Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/sjar

Spanish Journal of Agricultural Research 2010 8(3), 662-671
ISSN: 1695-971-X
eISSN: 2171-9292

Variation of five major glucosinolate genes in *Brassica rapa* in relation to *Brassica oleracea* and *Arabidopsis thaliana*

B. Yang, D. Qiu and C. F. Quiros*

Department of Plant Sciences. University of California, Davis, CA 95616. USA

Abstract

Glucosinolates and their derivatives isothiocyanates are important secondary metabolites in the Brassicacea that has biological activity, such as cancer protecting and biofumigant properties. The putative orthologs of five major genes in the glucosinolate biosynthetic pathway, *Bra.GSELONG.a*, *Bra.GSALK.a*, *Bra.CYP83B1*, *Bra.SUR1.a* and *Bra.ST5.a*, were cloned from both cDNA and genomic DNA from different subspecies of *Brassica rapa*. Interspecies comparative analysis disclosed high conservation of exon number and size for *GS-Elong, GS-Alk, GS-CYP83B1* and *GS-ST5a* among *B. rapa*, *B. oleracea* and *A. thaliana*. Splice site mutations caused the differences observed for exon numbers and sizes in *GS-SUR1* among the three species. However, the exonic sequences were highly conserved for this gene. There were not major differences of intronic sizes among the three species for these genes, except for intron 1 for *GS-Elong* in two subspecies of *B. rapa.* The cloning of the putative orthologs of all these major genes involved in the glucosinolate biosynthesis pathway of *B. rapa* and sequence analysis provide a useful base for their genetic manipulation and functional analysis.

Additional key words: chinese cabbage, comparative genomics, glucosinolates, rapeseed, turnip.

Resumen

Variación de cinco genes mayores de glucosinolatos en *Brassica rapa* en relación con *Brassica oleracea* y *Arabidopsis thaliana*

Los glucosinolatos y sus derivados, los isotiocianatos, son metabolitos secundarios importantes propios de la familia Brassicaceae, ya que poseen actividades biológicas tales como protección contra el cáncer y biofumigación. Se describe la clonación en base a DNA genómico y cDNA de cinco genes mayores implicados en la síntesis de glucosinolatos en subespecies diferentes de *Brassica rapa*, *Bra.GSELONG.a*, *Bra.GSALK.a*, *Bra.CYP83B1*, *Bra.SUR1.a* y *Bra.ST5.a*. El análisis comparativo de los genes *GS-Elong, GS-Alk, GS-CYP83B1* y *GS-ST5a* con sus homólogos en las especies *Brassica oleracea* y *Arabidopsis thaliana* reveló una alta conservación en número y tamaño de exones. Se observaron diferencias en el número de exones en *GS-SUR1* en las tres especies debido a mutaciones en los sitios de procesamiento, aunque las secuencias de estos exones mantienen una alta conservación. No se detectaron mayores diferencias en el tamaño y número de los intrones de estos genes en las tres especies, excepto para el intrón 1 de *GS-Elong* en dos subespecies de *B. rapa*. La clonación y el análisis de estos genes en *B. rapa* facilitarán su manipulación genética y análisis funcional.

Palabras clave adicionales: col china, colza, genómica comparativa, glucosinolatos, nabo.

Introduction

Glucosinolates and their breakdown products, the isocyanates, are important secondary metabolites in the Brassicacea. These compounds posses biological activity ranging from cancer protection to biofumigation. The synthesis of glucosinolates (GSL) can be divided into three independent stages: side-chain elongation, glycone formation and side-chain modification (Wittstock and Halkier, 2002). Several members of

* Corresponding author: cfquiros@ucdavis.edu
Received: 12-11-09; Accepted: 31-05-10.

