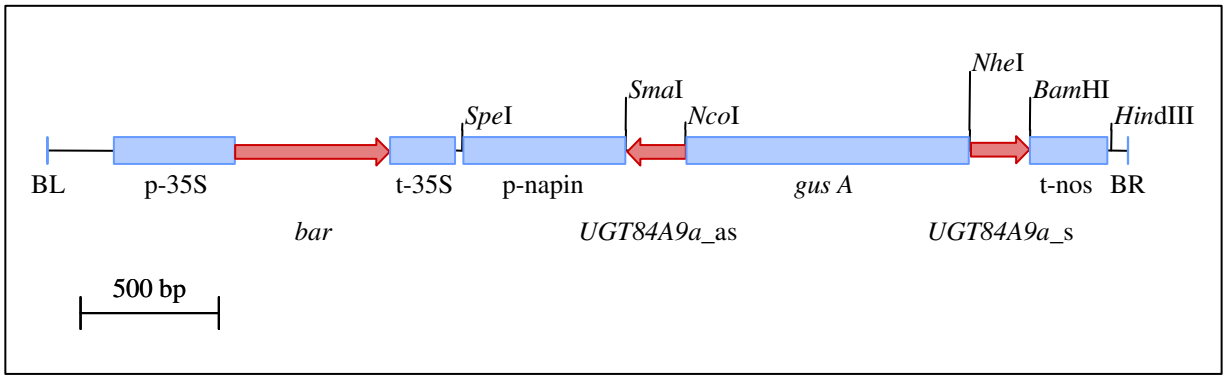



**Figure 1-6. Schematic overview of the *UGT84A9a* dsRNAi construct (Hüsken et al., 2005, modified). TheT-DNA region of the binary vector is depicted. Shown vector elements: left border (LB) and right border (RB) of the T-DNA, CaMV 35S promoter (p-35S), *bar* gene encoding phosphinotricine acetyl transferase (*bar*), CaMV 35S terminator (t-35S). Seed specific suppression cassette consists of napin promoter (p-napin), a 213 bp subsequence of *UGT84A9a* in antisense (as) and sense (s) orientation, a *'gus A* gene fragment (*gus A*) and the nos terminator (t-nos). Recognition sites used for constructuion are shown.**

Comparative studies of Arabidopisis and Brassica genome organization revealed the occurrence of a triplication event in the Brassica genomes after the separation from a common ancestor. Thus, for a singular gene in *A. thaliana*, like *UGT84A2*, two to six copies of the orthologous genes could be expected in *B. napus*. Therefore, the exact copy number of *UGT84A9* should be obtained and gene expression patterns should be evaluated especially regarding transcription during seed development. Furthermore, the genomic structure of the *UGT84A9* loci should be investigated and compared to the homologous genomic region of *A. thaliana* to assess the level of collinearity between these loci. Results from these experiments could be taken as input for TILLING (Targeting Induced Local Lesions IN Genomes; McCallum et al., 2000) of *UGT84A9*; a non-transgenic approach to reduce sinapate esters in *B. napus* seeds.

# 2 Material and methods

## 2.1 Material

### 2.1.1 Brassica

Clonings and gene expression studies of oilseed rape genes were carried out in the winter cultivar *Brassica napus* L. var. *napus* cv. Express. For transgenic approaches two spring cultivars were used; *Brassica napus* L. cv. Lisora and *Brassica napus* L. var. *napus* cv. Drakkar. The characterization of the progenitor genomes was conducted in *Brassica rapa* L. *sylvestris* cv. Rex and *Brassica oleracea* L. var. *medullosa* cv. Markola. For comparison of the genomic UGT84A9 loci, the yellow-seeded *B. napus* cultivar YN01-429 was investigated. All seeds were provided by Norddeutsche Pflanzenzucht Hans-Georg Lembke KG (NPZ Lembke, Hohenlieth, Germany), except the seeds of the cultivar YN01-429, which were contributed by Dr. Gerhard Rakow (Agriculture and AgriFood Canada, Saskatoon, SK, Canada).

### *2.1.2 Escherichia coli*

For all cloning techniques the *Escherichia coli* strains XL1 Blue (Stratagene, La Jolla, CA, USA) and Top10 (Invitrogen, Karlsruhe, Germany) were used:

XL1Blue: *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F'*pro*AB *lacI*$^{q}$ZΔM15 Tn10 (Tet$^{r}$)]

Top10: F$^{-}$*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*Z ΔM15 Δ*lac*X17 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str$^{r}$) *end*A1 *nup*G

Heterologous expression of recombinant proteins was carried out in *Escherichia coli* BL21-CodonPlus (DE3)-RP cells (Stratagene).

BL21-CodonPlus (DE3)-RP: B F$^{-}$ *omp*T *hsd*S($r_B-$ $m_B-$) *dcm*+ Tet$^{r}$ *gal* λ (DE3) *end*A Hte [*arg*U *proL* Cam$^{r}$]

### *2.1.3 Agrobacterium tumefaciens*

For the transformation of *Brassica napus* the *Agrobacterium tumefaciens* strain C58C1 (Hellens et al., 2000) with plasmid pMP90 (Koncz and Schell, 1986) was used:

C58C1pMP90: Rif$^{r}$, Gm$^{r}$, pTiC58ΔT-DNA

### 2.1.4 Oligonucleotides

Oligonucleotides were synthesized by MWG Biotech (Ebersberg, Germany). Primers shown in Table 2-1 were used in PCR, RACE-PCR and Genome Walking approaches for the cloning of *UGT84A9* related genes and for *UGT84A9* sequences.

**Table 2-1. Oligonucleotides used as primer in PCR, RACE-PCR and Genome Walker experiments for cloning of *UGT84A9* related genes and *UGT84A9* sequences.**

| Identifier | Sequence |
|---|---|
| FL_UGT84A9fw | ATG GAA CTA TCA TCT TCT CCT TTA CCT CCT CAT G |
| FL_UGT84A9rev | TTA TGA CTT TTC CAA TAA AAG TTC TTG AAC ACT TCC GTT TT |
| UGT84A10rev | CCC ACC GTA GCC TCA GCC TCC TTC TTC C |
| FL_UGT84A10fw | ATG GAA TCG TCG CTG ACT C |
| FL_UGT84A10rev | TCA ACA GAT GTT GGC AAC CAA |
| Gen-FL_UGT84A10fw | ATG GAA TCG TCG CTG ACT CAT GTG ATG CT |
| Gen-FL_UGT84A10rev | TCA ACA GAT GTT GGC AAC CAA CTT ATC CAC AA |
| UGT84A11fw | AGA CAA GCC AAG CAG ATC GTC GAA GGT GAG |
| UGT84A11fw-nested | ACG ACG AGA TCG GGA CGG TTG CGT ATT TG |
| UGT84A11rev | GGC AGG ACA TGA GTC TCC AGC TTC AGA T |
| FL_UGT84A11fw | ATG GAG TTG GAA TCT TCT TCA CAT TCG AGT CCA GTT C |
| FL_UGT84A11rv | TCA GCC ACT GAT CAC GCC TAA CTT CTC CA |

Primers given in Table 2-2 were used in EST cloning from flowers and seeds. RT-Homerace is an adapted poly-dT-oligo used for reverse transcription of mRNA. It provides a 5'overhang that increases the melting temperature ($T_m$) thereby allowing higher annealing temperatures in the following PCRs for EST cloning. Both GT-Short primers are degenerate primers, derived from the highly conserved amino acid sequence HCGWN , which is a part of the PSPG box of plant secondary metabolism glycosyltransferases (Vogts and Jones, 2000).

**Table 2-2. Primers used for EST-cloning.**

| Identifier | Sequence |
|---|---|
| RT-Homerace | CTA TCT CGG TCT AAT TTT TTT TTT TTT TT |
| GT-Short1 | CAY TGY GGN TGG AAT |
| GT-Short2 | CAY TGY GGN TGG AAC |

Primer for the determination of transcript levels in semi-quantitative RT-PCRs are given in Tables 2-3 and 2-4. Table 2-3 lists primer for the characterization of BnGT-related ESTs and BnHCA-GTs. For this experiment, primers derived from a certain EST were compared against all other EST sequences to prevent unwanted binding and false signals. The primers to detect genomic DNA contamination (Genomic DNA control fw/-rv) were derived from the 5' upstream region of *UGT84A9a*.

**Table 2-3. Primers for semiquantitative RT-PCR of *B. napus* GT-like seed EST, BnHCA-GTs and control genes. [1]RT-UGT84A9 primer did not match exactly to the rarely expressed locus *UGT84A9d*.**

| Identifier | Sequence |
|---|---|
| BnGT1 fw | TCG GTG CTG GAG AGC TTG TG |
| BnGT1 rv | ACG ACG ATC CTC CGT CCT TG |
| BnGT2 fw | GGC GTT CTG ATG CTG ACG TG |
| BnGT2 rv | ACG CTG CTT CCC TCA GCT TC |
| BnGT3 fw | GGG TTG CCT ATG GTA ACT TG |
| BnGT3 rv | CGC CTT TCT TCT GCT TCT TC |
| BnGT4 fw | AAT GAT GGC CGA TCA ACC |
| BnGT4 rv | CCG TCT CCT TCC ATT AGT TC |
| BnGT5 fw | AAG CTC GCC GTC TAC GGT TC |
| BnGT5 rv | GTG TTC CGG TGG TGG CAT TC |
| BnGT6 fw | TGC GTT TCG GCG TCC CAA TC |
| BnGT 6rv | TCC TCG GCA CAT CTT CAC CCT C |
| BnGT 7fw | TGG GAT GCC AGC TTA TAG |
| BnGT7 rv | CGT CTC CTC CAT AAC TTC TC |
| BnGT8 fw | AAC CTC TGG TCT ACC ACT AC |
| BnGT8 rv | CCA CCA ACA CTC CTA TCT TC |
| BnGT9 fw | AGT GGA CCG ATC AAC CTA TG |
| BnGT9 rv | TGA CTT GAC TGC CAA GTC TC |
| BnGT10 fw | GCT GGA ACT CGC TTC TTG |
| BnGT10 rv | GCT TCT TCG CCC GTT TAC |
| BnGT11 fw | CTC AGC GAC GAA CTT GGA ATC |
| BnGT11 rv | ACG ACA TCT CTG CCT TGT CTC |
| BnGT 12 fw | CCT GAC GCA GAG TCT AAG |
| BnGT12 rv | GAG AGG ACC AAG GCT ATG |
| RT-UGT84A9 fw[1] | CGC CAA CAA GAT TCA AGA CC |

| | |
|---|---|
| RT-UGT84A9 rv[1] | AAG GGA TCT CGT CGT GTT TC |
| RT-UGT84A10 fw | GAT ATC AAG GGA GAC ATC TCT GAC TC |
| RT-UGT84A10 rv | GTA GCC TCA GCC TCC TTC TTC C |
| RT-UGT84A11 fw | GTACATGCCACGGCTAGAACAA |
| RT-UGT84A11 rv | TCCGGTGAATGGAGAAGAAGGA |
| Napin fw | GCA CAA CAC CTA AGA GCT TG |
| Napin fw | GCA GCT GCT GTC CCT GCT GT |
| Ubiquitin fw | AGG CCA AGA TCC AGG ACA AAG |
| Ubiquitin rv | CGA GCC AAA GCC ATC AAA GAC |
| Genomic DNA control fw | TTC TAA GGT CTC GTG GGC TCA GTC AAC |
| Genomic DNA control rv | TAA CGT GGC CTT GTC CTG GGA AGG ATA C |

**Table 2-3 continued.**

Primers for further characterization of *UGT84A9* are given in Table 2-4. They were used to determine transcript levels in transgenic plants or to investigate the expression of different loci of *UGT84A9*. Primer pairs applied for distinction between different loci were tested on plasmids carrying cloned and characterized cDNAs as templates. Control primers such as napin, ubiquitin and genomic DNA were given in Table 2-3.

**Table 2-4. Primers for semi-quantitative RT-PCR and further characterization of *UGT84A9* loci. [1]UGT84A9fw7 and UGT84A9rev7 recognize all four loci of *UGT84A9*.**

| Identifier | Sequence |
|---|---|
| UGT84A9fw7[1] | CAC GAC GAG ATC CCT TCT TTC |
| UGT84A9rev7[1] | GTC ACG AAA CAA ACC ACA GAA G |
| UGT84A9-1fw | TGC AAG CAG AAC CAA CTT AAC CAT CCT |
| UGT84A9-1rev | CTC CAT GCA GTC ATC CCT CGT CTC A |
| UGT84A9-2fw | CGT AAG AAG ACA CGA TTT CAC CAT CTA |
| UGT84A9-2rev | TTC CAT ACA ATG GTC CGT GGT CTC G |

For confirmation of the semi-quantitave RT-PCR results, cDNA-AFLP was carried out. Primers and oligonucleotide adapters needed for this experiment are given in Table 2-5. Primers were designated according to the Keygene (Wageningen, Netherlands) nomenclature, thus, H stands for *Hin*dIII recognition site primer, M stands for *Mse*I recognition site primer and the numbers code for the specific nucleotides at the 3' end of the primer.

**Table 2-5. Primers and oligonucleotide adapters for cDNA-AFLP experiments. Primers were designated according to the Keygene (Wageningen, Netherlands) nomenclature. Primers marked with an asterisk were 5' fluorescence labeled.**

| Identifier | Sequence |
|---|---|
| *Hin*dIII-Adapter | CTC GTA GAC TGC GTA CC |
| *Hin*dIII-Adapter (second strand) | AGC TGG TAC GCA GTC TAC |
| *Mse*I-Adapter | GAC GAT GAG TCC TGA G |
| *Mse*I-Adapter (second strand) | TAC TCA GGA CTC AT |
| H03 | GAC TGC GTA CTA GCT TG |
| H04 | GAC TGC GTA CTA GCT TT |
| H22* | GAC TGC GTA CTA GCT TGT |
| H76* | GAC TGC GTA CTA GCT TGT C |
| H90* | GAC TGC GTA CTA GCT TTG T |
| M01 | GAT GAG TCC TGA GTA AA |
| M02 | GAT GAG TCC TGA GTA AC |
| M03 | GAT GAG TCC TGA GTA AG |
| M36 | GAT GAG TCC TGA GTA AAC C |
| M40 | GAT GAG TCC TGA GTA AAG C |
| M47 | GAT GAG TCC TGA GTA ACA A |
| M64 | GAT GAG TCC TGA GTA AGA C |

Primers used for the construction of protein expression vectors or gene-specific silencing cassettes of BnHCA-GTs and *BnSCT1* are given in Table 2-6. Restriction sites were added to the primers as indicated to allow further cloning of the PCR products.

**Table 2-6. Primers used for design of expression- and dsRNAi-constructs.**

| Identifier | Sequence |
|---|---|
| *Nco*I-UGT84A10fw | GTA CCA TGG AAT CGT CGC TGA CTC |
| *Sal*I-UGT84A10rev | GTA GTC GAC ACA GAT GTT GGC AAC CAA |
| *Nco*I-UGT84A11fw | TGA CCA TGG AGT TGG AAT CTT CTT CAC ATT CGA GTC CAG TTC |
| *Sal*I-UGT84A11rev | GTA GTC GAC GCC ACT GAT CAC GCC TAA CTT CTC CA |
| *Nhe*I-UGT84A9i_ffw | GTA GCT AGC AGT CAC CGA CGC TGC TTA CA |
| *Nco*I-UGT84A9i_rfw | GTA CCA TGG AGT CAC CGA CGC TGC TTA CA |
| *Sph*I-UGT84A9i_frev | GTA GCA TGC TTC CGC CTC CTC CTT CCA TT |

| | |
|---|---|
| *Bgl*II-UGT84A9i_rrev | GTA AGA TCT TTC CGC CTC CTC CTT CCA TT |
| *Sph*I-BnSCTi_ffw | ATA GTA GCA TGC GGA GTC GAT GCA ACA CTC AA |
| *Bgl*II-BnSCTi_rfw | ATA GTA AGA TCT GGA GTC GAT GCA ACA CTC AA |
| *Bam*HI-BnSCTi_frev | ATA GTA GGA TCC CCA AGG TCT CCA GTC ATC AA |
| *Sma*I-BnSCTi_rrev | ATA GTA CCC GGG CCA AGG TCT CCA GTC ATC AA |

**Table 2-6 continued.**

### 2.1.5 Plasmids

Plasmids used in this work are summarized in Table 2-7. Their main features and references are described.

**Table 2-7. List of plasmids**

| PLASMID | CHARACTERISTICS | REFERENCE |
|---|---|---|
| pGEM-T® Easy | 3 kb multi-copy vector, enables A/T cloning of PCR products, ampicillin resistance, allows blue/white screening | Promega, Mannheim, Germany |
| pCR®-BluntII-TOPO® | 3.5 kb multi-copy vector for cloning blunt end PCR products, kanamycin resistance | Invitrogen, Karlsruhe, Germany |
| pET28a(+) | 5.3 kb vector for expression of recombinant proteins in *E. coli*, kanamycin resistance, gene expression inducible by IPTG | Novagen, Darmstadt, Germany |
| pBNN | 3.8 kb construction vector for expression cassettes originated from pBlueScript, harbors the seed-specific napin promoter, a nos terminator and a multiple cloning site (MCS) between both elements, ampicillin resistance | AG Prof. E. Heinz, Uni Hamburg<br>Stratagene, LaJolla, CA, USA<br>Kridl et al., 1991<br>Chen et al., 2003a |
| pBNNGUS | 4.8 kb construction vector for dsRNAi vectors, β-glucuronidase (GUS) gene frament containing nucleotides 787-1,809 from *E. coli* was inserted into the MCS of pBNN, ampicillin resistance | Dr. C. Milkowski, IPB Halle<br>Chuang and Meyerowitz, 2000 |
| pLH7000 | binary cloning vector, confers resistance against spectinomycin (bacteria) and phosphinothricine (plants) | Hausmann and Töpfer, 1999 |

### 2.1.6 Chemicals, enzymes and antibodies

All chemicals were obtained from Roth (Karlsruhe, Germany), Sigma-Aldrich (Munich, Germany) and Merck (Darmstadt, Germany) unless stated otherwise. Enzymes for DNA modifications were purchased from Roche (Mannheim, Germany), New England Biolabs

(Frankfurt a. Main, Germany), Promega, Qiagen (Hilden, Germany) and Invitrogen. All solutions were prepared with double distilled water from a Milli-Q-facility (Millipore, Schwalbach, Germany).

For Western blot analyses a monoclonal Anti-His-Tag antibody from mouse (Novagen) was used as a primary antibody. Protein was detected using a secondary antibody against mouse IgG (immunoglobulin G) coupled with alkaline phosphatase (Sigma-Aldrich).

In Southern blot analyses a polyclonal antibody from sheep recognizing digoxigenin was used (Roche). This antibody was also linked to alkaline phosphatase.

### 2.1.7 Equipment

| | |
|---|---|
| Centrifuges : | Centrifuge 5417R (Eppendorf, Hamburg, Germany) |
| | Centrifuge 5810R (Eppendorf) |
| | Mikro 20 (Hettich, Tuttlingen, Germany) |
| | Universal 30 RF (Hettich) |
| Thermal cycler: | Eppendorf MasterCycler Gradient (Eppendorf) |
| Heaters: | Eppendorf comfort (Eppendorf) |
| | Eppendorf 5436 (Eppendorf) |
| | DRI-BLOCK® DB*3 (Techne, Staffordshire, UK) |
| Incubator: | GFL 3032 (GFL, Leipzig, Germany) |
| | CERTOMAT® BS-1 (Vitaris, Baar, ZG, Schweiz) |
| | WTC Binder (Binder, Tuttlingen, Germany) |
| Electrophoresis: | Electrophoresis system MINI (Roth) |
| | Electrophoresis system VARIA 1 (Roth) |
| | Mini-PROTEAN II (BioRad, Munich, Germany) |
| | Power Supply EPS 301 (GE Healthcare, Munich, Germany) |
| | BioDocAnalyze (Whatmann Biometra, Goettingen, Germany) |
| Spectrophotometer: | Beckmann DU 530 (Beckmann Coulter, Krefeld, Germany) |
| Ultrasonic disruptor: | UW 2070 (Bandelin, Berlin, Germany) |
| Bead mill: | Bead Beater (Biospec. Products, Bartesville, OK, USA) |
| HPLC: | Waters™ Millipore System (Waters, Eschborn, Germany) |
| Scales: | BL 1500 S (Sartorius, Goettingen, Germany) |
| | CP225D (Sartorius) |

| | |
|---|---|
| Electroporator: | Micropulser$^{TM}$ (BioRad) |
| X-ray film processor: | Optimax TR (MS Laborgeraete, Wiesloch, Germany) |
| UV crosslinker: | UV Stratalinker$^{TM}$ 1800 (Stratagene) |
| Hybridization incubators: | HB-2D (Techne) |
| | Hybrigene (Techne) |
| | GFL 7601 (GFL) |

## 2.2 Cultivation and transformation of bacteria

### *2.2.1 Escherichia coli*

*E. coli* was grown on LB medium using agar plate or suspension cultures at 37°C unless indicated otherwise. For selection of transformants carrying resistance conveying plasmids the corresponding antibiotic was added. Final concentrations for different antibiotics were: 50 or 100 $\mu g \cdot ml^{-1}$ carbenicillin, 50 $\mu g \cdot ml^{-1}$ chloramphenicol, 50 $\mu g \cdot ml^{-1}$ kanamycin, 100 $\mu g \cdot ml^{-1}$ spectinomycin and 12.5 $\mu g \cdot ml^{-1}$ tetracyclin.

For the preparation of competent cells, bacteria were grown to the mid-log phase ($OD_{600}$ = 0.6). To achieve electrocompetence, cells were washed several times in ice-cold water, then resuspended in a 10% ice-cold glycerol solution, aliquoted and stored at –80°C. Transformation was conducted using a Micropulser$^{TM}$ electroporator (BioRad) according to the manufacturers protocol. Chemical competence was produced by softening the cell walls by two ionic glycerol solutions. The first solution contained 30 mM potassium acetate, 100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride and 15% glycerol. The second solution contained 10 mM MOPS (4-morpholinepropanesulfonic acid), 75 mM calcium chloride, 10 mM rubidium chloride and 15% glycerol and served also the storage solution for the competent bacteria. For transformation 50 µl of competent cells were thawed on ice, then the DNA (1pg – 100ng) was added. The cells were incubated 30 min on ice, followed by a heat shock for 30 s at 42°C and a final incubation for 2 min on ice. Subsequently, 450 µl of liquid LB medium was added, and the bacteria were incubated at 37°C for 1 hour with gentle shaking. Thereafter, cells were plated on selective LB agar plates. Selected clones were preserved in a 7% dimethyl sulfoxide solution (DMSO; Sigma-Aldrich) at –80°C.

### *2.2.2 Agrobacterium tumefaciens*

*A. tumefaciens* was grown on LB media at 28°C. Preparation of compentent cells and transformation were done according to the *E. coli* protocols. After electroporation, bacteria were incubated at least 2 h at 28°C in a shaking incubator before plating on selective media. Transformant colonies appeared after two to three days. Strains were preserved in a 7% DMSO solution (Sigma-Aldrich) at –80°C.

## 2.3 Cultivation and transformation of Brassica

All Brassica species were cultivated in the greenhouse. Seeds were sown on a soil-vermiculite mixture (1:1), and the seedlings were covered by plastic hoods for the first ten days. Greenhouse conditions were as follows: temperature during daytime 16-19°C, temperature during night 14-17°C, moisture 40-70% and 16 h day - 8 h night timecycle. Winter cultivars were vernalized at 4°C for eight weeks after reaching the five to seven leaves stage. Transformation of *B. napus* var. Drakkar was done by Alexandra Hüsken at the Institute of Agronomy and Plant Breeding at the Georg-August-Universtity in Goettingen (Hüsken et al., 2005). Transformation of *B. napus* var. Lisora was done in Deutsche Saatenunion-Resistenzlabor by Dr. Orsini using a modified protocol of De Block and co-workers (De Block et al., 1989). Transgene copy number was determined in T1 plants by Southern blot analysis. For the production of further generations, plants were propagated by selfing.

## 2.4 Isolation and separation of nucleic acids

### 2.4.1 RNA isolation

RNA from plant tissue was isolated by selective binding to a silica-based matrix using the RNeasy Plant Mini Kit from Qiagen. Applied extraction buffer depended on the plant tissue (Table 2-8). RLT, containing guanidine thiocyanate, was the standard lysis buffer. If the analysed tissue had high amounts of secondary metabolites, buffer RLC, which contains guanidine hydrochloride instead, was used as recommended.

**Table 2-8. Extraction buffer for RNA isolation for different plant tissues from *B. napus*.**

| Plant tissue | Extraction buffer |
|---|---|
| seedlings younger than 3 days | RLC |
| seedlings older than 3 days | RLT |
| mature leaves | RLT (or RLC) |
| buds | RLT |
| flowers | RLT |
| seeds stage A – D | RLT |
| seeds stage E – G | RLC |

Tissue was disrupted either by a chilled glass homogenizer (VWR, Darmstadt, Germany) or by grinding under liquid nitrogen with a mortar and pestle. Following manufacturer's protocol, genomic DNA contamination was prevented by DNase I (Qiagen) on-column-digestion. Messenger RNAs (mRNAs) were enriched by binding to a poly-dT-matrix using the Oligotex mRNA Midi Kit from Qiagen. RNA concentration was measured by UV

absorption at 260 nm and purity was evaluated by the absorption ration $A_{260}/A_{280}$. The integrity of RNA was analyzed by gel electrophoresis on 2% agarose gels. Isolated RNA was stored at –80°C.

### 2.4.2 Isolation of genomic DNA

Genomic DNA (gDNA) from plant tissue was isolated by selective binding to a silica-based matrix using the DNeasy Plant Mini Kit for PCR applications or DNeasy Plant Maxi Kit for Southern blot analysis (Qiagen). For the DNeasy Plant Maxi DNA isolation, the manufacturer's protocol was modified by extenting the lysis and following precipitation step from 10 min to 30 min each. Each preparation was eluted twice with a volume of 100 µl or 1 ml elution buffer (EB), respectively. DNA concentration was measured by by UV absorption at 260 nm and purity was evaluated by the absorption ratio $A_{260}/A_{280}$.The integrity of gDNA was analyzed by gel electrophoresis on 0.7% agarose gels. Isolated genomic DNA was stored at 4°C.

### 2.4.3 Plasmid isolation

Plasmid DNA from *E. coli* was isolated from 2 ml overnight culture by selective binding to a silica-based matrix using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. Plasmid DNA was stored at 4°C as working solution or at –20°C as a stock.

Plasmid DNA from *A. tumefaciens* was isolated accordingly, except for using 4 ml of overnight culture and performing a prolonged cell lysis (30 min) at 37°C with 100 $\mu g \cdot ml^{-1}$ lysozyme added to the lysis buffer. Due to the low yield, 1 µl of the obtained plasmid DNA was transformed into *E. coli* for amplifiation and analyses were conducted with the multiplied plasmid DNA.

### 2.4.4 Isolation of BAC DNA

BAC DNA from 3 ml to 40 ml overnight cultures was isolated with the BACMAX DNA purification kit [Epicentre (Madison, WI, USA)].

For larger amounts of BAC DNA, an alkaline plasmid DNA extraction protocol was used (Birnboim, 1983). 300 ml overnight culture was centrifuged and the pellet was resuspended in 60 ml buffer containing 50 mM Tris-HCl at pH 8.0, 10 mM EDTA, 0.02 $mg \cdot ml^{-1}$ RNase A (Qiagen) .Cells were lyses by adding an equal volume of an alkaline solution (0.2 M NaOH, 1% SDS). Proteins, chromosomal DNA and other cell remnants were precipitated by adding 60 ml of a high salt acidic buffer (2,55 M potassium acetate pH 4.8). The mixture was centrifuged and the supernatant was cleared by filtration. Nucleic acids were ethanol precipitated. To remove remaining protein contamination, proteinase K was added to a final concentration of 50 $\mu g \cdot ml^{-1}$ and incubated at 37°C for 15 min, followed by a phenol/chloroform extraction. Remaining RNA was precipitated by addition of 3 ml LiCl-MOPS solution (5 M lithium chloride, 50 mM MOPS). BAC DNA was purified again by

ethanol precipitation and resuspended in 3 ml TE buffer (10 mM Tris-HCl at pH 7.5, 1 mM EDTA). Integrity was checked by gel electrophoresis on a 0.8% agarose gel.

### 2.4.5 Separation of nucleic acids by agarose gel electrophoresis

Nucleic acids were separated by electrophoresis in agarose gels with 0.5x TBE buffer (0.45 M Tris-HCl, 0,45 M boric acid, 1 mM EDTA). For the separation of DNA fragments different agarose concentration in gels were used depending on the molecule size of the fragments (Table 2-8).

**Table 2-9. Separation of nucleic acids by different agarose concentrations.**

| Agarose concentration | Separated nucleic acids |
|---|---|
| 0.7% | 0.8 – 12 kb (DNA) |
| 1% | 0.5 – 10 kb (DNA) |
| 1.5% | 0.2 – 3 kb (DNA) |
| 2% | 0.1 – 2 kb (DNA) and RNA |

Ethidium bromide was added to the molten agarose to a final concentration of 0.1 $\mu g \cdot ml^{-1}$, to visualize the nucleic acids under UV light. An adequate volume of 10x loading buffer [20% (v/v) glycerol, 0.1 M EDTA, 1% SDS, 0.2% (w/v) bromphenol blue, 0.2% (w/v) Xylene Cyanol] was added to the samples before loading on gel. Gels were run at approximately 10 $V \cdot cm^{-1}$. For size determination of the analyzed nucleic acid fragments, a mixture of DNA fragment with defined sizes (SmartLadder, Eurogentec, Cologne, Germany) was utilized.

### 2.4.6 Purification of DNA from agarose gels

DNA fragments intended for further use were cut out of the ethidium bromide-stained agarose gels under low dosage UV light. Fragments shorter than 4 kb were purified by the MinElute Gel Extraction kit (Qiagen), otherwise the QIAquick Gel Extraction kit (Qiagen) was used. In both cases the instructions of the manufacturer's protocol were followed. DNA concentration of the eluate was estimated by agarose gel electrophorese.

### 2.4.7 Purification of nucleic acids from solutions

PCR reactions, which resulted in a single, distinct product, were purified using the QIAquick® PCR Purification Kit (Qiagen). This column-format purification on a silica-based matrix removed enzymes, salts, remaining nucleotides and primers.

Protein contamination of nucleic acids was removed by phenol/chloroform extraction, followed by ethanol precipitation (Sambrook et al., 1989).

### 2.4.8 Determination of nucleic acid concentration and purity

Concentration of nucleic acid solutions was determined by measuring the absorption (A) at a wavelength of 260 nm. Purity of nucleic acids was estimated by the ratio of $A_{260}$ to $A_{280}$. This ratio should be between 1.8 and 2.0.

## 2.5 Enzymatic modification of nucleic acids

All molecular techniques were carried out according to standard laboratory protocols (Sambrook et al., 1989).

### 2.5.1 Restriction

Restriction analysis of plasmid DNA was conducted using 0.2-0.5 µg DNA and 5 U restriction endonuclease in a volume of 10 µl for a single digest or 20 µl for a double digest. Restriction enzymes were obtained from New England Biolabs and Roche. Buffers provided by Roche were used whenever possible. Reactions were incubated for at least 60 min at the appropriate working temperature for each enzyme used. Results were examined by agarose gel electrophoresis (chapter 2.4.5). For preparative restriction reactions the amount of plasmid DNA was increased to up to 2 µg. For restriction of genomic DNA for Southern blot analysis, 5-10 µg DNA was used and the incubation time was prolonged to overnight.

### 2.5.2 Polymerase chain reaction (PCR)

PCRs for standard applications were carried out using Taq PCR Master Mix Kit (Qiagen). Denaturation was done at 95°C for 30 sec, time was shortened for genomic DNA templates to 5 sec, annealing was conducted for 30 sec at 5°C below the melting temperature of the primers, unless stated otherwise. Elongation was performed at 72°C for 1 $min \cdot kb^{-1}$ of the amplicon. PCRs for cloning application were conducted with proofreading enzymes, using either ProofStart DNA polymerase (Qiagen) producing blunt end products or Platinum® PCR SuperMix High Fidelity (Invitrogen) attaching 3'-A overhangs. For genomic DNA templates, Advantage® Genomic Polymerase Kit (BD Clontech, Heidelberg, Germany), which resulted in products with A-overhangs, was also used. These enzymes or enzyme mixes all contained a proofreading activity. In all cases, manufacturer's protocols were followed.

### 2.5.3 Ligation

PCR products were routinely cloned into vector pGEM-T® Easy (Promega), if they contained an 3'-A overhang. The manufacturer's protocol was followed, except for an overnight incubation at 12°C instead of 4°C. Blunt end PCR products were cloned into vector pCR®-BluntII-TOPO® (Invitrogen).

For plasmid construction, vector and insert were restricted with the appropriate endonucleases (2.5.1) and purified from agarose gels (2.4.6). For the ligation reaction both components were mixed in a ratio of 1 (vector) to 3 (insert) in a reaction volume of 10 µl. T4 DNA ligase from

Promega or Roche were used. The reaction was incubated at 12°C overnight and then transformed into *E. coli* (2.2.1).

Alternatively, the donor vector and the destination vector were restricted with the appropriate endonucleases in a single reaction tube. After restriction, the enzymes were removed by phenol/chloroform extraction. The remaining aqueous phase containing the DNA was precipited with ethanol and resuspended in water. Then ligation buffer (10x concentrated) and T4 DNA ligase (Roche) were added in the volumes required for a final reaction volume of 10 µl. Ligation was performed at 12°C overnight. Subsequently, T4 DNA ligase was inactivated by incubation at 65°C for 15 min. If the vectors carried different antibiotic selection markers, selection was done by resistance against the appropriate antibiotic. Otherwise the unwanted vector was cleaved by a restriction endonuclease cutting only the backbone of this plasmid. Thereafter, the enzyme was inactivated by incubation at 65°C for 20 min and the reaction mixture was transformed into *E. coli*.

For the cloning of BAC fragments, a self ligated variant of the vector pGEM-T® Easy (Promega) was used. 1 µg of vector DNA was cut with 10 U of the required restrictase [*Nco*I, *Nco*I and *Sal*I or *Nco*I and *Sph*I (Roche)] in a final volume of 20 µl at 37°C for 1 h. Afterwards, restriction enzymes were inactivated at 65°C for 20 min. Subsequently, 2.5 µl shrimp alkaline phosphatase (SAP) and 2.5 µl 10x reaction buffer (Roche) were added and the mixture was incubated at 37°C for 30 min. SAP was inactivated at 65°C for 15 min. Approximatly 500 ng BAC-DNA were restricted with 40 U of the appropriate endonuclease in a final volume of 100 µl, incubated at 37°C overnight. The enzymes were removed by phenol/chloroform extraction, and the DNA was further purified and concentrated by ethanol precipitation, and subsequently resuspended in 10 µl $H_2O$. 1 µl of the DNA solution was used for agarose gel examination of the restriction reaction, and the remaining 9 µl were added to the ligation reaction containing 1 U T4 DNA ligase (Roche), 20 ng dephosphorylated vector pGEM-T® Easy and the appropriate volume of ligation buffer. As a negative control, a ligation reaction was set up with the empty, dephosphorylated pGEM-T® Easy and water instead of BAC fragments. Ligation reactions were incubated at 12°C overnight and inactivated at 65°C for 20 min. 1 µl of the ligation mixture was transformed into *E. coli*. Subsequently, a colony hybridisation was carried out, if there were at least a 100-fold more transformants originating from ligation reactions containing BAC DNA than in the empty vector control.

### 2.5.4 Synthesis of cDNA and semi-quantitative RT-PCR

1 µg total RNA was used for cDNA first strand synthesis with an oligo-$(dT)_{15}$ primer (Promega) using the Omniscript RT kit (Qiagen) in a 20 µl reaction. 0.8 µl of the reverse transcription reaction mix was included as template into a 20 µl PCR reaction (Taq PCR Master Mix Kit, Qiagen). RT-PCR was run for 30 cycles unless stated otherwise, reaction conditions were used as indicated in chapter 2.5.2. PCR products were then examined by agarose gel electrophoresis (chapter 2.4.5).

Double-stranded cDNA was synthesized from enriched mRNA using the SMART™ PCR cDNA Synthesis Kit (BD Clontech). This kit applied the SMART™ technology. During first strand synthesis the cDNA strand was labeled at the 3' end and the 5' end. The 5'end of the

cDNA was labeled by a modified poly-dT primer whereas the 3' end labelling was done by "template switching". When the reverse transcriptase reached the 5' end of the RNA template three cytosine residues were added to the freshly synthesized cDNA strand. Those residues allowed the SMART II A oligonucleotide to bind to the cDNA due to its 3' guanine overhang, lengthening the original template for the reverse transcriptase. Thus, the reverse transcription resulted in a first cDNA strand labeled on both sides with the same known sequence. Second strand synthesis was a PCR reaction using primers against the attached sequences. The obtained double-stranded cDNA was purified by a proteinase K treatment and nucleotide overhangs were removed by T4 DNA polymerase. After precipitation and resuspension the concentration of the cDNA was determined by absorption measurement (chapter 2.4.8) and the integrity was examined by agarose gel electrophoresis.

### 2.5.5 Sequence analysis

All sequencing reactions were carried out by MWG Biotech (Ebersberg). Obtained sequences were analyzed using the Clone Manager Professional Suite software (Sci Ed Software, Durham, NC, USA). Sequence comparisons to known sequences in various databases were conducted using the BLAST algorithm (Altschul et al. 1990) Online databases used were Genbank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) for general sequence comparison, TAIR (http://www.arabidopsis.org/) for comparisons to the model plant *A. thaliana*, Swiss-Prot (http://www.expasy.ch/sprot/) for protein analysis, and Brassica DB (http://ukcrop.net/brassica.html#brassicadb; Dicks et al., 2000; later transferred to http://brassica.bbsrc.ac.uk/) for Brassica sequences. Similarity trees were constructed using the program Treecon (Van de Peer and de Wachter, 1994), after aligning the sequences using the ClustalW algorithm. For the composite tree of the deduced protein sequences of the UGT-ESTs a multiple alignment was performed covering 70 amino acids, starting with the UGT signature motif HCGWN. The unrooted Neighbor-Joining tree from different HCA-GTs was constructed with an alignment of their full-length open reading frames (ORF). For distance calculations, the algorithm of Jukes and Cantor (1969) was used; insertions and deletions were taken into account.

## 2.6 Cloning of BnHCA-GT-like sequences

### 2.6.1 EST-cloning

Total RNA extracted from buds, flowers and different seed developmental stages covering the whole seed development was pooled and reverse transcribed. For EST amplification primers listed in Table 2-2 were used with the cDNA as template and Platinum® PCR SuperMix High Fidelity (Invitrogen). To increase specificity, a touch down cycling modus was applied with annealing temperatures ranging from 55°C for the initial 5 cycles through 52°C and 49°C for 5 cycles each to a final value of 45°C for the remaining 25 cycles. Denaturing was done at 95°C for 30 sec and elongation was conducted at 68°C for 60 sec. PCR products were cloned and characterized by sequence analysis.

### 2.6.2 Cloning of *UGT84A10*

Sequence analysis revealed that the *B. oleracea* genomic fragment specified by GenBank Accession number BH591016 (Ayele et al., 2005) harbors the 3' end of an open reading frame (ORF) encoding a BnHCA-GT like protein. To amplify the 5'end of the gene, 5'RACE PCR was conducted. Poly($A^+$)RNA from seeds and leaves was reversely transcribed using the SMART$^{TM}$ RACE cDNA Amplification Kit (BD Clontech), resulting in RACE-ready-cDNA. The 5' end of RACE-ready-cDNA was labeled by a modified oligo (dT) primer, whereas the 3' end was extended by "template switching" (2.5.4). 5'RACE was performed using the UGT84A10rev primer (Table 2-1). The resulting amplicon was cloned and sequenced using primers SP6 and T7. Sequence analysis of the obtained fragments indicated the presence a full-length ORF, which was amplified by PCR using 5'RACE ready cDNA from leaves as template and FL_UGT84A10fw and FL_UGT84A10rev as primers. The genomic sequence of *UGT84A10* was PCR-amplified using genomic DNA from *B. napus* cv. Express as template and Gen-FL_UGT84A10fw and Gen-FL_UGT84A10rev as primers.

### 2.6.3 Cloning of *UGT84A11*

GenBank fragment BH516441 (Ayele et al., 2005) was used to amplify the according full-length ORF by a combined strategy of RACE PCR and genome walking. 3'RACE was carried out using 3'RACE ready cDNA from leaves and seeds as template and the gene-specific primer UGT84A11fw (Table 2-1). Nested PCR was conducted with a 50-fold dilution of the first PCR reaction as template and the gene specific nested primer UGT84A11fw-nested. Since the 5'RACE failed to produce an amplicon and the *A. thaliana* homologues were known to be intron-less genes, a genome walker approach was applied. Genome walker libraries were constructed with genomic DNA from *B. napus* cv. Express with the Universal Genome Walker$^{TM}$ kit (BD Clontech). For the amplification of the 5' end UGT84A11rev was used as gene-specific primer. Sequence analysis of the cloned PCR products revealed the presence of an ORF covering the combined sequences, which was used to derive primers for amplification of the full-length cDNA. The cDNA and the genomic fragment of *UGT84A11* were cloned using the primers FL_UGT84A11fw and FL_UGT84A11rev.

## 2.7 DNA hybridization techniques

### 2.7.1 Southern blot

5-10 µg of genomic DNA or 100 ng of BAC DNA were restricted with the appropriate endonuclease in a volume of 100 µl. After an overnight incubation of the restriction reaction, the enzyme was removed by phenol/chloroform extraction and DNA was precipitated and resuspended in 20 µl of water. Cleaved DNA was separated by agarose gel electrophoresis using a 0.75% agarose gel and 1x TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA at pH 8.0). For size determination, DIG-labeled DNA Molecular Weight Marker II (Roche) was applied. The gel was run at 20 V overnight.

Before blotting, the gel was soaked for 10 min in 0.25 M HCl to depurinate the DNA, followed by a denaturation step (0.5 M NaOH, 1.5 M NaCl) for 30 min and a neutralisation (1 M Tris-HCl, 2 M NaCl at pH 7.5) for 30 min. Thereafter, the gel was equilibrated in 20x SSC (3 M NaCl, 0.3 M sodium citrate at pH 7.0). Blotting was carried out by capillary blot technique using 20x SSC as transfer buffer (Sambrook et al., 1989). DNA was transferred to a nylon membrane (Hybond NX, Amersham Biosciences, Piscataway, NJ, USA) and subsequently fixed by UV-crosslinking. Afterwards, the membrane was washed in water and put in a hybridization bottle.

Pre-hydridization was conducted by incubating the membrane with DIG Easy Hyb buffer (Roche) for 2 h at 42°C. The probe was labeled using the PCR DIG Probe Synthesis kit (Roche), following the manufacturer's protocol. 5-25 ng labeled PCR product per ml hybridization solution [DIG Easy Hyb (Roche)] were used as probe. The probe was denatured at 95°C for 5 min, then cooled on ice for 5 min and added to the hybridization solution. Hybridization was done at 42°C overnight.

After hybridization, the membrane was washed two times for 5 min in 2x SSC, 0.1% (w/v) SDS at room temperature and two times for 15 min in 0.5x SSC, 0.1% (w/v) SDS at 68°C to remove non-specifically bound probe molecules. The detection of the DIG-labeled probe was conducted using the DIG Luminescent Detection kit (Roche). All steps were performed at room temperature. Following the washes, the membrane was equilibrated for 1 min in DIG washing buffer [0.1 M maleic acid, 0.15 M NaCl, 0.3% Tween 20 (pH 7.5)]. After equilibration, the membrane was soaked for 2 h in 2x blocking solution [2% (w/v) blocking reagent diluted in maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl at pH 7.5)], followed by an incubation in FAB-antibody solution [2x blocking solution, 1.5 $U \cdot ml^{-1}$ anti-digoxigenin-AP, Fab fragments] for 30 min and two washing steps of 15 min each in DIG washing buffer. Since the alkaline phosphatase conjugated to the anti-digoxigenin antibody requires an alkaline pH for proper function, the membrane was equilibrated for 5 min in DIG-detection buffer (0.1 M Tris-HCl, 0.1 M NaCl at pH 9.5). Signals were generated using CSPD-solution [250 nM CSPD (disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[1.3.1.1.$^{3,7}$]decan}-4-yl)pheryl phosphate) in DIG-detection buffer] as substrate for the conjugated alkaline phosphatase. Chemoluminescence was captured on Lumi-Film Chemiluminescent Detection Film (Roche) during two hour exposition at 37°C.

### 2.7.2 Screen of a genomic BAC library

Three filters containing a genomic BAC library of *B. napus*, providing an eight-fold genome coverage, were obtained from the University of Giessen. The BAC library was originally provided by the German Resource Center for Genome Research (RZPD, Berlin, Germany) and was developed using genomic DNA from *B. napus* L. var. *napus* cv. Express. Filters were pre-hybridized in hybridization buffer [500mM sodium phosphate (pH 7.2), 7% (w/v) SDS, 10 mM EDTA] at 65°C for at least two hours.

The probe was labeled radioactively by incorporation of α-$^{33}$P-ATP, using the MegaPrime-DNA Labeling System (GE Healthcare Life Science). 25 ng DNA was used as starting material and labeling was carried out by the random priming method with random hexamer primer, α-$^{33}$P-dATP [3000Ci per mmol (ICN Biomedicals, Eschwege, Germany)] and Klenow

fragment in the reaction mixture. Non-incorporated nucleotides were removed by gel filtration chromatography using ProbeQuant G-50 Micro Columns (GE Healthcare Life Science). The probe was denatured at 95°C for 5 min then cooled on ice for 5 min and added to 10 ml of hybridization buffer. The hybridization was done at 65°C overnight.

Non-specifically bound probe was removed by washing for 30 min at 65°C in washing buffer [40 mM sodium phosphate buffer (pH 7.2), 1% (w/v) SDS] at least twice. Afterwards, the filters were dried briefly between two sheets of Whatman paper (Whatman, Dassel, Germany) and then wrapped into saran wrap, avoiding air bubbles. Signals were detected by exposing the filters to X-ray films for two days at –80°C. Signals were assigned to BAC clones according to evaluation schemes provided by the manufacturer (Deutsches Ressourcenzentrum für Genomforschung). The identified BAC clones were kindly supplied by Rod Snowdon (University of Giessen, Germany).

### 2.7.3 Colony hybridization

Colony hybridization was used to rapidly screen a large number of *E. coli* transformants, especially for the cloning of BAC sub-fragments. Ligation was conducted as described in chapter 2.5.3. Transformed *E. coli* were spread onto 145 mm Petri dishes (Greiner Bio One, Solingen, Germany) and incubated overnight at 37°C. Numbers of colonies were estimated and compared to a self-ligation of the dephosphorylated vector.