Abbreviations used: GSL (glucosinolates), MAM (methylthioalkylmalate synthase).

methylthioalkylmalate synthase (MAM) gene family are known to be involved in catalyzing the side-chain elongation cycle of aliphatic GSL biosynthesis in *A. thaliana*. *MAM1* (At5g23010) and *MAM3* (At5g23020) are tandem duplicates on chromosome 5 at the *GS-Elong* locus (Kroymann *et al*., 2001; Field *et al*., 2004; Textor *et al*., 2004; Benderoth *et al*., 2006; Textor *et al.,* 2007). Aliphatic GSL might be modified further by other genes, such as *GS-Alk,* which is responsible for the conversion of methylsufinylalkyl to alkenyl GSL (Mithen *et al.,* 1995). During the biosynthesis of the glycone moiety, cytochrome-P450-dependent monoxygenases in the *CYP79* gene family are responsible for catalyzing the conversion of amino acids to aldoximes. Five of seven *CYP79* homologs identified in the *Arabidopsis* genome have been shown to be involved in this process (Hansen *et al*., 2001; Reintanz *et al.*, 2001; Chen *et al*., 2003; Halkier and Gershenson, 2006). *CYP83A1* (At4g13770) and *CYP83B1* (At4g31500), two members of another cytochrome P450 gene family in *Arabidopsis*, have been suggested for the functional role in oxidation of aldoximes in the biosynthesis of GSL (Hansen *et al.*, 2003). *CYP83B1* has higher affinity for the indole-3-acetaldoxime derived from tryptophan and aromatic aldoximes derived from phenylalanine or tyrosine (Naur *et al.,* 2003). In *Arabidopsis*, S-alkyl thiohydroximate is then cleaved and converted to thiohydroximic acid by the C-S lyase *SUR1* gene (At2g20610), which lacks side chain specificity (Mikkelsen *et al*., 2004). The final step in the biosynthesis of the GSL core structure is catalyzed by a desulfoglucosinolate. Genetic and biochemical characterization of a small gene family, *AtST5,* coding for sulfotransferase in *A. thaliana* showed that *ST5a* (At1g74100) prefers tryptophan- and phenylalanine-derived desulfoglucosinolates as substrates, whereas *ST5b* (At1g74090) and *ST5c* (At1g18590) prefer long chain aliphatic desulfoglucosinolates (Piotrowski *et al.*, 2004). The subsequent side-chain modification reactions will determine the final structure of GSL. Methionine-derived aliphatic GSL are known to be the most extensively modified and several genetic loci have been mapped in *Arabidopsis* and *Brassica* cultivars (Parkin *et al.,* 1994; Mithen *et al.*, 1995; Giamoustaris and Mithen, 1996; Hall *et al.,* 2001).

Most of the research on GSL biosynthesis has been done in the model plant *A. thaliana.* Some work has been applied to crops of economic importance belonging to the species *B. oleracea*. The two major genes involved in the aliphatic GSL side-chain elongation and modification, *BoGSL-Elong* and *BoGSL-Alk* have been cloned (Li and Quiros, 2002, 2003). Inheritance analysis also demonstrated that another gene, *BoGSL-PRO*, is responsible for three-carbon side chain elongation (Li *et al*., 2001). Two *B. oleracea* BAC clones, which contain the orthologs of *AOP* (B21H13) and *MAM* (B21F5) families, were identified and used for comparative analysis between *A. thaliana* and *B. oleracea* (Gao *et al.,* 2004, 2006). The work presented here focused on the cDNA and genomic DNA cloning of the putative orthologs of five major genes involved in the *B. rapa* glucosinolate biosynthetic pathway: *Bra.GSELONG.a*, *Bra.GSALK.a*, *Bra.CYP83B1*, *Bra.SUR1.a* and *Bra.ST5.a*. The sequence analysis of the cDNA clones provides a useful base for further functional analysis of these genes and help to elucidate the glucosinolate biosynthesis pathway of *B. rapa* in the future.

Material and methods

Plant material

Eight varieties representing three subspecies and including five different *B. rapa* crops, Chinese cabbage, turnip, turnip tops, oriental cabbage and Pak choi, were used in this study (Table 1). The source of the plant material was the Brassica working collection from the Department of Plant Sciences at UC Davis. All plants were grown in the greenhouse and were sampled six weeks after germination for DNA and RNA isolation. Genes were named following the standard nomenclature proposed by Østergaard and King (2008).

DNA isolation

DNA was isolated using a modified CTAB method (Saghai-Maroof *et al*., 1984) in 2x DNA extraction buffer (1.4 M NaCl, 100 mM Tris-HCl, 20 mM EDTA, 2% CTAB, pH 8.0). The ground leaf samples were incubated for 1.5 hr at 65°C with occasional gentle mixing. An equal volume of chloroform: isoamyl alcohol (24:1) was added and mixed gently but thoroughly followed by centrifugation at 9,000 rpm for 10 min. The supernatant was mixed gently with an equal volume of isopropanol at room temperature. The pellet through centrifugation was re-suspended with TE buffer. RNA was removed with RNase (50 μg mL^{-1})

Table 1. Index for *Brassica rapa* varieties

Acc.#[a]	Crop	Botanical name	Variety	Origin
B0233	Ori. cabbage[b]	*B. rapa* L. var. *pekinensis*	Kwan-Hoo-Choi	USA, California
B0488	Ch. cabbage[c]	*B. rapa* L. var. *pekinensis*	Matshushima	USA, Maine
B0218	Turnip	*B. rapa* L. var. *rapa*	While Lady (F1)	Japan
B0219	Turnip	*B. rapa* L. var. *rapa*	Tokyo Top (F1)	Japan
B0493	Turnip	*B. rapa* L. var. *rapa*	Yorii Spring	USA, Maine
B1658	Pak Choi	*B. rapa* L. var *chinensis*	Japanese Greens, Taisai	UK, Wellesbourne
B1668	Turnip Tops	*B. rapa* L. var. *rapa*	Rapa 60 Giorni	UK, Wellesbourne
B1670	Turnip Tops	*B. rapa* L. var. *rapa*	Rapa Febbrario	UK, Wellesbourne

[a] Accession number. [b] Oriental cabbage. [c] Chinese cabbage.

(Roche Applied Science, IN) and RNase was removed with chloroform: isoamyl alcohol (24:1) followed by centrifugation. DNA was finally precipitated by mixing gently with 1/10 volume of 3 M sodium acetate solution, followed with 2x volume of 100% ethanol (–20°C, at least 30 min). Then the pellet was washed twice with 75% ethanol. DNA was dissolved in TE buffer and the sample concentrations were determined with a spectrophotometer (NanoDrop ND-1000, Thermo Scientific, MA).

RNA isolation and RT-PCR

RNA from leaf tissue was isolated with Trizol® Reagent (Invitrogen, CA) according to manufacturer's instruction and the concentration was determined with the aid of spectrophotometer (NanoDrop ND-1000). First-strand cDNA was synthesized using M-MLV reverse transcriptase following the manufacturer's protocol (Invitrogen), except that only 1/10 × concentration of the specified oligo$(dT)_{15}$ primer was used and the incubation time at 37°C was extended from 50 min to 2 hr. A 200 µL reaction mixture was used for 50 µg of total RNA. The reverse transcription product was purified by using the QIAquick PCR purification kit according to the manufacturer's instructions (Qiagen, CA), except that the cDNA was eluted with 50 µL Elution Buffer (EB) buffer twice in the last step. Consensus primers were designed based on sequences of the glucosinolate biosynthesis genes reported in either *A. thaliana* or *B. oleracea* to amplify the full-length corresponding gene in *B. rapa.* The rest of the primers in this study were designed directly based on the *B. rapa* BAC end sequence database (Table 2). PCR reactions were performed in a 50 µL final volume containing 200 ng of cDNA, 1 unit of high fidelity platinum *Taq* DNA polymerase (Invitrogen), 1x platinum *Taq* DNA polymerase buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 10 pmol of each primer with the following reaction conditions: 94°C for 2 min, followed by 30 cycles with 94°C for 30 s, 55°C for 30 s and 68°C for 2 min, final extension at 68°C for 30 s.