Round Hybond N+ nylon membranes with a diameter of 132 mm (Amersham Biosciences) were punched three times asymmetrically and then placed onto the colonies. Plates containing the colonies were punched in the same pattern to allow the assignment of positive signals to the respective colonies. After 2 min the membranes with the transferred *E. coli* cells were carefully removed from the plate. The plates containing the *E. coli* clones were again incubated at 37°C for 4 to 6 h to allow the colonies to re-emerge. The nylon membranes were placed on saran wrap containing 1 ml denaturation buffer (0.5 M NaOH, 1.5 M NaCl), allowing the membrane to get soaked with the buffer for 2 min. Then, the membranes were dried briefly and this procedure was repeated once. Subsequently, the membranes were soaked two times with neutralisation buffer (1 M Tris-HCl, 2 M NaCl at pH 7.5) for 2 min, and were then UV-crosslinked to fix the DNA. Hybridization and signal detection were done according to the southern blot protocol (2.7.1).

## 2.8 cDNA-AFLP

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). Double-stranded cDNA was synthesized as described in chapter 2.5.4. Primers and AFLP adapter were derived following the Keygene protocol. Primers consisted of a core region complementary to the AFLP adapter and one to three specific nucleotides at the 3’ end. Sequences are given in Table 2-5. The actual cDNA-AFLP experiments were performed by Dr. Sabine Mikolajewski from the Bavarian State Research Center for Agriculture in Freising. Identities of eluted transcript-derived fragments (TDFs) were

confirmed by PCR product sequencing using the primer with three specific nucleotides at the 3' end.

## 2.9 Heterologous gene expression and protein analysis

### 2.9.1 Heterologous expression of *BnHCA-GT* cDNAs in *E. coli*

*UGT84A10* and *UGT84A11* sequences were amplified from the cloned cDNA by PCR using ProofStart DNA Polymerase (Qiagen). Restriction sites required for cloning were introduced by primers as presented in Table 2-6 (primers 1-4). The produced amplicons were cloned (2.5.3) into pCR®-BluntII-TOPO® (Invitrogen) and evaluated by sequencing. Ligation of the cloned cDNA into the expression vector pET28a(+) (Novagen) was carried out as described in chapter 2.5.3. Expression constructs were then transformed into *E. coli* strain BL21 Codon Plus (DE3)-PR (Stratagene). Cells were grown according to the manufacturer's instruction to an $OD_{600}$ of 0.6. Protein expression was induced by adding 1 mM IPTG to the medium. After induction, the bacteria were incubated at 20°C overnight shaking at 150-200 rpm. Subsequently, the cells were centrifuged at 4°C and resuspended in a buffer containing 0.1 M MES [2-(N-morpholino)ethanesulfonic acid] at pH 6.0, 10% (v/v) glycerol, 0.01% (v/v) β-mercaptoethanol, 0.5 mM EDTA. Cells were disrupted by sonication and then the debris was sedimented by centrifugation at 4°C and 14.000 rpm for 20 min. The supernatant contained the soluble proteins which were subjected to the SGT activity assay (2.9.6).

### 2.9.2 Determination of protein concentration

1 ml Bradford reagent [0.0065% (v/v) coomassie G-250 in 5% (v/v) methanol, 10% (v/v) phosphoric acid] was added to 2 to 20 µl of crude protein extract and incubated for 5 min at room temperature (Bradford, 1976). Absorption was measured at 595 nm wavelength and compared to a blank value. Protein concentration was determined against a BSA calibration curve.

### 2.9.3 SDS polyacrylamide gel electrophoresis (SDS PAGE)

**Table 2-10. Composition of SDS gels (specifications for 2 gels )**

| Solution | Stacking gel | Separation gel |
|---|---|---|
| Acrylamide [30% (w/v); BioRad] | 0.7 ml | 3.7 ml |
| 0.5 M Tris HCl at pH 6.8 | 1.25 ml | - |
| 1.5 M Tris HCl at pH 8.8 | - | 2.5 ml |
| $H_2O$ | 3.0 ml | 3.8 ml |
| 10% APS | 25 µl | 50 µl |
| TEMED | 5 µl | 5 µl |

SDS-PAGE is a technique used for separation of denaturated proteins according to their molecular weight. Table 2-10 shows the composition of the used gels. Protein solutions containing 20 µg protein were mixed with equal volumes of protein sample buffer [0.125 M Tris HCl at pH 6.8, 5 mM EDTA, 15% (v/v) glycerol, 2% (v/v) SDS, 0.1% (v/v) bromphenol blue, 0.1% (v/v) β-mercapto ethanol]. Electrophoresis was carried out at 25 mA for 60 min.

### 2.9.4 Protein staining

After electrophoresis proteins were stained by boiling the gels for circa 2 min in coomassie solution [0.2% (v/v) Coomassie Brilliant Blue R-250, 10% (v/v) acetic acid, 40% (v/v) methanol]. Background was stripped by boiling the gels three times in 7% acetic acid for 1 min, followed by overnight incubation of the gels in acetic acid at room temperature. Gels were documented by photographing under white light.

### 2.9.5 Western blot

His-tagged proteins expressed in *E. coli* were detected by Western blot. After electrophoresis, proteins were transferred to nitrocellulose membranes by a wet blot techniques using XCell II blot module (Invitrogen) following the manufacturer's protocol. Protein transfer was performed at 30 V for 60 min. Thereafter, membranes were washed briefly in TBS (100 mM Tris HCl at pH 7.5, 150 mM NaCl) and then incubated in blocking solution [5% (w/v) milk powder in TBS] at room temperature for 60 min. After blocking, the membranes were incubated with the primary anti-His-tag antibody, applied 1:1000 diluted in blocking solution, at 4°C overnight. Non-specifically bound antibodies were removed by washing three times for 5 min at room temperature with TTBS [TBS at pH 7.5, 0.05% (v/v) Tween 20]. Then, the membranes were incubated with the secondary antibody, diluted 1:1000 in blocking solution, for 2 h at room temperature. The secondary antibody recognized immunoglobulin 1 from mouse and was conjugated with alkaline phosphatase for colorimetric detection. After incubation with the secondary antibody, the membranes were washed again three times for 5 min in TTBS. Subsequently, the blot was incubated for 5 min in TBS at pH 9.5 and then equilibrated with substrate buffer (TBS at pH 9.5, 50 MM $MgCl_2$). The color-reaction was started by adding of 0.16 $mg \cdot ml^{-1}$ BCIP (5-bromo-4-chloro-3-indolyl phosphate) and 0.32 $mg \cdot ml^{-1}$ NBT (nitroblue tetrazolium). After signals appeared the reaction was stopped by washing the membranes in TE-buffer (10 mM Tris HCl at pH 8.0, 1 mM EDTA).

### 2.9.6 SGT activity assay

Crude, desalted protein extracts were incubated in MES buffer [0.1 M MES at pH 6.0, 10% (v/v) glycerol, 0,01% (v/v) β-mercapto ethanol, 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 triflouroacetic acid (TFA). Reaction products were analyzed by HPLC. Sinapate esters were identified by comparison with the authentic compounds from the compound library of the IPB.

## 2.10 Generation and analysis of transgenic plants

### 2.10.1 Construction of dsRNAi vector

Primers given in Table 2-6 were used for the amplification of gene-specific *UGT84A9a* (primers 5-8) and *BnSCT1* (9-12) sequences and introduction of the appropriate restriction sites. The PCR reaction was carried out using ProofStart DNA polymerase (Qiagen). The cloning of antisense and sense fragments into pBNN (Dr. C. Milkowski, IPB Halle) was described previously (Hüsken et al., 2005). The dsRNAi cassette was transferred to the binary vector pLH7000 (Hausmann and Töpfer, 1999) using restriction sites of *Spe*I and *Pst*I in both vectors. Vector construction was checked constantly by restriction analysis.

### 2.10.2 Segregation analysis

In order to obtain homozygous transgenic plants, a segregation analysis was carried out in the T3 generation. Selective marker was the *bar* gene, which confers phosphinothricine resistance. For each line 10 T3 seeds were sown on a soil-vermiculite mixture. 14 days after sowing the seedlings were sprayed with a herbizide solution containing 0.04 $mg \cdot ml^{-1}$ BASTA® (Hoechst Schering AgrEvo GmbH). This treatment was repeated twice, every four to five days. Five days after the third treatment, the plants were evaluated by eye inspection. If all germinated plants of a given line were resistant to the herbicide treatment, this line was assumed to be homozygous. For these lines the experiment was repeated to confirm the result.

### 2.10.3 Transcript level evaluation

For the investigation of transcript levels of *UGT84A9a*i seeds, homozygous transgenic T5 plants were grown in the greenhouse (2.3). Flowers and two different developmental stages of the formed T6 seeds were harvested. For the investigation of seedlings, T6 seeds were sown on wet 3MM Whatman paper (Whatman) and harvested after 2 days. To measure *UGT84A9* transcripts in *UGT84A9a*i seedlings over a longer time period, homozygous transgenic T6 seeds were sown on a soil-vermiculite mixture and the seedlings were harvested after 2, 4, 7, 10 and 14 days. For the last three time points, the seedlings were separated in upper parts and roots. RNA was prepared as described (2.4.1) and semi-quantitative RT-PCR was performed (2.5.4).

### 2.10.4 Germination experiments

Germination studies were carried out independently for one batch of T6 and three batches of T7 seeds of homozygous *UGT84A9a*i lines and a wild type control. For each experiment, 100 seeds of each line were sown of wet 3MM Whatman paper (Whatman) in the greenhouse. Seeds were evaluated for germination once every 24 hours. A seed was defined as having germinated if the root tip or a part of a cotyledon was visible by eye inspection.

### 2.10.5 Extraction of sinapate esters from *B. napus* seeds

For single seed analysis, the fresh weight of each seed was determined before extraction. For batch analysis, 20 seeds were coarsely ground and 20 µg fresh weight were subjected to analysis. The seed material was transferred to 500 µl 80% (v/v) methanol and then ground by zirconium beads (1.0 mm diameter; Biospec Products) in a Bead Beater (Biospec Products). Subsequently, homogenates were centrifuged at 4°C and maximum speed for 5 min in a benchtop centrifuge. The supernatant was used for HPLC analysis.

For the analysis of total sinapate content, methanolic seed extracts were saponified by incubation with an equal volume of 1 M NaOH at 50°C for 1 h. Then, the solutions were neutralized with 50 µl glacial acetic acid and centrifuged at room temperature and maximum speed for 5 min in a benchtop centrifuge. 10 µl of the supernatant was subjected to HPLC analysis.

## 2.11 Detection of sinapate esters by HPLC

HPLC analysis was done on a Waters Alliance System (Waters, Eschborn, Germany), using a 5 µm Nucleosil C18 column [250 mm x 4 mm (Machery-Nagel, Düren, Germany)]. Components were separated by a 15 min gradient starting at 10% acetonitrile and going up to 50%, diluted in 1.5% phosphoric acid. Signals were detected by a PDA (Photodiode Array Detector). Sinapate derivates were identified by comparison of their spectra and their retention times to external standards.

# 3 Results

## 3.1 Search for alternative glucosyltransferases

The aim of the following experiments was to search for glucosyltransferases related to *UGT84A9* that possibly contribute to the formation of sinapoylglucose during seed development in *Brassica napus*. Two different strategies were follwed to achieve this goal. First, based on genomic sequence tags of *Brassica oleracea* that showed high identity to *UGT84A9*, two novel genes encoding BnHCA-GTs were cloned and designated *UGT84A10* and *UGT84A11*. Both new genes were functionally characterized.

The second strategy used was an directed EST cloning project. Degenerated primers recognizing the sequences encoding the well characterized HCGWN motif of the PSPG-box of plant secondary metabolism glycosyltranferases were used in an RT-PCR approach with cDNA from flowers and developing seeds as template. By applying this strategy, most glycosyltransferases with significant gene expression in the analyzed tissues should be found.

### 3.1.1 Cloning of *UGT84A10* and *UGT84A11*

For the cloning of *UGT84A10* and *UGT84A11*, RACE PCR and Genome Walker techniques were applied. For RACE PCR, cDNA libraries synthesized from RNA of mature leaves and early seed stages were used as template. The experimental strategy depends on oligo-nucleotides with known sequences, which are ligated to the cDNA ends to provide primer binding sites outside the gene. Genome walking applies the RACE principle to genomic DNA. For all experiments, *B. napus* L. var. *napus* cv. Express was used.

Starting point for the cloning of *UGT84A10* was sequence information of the GenBank entry BH591016. This database fragment harbours the 3'end of an ORF similar to *UGT84A3* from *Arabidopsis thaliana*. 5'RACE PCR was performed using a 5'RACE ready cDNA originating from leaves as template. The reaction resulted in a 1560 bp amplicon, which was cloned into pGEM-T Easy and sequenced from both ends using the sequencing primer T7 and SP6. The sequence obtained with the T7 primer contained a start codon indicating the start of an ORF. The SP6-derived fragment harboured a continuous ORF. Both partial sequences showed over 80% sequence identity to *UGT84A3*. Primers for the amplification of the putative full-length ORF were derived and applied to RACE-ready cDNA as template. A 1419 bp ORF was amplified and designated *UGT84A10*, according to the nomenclature by Mackenzie and co-workers (1997). Sequence analysis showed a high identity to *UGT84A3* and *UGT84A4* (Table 3-1).

Database fragment BH516441 (GenBank, NCBI) shows pronounced similarity to the 5' end of *UGT84A1*. 3'RACE PCR was performed with 3'RACE ready cDNA from leaves as template. Nested PCR resulted in a 702 bp fragment harbouring the 3'end of an ORF that showed 79% identity on nucleotide level to the 3' end of *UGT84A1*. Since BH516441 contains no continuous ORF, 5'RACE PCR was performed. As several attempts to amplify the 5'end of this ORF failed, genome walking was applied. As *UGT84A1*, the homologous *A. thaliana* gene, has no introns, none were expected for the *B. napus* homologue either. A 1311

bp fragment was amplified which harboured the 5' end of the coding sequence. Primers were derived from both ends of this ORF and the complete 1431 bp ORF was obtained as a continuous PCR fragment using cDNA from leaves as template. Sequence analysis revealed that this ORF, designated *UGT84A11*, shares 80% identical nucleotides with *UGT84A1* (Table 3-1).

**Table 3-1. Comparison of DNA and amino acid sequences of *UGT84A10* and *UGT84A11* with related UGTs.**

| | | Identity DNA | Identity amino acids |
|---|---|---|---|
| *UGT84A10* | *UGT84A1* | 64% | 62% |
| | *UGT84A2* | 65% | 60% |
| | *UGT84A3* | 76% | 71% |
| | *UGT84A4* | 77% | 73% |
| | *UGT84A9* | 65% | 60% |
| | *UGT84A11* | 64% | 60% |
| *UGT84A11* | *UGT84A1* | 80% | 78% |
| | *UGT84A2* | 65% | 59% |
| | *UGT84A3* | 65% | 62% |
| | *UGT84A4* | 64% | 61% |
| | *UGT84A9* | 66% | 58% |

### 3.1.2 Heterologous expression of recombinant *UGT84A10* and *UGT84A11*

*UGT84A10* carries an ORF for a protein of 472 amino acids with a calculated molecular mass of 53,056 Da and an isoelectric point (pI) of 5.53. The *UGT84A11* ORF encodes a protein of 476 amino acids. The molecular mass and the pI were calculated to be 53,619 Da and 5.66, respectively. To prove the functionality of the proteins, cDNAs of both genes were expressed in *E. coli*. Crude protein extracts were assayed for UGT activity using cinnamate and hydroxycinnamates as substrates in the presence of UDP-glucose (Figure 3-1).

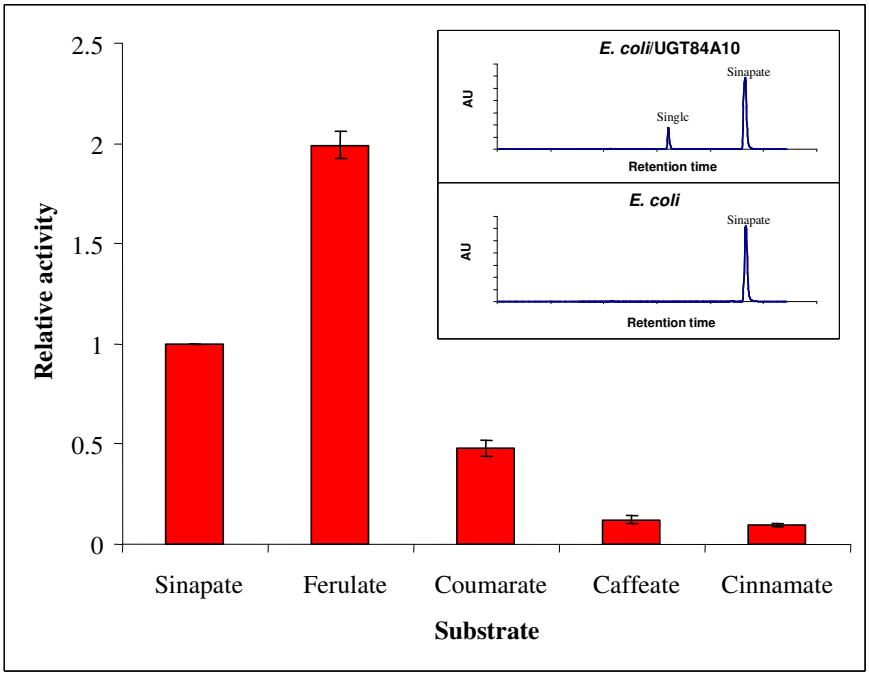



**Figure 3-1. Acceptor specificity of the recombinant UGT84A10 enzyme (Mittasch et al., 2007). Mean values were calculated from three independent measurements (mean ± SD). Relative activities were calculated according to the formation of sinapoylglucose [1.9 ± 0.5 pkat (mg protein)$^{-1}$], which was set to 1.0 for each measurement. Inserted figure shows the HPLC chromatograms of sinapate GT assays conducted with *E. coli* cell extracts harboring pET28a(+)-UGT84A10 (top) and the empty pET28a(+) (bottom).**

Cell extracts from *E. coli* harbouring the expression plasmid pET28a(+)-*UGT84A10* were found to convert HCAs to their respective glucose esters. Preferred substrates were ferulate, followed by sinapate and coumarate (Figure 3-1). The activity of recombinant UGT84A10 towards different hydroxycinnamates and cinnamate were calculated relative to the conversion of sinapate, which was set to 1.0 for each set of measurements. The formation of the respective glucosides was never observed. Crude protein extracts from cells containing pET28a(+)-*UGT84A11* were never found to catalyze the formation of HCA-glucose esters, although the recombinant protein could be detected in Western blot analysis ( Figure 3-2).

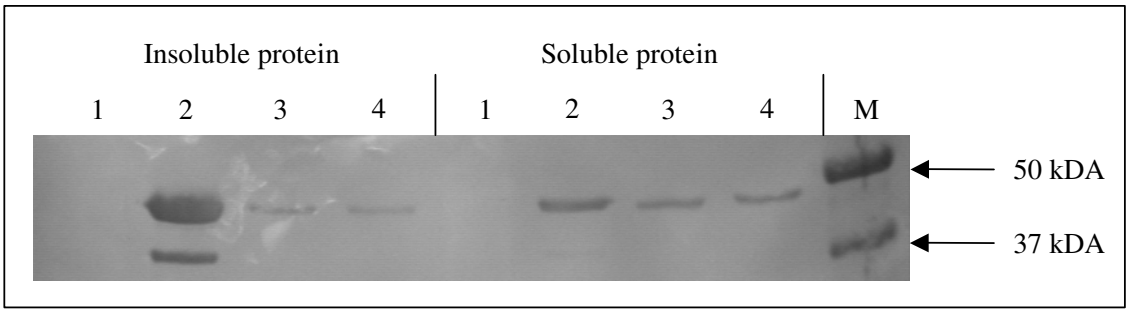



**Figure 3-2. Western blot analysis of recombinant His-tagged UGT84A11 protein. The primary antibody used was anti-His antibody produced in mouse, the secondary antibody was anti-mouseIgG antibody conjugated with alkaline phosphatase. Samples: 1, empty vector control; 2, UGT84A10; 3 and 4, UGT84A11; M, marker. No product formation was found in GT activity assay using recombinant UGT84A11.**

### 3.1.3 EST cloning from flowers and seeds of *B. napus*

The second approach to find UGT84A9-related enzymes was an EST cloning experiment. ESTs were amplified using degenerate primers derived from sequences encoding a highly conserved amino acid motif in the plant secondary product glycosyltransferases designated PSPG box (Hughes and Hughes, 1994). The RT-PCR was performed with a mixed RNA pool, representing flowers and seeds from the early globular to the desiccated developmental stage. This should ensure that the widest possible range of UGT-ESTs potentially involved in seed metabolism would be detected. The approach resulted in the cloning and sequence analysis of 132 ESTs. Of these, about 24% (32 ESTs) shared sequence similarity with plant secondary product UGTs. The implementation of a touchdown protocol for the PCR reaction led to an increase in the proportion of putative UGT fragments in the RT-PCR product pool from 17 to about 37%.

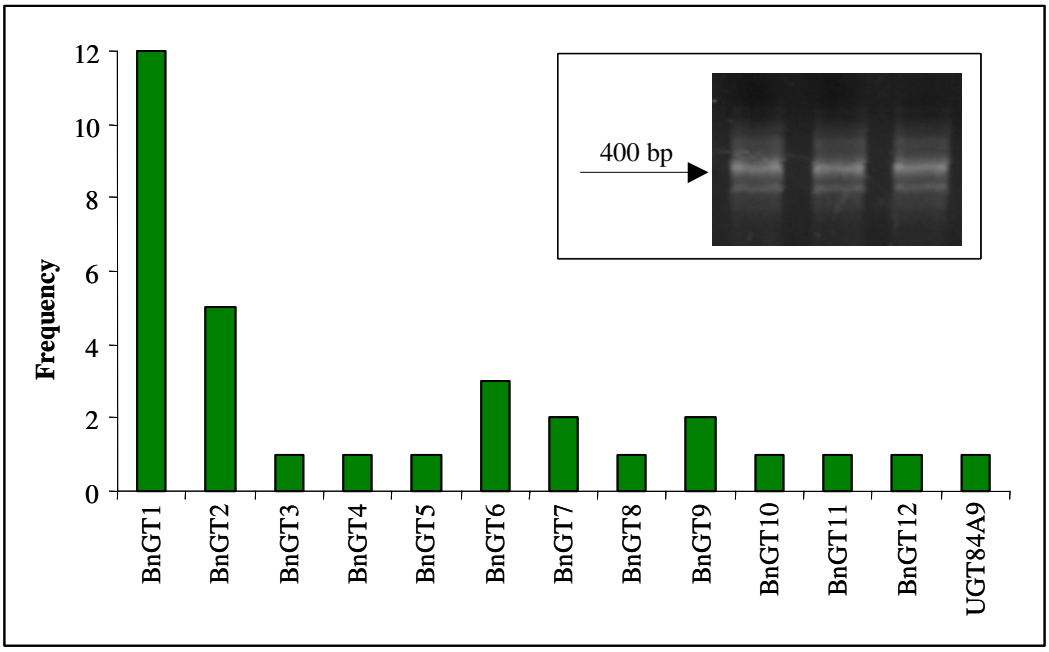



**Figure 3-3. Frequency of UGT-like EST classes from seeds and flowers of *B. napus* (n = 32). The inserted figure shows the agarose gel electrophoresis pattern of cDNA fragments generated by a touchdown RT-PCR with RNA from flowers and seeds of *B. napus*. The expected fragment size was about 400 bp.**

The 32 cloned putative UGT cDNAs represent 13 distinct classes of ESTs (Figure 3-3). Sequence analysis revealed that one class specifies *UGT84A9*, the others were numbered BnGT1 to BnGT12. BnGT1 was the most frequent EST, followed by BnGT2. All other classes were found in low abundance. Each class contains the 3' end of an ORF, the deduced amino acid sequence of which starts with the UGT signature motif HCGWN, with exception of BnGT5 and BnGT12. These two EST classes accumulated nonsense mutations interrupting the ORFs and indicating pseudogenes, thus they were excluded from further analysis. BnGT4 and BnGT7 carry remarkably short ORFs, compared to functional UGTs, which are characterized by more than 100 amino acid residues reaching from the HCGWN motif to the C-terminus. The reason for this is most likely internal priming during RT-PCR. Since both GTs contain a continuous ORF they were included in further analysis.

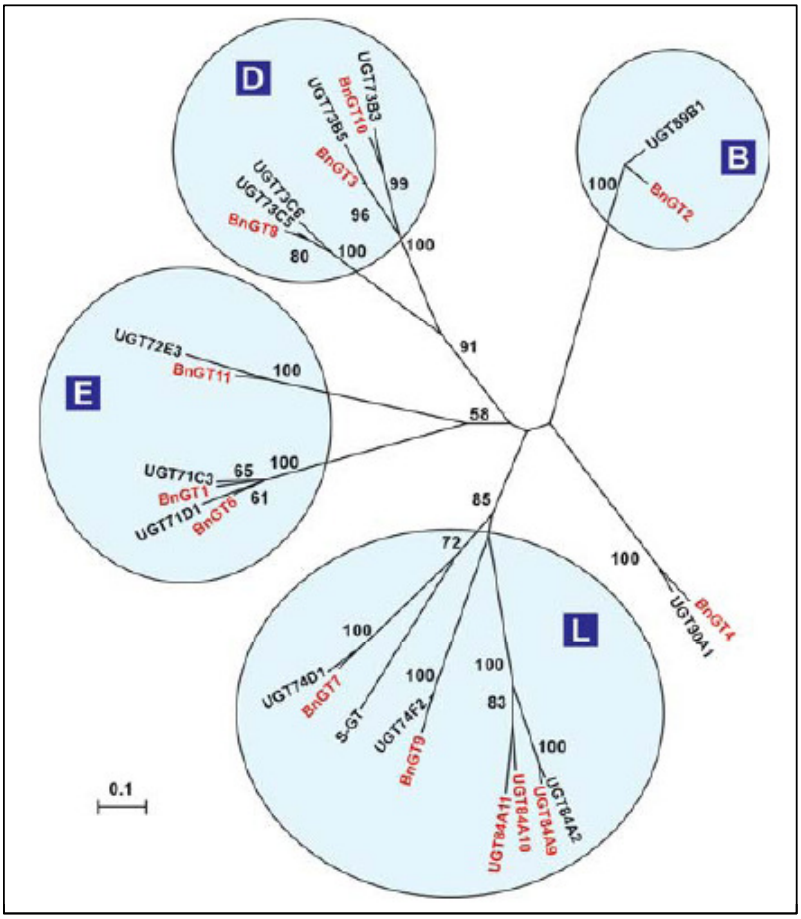



**Figure 3-4. Composite tree of of UGT-like ESTs from *B. napus* and *A. thaliana* homologues (Mittasch et al., 2007). Bootstrap values over 50% are indicated. Scale represents 0.1 fixed mutations per site. Groups B, D, E and L established for the *A. thaliana* UGT family (Li et al., 2001) are indicated. ESTs and full-length cDNAs isolated during this work are marked in red, their deduced amino acid sequences were used for comparison. *A. thaliana* sequences are named according to the UGT nomenclature. UGTs with proven activity: UGT89B1, hydroxybenzoate GT (Lim et al., 2002); UGT73C6, flavonol 3-*O*-glycoside GT (Jones et al., 2003); UGT73C5, brassinosteroid GT (Poppenberger et al., 2005); UGT72E3, hydroxycinnamate, hydroxycinnamyl alcohol and aldehyde GT (Lim et al., 2005); UGT74F2, anthranilate GT (Quiel and Benders, 2003); UGT84A2, sinapate GT (Milkowski et al., 2000b; Lim et al., 2001). S-GT refers to thiohydroximate *S*-glucosyltransferase from *B. napus* (Marillia et al., 2001).**

To establish the relationship between the newly cloned *B. napus* ESTs and proven UGTs from *A. thaliana*, a distance-based unrooted tree was constructed based on deduced amino acid sequences (Figure 3-4). The isolated sequences could be assigned to established groups of the UGT family (Li et al., 2001) with exception of BnGT4. Three classes of ESTs belong to groups D, E and L, respectively, whereas only one class was placed in group B. Other groups of the UGT family were not found in this EST pool. Group L that contains the ester-forming UGTs was of special interest, since it also harbors UGT84A1-4, which are involved in the sinapate metabolism in *A. thaliana*. In this study we found no ESTs of genes potentially involved in sinapoylglucose formation in flowers or seeds, except the previously identified UGT84A9. The two other ESTs, which were assigned to group L are BnGT7 and BnGT9. For the *A. thaliana* homologue of BnGT7 (UGT74D1) no function has been described yet, however, the related *A. thaliana* enzymes UTG74B1 and UGT74C1 were reported to function in the glucosylation of glucosinolate precursors (Grupp et al., 2004; Gachon et al., 2005). In a sequence alignment covering 70 amino acids downstream of the HCGWN motif, BnGT7

shows 91% sequence identity to the non-characterized UGT74D1, but only 56% and 67% to UGT74B1 and UGT74C1, respectively. S-GT (thiohydroximate *S*-glucosyltransferase; Marillia et al., 2001), which is the only other characterized group L glycosyltransferase from *B. napus* so far, also shows 56% sequence identity to BnGT7 in this alignment. This indicates that BnGT7 is not an allele of S-GT and that there is only a distant relationship to UGT74B1 and UGT74C1. In group D, BnGT8 clusters with two functionally proven enzymes, UGT73C5 and UGT73C6. UGT73C5 was shown to glucosylate brassinosteroids (Poppenberger et al., 2005), whereas UGT73C6 was characterized as an UDP-glucose:flavonol-3-*O*-glycoside-7-*O*-glucosyltransferase (Jones et al., 2003). The other two classes of ESTs in this group, BnGT3 and BnGT10, were found to be related to UGTs involved in establishing resistance to the pathogen *Pseudomonas syringae* (UGT73B5 and UGT73B3, respectively; Langlois-Meurinne et al., 2005). BnGT11, clustered to group E, shows homology to UGT72E3, which was found to catalyze the *O*-glucosylation of hydroxycinnamates, hydroxycinnamoyl alcohols and aldehydes *in vitro* (Lim et al., 2005). The *A. thaliana* homologue of BnGT4, UGT90A1, was neither functionally characterized nor assigned to one of the UGT groups until now, although it was suggested that there is a relation to group D and G (Paquette et al., 2003).

### 3.1.4 Transcription analysis of glucosyltransferases

To assess the contribution of the cloned ESTs and HCA-GTs to seed metabolism, semi-quantitative RT-PCR with specific primers was performed (Figure 3-5). Gene expression was analyzed in several tissues starting with two days old seedlings and covering complete plant development. For this experiment 1 µg of total RNA was reversly transcribed for each tissue. The RT-reaction was diluted 25-fold and used for semi-quantitative RT-PCR.

It was found that the majority of potential UGTs is differentially expressed (Figure 3-5). Only BnGT6 was abundantly detected in all tissues analyzed. With regard to seed maturation, nearly constant high expression levels were found for BnGT1 and BnGT2, whereas transcriptional up-regulation was observed for BnGT3, BnGT7 and *UGT84A9*. However, for all these genes the highest expression was reached in early seedling stages. *UGT84A10* was expressed during seed ripening but more pronounced in flowers. Transcripts of BnGT4, BnGT8, BnGT9 and BnGT11 could hardly be detected during seed development, but were observed during flowering. Unlike the *napA* gene (Josefsson et al., 1987), a gene that encodes a major seed storage protein and shows the exemplary expression pattern of genes induced in seed maturation, all analyzed UGTs reach their highest expression level in different tissues than seeds.

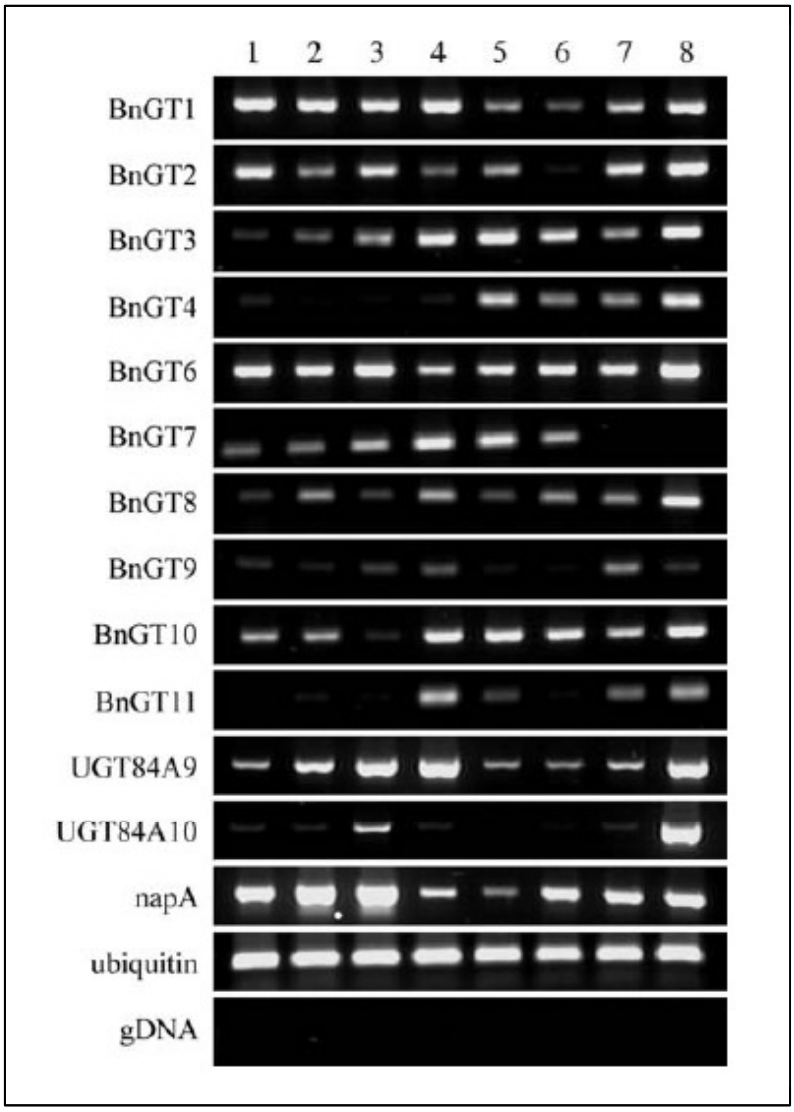



**Figure 3-5. RT-PCR analysis of *Brassica napus* UGT expression (Mittasch et al., 2007). Results for *napA* (GenBank Accession J02798; Josefsson et al., 1987), encoding a major seed specific storage protein, are included. Expression of ubiquitin is given as positive control. To demonstrate absence of genomic DNA contamination, RT-PCR with primers recognizing sequences upstream of *UGT84A9* was performed (gDNA). Analyzed tissues: *lane* 1, seeds at early globular to torpedo stage; *lane* 2, seeds at cotyledon stage; *lane* 3, seeds at well-developed mature embryo stage; *lane* 4, seedlings (2 days after sowing); *lane* 5 seedlings (7 days); *lane* 6, rosette leaves; *lane* 7, flower buds; lane 8, flowers.**

## 3.2 Characterization of *UGT84A9*

Since our investigation of *UGT84A9*-related genes and our analysis of seed ESTs indirectly revealed that *UGT84A9* itself plays the major role in sinapoylglucose formation during seed development, we decided to characterize *UGT84A9* further. Two main questions arose in this respect. First, how many copies of *UGT84A9* exist in the *B. napus* genome? Which subsequently led to the second question: Which of these *UGT84A9* gene copies are expressed during seed maturation?

### 3.2.1 Identification of two major sequence types of *UGT84A9*

*UGT84A9* full-length clones were obtained by PCR techniques using genomic DNA of *B. napus* and its ancestors, *B. rapa* and *B. oleracea*, and cDNA derived from RNA of seeds and flower tissue as templates. After sequencing, two sequence types could be clearly distinguished, referred to as sequence type 1 and sequence type 2, sharing a nucleotide sequence identity of 89% (Figure 3-6). Sequence type 1 resembles the previously characterizied cDNA (GenBank Accession number: AF287143; Milkowski et al., 2000a; Milkowski et al., 2004) and was mainly found in the genomes of *B. napus* and *B. oleracea*, and as transcript in seeds. Sequence type 2, however, was detected in the genomes of *B. napus* and *B. rapa*, and among transcripts from flowers.

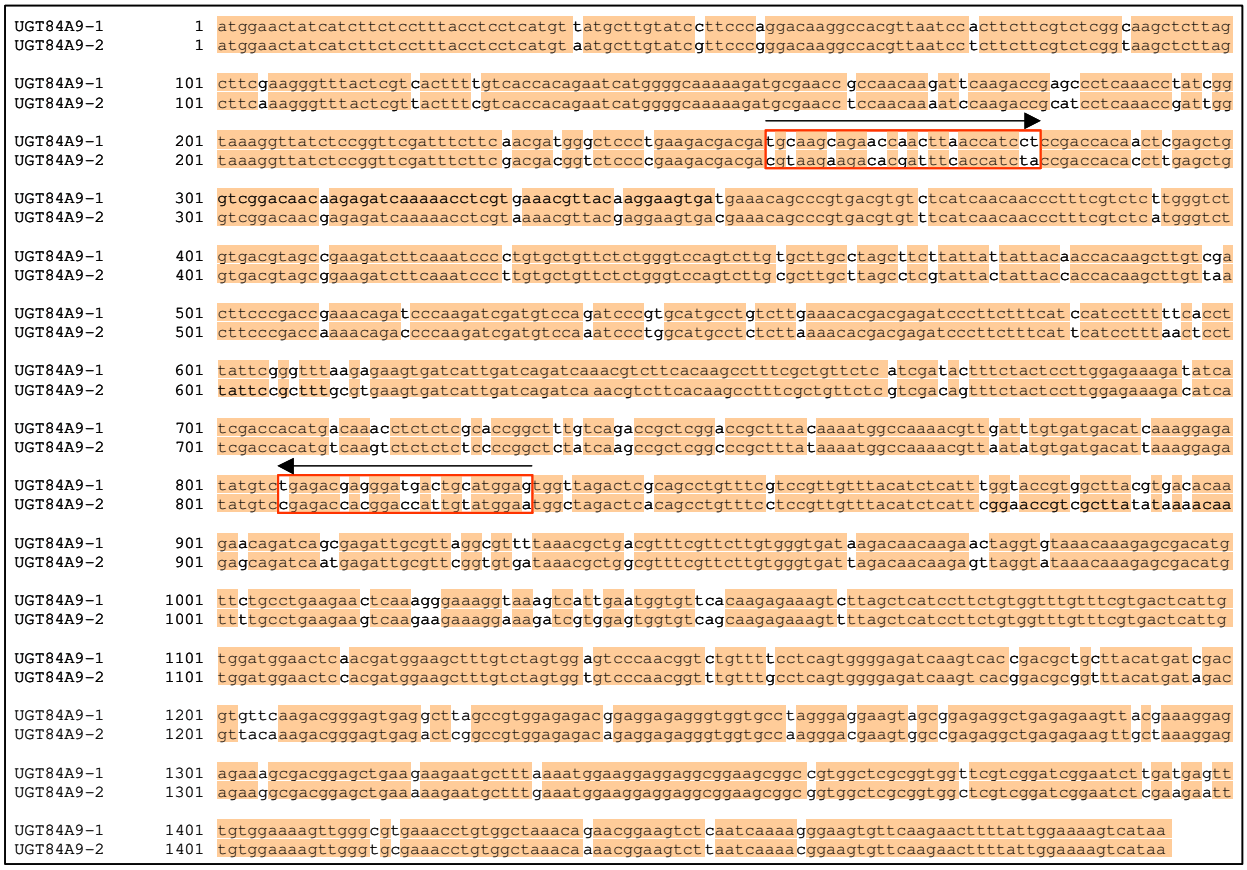



**Figure 3-6. Alignment of both *UGT84A9* sequence types. Sequence type 1 is indicated as UGT84A9-1, sequence type 2 as UGT84A9-2. Regions of sequence identity are shaded in orange. Primer sequences distinguishing both sequences types are boxed in red, primer orientation is indicated by the arrows above the sequence. The resulting amplicon has a size of 576 bp.**

Primer pairs were derived to distinguish between both sequence types by specific PCR amplification (Figure 3-6). The primers were named UGT84A9-1fw/rev, and UGT84A9-2fw/rev, respectively. Their sequences are given in Figure 3-6. Test PCRs using the cloned Brassica fragments as templates showed that these primers recognize specifically the appropriate sequence type in each case by producing an amplicon of 576 bp (Figure 3-7).

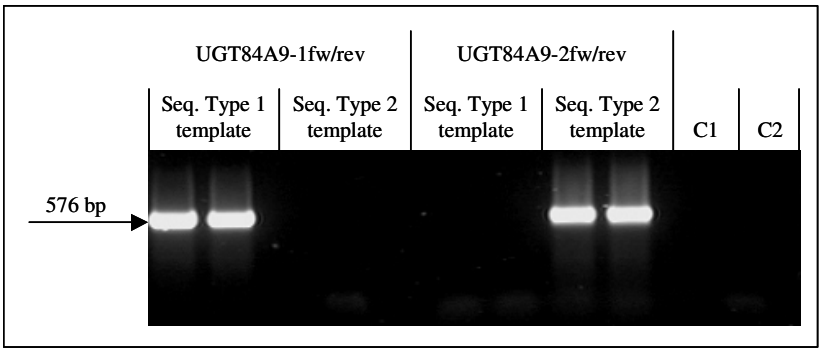



**Figure 3-7. PCR to assess the specificity of the primer pairs UGT84A9-1fw/rev , and UGT84A9-2fw/rev. Plasmids harboring the respective sequence type were used as template. C1, water control using sequence type 1 primer; C2, water control using sequence type 2 primer.**

### 3.2.2 Transcript analysis by semi-quantitative RT-PCR

Expression of both sequence types was first evaluated by semi-quantitative RT-PCR ( Figure 3-8). Specific primer pairs (chapter 3.2.1) were used to distinguish between the two sequence types of *UGT84A9*.

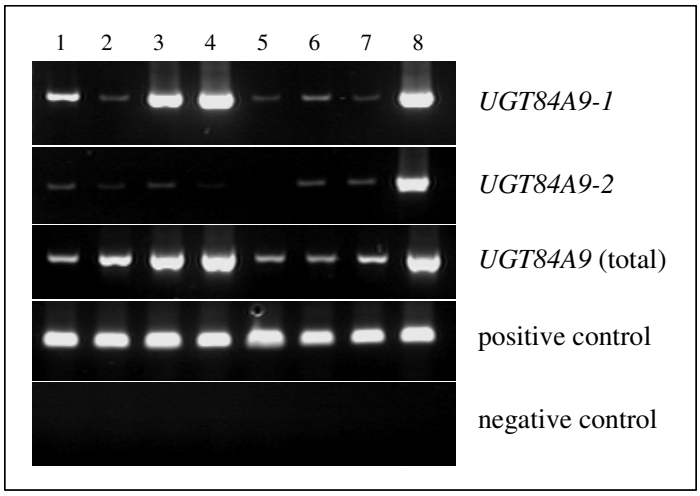



**Figure 3-8. Semi-quantitative RT-PCR of *UGT84A9* sequence types. Primer used: UGT84A9-1fw/rev, UGT84A9-2fw/rev, UGT84A9fw/rev; positive control – ubiquitin, negative control – genomic DNA. Primer pairs UGT84A9-1fw/rev, UGT84A9-2fw/rev bind specifically to the respective sequence type, whereas the primer pair UGT84A9fw/rev binds to all *UGT84A9* variants. Analyzed plant organs: *lane* 1, seeds at early globular to torpedo stage; *lane* 2, seeds at cotyledon stage; *lane* 3, seeds at well-developed mature embryo stage; *lane* 4, seedlings (2 days after sowing); *lane* 5, seedlings (7 days); *lane* 6, rosette leaves; *lane* 7, flower buds; *lane* 8, flowers.**

Tissues covering the complete plant development were analyzed ( Figure 3-8). 1 µg of total RNA for each tissue was used in cDNA synthesis. The resulting cDNA was diluted 25-fold and applied as template in a PCR reaction. It could be observed that *UGT84A9* sequence type 1 was increasingly expressed during the whole seed development, in very young seedlings and in flowers, whereas *UGT84A9* sequence type 2 transcripts were only detected in flowers. Thus, the gene expression pattern of sequence type 1 resembles the gene expression pattern of total *UGT84A9* in seed maturation and early seedling growth (Milkowski, et al., 2004).

### 3.2.3 Expression analysis by cDNA-AFLP

To confirm the results of the semi-quantitative RT-PCR, cDNA-AFLP (amplified fragment length polymorphism) analysis was done in cooperation with Dr. Sabine Mikolajewski (Bavarian State Research Center for Agriculture, Freising-Weihenstephan). For a cDNA-AFLP analysis, adapters are ligated to restriction fragments of double-stranded cDNA. The adapter sequence is used for primer design, with one to three specific bases added on the 3' end. Therefore, cDNA-AFLP is independent on amplification efficiency of individual primers (Bachem et al., 1996), and thus allows the direct comparison of the expression of both *UGT84A9* sequence types. For this analysis, two restriction enzymes were used, *Hin*dIII, a nucleotide hexamer cutter, and *Mse*I, a nucleotide tetramer cutter, which cut several times in the 1494 bp ORF of *UGT84A9*. For *UGT84A9* sequence type 1 this restriction results in two fragments appropriate for AFLP analysis, with a size of 120 bp and of 209 bp, respectively (Figure 3-9). Using the same restriction with *UGT84A9* sequence type 2 one 327 bp fragment can be detected in cDNA-AFLP analysis. The second fragment generated for this sequence type with a size of 7 bp is too short to give a signal in gel electrophoresis.

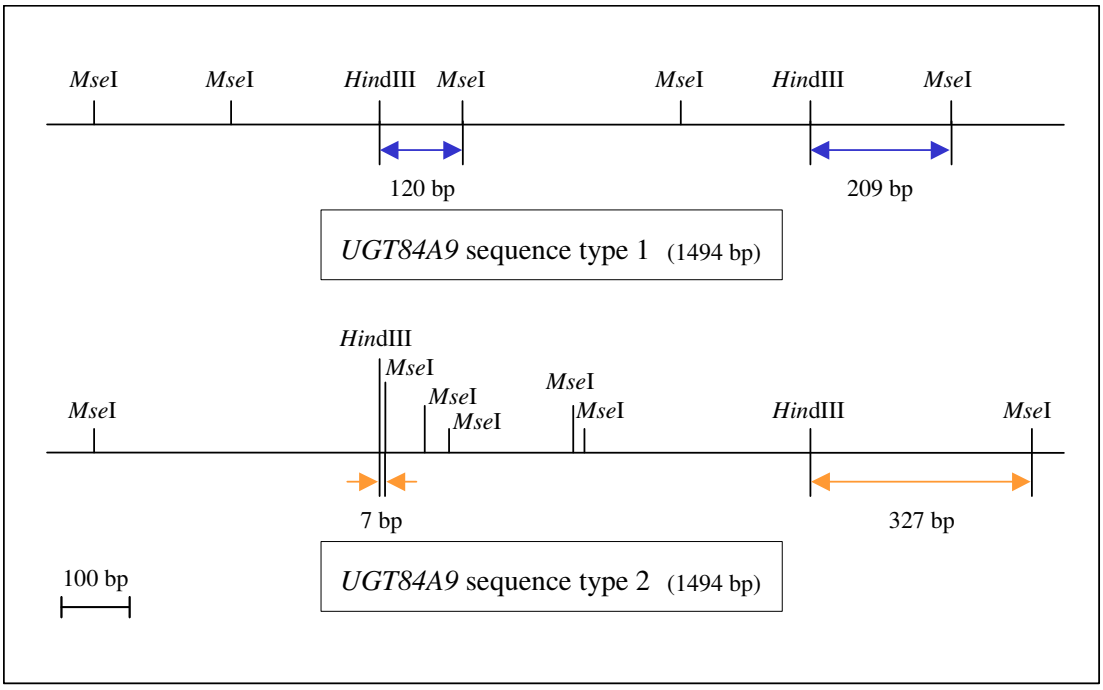



**Figure 3-9. Restriction map of *UGT84A9* sequence type 1 and 2 displaying the restriction sites for *Hin*dIII and *Mse*I. Restriction fragments potentially useful for cDNA-AFLP analysis are indicated in blue for sequence type 1 and in orange for sequence type 2.**

Due to the length of the two ligated adapter sequences [17 bp for the *Hin*dIII adapter and 14 bp for the *Mse*I adapter (Table 2-5)], fragments detected in cDNA-AFLP analysis are 31 bp larger than their respective restriction fragments. Figure 3-10 shows a part of a cDNA-AFLP gel. Specific primer combinations were used to amplify the second fragment of both sequence types (H90-M40 for *UGT84A9* sequence type 1 and H90-M64 for sequence type 2). The four tissues with the most prominent *UGT84A9* expression in RT-PCR were analyzed: flowers, seeds at cotyledon stage, seeds at mature embryo stage and two days old seedlings.

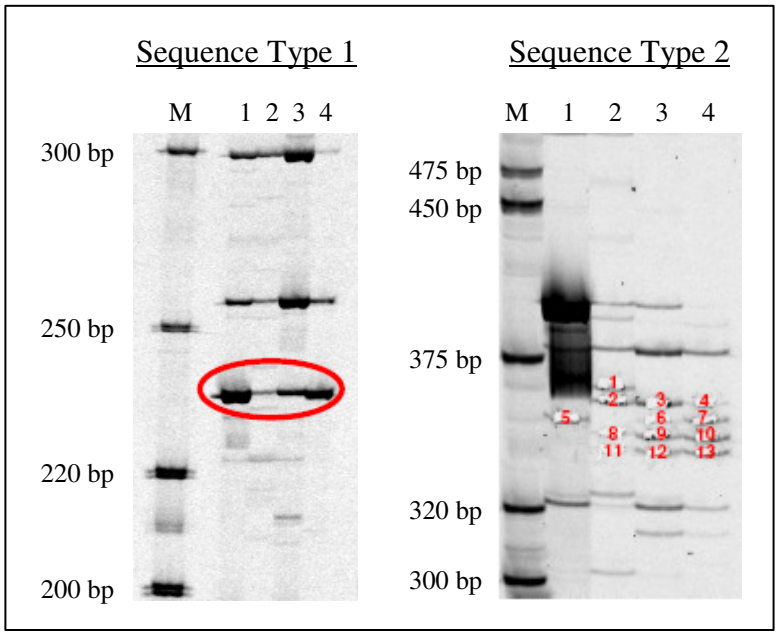



**Figure 3-10. Section of a cDNA-AFLP gel. TDFs (transcript derived fragments) marked with red numbers were eluted from the gel and their identity was confirmed by sequencing. Analyzed tissues: 1, seedlings (2 days after sowing); 2, flowers; 3, seeds at cotyledon stage; 4, seeds at well developed mature embryo stage.**

For sequence type 1, TDFs (transcript derived fragments) representing *UGT84A9* expression could be easily identified by their fragment size, whereas several weak TDFs emerged in the size range expected for *UGT84A9* sequence type 2 transcripts (Figure 3-10). The red numbered TDFs were eluted from the gel, amplified using their respective specific primers, and the resulting PCR products were sequenced. Sequencing confirmed that the red circled TDFs in the left part of Figure 3-10 represent *UGT84A9* sequence type 1 transcripts and that the red-numbered TDFs 2, 3 and 4 on the right gel are derived from *UGT84A9* sequence type 2 transcripts.

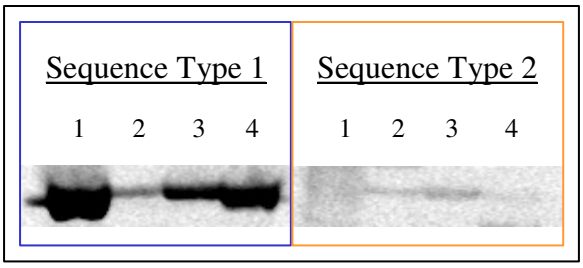



**Figure 3-11. Results of cDNA-AFLP analysis. Analyzed tissues: 1, seedlings (2 days after sowing); 2, flowers; 3, seeds at cotyledon stage; 4, seeds at well developed mature embryo stage.**

In each plant organ analyzed, sequence type 1 transcripts are more abundant than sequence type 2 transcripts, which could hardly be detected at all (Figure 3-11). Contrary to the semi-quantitative RT-PCR results, *UGT84A9* sequence type 1 gave only a faint signal in flower tissue, indicating an expression weaker than expected. The discrepancy was even greater for sequence type 2. While with semi-quantitative RT-PCR a distinct signal was detected, almost no transcript was found in the cDNA-AFLP analysis. A reason for this difference could be an abundant UGT transcript in flower tissues that is non-specifically amplified by the primers used in RT-PCR. Nevertheless, it was confirmed that *UGT84A9* sequence type 1 is the predominantly expressed sequence type during seed development.

### 3.2.4 Screening of a genomic BAC library

For the determination of *UGT84A9* gene copy numbers a genomic BAC library from *B. napus* was screened. The library was developed using genomic DNA from *B. napus* L. var. *napus* cv. Express which is almost completely homozygous.

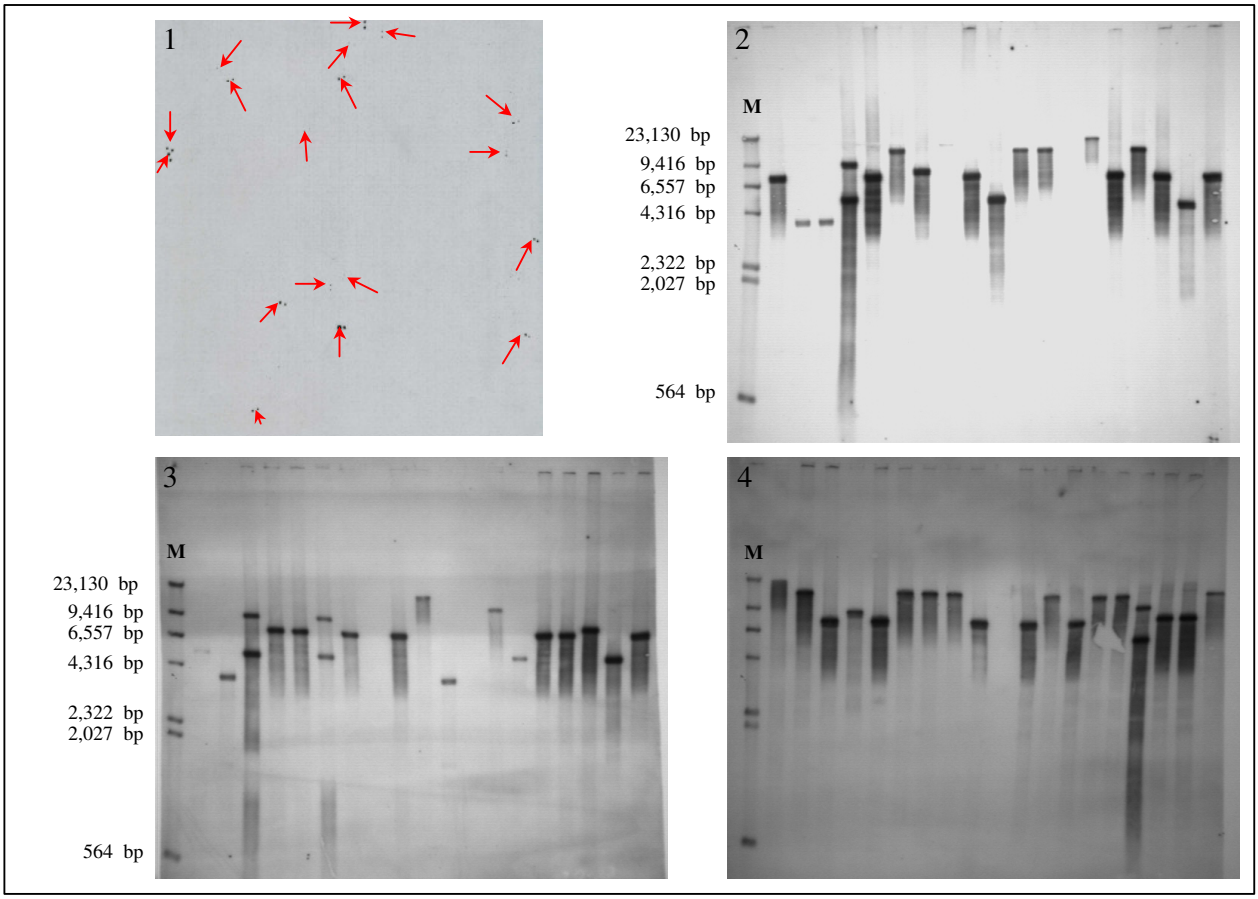



**Figure 3-12. Screen of a BAC library for *UGT84A9*-containing clones and subsequent confirmation of the isolated BACs. Picture 1 shows an exemplary X-ray film from one filter hybridization of the BAC library screen. 18 clones could be identified on this filter (red arrows). Overall, 3 filters were hybridized and 57 BACs isolated. Pictures 2-4 show the results of the subsequent Southern blot of the 57 BACs using the restriction enzyme *Nco*I. Analyzed BACs: picture 2, DIG-labeled size marker (M) and BACs 1-19; picture 3; DIG-labeled size marker (M) and BACs 20-38; picture 4, DIG-labeled size marker (M) and BACs 39-57. BACs 13, 27, 31 and 48 showed no hybridization with *UGT84A9* probe.**

Three filters containing 27,648 BAC clones, providing an estimated eightfold genome coverage, were hybridized with a probe derived from the *UGT84A9* sequence (sequence type 1) published by Milkowski and co-workers (2000a; Figure 3-12). This screen resulted in the identification of 57 BAC clones potentially carrying an *UGT84A9* ORF, numbered BAC 1 to BAC 57. To exclude false positive BAC clones, DNA from all 57 BAC clones was cleaved with *Nco*I, a restriction enzyme that does not cut inside the known *UGT84A9* ORFs, and was hybridized with the previously used *UGT84A9* probe. BAC 13, 27, 31 and 48 could not be detected in this Southern blot, and were therefore excluded from further analysis. It is noticeable, that BAC 4, 22, 25 and 54 contain an additional *Nco*I restriction site inside the putative *UGT84A9* frame.

### 3.2.5 BAC characterization by PCR

In order to get more information on the isolated BAC clones, PCR experiments were performed. Primer pairs for the amplification of the full-length *UGT84A9* ORF and primers distinguishing between the two different sequence types were applied.

**Table 3-2. Schematic overview of the isolated BAC clones. Primer pairs used for PCR-based characterization: FL_UGT84A9fw/rev, UGT84A9-1fw/rev, UGT84A9-2fw/rev.**

| | | | |
|---|---|---|---|
| Full-length ORF | full-length *UGT84A9* ORF | | partial *UGT84A9* ORF |
| | 38 | | 15 |
| Discrimination of sequence types | sequence type1 | sequence type 2 | not defined |
| | 27 | 17 | 9 |
| Summary | full-length *UGT84A9-1* | | full-length *UGT84A9-2* |
| | 23 | | 15 |

PCR analysis revealed, that the pool of 53 BACs contained 38 BACs with a full-length *UGT84A9* ORF (Table 3-2). It could be shown that 27 BACs carried *UGT84A9* sequences of sequence type 1 and 17 BACs of sequence type 2. 23 BACs containing a full-length *UGT84A9* sequence type 1 ORF, and 15 BACs with a full-length *UGT84A9* sequence type 2 ORF could be identified. Moreover, partial ORFs could be identified for each sequence type; four for sequence type 1 and two for sequence type 2. The genomic BAC library was constructed using a partial *Hin*dIII restriction. Since the ORFs of both sequence types contain two *Hin*dIII restriction sites, partial *UGT84A9* sequences can be expected on some of the BAC clones.

### 3.2.6 Copy number evaluation by BAC restriction analysis

To determine the number of loci for *UGT84A9* in the *B. napus* genome, a restriction analysis of 53 isolated BAC clones was carried out using Southern hybridization for the detection of fragments carrying *UGT84A9* sequences. For this analysis four restriction enzymes were used, *Hin*dIII, *Xho*I, *Sal*I and *Bgl*II. Selection of the restriction enzymes was done according to the already published *UGT84A9* sequence (Milkowski et al., 2000a). As this sequence was derived from a cDNA library screen, nucleotide substitutions potentially caused by PCR could be minimized. *Hin*dIII restriction of the ORF resulted in three fragments due to the two recognition sites for this enzyme. For the other three enzymes, two fragments were expected in each blot.

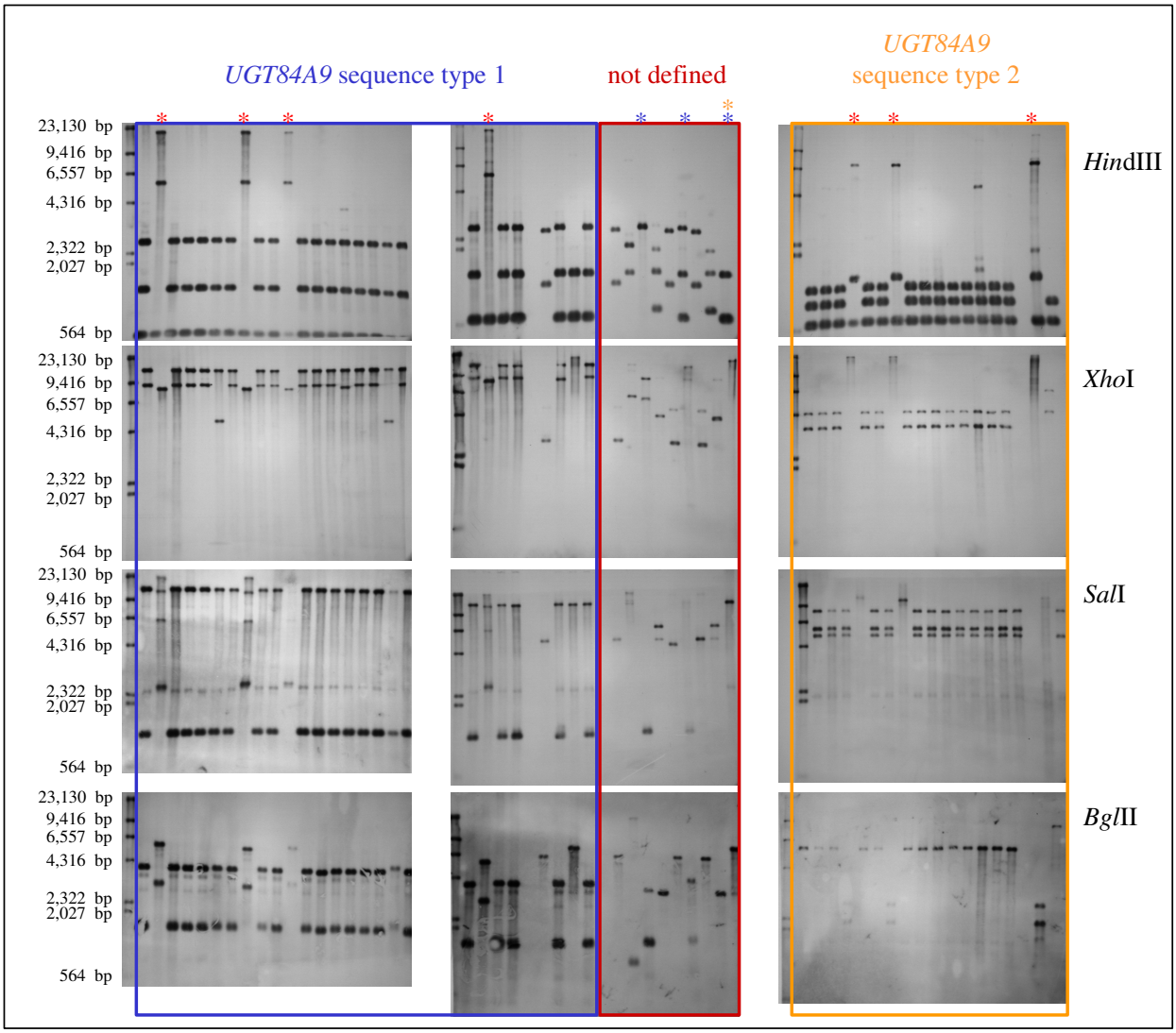



**Figure 3-13. Southern blot analysis of BAC clone restriction patterns. Restriction enzyme used is indicated at the right of the three respective gels. BACs were arranged according to the results of the PCR characterization. BACs expected to contain *UGT84A9* sequence type 1 are marked by a blue frame, BACs expected to carry *UGT84A9* sequence type 2 are marked by an orange frame, BACs not assigned to either sequence type are marked by a red frame. Aberrant restriction patterns are labeled with a red asterisk in both sequence types. Restriction patterns among the non-predefined BACs that could be assigned to one of the two *UGT84A9* sequence types are indicated by an asterisk in the appropriate color. The left panel of blots contains DIG-labeled marker and the type 1 BACs 1, 4, 5, 9, 15, 17, 19, 22, 23, 24, 25, 28, 34, 35, 36, 41, 43, 47 and 49; the middle panel contains DIG-labeled marker and the type 1 BACs 51, 54, 55, 56, 3, 7, 10 and 38, and the non-defined BACs 2, 8, 18, 20, 21, 26, 30, 33 and 37; the right panel contains DIG-labeled marker and the type 2 BACs 6, 11, 12, 14, 16, 32, 39, 40, 44, 45, 46, 50, 52, 53, 57, 29 and 42.**

For each *UGT84A9* sequence type two different restriction patterns could be consistently detected in restriction analysis for all four enzymes used, indicating two different loci for each sequence type in the haploid *B. napus* genome (Figure 3-13). These four loci were designated *UGT84A9a-d*, respectively. *UGT84A9a* and *UGT84A9b* belong to sequence type 1, whereas *UGT84A9c* and*UGT84A9d* are related to sequence type 2. For three BACs (7, 29 and 38) previously thought to carry an incomplete ORF, the complete restriction patterns of a specific *UGT84A9* locus were detected. BAC 7 and BAC 38 were found to harbor a complete *UGT84A9a* gene, whereas BAC 29 carries an complete *UGT84A9d* ORF. 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. A third BAC, 37, which most likely contains a partial *UGT84A9* ORF, showed restriction patterns similar to 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. To characterize the BACs further, the ends of BACs containing partial ORFs or of BACs that could not be classified to one of the four loci were sequenced. Furthermore, fragments containing *UGT84A9* sequences were cloned from BAC 1, 4, 6 and 14, representing the four different loci found in the restriction analysis.

### 3.2.7 Investigation of the non-defined BACs by BAC end sequencing

For further characterization, BACs with an incomplete putative ORF and BACs not yet assigned to one of the four *UGT84A9* loci were subjected to end sequencing. BAC end sequencing was done in cooperation with Prisca Viehöver and Prof. Bernd Weisshaar (Department of Biology, Bielefeld university). For 11 BACs end sequences were obtained using T7 and SP6m primers.

**Table 3-3. Results of BAC end sequencing. Abbreviations: At *Arabidopsis thaliana*, KBr genomic *Brassica rapa* libraries (Park et al., 2005; Lim et al., 2007)**

| BAC | PCR Characterization | Restriction Patterns | Closest homologues of BAC end sequence | Final Assignment |
|---|---|---|---|---|
| 3 | Sequence type 1 | Novel 1 | KBrB089H07 | putative BnHCA-GT |
| 10 | Sequence type 1 | *UGT84A9a* or *–d* | *UGT84A9a* | *UGT84A9a* |
| 42 | Sequence type 2 | *UGT84A9c* | *UGT84A9c* | *UGT84A9c* |
| 2 | None | Novel 1 | KBrH010F15, At F7B19 (chromosome 2) KBrB087B10 | putative BnHCA-GT |
| 8 | None | Novel 2 | KBrB089H07 | putative BnHCA-GT |
| 18 | None | *UGT84A9a* | *UGT84A9a* | *UGT84A9a* |
| 20 | None | Novel 2 | KBrB089H07 | putative BnHCA-GT |
| 21 | None | Novel 1 | KBrB089H07, At4g15475 | putative BnHCA-GT |
| 30 | None | Novel 1 | KBrB089H07, At4g15470 | putative BnHCA-GT |
| 33 | None | Novel 2 | KBrB089H07, At4g15890 | putative BnHCA-GT |
| 37 | None | *UGT84A9a* or *–d* | *UGT84A9a* | *UGT84A9a* |

Three of the eleven BACs (BAC 3, 10, 42) were assigned to either sequence type by PCR characterization (Table 3-3). For BAC 10, which was classified as sequence type 1, and BAC 42, classified as sequence type 2, the pre-characterization could be confirmed by restriction analysis and BAC end sequencing. For the assignment to an individual *UGT84A9* locus we used sequence alignments with all four *UGT84A9* sequences described in chapter 3.2.8 and the appropriate BAC end sequence. For BAC 3 a novel restriction pattern was found,

corresponding to the patterns observed in BACs that could not be grouped to one of the *UGT84A9* sequence types. In both BAC end sequences no *UGT84A9* sequence could be detected. The closest homologue found in a Blast search of the SP6m sequence was KBrB089H07, a genomic *Brassica rapa* BAC, published by the Brassica Genome Project (Korea).

The eight remaining BACs analyzed by end sequencing were not assigned to a sequence type by PCR. Restriction analysis revealed for BAC 18 a typical *UGT84A9a* pattern and for BAC 37 a pattern that could either point to *UGT84A9a* or *–d*. The BAC end sequences confirmed that BAC 18 harbors a partial *UGT84A9a* ORF. BAC 37 could also be assigned to *UGT84A9a*, carrying a partial ORF as well. The six remaining BACs all showed novel restriction patterns in the Southern blot (Figure 3-13) that could be grouped into two classes, named Novel 1 and Novel 2. For five of the six BACs the closest homologue to their end sequences is again the genomic *B. rapa* BAC KBrB089H07. Furthermore, *A. thaliana* genes located near *UGT84A1* (*At4g15480*), *UGT84A3* (*At4g15490*) and *UGT84A4* (*At4g15500*) were detected as closely related to these BAC ends. Sequence analysis of KBrB089H07 revealed that it harbors a putative *B. rapa* hydroxycinnamate UGT (BrHCA-GT). The closest *A. thaliana* homologue of this putative BrHCA-GT is *UGT84A1*, showing 85% sequence identity on nucleotide level and 89% on amino acid level. Since KBrB089H07 was found as closest homologue to BAC end sequences of BACs showing either restriction pattern Novel 1 or Novel 2, probably both *B. napus* homologues of the putative BrHCA-GT, one from the A-genome and one from the C-genome, are present in the isolated BACs.

### 3.2.8 Sequence analysis of the four *UGT84A9* ORFs

To determine the exact sequence of the four *UGT84A9* loci, fragments carrying *UGT84A9* ORFs were cloned from BACs 1, 4, 6 and 14 and sequenced. Additionally, PCR products of the four *UGT84A9* loci amplified from the same BACs were sequenced and these sequences were then used as starting points to read the surrounding sequences on each BAC. Direct sequencing of BAC DNA was done in cooperation with Prisca Viehöver and Prof. Bernd Weisshaar (Department of Biology, Bielefeld university).

Comparison of the four obtained *UGT84A9* sequences revealed that sequences of the same type are very similar to each other, 97-98% identity on nucleotide level, whereas sequences of the different types are much less similar (Table 3-4). They share only an identity of 87-88% to each other. An alignment of the deduced amino acid sequences showed no amino acid exchanges inside the highly conserved PSPG box.

**Table 3-4. Sequence comparison of the four *UGT84A9* loci. *UGT84A9a* and *–b* belongs to sequence type 1, *UGT84A9c* and *–d* to sequence type 2. Percentage identity on nucleotide level is shown.**

| | *UGT84A9a* | *UGT84A9b* | *UGT84A9c* | *UGT84A9d* |
|---|---|---|---|---|
| *UGT84A9a* | - | 97% | 88% | 88% |
| *UGT84A9b* | | - | 87% | 88% |
| *UGT84A9c* | | | - | 98% |
| *UGT84A9d* | | | | - |

To determine the origin of the four *UGT84A9* loci, full-length SGT clones were obtained by PCR using genomic DNA from the progenitor species of *B. napus*, *B. rapa* and *B. oleracea* as template, representing the A- and C-genome, respectively. The older SGT nomenclature was applied here, since no UGT names for the *B. rapa* and *B. oleracea* genes were available so far. Primer pair FL_UGT84A9fw/rev was used for amplification. Altogether, ten full-length clones from *B. rapa* and nine full-length clones from *B. oleracea* were sequenced. From these sequences, two consensus *SGT* sequences were identified for each *Brassica* genome. The consensus sequences were used together with the four *UGT84A9* sequences and the other known HCA-GT like sequences from *B. napus* and *A. thaliana* to construct an unrooted neighbor-joining similarity tree (Figure 3-14).

Tree topology revealed that each *UGT84A9* sequence type is closely related to one *B. rapa SGT* sequence and one *B. oleracea SGT* sequence, respectively (Figure 3-14). The closest homologue of *UGT84A9a* is the *B. oleracea SGT* sequence similar to sequence type 1 (*BoSGT I*), whereas *UGT84A9b* seems to be closely related to the corresponding *B. rapa SGT* sequence (*BrSGT I*). For sequence type 2, the relationship of the sequences could also be elucidated. *UGT84A9c* was detected in one cluster together with the *SGT* sequence of type 2 from *B. oleracea* (*BoSGT II*), whereas *UGT84A9d* grouped with the *B. rapa SGT* sequence type 2 (*BrSGT II*). So the ancestry of the four *UGT84A9* loci became apparent: *UGT84A9a* and *–c* were both derived from the *Brassica* C-genome, whereas *UGT84A9b* and *–d* were originated from the A-genome. Obviously, both sequence types were already present in the diploid progenitors of *B. napus*, *B. rapa* and *B. oleracea*. The two sequence variants of *UGT84A9* and their *Brassica* homologues are closer related among themselves than to the Arabidopsis *SGT* (*UGT84A2*), indicating an evolution of these sequences after the separation from the Arabidopsis branch.

During analysis of the BAC ends of genomic *UGT84A9* BACs, a putative *B. rapa* HCA-GT was identified as closest known homologue to an unknown *B. napus* HCA-GT supposed to be harbored by BACs that could not be assigned to an *UGT84A9* locus. It was shown that this BrHCA-GT is closely related to the *A. thaliana* gene *UGT84A1* and the *B. napus* gene *UGT84A11* identified in this work. *UGT84A10*, the other UGT isolated in this study, was found to be related to the *A. thaliana* genes *UGT84A3* and *–4*. Closest homologue to this gene is *UGT84A4*.

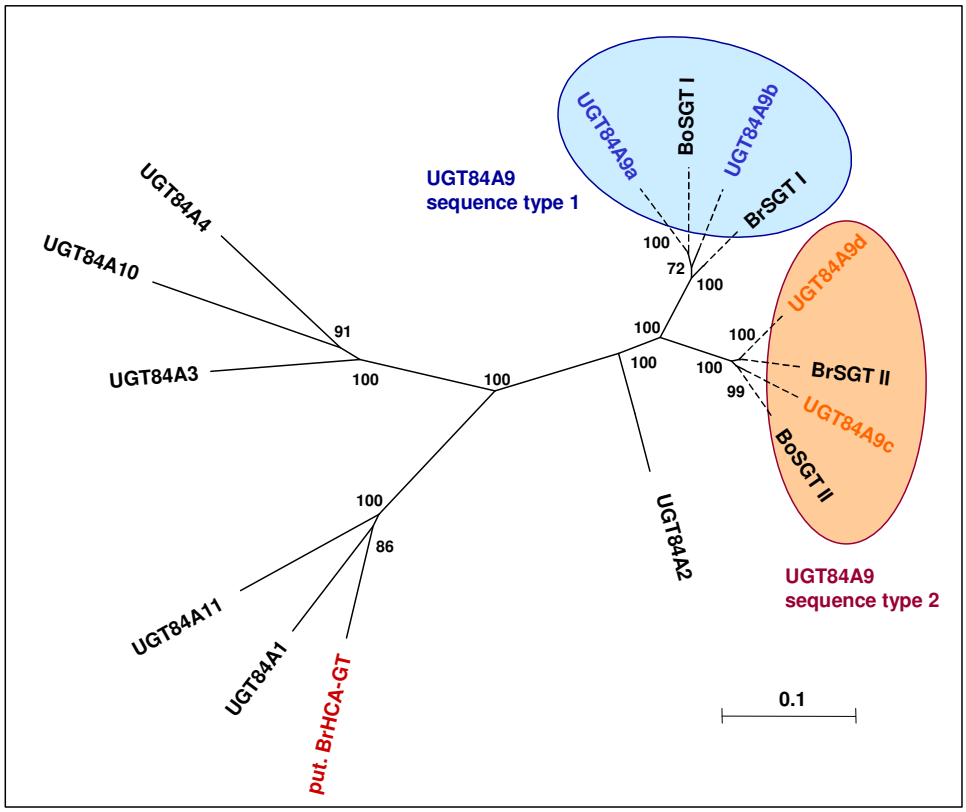



**Figure 3-14. Unrooted neighbor-joining similarity tree of known HCA-GT sequences. Distance estimation was done by Cantor and Jukes algorithm (1969); insertion and deletions were taken into account. Bootstrap values above 50 are given. Scale represents 0.1 fixed mutations per site. Dashed lines were prolonged after tree construction for clarity. *UGT84A9* sequences grouped to sequence type 1 are marked in blue and closely related sequences are highlighted with a light blue ellipse. *UGT84A9* sequences belonging to sequence type 2 are given in orange and closely related sequences are highlighted with an orange ellipse. Abbreviations: Bo, *Brassica oleracea*; Br, *Brassica rapa*; HCA-GT, hydroxycinnamate glucosyltransferase; SGT, UDP-glucose:sinapate glucosyltransferase (EC 2.4.1.120); put., putative.**

To confirm the number of *UGT84A9* loci, PCR product sequencing was performed. A primer pair was derived from the sequences obtained from BACs 1, 4 ,6 and 14, recognizing all four *UGT84A9* loci. These primers, designated UGT84A9fw7 and –rev7, were used to amplify a 532 bp product, located between nucleotides 562 and 1094 of each ORF. All BACs carrying *UGT84A9* were used as template, with exception of the aforementioned BACs 1, 4, 6 and 14, which were already characterized by sequencing. Altogether, 42 BACs were subjected to PCR, from which 39 amplicons were derived. PCRs using BAC 10, 37 and 42 as template failed to produce a product. These BACs were determined to carry a partial *UGT84A9* ORF during restriction analysis. The 39 obtained PCR products were sequenced, and the resulting sequences were then compared to the four known *UGT84A9* sequences. For each BAC, its previous classification could be confirmed. No additional polymorphism was detected.

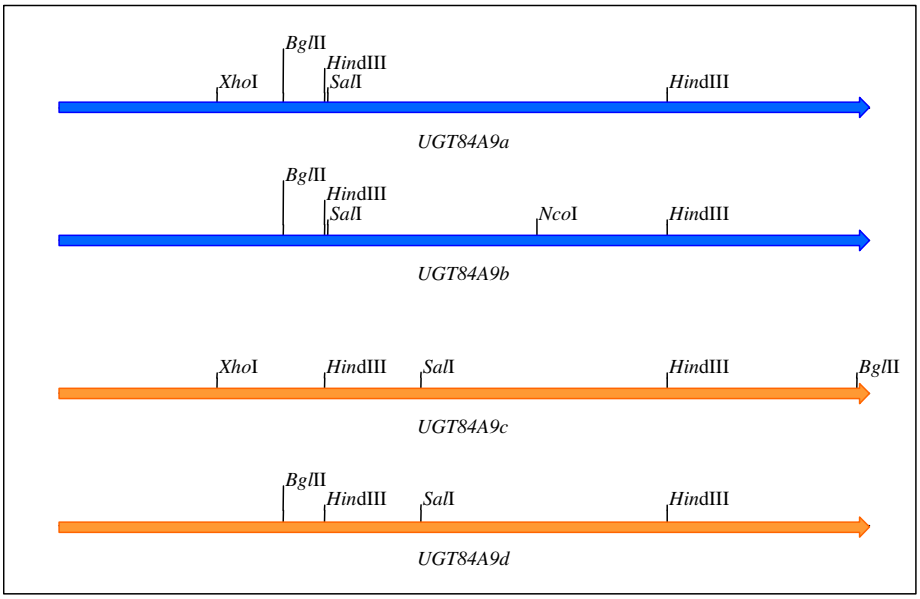



**Figure 3-15. Restriction map of the four *UGT84A9* ORFs. Restriction sites of enzymes used for BAC characterization by Southern blot analysis are indicated. ORFs characterized as sequence type 1 are drawn in blue, whereas ORFs belonging to sequence type 2 are marked in orange.**

For further characterization of the *UGT84A9* ORFs and to reinforce the BAC restriction analysis, restriction maps of the four ORFs were analyzed (Figure 3-15). For *Hin*dIII, the three fragments, due to the two internal restriction sites, had been detected in all four *UGT84A9* ORFs (Figure 3-13). *Xho*I has one recognition site in the *UGT84A9* ORFs derived from the C-genome, *UGT84A9a* and *–c* but none in the other two loci. This pattern was confirmed in the Southern blot characterization of the BAC clones, where for BACs carrying *UGT84A9a* and *–c* two hybridized fragments had been detected, but only one for the other two loci. In the *Sal*I restriction analysis, two fragments should have been detected for each loci. Southern blot analysis had shown this pattern for three of the four loci, but for *UGT84A9c* three bands has been found. The reason for this could be a small insertion of *UGT84A9c* sequences elsewhere in the *B. napus* genome. Sequencing of these three fragments could solve the puzzle. *Bgl*II restriction would have shown two fragments for all clones. As, however, the *Bgl*II restriction site in *UGT84A9c* is very close to the 3' end of the ORF and the *UGT84A9a* ORF was used as probe, the resulting 24 bp overlap between 3'fragment and probe was not sufficient for hybridization and detection of the *UGT84A9c* 3'fragment. Thus only the 5' fragment of *UGT84A9c* was detected, whereas for the other three loci both fragments has been found.

### 3.2.9 Investigation of the genomic context of the four *UGT84A9* loci

To characterize the four *UGT84A9* gene loci further, the up- and downstream regions of the ORFs were sequenced, by reading directly on the BACs and by sequencing BAC sub-fragments containing *UGT84A9* sequences. For all four loci genomic sequences were obtained, although the sizes of the sequenced fragments varied greatly. From BAC 1, harboring *UGT84A9a*, one fragment was cloned, including the complete *UGT84A9a* ORF, whereas for the other three loci two sub-fragments, one covering the upstream region, one the downstream one, were combined to gain the complete information. Altogether, a 7.7 kb

fragment harboring *UGT84A9a* was sequenced, 9.4 kb for *UGT84A9b*, 3.4 kb for *UGT84A9c* and 20.3 kb for *UGT84A9d*. The fragments containing *UGT84A9a* and *–b* were obtained using *Nco*I restrictions. The fragment harboring the 3' end of *UGT84A9b* and downstream sequences could not be sequenced completely due to a low GC content of the sequence and the occurrence of highly repetitive sequences. For the cloning of the *UGT84A9c* and *–d* fragments double digestions with *Nco*I and *Sal*I or alternatively *Sph*I were used, since the respective *Nco*I fragments were too large. The sequence information for the upstream genomic context of *UGT84A9c* came from a *Nco*I/*Sph*I and a *Nco*I/*Sal*I fragment. Neither cloned fragment could be sequenced completely, since they could not be stably stored as *E. coli* freeze stocks. The sequence information for the downstream region was obtained from a *Nco*I/*Sal*I restriction fragment. For the characterization of *UGT84A9d* two *Nco*I/*Sal*I fragments were sequenced, one covering the upstream region and one covering the downstream one.

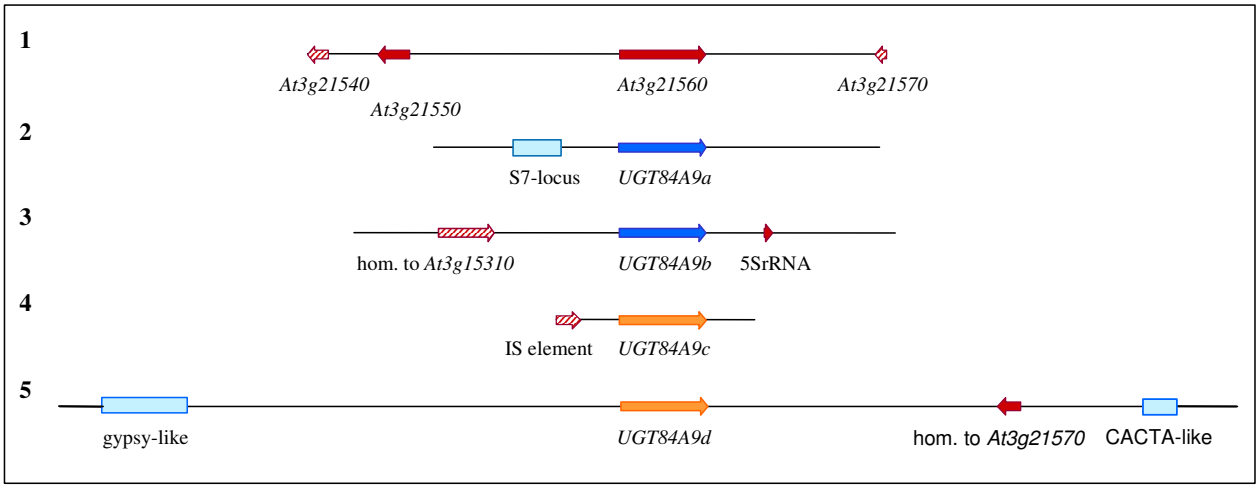



**Figure 3-16. Schematic overview of four BAC fragments harboring the four *UGT84A9* ORFs and comparison to their homologous Arabidopsis region. Red arrows indicate complete ORFs, hatched red arrows picture incomplete ORFs. *UGT84A9* ORFs are drawn as blue arrows (sequence type 1) or orange arrows (sequence type 2). Blue bars highlight regions of homology. 1, 10 kb section of *Arabidopsis thaliana* chromosome 3, features *At3g21540* (362 – 1 bp complementary strand, incomplete ORF), *At3g21550* (1768-1214 bp complementary strand), *At3g21560* (*UGT84A2*, 5390-6880bp), *At3g21570* (10000-9809bp complementary strand, incomplete ORF); 2, BAC 1 fragment (7.7 kb), features region homologous to S7-locus from *B. oleracea* (1365-2198 bp), *UGT84A9a* (3208-4701 bp); 3, BAC 4 fragment (9.4 kb), features incomplete ORF homolog to *At3g15310* (1458-2434 bp), *UGT84A9b* (4581-6074 bp), 5SrRNA homologous to *B. nigra* gene II 5SrRNA (7099-7225 bp); 4, BAC 6 fragment (3.4 kb), features IS element (1-431 bp, incomplete), *UGT84A9c* (1118-2611 bp); 5, BAC 14 fragment (20.3 kb) features gypsy-like region showing weak homology to *At3g30418* annotated as gypsy-like retrotransposon (737-2233 bp), *UGT84A9d* (9688-11181 bp), ORF homologous to *At3g21570* (16641-16225bp complementary strand), CACTA-like region showing weak homology to *At3g32226* annotated as CACTA-like transposase family (18741-19304 bp).**

Blast analysis using NCBI and TAIR databases revealed that there is little collinearity between the *A. thaliana* chromosome 3 region harboring *UGT84A2* (*At3g21560*) and the four characterized BAC fragments containing the four *UGT84A9* loci (Figure 3-16). With the exception of the homolog of *At3g21570* on the fragment of BAC 14, none of the neighboring *A. thaliana* genes was found. In the *A. thaliana* genome, upstream and downstream of *At3g21560* large non-coding regions are located. It is noticeable that in three out of the four BAC fragments sequences were detected that show similarities to transposons. On the

fragment cloned from BAC 1 a 1.2 kb region (1365-2198 bp) shows 91% sequence identity to a section of the non-coding A-region of the *B. oleracea* self incompatibility locus S7 (Fujimoto et al., 2006). The non-coding B-region of this locus was shown to contain transposable elements. A WU-Blast search in the TAIR database revealed 60% sequence identity for the same region of BAC 1 fragment (1393-2181 bp) to *At5g38383* was found, a gene annotated as belonging to the gypsy-like retrotransposon family Athila. Upstream of *UGT84A9c* on the fragment cloned from BAC 6 an IS element (insertion sequence element, 1-431 bp) was found. In the *A. thaliana* genome five genes (*At1g76035*, *At3g25185*, *At1g47875*, *At1g35666*, *At1g36930*) were found which share 99-100% sequence identity with this IS element. The whole region of homology is given, although the homology of the last two mentioned genes ended at 419 bp of the BAC 6 fragment, because their ORFs ended there. These genes were all annotated as transposable elements and are thought to be cloning artefacts derived from the sequencing of *A. thaliana*. Maybe this element is a cloning artefact as well, but on the other hand, it is possible that the IS element is a genuine part of the *B. napus* genome. An insertion of this IS element on BAC 6 during BAC library construction is unlikely, since BAC 6 exhibited no aberrant restriction pattern in the Southern blot characterization of the complete BAC collection (Figure 3-13). The upstream region of *UGT84A9c* was characterized by sequencing two restriction fragments of BAC 6, one derived from a restriction with *Nco*I/*Sal*I and another from a *Nco*I/*Sph*I restriction. The *Nco*I/*Sph*I fragment harboured the IS element sequence, whereas for the *Nco*I/*Sal*I fragment further sequencing was not successful after the start of the IS element (431 bp). As mentioned earlier, both cloned fragments could not be sequenced completely. The fact that *E. coli* possesses very similar IS elements could explain why the cloned fragments were not stable in *E. coli* cultures. On the BAC 14 fragment two regions with similarities to transposon-like sequences were identified during a WU-Blast search on the TAIR database. A region upstream of *UGT84A9d* (737-2233 bp) shows 53% sequence identity to *At3g30418*, a gene described as beloning to the gypsy-like retrotransposon family Athila. Downstream of *UGT84A9d* a second region was identified which shows similarities to four *A. thaliana* CACTA-like transposase genes (18741-19304 bp). The 3' ends of these genes (*At2g12990*, *At1g36270*, *At3g32226*, *At3g43126*) exhibit 60-61% sequence identity to the BAC 14 region.

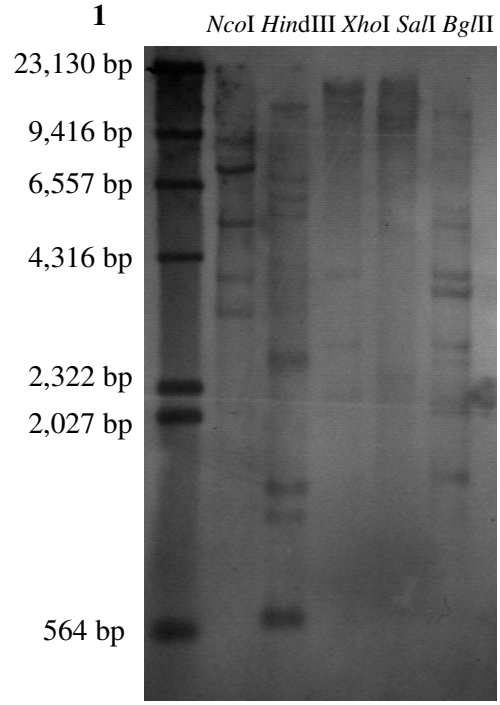
### 3.2.10 Analysis of YN01-429 by Southern hybridization

The genomic organization of the *UGT84A9* genes was elucidated in *B. napus* L. var. *napus* cv. Express. Next it was tested, whether these results could be applied to another breeding cultivar, YN01-429. YN01-429 is a yellow-seeded spring cultivar from Canada (AAFC, Canada). Southern blot experiments were conducted to determine the number of *UGT84A9* loci and to compare the fragment sizes as an indicator for the genomic organization. For this study the nearly homozygous F9 generation of this line was used.

In the genomic Southern blot analysis of *B. napus* cv. YN01-429 the same restriction enzymes were applied as for the BAC characterization of *B. napus* cv. Express (chapter 3.2.6) to ensure comparability of the results (Figure 3-17). As probe, the full-length ORF of *UGT84A9a* was used. Detected fragments were regarded as positive, when they appeared in two independent experiments. The number of detected fragments was the same for both cultivars (YN01-429

and Express) in four of five restriction analyses. Only for the *Sal*I restriction, the number of fragments was lower in YN01-429 than in Express. This is most likely due to GpC methylation of the *Sal*I recognition site in the genomic DNA. GpC methylation does not occur in BAC DNA, which was used in the experiments characterizing the cultivar Express, thus explaining the lower fragment number in YN01-429. Southern blot analysis of *B. napus* cv. Express genomic DNA restricted with *Sal*I was not conducted. For the *Bgl*II restriction three weak signals above 5.4 kb were detectable (Figure 3-17). Theses signals could not be confirmed in a second Southern blot hybridization.

**1**




**2**

| | **Number of fragments** | |
|---|---|---|
| Restriction enzyme | YN01-429 | Express |
| *Nco*I | 5 | 5 |
| *Hin*dIII | 9 | 9 |
| *Xho*I | 6 | 6 |
| *Sal*I | 5 | 8 |
| *Bgl*II | 7 | 7 |

**Figure 3-17. Genomic Southern blot analysis of *B. napus* cv. YN01-429. 1, results of Southern hybridization. Full-length *UGT84A9a* ORF was used as probe. Restriction was conducted with the same enzymes used for BAC characterization of *B. napus* cv. Express. 2, Comparison of the fragment numbers between the *B. napus* cultivars Express and YN01-429.**

Sizes of the fragments derived from Southern blot experiments of both cultivars were compared (Table 3-5). Fragments derived from restriction analyses using *Sal*I were excluded from this study, since *Sal*I does not work properly with genomic DNA. Results obtained using *Nco*I, *Hin*dIII, *Xho*I and *Bgl*II restrictions were evaluated. Fragment sizes were estimated by comparison to a pre-labeled size marker [DIG-labeled DNA Molecular Weight Marker II (Roche)].

In all restrictions, fragments of the same size were found in both cultivars, but also a considerable number fragments differing in size (Table 3-5). At least 40% of the detected fragments for a given restriction differs in size between the two cultivars. This indicates frequent occurrence of sequence polymorphisms between Express and YN01-429, shifting restriction sites of the different enzymes.