The consensus primers YB9 and YB40 (Table 2), designed for amplification of the full-length *Bra.GSELONG.a* in *B. rapa,* were based on 5' and 3' UTR region sequence of the corresponding *B. oleracea* gene *BoGSL-Elong*. Primers IPM2 and PM44 (Table 2), designed for subsequent cloning and sequencing of this gene, were based on the first and the last exon sequences of *BoGSL-Elong,* respectively (Li and Quiros, 2002). The consensus primers ODD48 and ODD62 (Table 2), designed for amplification of the full-length *Bra.GSALK.a* gene in *B. rapa,* were based on 5' and 3' UTR region sequence of *BoGSL-ALK*. ODD57 and ODD62 (Table 2), designed for subsequent cloning and sequencing, were based on the first and the last exon sequences of

Table 2. Primer sequences used for DNA amplification

Primer	Sequence (5' → 3')	Gene
IND19	GAGAGAAGATGAGCGAAGAA	*Bra.SUR1.a*
IND21	GAATGGAAAGGCATGATAGA	*Bra.SUR1.a*
IND22	TCTGTTTCAACAAACAATGG	*Bra.ST5.a*
IND24	CACCAAACAACAAAAAGACC	*Bra.ST5.a*
IND12	ACAAATGGATCTCTTCTTGATT	*Bra.CYP83B1*
IND16	ATATTATTGTTCGCCACACC	*Bra.CYP83B1*
ODD48	TTCCATCATTTACTTTCTCAG	*BoGSL-Alk*
ODD57	ATGGGTGCAGACACTCCT	*BoGSL-Alk*
ODD62	TTATGCTCCAGAGACGGCA	*BoGSL-Alk*
YB2	TCATCAGTCGCAAACGGTAA	*BoGSL-Alk*
IPM2	GTGACGGTGAACAATCTCC	*BoGSL-Elong*
PM44	ATCACCACACGTTACTAATG	*BoGSL-Elong*
YB9	CCATCTCATTGCAAAACAAT	*BoGSL-Elong*
YB40	CGTGGTCACATAATTCATACC	*BoGSL-Elong*

BoGSL-ALK (Li and Quiros, 2003). Three *B. rapa* genes corresponding to the *A. thaliana* orthologs in the glycone formation pathway, *CYP83B1, SUR1* and *ST5a*, were amplified from cDNA of young leaves from several different varieties of *B. rapa*. All the primers designed for the amplification of these three genes were based on the public *B. rapa* BAC end sequence database from GenBank and located on the 5' or 3' end UTR region of the genes (Table 2).

Cloning and sequencing

The amplified fragments were resolved in 1% w/v agarose gel, eluted and purified using QIAquick gel extraction kit (Qiagen) according to the manufacturer's protocol. pGEM-T Easy vector system I (Promega, WI) was used to clone the fragments. Both the ligation reaction and transformation followed the manufacturer's protocols. After the preparation of the cloned products with QIA prep spin mini prep kit (Qiagen), the nucleic acid sequences were determined using an ABI 3730 capillary electrophoresis genetic analyzer at the UC Davis DBS (Division of Biological Sciences) DNA sequencing facility. To avoid PCR-based mutations, clones of fragments were prepared from three independent amplifications. Three clones from each amplification product were sequenced.

Sequence analysis

The sequence was viewed using ContigExpress of Vector NTI Advance 9 (Invitrogen). The coding regions of the genes were deduced based on the comparison of genomic DNA sequences and cDNA sequences of the genes. The alignments among different crops or species were performed by using AlignX of Vector NTI Advance 9 (Invitrogen). The dotplot matrices were constructed using DNADOT (http://arbl.cvmbs.colostate.edu/molkit/dnadot/).