**Table 3-5. Comparison of fragment sizes derived from Southern blot experiments between the cultivars Express and YN01-429.**

| | *Nco*I | *Hin*dIII | *Xho*I | *Bgl*II |
|---|---|---|---|---|
| Fragments found in both cultivars [kb] | 9.2, 7, 5.2 | 6, 2.5, 1.4, 1.2, 0.6 | 14 | 4.1, 2.6, 2.5, 2.1 |
| Fragments found only in Express [kb] | 14, 18 | 18, 6.9, 1.5, 0.9 | 20, 9, 8.5, 4.3, 3.7 | 6, 5.8, 1. |
| Fragments found only in YN01-429 [kb] | 4, 3.2 | 14, 6.5, 5.5, 2.4 | 18, 12, 4.1, 2.9, 2.2 | 5.4, 3.9, 1.5 |

In summary, it can be stated, that the number of *UGT84A9* loci remained constant between both cultivars, but there are sequence polymorphisms observable. This complicates a transfer of sequence information between the two cultivars for applications like TILLING.

## 3.3 Seed-specific silencing of *UGT84A9*

### 3.3.1 Transcript levels in seeds and seedlings

Seed-specific silencing of *UGT84A9* has been established via dsRNAi; total sinapate ester content could be reduced to 24% in T3 seeds (Hüsken et al., 2005). To confirm the stability of the dsRNAi silencing effect, transcript levels in seeds and seedlings from plants of transgenic offspring generations were investigated by semi-quantitative RT-PCR. Another interesting question was, how the previously isolated alternative BnHCA-GT, *UGT84A10*, would be expressed in the transgenic background. Two homozygous transgenic *UGT84A9a*i lines were chosen, that had shown a strong suppression phenotype in previous experiments, lines 13 and 44. As a wild type control *B. napus* cv. Drakkar designated as line 142 was used.

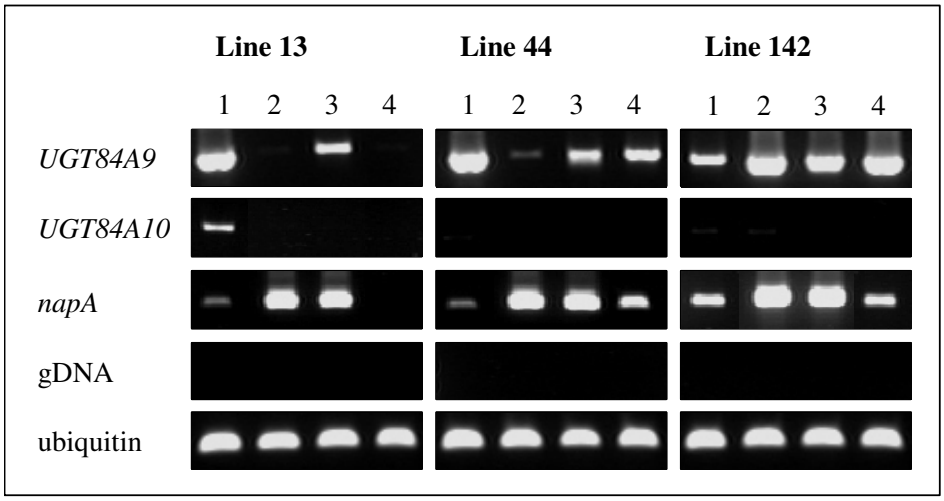



**Figure 3-18. *UGT84A9* transcript levels in *UGT84A9a*i lines of *B. napus*. Lines 13 and 44 are *UGT84A9a*i lines, whereas line 142 served as wild type control. *UGT84A9* transcripts were investigated using *UGT84A9*fw7/rev7 primers. *NapA* expression is used to demonstrate the expression of the napin promoter driven dsRNAi construct. PCR with primers recognizing genomic sequences upstream of *UGT84A9a* was used to demonstrate the absence of genomic DNA in the samples (gDNA). Expression of ubiquitin is given as a positive control. Analyzed plant organs: 1, flowers; 2, seeds at well developed mature embryo stage; 3, seeds at desiccation stage; 4, seedlings (2 days after sowing).**

First, *UGT84A9* transcript levels of all sequence types were investigated in flowers, developing seeds and seedlings of *B. napus* by semi-quantitative RT-PCR (Figure 3-18). This method allows to obtain a rough insight into transcriptional regulation and thus, allows to derive major tendencies of gene expression.

In the two seed developmental stages and to a lesser extent in the two-day-old seedlings *UGT84A9* expression is clearly reduced in the dsRNAi plants. It was shown that *UGT84A9* gene expression is suppressed stronger in line 13 than in line 44. However, *UGT84A9* transcripts were still detected in the desiccating seeds of both *UGT84A9a*i lines. This could be due to a high abundance of this transcript, so that the dsRNAi effect is not strong enough to silence the gene completely. *UGT84A10* is a homologue of the *A. thaliana* genes *UGT84A3* and *UGT84A4*, both capable of catalyzing the formation of sinapoylglucose (Milkowski et al., 2000b), and could provide redundancy for *UGT84A9*. However, no increased levels of *UGT84A10* expression were found in the transgenic lines compared to the wild type, except for a weak signal in flowers of line 13. In the flowers of the transgenic plants *UGT84A9* appears to be expressed strongly, compared to the relative weak signal in the wild type line 142. This could be due to the limitations of the method. To monitor the expression of the dsRNAi cassette, which is driven by a napin promoter, *napA* expression was measured. In both transgenic lines (line 13 and line 44) the reduction of *UGT84A9* transcript correlates with the increase of *napA* expression in seeds at the mature embryo stage. In two-day-old seedlings differences in the *napA* transcript abundance could be observed between line 13 and the other two lines. This should be due to biological variation, since the transgenic effect should not impair *napA* gene expression. The *napA* transcripts detected in the seedlings of line 44 and the wild type line 142 are most likely residual mRNAs of seed development, because *de novo* transcription of napin genes in seedlings has not been reported until now.

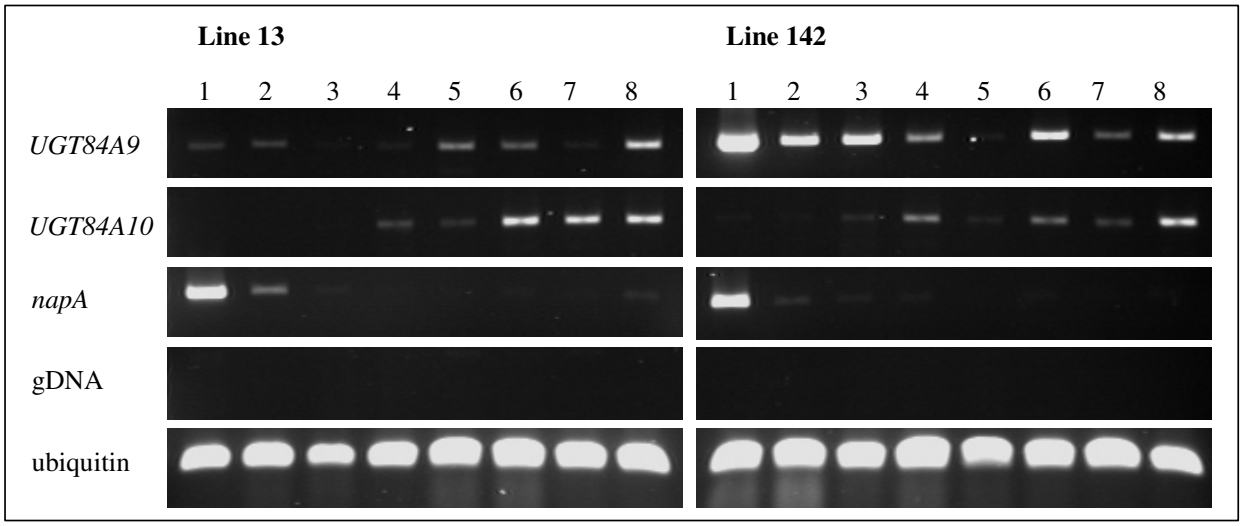



**Figure 3-19. *UGT84A9* and *UGT84A10* transcript levels in seedlings of *UGT84A9a*i *B. napus* plants. Line 13 is an *UGT84A9a*i line, whereas line 142 served as wild type control. *UGT84A9* transcripts were amplified using UGT84A9fw7/rev7 primers. *NapA* expression is used to demonstrate the expression of the dsRNAi construct. PCR with primers recognizing genomic sequences upstream of *UGT84A9a* was used to demonstrate the absence of genomic DNA in the samples (gDNA). Expression of ubiquitin is given as a positive control. Analyzed plant organs: 1, seedlings (2 days after sowing); 2, seedlings (4 days); 3, upper parts of seedlings (7 days); 4, upper parts of seedlings (10 days); 5, upper parts of seedlings (14 days); 6, roots of seedlings (7 days); 7, roots of seedlings (10 days); 8, roots of seedlings (14 days).**

To investigate, for how long the suppression of *UGT84A9* is sustained after germination transgenic *UGT84A9a*i seedlings were analyzed over a time period of fourteen days after sowing and compared to the wild type (Figure 3-19). Seedlings harvested seven days after sowing and at later time points were divided into upper parts and roots, and investigated separately. In the wild type line (line 142), after a strong signal in two-day-old seedlings a decrease of *UGT84A9* expression was observed. In contrast, very low amounts of *UGT84A9* transcripts were detected in the transgenic line (line 13). A slight increase of expression was found in the upper parts of forteen-day-old seedlings, indicating that the dsRNAi effect diminishes. *UGT84A9* transcripts were not only detected in the emerging primary leaves, but also in the cotyledons of fourteen-day-old seedlings. In general, the suppression state was kept during the first two weeks of seedling growth. In the roots of both lines, *UGT84A9* is barely expressed. Only little difference was observed in *UGT84A10* expression in both lines. *UGT84A10* was mainly expressed in the roots of the seedlings. Its expression was higher on the seventh and tenth day after sowing in the transgenic line, although compared to the dsRNAi effect this change is minor. *UGT84A9* transcript levels were still reduced in the dsRNAi line at these time points. *NapA* expression was monitored again, to estimate the expression of the dsRNAi construct. The high *napA* transcript levels 2 days after sowing was most probably due to leftover RNAs of the seed filling phase. Thereafter, almost no *napA* transcript was detected.

### 3.3.2 Germination studies

To evaluate whether the ongoing suppression of *UGT84A9* has an influence on early plant development, germination studies were carried out using homozygous transgenic T6 seeds and three seed batches of their T7 offspring. In these experiments both *UGT84A9a*i lines, 13

and 44, were investigated and compared to the wild type line 142. To assess the germination efficiency of different lines, 100 seeds per line and experiment were placed on moist filter paper. As criterion for germination the appearance of the radicle was used, which was evaluated by eye inspection.

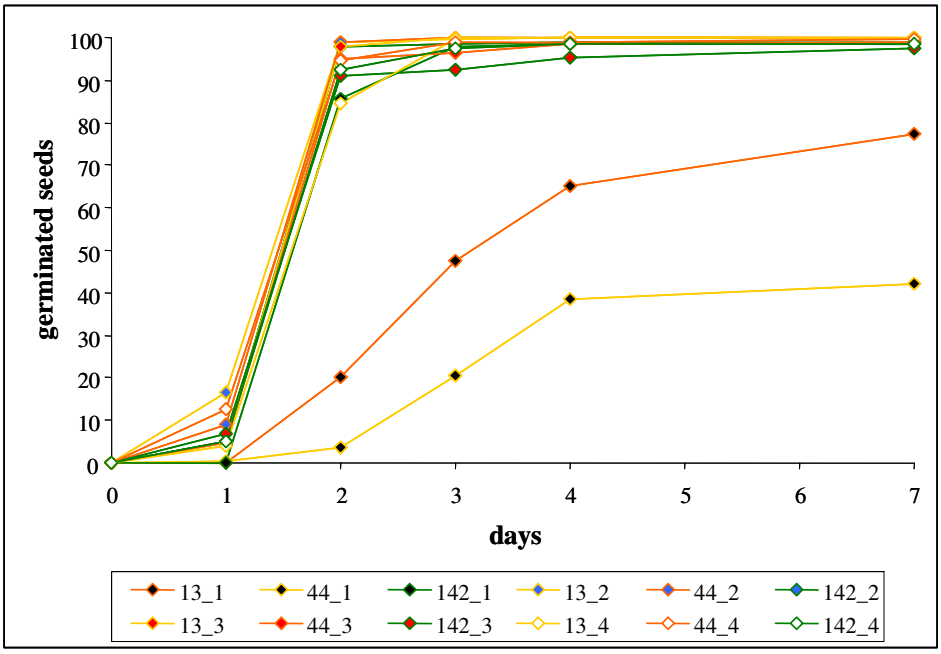



**Figure 3-20. Germination studies of *UGT84A9a*i seeds compared to wild type. Average values of two independent studies are shown. Analyzed lines are depicted as orange (line 13), yellow (line 44) or green (line 142) curves. Different seed batches are represented by differently colored diamonds: batch 1 black, T6, harvested 02/2006; batch 2 blue, T7, harvested 01/2007; batch 3 red, T7, harvested 04/2007; batch 4 open, T7, harvested 08/2007.**

The results of the germination studies revealed no difference in the germination rate of *UGT84A9a*i lines compared to the wild type Drakkar with the exception of the T6 seeds (Figure 3-20). Two days after sowing at least 80% of the T7 seeds are germinated; at the end of the experiments nearly all seedlings developed. In the only available seed batch of the T6 generation the results were different. The germination rate in both transgenic lines was much lower than in the wild type due to an infection of the seeds with mildew. Similar infections were not observed in any of the T7 seed batches. It could not be excluded that the infection of the transgenic T6 seeds was a storage or a harvest effect. There were no independent seed batches from this generation to repeat the experiment.

## 3.4 Simultaneous silencing of *UGT84A9a* and *BnSCT1*

### 3.4.1 Vector construction and estimation of transgene copy numbers

For a further reduction of sinapine content in *B. napus* seeds, simultaneous seed-specific silencing of *UGT84A9a* and *BnSCT1*, the gene encoding the final enzyme in sinapine biosynthesis, was attempted. For vector construction, pBNN (Hüsken et al., 2005) was used as starting point. Two additional restriction sites were introduced by PCR to connect the two gene-specific fragments to a consecutive sense or anti-sense part, respectively. For the anti-sense part a new *Bgl*II restriction site was established, for sense part *Sph*I was used. The dsRNAi construct was transferred to the binary vector pLH7000 (Hausmann and Töpfer, 1999). Figure 3-21 shows a schematic overview of the *UGT84A9ai*/*BnSCT1*i construct.

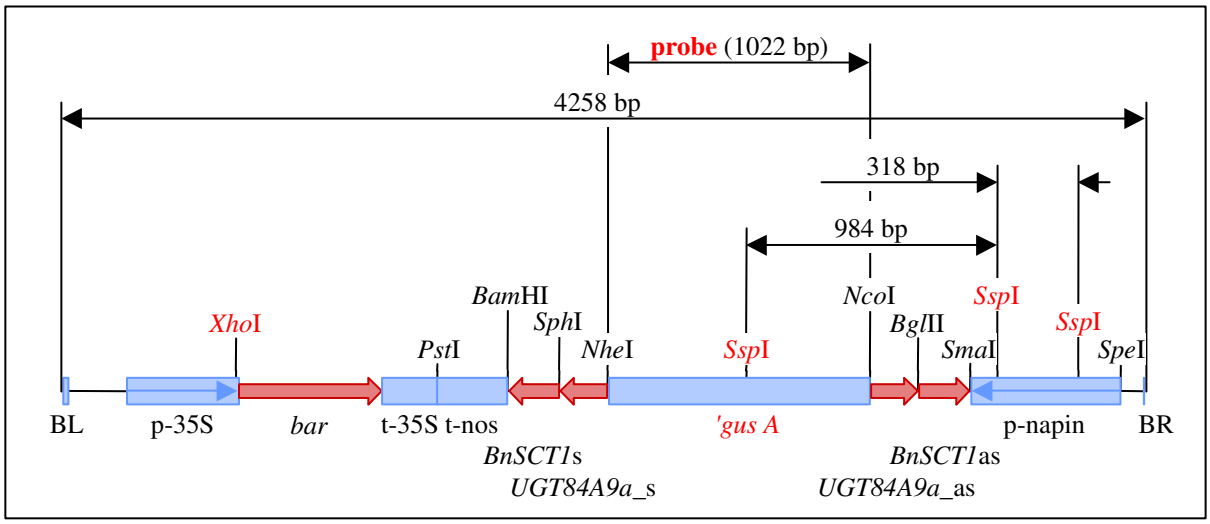



**Figure 3-21. Schematic overview of the *UGT84A9ai*/*BnSCT1*i construct. The T-DNA part of the binary vector is depicted. Indicated vector elements: left border (LB) and right border (RB) of the T-DNA, CaMV 35S promoter (p-35S), *bar* gene encoding phosphinothricine acetyl transferase (*bar*), CaMV 35S terminator (t-35S). The seed specific suppression cassette consists of the napin promoter (p-napin), a 197 bp subsequence of *BnSCT1* in antisense (as) and sense (s) orientations, a 180 bp subsequence of *UGT84A9a* in antisense (as) and sense (s) orientations, a *gus A* gene fragment (*'gus A*), and the nos terminator (t-nos). Recognition sites for restriction endonucleases used for cloning are shown in black. Probe and recognition sites for enzymes used for Southern blot analysis are given in red.**

Transgenic plants were produced by Dr. Orsini and co-workers (Saaten-Union Resistenzlabor). The copy numbers of the integrated T-DNA were estimated by Southern blot analysis. For this, genomic DNA was extracted from the leaves of hemizygous T1 plants and cleaved with *Xho*I and *Ssp*I separately (Figure 3-21). As probe, a digoxigenin-labeled amplicon of the *'gus A* fragment was used. *Xho*I does not cleave inside the probe sequence, thus, the number of hybridization signals directly indicates the transgene copy number in the genome. In case of a tandem insertion one detected fragment must be in the size of the inserted T-DNA (4,289 bp), since *Xho*I cuts once inside the transferred cassette. In contrast, for *Ssp*I there are three restriction sites inside the T-DNA, one of them inside the probe. A single copy insertion would result in two visible fragments in the Southern blot, one of them with the size of 984 bp, which results from the cleavage inside the integrated T-DNA. Multiple insertions of the T-DNA could be identified by the 984 bp fragment, which should

be present in each Southern blot, and a number of fragments that represents the number of inserted transgene copies. Tandem insertion would result in a 984 bp fragment, a 2,956 bp fragment, which occurs if two consecutively arranged T-DNA cassettes are cleaved with *Ssp*I and then hybridized with the *'gus A* probe, and one fragment of variable length.

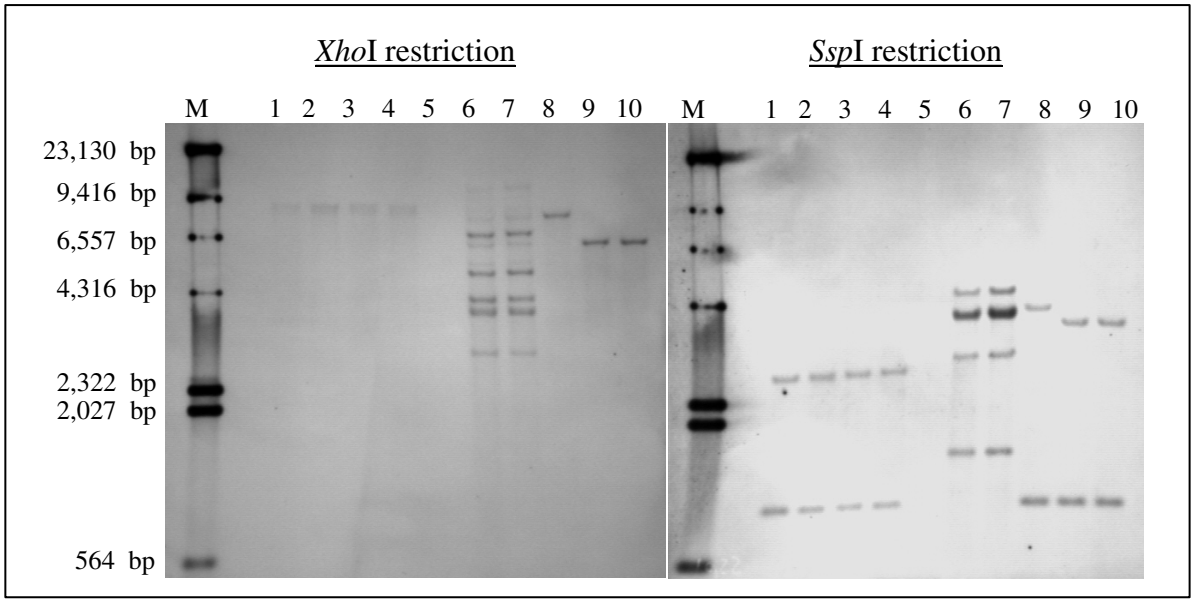



**Figure 3-22. Southern blot analysis of transgenic *UGT84A9ai/BnSCT1i* lines. Genomic DNA of T1 plants was restricted with *Xho*I (non cutter in the probe region) and *Ssp*I (single cutter in the probe region). Pre-labeled marker was applied in lane M. Analyzed lines: 1, T1505.1a; 2, T1505.1b; 3, T1505.1c; 4, T1505.1d; 5, T1505.2; 6, T1505. 3a ; 7, T1505.3b; 8, T1505.4 ; 9, T1505.5a ; 10, T1505.5b.**

Ten transgenic lines were analyzed by Southern blot (Figure 3-22). The lines were named according to the transformed construct as T1505 and were then numbered after the produced calli in the transformation experiments. If a callus produced more than one scion, a lower-case letter was added to the name. All lines originating from T1505.1(a-d), T1505.4 and T1505.5(a,b) were shown to harbor exactly one copy of the integrated DNA. Line 1505.2 was not transgenic, since no fragments were detected in the Southern blot. For the two lines T1505.3a and T1505.3b the exact copy number could not be determined. *Xho*I restriction produced five distinct bands, three of them smaller than the integrated sequence. In contrast, after *Ssp*I restriction, only four fragments could be detected and the characteristic 984 bp fragment was missing. Probably, the T-DNA was not integrated correctly in these lines. It is noticeable, that Southern blot analysis produced indentical restriction pattern for different scions originating from the same callus, indicating that these lines descended from the same transformation event and are syngeneic (genetically identical).

### 3.4.2 Sinapine content of T2 single seeds

To evaluate the suppression effect of the *UGT84A9ai*/*BnSCT1*i lines, we measured the sinapine content in the T2 seeds. Transgenic T1 plants are hemizygous for the transferred DNA, thus in the T2 generation the plants are segregating into transgenic homozygous, heterozygous and wild type plants. For this reason, single seeds instead of seed batches were investigated. The wild type seeds generated by segregation served as an internal control.

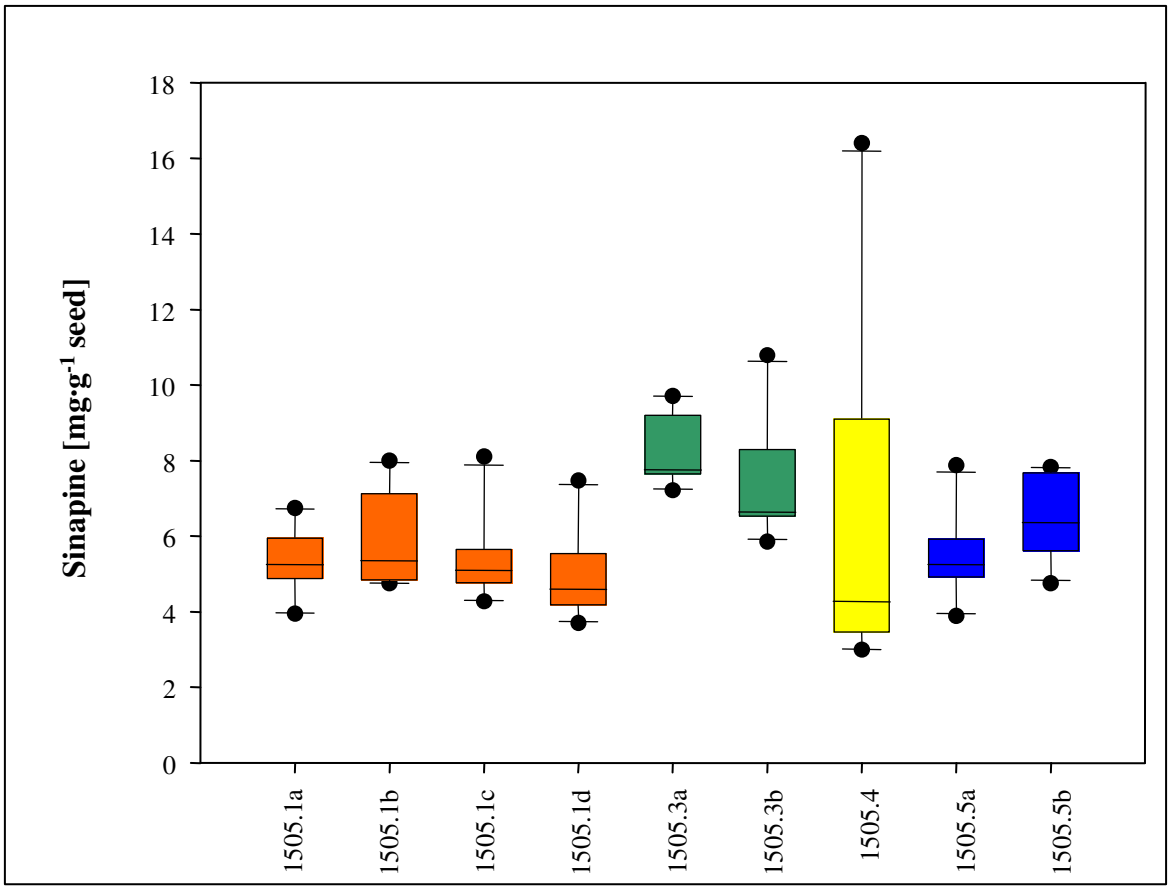



**Figure 3-23. Sinapine content in single T2 seeds of *UGT84A9ai*/*BnSCT1*i lines, depicted as a box plot of all 9 transgenic lines. The boundary of the box closest to zero indicates the 25th percentile, a line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles. Highest values per box represent putative wild type seeds.**

From nine transgenic lines T2 seeds were harvested. From the lines analyzed by Southern blot line T1505.2 failed in setting seeds. For each line, 10 individual seeds were analyzed, and a box plot of their sinapine content was calculated (Figure 3-23). This form of representation was chosen, since the sinapine value of the segregating dsRNAi seeds did not follow a normal distribution. Therefore, median values and percentiles should be used to describe the data, instead of means and standard deviations. The 10th and 90th percentiles allow a display of extremely low and high sinapine values, most likely representing homozygous dsRNAi and wild type seeds originating from segregation, respectively. For all single copy lines (1505.1a-d; 1505.4; 1505.5a,b) a lower median for sinapine content could be detected than in the two multi copy lines (1505.3a,b). It is noticeable, that Line 1505.4 shows a high diversity in its sinapine content, displaying both the lowest and the highest value overall. Since the Southern

blot experiments showed that different scions of a callus produced the same hybridization patterns, thus indicating they originated from the same transformation event, the single copy lines were evaluated on the basis of their transformation origin. For the four 1505.1 lines and the two 1505.5 lines, the line with the lowest median was selected, lines 1505.1d and line 1505.5a, respectively. Line 1505.4 was added to the selected lines, because it also is a single copy line with a low median on sinapine content. These lines also showed the lowest sinapine values overall in this analysis. Consequently, these three lines were selected for further breeding and the production of homozygous plants.

### 3.4.3 Generation of homozygous T3 plants

To measure sinapine and other sinapate esters in homozygous transgenic lines, homozygous T3 plants were generated by segregation. During meiosis in T2 seed development, the transgene segregates into homozygous, heterozygous and wild type state. For a single locus trait the Mendelian ratio is 1:2:1 for each state. Thus, one third of the T2 plants should be homozygous, since the wild type T2 plants were eliminated by earlier phosphinothricine selection. The homozygous plants should be detectable, because all of their produced T3 seeds and their seedlings should be resistant to phosphinothricine. Altogether, the T3 seeds of 50 T2 plants originating from each T1505.1d and T1505.4 and the seeds of 35 T2 plants originating from T1505.5a were used for the segregation analysis. Ten descendants of a given T2 plant were tested for phosphinothricine resistance.

**Table 3-6. Segregation analysis of T3 *UGT84A9ai*/*BnSCT1i* plants.**

| | Total number | Homozygous | Heterozygous | Wild type |
|---|---|---|---|---|
| T1505.1d | 50 | 16 | 27 | 7 |
| T1505.4 | 50 | 16 | 29 | 5 |
| T1505.5a | 35 | 5 | 22 | 8 |

After the first screen, 17 putative homozygous lines were found for T1505.1d, 18 lines were found for T1505.4 and five lines were identified for T1505.5a. These lines were checked again with the same method to confirm the homozygous status. Finally, 16 homozygous lines were confirmed for T1505.1d and T1505.4, respectively, and five homozygous lines were identified for T1505.5a. For the two lines T1505.1d and T1505.4, the results are in the range of the expected Mendelian ratio, for line T1505.5a the number of identified homozygous sub-lines is too low. The homozygous sub-lines identified for each independent transgenic line should be syngeneic, because all originate from one transformation event. Altogether, 148 homozygous transgenic plants were produced for line T1505.1d and T1505.4, respectively, and 34 homozygous transgenic plants were identified for T1505.5a.

### 3.4.4 Sinapate ester content in T4 seeds of homozygous plants

The homozygous T3 plants produced in the previously described segregation analysis were raised in the greenhouse and the T4 seeds were harvested and then analyzed for their sinapine content by HPLC. Seeds from homozygous transgenic single plants were investigated. Contrary to the study of T2 seeds, no single seeds were measured, but 20 µg of pooled seed material were used for methanol extraction. Sinapine was the most abundant sinapate ester found, contributing the main part to the whole sinapate ester pool. Other sinapate esters were mostly found in trace amounts, never reaching 10% of the concentration of sinapine.

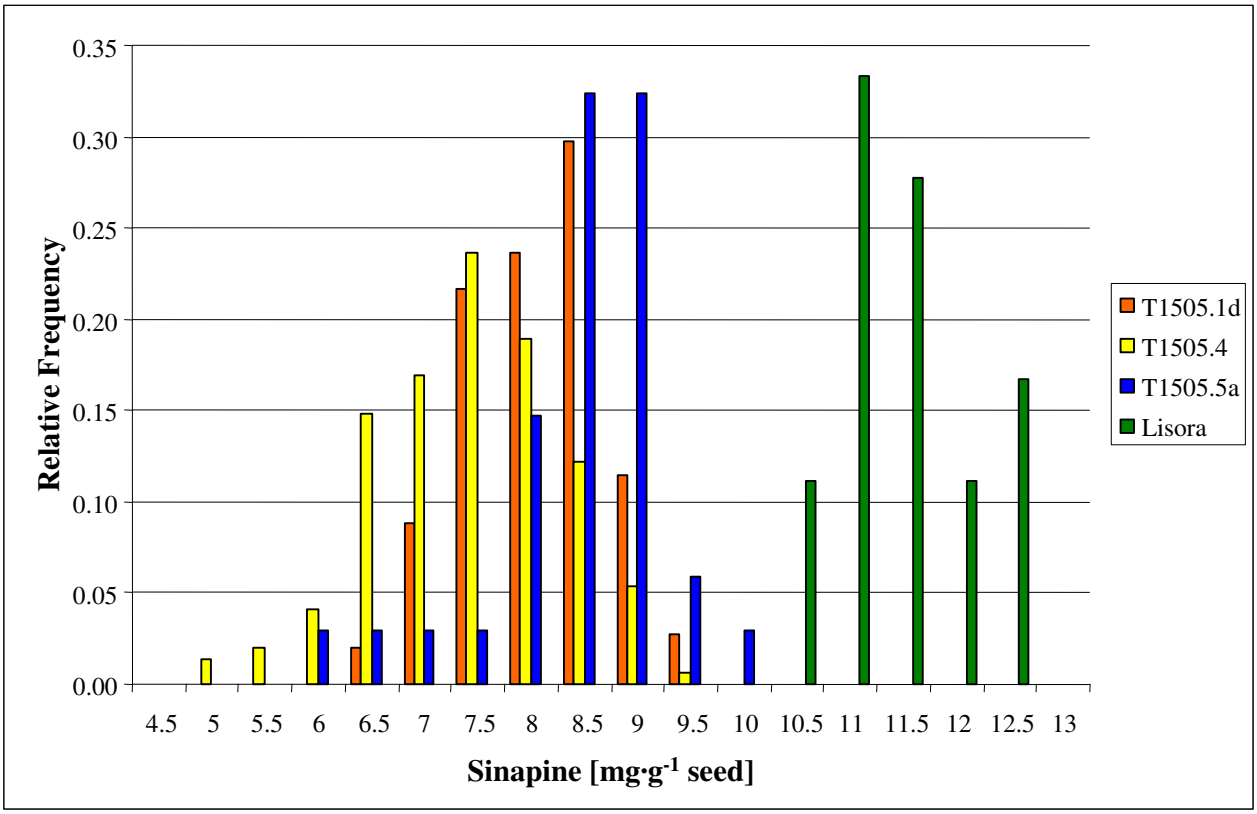



**Figure 3-24. Sinapine content in T4 seeds of single plants of *UGT84A9ai*/*BnSCT*1 lines. For line T1505.1d and T1505.4, 148 plants were analyzed, respectively, for line T1505.5a 34 plants, and 18 Lisora plants served as wild type control. Sinapine content was divided into distinct classes with an 0.5 mg per g increment. Relative frequency per class is given, since the number of analyzed plants differed for each line.**

The sinapine content of the T4 seeds is depicted in Figure 3-24. Sinapine content was divided into distinct classes with an increment of 0.5 mg sinapine per g seed material. For lines T1505.1d and T1505.4, 148 single plants were investigated, respectively. 34 plants were measured in the line T1505.5a, and 18 control plants were used for comparison. The relative frequency of each class is given, since the number of investigated plants differs for each analyzed line. It was shown that the sinapine content of all transgenic plants is lower than that of the wild type plants. Line T1505.4 contained the lowest amounts of sinapine, in average (7.19 $mg \cdot g^{-1}$ seed material) as well as in absolute values (4.63 $mg \cdot g^{-1}$ seed material).

To compare the strength of suppression of sinapate esters in the *UGT84A9ai*/*BnSCT*1 lines to the *UGT84A9a*i lines (Hüsken et al., 2005), the total content of sinapate derivatives was

determined. For these experiment seeds were pooled for each of the homozygous sub-lines (3.4.3). Methanolic extracts of *B. napus* seeds were incubated with sodium hydroxid to saponify the various sinapate esters. Sinapate content was then measured by HPLC.

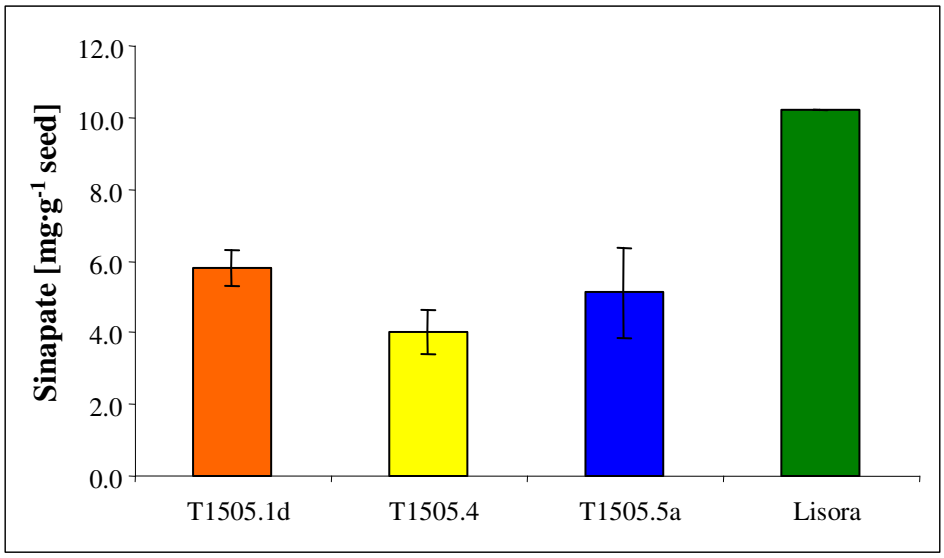



**Figure 3-25. Sinapate content in T4 seeds of *UGT84A9ai/BnSCT1* lines after saponification of sinapate esters. Sinapate content was measured for each homozygous sub-line and is shown as average value. Standard deviation is given. For T1505.1d and T1505.4, 16 sub-lines were investigated, for T10505.5a five sub-lines. Lisora is used as wild type control.**

Average values of the total sinapate content show a clear decrease in the dsRNAi lines compared to the wild type control (Figure 3-25). Best performing line was T1505.4, for which an average reduction by 60% of the wild type sinapate value was detected. Minimum total sinapate content was 3.23 $mg \cdot g^{-1}$ seed material, measured in sub-line T1505.4.4. This is a reduction by 68% compared to the average wild type total sinapate content. For the *UGT84A9a*i plants a maximum reduction of total sinapate content by 76% was reported, the lowest value of total sinapate ester content measured was 1.82 $mg \cdot g^{-1}$ seed material (Hüsken et al., 2005). Thus, it can be concluded that the simultaneous suppression of *UGT84A9a* and *BnSCT1* did not improve the reduction of sinapate ester content in comparison to the *UGT84A9a*i plants.

# 4 Discussion

## 4.1 UDP-glucose:hydroxycinnamate GTs in *Brassica napus*

The aim of this work was to isolate and characterize genes from *B. napus* encoding enzymes involved in the formation of 1-*O*-sinapoylglucose. Previously, an UDP-glucose:sinapate glucosyltransferase was cloned from *B. napus* (Milkowski et al., 2000a), designated *BnSGT1* and later renamed *UGT84A9* according to the conventional UGT nomenclature (Mackenzie et al., 1997). However, in the closely related model plant *Arabidopsis thaliana* four genes were identified, encoding proteins capable of converting sinapate to sinapoylglucose *in vitro* (Milkowski et al., 2000b; Lim et al., 2001). This raised the question, whether of additional HCA-GT genes exist in *B. napus* also. A first dsRNAi approach to reduce *UGT84A9* gene expression showed a maximum decrease of the total sinapate ester content in homozygous T3 seeds by 76%, thus proving the importance of sinapoylglucose as precursor for the formation of these compounds. Sinapate esters, mainly sinapine, spoil the otherwise valuable protein fraction of *B. napus* seeds by causing a bitter taste, unwanted color and problems like the "tainted eggs" (Naczk et al., 1998). Therefore, the newly isolated and characterized genes could serve as targets for molecular breeding approaches to reduce the sinapate ester content in seeds. In this work, the first approach was to identify genes related to *UGT84A9*, that potentially contribute to the formation of sinapoylglucose in the seeds of *B. napus*. To achieve this goal, two cDNA-based methods were applied, a RACE PCR approach and EST cloning.

These strategies led to the cloning of two full-length cDNAs encoding putative BnHCA-GTs which were named *UGT84A10* and *UGT84A11*, respectively. To determine whether both cDNAs encode functional BnHCA-GTs, they were heterologously expressed in *Escherichia coli*. The recombinant proteins were tested for glucosyltransferase activity towards cinnamate and four hydroxycinnamates. UGT84A10 was found to be active mainly towards ferulate and sinapate. It showed, however, a 150-fold lower activity towards sinapate than UGT84A9 (Milkowski et al., 2000a). Compared to its closest homologue from *A. thaliana*, UGT84A4, UGT84A10 exhibited a 9-fold lower activity towards sinapate (Milkowski et al., 2000b). The recombinant UGT84A11 could not convert any of the hydroxycinnamate substrates to the respective glucose esters under our established assay conditions, although Western blot analysis confirmed the expression as soluble protein. An alignment of the deduced amino acid sequences of HCA-GTs from *B. napus* and *A. thaliana* revealed several amino acid exchanges in UGT84A11 located in sequence blocks that are highly conserved among proven HCA-GTs (Mittasch et al., 2007). These findings led us to the conclusion that *UGT84A11* does not code for a functional BnHCA-GT.

Gene expression studies showed that the transcript levels for *UGT84A10* were not enhanced in seeds. *UGT84A10* transcript was mainly detected in flowers and roots. Among several plant organs investigated a faint RT-PCR signal was detected only in maturating seeds. In developing *B. napus* seeds of *UGT84A9a*-dsRNAi plants no enhanced transcript levels were detected for *UGT84A10*, contradicting the hypothesis that UGT84A10 might substitute for the loss of sinapate-UGT activity. Sinlapadech et al. (2007) proposed that *UGT84A4*, the *A. thaliana* homologue of *UGT84A10*, is redundant to *UGT84A2*, the homologue of *UGT84A9*,

in the leaves of *A. thaliana*. But since *UGT84A10* gene expression could barely be detected in *B. napus* seeds and the UGT84A10 enzyme activity towards sinapate is ninefold lower than that of UGT84A4, a redundancy of both enzymes in *B. napus* seeds is highly unlikely. However, in gene expression studies in *B. napus* seedlings significant amounts of *UGT84A10* transcript could be detected in the roots. A comparison of *UGT84A9a* silenced plants to the respective wild type control showed enhanced *UGT84A10* transcripts in the roots of the dsRNAi plants. Genevestigator studies of *UGT84A4* revealed an elevated expression in radicles and roots of Arabidopsis, especially in the elongation zone and in lateral roots. Semi-quantative RT-PCR data for *UGT84A4* support this analysis (Sinlapadech et al., 2007). Therefore, we assume that *UGT84A10* rather plays a role in the root than in the seed phenylpropanoid metabolism.

A screen of ESTs derived from flowers and seeds of *B. napus* resulted in twelve new UGT-like sequences, of which ten carried the 3' end of an ORF. As expected, *UGT84A9* was also detected in this screen. Two isolated ESTs, BnGT7 and BnGT9, were assigned to the phylogenetic group L (Li et al., 2001), which contains the ester-forming UGTs, but both ESTs missed any close relationship to *UGT84A9* or other HCA-GTs (UGT84Ax). For the *A. thaliana* homologue of BnGT7, UGT74D1, no activity has been described so far. BnGT9 is closely related to UGT74F2, which was characterized as anthranilate UGT (Quiel and Bender, 2003). Our data revealed that two different ESTs, BnGT8 and BnGT11, could be of potential interest for hydroxycinnamate metabolism in *B. napus*. BnGT8 was assigned to group D and showed the highest similarity to the *A. thaliana* glycosyltransferase UGT73C5, that is involved in the glucosylation of brassinosteroids (Poppenberger et al., 2005). A second BnGT8 homologue, UGT73C6, was characterized as UDP-glucose:flavonol 3-*O*-glycoside 7-*O*-glucosyltransferase (Jones et al., 2003). Since kaempferol 3-*O*-sophoroside 7-*O*-*β*-glucopyranoside and related 7-*O*-glycosides were found in *B. napus* seeds (Baumert et al., 2005), the product encoded by the full-length BnGT8 ORF is a candidate to be tested for an involvement in these syntheses. Although BnGT8 was found to be expressed in *B. napus* seeds, its transcripts were more abundant in flowers. For BnGT11, assigned to group E, the closest *A. thaliana* homologue UGT72E3 was shown to glucosylate hydroxycinnamates, hydroxycinnamyl alcohols and aldehydes *in vitro* (Lim et al., 2005). However, no BnGT11 transcript was detected in the different developmental stages of *B. napus* seeds. Gene expression was found in young seedlings and flowers, instead. Therefore, it is unlikely for BnGT11 to be involved in sinapate metabolism in *B. napus* seeds.

Given the characterization of UGT84A10 and UGT84A11 and the evaluation of the ESTs derived from flowers and seeds it can be concluded, that the formation of sinapoylglucose in *B. napus* seeds is mainly catalyzed by UGT84A9. We found no experimental evidence for the involvement of a second HCA-GT in the seed sinapate metabolism of *B. napus*. These results are in contrast to the situation in the model plant *A. thaliana*, where a functional redundancy of four HCA-GTs, UGT84A1-A4, was described (Sinlapadech et al., 2007; Dirk Meissner, personal communication). A possible explanation for this difference could be that the multiple polyploidization events *B. napus* experienced after the separation from the Arabidopsis branch provided new copies of each gene. As the additional copies of *UGT84A9* could provide the necessary redundancy in seed sinapate metabolism, potential homologues of *A. thaliana UGT84A1*, *-A3* and *-A4*, were probably subjected to sub- and neofunctionalization.

## 4.2 Organization and regulation of genes encoding UGT84A9

Given the importance of *UGT84A9* for the sinapate ester metabolism of *B. napus*, and therefore, as a target for molecular breeding approaches towards low sinapine oilseed rape lines, the gene was characterized in its genomic context. Since the results should be used in a TILLING approach, gene copy number and the expression pattern of individual loci were of major interest. However, the elucidation of *UGT84A9* loci structure would also be of academic interest, since this would provide additional insight into the complexity of the allotetraploid genome of *B. napus*. In this work, four copies of *UGT84A9* were identified in the *B. napus* genome and named *UGT84A9a*, *UGT84A9b*, *UGT84A9c* and *UGT84A9d*. *UGT84A9a* and *UGT84A9b* belong to sequence type 1, whereas *UGT84A9c* and *UGT84A9d* were grouped into sequence type 2. The sequence types share 88% identity on nucleotide level. The gene copy number was determined by a thorough characterization of 57 BACs that hybridized with *UGT84A9a*. Previous results from a genomic Southern blot analysis indicated the presence of two gene copies in the *B. napus* genome, but additionally several weaker signals were detected in each hybridization (Milkowski et al., 2004). Since the probe for this blot was generated using sequence type 1 as template, the weaker signals probably represent copies of *UGT84A9* sequence type 2.

UGT84A9 enzyme activity to form sinapoylglucose in seeds is mainly provided by sequence type 1 composed of *UGT84A9a* and *UGT84A9b* as could be demonstrated by semi-quantitative RT-PCR and cDNA-AFLP. We infer the level of enzymatically active UGTs from transcript abundance for two reasons. First, previous studies have shown that for *UGT84A9* that the increase in gene transcripts during seed development was followed closely by an increase in the respective UGT84A9 activity in protein extracts of *B. napus* (Milkowski et al., 2004). Secondly, we were able to distinguish between the transcripts of both sequence types by RT-PCR as well as by cDNA-AFLP. This could neither be accomplished on protein nor on enzyme activity level. Both techniques, RT-PCR as well as cDNA-AFLP, exhibited similar results except for *UGT84A9* expression in flowers. While semi-quantitative RT-PCR showed a strong signal in flowers for both sequence types, less transcript was detected in cDNA-AFLP analysis. A possible explanation for this could be unspecific primer binding to an abundant UGT in RT-PCR of the flower sample. The cDNA-AFLP method was deliberately developed to be independent of the amplification efficiency of individual primers (Bachem et al., 1996). Thus, the results of the cDNA-AFLP reflect more precisely the transcript abundance. A comparison between the nucleotide sequences of *UGT84A9a* and *–b* and the cDNA clones obtained from *B. napus* seeds revealed that indeed both genes of the sequence type 1 were expressed in the seeds. Thus, we conclude, that UGT84A9 acitivity is mainly provided by *UGT84A9* sequence type 1. Therefore, for conventional breeding approaches and for the TILLING project it is reasonable to focus on the genomic loci *UGT84A9a* and *UGT84A9b*.

Sequence analysis revealed that *UGT84A9a* and *UGT84A9c* most likely originated from the Brassica C-genome, whereas *UGT84A9b* and *UGT84A9d* are probably descended from the Brassica A-genome. This would imply, that *UGT84A9a* and *–b* are orthologous genes, as well as *UGT84A9c* and *–d*, whereas *UGT84A9a*/*-c* and *UGT84A9b*/*-d* would form paralogous gene pairs (Figure 3-14). During evolution, the genus Brassica underwent a genome triplication after its separation from the Arabidopsis branch (Lagercrantz and Lydiate, 1996;

Lagercrantz, 1998, Town et al., 2006). Thus, it can be expected to find six copies of each gene in *B. napus*, from which the homologue is singular in *A. thaliana* like *UGT84A2*. In recent studies, however, it could be shown that about 35% of the initial gene copies from the triplicated genome of Brassica have been lost, most likely by deletion mechanisms (Town et al., 2006). Additional genes originating from duplication or, as it was the case in Brassica, from triplication events may undergo several different developments. They could gain new functional properties, 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 similarity between orthologues than between paralogues, as it was found in this work for the four *UGT84A9* genes. Bioinformatical and gene expression studies on duplicated genes in *A. thaliana* have shown that either one paralogous copy gets lost or – in case of remaining – these copies most frequently acquire different functions (Blanc and Wolfe, 2004). In some cases, a "regulatory hypofunctionalization" was observed, i.e. the gene expression of one paralogue was greatly reduced in almost all organs or treatment conditions (Duarte et al., 2006). In this work, the same effect could be detected for the two paralogous gene pairs *UGT84A9a*/*-c* and *UGT84A9b*/*-d* in *B. napus*. The two genes of sequence type 1 (*UGT84A9a*/*-b*) mainly contributed to *UGT84A9* gene expression in all observed tissues exhibiting UGT84A9 enzyme activity, whereas sequence type 2 gene expression could only be detected in flowers. A similar pattern, restricting the major gene expression to one orthologous gene pair, could be observed for other genes in *B. napus* (Milena Ouzunova, personal communication). Two possible explanations were given for this effect: either the second paralogue is needed for genetic robustness or the loss of selective maintenance is so recent that the minor paralogue is not completely silenced in all organs (Duarte et al., 2006). To solve this puzzle, additional information about the functionality of the two genes of *UGT84A9* sequence type 2 would be helpful, since functionality would be required for redundancy.

BAC sequencing of fragments containing the four *UGT84A9* loci and comparison to the corresponding region on *A. thaliana* chromosome 3 revealed that there were only remnants of collinearity between the two species at these loci. However, on three of the four characterized genomic loci of *B. napus* transposon-like sequences or incomplete transposon-like elements were found. Transposons 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). For the *UGT84A9* loci, we found homologies to *gypsy*-like elements and to a *CACTA*-like transposase. *Gypsy*-like elements are the main type of plant LTR (long terminal repeat) retrotransposons, which belong to the class I of transposable elements and share high similarity to retroviruses (Sabot and Schulmann, 2006). *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).

Comparative sequencing of genomes or genome regions of several individual strains or haplotypes of a given species revealed that there was extensive variation among those individuals. Based on the genome of *Streptococcus agalactiae* (Tettelin et al., 2005) the pan-genome theorie was developed. This model describes the bacterial genome as consisting of the core-genome, shared by all *S. agalactiae* strains and a dispensable genome which is individual for each strain. The core genome consists almost exclusively of housekeeping genes, whereas the dispensable genome harbors additional housekeeping genes and genes with unknown function as well as transposable elements. Later, similar variations among haplotypes in one species were found in *Zea mays* (Wang and Dooner, 2006) and other grass species (Morgante et al., 2007). This led to the extension of the pan-genome theory to plants. As most of the observed variations in plant genomes were attributed to transposon activities, the transposon-like sequences found on the three BAC fragments could indicate that the four *UGT84A9* loci are located in the dispensable genome of a possible *B. napus* pan-genome. This would also explain the differences in *UGT84A9* loci structure of the spring cultivar YN01-429 compared to the winter cultivar Express indicated by alteration of fragment size in Southern blot analysis.

## 4.3 Suppression of sinapine biosynthesis

During this work, transgenic *B. napus* plants carrying two different dsRNAi constructs have been characterized. In a first approach, transgenic lines harboring a seed-specific *UGT84A9a*i suppression construct were analyzed with regard to transcription of BnHCA-GT genes and germination, to determine whether the changed sinapate ester metabolism has any negative impact on the fitness of the transgenic plants. These lines were created and initially characterized with regard to the seed sinapate ester content by Hüsken et al. (2005).

To prove the stability of the gene suppression effect in *UGT84A9a*i plants, transcript levels of T5 seeds and seedlings were analyzed. For both investigated transgenic lines (line 13 and line 44, 3.3.1), a clear reduction of *UGT84A9* transcripts could be detected in seeds compared to the wild type control. These results show, that the dsRNAi effect was stably inherited in the offspring generations. For *A. thaliana* the maintenance of hairpin-RNA meditated silencing has been proven up to T5 offspring (Stoutjesdijk et al., 2002). In crop plants, stable heritage of the dsRNAi effect was shown for T2 (Byzova et al., 2004) or T3 generations (Townsend and Llewellyn, 2007). Here, we could demonstrate that the effect was still present in T5 and later generations. It is remarkable, that *UGT84A9* transcripts were detected in seeds at desiccation stage of both transgenic lines (Figure 3-18). Wild type *B. napus* seeds at this developmental stage contain high *UGT84A9* transcript levels (Milkowski et al., 2004). Apparently the dsRNAi effect was not strong enough to silence the gene completely, and thus the remaining *UGT84A9* gene expression contributes to the residual formation of sinapoylglucose. In young transgenic seedlings, reduced *UGT84A9* gene expression was observed as well. Fourteen days after sowing a weakening of the dsRNAi effect was found. This indicates, that the seed-specific silencing of *UGT84A9* is indeed limited to the seeds and young seedlings, and does not affect other stages of plant development. By use of inducible-promoter-driven constructs, temporal- and tissue-specific silencing of endogenous genes has been demonstrated in tobacco (Chen et al., 2003b). In the following years, tissue-specific

silencing has been shown for several plants: *A. thaliana* and *B. napus* (Byzoga et al., 2004) as well as cotton (Sunilkumar et al., 2006). There were reports about a systemic gene silencing effect induced by dsRNAi or siRNA (Palauqui et al., 1997, Mlotshwa et al., 2002, Klahre et al., 2002), however, this phenomenon was restriced to the silencing of transgenes, for endogenous mRNAs systemic silencing was not observed (Himber et al., 2003). A possible explanation could be that the transgenes are often highly expressed, contrary to the endogenous genes.

Germination studies of T6 and T7 seeds of *UGT84A9a*i plants were carried out to determine whether the reduction of sinapate esters has a negative impact on the ability of the transgenic seeds to germinate or whether the development of the transgenic seedlings is affected. The biological role of sinapine, the most abundant sinapate ester, is not yet resolved. Initially, sinapine was thought to be a storage form of choline (Strack, 1981), but the *A. thaliana* mutant *sin1* (sinapoylmalate biosynthesis locus 1), blocked in the biosynthesis of sinapate esters and therefore, displaying strongly reduced sinapate ester concentration in its seeds, showed no reduced germination efficiency. Choline concentration was found to be elevated in the seeds of the *sin1* mutant (Chapple et al., 1992). The Arabidopsis *sng2* (sinapoylglucose accumulator locus 2) mutant impaired in SCT activity accumulated sinapoylglucose and choline in seeds instead of sinapine. During germination this mutant was characterized by an elevated sinapoylglucose content until day three after sowing compared to the wild type, while the total amount of sinapate esters was nearly equivalent. After day three, the *sng2* mutant started to produce sinapoylmalate in comparable amounts to the wild type. The lack of sinapine had no obvious negative effect on the seedling development (Shirley et al., 2001). In our germination studies, comparing *UGT84A9a*i plants with reduced sinapate ester levels to *B. napus* wild type plants (cv. Drakkar), no negative effects on germination and seedling development were found. A single batch of T6 seeds, however, showed a different result. From this batch, both transgenic lines investigated had a significantly lower germination rate and germinated later than the wild type control. This was due to a mildew infection which was found on the surface of the transgenic seeds, but only to a very small extent on wild type seeds. It could not be ascertained, if this happened due to harvest issues or if there is a greater susceptibility of the transgenic lines for mildew infections, because it was only this one T6 seed batch available. In *A. thaliana* it could be demonstrated that a loss of function of the cinnamyl-alcohol dehydrogenases and the *O*-methyltransferases involved in phenylpropanoid metabolism led to a increased susceptibility to various pathogens. These mutants also exhibited changes in lignin compositions. Another mutant impaired in the *CCR1* gene showed lignin levels reduced by 30% but wild type responses to all tested pathogens (Nikolaus Schlaich, personal communication). This illustrates the fact, that the impact of HCA metabolism on pathogen susceptibility needs further experimental proof.

As a second approach for sinapate ester reduction in *B. napus* seeds, a dsRNAi construct was designed that suppressed UGT84A9 and BnSCT activity seed-specifically and simultaneously. The reduction of sinapate esters achieved with the simultaneous silencing equaled that of the seed-specific silencing of *UGT84A9* alone (Hüsken et al., 2005). Total sinapate ester content was reduced to about 40% in average in the best performing homozygous double-suppression line compared to the wild type control. The minimum total

sinapate content (after saponification) observed was 3.23 mg per g seed, indicating a reduction to 32% of the wild type content.

In a first characterization, T1 plants of ten putative transgenic lines were investigated by Southern blot analysis. Those lines originated from five hypocotyl dishes which produced altogether ten scions. Seven plants exhibited the characteristic hybridization pattern typical for a single insertion of the transgene. Strikingly, all scions of a given hypocotyl dish showed the same pattern in the Southern blot (Figure 3-22), thus giving rise to the assumption, that these scions did not produce independent lines, but were syngeneic. This assumption was encouraged by the similar median values for sinapine content for scions of the same callus in the single-seed-analysis of the regarding T2 seeds. Finally, we were able to investigate three independent single-insertion lines for the *UGT84A9a*i/*BnSCT1*i construct. This low number of independent lines generated could be one factor for the relatively weak reduction of sinapate esters found. Several studies described a higher number of independent lines and reported a great variation in the strength of suppression (Byzova et al., 2004; Yu et al., 2007). Therefore, to evaluate the suppression potential of the double dsRNAi approach would require the generation and investigation of more independent transgenic lines.

Second, it is known for *A. thaliana* that plants, which are impaired in sinapine formation, accumulate sinapoylglucose instead. For the *A. thaliana* mutant *sng2*, which is deficient in the *AtSCT* gene, no reduction of the total sinapate ester content was reported, because sinapoylglucose accumulated in the seeds (Shirley et al., 2001). Accordingly the dsRNAi approach designed to suppress the *AtSCT*, resulted in a negative correlation of sinapine and sinapoylglucose (Weier et al., 2008). If regulation of the sinapate esters is similar in *B. napus*, it can be expected, that a suppression of BnSCT does not effect the total sinapate ester content significantly. Thus, the reduction of sinapate esters in the transgenic *B. napus* seeds of the *UGT84A9a*i/*BnSCT1*i lines is mainly due to the suppression of UGT84A9.

## 4.4 Conclusions and future work

Our results clearly confirmed UGT84A9 as the primary enzyme catalyzing the formation of 1-*O*-sinapoylglucose in seeds, and therefore, the encoding genes as major targets for molecular breeding approaches toward low sinapine *B. napus* lines. The dsRNAi-mediated silencing of UGT84A9 was stably inherited and limited to seeds and early seedling development. The decreased sinapate ester content did neither affect germination nor the development of the transgenic plants under greenhouse conditions. Characterization of *UGT84A9* revealed four distinct loci present in the *B. napus* genome designated as *UGT84A9a*, *UGT84A9b*, *UGT84A9c* and *UGT84A9d*, respectively. These genes could be grouped in the distinguishable sequence types 1 (*UGT84A9a* and *UGT84A9b*) and 2 (*UGT84A9c* and *UGT84A9d*). Sequence comparison showed that *UGT84A9a* and *UGT84A9c* were derived from the Brassica C-genome, whereas *UGT84A9b* and *UGT84A9d* traced back to the A-genome. Gene expression analysis demonstrated that *UGT84A9* sequence type 1 transcripts were mainly present in *B. napus* seeds, and thus contribute to sinapoylglucose formation.

Based on these findings, future experiments could be proposed. In this regard, a major point would be the evaluation of *UGT84A9a*i plants in the open field including the validation of agronomical traits. Furthermore, the *UGT84A9a*-dsRNAi construct should be transformed

into a high performance cultivar to estimate its potential influence on the yield. Based on EST data available in public databases the dsRNAi approach could be extended to other genes of the phenylpropanoid metabolism. Recently developed Brassica microarrays (Yin et al., 2006; Gopalan Selvaraj, personal communication) could be screened to identify seed-specifically upregulated biosynthetic genes from the phenylpropanoid metabolism. Co-expressed genes could be evaluated for their possible role as regulatory elements. Another method recently adapted to *B. napus* is the SAGE-technology (Serial Analysis of Gene Expression; Obermeier et al., 2007). By assembling short expressed sequence tags into concatemers that are cloned and analyzed, this method allows the identification and quantification of tissue-specific gene expression and is especially suited for large genomes for which the development of arrays has turned out to be difficult. These experiments would help to identify target genes to modify the phenylpropanoid metabolism and could also provide new insights into the regulation of the whole pathway.

One of the main goals of this work was to provide information to be used in a *B. napus* TILLING project designed to create non-transgenic *B. napus* plants with a low content of sinapate esters in seeds. Our results indicate that the two *UGT84A9* sequence type 1 genes, *UGT84A9a* and *UGT84A9b*, should be the primary targets for this approach. Since the sequence information about the *UGT84A9* loci was obtained from *B. napus* cv. Express, this established high-performing winter cultivar should be used for a first TILLING experiment and the evaluation of the low sinapate ester phenotype. Afterwards, the results could be applied to YN01-429, a yellow-seeded spring cultivar of Canadian origin optimized for low fiber content in seeds. However, our results suggest a remarkable genomic variability at the *UGT84A9* loci, thus requiring additional sequence information for this cultivar. Another interesting aspect would be the investigation of *UGT84A9* loci of different cultivars to find additional evidence for the hypothesized pan-genome of *B. napus*. On the other hand, the information derived from genomic BAC sequences could be utilized to place the four *UGT84A9* loci on an existing *B. napus* linkage map. Their map position can be compared to known QTLs (quantitative trait loci) for the low sinapine trait to either characterize them more thoroughly or to identify alternative QTLs which could provide new information about other candidate genes involved. This could also help to identify regulatory factors of the sinapate metabolism, such as involved transcription factors. Another interesting aspect with regard to genome evolution is the question whether *UGT84A9* sequence type 2 could provide redundancy to the phenylpropanoid metabolism on the enzyme level. To answer this question would require functional expression of both *UGT84A9c* and *UGT84A9d* followed by analysis of substrate specificities.

# 5 Summary

The formation of sinapoylglucose is cricital for the biosynthesis of the various sinapate esters, including sinapine as the major phenolic seed compound in *Brassica napus* (Baumert et al., 2005; Hüsken et al., 2005). In the model plant *Arabidopsis thaliana* four genes have been identified, encoding the glucosyltransferases UGT84A1-A4 that catalyze this reaction (Milkowski et al., 2000b; Lim et al., 2001). From the related crop plant *B. napus*, however, only one homologous gene, *UGT84A9*, was known and characterized so far (Milkowski et al., 2000a; Milkowski et al., 2004). Therefore, this work was aimed at a comprehensive characterization of *B. napus* genes involved in the formation of sinapoylglucose, especially in seeds. Since *B. napus* is an allotetraploid plant, the determination of gene copy number and transcript abundance for different gene loci were included.

- Starting from database sequence information, two full-length putative BnHCA-GT ORFs, *UGT84A10* and *UGT84A11* (accession numbers: AM231594, AM231595), were cloned and functionally characterized. UGT84A10 was found to convert mainly ferulate and sinapate to their respective glucose esters, although its activity towards sinapate was substantially decreased compared to the homologs UGT84A9 (by a factor of 150) and UGT84A4 (by a factor of 9). UGT84A11 showed no activity towards any of the hydroxycinnamates under established assay conditions and was therefore considered not to be a functional BnHCA-GT.
- An EST-cloning project conducted with cDNAs from *B. napus* flowers and seeds enriched for UGT-like sequences resulted in 132 sequenced ESTs. 32 of these harbored UGT-like sequences, that were assigned to 13 groups. One group resembled *UGT84A9*, the remaining twelve groups were designated BnGT1-12 (accession numbers: AM231595-AM231607). Besides *UGT84A9*, no other sequences with similarity to ester-forming HCA-GTs were found to be expressed in *B. napus* seeds and flowers.
- Gene expression analysis by semi-quantitative RT-PCR revealed that none of the *B. napus* ESTs isolated in this work was expressed seed-specifically. *UGT84A10* showed strongest gene expression in the roots, similar to its Arabidopsis homolog *UGT84A4*.
- Cloning of *UGT84A9* cDNA from *B. napus* seeds and flowers and of genomic sequences from *B. napus* and its ancestors, *B. oleracea* and *B. rapa*, revealed the existence of two distinct sequence types for *UGT84A9*, further referred to as *UGT84A9* sequence type 1 and *UGT84A9* sequence type 2. Gene expression analysis conducted by semi-quantitative RT-PCR and cDNA-AFLP showed that the main *UGT84A9* activity in seeds is provided by sequence type 1. *UGT84A9* sequence type 2 transcripts were only found in flowers.
- By analyzing a genomic BAC library from *B. napus*, four distinct *UGT84A9* gene copies were identified, referred to as *UGT84A9a*, *UGT84A9b*, *UGT84A9c* and *UGT84A9d*. *UGT84A9a* and *–b* were shown to belong to sequence type 1, whereas *UGT84A9c* and *–d* were classified as sequence type 2. Sequence comparisons with *UGT84A9*-like sequences from *B. napus* progenitor species *B. rapa* and *B. oleracea*

showed that *UGT84A9a* and *–c* were inherited from *B. oleracea*, whereas *UGT84A9b* and *–d* originated from *B. rapa*.

- Characterization of the genomic upstream and downstream regions of the four *UGT84A9* loci revealed almost no collinearity with the homologous genomic region from *A. thaliana*. Instead, remains of different transposon activities were found at the *B. napus* loci.
- Transcript analysis of *B. napus* plants seed-specifically silenced for *UGT84A9a* activity demonstrated that the dsRNAi effect was restricted to seeds and early seedling development. Gene expression of the previously identified alternative BnHCA-GT *UGT84A10* did not increase in *UGT84A9*-silenced seeds, excluding an involvement of UGT84A10 in a hypothesized metabolic compensation. No difference in germination efficiency between *UGT84A9a*i and control plants was observed.
- A seed-specific RNAi approach targeting simultaneously *UGT84A9a* and *BnSCT1* did not result in further reduction of the sinapate ester content compared to the seed-specific inhibition of *UGT84A9a*.

A systematic search for alternative BnHCA-GTs at the transcript level gave no evidence for an additional enzyme besides UGT84A9 contributing to the conversion of sinapate to sinapoylglucose in *B. napus* seeds. The following in-depth characterization of *UGT84A9* revealed the presence of four distinct loci, designated as *UGT84A9a-d*. Each of these loci could be assigned to either of the two progenitor genomes of *B. napus*, thus confirming its heritage. The genes *UGT84A9a* and *UGT84A9b* were identified as targets for molecular breeding approaches including TILLING towards low sinapate ester content in seeds. However, the complexity of the genomic context implies that additional sequence information should be raised in order to transfer the results to other breeding cultivars of *B. napus*.

# 6 Literature

Akhunov, E. D.; Akhunova, A. R.; Dvorak, J: Mechanism and rates of birth and death of dispersed duplicated genes during the evolution of a multigene family in diploid and tetraploid wheats. In: *Mol. Biol. Evol*. Vol. 24(2): 539-550, 2007

Al-Shehbaz, I. A.; Beilstein, M. A.; Kellogg, E. A.: Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. In: *Pl. Syst. Evol*. Vol. 259: 89-120, 2006

Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J.: Basic local alignment search tool. In: *J. Mol. Biol*. Vol. 215: 403-410, 1990

Ayele, M.; Haas, B. J.; Kumar, N.; Wu, H.; Xiao, Y.; Van Aken, S.; Utterback, T. R.; Wortman, J. R.; White, O. R.; Town, C. D.: Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in *Arabidopsis*. In: *Genome Research*. Vol. 15: 487-495, 2005

Bachem, C. W. B.; van der Hoeven, R. S.; de Bruijn, S. M.; Vreugdenhil, D.; Zabeau, M.; Visser, R. G. F.: Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. In: *Plant J*. Vol. 9 (5): 745-753, 1996

Bairoch, A.: PROSITE: a dictionary of sites and pattern in proteins. In: *Nucl. Acids Res*. Vol. 19: 2241-2245, 1991

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

Beilstein, M. A,: Al-Shehbaz, I. A.; Kellogg, E. A.: Brassicaceae phylogeny and trichome evolution. In: *Am. J. Bot. Vol*. 93(4): 607-619, 2006

Bell, J. M.: Factors affecting the nutritional value of canola meal: A review. In: *Can. J. Anim. Sci*. Vol. 73: 679-697, 1993

Bharti, A. K.; Khurana, J. P.: Mutants of Arabidopsis as tools to understand the regulation of Phenylpropanoid pathway and UVB protection mechanisms. In: *Photochemistry and Photobiology*. Vol: 65 (5): 765-776, 1997

Birnboim, H. C.: A rapid alkaline extraction method for the isolation of plasmid DNA. In: *Methods Enzymol*. Vol. 100: 243-255, 1983

Blanc, G.; Wolfe, K. H.: Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. In: *Plant Cell*. Vol. 16: 1679-1691, 2004

Boerjan, W.; Ralph, J.; Baucher, M.: Lignin biosynthesis. In: *Annu. Rev. Plant Biol*. Vol. 54: 519-546, 2003

Bouchereau, A.; Hamelin, J.; Lamour, I.; Renard, M.; Larher, F.: Distribution of sinapine and related compounds in seeds of *Brassica* and allied genera. In: *Phytochemistry*. Vol. 30 (6): 1873-1881, 1991

Bradford, M. M.: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein binding. In: *Anal. Biochem*. Vol. 72: 248-254, 1976

Byzova, M.; Verduyn, C.; De Brouwer, D.; De Block, M. : Transforming petals into sepliod organs in Arabidopsis and oilseed rape : implementation of the hairpin RNA-mediated gene silencing technology in an organ-specific manner. In: *Planta*. Vol. 218: 379-387, 2004

Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. : A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. In: *Biochem. J*. Vol. 326: 929-942, 1997

Chapple, C. C. S.; Vogt, T.; Ellis, B. E., Somerville C. R. : An Arabidopsis mutant defective in the general phenylpropanoid pathway. In: *Plant Cell*. Vol. 4: 1413-1424, 1992

Chen, P. Y.; Wang, C. K.; Soong, S. C.; To, K.Y. :Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants. In: *Mol. Breed*. Vol. 11: 287–293, 2003a

Chen, S.; Hofius, D.; Sonnewald, U.; Börnke, F.: Temporal and spatial control of gene silencing in transgenic plants by inducible expression of double stranded RNA. In: *Plant J.* Vol. 36: 731-740, 2003b

Chuang, C.-F. ; Meyerowitz, E. M.: Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. In: *Proc. Natl. Acad. Sci USA*. Vol. 97 (9): 4985-4990, 2000

Clauß, K.; Baumert, A.; Nimtz, M.; Milkowski, C.; Strack, D.: Role of a GDSL lipase-like protein as sinapine esterase in Brassicaceae. In: *Plant J*. Vol. 53: 820-813, 2008

De Block, M.; De Brouwer, D. ; Tenning, P.: Transformation of *Brassica napus* and *Brassica oleracea* using *Agrobacterium tumefaciens* and the expression of *bar* and *neo* genes in the transgenic plants. In: *Plant Physiol*. Vol. 91: 694-701, 1989

Dicks, J.; Anderson, M.; Cardle, L.; Cartinhour, S.; Couchman, M.; Davenport, G.; Dickson, J.; Gale, M.; Marshall, M.; May, S.; McWilliam, H.; O'Malia, A.; Ougham, H.; Trick, M.; Walsh, S.; Waugh, R.: UK CropNet: a collection of databases and bioinformatics resources for crop plant genomics. In: *Nucl. Acids Res*. Vol. 28: 104-107, 2000

Dixon, R. A.: Natural products and plant disease resistance. In: *Nature*. Vol. 411: 843-847, 2001

Downey, R. K.: A selection of *Brassica campestris* L. containing no erucic acid in its seed oil. In: *Can. J. Plant Sc*i. Vol. 44: 295, 1964

Downey, R. K.; Craig, B. M.; Youngs, C. G.: Breeding rapeseed for oil and meal quality. In: *J. Am. Oil Chem. Soc*. Vol. 46: 121-123, 1969

Downey, R. K.: Tailoring rapeseed and other oilseed crops to the market. In: *Chem. Ind*. Vol. 1: 401-406, 1976

Downey, R. K.; Bell, J. M.: New development in canola research. In: Shadidi (Hg.) *Canola and Rapeseed. Production, Chemistry, Nutrition and Processing Technology*. Van Nostrand Reinhold, New York: 211-220, 1990

Duarte, J. M.; Cui, L.; Wall P. K; Zhang, Q; Zhang, X.; Leebens-Mack, J.; Ma, H.; Altmann, N.; dePamphilis C. W.: Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. In: *Mol. Biol. Evol*. Vol. 23(2): 469-478, 2006

Ehlting, J.; Hamberger, B.; Million-Rouseau, R.; Werck-Reichhart, D.: Cytochromes P450 in phenoilc metabolism. In: *Phytochem. Rev*. Vol. 5: 239-270, 2006

Fenwick, G. R.; Curtis, R. F.: Rapeseed meal in rations for laying hens: a review of the effects on egg quality. In: *J. Sci. Food Agric*. Vol. 31: 515-525, 1980

Fenwick, G. R.: The assesment of a new protein source – Rapeseed. In: *Proc. Nutr. Soc*. Vol. 41: 277-288, 1982

Fernandez de Simon, B.; Pérez-Ilzarbe, J.; Hernandez, T.; Goméz-Corovés, C.; Estrella, I.: Importance of phenolic compounds for the characterization of fruit juices. In: *J. Agric. Food Chem*. Vol. 40: 1531-1535, 1992

Franke, R.; Humphreys, J. M.; Hemm, M. R.; Denault, J. W.; Ruegger, M. O.; Cusumano, J. C.; Chapple, C. C. S.: The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. In: *Plant J*. Vol. 30 (1): 33-45, 2002

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

Gachon, C. M. M.; Langlois-Meurinne, M.; Henry, Y. ; Saindrenan, P.: Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis* : functional and evolutionary implications. In: *Plant Mol. Biol*. Vol.58: 229-245, 2005

Goldberg, R. B.; Barker, S.; Perez-Grau, L.: Regulation of gene expression during plant embryogenesis. In: *Cell*. Vol. 56: 149-160, 1989

Goldberg, R. B.; de Pavia, G.; Yadegari, R.: Plant embryogenesis: zygote to seed. In: *Science*. Vol. 266: 605-613, 1994

Grupp, C. D.; Zipp, B. J.; Ludwig-Müller, J.; Masuno, M. N.; Molinski, T. F.; Abel, S.: Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. In: *Plant J*. Vol. 40: 893-908, 2004

Hahlbrock, K.; Scheel, D.: Physiology and molecular biology of phenylpropanoid metabolism. In: *Annu. Rev. Physiol. Plant Mol. Biol*. Vol. 40: 347-369, 1989

Halaweish, F. T.; Dougall, D. K.: Sinapoyl glucose synthesis by extracts of wild carrot cell cultures. In: *Plant Science*. Vol. 71: 179-184, 1990

Hans, J.; Brandt, W.; Vogt, T.: Site directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-*O*-glucosyltransferase from *Dorotheanthus bellidiformis*. In: *Plant J*. Vol. 39: 319-333, 2004

Hausmann, L.; Töpfer, R.: Entwicklung von Plasmid-Vektoren. In: *BioEngineering für Rapssorten nach Maß. Vorträge für Pflanzenzüchtung*. Vol. 45: 153-171, 1999

Hellens, R.P.; Edwards, E.A.; Leyland, N.R.; Bean, S.; Mullineaux, P.M.: pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. In: *Plant Mol. Biol*. Vol. 42: 819–832, 2000

Henry; Fils; Garot: Acide sulfo-sinapique. In: *J. Pharm*. Vol. 11: 473-475, 1825

Himber, C.; Dunoyer, P.; Moissiard, G.; Ritzenthaler, C.; Voinnet, O.: Transitivity-dependent and –independent cell-to-cell movement of RNA silencing. In: *EMBO J*. Vol. 22(17): 4523-4533, 2003

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

Hughes, J; Hughes, M. A.: Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* Crantz) cotyledons. In: *DNA Seq*. Vol. 5: 41-49, 1994

Humphreys, J. M.; Hemm, M. R.; Chapple, C. C. S.: New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. In: *Proc. Natl. Acad. Sci USA*. Vol. 96: 10045-10050, 1999

Humphreys, J. M.; Chapple, C. C. S.: Rewriting the lignin roadmap. In: *Curr. Opin. Plant Biol*. Vol. 5: 224-229, 2002

Jones P.; Messner B.; Nakajima, J.-I.; Schäffner, A. R; Saito, K.: UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. In: *J. Biol. Chem*. Vol. 278: 43910-43918, 2003

Josefsson, E.; Appelqvist, L.-Å. : Glucosinolates in seed of rape and turnip rape as affected by variety and environment. In: *J. Sci. Food Agric*. Vol. 19: 564-570, 1968

Josefsson, L.-G.; Lenman, M.; Ericson, M. L.; Rask, L.: Structure of a gene encoding the 1.7S storage protein, napin, from *Brassica napus*. In: *J. Biol. Chem*. Vol. 262 (25): 12196-12201, 1987

Jukes, T. H.; Cantor, C. R: Evolution of protein molecules. In: *Mammalian Protein Metabolism*. H. N. Munro (ed). Academic Press, New York: 21-132, 1969

Klahre, U.; Crété, P., Leuenberger, S. A.; Iglesias, V. A.; Meins Jr., F.: High molecular weight RNAs and small interfering RNAs induce systemic posttranscriptional gene silencing in plants. In: *Proc. Natl. Acad. Sci USA*. Vol. 99(18): 11981-11986, 2002

Kobayashi, S.; Goto-Yamamoto, N.; Hirochika, H.: Retrotransposon-induced mutations in grape skin color. In: *Science*. Vol. 304: 982, 2004

Koncz, C.; Schell, J.: The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. In: *Mol. Gen. Genet*. Vol. 204: 383-396,1986

Körber-Grohne, U.: Ölfrüchte. In: *Nutzpflanzen in Deutschland – von der Vorgeschichte bis heute*. Nikol Verlagsgesellschaft. Hamburg: 149-170, 1995

Kridl, J. C.; McCarter, D. W.; Rose, R. E.; Scherer, D. E.; Knutzon, D. S.; Radke, S. E.; Knauf, V. C: Isolation and characterization of an expressed napin gene from *Brassica rapa*. In: *Seed Sci. Res*. Vol.1: 209–219, 1991

Kroll, H.: Ein archäologischer Rapsfund des 16. Jahrhunderts, entdeckt in Heide in Holstein, Norddeutschland. In: *J. Agron. Crop Sci*. Vol. 173:17-21, 1994

Kwon, S.-J.; Park, K.-C.; Kim, J.-H.; Lee, J. K.; Kim, N.-S: *Rim 2*/*Hipa* CACTA transposon display; A new genetic marker technique in Oryza species. In: *BMC Genetics*. Vol. 6(1): doi: 10.1186/1471-2156-6-15, 2005

Kwon, S.-J; Hong, S.-W.; Son, J.-H.; Lee, J. K.; Cha, Y.-S.; Eun, M.-Y.; Kim, N.-S : CACTA and MITE transposon distributions on a genetic map of rice using $F_{15}$ RILs derived from Milyang 23 and Gihobyeo hybrids. In: *Mol. Cells*. Vol. 21(3): 360-366, 2006

Lacombe, E.; Hawkins, S.; Van Doorsselaere, J.; Piquemal, J.; Goffner, D.; Poeydomenge, O.; Boudet, A.-M.; Grima-Pettenati, J.: Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. In: *Plant J*. Vol. 11 (3): 429-441, 1997

Lagercrantz, U.; Lydiate, D. J. : Comparative genome mapping in Brassica. In: *Genetics*. Vol. 144: 1903-1910, 1996

Lagercrantz, U.: 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. In: *Genetics*. Vol. 150: 1217-1228, 1998

Landry, L. G.; Chapple, C. C. S.; Last, R. L.: Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. In: *Plant Physiol*. Vol. 109: 1159-1166, 1995

Langlois-Meurinne, M.; Gachon, C. M. M.; Daindrenan, P.: Pathogen-responsive expression of glycosyltransferase genes UGT73B3 and UGT37B5 is necessary for resistance to *Pseudomonas syringae* pv *tomato* in Arabidopsis. In: *Plant Phys*. Vol. 139: 1890-1901, 2005

Lee, J.-K.; Park, J.-Y.; Kim, J.-H.; Kwon, S.-J.; Shin, J.-H.; Hong, S.-K.; Min, H.-K.; Kim, N.-S.: Genetic mapping of the *Isaac*-CACTA transposon in maize. In: *Theor. Appl. Genet*. Vol. 113: 16-22, 2006

Li, J.; Ou-Lee, T.-M., Raba, R. ; Amundson, R. G. ; Last, R. L.: Arabidopsis flavonoid mutants are hypersensitive to UV-B irradiation. In: *Plant Cell*. Vol. 5: 171-179, 1993

Li, Y.; Baldauf, S.; Lim, E.-K.; Bowles, D. J.: Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. In: *J. Biol. Chem*. Vol. 276: 4338-4343, 2001

Lim, E.-K.; Li, Y., Parr, A.; Jackson, R.; Ashford, D. A.; Bowles, D. J.: Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in Arabidopsis. In: *J. Biol. Chem*. Vol. 276 (6): 4344-4349, 2001

Lim E.-K.; Doucet, C. J. ; Yi, L.; Elias, L.; Worrall, D.; Spencer, S. P.; Ross, J.; Bowles, D. J.: The activity of *Arabidopsis* glycosyltransferases towards salicylic acid, 4-hydroxybenzoic acid and other benzoates. In: *J. Biol. Chem*. Vol. 277 (1): 586-592, 2002

Lim E.-K.; Jackson, R. G.; Bowles, D. J.: Identification and characterization of Arabidopsis glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. In: *FEBS Lett*. Vol.579: 2802-2806, 2005

Lim, K.-B.; Yang, T.-J.; Hwang, Y.-J.; Kim J. S.; Park, J.-Y.; Kwon, S.-J; Kim, J.-A.; Choi, B.-S.; Lim, M.-H.; Jin, M.; Kim, H.-I.; de Jong, H.; Bancroft, I.; Lim Y. P.; Park, B.-S.: Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related Brassica species. In: *Plant J*. Vol. 49: 173-183, 2007

Lippman, Z.; Gendrel, A.-V.; Black, M.; Vaughn, M. W.; Dedhia, N.; McCombie, W. R.; Lavine, K.; Mittal, V.; May, B.; Kasschau, K. D.; Carrington, J. C.; Doerge, R. W.; Colot, V.; Martienssen, R.: Role of transposable elements in heterochromatin and epigenetic control. In: *Nature*. Vol. 430: 471-476, 2004

Lorenzen, M.; Racicot, V., Strack, D.; Chapple, C. C. S.: Sinapic acid ester metabolism in wild type and a sinapoylglucose-accumulating mutant of Arabidopsis. In: *Plant Physiol*. Vol. 112: 1625-1630, 1996

Mackenzie, P. I.; Owens, I. S.; Burchell, B., Bock, K. W.; Bairoch, A.; Bélanger, A.; Fournel-Gigleux, S.; Green, M.; Hum, D. W.; Iyanagi, T.; Lancet, D.; Louisot, P.; Magdalou, J.; Chowdhury, J. R.; Ritter, J. K.; Schachter, H.; Tephly, T. R.; Tipton, K. F.; Nebert, D. W.: The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. In: *Pharmacogenetics*. Vol. 7: 255-269, 1997

Marillia, E.-F.; MacPherson, J. M.; Tsang, E. W. T.; Van Audenhove, K.; Keller, W. A.; GrootWassink, J. W. D.: Molecular cloning of a *Brassica napus* thiohydroximate *S*-glucosyltransferase gene and its expression in *Escherichia coli*. In: *Physiol. Plant*. Vol. 113: 176-184, 2001

McCallum, C. M.; Comai, L.; Greene, E. A.; Henikoff, S.: Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. In: *Plant Physiol*. Vol. 123: 439-442, 2000

Michelmore, R. W.; Meyers, B. C.: Clusters of resistance genes in plants evolved by divergent selection and a birth-and-death process. In: *Genome Res*. Vol. 8: 1113-1130, 1998

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

Milkowski, C.; Baumert, A.; Strack, D.: Identification of four *Arabidopsis* genes encoding hydroxycinnamate glucosyltransferases. In: *FEBS Letters*. Vol. 486: 183-184, 2000b

Milkowski, C.; Baumert, A.; Schmidt, D.; Nehlin, L.; Strack, D.: 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. In: *Plant J*. Vol. 38: 80-92, 2004

Miura, A.; Kato, M.; Watanabe, K.; Kawabe, A.; Kotani, H., Kakutani, T.: Genomic localization of endogenous mobile CACTA family transposons in natural variants of *Arabidopsis thaliana*. In: *Mol. Gen. Genomics*. Vol. 270: 524-532, 2004

Mittasch, J.; Strack, D.; Milkowski, C.: Secondary product glycosyltransferases in seeds of *Brassica napus*. In: *Planta*. Vol. 225: 515-522, 2007

Mlotshwa, S.; Voinnet, O.; Mette, M. F.; Matzke, M.; Vaucheret, H.; Ding, S. W.; Pruss, G., Vance, V. B.: RNA silencing and the mobile signal. In: *Plant Cell*. Vol. 14 Suppl:S289-301, 2002

Mock, H.-P., Strack, D.: Energetics of the uridine 5'-diphosphoglucose:hydroxycinnamic acid acyl-glucosyltransferase reaction. In: *Phytochemistry*. Vol. 32 (3): 575-579, 1993

Morgante, M.; De Paoli, E.; Radovic, S.: Transposable elements and the plant pan-genomes. In: *Curr. Opin. Plant Biol*. Vol. 10: 149-155, 2007

Nair, R. B.; Joy IV, R. W. ; Kurylo, E. ; Shi, X.; Schnaider, J.; Datla, R. S. S.; Keller, W. A.; Selvaraj, G.: Identification of a CYP84 family of Cytochrome P450-dependant mono-oxigenase genes in *Brassica napus* and perturbation of their expression for engineering sinapine reduction in the seeds. In: *Plant Physiol*. Vol. 123: 1623-1634, 2000

Nair, R. B.; Bastress, K. L. ; Ruegger, M. O. ; Denault, J. W. ; Chapple, C. C. S.: The *Arabidopsis thaliana REDUCED FLUORESCENCE1* gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. In: *Plant Cell*. Vol. 16: 544-554, 2004

Nazck, M.; Amarowicz, R.; Sullivan, A.; Shahidi, F.: Current research developments on polyphenolics of rapeseed/canola: a review. In: *Food Chemistry*. Vol. 62 (4): 489-502, 1998

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

Nurmann, G.; Strack, D.: Formation of 1-sinapoylglucose by UDP-glucose:sinapic acid glucosyltransferase from cotyledons of *Raphanus sativus*. In: *Z. Pflanzenphysiol*. Vol. 102: 11-17, 1981

Obermeier, C.; Hosseini, B.; Snowdon, R.: Serial Analysis of Gene Expression (SAGE) applied to the study of Brassica napus seed development. In: *Proceedings of the 12-th International Rapeseed Congress*. Bd. 2: 120-123, 2007

Ohlson, R.: Functional properties of rapeseed oil and protein product. In: *Proceedings of the 5$^{th}$ international rapeseed congress, Malmö, Schweden*: 152-167, 1978

Osborn, T. C.; Pires, J. C.; Birchler, J. A.; Auger, D. L.; Chen, Z. J.; Lee, H.-S.; Comai, L.; Madlung, A.; Doerge, R. W.; Colot, V.; Martienssen, R. A.: Understanding mechanism of novel gene expression in polyploids. In: Trends in Genetics. Vol. 19 (3): 141-147, 2003

Palauqui, J.-C.; Elmayan , T.; Pollien, J.-M.; Vaucheret, H.: Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. In: *EMBO J*. Vol. 16(15): 4738-4745, 1997

Paquette, S.; Møller, B. L. ; Bak, S. : On the origin of family 1 plant glycosyltransferases. In: *Phytochemistry*. Vol. 62: 399-413, 2003

Park, J. Y.; Koo, D. H.; Hong, C. P.; Lee, S. J.; Jeon, J. W.; Lee, S. H.; Yun, P. Y.; Park, B. S.; Kim, H. R.; Bang, J. W.; Plaha, P.; Bancroft, I.; Lim, Y. P.: Physical mapping and microsyntheny of *Brassica rapa* ssp. *pekinensis* genome corresponding to a 222 kbp gene-rich region of *Arabidopsis* chromosome 4 and partially duplicated on chromosome 5. In: *Mol. Gen. Genomics*. Vol. 274: 579-588, 2005

Parkin, I. A. P.; Sharpe, A. G.; Keith, D. J.; Lydiate, D. J.: Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). In: *Genome*. Vol. 38: 1122-1131, 1995

Parkin, I. A. P.; Gulden, S. M.; Sharpe, A. G.; Lukens, L.; Trick, M.; Osborn, T. C.; Lydiate, D. J.: Segmental structure of the *Brassica napus* genome based on comparative analysis with Arabidopsis thaliana. In: *Genetics*. Vol. 171: 765-781, 2005

Poppenberger, B.; Fujioka, S.; Soeno, K.; George, G. L.; Vaistij, F. E.; Hiranuma, S.; Seto, H., Takatsuto, S.; Adam, G.; Yoshida, S.; Bowles, D.: The UGT73C5 of Arabidopsis thaliana glucosylates brassinosteroids. In: *Proc. Natl. Acad. Sci USA*. Vol. 102 (42): 15253-15258, 2005

Quiel, J. A.; Bender, J.: Glucose conjugation of anthranilate by the Arabidopsis UGT72F2 glucosyltransferase is required for tryptophan mutant blue fluorescence. In: *J. Biol. Chem*. Vol. 278 (8): 6275-6281, 2003

Rana, D., van den Boogaart, T.; O'Neill, C. M.; Hynes, L.; Bent, E.; Macpherson, L.; Park, J. Y.; Lim, Y. P.; Bancroft, I.: Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. In: *Plant J*. Vol. 40: 725-733, 2004

Röbbelen, G.: Einführung. In: *BioEngineering für Rapssorten nach Maß. Vorträge für die Pflanzenzüchtung*. Vol. 45: 9-27, 1999

Ross, J.; Li, Y.; Lim, E.-K.; Bowles, D. J.: Higher plant glycosyltransferases. In: *Genome Biol*. Vol. 2: 3004.1-3004.6, 2001

Sabot, F.; Schulmann, A. H.: Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. In: *Heredity*. Vol. 97: 381-388, 2006

Sambrook, J.; Fritsch, E. F.; Maniatis, T.: In: *Molecular Cloning - a Laboratory Manual*. Cold Spring Harbor Press. New York, 1989.