Results

Bra.GSELONG.a, side chain elongation gene

This gene was amplified from both cDNA and genomic DNA of young leaves of turnip 'Yorii spring' (B0493) and Chinese cabbage 'Matsushima' (B0488). It has 10 exons in both *B. oleracea* and *B. rapa*. Its genomic sequences were quite different for turnip and Chinese cabbage although both crops belong to the species *B. rapa* but to different subespecies, *rapa* and *pekinensis*, respectively. The turnip allele is much longer (EF611254, 3896bp) than that of Chinese cabbage allele (EF611255, 2833bp). This difference is due to the large size difference for introns 1 and 2 and to a lesser extent for the other introns (Fig. 1a). The exonic regions and the encoded amino acid sequences of both alleles are 100% identical with each other and 98% similar with those of *B. oleracea BoGSL-Elong* (AF399834) (Tables 3 and 4). Because of the large differences in size for introns 1 and 2 in the *B. rapa* alleles, the turnip gene has higher similarity to the *BoGSL-Elong* (96% in genomic DNA) than to Chinese cabbage allele (73% in genomic DNA, Table 3). Dotplot matrices of *BoGSL-Elong* versus turnip and Chinese cabbage *Bra.GSELONG.a* genomic DNA sequences show that differences between the genes of both species are primarily due to their intronic regions (Figs. 2a and 2b).

The identity between the amino acid sequences of *Bra.GSELONG.a* alleles and the *A. thaliana* members of the *MAM* gene family, to which the gene *GS-Elong* belongs, ranged from 55% to 80% (Table 4). The *Bra.GSELONG.a* alleles showed higher similarity with *MAM1* and *MAM2* than to the other three members (*MAM3, MAML-3* and *MAML-4*).

Table 3. Identity for nucleotide genomic (g) and coding (CDS) sequences (in %) of *GS-Elong* gene and its counterparts in *B. oleracea* (AF399834) and two varieties of *B. rapa* (B0493 and B0488)

Sequences	*Bra.GSElong.a1 CDS*	*Bra.GSElong.a2 CDS*	*BoGSL-Elong gDNA*	*Bra.GSElong.a1 gDNA*	*Bra.GSElong.a2 gDNA*
BoGSL-ElongCDS	98	98	84	84	77
Bra.GSElong.a1CDS		100	83	85	77
Bra.GSElong.a2CDS			83	85	77
BoGSL-Elong gDNA				96	73
Bra.GSElong.a1 DNA					74

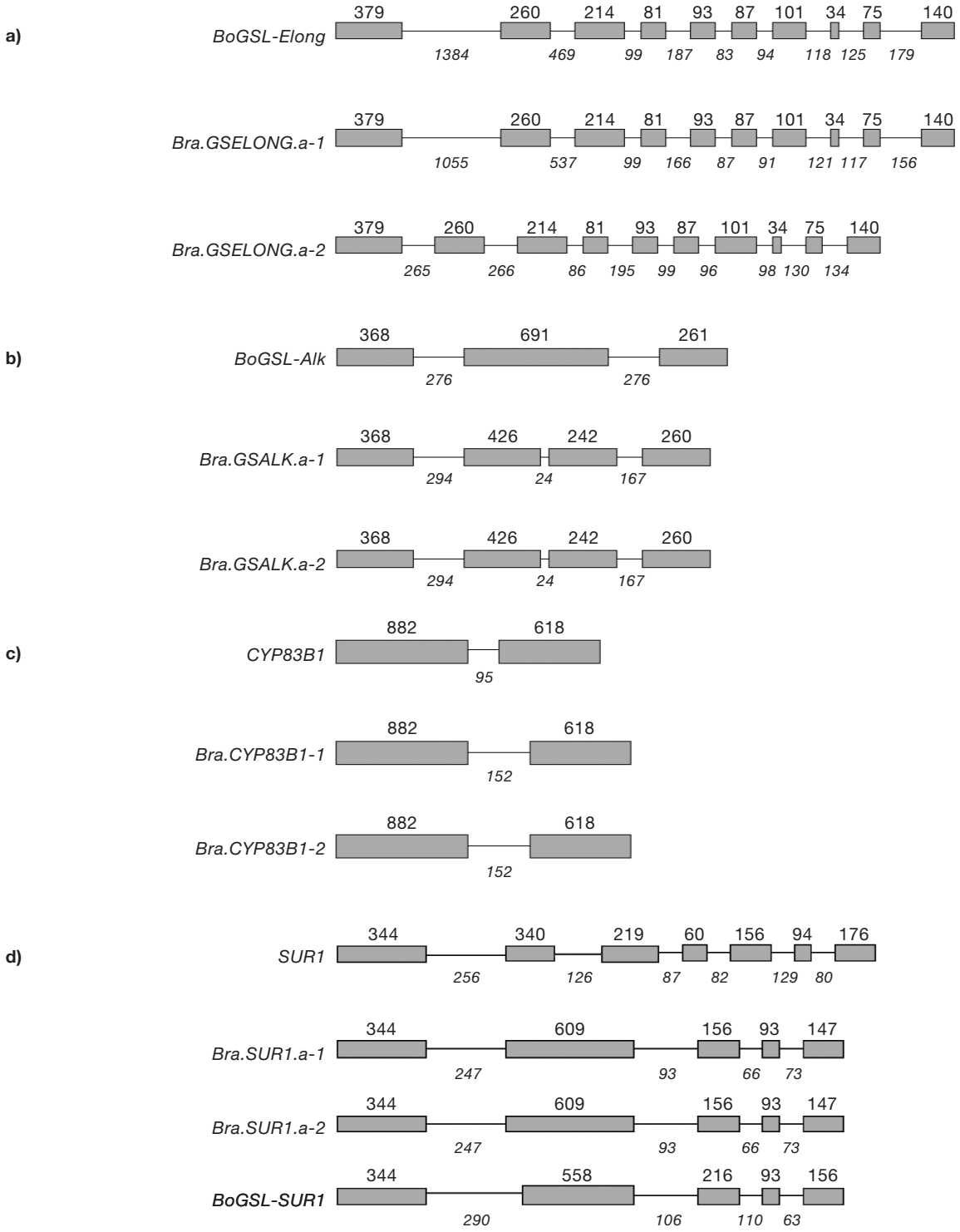

Figure 1. Gene structure diagrams comparing exon number and size among different crops and species. Exons are represented by boxes, introns are represented by lines. a) *BoGSL-Elong* and its two corresponding *B. rapa* alleles, *Bra.GSELONG.a-1* and *Bra.GSELONG.a-2* for turnip and Chinese cabbage, respectively. b) *BoGSL-Alk* and its two corresponding *B. rapa* alleles, *Bra.GSALK.a-1* and *Bra.GSALK.a-2* for turnip and Chinese cabbage, respectively. c) *CYP83B1* in *A. thaliana, B. oleracea* and its two corresponding *B. rapa* alleles *Bra.CYP83B1-1* and *Bra.CYP83B1-2* in turnip and Chinese cabbage, respectively. d) *SUR1* in *A. thaliana, BoGSL-SUR1* and its two corresponding *B. rapa* alleles *Bra.SUR1.a* in turnip and Chinese cabbage.

Table 4. Identity for amino acid sequences (in %) of *GS-Elong* gene and its counterparts in *B. oleracea, B. rapa* and *A. thaliana* (*MAM* gene family)

	Bra.GSELONGa-1	***Bra.GSELONG.a-2***	***MAM1***	***MAM2***	***MAM3***	***MAML-3***	***MAML-4***
BoGSL-Elong (AF399834)	98	98	81	80	78	56	56
Bra.GSELONG.a-1		100	80	80	77	55	56
Bra.GSELONG.a-2			80	80	77	55	56
MAM1				96	79	56	57
MAM2					76	55	56
MAM3						56	55
MAML-3							87

Glycone formation genes: *Bra.CYP83B1, Bra.SUR1.a and Bra.ST5.a*

The sequences of *B. rapa* corresponding to genes *CYP83B1*, *SUR1* and *ST5a* were amplified from both cDNA and genomic DNA of young leaves of turnip 'Yorii spring' (B0493, Table 1) and Chinese cabbage 'Matsushima' (B0488).