Schoch, G.; Goepfert, S.; Morant, M.; Hehn, A.; Meyer, D.; Ullmann, P.; Werck-Reichhart, D.: CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. In: *J. Biol. Chem*. Vol. 276 (36): 36566-36574, 2001

Shahidi, F.; Naczk, M.: An overview of the phenolics of Canola and Rapeseed: chemical, sensory and nutritional significance. In: *J. Am. Oil Chem. Soc*. Vol. 69 (9):917-924, 1992

Sheahan, J. J.: Sinapate esters provide greater UV-B attenuation than flavonoids in *Arabidopsis thaliana* (Brassicaceae). In: *Am. J. Bot*. Vol. 83: 679-686, 1996

Shirley, A. M.; McMichael, C. M.; Chapple, C. C. S.: The *sng2* mutant of *Arabidopsis* is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. In: *Plant J*. Vol. 28 (1): 83-94, 2001

Sinlapadech, T., Stout, J.; Ruegger, M. O.; Deak, M.; Chapple, C. C. S.: The hyper-flourescent trichome phenotype of the *brt1* mutant of Arabidopsis is the result of a defect in a sinapic acid: UDPG glucosyltransferase. In: *Plant J*. Vol. 49: 655-668, 2007

Smith, N. A.; Singh, S. P.; Wang, M. B.; Stoutesdijk, P. A.; Green, A. G.; Waterhouse, P. M.: Total silencing by intron-sliced hairpin RNAs. In: *Nature*. Vol. 407: 319-320, 2000

Snowdon. R. J.: Cytogenetics and genome analysis in *Brassica* crops. In: *Chromosome Res*. Vol. 15 (1): 85-95, 2007

Stefansson, B. R:, Hougen, F. W.; Downey, R. K.: Note on the isolation of rape plants with seed oil free from erucic acid. In: *Can. J. Plant Sci*. Vol. 41: 218-219, 1961

Stoutjesdijk, P. A.; Singh, S. P.; Liu, Q.; Hurlstone, C. J.; Waterhouse, P. A.; Green, A. G.: hpRNA-mediated targeting of the Arabidopsis *FAD*2 gene gives highly efficient and stable silencing. In: *Plant Phys*. Vol. 129: 1723-1731, 2002

Strack, D.; Nurmann, G.; Sachs, G.: Sinapine Esterase II. Specificity and change of sinapine esterase activity during germination of *Raphanus sativus*. In: *Z. Naturforsch*. Vol. 35c: 963-966, 1980

Strack, D.: Sinapine as a supply of choline for the biosynthesis of phosphatidylcholine in *Raphanus sativus* seedlings. In: *Z. Naturforsch*. Vol. 36c: 215-221, 1981

Strack, D.; Knogge, W.; Dahlbender, B.: Enzymatic synthesis of sinapine from 1-O-sinapoyl-β-$_D$-glucose and choline by a cell-free system from developing seeds of Red Raddish (*Raphanus sativus* L. var. *sativus*). In: : *Z. Naturforsch*. Vol. 38c: 21-27, 1983

Strack, D.; Mock, H.-P.: Hydroxycinnamic acids and lignins. In: *Methods in plant biochemistry: enzymes of secondary metabolism*, 45-97, 1993

Sunilkumar, G.; Campbell, L. M.; Puckhaber, L.; Stipanovic, R. D.; Rathore, K. S.: Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. In: *Proc. Natl. Acad. Sci USA*. Vol. 103: 18054-18059, 2006

Tettelin, H.; Masignani, V.; Cieslewicz, M. J.; Donati, C.; Medini, D.; Ward, N. L.; Angiuoli, S. V.; Crabtree, J.; Jones, A. L.; Durkin, S. A.; DeBoy, R. T.; Davidsen, T. M.; Mora, M.; Scarselli, M.; y Ros, I. M.; Peterson, J. D.; Hauser, C. R.; Sundaram, J. P.; Nelson, W. C.; Madupu, R.; Brinkac L. M.; Dodson, R. J.; Rosovitz, M. J.; Sullivan, S. A.; Daugherty, S. C.; Haft, D. H.; Selengut, J.; Gwinn, M. L.; Zhou, L.; Zafar, N.; Khouri, H.; Radune, D.; Dimitrov, G.; Watkins, K.; O'Connor, K. J. B.; Smith, S.; Utterback, T. R.; White, O.; Rubens, C. E.; Grandi, G.; Madoff, L. C.; Kasper, D. L.; Telford, J. L.; Wessels, M. R.; Rappuoli, R.; Fraser, C. M.: Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". In: *Proc. Natl. Acad. Sci USA*. Vol. 102(39): 13950-13955, 2005

Thomas, T. L.: Gene expression during plant embryogenesis and germination: an overview. In: *Plant Cell*. Vol. 5: 1401-1410, 1993

Tkotz, N.; Strack, D.: Enzymatic synthesis of sinapoyl-L-malate from 1-sinapoylglucose and L-malate by a protein preparation from *Raphanus sativus* cotyledons. In: *Z. Naturforsch*. Vol. 35c: 835-837, 1980

Town, C. D.; Cheung, F.; Maiti, R.; Crabtree, J.; Haas, B. J.; Wortmann, J. R.; Hine E. E.; Althoff, R.; Arbogast, T. S.; Tallon, L. J.; Vigouroux, M.; Trick, M.; Bancroft, I.: Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. In: *Plant Cell*. Vol. 18 (6): 1348-1359, 2006

Townsend, B. J.; Llewellyn, D. J.: Reduced terpene levels in cottonseed add food to fiber. In: *Trends in Biotechn*. Vol. 25(6): 239-241, 2007

Truco, M. J.; Hu, J.; Sadowski, J.; Quiros, C. F.: Inter- and intra-genomic homology of the *Brassica genomes*: implications for their origins and evolution. In: *Theor. Appl. Genet*. Vol. 93: 1225-1233, 1996

Tzagoloff, A.: Metabolism of sinapine in mustard plants. I. Degradation of sinapine into sinapic acid & choline. In: *Plant Physiol*. Vol. 38: 202-206, 1963

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

Van de Peer, Y. ; de Wachter, R. : TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. In: *Comput. Appl. Biosci*. Vol. 10: 569-70, 1994

Vogt, T.; Jones, P.: Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. In: *Trends Plant Sci*. Vol. 5 (9): 380-386, 2000

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

Wang, Q.; Dooner, H. K.: Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. In: *Proc. Natl. Acad. Sci USA*. Vol. 103(47): 17644-17649, 2006

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

Wendel, J.F.: Genome evolution in polyploids. In: *Plant Molec. Biol*. Vol. 42: 225-249, 2000

Winter, M.; Herrmann, K.: Esters and glucosides of hydroxycinnamic acids in vegetables. In: *J. Agric. Food Chem*. Vol. 34: 616-620, 1986

Yin, H.; Li, S.; Zhao, X.; Du, Y.; Ma, X.: cDNA microarray analysis of gene expression in Brassica napus treated with oligochitosan elicitor. In: *Plant Physiol. and Biochem*. Vol. 44: 910-916, 2006

Yu, B.; Lydiate, D. J.; Young, L. W.; Schäfer, U. A.; Hannoufa, A.: Enhancing the carotenoid content of B. napus seeds by downregulating lycopene epsilon cyclase. In: *Transgenic Res*. DOI 10.1007/s11248-007-9131-x, 2007

Zum Felde, T.; Becker, H. C.; Moellers, C.: Genotype × Environment interactions, heritability, and trait correlations of sinapate ester content in winter rapeseed (*Brassica napus* L.). In: *Crop Sci*. Vol. 46: 2195-2199, 2006

Zum Felde, T.; Baumert, A.; Strack, D.; Becker, H. C.; Moellers, C.: Genetic variation for sinapate ester content in winter rapeseed (*Brassica napus* L.) and development of NIRS calibration equations. In: *Plant Breeding*. Vol. 126 (3): 291-296, 2007

# 7 Supplement

## 7.1 Cloning of *UGT84A10* and *UGT84A11*

**Figure 7-1. T7 derived sequence of the 5'RACE fragment of *UGT84A10*. Features given: SMART II A oligonucleotide (1-21 bp) in green, 5' end of the UGT84A10 ORF (209-432 bp) in red.**

```
  1  tatcaacgca gagtacgcgg ggattttatt tatttgatct aaattgtttt ggttagtgac
 61  tctttaaaaa ctgtttatat tcttcatttt ctttttctta attatgtttc gtccataccg
121  tcactaatcc ctattaaaaa gggatcaccg aggtggagag ggtgaaaaag atttagcagt
181  cttgtttttg ttatcaagaa aagtagagat ggaatcgtcg ctgactcatg tgatgctcgt
241  ctcattccca tcacaaggtc acataaaccc acttcttcgt ctcggaaagc tcatagcctc
301  taaaggctta ctagtcactt ttgtcacaac agagaagcca tttggcaaga agatgcgtca
361  agccaacggg attcaagacg gctcgcttaa acccgtcggt ttaggtttcc tccggtttga
421  gtt
```

**Figure 7-2. SP6 derived sequence of the 5'RACE fragment of *UGT84A10* harboring a continious *UGT84A10* ORF. The whole sequence shows 93% identity to BH591016 on nucleotide level.**

```
  1  tcccaccgta gcctcagcct ccttcttcca cctccgagca ttttctctca gctccacggc
 61  cttatctcca accaccgctt caagtagctt ctccgccaca acctccctag aaacaatttt
121  ttcatcagcc tctccgcggc caagtcttac tcctgtctta aaaacatcaa ccaagtacaa
181  agcatttgtc acttggtcgc cccactgcgg tagacaaaca atgggaacgc cagaggacag
241  agcctccact gttgagttcc atccgcagtg acttagaaag caagcaaccg caggatgagc
301  caatactctc tcttgtggac accattcaac gatcattcct tgctcttcga cctctcgagg
361  caacacatgt ggtgctcggg ataacccttc cgtaggaggc cgaaccaccc ataagaagga
421  taagccggag tttagaaggg cgtgagcgat ctcgtcgatc tgctcttgct tcacgtggac
481  tacagtccca aaggagatgt aaacgatgga ggatggatct ttcgagtcaa gccactccat
541  gcactgattt gcagagtcag agatgtctcc cttgatatcg gagctgtg
```

**Figure 7-3. Database fragment BH591016 (UK CropNet) containing the 3' end of an ORF for a putative BnHCA-GT. Features given: region of homology to the 5' RACE fragment for *UGT84A10* (1-585 bp) is in red, putative stop codon (650-652 bp) in green**

```
  1  gagctctgat atcaagggag acatctctga ctctgcaaat cagtgcatgg attgtcttga
 61  ctcgaaagaa ccatcctcca tcgtttacat ctcctttggg actatagtcc acttgaagca
121  agagcaaatc gacgagatcg ctcacggcct tctaaactca ggcttatcct tcttatgggt
181  ggttcggcct cctatggaag ggttatccct ggcaccacat gtgttacctc aagagctcga
241  agataaagga atgatcgttg aatggtgtgc acaagagaga gtgttggctc atcccgcggt
301  tgcttgcttc ctaagtcatt gcggatggaa ctcaaccgta gaggctctat cttctggcgt
361  tcccattgtt tgtctaccgc agtggggaga tcaagtgaca aatgctgtgt acttggttga
421  tgtgttcaag acaggagtaa gacttggccg cggagaggct gatgagaaga ttgtttcaag
481  ggaggttgtg acggagaagc tacttgaagc gacggttgga gagaaggcgg tggagctgag
541  agaaaatgct cggaggtgga agaaggaggc tgaggctacg gtggggtatg gtggatcatc
601  tgataaaaac tttggagagt ttgtggataa gttggttgcc aacatctgtt gatatctata
661  attatttcac cttcagttac tc
```

**Figure 7-4. Sequence of the 3' RACE fragment for *UGT84A11*. Feature given: 3'end of *UGT84A11* Orf in red**

```
   1  cgacgagatc gggacggttg cgtatttgaa gcaagaacag atggaagaga tggctcatgg
  61  agttttgaaa acgggtttat cgttcttgtg ggtgatacga cttccattac cagatctgaa
 121  gctggagact catgtcctgc ctcaagaact gaaagaagcg agtggtaagg gattagggaa
 181  gattgtggag tggtgtccac aagagcaagt gctggctcat tcttcagtgg cctgttttgt
 241  gactcactgc ggatggaact cgacaaccga ggctttgact tcgggggttc cggtggtttg
 301  ttttccgcag tggggagatc aagtcaccaa cgcggtttat cttattgatg tcttcaagac
 361  gggagtgagg cttggctgtg gggcggcgga cgagaggatt gtgccgaggg aggacgtagc
 421  ggagaagctt ttagaagcga cggttgggga gaaagcgaag gagctgagga agagtgcttt
 481  gaaatggaag gcggaggcgg aagcagctgt ggctccgggg ggttcgtcgg aaaagaatct
 541  ccgggagttt gtggagaagt taggcgtgat cagtggctga atgagataaa tatttcttag
 601  gtaagaataa attctaaata aatattaaaa taatgtccta taaaaggcta taaagtattt
 661  atgttactgt ttattgaaaa catgtactct ctaccctatc tt
```

**Figure 7-5. Sequence of the genome walking fragment carrying the 5'end of the *UGT84A11* ORF. Features given: Genome walker adaptor (1-48 bp) in green, UGT84A11 ORF (319-1311 bp) in red**

```
    1  gtaatacgac tcactatagg gcacgcgtgg tcgacgaccc gggctggtat ctctgtgacc
   61  atgttgataa ttttctaatt aaattaaagt tatacgttcg tatcttttaa attaaattaa
  121  gattgagaaa aacgagagtg gtaagtaagg tgtgaggtgc tgacaaagaa aaaaacaaag
  181  gaaagatgtg cgttagagaa agtgagtgag taggtctata aacgtttcgg cgcttctata
  241  aataaagaat ggtgctgagt aggtcttttt tgtaatacaa gtcttcataa aaataattgt
  301  taccaaattt gaacaaagat ggagttggaa tcttcttcac attcgagtcc agttcatgta
  361  atgctcgtgt cttttcacgg acaaggcagc gtcggtcctc ttcttcgttt cggcaagcta
  421  attgcttcaa agggaacagt cgttacgttc gtcactacgg agtactgggg gaagaaaatg
  481  agacaagcca accagatcgt cgaaggtgag cttaaaccgg ctggttccgg ctcaatccgg
  541  tttgagtttt tctacgatgg atgtgcagaa gatgatgtcc ggagaggtac caccttgtac
  601  atgccacggc tagaacaaac cgggaaacga gaagtatcga aactcgtgag gagatacgag
  661  gagaagaacg agcccgtctc ttgtctgatc aataacccgt ttgtcccgtg ggtcggggac
  721  gtggcggaag agttaaacat cccgtgcgca gtgctgtgga tacagtcttg tgcttgtttc
  781  tctgcttatt atcattacca aaatggctct gttcctttcc caaccgaatc agctcctgag
  841  ctcgatgtaa agctcccatg cgttcctgtc ttaaagcacg acgagatcca tacctttctc
  901  catccttctt ctccattcac cggaatgcga gacgcgattc ttgggcagtt taagaatctt
  961  agtaagtcct tttgtgttct aatcaattcc ttcgacgctt tggaacaaga agtgatcgat
 1021  cacatgtcga aactctttcc aataaaaacc attggaccgg tttttaaact cgctaagacg
 1081  gtgatctccg atgtaagcgg cgacttctgc aaacccgcgg atcaatgcct cgactggtta
 1141  gactcgaggc cagaatcgtc agttgtctac atctctttcg ggacggttgc gtatctgaag
 1201  caagaacaga tggaagagat ggctcatgga gttttgaaaa cgggtttatc gttcttgtgg
 1261  gtgatacgac ttccattacc agatctgaag ctggagactc atgtcctgcc a
```

## 7.2 BAC fragments

**Figure 7-6. *Nco*I-fragment of BAC1 carrying *UGT84A9a*. Features given: S7-locus region (1365-2198 bp) in light blue, *UGT84A9a* (3208-4701 bp) in blue.**

```
   1  aatagaagct ctccatcttc atgtgttgca gcattgtcct aacatctcct tcgattccta
  61  gctcttctat catcttggtg tcagcgaagc gtgttggaac ccaagaaaca ccattaaaca
 121  agtgatgata tagctcatca gctgttggtg tcttgctctt gtctctatca agagcaatcc
 181  cctgttcttt agacatcttg gctttctttg aaggagctgg ctcttctcct ccttcatccc
 241  tagcagaaga tgccacagct ttaccctctc tcttctctct ttcttttgcc ttcttctcag
 301  ctagaacttg ttgcctagat gttctctcca ctcctcgttg acttggctca gaaggttcta
 361  ggggattttc cctaagtgca gtttccctct ttgcagttgc tgccttctta gccttcatat
 421  ctgttttgtt tgtttttgtc attgtacctt taacattcaa accttggtaa agtattagca
 481  agaggtacaa aagatgagta aaaatgaagt tgcaaatgtg aattcaatca attgaataat
 541  aagctttagg caagactatc acattgcaac atcaatagag aaccaggtct aagagataga
 601  gttataaaca aagaatcaag ttcctcaatg ttctcaagca tgaaaatggc aattaagtgt
 661  ttttcattga acaattgtgc tctaaattga gagaacttcc taagcaaact ctaaaacatc
 721  ccaatacctc aacccaaaca acattcaaac tcaatgtgga caatttcgaa aatttccaaa
 781  tctcaaaaac cctaattttc caatttcgga atcaactcaa ttctaccatg aaattagcaa
 841  cccaacatga tatataagtt tttcctagtc aaattcacaa gaatcaagca ataagctaaa
 901  ttctagtttt aatttctaga gtttatgaac ccacaaaatt catctttgaa tctccatacc
 961  ttgcttggat tagaatgaaa agatgaactg aattaagaaa atagactata ggggcgaaag
1021  cttggattca gaaacaagaa tgagctgatt tggttaagat tagagtgagt tatggggttt
1081  ttggagttgg gaaatttgag agagtgtggt tgtgtttagg ggaaaagaga ggaaaacagg
1141  agagagagag acacgtgagc ggggtagggt tagggtgaag tcaggtcggg tcgtcgggta
1201  tgggtcgggt cgtcgggttt cagatcaaga aatcatacct gggtggttgt ggagcgactt
1261  cggtacctcg ctgtgagggc tcgctctgaa accgtcgaca ggttcttcga gtggtggcgc
1321  gacttcgata ccccgctgcc ataccccggc aacggcgtca aatttgaaat ggacttttca
1381  agggtccaaa gttcaaatca aagtagtatt ttagatgtca atctatttct aagtgtttca
1441  attcaaagag aattcaggtt catacttaag ctatgtgaat tcaatgtaat gaagtgattc
1501  aatcaagcta acaagattct aacacaacta actagcaact ttcaagcaaa tggataaagg
1561  aggaatcatg ggtatagaaa tttgatttca gatgactaag atttaatcta aaatagcaag
1621  gtttcaatca acacattccc ttaagtctag atagcaattc taagcaagtt ctttgtcaag
1681  acaaaagctc atttactctc attgatcaaa catcaaattc ctttggtttg tgtcaatcaa
1741  acaatcatta agaatggatt attcaacttt ctaaactttc ctaacatcaa atccctttgg
1801  tagggtaagc taagagcatg ttgagttggc tcagacattt catcgaacac cttccgagca
1861  atgaaatgtc tagagttcta atttaaaatg gccaactcta gattagcatt aagatcactc
1921  aatcaagcaa ggaaatatat taatctactc taaaacattc tagatcatca cttaatcatc
1981  ctaatcatcc tagcccatga atccaaagat gactactcac tcattatcat ggtagacact
2041  aaatcattag ttgatctaag tctaaacatg attaatgatc aaagcaatca agtaaacaca
2101  agagaatatg atcaagatta gatctttcac caaaaagtag atttttgatt agatagaaaa
2161  caagataaga tcctaacatt aggtttagag agtcattagg tttagagagt atctatatga
2221  cttctaaggt ctcgtgggct cagtcaacaa cctgaaaggt ccaaataaaa cataaatttt
2281  tgattagaag aaggggaacg ccacgggttc attaggtcgc tccgccaggt cgctctgctt
2341  ctagagcggg ttcgtcacct cactccggga cgtcgctcca atatgcttgc ttcatttctt
2401  catattgtgc gcgggttcat gaggtcactc caccaggtcg ctctgcttct agagcgggtt
2461  tgccacctcg ctctgtcttg gtcgctctaa ttttcgaaag tcgtcgacga ggtgttggtt
2521  cagagcgggt gcgtaacccc gctgcgggac gtcgctcttc cttcgtttcg tctggagcgg
2581  gtgtcgtagg tcgctcctgt agcccgctcc agtgttctac tcgctaaaac aacactttat
2641  actctctttt tgatctcaaa tgcttccaag tacctgtaag aactccaatg gacactccaa
2701  cacttgataa agacttatgc aatgcaaaaa agaaactaaa atgcataatc cctaatctaa
2761  atgaccaaaa catgcaagaa tgaaaaatta aaacaatgca aatgtgcaag atatcacctt
2821  tgtaaaatac aaatgttgtc agtttgatcc atgtattttt ttgtatatat tgtggacatg
2881  cgggtaggac cgtaggatga acgttaacgt tcaatcgttt ggacctaaat gtggtcatac
2941  atacaaagaa aaaatgatcg aatattcctt aacacgacat ggaatacagt aatacagagg
3001  taacgttaca ctgacatgtc tcttctttgg ttaataaaaa gaacatatct tcttttgttt
3061  tcaaaatgta aataatcgag aaactaacga gtgtcaaatt atccacgtca ccttctttgt
3121  cttgtattta agcatatccc ttgacatgtc acatacttga gaaggataaa aaaacagagc
3181  acacagaaga gaaccccttt ttgaaggatg gaactatcat cttctccttt acctcctcat
3241  gttatgcttg tatccttccc aggacaaggc cacgttaatc cacttcttcg tctcggcaag
3301  ctcttagctt cgaagggttt actcgtcact tttgtcacca cagaatcatg gggcaaaaag
3361  atgcgaaccg ccaacaagat tcaagaccga gccctcaaac ctatcggtaa aggttatctc
3421  cggttcgatt tcttcaacga tgggctccct gaagacgacg atgcaagcag aaccaactta
```

```
3481  accatcctcc gaccacaact cgagctggtc ggacaacaag agatcaaaaa cctcgtgaaa
3541  cgttacaagg aagtgatgaa acagcccgtg acgtgtctca tcaacaaccc tttcgtctct
3601  tgggtctgtg acgtagccga agatcttcaa atcccctgtg ctgttctctg ggtccagtct
3661  tgtgcttgcc tagcttctta ttattattac aaccacaagc ttgtcgactt cccgaccgaa
3721  acagatccca agatcgatgt ccagatcccg tgcatgcctg tcttgaaaca cgacgagatc
3781  ccttctttca tccatccttt ttcaccttat tcgggtttaa gagaagtgat cattgatcag
3841  atcaaacgtc ttcacaagcc tttcgctgtt ctcatcgata ctttctactc cttggagaaa
3901  gatatcatcg accacatgac aaacctctct cgcaccggct ttgtcagacc gctcggaccg
3961  ctttacaaaa tggccaaaac gttgatttgt gatgacatca aaggagatat gtctgagacg
4021  agggatgact gcatggagtg gttagactcg cagcctgttt cgtccgttgt ttacatctca
4081  tttggtaccg tggcttacgt gacacaagaa cagatcagcg agattgcgtt aggcgtttta
4141  aacgctgacg tttcgttctt gtgggtgata agacaacaag aactaggtgt aaacaaagag
4201  cgacatgttc tgcctgaaga actcaaaggg aaaggtaaag tcattgaatg gtgttcacaa
4261  gagaaagtct tagctcatcc ttctgtggtt tgtttcgtga ctcattgtgg atggaactca
4321  acgatggaag ctttgtctag tggagtccca acggtctgtt ttcctcagtg gggagatcaa
4381  gtcaccgacg ctgcttacat gatcgacgtg ttcaagacgg gagtgaggct tagccgtgga
4441  gagacggagg agagggtggt gcctagggag gaagtagcgg agaggctgag agaagttacg
4501  aaaggagaga aagcgacgga gctgaagaag aatgctttaa aatggaagga ggaggcggaa
4561  gcggccgtgg ctcgcggtgg ttcgtcggat cggaatcttg atgagtttgt ggaaaagttg
4621  ggcgtgaaac ctgtggctaa acagaacgga agtctcaatc aaaacggaag tattcaaaaa
4681  cttttattgc aaaagtcata aatgtgtttt gtattatctg ttgtttgttt tctgcaaatc
4741  ctgataattg tgtttgtttg tatgctccct ccgttctttg ttctcgtttt ttttttgtat
4801  gggtttgggg agactagtta tccttcttga ataaatgtta gatcacgact tgtgttctta
4861  agattaatta gtttgtttat ttttaaattc ggaaatatat tttgtatatg atttatgtgg
4921  atgcacaatt gcacactaaa attaagaaga gtcggtgttt ttttttttt ggaaattaaa
4981  agaaaaataa cagttactta ataatagtcg gtctgcacac ataagacaat tttgtaagcg
5041  aaatcaaaac ttgagagatc atatgtaagg tgatatacaa aaaaattgtt taggaatagt
5101  ttgattggtg actgccttaa aattttgtga gtttacaaaa catctaaaaa gtgaaaagta
5161  attataaata ctggttttgc ttattatttc tataaaaaat atgtattaat atattacttc
5221  cacaattcac aactttgaat actcagaatc tcagataacc aaggattatg tatgtatcct
5281  catcgtctcg aatatttcga cgttaacaaa ctcagacaga cagaaacacg tgttggatag
5341  tatctgcatg gggttgaaaa tatatgagtg tgactgtatt agagtaaatg ttttggagca
5401  aaagtatttt actcagaatt cagctgaaag attaccaatc agataaaaaa ttattctttt
5461  atttttattc tagacactgt tgaaaaagcg gaaaaagtac cctgaacttt actcttgatt
5521  tgggaggtag aatatattaa cgctaattac tgttgctttt gtccttttcc acctttttata
5581  ttttcctctc atttttattt ttcaatttac tccactttat tctcataatt accggtcaca
5641  aactatattt ttacctgatt gtgtatacgt tttttgtttt gctcgtgtat acgttagttt
5701  tcttaaaaca tactaagaga taatacaact cgtgtgtaat taaagcttta tttaatttta
5761  tcagctgtag attgttaagt ttgactaata acgactattg aaattttctc acaggcacgt
5821  aggaatcata tgcaccaata gattatacag tgtatgttcg ttaggtccgt gagtaaaata
5881  aagtgcacac agctttttgg aaactacaaa aaggaagatt agaggattcc gcatcgacca
5941  ataataataa ttatatgaaa aaataataat tattattata tagcattcat tgtggacggg
6001  agaagataga actcttaatt acaaagcctc gtcgcctatt agtaagccac tgcatatttt
6061  agccgattca ctgctggagt ataacattgg tactcggacc cgttatctat cgacatgggg
6121  ggcactgtta catttaatgc tgtaacaggt gaatcattgc attctccgtt aaaagctaag
6181  ccatacggtt ttagctagtg tcgttgtgcg tttagtacta ttttgatgaa ttgaattatt
6241  gcctcaaaaa gatcagttca tcttttattg gtagacaaag cggtccaaat cgaagtcata
6301  ataagaaata catccaagag gggaggtgtt gtggtaagta agatgtgtgg tatgtgtggc
6361  agaaaaaaat gtggcaggta ggcaaaacct gcgtcatatg caatttctag aatgattcat
6421  ttaaaacgca aacgttatac cacccccttg gtggtaagta gaatgtgtgg ccaaaaacaa
6481  tgtggtaggt tgggaaaaaa atttaagtca tagtcgattt cttaaattac gtacatctcc
6541  taccaaacaa gccacgtcaa ttattgataa atgaatttat gaagattttc aaaacatttg
6601  ttatattata tagatttgtg aattatgaca cttgtaaact ctgttattgt tgattttaaa
6661  ttttattcaa aagtactgtc atttagatag ttcggtttca taatcgattg gtctgattat
6721  gtaatttcac aaaatcatat ataactgtaa tatgaatatc atatcagttt atatataaaa
6781  tgcatttttа aaaatactat aaaatatttt ttttttaatt caagaagttc aatactgaat
6841  cttataatat ttttggccta aagataagat ccggttgcta aatcccttcc attgctctgc
6901  taaaactcga atatggtact ttttctgaaa cactcgaata tggtacttat atctttaagg
6961  tataaggttt taccagtaaa aataatagct cctagtaata caatagatta ttatctactg
7021  gatctctgga tatatatatt tcttttgtaa atataacgaa tttataacat atttacattt
7081  tgtaattttt gtatttttтg gctgacatta taaagataaa ctagattatg aaccgtgatt
7141  taaaagagca taaacaattt cttaacaaaa tgtaatttac aaaattaata tgttgtaaca
7201  ttgttatttt agtttatttg aaaactatat aattatatta ctttctaatg taattgtata
```

```
7261  aaatatgaaa ttatataata tattacttaa tttgtttttc aattatttca tccccgcttt
7321  attatgaatt tttaataaaa tctcagtaat ttgcttgatt agttatgtga tatacatttt
7381  gatataaata aaataatcta ttaatttaat ttatttcata ttaaatgaaa tcattttata
7441  atatgtaaac ctgctaaaaa tagttataaa aacttattaa aaaggtaaga tctaatgtta
7501  gatatttttt gaacatgtgt tgtttggaaa catatatagt agtataatgt tattttccat
7561  aattattctt aataaagatc taatttttta ttattactaa taacatttaa tgtattataa
7621  tgtatcatta aaggttaggt ctaaataaat gtttctgttt tactaagaga gatggttata
7681  acaatttcga aatattatga aagtataa
```

**Figure 7-7. Combined *Nco*I-fragments carrying *UGT84A9b*. Features given: incomplete ORF homolog to *At3g15310* (1458-2434 bp) in red, *UGT84A9b* (4581-6074 bp) in blue, 5SrRNA ORF (7099-7225 bp) in red**

```
   1  ccatgggact aagcgccacc gttcatagct acaaccgcct tgttcatagc catcatcgct
  61  tccatggaac actgttggag ggaactgaga catgatcaga agtggaagtc agtttttaag
 121  tcaagagatg gcggcaagga gaaaaggaag gaagctgagg aggtgatacc tgaggaggag
 181  gttagaccgc ctggtgttaa ggcttccaaa gaagccaaac gcaagaggca cggtcatgaa
 241  gctgctttcg atcaaattga gagtatcttg gctgagagaa agaaaatttc tcagcagaaa
 301  ctcctagatc gtctccttgc caaaaaggaa actgatctat ctcctaacga aatcaccctc
 361  aaaaacaaac tcatttctga actccttgat tgaattggtt aacttgcttg ctcgaatctc
 421  tgtttttttct tttgttgaac tcgatatgtt tagttattaa ctttgaagtt gccttttttct
 481  tttgcaggtt tctatgtcac gggttgctgt gtctgttgct ggatcacggg ttgctatgtc
 541  tggatcacgg gttgctatgt ctgttgctgg tattaataag tctgttgttg ctttatttgt
 601  acgggatcag tgtagtattt tgttagtttc tacactgatg ttgctttatg atgaacggat
 661  gcagtggagt ctttctcgca ctgctactat gtccttttgt ttcacgggtt gtactatgtt
 721  ttgacactag tttatttaag tcacgggttg tactatgtaa caaagaagca tatataaata
 781  tgctcttgat gtatgtttct aaatcagaac tcaactactt ctctttctgt ctctgtcttt
 841  gaatttcaaa atttcaagaa cacaacttct gaaacacttg atcaactact tctcttcctt
 901  gatgtctttc tgaaacactt gatcaactat aaacgaagag gacaactatg cttctactgc
 961  ctcacccaac cgagaacaac acaaccaacg taagctcttc attacattat aatactaagg
1021  ttaaacgaga ttgacatgat cagacttacg caatgttaat tgcaaattga taaaattcat
1081  aattgataat ttgttatatt ttggactaga gaataaaatt gataatttgt tatattttgg
1141  agtagagaat aaaattgata atttgttata gtttggacta gagaataaaa ttgataattt
1201  gttatatttt ggactagaga ataaaattga taatttgtta tattttggag tagagaataa
1261  aattgataat ttttatttttt tgggctagag aaaaaaaaat gataatttgt tatattttgg
1321  agtagagaat aaaattgata atttttattt ttttgaaaaa agaaagaatg tcaagctcct
1381  caaatgatga agtatatgaa gtttttgagg aaatggtgga ccaacaaatc gatgatttca
1441  tcgactctgt tctaaccaaa gagccgaaga gacgagtata catcgaaaga gatcgggaac
1501  aaggacacaa tcaactatgg caggattact tcagtgaaaa tcctacatac acacacgaca
1561  tgtttaggcg gcgttttcga atgaacaaat ctttgttcct tcgcattgtc gaacgcctag
1621  gtaatgaaat tccatacttt cagcaacgaa gaaatggtca tggaaggaac ggcctctcta
1681  cacttcaaaa atgcacttca gccatacgat cgttggcata cggtcaagcg ggagatgtca
1741  atgacgaata tctccgactt gcggcaagta ctgcaatttt atgtttgtag aattttgcgg
1801  aagcgataat acaattgttt ggagatgagt atctaagaac acctacaaca gaagatcttc
1861  aaatactact tgacgctgga gaggcacgcg ggtttccggg gatgataggc agcatcgatt
1921  gtatgcattg ggagtggaaa aactgcccta cggcttggaa aggtcagttc acacgtggtt
1981  caggaaagcc gacaattgtc ttagaagctg tggcatcaca agatctttgg atatggcact
2041  cattttcgg attaccaggt acccttaacg gtatcaatgt tcttgatcgg tcaccagttt
2101  ttgatgacat tttacaaggt cgagcaccta aagtgaagtt caaggtcaac aaccacactt
2161  atcatatgac ctactatctc accgatggaa tttatcataa ttgggcaaca tttatccaat
2221  ccatcccgct tcctcaaggt cctaaagccc aaaaatttgc agaaaggcaa gaatcagtca
2281  gaaaagatgt cgaacgggca tttggagtgt tgcaatcgag gtttgcaatt gtaaaaaacc
2341  tgctctaaaa tggaacatgg aaaagatagg aaagataatg agaacgtgtg tcatattgca
2401  caatatgata gtagaggatg aacgacacgg atacactcta gctgatacat ctgagttcca
2461  aggtgaaaac gagggaaagt gtgaatgtcg atatgcttaa cattcgcaat caaattcggg
2521  atccaggtgt acatgagcgt ttgaaagctg atttagttga aaatgtttgg gaaaaatttg
2581  gtaatcgtga tgaataatct ttgtatgttc atgatttttc aatgtattga ataaaaactt
2641  ttcataaata atgtaaaata tttagtaagt ttctttaaaa aaataatttt tttaaatgaa
2701  aaattatata tatttttttc agaaaaaata ttattttatt tctatgaacc cggatttgag
2761  gttcattgat ggaggagcac aattgatccg ggttcaacac tgttcacggg cccacaatta
2821  aatacataaa aaatacatat gaatcccaat agggggttga tgtacatgct cttagataac
2881  caacggtata acaacacgtc acctttttaa tttcgtggaa ctaaactaat aaagtctccg
2941  agatttccat gactgatgat tctatacaaa catagaagaa gaaaacatat ataagtagag
3001  gtatttagtg tttcaaaaaa agaaaaagta gaggtattta attcgccacg ccggtctaac
3061  ttgttttgtt actttcttgt ttgataataa tgtccactaa cagaatatta ttgtccaaaa
3121  gtatctcgaa ttagtcaact agggtcaatt tcaccatcga tggttagaaa agtagatatt
3181  attttggatg aagctagact tgaatttgaa acgatattat attttggtga attgtgtggt
3241  tcactggttt tgctatgcat tgttgattat tgttctcaaa tctgatggaa tactctttta
3301  agtaaccaaa aagcaaataa ctttgtagat attataaaag tgccattaac caaaagcatg
3361  tattaatatg gtacaggagt atcaagtttg ataccattta aatttattca gtaaaaaaaa
3421  gtaaaattat ttccaaaagt ctatcttctc agtcgtaata ccgacataag ttctatactc
3481  aaaactatac tacaaaacaa atatattggt taaataactt ttatttgggc ttcgttttta
3541  tatactatca gctaatttca gactattttc aatatatttt gctaatttat ttttaaaata
```

```
3601  gataaattta atttataata gaaataaagt tttattccaa tgtttttcaa atgataattt
3661  ttaattataa aacaaaataa aaattactat ttttctattt ttttagagaa aataatgaca
3721  ttctttattt tacatgaaaa aatatagtta aactggagtg gatttacctt taaatgctat
3781  tataaaaaca aatataaaag tgggtccgaa taattcttac agagaaaatt aacaagagta
3841  tcatacataa ggcacgtcat ttggattaaa attaaacact tcagctaacg ttacttgcca
3901  acttcaagta cgactgatag tgggggagct accatagtgg catgtgggtc aaatgaccca
3961  tgtaaaattt tgtaaatttt tttttacaaa atttacatgt atttcttgca taaaaatcta
4021  aagagtaagc tggttgaccc ttgtaaaata caaatgttat cagtttgatc catgtatttt
4081  ttgtatattg tggacatgga ggtaggaccg taggatgaac atccaatcgt ttggacctaa
4141  atgtgggtca taaatacaaa gaaaaaatga tcgaatattc cttaacacca catggaatac
4201  agtaatacag aggtaacgtt acactgacat gtctcttttt aggttaataa aaaaaaaatc
4261  ttcttttgtt ttcaaaatgt aaataaccga gaaactaacg agtgtcaagt tatccacgtc
4321  acctacttaa ccaagttacc tttcaagttt caagggcctc atgcattcaa tataattgaa
4381  taagacagta tggacaacaa tgcaacggtg gaaattgata actccttaat tatacacctt
4441  ctttgtcttg tatttaagca tatcccttga ctcttaacat gtcacatact tgagaaggat
4501  aaaaaaaaca gagcacacag aagagaaccc ctttttttt ttttttttgc taaaatttga
4561  gagaacccct tttttgaagg atggaactat catcttctcc tttacctcct catgttatgc
4621  ttgtatcgtt cccagggcaa ggccatgtta atccacttct tcgtctcggc aagctcttgg
4681  cttcgaaggg tttactcgtc acttttgtca ccacagaatc atggggcaaa aagatgcgaa
4741  ccgccaacaa gattcaagac cgagccctca aacctatcgg taaaggttat ctccggttcg
4801  atttcttcga cgacggtctc cctgaagacg acgatgcaag cagaaccaac ttaaccatcc
4861  tccgaccaca actagagctg gtcggacaac aagagatcaa aaacctggtg aaacgttaca
4921  aggaagtgat gaaacagccc gtgacgtgtc tcatcaacaa ccctttcgtc tcttgggtct
4981  gtgacgtagc cgaagatctt caaatcccct gtgctgttct ctgggtccag tcttgtgctt
5041  gcctagcttc ttattattat taccaccaca agcttgtcga cttcccgact gaaacagatc
5101  ccaagatcga tgtccagatc ccatgcatgc ctgtcttgaa acacgacgag atcccttctt
5161  tcattcatcc tttttcacct tattcgggtt taagagaagt gatcattgat cagatcaaac
5221  gtctccacaa gcctttcgtt gttctcatcg atactttcta ctccttggag aaagatatca
5281  tcgaccacat gacaaacctc tctcgcaccg gcgttgtcag accgctcgga ccgctttaca
5341  aaatggccaa aacgttgatt tgtgatgaca tcaaaggaga tatgtctgag acgagggatg
5401  actgcatgga gtggttagac tcgcagcctg tttcctccgt tgtttacatc tcatttggta
5461  ccatggctta cgtgacacaa gaacagatca gcgagattgc gtttggcgtt ttaaacgctg
5521  gcgtttcgtt tttgtgggtg ataagacaac aagaattagg tgtaaacaaa gagcgacatg
5581  ttttgcctga agaactcaaa gggaaaggta aagtcgttga atggtgttca caagagaaag
5641  tcttggctca tccttctgtg gtttgtttcg tgactcattg tggatggaac tcaacgatgg
5701  aagctttgtc tagtggagtc ccaacggtct gttttcctca gtggggagat caagtcaccg
5761  acgctgctta catgagcgac gtgttcaaga cgggagtgag gcttagccgt ggagagacgg
5821  aggagagagt ggtgcctagg gaggaagtag cggagaggct gagagaagtt acgaaaggag
5881  agaaagcgac ggagctgaag aagaatgctt taaaatggaa ggaggaggcg gaagcggccg
5941  tggctcgccg tggctcgtcg gatcggaatc ttgatgagtt tgtggaaaag ttgtgcgtca
6001  aacatgtggc taaacagaac ggaagtctca atcaaaacgg aagtattcaa aaacttttat
6061  tgcaaaagtc ataaatgtgt tttgtattat ctgttgtttg ttttctgcaa atcctgataa
6121  ttgtgtttgt ttgtatgctc cgttctttgt tctcgttttt ttttgtatgg gtttggggag
6181  actagttatc cttcttgaat aatgttagat cacgacgtgt gttcttaaga ttaattagtt
6241  tgctatattg ttaaattcgg aaatgtattt tgtatatgat ttatggggac gtggtgtaat
6301  ctataccatg cacaattgca cactaaaatt aagaagacta ggtgtttttt ttttttggaa
6361  attaaaagaa aaataacagt taaagttcat tactcgagag atcatatgta aggtgatata
6421  caaaaagttg tttgggaata gtctgattgg agactgcctt aaaattttgt aaatttacaa
6481  acatctacaa agtgaaaagt aattataaat actggttttg cttattattt ctataaaaat
6541  ataagtatta atatattact tccacaattc acaactttga ataatcagaa tctcagataa
6601  ccaaggatta tgtatgtatc ctcatcgtct cgaatatttc gacgttaaca actcagacag
6661  acagaaacac gtgttgggta gtatctgcat ggggttgaaa gtatttttg ttttgcttgt
6721  gtatacgtta gttttcttaa aacataagag ataatacaac tcatgtgtaa ttaaaagttt
6781  atttaatttt atcagttgta gattattaag tttgactaat aacgactatt gaaattttct
6841  cacatgcatg taggtatgta ccaatagatt atacagtgta taaagtgcac acagcttttt
6901  gttacctaaa cgactattgt actacatgcc tgtgagaaaa aaagctgtgc acacagcttt
6961  ttggaaatta caaaaaggaa gattagagga ttcgcatcca ccaaaaataa taattatata
7021  atgttcaaaa aaaaaattac ttatatagca ttcgttgtgg atgggagaag atagaagaac
7081  tcttattaga ctatgattgg gtgcgatcat accagcacta atgcaccgga tcccatcaga
7141  actccgtagt gaagcgtgtt tggacgagag tagtactagg atggatgacc tcccggggag
7201  tcctcgtgtt gcaccccttt tttatagtgt ttcaaaaaaa aaaaaagact atgattaacc
7261  tcggtttttt aactaggatt attacttttc ggttaagaga cgtttcataa cttttaatta
7321  agaaaaacta agaactatct tttaaataag agatataaga aacgtttctt aaccgaaagt
```

```
7381  gtaaaaaaaa aaaagttcaa atcatgaatt aaaaatctgt gctcttgatt acaaagcctc
7441  gtcgcctatt aataagccac tgcatatttt agccgattca ctgctggagt ataacattgg
7501  tactcggacc cattatctat cgacatgggg ggcactgtta catttaatgc tgtaacaggt
7561  gatcattgca ttctccgtta agctagccat acggttttag ctagtgtcgt tgtgcgttta
7621  gtactatttt gatacattga attattgcct caaaaagata agttcatctt ttattggtag
7681  acaaagcggt ccgaatcgaa gtcataataa gagatatcca agaggggagg tgttgtggta
7741  agttaagatg tgtggtatgt gtggcagaaa aaaaatgtgg caggtaggca aaacctgcgt
7801  catatgcaat ttctagaatg attcatttta aaacgcaaac gttataccac ccccttggtg
7861  gtaagtagga tgtgtggcca aaaacaatgt ggtaggttgg gaaaaaaact gtcatagtcg
7921  atttcttaaa ttacgtacat ctcctaccaa gcaagtcacg tcaattatta atgaatgaaa
7981  ttataaagat tttcaaaaca tttgttatat tttcaaattt agtcatcaga gaatttttt
8041  atagagtatt cacgtgtatt aaaaatataa taaattttta caattaaatt tatttttat
8101  ttcactatag atttttcaat aactattaac caatagtatt taatcaaatc aaatattttg
8161  aattaatctt tcttaaaaaa tatataacat atcaacatta attactaaaa ataccttaca
8221  attttagaaa atctattatt ttggaacaaa aaataactct agaaaatcat atttttgaga
8281  gcggagtgag tatatagatt tatgaattat gacacttgta aacttcatct ctcaaaacat
8341  attttaattt tttgtgattc attaaaaata tacagtagaa cctctgtaaa ttaataatgt
8401  cggaactttg aaatttttact attttataga gatattatta atttacaaaa aaattcttat
8461  ttaaatttta ttttttggat atatttattc taagataaga aaaatatttg atttaatata
8521  tagacattaa ttgttatttt tttttgaaat ttgacattta tattaattta tattaatttt
8581  attatattat ttggtgtata taatatatag gaaaattagg ggccatgtac atagcagaaa
8641  aaatattgag tctataccca taacaagttg atgactatag tattctctat atattacgta
8701  ttaatgacaa aactaacatg cgcgtgtatt ttattagata aataattttt taatattata
8761  ttcatttttg taagaattac aaatgaaact gaaccgtatg cacatcatat atactcatct
8821  taataatgaa ataatgagaa aacattgcta agaagaagta gaaacgacga cggctccaag
8881  aatggtggta aatggagctc gagctcggcg gtggttttgt tcgtgtaact cgaatgcagc
8941  aatgatcgat acagaaggtt gattatcaaa ctgttcgtct actttaattg ccagagaaat
9001  gaaatttttt gaagaagata aacagttact atacttatca aaaaaaaaat aaacagttac
9061  cattacaata tactattcta ttttagtagg aatcatagaa tacaacatat tttttttcaa
9121  agtgtttttg ttaggttttt atattagaga aacattcaag tgattaatga tctggtgatt
9181  caattcatat ttattcttac tccatcacat ttgaaaaatt atcagtttaa attagaatgc
9241  aatattttgt tatgttctat tctatatact aaaggggttt ttgccaaaac taacccacaa
9301  cttgatttta attccaaacc tatacccaaa cttgaatcaa atgcaaaact aacctaa
```

**Figure 7-8. Combined incomplete *Nco*I/*Sph*I and *Nco*I/*Sal*I BAC 6 fragments carrying *UGT84A9c*. Features given: IS-element (1-431 bp) in red, *UGT84A9c* (1118-2611 bp) in orange**

```
   1  catgggttct agcaactaac ttacctgttg aaattcgaac acccaaacaa cttgttaata
  61  tctattcgaa gcgaatgcag attgaagaaa ccttccgaga cttgaaaagt cctgcctacg
 121  gactaggcct acgccatagc cgaacgagca gctcagagcg ttttgatatc atgctgctaa
 181  tcgccctgat gcttcaacta acatgttggc ttgcgggcgt tcatgctcag aaacaaggtt
 241  gggacaagca cttccaggct aacacagtca gaaatcgaaa cgtactctca acagttcgct
 301  taggcatgga agttttgcgg cattctggct acacaataac aagggaagac ttactcgtgg
 361  ctgcaaccct actagctcaa aatttattca cacatggtta cgctttgggg aaattatgag
 421  gggatctctc aggttcaaac ctccatatat atagaaaatt ttcgaatctg gtagttgaat
 481  tttgctagga atttacgacg aaaaattagg ttggtggcaa aaaaaacgtg ttaagttggt
 541  ggatttctcg ttctttcctc gtaagttatt tcctcgtaaa aatcatgcta aaattacgac
 601  gaatttgcga cgaaacacat ttttctcgga aaaaccacgt caacttacga cgattctacg
 661  aggaaaagct ttactcggta ttttacaaag ccattacgag gaaactttat ttcctcgcaa
 721  tttcatcgtt aagtcatcgt aatttcacga ggaatagttt tcctcgttaa atttccttgc
 781  taaaactgtg ttttcttgta gtgaaacaaa cgagtgatac gttatataaa tacctcaccc
 841  cactgaacca agtcaccttt acgagcgtca tacctttttg gtcaatttaa gggcttttta
 901  tacaattaat aaaattgaat aatacggtgt ccaaaagcag cggtagatat tgagaaatat
 961  tgagatatcc gtaattatta atctttttt tttttaattt gattcacctt ctttgtcttg
1021  tatttaagcg catctcactc ttgtcttcca acacgtaaca tagtagctta acaaagaaaa
1081  aaaatcccac ataaaaagag ataccatttt ttaaaggatg gaactagaat cttctccttt
1141  acatcctcat gtaatgcttg tatcgttccc gggacaaggc cacgttaatc cacttcttcg
1201  tctcggtaag ctcttagctt caaagggttt actcgttact ttcgtcacca cagaatcatg
1261  gggcaaaaag atgcgaacct ccaacaaaat ccaagaccgc atcctcaagc cgattggtaa
1321  aggttatctc cggttcgatt tcttcgacga cggtctcccc gaagacgacg acgtaagaag
1381  acacgatttc accatctacc gaccacacct cgagctggtc ggacaacgag agatcaaaaa
1441  cctcgtaaaa cgttacgagg aaatgacgaa acagcccgtg acgtgtctca tcaacaaccc
1501  tttcgtctca tgggtctgtg acgtagcgga agattttcaa atcccttgtg ctgttctctg
1561  ggtccagtct tgcgcttgct tagcgtcgta ttactattac caccacaagc ttgttaactt
1621  cccgaccaaa acagaccccg agatcgatgt ccaaatccct ggcatgcctc tcttaaaaca
1681  cgacgagatc ccttctttca tacacccatt aactccttat tccgctttgc gagaagtgat
1741  cattgatcag atcaaacgtc ttcacaagcc tttcgctgtt ctcgtcgaca gtttctactc
1801  cttggagaaa ggcatcatcg accacatgtc aagtctctct ctccccggct ctatcaaacc
1861  gctcggcccg ctttataaaa tggccaaaac gttaatatgt gatgacatta aaggagatat
1921  gtccgagacc acggaccatt gtatggaatg gctagactca cagcctattt cctccgttgt
1981  ttatatctca ttcggaaccg tcgcttatat aaaacaagag cagatcaatg agattgcgtt
2041  cggtgtgata aacgctggcg tttcgttctt gtgggtgatt agacaacaag agttaggtat
2101  aaacaaagag cgacatgttt tacctgaaga agtgaagaag aaaggaaaga tcgtggagtg
2161  gtgtcagcaa gagaaagttt tagctcatcc ttctgtggtt tgtttcgtga ctcattgtgg
2221  ttggaactcc acgatggaag ctttgtctag tggtgttcca acggtttgtt tgcctcagtg
2281  gggagatcaa gtcacggacg cggtttacat gatagacgtc acaaagacgg gagtgagact
2341  cggccgtgga gagacagagg agagggtggt gccaagggag gaagtggccg agaggctgat
2401  agaagttgct aaaggagaga aggcgacgga gctgaaaaag aatgctttga aatggaagga
2461  ggaggcggaa gcggcggtgg ctcgcggtgg ttcgtcggat cggaatcttg aagagtttgt
2521  ggaaaagttg ggtgcgaaac ctgtggctaa acaaaacgga agtcttaatc aaaacggaac
2581  tgttcaagat cttttattgg aaaagtcata aatgtaattt gtttttctat ttgttggttg
2641  tttctgtcag tccttaatcg tgttcgtatg tatgctcctc gtcgttcttt gtattttgtt
2701  ttcttgtaag ggtctaaggg agacttatta tccttctgga agaatgttta gttcgcgtat
2761  tatgttatat tatgattaat tctcttgttt tgttgctaaa tcaggaaaag agtttgtata
2821  gaacgttctg taccgtattc ggcactcatt ccccttaac cgacaaaaaa gattgatcta
2881  gctagtaagt aaagtatttg accaaaaaaa aaaagctagt aagtaaagta aaagagttaa
2941  tggtatacag taattatatt attatatagt actattaata tacgtttatc aataatccga
3001  ttggagatta caagacattc ataaaacgaa aaaacaaata gaatcaaata tgggttttgc
3061  atgtatattt ctattaattt ccgcaattca caaactttga aactcagaat ctcagataac
3121  caggattatc agtatgtatc attaacgtct cgaatctttg gaagttaaca acacagattg
3181  acagaaacgt aatttttgga cttggatagt atctgcatgg attttttttt ttacagcaaa
3241  caccatgttt tattaattcg tgacgaagag gctagcttgt gaggctaacc acgtcggaat
3301  caaagagttt acatgggata gaatagaatt ttggagacga gctcctttag caaggaggtc
3361  ggctcttcca ttgcgagtac gagctataaa ggaaataggg aaagcattaa actcagaaac
3421  taaaaagctg aattcgatc
```