The coding and genomic DNA sequences of *CYP83B1* for these two crops (EF611258, EF611259) are both 100% identical (Table 5). Both *CYP83B1* from *A. thaliana* and its corresponding gene *Bra.CYP83B1* have two exons of conserved sizes and one intron, 57 bp shorter in *A. thaliana* .These sequences share 89% similarity for the genomic DNA sequences and 90% similarity for the coding sequences (Table 5).

The genomic sequences for the *Bra.SUR1.a* alleles of turnip (EF611281) and Chinese cabbage (EF611282) have 100% similarity (Table 6). There are 83% similarity between *A. thaliana SUR1* and *Bra.SUR1.a* genomic DNA sequences, and 87% similarity for their coding sequences. Also there is 81% similarity between *A. thaliana SUR1* and *BoGSL-SUR1* genomic DNA sequences, and 87% similarity for their coding sequences. For the *BoGSL-SUR1* and *Bra.SUR1.a*, there is 89% similarity in genomic DNA sequences, and 92% similarity for their coding sequences (Table 6). *A. thaliana SUR1* has seven exons, whereas *Bra.SUR1.a* and *BoGSL-SUR1* have five exons. This change seems to be due to loss of introns 2 and 3 and fusion of exons 2, 3 and 4 in *Bra.SUR1.a* and *BoGSL-SUR1* plus other minor base substitutions in the rest of introns and exons