**Figure 7-9. Combined *Nco*I/*Sal*I fragments of BAC 14 carrying *UGT84A9d*. Features given: gypsy-like region (737-2233 bp) in light blue, *UGT84A9d* (9688-11181 bp) in orange, ORF homolog to *At3g21570* (16641-16225 C) in red, CACTA-like region (18741-19304 bp) in light blue**

```
   1  catggatttt tcaattgcaa aaactaaccc aaattataga cctacttgaa agtctacaaa
  61  cgtaaactga aaagaaagtc tattagtatg tagatggaat ataaagtcta catagaaaat
 121  ctataaaaag tggatttgtg tattattcgt aatgtatttt caagattttt tagtttgaca
 181  gtgtgtgtta gttaatccta atggtctcga actattttga aaatattttt ttttataatt
 241  atgaaattat caacattttg gtttcagtat ctagttagct aataagtgta gacgtctttg
 301  taaatctact ttgtaatatt atcaaacatg aaatttttct tcacttattt tgtacacatg
 361  gtaaacttgt gcatgatggt tgtacacatg atgaacttgt tgagatgaaa caagaagatt
 421  atgatctcga caagaaaacc gagatggtgg agctaatgta ttccttacca gacgtaattt
 481  taaagcaaat gggtcatgat actcctccta tacatgatcg acaagtacaa aacctgatag
 541  agttatgtaa gacgcatatt gtacgccttt gtttatcaag acagtgccaa gtagactaca
 601  aatttttggt ttgacaagat attagataat aagtgtagac atctttgtaa gtctacaatt
 661  gtagactgct gttgaagttt atgtcgtata tactattaaa agtacatttg tgtattagtc
 721  atcatatttt tcatgattta ataatattac agtgttagta gtcttcttct gcattttgtt
 781  gtttgaagta tacacttgat acttattgca aaattttgtg attagtttta ctcttcaaac
 841  ttcttaagaa ttttgtatag attttttttct tctcacatgt gtgttagtta ttgttttagc
 901  ttatgaattt tttaattgtt tggttgatta gagtatcttg tgaaaaaaat gatagctaac
 961  aaaaattaat aatatcatac aagtgaaaca cgactaaaaa taatacaatg ttatagatta
1021  tgcatcatat aattatttac aaatgaatat tttattatga taaattatta caaagccatt
1081  tattaaaaaa tatttttgaa cgtgattgat ttttcataat cttgatgcct caaatgaagt
1141  cttggaaaaa aatatctgca aaataagata aagttatgag aaaaacataa gataaaaaaa
1201  attaaataat attttagtag actttttatt aagtctactt aaaatttatg gtttgtagac
1261  tttagttgaa ctctactagc cctcactttt aagtagactt catttactac tattttgtct
1321  tataaattct attttctaaa aaaattagat ctaagacttt gtgagaaaaa tcttaaatct
1381  gaaaaatata tgcattaaaa atatgaaaat gagtttagat ctaatttttt ttagaaaata
1441  gaatttataa gacaaaatag tagtaaatta aagatataat tttgaatatg taatttttatt
1501  taaatttaaa tacaaaaaca catatttaaa atatatagat gtaattctta catttttggg
1561  tggagagata acaacaatgg aagaaaatac ctgcaaaata agataaaaat attaataaaa
1621  cataggaaac aaaattgaag tcatatttttt gtagacttcc aatgaagtct gcttaaaatt
1681  agtggcttag tatacttttt tttaaagttt actagcctta gcttataagc atacttctaa
1741  tgaagtatac attatctaat aattttcaga tctgaaaaaa tatgtacatt ttgaaatgtg
1801  aaaaatagtt ttaaatctta aaatacttaa aatatagaat ttgtaagata aaatagtagc
1861  aaagtaaatg cacatatttt gaatcaataa atttatttga atatgaacat ataaatacac
1921  atttaaaatt tatagattta aaacttacat ttttagagga gaagatgaaa accatgaatg
1981  atagtaccta taaaaacaag ataatattgt aagaaagaca taacataaaa atacaaattt
2041  aaaattttca gttctaaact tcatattaac ccttgtatat atatgattag tttaaagttt
2101  cacaaaaact attttttttac attttaaaaa ttaatttata agatcaactt acgacgtaga
2161  cttgtattgg catctacata tatgtaaact ttcattttta catataagaa tttcgaagga
2221  caaaattgta aaataacttt ctgtttttttt gtttggttag ttatatagta cttttattac
2281  tttttaggtt agtattgcat ttgattgaat atgtggtcta attttatggt tagaatcaat
2341  atttggattg gttttggcaa aaatcccata aaataatgag atgtattaaa aagtaggaag
2401  aatttgtgaa actaaccaag atctcaataa ccataaagaa tatagcactt atcattgaga
2461  aattaaataa ataattaaat cactgttttt ttttcaaatt tttacctttt ccacaacaaa
2521  cggaagattg ttataccatt taactattat tatgctttct ttatatagta taggtacata
2581  atatgataat agtataatat atatgatatt atatattgat agtttttaag tacattaaaa
2641  ttagactgag caatgaatac aaattatgtg taaacagtat ttataattaa tcagtattga
2701  tataaacgac tgaatgcatg gcaagttttg aaaatctgtt tttaaagaaa atagtataaa
2761  actagtaaaa aaacagtaca tacttaaaat ataatcaaga aggagatata tatgatttat
2821  gagcaaatga taaacaaata gaataaaatg acagtatact gtatatagta cagttgaaca
2881  acttttaaat tatttattta tatttttttg gaaaagtgta taataagaaa aaaattcaaa
2941  tcatattaca tgtaaaaaaa gatgaaactt atctaattaa attttacatt tgattaatta
3001  tatagtttga agaattttaa gcaagtaaga ttctaggagg aagaaagaca caaaagaata
3061  gaaaatctat tatataatag gagagtttct ctcctcctga tgtgacacgt cagcggtctg
3121  tgggtcctac tttttaaaag tgtgaaaaaa tgtcaaatat gtggctcgaa cccgggttat
3181  tgagatataa gcacgcacaa ctataccatt gcactaaagg atactttgta cattgatggt
3241  cgaaactaat ataacattgt ggaccccacc tcttttgttt tttaattgat ggataacgaa
3301  tggacttcga taaaaagtgg gctttgggct tattgatttt ctgtttatta gactttgtat
3361  ctgcccaaag tggtgatacg gacaaagcta agaagattag ggaagccgcc atcttcacca
3421  tctccttcgt cgcttgcgat tccccttccg gaaatcagct attatggtca atctttaagg
3481  ctctacgcac gttttgcgcc taccaaacgc ttagtttctc ctctaacgcc ttcagagctc
```

```
3541  tcatatatat agaatccctc gaggtaaaag ctccatcttg tctcgaactt ttctttctca
3601  gttttccgat cgctttactt tctcacatcc gtcacgaaaa tgacttatcc agctgctcca
3661  accgcttcag ccgccgttcc ctactccacc ttcaactctc tccgccttag aaggtccact
3721  cagtccattg ttggtcggct catccgattc taggattcca ggaacattta caagaatgga
3781  gagtttatgg gcatcaccat tctcctcctt tatgaactgg tgatgatggt ttctaactta
3841  ttttttatca ttttaactct gctttgctta ctctttgttc gtatatgaaa cgtgtcttta
3901  tctttttgtt ttgtttttgt ctatgggttt ttttatgaac aactaatgat ttttctctgt
3961  ttttcttaca ggatttttgtg atctacgctt tcatctctgc aatctccaaa caaaacagga
4021  aaactgaatg aattatgttt ttagcaaccg aatctaaaag aaataaaaga cattttactt
4081  aacaaattaa aaacttttct tttgcagatt ccgacgatga cgaagcaccc gaggaccaac
4141  gacaatggaa tgatgcggac gaagcctaac tgatgtaagt aaactcctct tcaataacgc
4201  atcacactat tatcgctccc caatctcaca aacttaagct ctaatcatgc cagggatccc
4261  acaggtgatg ttgtcccgac gacggagaaa aaaaaagaaa gataacctaa cgtttttttt
4321  tttttcagtt ttgaagttca aatcatgttt tagacataga gtccaacata atttaagccc
4381  aaacatttta aacccacaat tacttttttca taatgtaaaa acgtcaaccg tcgtttatgt
4441  tttgttaaac tattttaatc tttatgtttt gttaaactat ttcacaccga gggcagatct
4501  ccttagaagg ggagccactc gccacgatgc ttccagtttc accattcttg aaactcttat
4561  gaaccacaaa gccaatataa gagctttgtt tcaatccaac ggctggattt tgtcacagac
4621  caccgctaag cctgaagagg gaagagaagt ggagtgctgc gttttggaag atggttactt
4681  ggagtttaca agatgaagat ggaacccgtg acgagaaaga cgaggagaag tgttacttgg
4741  aatatcctaa cgttgaggat gtagacaatg tgtgatctta gcgatgaaga ttagtgtagc
4801  tttaatatat gctactacgg tccttgagct ccctaacaag tactagcaaa caatgtgtat
4861  tttgaaaaat aaaataattt tgtaatgcaa gtagtttgaa actttgagta tttattcaaa
4921  caacggatat atacagagtt gataacttga tatggacccc caaataagtt ggtgttgcgc
4981  tgtcacaata tctgtctttt taatttcata tacatgacgt aagtttcttc atcttcttct
5041  ctttcccttt ccgttgttcc acgctctctg tttcatgttt aggccttttt tacattctct
5101  atattttacc ttaattatac tttttgaatt ttccttctgt tttgggagat gtctctaact
5161  tgagttaatc gtcaaaagct aggatcaatt ttgttgtctt acttccgata aattttattc
5221  tttaatttgc acaaatctta aacaacacca acatcccgta aatcatactg tttcaccgat
5281  atccaaagta accagctccg cacttgcgcg gggtctcagc acctagttag gataaggaca
5341  atgtaatata tattgtagaa atgatggagc taaaattttt tggatagtta gttaactata
5401  atttttgttt gatatacaat gcaatttcca taaaaagtaa ttgtcaaagg agtgcatccg
5461  tttggaaaaa gagaaaattt gacaaaagag tgggttgatt tcactaattt tattggatgc
5521  acttgttata attattagta atttagaaag gcttttgata tttaattgga gtcatttctt
5581  cgctctccct tctctcttgc aaacaactca gtgaccggtg tcactaccgg taccgcgaga
5641  aagaccattc ttttgcttgt tgcaaagcct tggtcttcat agtgtgcttg ccttgtcttt
5701  ggcttggctg gggctccggt ttcttgttcc aaatatttgg ggtccgatct ggggcgagat
5761  accccaaacc taccttagtt ctctcctttc aactcacaat ggttttgaga tctctagtta
5821  accggcgttt gtggttctgg atttcctggt cttatatctc tttggcttgc atgtctgctc
5881  tgaggacaat gggctgagat cgaggtttga ggctggtgga ggtggaggtt tgtgaggtga
5941  ggcttcgcct attccgtacc ggcaaagcct gttaggtctg ttggctactt tttctccttg
6001  tgttttgcct cttattttcg cgtttcgtca agccggattt ccattcttct ctggttcgtg
6061  gggcttatcg ttttactctc catcggaagt ttctcctcca ttcatttggt ggtcttttgc
6121  gttccgcttt tggcgatgtt ctgcagtgac gtggtgaccc ccttggtggt tgcacacgcg
6181  cgcgcacaca cacacattga ccttgagctt tcattcctat aaacactttg actcatgtac
6241  acctccattt tcattagttt tagtctcatt tatactcgaa aacatttgta agttgtaaac
6301  cttttgatat atgatataga aaaaaattca taaattacct caactaaatt tcgctaagct
6361  ttttaatatc caaacttttt cattacacaa ctaattacct aaaccaattc gttttcacat
6421  tttaactcat aaacttttaa gactatgcct attttaatat tcaagctatg tttattgaat
6481  ctaaatatta ataagtttca aatatttttt ttcttaaata tctaaaattc acaaacaatt
6541  cttaaaaaat ttctagtttt attttaagtt ttagaaataa aagtaactaa attattcatg
6601  atcttcaatt ttgtgacaaa atttttaatt tttaaatatc aaatttagtt tgtgtttaga
6661  aatagaaaaa aaattccttt tctccttcaa aattattatt attatctaaa tttaattttt
6721  tttccataac ttagttactc tctttcctaa aaataaaata aaaggataaa ttttaagatt
6781  tttcttagat tttgagattt aagaaaaatt tccgaaactt attaatattt ttgataaaat
6841  aaacacattt tgagtattaa tatggcacct tttataaaag tttgtgtatt aaaatgagta
6901  aaataattaa tttaggtact tagttgggta atgaaaaaat atgggtggtt ttttttttt
6961  ttttttttt ttttgacaga atatgggtgt taaaaaggtt aacaaaattt agttgaatta
7021  attttatgga attttaccta taatatataa tctagtttcg tcaacttgag tctgattaca
7081  tcgttattgt tttaaagata ttgttgtcaa caagtttttt cttctcttat tttacaaaca
7141  atcactaccc aaattcgtag tgtgagtttt aggatacaca ttttgacgtc atctgaaggg
7201  gaaggggagg gggggggga ggataaacaa aaacattctt gctaagaaca ctgtttaagg
7261  ttcataagca cgatcatgga tccctcgctc acgactcccg tttccaccaa accacagatc
```

```
 7321  aattatatca atctccctag atctgacatg agaaacaaaa ccaacatatt gcccacagag
 7381  actggcaagc cgatttaact tgtacagcaa gccttcctca aatggcatca acacccttaa
 7441  acacctcggg acatcaattt tcgattgtca aaaatcatgg aaatgatcct agtgctcgga
 7501  cctttacacc gagaccgata tctcccttgc tgatagctca tacttgggaa gaaatggacg
 7561  acctgtgaca tattctcacc attcttatca taagtcccat aggcaagaat tcgcttagag
 7621  aactgtagtt agtcgcctag agcaaaaagg gatctcgcag cgcatcaagt agaagtacac
 7681  gagtcgaagt gatcaagtgt ttgaggaacc gatctgaagt tttagatcat atccccacat
 7741  atcaagcaca atcataccac cgaggataaa caaaaacatt cttgctaaga acactgttta
 7801  aggttcataa gcacgatcat ggatccctcg ctcacgactc ccgtttccac caaaccacag
 7861  atcaattata tcaatctccc tagatctgac atgagaaaca aaaccaacat attgcccaca
 7921  gagactggca agccgattta acttgtacag caagccttcc tcaaatggca tcaacaccct
 7981  taaacacctc gggacatcaa ttttcgattg tcaaaaatca tggaaatgat cctagtgctc
 8041  ggacctttac accgagaccg atatctccct tgctgatagc tcatacttgg gaagaaatgg
 8101  acgacctgtg acatattctc accattctta tcataagtcc cataggcaag aattctgtga
 8161  ttctaccgac gaaaaagagc cataattaga ggtatttaat ttcgctacgc cgtttcttat
 8221  attctccttt gtactttatt gtttgataat aatgtccatt aacggaataa ttttttttta
 8281  ttaacttgaa ccattaacgg aatattcctg tccaaaaata taccacattc gtcaactagc
 8341  tgatattatt aaggttttcg aagacttgaa tttcaacacg atactacatt acatcttggt
 8401  gaatggtggg tggctcactg ggtttgctat gcattattga caacgactgt ttctcaaata
 8461  tgaattaatc ggaagcatat aacttaatac tctagatatt aagagaaatt ttatgatgcc
 8521  aaaaatattt tatgtgtatt aatcgtcacc tttggaatta gggtctgatt ggttgcaact
 8581  gtgattttag ttttttgctg tagaatttat gttgtgtatt tttttactgt agctgtaaaa
 8641  tttttgttgt gattttgtat gttataaaac ttacagattt attaataagt ttctaaagca
 8701  ctgttatttt gttgtggaaa tattagtgtc tacagccaaa aaaaaagaaa aaaaagaaac
 8761  aaagcttgac tgttaaatat tgtgactgta gaaatatttt ttgctgtcca cagcaactaa
 8821  agcactaacc aatcaacttc ttagttttat ggattcagtt ggtacagagt atttaattta
 8881  cgcctgctac aacagaggta ataattattt catccatttc ataatactta atgttttgtc
 8941  ttagtgcaca aaaattaaaa aaagttagat ttaaaaaaaa cattttaga aatatagttt
 9001  tctaatcatt tgatcaatta taaaaaaaaa ctgtaaaatc taattagtta tatagtttac
 9061  aataaaatta aagttaattt taaaacctca aaatttcacg taaattaaaa taaaatattt
 9121  cttctaaaac atcatctata ttgaaaagga tggattattt tccagtttaa tcggttaaat
 9181  gatttatttc ctcaatttta tattcccaat cgtaataccg tgagaaaggt tttaaagaca
 9241  aaccatgaac gttttggacc taaaatgtca tatatggaac aaggtaatgt ttacttgtac
 9301  ttgcgactat ctcttcttga ttttggctta tataaaaaat atcttctttg ggttttcaca
 9361  atgtaaataa cctagaaaca aacgagtgat acgttatata aatacctcac cccactgaac
 9421  caagtcacct ttatgagcgt catacctttt tggtcaattt aaaggctttt tatacaatta
 9481  ataaaattga ataatacggt gtccaaaagc agcggtagat attgagaaat attgagatat
 9541  ccgtaattat taatcttttt ttttaacttg attcaccttt ctttgtcttg tatttaagcg
 9601  catctcactc ttgtcttcca acatgtaaca tagtagctta acaaagaaaa aaaatcccac
 9661  ataaaaagag ataccatttt ttaaaggatg gaactagaat cttctccttt acatcctcat
 9721  gtaatgcttg tatcgttccc gggacaaggc cacgttaatc ctcttcttcg tctcggtaag
 9781  ctcttagctt caaagggttt actcgttact ttcgtcacca cagaatcatg gggcaaaaag
 9841  atgcgaacct ccaacaaaat ccaagaccgc atcctcaaac cgattggtaa aggttatctc
 9901  cggttcgatt tcttcgacga cggtctcccc gaagacgacg acgtaagaag acacgatttc
 9961  accatctacc gaccacacct tgagctggtc ggacaacgag agatcaaaaa cctcgtaaaa
10021  cgttacgagg aagtgacgaa acagcccgtg acgtgtttca tcaacaaccc tttcgtctca
10081  tgggtctgtg acgtagcgga agatcttcaa atcccttgtg ctgttctctg ggtccagtct
10141  tgcgcttgct tagcctcgta ttactattac caccacaagc ttgttaactt cccgaccaaa
10201  acagacccca agatcgatgt ccaaatccct ggcatgcctc tcttaaaaca cgacgagatc
10261  ccttctttca ttcatccttt aactccttat tccgctttgc gtgaagtgat cattgatcag
10321  atcaaacgtc ttcacaagcc tttcgctgtt ctcgtcgaca gtttctactc cttggagaaa
10381  gacatcatcg accacatgtc aagtctctct ctccccggct ctatcaagcc gctcggcccg
10441  ctttataaaa tggccaaaac gttaatatgt gatgacatta aaggagatat gtccgagacc
10501  acggaccatt gtatggaatg gctagactca cagcctgttt cctccgttgt ttacatctca
10561  ttcggaaccg tcgcttatat aaaacaagag cagatcaatg agattgcgtt cggtgtgata
10621  aacgctggcg tttcgttctt gtgggtgatt agacaacaag agttaggtat aaacaaagag
10681  cgacatgttt tgcctgaaga agtcaagaag aaaggaaaga tcgtggagtg gtgtcagcaa
10741  gagaaagttt tagctcatcc ttctgtggtt tgtttcgtga ctcattgtgg atggaactcc
10801  acgatggaag ctttgtctag tggtgtccca acggtttgtt tgcctcagtg gggagatcaa
10861  gtcacggacg cggtttacat gatagacgtt acaaagacgg gagtgagact cggccgtgga
10921  gagacagagg agagggtggt gccaagggac gaagtggccg agaggctgag agaagttgct
10981  aaaggagaga aggcgacgga gctgaaaaag aatgctttga aatggaagga ggaggcggaa
11041  gcggcggtgg ctcgcggtgg ctcgtcggat cggaatctcg aagaatttgt ggaaaagttg
```

```
11101  ggtgcgaaac ctgtggctaa acaaaacgga agtcttaatc aaaacggaac tgttcaagaa
11161  cttttattgg aaaagtcata aatgtaattt gtttttttat ttgttggttg tttctgtcag
11221  tccttaattg tgttcgtatg tatgctcctc gtcgttcttt gtattttgtt ttcttgtaag
11281  ggtctaaggg agacttatta tccttctgga acaatgttta gttcgcgtat tatgttatat
11341  tatgattaat tctcttgttt tgttgctaaa tcaggaaaag agtttgtata gaacgttctg
11401  taccatgcaa aaccaaaatt gctaaaacct agtaatattc ggcactcatt ccccttagc
11461  cgacataaaa gattgatcta gctagtaagt aaagtatttg accaaaaaaa aaagaaggta
11521  gtaagtaaag taaaagagtt aatggtatac agtaattata tttttatata gtactattaa
11581  tatacgttta tcaataatcc gattggagat tacaagacat tcttaaaacg aaaaaacaaa
11641  tagaatcaaa tatgggtttt gcatgtatat ttctagtaat ttccgcaatt cacaaacttt
11701  gaaactcaga atctcagata accagggatt atcaatatgt atcattaacg tctcgaatct
11761  ttggaagttg acaacacaga ttgacagaac tcccattgga agtgctctaa gaaccatctc
11821  ttaacagaaa aatattaaaa aaatgtcaaa tcatgagtta attaagaact ccggctaaga
11881  aatcggagtt aatcattctc taacggttgg tattgcgcac tgaggttcta ttctatataa
11941  aactaataat ttaaataaat taaaatactt gataaaaagt aaattcattt gtatgctgat
12001  atgtgtatat gatttttttt ttaaatattt aatattaaac tcttttactc tatttctctc
12061  ccgtcattct cttattttgt tttctcttta attattgata tactcaaaag tagcattttt
12121  taacgccgat ttattagaaa atattacaat tatgagaaat attaaataga tgattcgaca
12181  accgacaata ctacctgcct tatgaggatc tatgtctaat tgcatcacga gagccgtctt
12241  ataaagatct acgcctgacc gggtttactt gcaccatgtt gaagattctt tgtaagcctt
12301  tttttagtac cctacataat tgtttaataa atcattttct ttgagaattg aaacttagat
12361  ctgctgatgt aaaaacatta atgaattttt agtcaaacca ttggactaat gaggtttcca
12421  cacccaaaaa atggcatcac tggaagtgtt ttacagagat attacaaact tgcactgatt
12481  aaaataaagt gcacacagct ttttgactca ctcagtatat gttttacatt atactcagta
12541  tatttttggt attaaagacc caacatctat cgacctgttt actactgtta tatctaaggc
12601  taacagttga aacattgcac tgtcccttcg ctacaagcca aaagcgaaaa gaaatgtgtc
12661  cccgtcaagt taaccatatg gtttagctag tgtcgttgtg catttagtac cgtaggcgtt
12721  ctgatatttt gaattattgc ctcaaaaaga cttaattaat ctattggtag acaaaacgtt
12781  ccaaattgaa gtcgtagtga gagtgataga gttacccaag aggggatgtg gtggtaagtt
12841  aagtgtgtgg cagaaaaaaa aatgtggtag gtatccaaaa cttatgacat attcgatttc
12901  tagaatacta gcaccctctt ggtggtaagt aagatgtgtg gcgaaaaaga tgtgggggta
12961  ggtcggcaaa aaatgacttc aaaatcgata tattaaagta cgtacgtgcc aacaacgtca
13021  cgttgattat tggaaaatga atctgtattt cagagcattt atataatgtg gttaatggtt
13081  tatatatatg acatattttc gaaacagaat tatgacatat tcttcgaaaa gaattatatg
13141  acatatttta aaatattatt ataggttgat gattctgaaa tgtttatttc tcaaaatttc
13201  aaatgtctaa acaaaataaa gattttcttg aaccatatca gaaatttaga atgatcccca
13261  cttctcgttt cgaagtgtgt ttgaagattg taaaaaaaaa aatagagctt tgaataatat
13321  tcatgctcgt ttttaaaaaa aaatctaggc atgggcattc ggggtcccaa tcggatttcg
13381  gttttatcca ttcgggtttt gatttttcga gtttatcaaa atcaacccca ttcgggttat
13441  ataaaagttc ggttcgggac cgattcgggt tatatcgggt tcgggtcggg gttagtaaat
13501  cttcaaagaa ccggtataat ccattgtact ttcgggttcg ggtcccaatc ggttcttcgg
13561  tttaaaaata catgatttgt acctattttg taactaaaac ataaatgaaa tcggttctgc
13621  ggatttaaaa tacataattt gtacatattt taatagccaa aacataagta aaatcgattc
13681  aaaaataaga aaaaacatca aacgtgatca ttcaaaatca agcgaaaggt aaacatagtt
13741  agtgataaaa agaaaatcag ataaatgaaa tcataaaaca aaaactaagt tctcatgaaa
13801  tgagaaacac tattcaatga aaacaaaacc aaaataaaaa aaactttagg attcaaccac
13861  tacattccac catcaacatt catataatag ataattattt tagaagttca ataatatctt
13921  aaagtatttt ggatacatat taagaattaa gatcatattt ggtagaaggt ctttttgtga
13981  ttataaatgt ttcgggttcc atcggatatc catttaggcc cgagttcggt tcggataata
14041  cccataaccc aatataccaa aaaacaggat ccattcggta tttatgtcgg gttcggatcg
14101  attcggattc atttttatcg gatcggattc ggtttaaatt ttcggattcg gtttatttgc
14161  ccagccctaa aaaaatctat taataatcat gtctcttcgg tatagtggtt gtaggttcta
14221  tggtatttgc cttgttggtg gaaggttcga actcagtttt cttttgttaa tttggttttt
14281  ccccttttct ggggataaca aatgagatgt ggtcgcaatc ttagttgaga tacgatagtt
14341  ttgactaaaa agatactatc agtttaaaaa atctcttagt tatttcattt tttacatttt
14401  agtttattta aaaaaaaaaa caactttgtt tcttttatat ataaaatact tttgttttca
14461  gcattttgtg aaaaaatagt atcttttgtt attcacttat ttgattgtta tgtccgctgg
14521  aaatattgtt tcttccaatc ttccatgtaa aattattttt cttctcttta ttttttgtaa
14581  attttacaat ctattttctg atttcagaaa tctgaagaaa catatatcat aaattaactt
14641  atgttttttc acaaatttcc agaaaatcaa caaaattatg caagttaaaa taataggatt
14701  cgatttaatt ataaatttcc ttaaattagc atttctctgg tagaaaactg ataagaatta
14761  tccatgtttc tattgcttat gaaccattaa atggaatggt cgattatatc aattgagtcg
14821  gctaaaataa tttattgaaa gttaaaaata attcataaaa aatgtaatac atatattatt
```

```
14881  tttaataata ctttatgttg ttctaatttt atatctaaat aaatttatat ttaaggtgca
14941  gcattgtttc gatgcgcttt aggcgtcttt ccagtccgga cgggcatatc acaactgcct
15001  ctcaatttca accaaactta gctaaacacc atgttcgaaa ttacgctagg cgtaacgcgt
15061  gcaggtggtc cgggcctagc gagttgaaag agaatcggag aataatcagg gattactcgg
15121  aaactagttt tatatgattt tattatatat cattattatc tgtgtattaa aacctaaaat
15181  gcgttgtagt ctggtggttt tccgtgttcc cacatgcagt tctagtgacg agtttgaata
15241  cctgtgtatg cgtttgggtt tttatttttt tgtgtatttt aatgaagtcg aaatgactta
15301  aatatcctca tcttctttcc cagttttatc gatctgaaac cctaccttcc gtttctatca
15361  tttttcttca cgattttggg ctgttttaac tcggtttcct tgtattttta ccttacaata
15421  acagttcttc aaccttattt ataatttgaa acttataaat ttaaaaatcc gagtttttga
15481  gcgatttggt acaagacgcc acggtgaaga accgattcac atttactcac cgcgtaaacg
15541  ttcaacgcgg acgcgtttta gaacatgggc taaacataat cataacattt tgatatacat
15601  cttttaagtt tagtactctt tccattccta aaaaatccat attctagaaa atacttttgt
15661  ttcaaaaaga tacatatttt atattttcag tgtaattttt gtcaactaat aatgaaaaat
15721  tgtgaagttt aagaacgtta attgcagttc ttaaaatttt attggtttaa aaatattaaa
15781  aatctaaaat tacaaaaaac tatgtattta taactaagtt ttaatatgtt ttattaaaaa
15841  ttgtgaaaat tctaaaacat ggatctttta tgaaaggagt aattattatt agctgacaaa
15901  aaaaaagcta atagttttta ttaactttta atataatttg gttttgccca tattttgtaa
15961  ctaattaata ttccgatctc gaataatact tttcttttaa atccatagaa ccaaataaaa
16021  ctgaaaagtc agctaatatt taacatttgt ttctaaatct gaccctgtac ttttttatac
16081  gatcactcag acccacccca ttctataaca aattattcaa aggaaggaca gattagtaaa
16141  acaataacgg agacgattcc cagatttaaa tacatatgag aaccgttata tagctgacga
16201  tcccgttaag tttcccgcgt taactcaaaa ttcaagccgg agaagtcgcc ggagatcctg
16261  agacgccgaa tcatcaacgg aaacaacaac gacctggctt ttcacgtttc tcctcgagaa
16321  cgacggcaaa ggtagcgagc tcggcgccgg agaaacaagg aaaatggaag acccggcgta
16381  gatatcgtcg ggagagactt ccggtgcatt tgatctccga gccttgatgt tcgaaaacga
16441  atcaaaagac tctcctcttc tcctaacgac gtccttcgcg gtcgttctga agtcaacgga
16501  tccgaaccgt ctcctttggt caggtcggag aggcggtctc cggtgagaat gatttctccg
16561  gggatggtga aaaatggtcg ccggagaatc tctcatccgg ttaaccaaac agtcctgagg
16621  cctgatgact tcggttccca tcgttataag agtagaaagg gcgaatcttg aattaacgtt
16681  cccgagaggg agaagagatg gtggtggtgg tggaaaggaa agagaggaag aagaaaaggg
16741  taacacggtc cttttcttcg gttcctgctc aagtttatgt ttgtaaatga ttagtgagtt
16801  tatgttgttg tctgaaatgg aaaattaggt tataaataga gaagatgaag aggtttgata
16861  aaaatggtgg gaggagaggg agaaaactat ttggcatctc tgtctgacca tcgtttcctc
16921  gtaacgacga atctggaacc tccaagcaac caacttattt ttttgtttcg tctaattata
16981  ttattctttt agttttaaat ttagttagtt tgtctgtcag tttttcgttt tacaaaatgt
17041  ggaggtgggg gattggctcc gtttgtgttt aacagtaaca ttataaaata ttctaaaata
17101  gcataaagtt agtttaatac attttcgtta atcaaatttt atgttaaaag ctgtattagt
17161  tctacaaagg acggttttga aaagataagt aaaaatgatt ttttttttt taagtttcat
17221  gatattagga aatcaaactc agtatttcct ttgaagtagg gttcgttttg gtgtatataa
17281  aggcaacatt atagaatcag agagagaaaa aacaattatc attgataatc tacttgatct
17341  aatgctcttt cttgttggaa aattcttcct gaaagctgat gggttaatat gaattaaggt
17401  acataaattt atctgttata atatatcagg tttgttgttg ctaggcattt tatctgttaa
17461  attagattcg attcgttatt cgtttcgatt cgatccaaaa attctggata tccgtacttt
17521  cgaatcaaaa ccaatactaa aaatcaatat ccgttaaaat cgaagcaaat cacaaatatt
17581  aaaattttgg gaaacgaata tccgatccga tctggcaata taaaaataca tgtatatctt
17641  gattatattt aaagttttta atgtataaaa ttatataatt attattctaa catatgattt
17701  gataaatttc atttacatta ttacttatat aaaagtatta cataaataga agataacaca
17761  tttatgacaa ttataatttt tttcttgagt tttgtgatat tataattgtt aactaagttc
17821  aaaaaattta caaaatatgt agattcacta ctttttaat ttctatcata tatcatgttt
17881  tataaaacaa atttgtatca aattttaaga ttatttgtat taatcaaaac atatgagata
17941  tccttaagta ttcgtaagta tccacaaata tctatttatt ttctggatat ctgttttttc
18001  gaatatccgt atttttccga agcaaagcaa atcgaaaaat tagatatccg tgacatacga
18061  agcaaatcac aaataccttc aaaaacctgg atatccgatc cgcgcccagg cctagtcata
18121  actttgggag aaaaaagcca atgaaatatt aatatccccc atatattaaa ggagaagcat
18181  ttttaaaact taccttaaaa gttatcggta aggattgtga cgtggtgacg tggcttcacg
18241  agaagctttt atttagtaaa acgcttaggt gtcacgcaga gaacgtttct caccaaaaca
18301  gtgtatttaa tgcgttcaca ctaacatttc tttaaaagaa cttcatacca tctacacttt
18361  tctcgacgta cactttacct tcagcttcaa attcgattag ttgatttatc actacaagaa
18421  aacgtaagtt taacgaggat aattaacgag gaaaaataat cctcgtaaaa gtacgttgat
18481  tttgcgagga aagtacgtgg aaaaagagaa gcatcgttat ttcgtcgtaa ggtaacgaca
18541  aaacaatttc gtcataaaga cgatgtaaat tgacgtggct tttacgagga aacactattt
18601  cctcgtaaat accacgtaaa cttcgcgagt tctttacgac gaaatacttt acgtgtactt
```

```
18661  aacgaggaaa ttttgaatcc accaacttgg taggtagctc acgtttgttt tggccatcca
18721  attaattttc gtcgtaattt cacacgaggc tatacttttc acacatatga ctatggtaga
18781  cagcgggcga ccagtaacta tggaatatgt gtgaaagggg aaatagattt ctacgggatc
18841  ttgacggaga ttattgaagt cgaatttcca gggatatgaa gctgaaatgc gtcctcttca
18901  aatgtgaatg gttcgatccc gtcgtcaaca gaggtgttcg gtctaacaaa ttcagtgtag
18961  ttgatgtcaa cggtggacga aggtacaaca aattcgagcc tttcatctta gcttcacaag
19021  cagaccaagt tagcttcctt ccataccctc gaatgagaga ttcaggtata aattggttag
19081  ctgtgatcaa agttacacct cgaggacgaa ttatcagtgg agaagaacca ccattgcaag
19141  aagaacagat aaatgaagtc gaggaacctg aacaagaaat tgatgacatc cttctcattg
19201  atccgcataa tcacgagtac gaagatctta ccaacgatgc cacagacgaa gctgttgaag
19261  acgagtttaa tgaaaatgat gatgtttcta gtgatgacga gaatgttgat gtatccgatt
19321  gatgtatttg attttatgta gtttgtttta tgaataagat aatgtgggag tttgatttat
19381  gaataaggta atgtgaaagt ttattttaat tatgaataag gacttgtggt ttggggtttg
19441  gaatatatta taaataaggt ttgaggtgaa agaatagatt gaggttcaag ggtagatggg
19501  tttggggttt ggaatatatg aagtagaaga taaggaagat ggggtttggg gtttcggatt
19561  ttagggattt aaacgtaata ctcgttaatt cctcgtaagc caaaatcgtc gtaaggttgt
19621  cgtaaggcca cgaggagttt acgacgaaat taaaaaaaat aaagaaagcg ggccacgcta
19681  attccacgca agccaaaatc gtcgtaaggt tgtcgtaagg ccacgaggag tttacgacga
19741  aattaaaaaa ataaagaaag cggggcacgc taattccacg taagactctc tgagtgaatg
19801  aaaataagga attgtggttt ggaatgaaaa taaagatagg gtttggaata tatgaagtag
19861  aagataagga atatggggtt tggggtttgg ggtttcggat tttagggatt taaacggaat
19921  actcgttaat tcctcgtatg ttaacgtcga acttacgacg aattcatggc tttttactct
19981  cagccacaaa aacaacagcc aaatattttc caaagccatt attttaaaag tcttaacagt
20041  aagtaacagc cctaaacaat caaagcctta aatttactaa aaccagtagt aactaacaac
20101  gcaatggctt gaaagcatcg atcaagtcaa catatatgtt gacgaggttc tgtacgtact
20161  aaggagagtt gaaatattca cagccacaaa agcaacatcc aaatattttc taaagccatt
20221  attttaaaag tcttaacagt aacagcccta aacaatcaaa accgagagtt gaaatattca
20281  cctgatgagt gatggagcat aagtagaggt gttgattcat catctggaga ataaatggat
20341  gcggtac
```

## 7.3 Characterization of *UGT84A9* loci

**Figure 7-10. *SGT* sequence type 1 from *B. oleracea***

```
   1  atggaactat catcttctcc tttacctcct catgttatgc ttgtatcctt cccaggacaa
  61  ggccacgtta atccacttct tcgtctcggc aagctcttag cttcgaaggg tttactcgtc
 121  acttttgtca ccacagaatc atggggcaaa aagatgcgaa ccgccaacaa gattcaagac
 181  cgagccctca aacctatcgg taaaggttat ctccggttcg atttcttcaa cgatgggctc
 241  cctgaagacg acgatgcaag cagaaccaac ttaaccatcc tccgaccaca actcgagctg
 301  gtcggacaac aagagatcaa aaacctcgtg aaacgttaca aggaagtgat gaaacagccc
 361  gtgacgtgtc tcatcaacaa ccctttcgtc tcttgggtct gtgacgtagc cgaagatctt
 421  caaatcccct gtgctgttct ctgggtccag tcttgtgctt gcctagcttc ttattattat
 481  tacaaccaca agcttgtcga cttcccgacc gaaacagatc ccaagatcga tgtccagatc
 541  ccgtgcatgc ctgtcttgaa acacgacgag atcccttctt tcatccatcc tttttcacct
 601  tattcgggtt taagagaagt gatcattgat cagatcaaac gtcttcacaa gcctttcgct
 661  gttctcatcg atactttcta ctccttggag aaagatatca tcgaccacat gacaaacctc
 721  tctcgcaccg gctttgtcag accgctcgga ccgctttaca aaatggccaa aacgttgatt
 781  tgtgatgaca tcaaaggaga tatgtctgag acgagggatg actgcatgga gtggttagac
 841  tcgcagcctg tttcgtccgt tgtttacatc tcatttggta ccgtggctta cgtgacacaa
 901  gaacagatca gcgagattgc gttaggcgtt ttaaacgctg acgtttcgtt cttgtgggtg
 961  ataagacaac aagaactagg tgtaaacaaa gagcgacatg ttctgcctga agaactcaaa
1021  gggaaaggta aagtcattga atggtgttca caagagaaag tcttagctca tccttctgtg
1081  gtttgtttcg tgactcattg tggatggaac tcaacgatgg aagctttgtc tagtggagtc
1141  ccaacggtct gttttcctca gtggggagat caagtcaccg acgctgctta catgatcgac
1201  gtgttcaaga cgggagtgag gcttagccgt ggagagacgg aggagagggt ggtgcctagg
1261  gaggaagtag cggagaggct gagagaagtt acgaaaggag agaaagcgac ggagctgaag
1321  aagaatgctt taaaatggaa ggaggaggcg gaagcggccg tggctcgcgg tggttcgtcg
1381  gatcggaatc ttgatgagtt tgtggaaaag ttgggcgtga aacctgtggc taaacagaac
1441  ggaagtctca atcaaaacgg aagtgttcaa gaacttttat tggaaaagtc ataa
```

**Figure 7-11. *SGT* sequence type 2 from *B. oleracea***

```
   1  atggaactat catcttctcc tttacctcct catgtaatgc ttgtatcgtt cccgggacaa
  61  ggccacgtta atccacttct tcgtctcggt aagctcttag cttcaaaggg tttactcgtt
 121  actttcgtca ccacagaatc atggggcaaa aagatgcgaa cctccaacaa aatccaagac
 181  cgcatcctca agccgattgg taaaggttat ctccggttcg atttcttcga cgacggtctc
 241  cccgaagacg acgacgtaag aagacacgat ttcaccatct accgaccaca cctcgagctg
 301  gtcggacaac gagagatcaa aaacctcgta aaacgttacg aggaagtgac gaaacagccc
 361  gtgacgtgtc tcatcaacaa ccctttcgtc tcatgggtct gtgacgtagc ggaagatttt
 421  caaatccctt gtgctgttct ctgggtccag tcttgcgctt gcttagcgtc gtattactat
 481  taccaccaca agcttgttaa cttcccgacc aaaacagacc ccgagatcga tgtccaaatc
 541  cctggcatgc ctctcttaaa acacgacgag atcccttctt tcatacaccc tttaactcct
 601  tattccgctt tgcgagaagt gatcattgat cagatcaaac gtcttcacaa gcctttcgct
 661  gttctcgtcg acagtttcta ctccttggag aaaggcatca tcgaccacat gtcaagtctc
 721  tctctccccg gctctatcaa accgctcggc ccgctttata aaatggccaa aacgttaata
 781  tgtgatgaca ttaaaggaga tatgtccgag accacggacc attgtatgga atggctagac
 841  tcacagccta tttcctccgt tgtttacatc tcattcggaa ccgtcgctta tataaaacaa
 901  gagcagatca atgagattgc gttcggtgtg ataaacgctg gcgtttcgtt cttgtgggtg
 961  attagacaac aagagttagg tataaacaaa gagcgacatg ttttgcctga agaagtgaag
1021  aagaaaggaa agatcgtgga gtggtgtcag caagagaaag ttttagctca tccttctgtg
1081  gtttgtttcg tgactcattg tggatggaac tccacgatgg aagctttgtc tagtggtgtc
1141  ccaacggttt gtttgcctca gtggggagat caagtcacgg acgcggttta catgatagac
1201  gttacaaaga cgggagtgag actcggccgt ggagagacag aggagagggt ggtgccaagg
1261  gaggaagtgg ccgagaggct gagagaagtt gctaaaggag agaaggcgac ggagctgaaa
1321  aagaatgctt tgaaatggaa ggaggaggcg gaagcagcgg tggctcgcgg tggttcgtcg
1381  gatcggaatc ttgaagagtt tgtggaaaag ttgggtgcga aacctgtggc taaacaaaac
1441  ggaagtctta atcaaaacgg aagtgttcaa gaacttttat tggaaaagtc ataa
```

**Figure 7-12. *SGT* sequence type 1 from *B. rapa***

```
   1  atggaactat catcttctcc tttacctcct catgttatgc ttgtatcttt cccagggcaa
  61  ggccatgtta atccacttct tcgtctcggc aagctcttgg cttcgaaggg tttactcgtc
 121  acttttgtca ccacagaatc atggggcaaa aagatgcgaa ccgccaacaa gattcaagac
 181  cgagccctca aacctatcgg taaaggttat ctccggttcg atttcttcga cgacggtctc
 241  cctgaagacg acgatgcaag cagaaccaac ttaaccatcc tccgaccaca actagagctg
 301  gtcggacaac aagagatcaa aaacctggtg aaacgttaca agggagtgat gaaacagccc
 361  gtgacgtgtc tcatcaacaa ccctttcgtc tcttgggtct gtgacgtagc cgaagatctt
 421  caaatcccct gtgctgttct ctgggtccag tcttgcgctt gcttagcttc ttattactat
 481  taccaccaca agcttgttga cttcccgact gaaacagatc ccaagatcga tgtccagatc
 541  ccatgcatgc ctgtcttgaa acacgacgag atcccttctt tcattcatcc tttttcacct
 601  tattcgggtt taagggaagt gatcattgat cagatcaaac gtctccacaa gcctttcgct
 661  gttctcatcg atactttcta ctccttggag aaagatatca tcgaccacat gacaaacctc
 721  tctcgcgccg gcgttgtcag accgctcgga ccgctttaca aaatggccaa aacgttgatt
 781  tgtgatgaca tcaaaggaga tatgtctgag acgagggatg actgcatgga gtggttagac
 841  tcgcagcctg tttcctccgt tgtttacacc tcatttggta ccacggctta cgtgacacaa
 901  gaacagatca gcgagattgc gtttggcgtt ttaaacgctg gcgtttcgtt tttgtgggtg
 961  ataagacaac aagaattagg tgtaaacaaa gagcgacatg ttttacctga agaactcaaa
1021  gggaaaggta aagtcgttga atggtgttcg caagagaaag tcttagctca tccttctctg
1081  gtttgtttcg tgactcattg tggatggaac tcaacgatgg aagctttgtc tagtggagtc
1141  ccaacggtct gttttcctca gtggggagat caagtcaccg acgctgctta catgatcgac
1201  gtcttcaaga cgggagtgag gcttagccgt ggagagacgg aggagagggt ggtgcctagg
1261  gaggaagtag cggataggct gagagaagtt acgaaaggag agaaagcgac ggagctgaag
1321  aagaatgctt tgaaatggaa ggaggaggcg gaagcggcgg tggctcgcgg tggctcgtcg
1381  gatcggaatc tcgaagaatt tgtggaaaag ttgggtgcga aacctgtggc taaacaaaac
1441  ggaagtctta atcaaaacgg aagtgttcaa gaacttttat tggaaaagtc ataa
```

**Figure 7-13. *SGT* sequence type 2 from *B. rapa***

```
   1  atggaactat catcttctcc tttacctcct catgtaatgc ttgtatcgtt cccgggacaa
  61  ggccacgtta atcctcttct tcgtctcggt aagctcttag cttcaaaggg tttactcgtt
 121  actttcgtca ccacagaatc atggggcaaa aagatgcgaa cctccaacaa aatccaagac
 181  cgcatcctca aaccgattgg taaaggttat ctccggttcg atttcttcga cgacggtctc
 241  cccgaagacg acgacgtaag aagacacgat ttcaccatct accgaccaca ccttgagctg
 301  gtcggacaac gagagatcaa aaacctcgta aaacgttacg aggaagtgac gaaacagccc
 361  gtgacgtgtt tcatcaacaa ccctttcgtc tcatgggtct gtgacgtagc ggaagatctt
 421  caaatccctt gtgctgttct ctgggtccag tcttgcgctt gcttagcctc gtattactat
 481  taccaccaca agcttgttaa cttcccgacc aaaacagacc ccaagatcga tgtccaaatc
 541  cctggcatgc ctctcttaaa acacgacgag atcccttctt tcattcaccc tttaactcct
 601  tattccgctt tgcgtgaagt gatcattgat cagatcaaac gtcttcacaa gcctttcgct
 661  gttctcgtcg acagtttcta ctccttggag aaagacatca tcgaccacat gtcaagtctc
 721  tctctccccg gctctatcaa gccgctcggc ccgctttata aaatggccaa aacgttaata
 781  tgtgatgaca ttaaaggaga tatgtccgag accacggacc attgtatgga atggctagac
 841  tcacagcctg tttcctccgt tgtttacatc tcattcggaa ccgtcgctta tataaaacaa
 901  gagcagatca atgagattgc gttcggtgtg ataaacgctg gcgtttcgtt cttgtgggtg
 961  attagacaac aagagttagg tataaacaaa gagcgacatg ttttgcctga agaagtcaag
1021  aagaaaggaa agatcgtgga gtggtgtcag caagagaaag ttttagctca tccttctgtg
1081  gtttgtttcg tgactcattg tggatggaac tccacgatgg aagctttgtc tagtggtgtc
1141  ccaacggttt gtttgcctca gtggggagat caagtcacgg acgcggttta catgatagac
1201  gttacaaaga cgggagtgag actcggccgt ggagagacag aggagagggt ggtgccaagg
1261  gacgaagtgg ccgagaggct gagagaagtt gctaaaggag agaaggcgac ggagctgaaa
1321  aagaatgctt tgaaatggaa ggaggaggcg gaagcggcgg tggctcgcgg tggctcgtcg
1381  gatcggaatc tcgaagaatt tgtggaaaag ttgggtgcga aacctgtggc taaacaaaac
1441  ggaagtctta atcaaaacgg aagtgttcaa gaacttttat tggaaaagtc ataa
```

**Figure 7-14. Putative BrHCA-GT identified on BAC KBrB089H07**

```
   1  atgggatcta tttcgggaat ggagttcgaa tcgtcaccat cttcaaatcc agttcatgta
  61  atgctcgtgt cgtttcaagg acaaggccac gtcaaccctc tccttcgtct cggcaagctc
 121  attgcttcga agggagcact cgttaccttc gtcaccactg agctttgggg aaagaaaatg
 181  agacaagcca acaagatcgt cgacggcgag cttaaaccgg ttggttccgg ttcactacgg
 241  tttgagttct tcgacgaaga atgggctgat gacgatgaac ggagaggcga tttcctttta
 301  tacatctcgc acctcgaaca ggtcgggaaa aaagaagtgt cgaagctcgt gaggaggtac
 361  gaggaagcga acgagcccgt gtcgtgtctt atcaataacc cgtttatccc atgggtctgc
 421  cacgtggcgg aagagttcaa cattccttgc gcagtgctgt gggtccagtc atgcgcttgt
 481  ttatccgctt attaccatta ccaaaacggc accgtttctt tcccgacgga gacagagccc
 541  gagctcgacg ttaagcttcc ttgcgtccca cgcttgaagc acgacgagat ccctagcttt
 601  ctccatcctt cttcgcggtt cgccggattc cgacaagcga ttctcggaca gttcgagaat
 661  ctgagcaagt ccttttgcgt cctagtcgac tccttcgacg ctttggaaca agaagtggtc
 721  gactacatgt cgaatctctg tccgatcaaa accgttggac cgcttttcaa agtggctaag
 781  acggtgacct ctgaggtaag cggtgacatc tgcaaaccgg cggatcaatg cctccagtgg
 841  ctagactcga ggcctcgatc gtcggttgtg tacatctcgt tcgggacggt tgcttatctg
 901  aagcaagaac agattgagga gatcgctcac ggagtcttga aagcgggttt atcgttctta
 961  tgggtgttta gacctccacc gcatgatctg aaggtagaga ctcatgtctt gccgcaagag
1021  ttaaaagaag agtgtgggag aggtaacggg aagattgtgg attggtgtcc acaagagcaa
1081  gtgttagctc atccttcagt ggcgtgtttc gtcactcact gtggatggaa ctcgacgatg
1141  gaggcattgt ctttgggagt tccggtggtt tgttgtccgc aatggggaga tcaagtaacc
1201  gacgcggttt atctaattga tgtttttaag acgggagtga ggcttggccg tggagcgacg
1261  gaggagaggg ttgtgccgag ggaagaagtg gcagagaagc ttttggaagc gacggttggg
1321  gagaaggcgg aggagttgag gaagaatgct ttgaagtgga aggcggaggc agaagcggcg
1381  gtggctccag gggggtcgtc ggataagaat ttccgtgagt ttgtggagag gttaggcgtg
1441  ggagcatcca aggttaagga gaacagacat taa
```

## 7.4 BAC end sequences

**Figure 7-15. BAC 2 end sequence obtained with primer SP6m.**

```
   1  tatggttcat ttgacagttc atcttggaag ggaagctaaa ctatgtggtc cagtccattt
  61  tcgctggatg tatccatttg agaggtaact tatactatat ttaaatttta aaatctctag
 121  actgtgacta aaacttttt tactctatgt agacacatga agatcctgaa agactatgtt
 181  agaaacactg cgaggccaga gggttgtatt gctgagtcct atcttgcaga agagtgcatg
 241  cagttctgca gtgcgtttct taaaaagaca accaatgtgg aggagaaatt agataggaat
 301  gctgactatg agagccagaa aatcctagag ggacgtccaa tatcggcacc aaaatcaatt
 361  aacctctctg aaatggagaa gaaaacagcc caccttgctg tcctacagaa catatctgtt
 421  gttgaccctt atgttgagta agtgactgca ttttttaact tcgtaaccta taagctacgt
 481  aggaactgaa acatttgttt tttgtgtcac aagcatgcat ctacaatttc tacaagacac
 541  aaatgcaaga tgtagacgtg atgcaacata tttgtggagt acacataccc tgaattttgc
 601  ttcttggtta aaagagcagg caaatttata acagttgttg cgtttctaaa ttaattgttc
 661  ttctattctg acaattgctt atacttttg ctagataaaa gtggattctg aaaacacatg
 721  gaagagactc taaaatggct ggcatatggt ccatgtagta ctgcaaggt
```

**Figure 7-16. BAC 2 end sequence obtained with primer T7.**

```
   1  agcttgtgat acaagtttcc ttttccttgt caagatatga ttatctttaa ggattaggaa
  61  atatgataat gatagatgga ataggaatat gntaggtcct atatatattg aatgcataat
 121  atgatgcata gtgtgatttt gatatttggt gaattgtgat cttaagatct tgtgattgat
 181  taataagaga aggacttctt attattgagt ttgtgttttt acatttggta tcagagccat
 241  catataaagc ttttgaaacc tacgaaatac aaccatggga gatgtcaagg atgagaccgt
 301  gtcgcgcaaa gaaaccggac caaactctgt taagtttccg atgttaactc catcgaacta
 361  tactgtatgg gctttacgaa tgaagatcgc actcaaagtt agcgaggtct gggaaacaat
 421  tgatcccgga aacaaagacg agaagaagaa caatatggct attgcattct tgttccaatc
 481  cataccagag acgttagtct tacaaattgg ggatatagat acagcaaaag cagcgtggga
 541  tgcaattaaa gcaagacatg tcggagctga acgagtttag agaggctagg ttacaaacct
 601  taatggcaga gttcgatagg atcaagatga atgatggtga tacagttgat atgttctcgg
```

```
661  acaaactatc ggaaattgtc tccaaggccg cttctttagg agaaagtaat tgaagaacct
721  aagcttgtaa agaaacttct taag
```

**Figure 7-17. BAC 3 end sequence obtained with primer SP6m. The whole sequence shows 83% identity to *B. rapa* BAC KBrB089H07 (15,591-16,307 bp).**

```
  1  gaaagaatac tcagacagcc aagggttact cttcaacttc tgggtgaagt gcccagatct
 61  gaaactctag tccataacat tattgtatca tcaccatact tacttccaca cccccaaaaa
121  aaaaacactt cacaaaaagc ataaactcca cgatgaaaaa gtaacagaaa aggtcttaag
181  ttttgtgtaa gacgaaaagc atataaatac acagccaaac atagacgaaa ccgagaacta
241  gggttccccc ggcggttata atttatttta gccttttctt gacccattta gtttcactga
301  gcccatctta cgtggcccaa acagacaact tcattactgt ataaactcac acaccttccc
361  ttactatttc tggtaaacca gtcaacacac aaaacccaaa aagggaaaga aagataaaca
421  tataaaatga agaaaatcca acttgtgttc atacccgcgg cggaaattgg ccatctccgg
481  tcaaccgtcg agttagctaa gaaactcctc gacgaagacg atcttctctc aatcaccgta
541  ctcatcatcc ctcgcccttc cgccgacgat tccgaccacg tttacacccc ttttcacgcc
601  gtctcaagaa cgccttccgt cacgtcacca tctccgccgc agatcaacca cccgacgttc
661  ctctcgatcc taaagccttc ctggacaacc aaaagccgaa gtgagagacg cagtct
```

**Figure 7-18. BAC 3 end sequence obtained with primer T7.**

```
  1  agcttcttat tctacacagt cctacctttt ctttcacctt tagaattagt tacaatagag
 61  aagtgattac taaccgctgt aataaatctg ttgtcatgtc atggttgttg tcttcttttc
121  catcttgatt actactactg gatagatgta gaaaatgatc tgtgtgcctt cactttgcta
181  tgtcatgtta caaatgatga tcatctctct ccttcaacat gtctctccct ggtcttttct
241  atatatccca tctgtcccat gtccctatgt ttctccataa agagttgttg gtctatatta
301  ataatacatc atgcgagatc aacaattata acttataaaa gtgggaaaac tcagacaatc
361  aagtgaagca gacatctgta gttacaaagt gtctatgacc ttcatattta atcagacgaa
421  aaaataatga atcaagtttg ttaactctga ccattctact cttgagctaa agatcttgat
481  ttctgtttca tcttacttag atcaagtaaa ctatgatatt aaacagacaa ttatagatta
541  gactcacctt ttgttattat taaaaagctg aaactcctgc caactactac ttgcaaaccg
601  tgaacaaaac agatcttctc cgcattcctt acccattgat aacttcaacg aggtgttttc
661  ttcatgattc tacaggttac gggctctagc aatcagatag aagggtcttc gtattccaaa
721  ccttagatca tttttgaaaa atcagaaaac ataatataga ttaggacaaa aaatcgaaac
781  aaatcagtgg tatcaagtag attaggacaa ccttcatatg gcatgattct taacgcctaa
841  cagtagatct atcttctgta atccttttttg gttgaaaaaa
```

**Figure 7-19. BAC 8 end sequence obtained with primer SP6m. Region of homology to *B. rapa* BAC KBrB089H07 is given in blue (43-744 bb vs 61,769-62,497 bp; 61% identity).**

```
  1  gttttctcta gttgcgtttt tgttgccagc gttgagaacc ttataaatga ctcaaccata
 61  aggtcaaggc tgggacttgg gataatttga gagaatcaac aaatcagttt acttttagat
121  tcatcttcgc atccatttcc tatgcaaaaa tcaaatttga ctcggttcct aatcgtttta
181  gaaatatttt attgaaaaaa tggtccaatt atataatttc ctactcgtga atttacgaat
241  taaccagttg ctgctttgac tccacacaca tataagctcc cttaaatcaa ggacataaag
301  gtaaattgac ataccggcca atctgtcgat tccctggttt tcatttttct tgttctcttc
361  aagaaaaaag aaaattgata ttctgtattt tttgcacaaa aaataaaata aaaaaaaata
421  ttctgtaatt cgaatccgaa ggaagccaga gatttactgt tgaaaaaaaa actagagaag
481  ggagagaaat agtgagagtc gagagacaat tatttaattt gtttcattac aaccttaaat
541  cgtcatactc tgttctgtat cttctcggat tattttctct aattagctgc tgtgcttcgt
601  tgccttgtct tcgttctctc ctgccttgtc tctctttgtc gattcagttt ctaaggagag
661  agaaaaaaag ggtttatcac aagatccctt gaagctcgca tttcttaggg tttgagtgcc
721  ccagatctca attgtaagct cttcttcatc ctcctcctcc ttcatatctc tcccaactta
781  accctctttt ttttttttt tttttttcct t
```

**Figure 7-20. BAC 8 end sequence obtained with primer T7.**

```
  1  agcttagctc catagccaat gaagcacaag ttctaggcaa aaaaatacga agatgcaaaa
 61  ttggtgaaga aacttttttgc gatgtttacc tactaagttt gcagcacatg aagcggcaat
```

```
121  aaacttgacg atgaataccg atgaactaaa gtttgctgag gtagtaggga tattaaaggc
181  tgaggagatg gaattgaaat ttaagttttc aaagcctgct caagactcta gagttacagt
241  tgatgaagat atccaaagag ctcaaaaact tgaagaatct attcacctga tgattcagaa
301  acttgaagaa accatgactc tcttgtataa gatgtccaac gacagattgc ctagaaaaga
361  agaaagacat aaaatgcagg gaggggtgtg ttttaattgt catggtattg gacatgtcaa
421  aacagattgt ccatcattca agagaaaaga aatcaagtgt aatggatgta atggttatgg
481  gcacataaaa tcagtgtgtg aacacaaaga agatgaagat gaagagaagt tgcggtgata
541  agaaggcaga gagaaaagaa tatgctctca attttgtggc tcttcatgga gttatggggt
601  ctgaaaaaga catcactaaa ggtgaatttc gtgcgtcatg tgtgtcacac accaattctg
661  atgcgtcaca cagtgataca gaaagtgatc tacatgtgga tcttgtagct gagtatcagg
721  ttctctttgg taagtttgca gaactcagtc acgaaaactt gcagctcatc aggataaagc
781  tatgttaaag ctcaagtca
```

**Figure 7-21. BAC 10 end sequence obtained with primer SP6m.**

```
  1  agcttgagtc gaattcgcca ttcctcttta aaaggcttga gatcattgat aagaaccatt
 61  tttcttgagg cagacatgtt tgttttatga gatgtgtgtg gatgttaaag ccgtcttgga
121  ttcggtatat atagataagt gaataattta ggtcaagagg agttatggac tggttagatt
181  tagtgcgtat attctctgat tttataattg atctgttgta aatattatag gattcgttgt
241  aaatattgtt aaataggttg cccgaaatta tggtaactaa tttagtttta atccgaacta
301  aattatgaga caatgtaatc tacagcaaaa agatatgcac cttggtattc acgtttactt
361  aggtaacatc ttttgtaaat caaatatttt gagaagatat agaactgatt tagaaatatc
421  gttggaactt atcttgcaag gagttattgc gtggttagat tttatgcgta tattctctgc
481  tttaaatatg ttaactaatt cctatatttt gtaaatatta tgagagtggt tttttaatat
541  tgttatatat gttgacgcag atcatggtaa ctaattacga aggcagcgga cggaaattat
601  gaggcaaggt aatctcaagc aaaaagatat gttacttgga gaaaaatata acgtggaaag
661  aaaggttaga attaaataat gatcattaaa ttagtcaatc aaatctttgg agatatatag
721  gactgattta taaaatatgg ttatataatc tcgcgaaatt tatggtaacc atccgaagat
781  atgagacatg ataaggataa ttacagactc tggttatgaa atatccaaaa ttatatgatt
841  acacattaaa taaagacgat cacataaaca atattgactt acattggtgc taaataagtg
901  aa
```

**Figure 7-22. BAC 10 end sequence obtained with primer T7. The whole sequence shows 100% identity to *UGT84A9a* (491-970 bp).**

```
  1  agcttgtcga cttcccgacc gaaacagatc ccaagatcga tgtccagatc ccgtgcatgc
 61  ctgtcttgaa acacgacgag atcccttctt tcatccatcc tttttcacct tattcgggtt
121  taagagaagt gatcattgat cagatcaaac gtcttcacaa gcctttcgct gttctcatcg
181  atactttcta ctccttggag aaagatatca tcgaccacat gacaaacctc tctcgcaccg
241  gctttgtcag accgctcgga ccgctttaca aaatggccaa aacgttgatt tgtgatgaca
301  tcaaaggaga tatgtctgag acgagggatg actgcatgga gtggttagac tcgcagcctg
361  tttcgtccgt tgtttacatc tcatttggta ccgtggctta cgtgacacaa gaacagatca
421  gcgagattgc gttaggcgtt ttaaacgctg acgtttcgtt cttgtgggtg ataagacaac
```

**Figure 7-23. BAC 18 end sequence obtained with primer SP6m.**

```
  1  agcttctctc ttattcacaa cctttgtttt caaccttatc tcttcttcac aaccattgtt
 61  ttcaaccttt ataaatttca caacctttat aaagtttatc tcttattcac aacctttata
121  aagtttatct cttcttcaca acctttgttt tcaaccttta taaatttcac aacccttata
181  aagttatatt atggtatctt cttttcataa cacttttgaa ggatcaattg atcatacttt
241  tgatcaatat tttgatcaac attttgatca aacattcgag aatttttcca ttaactatgg
301  tgatcacgaa gaagaaagga aaagaaggaa aaaaaagagt ttatatcgaa agaaatcatg
361  aagagggtga cttacgttta tggaatgatt atttcagtga aactccaaca tatcctgaaa
421  atctattccg ccgccgattt agaatgaaca agccattgtt catgcatatt gttgatcgac
481  tctccaataa agttcaattc tttcgacaaa agaaagatgg tctcagaagg ctttgtctct
541  ctccccttca aaagtgtaca acagccattc gtgtcttggc atatggttat gcggctgata
601  cggttgacga atacctccga ctcggtgaaa ctactactcg gtcatgtgtc aaaaattttg
661  tggaaggaat aatatattta ttcagcgatg agtacctaag aagaccaaca cccactgatc
721  ttcaacgtct acttggagca tcgtggattt cccgggatga taggaagaat cgattgcatg
```

```
781  cactggggag tggaagaatt gtcccaccgc ttggaaaggt caatattcac gtggttcggg
841  aaaacccaca atcgttttaa aggcggttgc ttcgtatgat ttatggatat gacatgcgtt
901  ttttggacct ccaggtactt taaa
```

**Figure 7-24. BAC 18 end sequence obtained with primer T7. The whole sequence shows 99% identity (one mismatch) to *UGT84A9a* (494-77 bp C).**

```
  1  agcttgtggt tgtaataata ataagaagct aggcaagcac aagactggac ccagagaaca
 61  gcacagggga tttgaagatc ttcggctacg tcacagaccc aagagacgaa agggttgttg
121  atgagacacg tcacgggctg tttcatcact tccttgtaac gtttcacgag gtttttgatc
181  tcttgttgtc cgaccagctc gagttgtggt cggaggatgg ttaagttggt tctgcttgca
241  tcgtcgtctt cagggagccc atcgttgaag aaatcgaacc ggagataacc tttaccgata
301  ggtttgaggg ctcggtcttg aatcttgttg gcggttcgca tctttttgcc ccatgattct
361  gtggtgacaa aagtgacgag taaacccttc gaagctaaga gcttgccgag acgaacaa
```

**Figure 7-25. BAC 20 end sequence obtained with primer SP6m.**

```
  1  agcttctggc ggatctcgga tctcggatct gacccagatc cgagcgcaat cggcgactaa
 61  cagtagatcc agggccccca aggccatcgc agacccttct cgcctcgctg atggtccaga
121  gaaagcaaag acccttttgat ttcagttaaa tccggcgaag aagatttcat cggacccaga
181  cacatcaagc aaggaggaga gaaaaggaga agatagcaga gggtgaaggc gatgaaaagg
241  aaccggtaga cagaggaggg agaccgacgg cgaaaggtgc cgtcggccac ctgaacgagc
301  agttaaagcg gcggtgacct gaagtttacg atgggagaag ggagagacta aagagagagt
361  ttcacttctc tctcctctct cttgcgtttt tggttttccc ctgattgctc tttgttctat
421  taagcgataa acaactcctg atcaaaacat caccaccaca aaaaaatcaa tcataattaa
481  tagaaatcta ctaatatata ggacggaaga gaaagagtag agaaatgact tatccatgat
541  cttcaaggtt atgctattat cttttgtgat atcgaatgga gctgatgaag aagagtatat
601  ttacacacaa tggcagacct ttgaaccaat catcatgaat catgatgtac tttaggcaga
661  gcgtaataaa aggcagtgct actaacaagt gggacgtaca ctgcacaatc agttatttag
721  aaccgatctt cctcttttcg tcaagacgac ccaaagctcg caagtaccct ccgccaaatg
781  tggtttctta cgccacatcc cttctactca agaccgtcac gcgacaacac accatcaaat
841  ttaatggttc aactaatgtc tttcagcttg atttttaagt tacaataatt attttctaat
901  tcaaggtcct ttgtaacgta aaacttttc attcttctgc atcattcaat aacgaatatc
961  ctgtcaagaa aaaataaagt agttc
```

**Figure 7-26. BAC 20 end sequence obtained with primer T7. The whole sequence shows 99% identity to *B. rapa* BAC KBrB089H07 (18,236-18,764 bp).**

```
  1  agcttctaca tcttcaccat cctttgagtc catcaacacc acgaacacta tgttatcgta
 61  gatgttatga gaagcaggtc caatccatac cggattaccc catccaaagt caacctcata
121  gaaaggcttc ctacaccagc tagacattgt gtacaagtct atacccggtc ttaagtactc
181  cgacatgaaa cttcccatcg cagtcaacaa gttctgacca agtgtggtgg tggtgatggt
241  agtatcatta ttattacagc cttggagatt ctccttaatc atctcgttga caccttcttt
301  aatctttcta aactcagcca cgatttcacc gatctccaac tcgctctctg cctctctctt
361  gagaaagaaa gcagattgta ggttccccat tgcgttttcc ggcaaaacat tagtggggat
421  cctaagcctc aagtccatat gcttgagtca tcagcgttga ctttggagaa agtgagttag
481  accgtgaggt ttttgcggcg catctccaaa taagcgacat gatggctttc
```

**Figure 7-27. BAC 21 end sequence obtained with primer SP6m. The whole sequence shows 95% identity to *B. rapa* BAC KBrB089H07 (98,529-99,373 bp). Region of homology to *At4g15475* (genomic sequence) is given in blue (6-749 bp vs 1090-2054bp, 61%).**

```
  1  agcttctcga ggtgctgttt cttgattccg aatgcatcca tgataaagga gtggtggcgg
 61  tggtgaaggg atgcagacgt ttgaagagtc ttaagcttca gtgtgtcaac gtgacggatg
121  ctgcgttttc cgtggtggga gacttttgtg tctcgttgga gatgttggct ttgtatagct
181  ttcagcagtt tactgacaag ttagtgcaat ctttactcct ttttagcata atctaatctc
241  tatgatatgc ttatatgctt tttttttaa tttgttgttt ccaggggaat gaagtctatt
```

```
301  ggtaaagggt gtaagaatct gaaagatctc actctaagcg actgctactt tgtgagctgc
361  gatggtttag aagctattgc tcatggctgc aaggagctta cacgtataga gattaatgga
421  tgtcacaaca ttggtactcg tggactagag gcagttggaa aatcttgccc gtgcgttttc
481  tttctcaatt ttcttattct tggtgacaaa ctcattgact aaaccgttct ttttctcggg
541  ggatgtatgc tgcaggcgcc tcactgagtt atatttgtta tattgccaaa ggattggtaa
601  ttcggctctt taccaaattg gaaaagggtg taagtccctt gagacgcttc acttaataga
661  ctgcgcaggc attggagaca ctgctatgtg cagtattgct aagggctgtc gcaacctcaa
721  gaaactccat atccgcagat gctacgaggc agataatctt cacattctct tctccatggt
781  ttctcctatt ctatagacat acaaactctt tttttgtttg ttttgtaggt tggaaacaaa
841  gg
```

**Figure 7-28. BAC 21 end sequence obtained with primer T7.**

```
  1  agcttatact cctgctcgga tccgccggtt tccacccgta tcatgaaggc cgtctccaac
 61  accttcatct ggtagtttcc ggagcgttct ttgctaggga gatgaaatct cgacaatatt
121  ccctaacagt ccccctcctg ctttaatgtc atcaacctct ctctggctgt cgtttcgtaa
181  gatgtggaaa actgttcgag aacgcgatat ttcatctgct cccagttcac gaacggattc
241  cgatctcgct cccaacggta ccacattaga gcctcctcgt caaaacacat cctcaccgct
301  cggagcttcc cttcctccga gaactcccct aattcgaaat actgctccac gcgtacgacc
361  caattctcgg cgttttcccg ttgaacagag gaatctcgat cttccgggcc ttgcgcg
```

**Figure 7-29. BAC 30 end sequence obtained with primer SP6m.**

```
  1  agcttcatgt ggagtgtttt ttatatatat ttacacttgt ttgtttgttt gatggttaca
 61  ggatgggcat gattctgagg atactaagca gagcactgct gatatgactg cttttgtgag
121  tagatttgtt ccccattgtc tctctgtttt agacttatac cacttggttt tgttactgat
181  tccttgaatt tttttcatat ccaggtccaa aatctcctgc agcagatggt gggctttta
241  actctcactt ctttaggtct tggaatcaca gcttcaattg gtagattctg atttcaagtt
301  tgctttcatt gttgtgtgtt tcagcaaagc aggtttcaga caatgtccga ctccatcatc
361  acaaagaata tccttttttct atcattcata tatttaatgt ttcttccttt agaataattt
421  gaaaatgttt ccccctctct cctagagggt ggaggagatg ttgatctcta gtattagttt
481  atttctgcca ttaaatctac cgaaagtaat agagttatct actgctagat tagatagagt
541  atttcttgag attctaaatg tctattctgt agataatcac cacatggttc tatgtttagc
601  tctctctctt gttaagttgc tctgttatgc atacttagtt ggtgtgagtc ctacacggaa
661  tcacctttac taccatgatc agtgatataa gtttcggaat ctgtaggttt gatgatttgt
721  gttggtttaa accgtaattt ccttgacttt gaatgaattt cactccaatg acatgggagg
781  cagaatcaat gagctggaac aaagcatcaa tgatctaaga gccgagatgg gagttgaaag
841  gcactccacc tcctgcctcc aaatcaggcg atgaaaccca aaacaccggg ctagttagtt
901  cccacttgat aaaaaaccta caggtccat
```

**Figure 7-30. BAC 30 end sequence obtained with primer T7. The whole sequence shows 95% identity to *B. rapa* BAC KBrB089H07 (93,416-93,727 bp) and 81% identity to *At4g15470* (genomic sequence, 1,483-1,802 bp).**

```
  1  agctttgata ttgacactgt ccgcggttgg atctctaacc gcatacacat ttctgggctg
 61  caaagaaggg aaaagacttc accttccttg gacccattct cttcaccagc ctcatcattc
121  ttgtagtgac cagcttcatg caggtctgtc tcttaggaac aagacaaaac accccaagtt
181  ccacacacat gttaaaaaga aaaaaggctt aattttgttt tgttatcata ataatattgc
241  agatgttctt ccctcttggc ccgacctcag ttgccatata tggaggaatc agcgcattgg
301  tcttctgtgg at
```

**Figure 7-31. BAC 33 end sequence obtained with primer SP6m. The whole sequence shows 99% identity to *B. rapa* BAC KBrB089H07 (55,548-56,136 bp).**

```
  1  cgcctaagag cttagcgcga ttttttgggg aaaacaactt tgtttgcaaa gagtaaggag
 61  tcccataagc atgttcggaa cttgacatcc cagttcctag gttcacaagg tttgagattg
121  aggatgatac aagacattga ttctttggct cgcacacaca tggagttagg agctaagaat
181  ggtggtcttg atatcaagga aacatcaagc aaggtaaaaa gatctcgtcc atttggagct
```

```
241  gtatgtaaaa taacatcaac tagattagtt aagatggatc atcacatgtt ataaatatct
301  tcaattatta cagatcttaa ttgaatgtct tgcgaagaaa gttatgggtg atatggaacc
361  ccatgcagca aaagaactaa cacaatgttg gggaggtttt cctagagggt ggtttcgttt
421  ctcatggagc attcctggga atggtgtcta tagaatgttg aaggtacgtg ttttcttggt
481  tcatataatc tttgatagct ctatatatgg aaaatgaatc gagggcgatg ttgttttata
541  taacaggcaa gacaccatat gatgaattta ctacaagaga ctgtgttgaa
```

**Figure 7-32. BAC 33 end sequence obtained with primer T7. Region of homology to *At4g15890* (genomic sequence) is given in blue (70-799 bp vs 5,343-4,635 bp C, 86%)**

```
  1  agcttaacta gatagggaca acatgagaaa cttgcttaca ttcgttttcg agcatacaca
 61  atagacatta gtaatttagt tacaacaaaa gctatttcct gagtctcgtc tcgctctcta
121  ttttagttat cgtcatcaga catgaaagtt ggtcttcttc tgagagactc cagcacatca
181  gccgatggtc gaccactctg tttacctcct ccaaggctcg accttagctt cggaacgctc
241  attgcgctca gaggcggcgt tgcagctcca ccgttcactt ctctcagtac cgaccttgcc
301  gctttagccg ccattgcatc agcagccgca gataccactc tcgattccag tcccgaatcc
361  ctaacattct cctccctttc tgcttctgtt gttggttgct ctggttcgcg gcttctatga
421  cttgtcccgc aaccaacccc ttggttagcc atggaccttg cctttgcagc ctcgtgcttc
481  tgcatatcgt agacgagctc tttctgaagc tgtttatcag cgacgagcat ttcttctatc
541  tcgttcttgt agtctttcag gagcacacgg aggcagtcca tgagcgagcc tgtgagagga
601  ctgttcttgc tctccaatag tctcttgagc tcaatgaaga ttggaatcgt gttctgaatc
661  aacccttttc taacagcttg cgttatggct cttcctttgg cagccgcccc gttggagtct
721  ccaccttctt cttccatctc tactgtctca gaacgaagat cctcttgaaa ctgataaacg
781  gatctctttg ccgccaaga
```

**Figure 7-33. BAC 37 end sequence obtained with primer SP6m.**

```
  1  cttctcctca tcactcaaca caagctccta gaaaaagaaa cagaaacatg ttaaacacct
 61  ctgccacctt taagacataa ataatctctc catatatata tttacagaaa ttttatataa
121  ttttttacag taaagaataa cctggtcttt tgaaatatat tctccggtta tattcaatat
181  cctccgacag acactcccac gtgtgttccc aaacattctc tggtgatgct aaactatcat
241  tattaagcat catagcaaaa cactgtcgta aatcacttgc agagctatca taacttctcc
301  tcaacaaatc atcaatatac tcgtgatcat cgtctaataa gccacgtgca caacatgctt
361  ctttgtaact cttgtaaaga acaccttcaa atgtctttat atcctcataa ctggtaggac
421  ctcttacaat gttcaacaac acgcggagat aatatgaatc ttcttgttta cgtggagcat
481  agttaatcct gccaagacta aatcctcttt tccttctttt gaatttcttc aacttcttat
541  cgtaagtaaa gaagttcggt atttgcgcat atgtgagggt tttcgcaagt gaatcgatct
601  tacacaactc aaaccatgct aagaacattg tattctcaat gagcttacga ctaaccacgg
661  cttccatctt atctttacct ctgaatataa catgttgttt tccaggaaga tgaaaactaa
721  agcttctcga ctgcagttag atcgataatg aattggaaaa
```

**Figure 7-34. BAC 37 end sequence obtained with primer T7. The whole sequence shows 98% identity to *UGT84A9a* (491-1,000 bp).**

```
  1  agcttgtcga cttcccgacc gaaactgatc ccaagatcga tgtccagatc ccgtgcatgc
 61  tgctcttgaa acacgacgag atcccttctt tcatccatcc tttttcacct tattcgggtt
121  taagagaagt gatcattgat cagatcaaac gtcttcacaa gcctttcgct gttctcatcg
181  atactttcta ctccttggag aaagatatca tcgaccacat gacaaacctc tctcgcaccg
241  gctttgtcag accgctcgga ccgctttaca aaatggccaa aacgttgatt tgtgatgaca
301  tcaaaggaga tatgtctgag acgagggatg actgcatgga gtggttagac tcgcagcctg
361  tttcgtccgt tgtttacatc tcatttggta ccgtggctta cgtgacacaa gaacagatca
421  gcgagattgc gttaggcgtt ttaaacgctg acgttttcgt tcttgtgggg tgataagaca
481  acaagaacta ggtgtaaacc aagaacgaca tg
```

**Figure 7-35. BAC 42 end sequence obtained with primer SP6m.**

```
  1  agctttaata ctcagatttc gaatcatttt cttcgttctc taaattttta gtggaagaga
 61  atgagggttt gcatataaac cttacgtatg gtttgacgat aaactctttt ccatgtcaaa
```

```
121  tattcttact aaaccagttt aaatcaaacc aacttctata ctagtttcaa ataaaccaac
181  tggtttgtct gctgacacca aagcgtcaca agagcgcttt aggaaaatca taaaagccag
241  ggaaagcgag gaaaagcgat cgctttattt aactagctgg ccaagtggtg ttgaggcgat
301  gaggtcatcg caaggggtcg cctagcgcca tggcgacgga ggctgtaaca cccgcccttt
361  ccccatttcc aaccggggtt cctgggcgcg gggttaagga cacacatgtc tactttcccg
421  atatcttcgt gtgtcatgtt actagtccca tgagtttcag gtgcgtttcc tttcttacca
481  tcgacacttc cacttaaggt tctgcaaaaa gggagggaaa aagaggggtg agtatgaata
541  ctcagtgaag cgattctaga ccagtctccc ccccatgagc caatgcgaaa cgagaaccga
601  ctagccacta cacccaattc gatgcaactt acttaaccag ttactacgat accatttcac
661  gaaatcaagc aatgtaatga actcagtcat cacccaattc acattaatcg ttttaagact
721  ccaatcatcc tagaactcta ggactttaca cagctcaagc aaaccaaccc ctcctctttt
781  cttgttgcat tcctgtggga gccgcctgcc ctggtgggct cgacatccgg ttgatcccac
841  tcaaattacc ctaaggagtg atttatactc ttctcaaata aaatgtc
```

**Figure 7-36. BAC 42 end sequence obtained with primer T7. The whole sequence shows 99% identity (one mismatch) to *UGT84A9c* (1,076-614 bp C).**

```
   1  gaaggatgag ctaaaacttt ctcttgctga caccactcca cgatctttcc tttcttcttc
  61  acttcttcag gtaaaacatg tcgctctttg tttataccta actcttgttg tctaatcacc
 121  cacaagaacg aaacgccagc gtttatcaca ccgaacgcaa tctcattgat ctgctcttgt
 181  tttatataag cgacggttcc gaatgagata taaacaacgg aggaaatagg ctgtgagtct
 241  agccattcca tacaatggtc cgyggtctcg gacatatctc ctttaatgtc atcacatatt
 301  aacgttttgg ccattttata aagcgggccg agcggtttga tagagccggg gagagagaga
 361  cttgacatgt ggtcgatgat gcctttctcc aaggagtaga aactgtcgac tagaacagcg
 421  aaaggcttgt gaagacgttt gatctgatca atgatcactt ctc
```

## 7.5 Construction of dsRNAi vectors

**Figure 7-37. *UGT84A9a*. Gene specific fragment for *UGT84A9ai*/*BnSCT1*i construct (1173-1353 bp) is given in red, gene specific fragment for *UGT84A9a*i (*BnSGT1*i) construct (1-213 bp) is given in blue (Hüsken et al., 2005).**

```
   1  atggaactat catcttctcc tttacctcct catgttatgc ttgtatcctt cccaggacaa
  61  ggccacgtta atccacttct tcgtctcggc aagctcttag cttcgaaggg tttactcgtc
 121  acttttgtca ccacagaatc atggggcaaa aagatgcgaa ccgccaacaa gattcaagac
 181  cgagccctca aacctatcgg taaaggttat ctccggttcg atttcttcaa cgatgggctc
 241  cctgaagacg acgatgcaag cagaaccaac ttaaccatcc tccgaccaca actcgagctg
 301  gtcggacaac aagagatcaa aaacctcgtg aaacgttaca aggaagtgat gaaacagccc
 361  gtgacgtgtc tcatcaacaa ccctttcgtc tcttgggtct gtgacgtagc cgaagatctt
 421  caaatcccct gtgctgttct ctgggtccag tcttgtgctt gcctagcttc ttattattat
 481  tacaaccaca agcttgtcga cttcccgacc gaaacagatc ccaagatcga tgtccagatc
 541  ccgtgcatgc ctgtcttgaa acacgacgag atcccttctt tcatccatcc tttttcacct
 601  tattcgggtt taagagaagt gatcattgat cagatcaaac gtcttcacaa gcctttcgct
 661  gttctcatcg atactttcta ctccttggag aaagatatca tcgaccacat gacaaacctc
 721  tctcgcaccg gctttgtcag accgctcgga ccgctttaca aaatggccaa aacgttgatt
 781  tgtgatgaca tcaaaggaga tatgtctgag acgagggatg actgcatgga gtggttagac
 841  tcgcagcctg tttcgtccgt tgtttacatc tcatttggta ccgtggctta cgtgacacaa
 901  gaacagatca gcgagattgc gttaggcgtt ttaaacgctg acgtttcgtt cttgtgggtg
 961  ataagacaac aagaactagg tgtaaacaaa gagcgacatg ttctgcctga agaactcaaa
1021  gggaaaggta aagtcattga atggtgttca caagagaaag tcttagctca tccttctgtg
1081  gtttgtttcg tgactcattg tggatggaac tcaacgatgg aagctttgtc tagtggagtc
1141  ccaacggtct gttttcctca gtggggagat caagtcaccg acgctgctta catgatcgac
1201  gtgttcaaga cgggagtgag gcttagccgt ggagagacgg aggagagggt ggtgcctagg
1261  gaggaagtag cggagaggct gagagaagtt acgaaaggag agaaagcgac ggagctgaag
1321  aagaatgctt taaaatggaa ggaggaggcg gaagcggccg tggctcgcgg tggttcgtcg
1381  gatcggaatc ttgatgagtt tgtggaaaag ttgggcgtga aacctgtggc taaacagaac
1441  ggaagtctca atcaaaacgg aagtattcaa aaacttttat tgcaaaagtc ataa
```

**Figure 7-38. *BnSCT1*. Gene specific fragment for *UGT84A9ai*/*BnSCT1*i (1052-1248) is given in red.**

```
   1  atgagaaatc tttacttttct agtcttattt ccgttgagca tcttgatttt ggttgatgct
  61  tctttgcatg tgaagtatct tcctggtctt gaaggtcctc ttccttttga gctcgagact
 121  gggtatgtga gtgttggtga atctggagat gttgagctct tttactactt tgtgaaatca
 181  gagagcaatc cagataaaga tcctctcatg atttggctaa ctggtgggcc tggatgcagc
 241  tcaatttgtg gtttactctt tgcaaatggt cctttggctt ttaaagggga tgagtataat
 301  gggacactgc ctcctttaga gctaacatct ttttcttgga caaaggtggc taacatttta
 361  tatttggaat ctcctgctgg ttctggatat tcttatgcca aaactcggcg tgctgctgag
 421  acgagcgaca ccaaacaaat tcaccaaatc gaccagttcc ttaggagttg gtttgtggac
 481  caccctgagt ttatatccaa ttcattttac gttggtggag attcatattc cgggaagatt
 541  gttccaggag ttgtgcaaca gatttcactt ggaaatgaaa aaggtctcac accactcata
 601  aatattaagg gatatgttct tggaaaccct gcagtacgta caaacttaga accaaatcat
 661  agagtttcat ttgcgcatcg gatgggactt atttcagatg agctccatga gtcacttgaa
 721  agaaactgtg gaggcaaatt ctttaacgta gatccaagta atgcaaaatg ttcaaatggg
 781  cttctagctt atcatcagtg tatctcagag atatacatag agcagatttt gttaccaaac
 841  tgcaaagtag attatgtctt agcagacata tcacaaacct taccaaatat cagaaccagt
 901  cgaagaagag aactcaagga gttttcaaga aatgattcat catcgttgcc tcctccaagc
 961  tgctttactt ataggtattt tctgtctgcc ttttgggcaa atgatgaaaa tgtacgcaga
1021  gctttaggcg tgaagaaggg cttcggaaaa tggagtcgat gcaacactca aaacatacca
1081  tatacatatg atattcacaa tgccattcca tatcacgtta ataatagccg taaaggcttc
1141  cgcgctctca tctacagtgg tgatcatgat atgatgatac ctttctcttc aactgaagca
1201  tggatcaaat ctctcaacta ttccattgtt gatgactgga gaccttggat gatgaatagc
1261  aatcaagttg ctggatatac aaggacctat gcaaataaga tgacatttgc aaccatcaag
1321  ggaggaggac acaccgctga gtataatcca gaccaatgct cacttatgtt caaaagatgg
1381  attgatggtg aatctctctg a
```

## Danksagung

Diese Arbeit wurde am Leibniz-Institut für Pflanzenbiochemie in der Abteilung Sekundärstoffwechsel angefertigt. Mein besonderer Dank gilt dem Leiter dieser Abteilung, Herrn **Prof. Dr. Dieter Strack** für die Überlassung des Themas, der Bereitstellung des Arbeitsplatzes und die Unterstützung bei der Fertigstellung der Arbeit.

Ganz herzlich möchte ich **Herrn Dr. Carsten Milkowski** für die fachliche Betreuung, die ständige Diskussionsbereitschaft und die vielen praktischen Ratschläge danken.

**Herrn Dr. Alfred Baumert** danke ich für die Einarbeitung in die Handhabung der HPLC und die Hilfe bei technischen und biochemischen Fragen.

**Frau Prisca Viehöver** und **Herrn Prof. Dr. Bernd Weisshaar** (Universität Bielefeld) danke ich für die Unterstützung bei der BAC Sequenzierung.

**Herrn Prof. Dr. Peter Ian Mackenzie** (Flinders University School of Medicine, Adelaide, AU) danke ich für die korrekte Benennung der *UGT84A9* Loci nach der UGT-Nomenklatur.

**Herrn Dr. Rod Snowdon** (Universität Giessen) möchte ich für die Bereitstellung der genomischen BAC Bank und die vielen Ratschläge danken.

**Herrn Dr. Jose Orsini** (Resistenzlabor, DSU) danke ich für die Transformation der Rapspflanzen.

**Frau Dr. Sabine Mikolajewski** (Bayerische Landesanstalt für Landwirtschaft, Freising-Weihenstephan) danke ich für die Hilfe bei der Durchführung der cDNA-AFLP Experimente.

Der **Norddeutschen Pflanzenzucht** (NPZ, Hohenlieth) und **Herrn Dr. Rakow** (Agriculture and AgriFood Canada, Saskatoon, SK, Canada) danke ich für die Bereitstellung des Saatgutes.

Ganz besonderer Dank gilt **Frau Sylvia Vetter** und **Frau Claudia Horn** für die seelische, moralische und technische Unterstützung der Arbeit.

Außerdem möchte ich mich bei der gesamten **Abteilung Sekundärstoffwechsel** bedanken, insbesondere bei allen ehemaligen und derzeitigen Mitgliedern der **AG Phenylpropanstoffwechsel, Kathleen Clauß, Ingrid Otschik, Dirk Meißner, Jessica**

**Leuchte, Felix Stehle, Jenny Teutschbein** und **Diana Weier,** für die Diskussionen, Ermutigungen und Ratschläge, die alle zum Gelingen der Arbeit beigetragen haben.

Zum Schluß möchte ich mich noch ganz herzlich bei meiner Familie, meiner Mutter **Ursula Mittasch**, meinem Bruder **Konrad Mittasch** und meinem Freund **Jorrit-Jan Krijger** bedanken, ohne deren Unterstützung diese Arbeit nicht möglich gewesen wäre.

## Publikationen und Tagungsbeiträge

### Publikationen

Mittasch J., Strack D., Milkowski C.: Secondary product glycosyltransferases in seeds of *Brassica napus*. In: *Planta.* Vol. 225: 515-522, 2007

Weier, D., Mittasch, J.; Strack, D.; Milkowski, C.: The genes *BnSCT1* and *BnSCT2* from *Brassica napus* encoding the final enzyme in sinapine biosynthesis - molecular characterization and suppression. In: *Planta*. Vol 227: 375-385, 2008

### Tagungsbeiträge

Mittasch J., Mikolajewski S., Strack D., Milkowski C.:Characterization of *UGT84A9* Loci in *Brassica napus*. Poster: Brassica 2008 5th ISHS International Symposium on Brassicas and 16th Crucifer Genetics Workshop, Lillehammer, Norwegen, 08.-12.09.2008

Mittasch, J.; Baumert, A.; Strack, D.; Milkowski, C.: Towards low sinapine oilseed rape. In: *Proceedings of the 12th International Rapeseed Congress*: 102-105, 2007

Mittasch, J.: Reduction of sinapine content in *Brassica napus* seeds. Vortrag: International Students Conference 2006 Gatersleben, 30.05.2006

Mittasch, J.; Meißner, D.; Baumert, A.; Strack, D.; Milkowski, C.: Ester-forming glucosyltransferases in phenylpropanoid metabolism of Brassiceae. Poster: Botanikertagung Braunschweig, 05.-10.09.2004

# Lebenslauf

Persönliche Daten:

| | |
|---|---|
| Name: | Juliane Mittasch |
| Anschrift: | Torstrasse 27 |
| | 06110 Halle |
| Geburtsdatum: | 03.12.1976 |
| Geburtstort: | Lauchhammer |

Schulischer Werdegang:

| | |
|---|---|
| 09/1983 – 07/1991 | Politechnische Oberschule „Fritz Kube“ in Kroppen |
| 08/1991 – 06/1996 | Gymnasium mit erweiterter mathematisch-naturwissenschaftlich-technischer Ausbildung „Max Steenbeck“ in Cottbus<br>Abschluss: Allgemeine Hochschulreife – Abitur |

Studium und Beruflicher Werdegang:

| | |
|---|---|
| 10/1996 – 04/2002 | Biologiestudium an der Martin-Luther-Universität Halle-Wittenberg<br>Abschluss: Diplom |
| 06/2002 – 11/2002 | Wissenschaftliche Mitarbeiterin im Landwirtschaftlichen Institut der Martin-Luther-Universität Halle-Wittenberg |
| 12/2002 – 05/2008 | Promotion am Leibniz Institut für Pflanzenbiochemie in Halle, Abt. Sekundärstoffwechsel |

## Eidesstattliche Versicherung

Hiermit erkläre ich, dass ich mich mit der vorliegenden Dissertationsschrift erstmals um die Erlangung eines Doktorgrades bewerbe. Ich versichere an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt, und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Halle, den