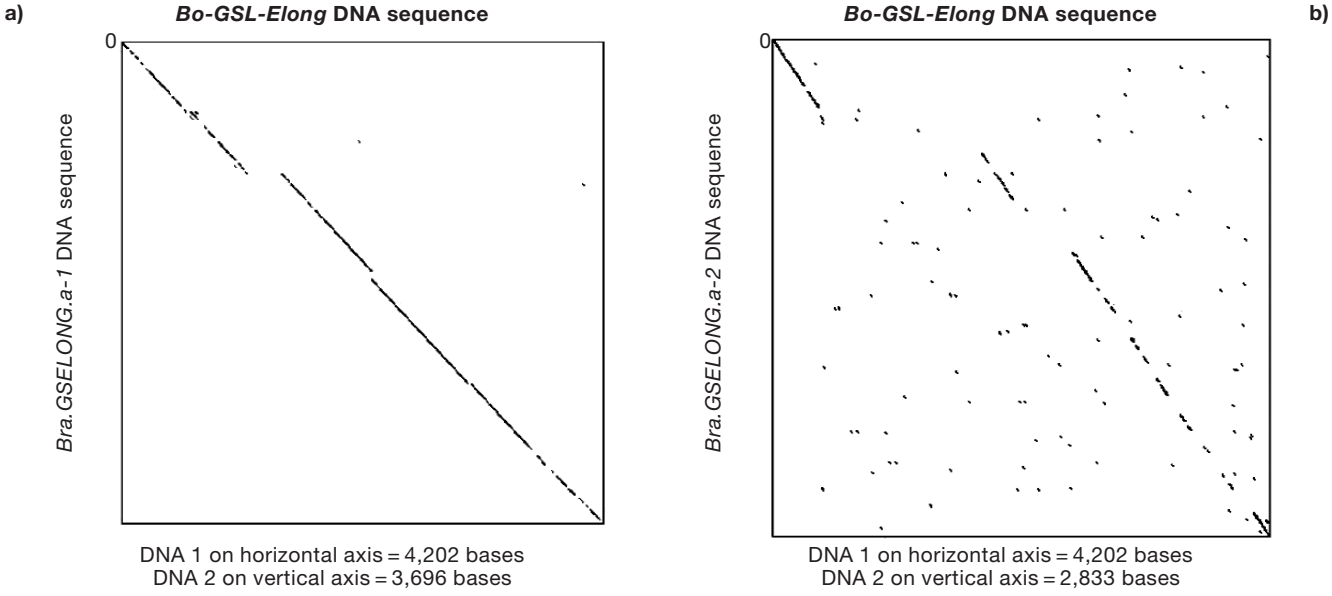

Figure 2. DotPlot Matrices for *GS-Elong* and *GS-SUR1* gene sequences. a) DotPlot Matrix of *GS-Elong* gene sequences between *B. oleracea* (*Bo-GSL-Elong*) and *B. rapa* turnip Tur, B0493 (*Bra.GSELONG.a-1*), window size 15, mismatch limit 0. b) DotPlot Matrix of *GS-Elong* gene sequences between *B. oleracea* (*Bo-GSL-Elong*) and *B. rapa* Chinese cabbage CC, B0488 (*Bra.GSELONG.a-2*), window size 0, mismatch limit 0.

Table 5. Identity of nucleotide sequences (in %) for genomic (g) and coding sequences (CDS) of *CYP83B1* and its counterparts in *A. thaliana* and two varieties of *B. rapa* (B0493 and B0488).

Sequences	*Bra.CYP83B1-1 CDS*	*Bra.CYP83B1-2 CDS*	*CYP83B1 gDNA*	*Bra.CYP83B1-1 gDNA*	*Bra.CYP83B1-2 gDNA*
CYP83B1CDS	90	90	100	90	90
Bra.CYP83B1-1CDS		100	90	100	100
Bra.CYP83B1-2CDS			90	100	100
CYP83B1gDNA				89	89
Bra.CYP83B1-1gDNA					100

Table 6. Identity of nucleotide sequences (in %) for *SUR1* and its counterparts in *A. thaliana, B. oleracea* and two varieties of *B. rapa* (B0493 and B0488)

Sequences	*BoGSL-SUR1 CDS*	*Bra.SUR1.a-1 CDS*	*Bra.SUR1.a-2 CDS*	*SUR1 gDNA*	*BoGSL-SUR1 gDNA*	*Bra.SUR1.a-1 gDNA*	*Bra.SUR1.a-2 gDNA*
SUR1 CDS	87	87	87	100	83	86	86
BoGSL-SUR1CDS		92	92	88	100	90	90
Bra.SUR1.a-1 CDS			99	88	90	99	99
Bra.SUR1.a-2 CDS				88	90	100	100
SUR1 gDNA					81	83	83
BoGSLSUR1 gDNA						89	89
Bra.SUR1.a-1 gDNA							100

(Fig. 1d). The conservation of the exonic regions of *Bra.SUR1.a* and *A. thaliana SUR1* was high. Only the first and fourth exon had the same size between the *Bra.SUR1.a* and *BoGSL-SUR1.*

A. thaliana ST5a and its *B. rapa* counterpart *Bra.ST5.a* (EF611261, EF611262) are intronless. The shared similarity between their encoding amino acid sequences is 92% (Table 7). The coding DNA sequences for *Bra.ST5.a*, were also amplified from the other six varieties of *B. rapa* (EF611263 to EF611270) listed in Table 1. Two copies were found in turnip tops 'Rapa 60 Giorni'. The longer copy, *Bra.ST5.a* (EF611269, 1020bp), has higher identity with the corresponding *A. thaliana ST5a* gene than the shorter copy, *Bra.ST5.b* (EF611270, 261bp). For all the other seven varieties *Bra.ST5.a* was cloned and sequenced. They are all identical to each other, but the *Bra.ST5.a* allele of 'Rapa 60 Giorni» has two amino acid substitutions.

Bra.GSALK.a, side chain modification gene

The *B. rapa* gene *Bra.GSALK.a*, corresponding to *A. thaliana GS-ALK* was amplified from genomic DNA

Table 7. Identity of amino acid sequences (in %) for *ST5a* gene and its counterparts in *A. thaliana* (At) and eight accessions[a] of *B. rapa*

Sequences	B0218	B0219	B0233	B0488	B0493	B1658	B1670	B1668 *ST5a*	B1668 *ST5b*
AtST5a	92	92	92	92	92	93	93	93	73
B0218		100	99	99	99	100	100	100	76
B0219			99	99	99	100	100	100	76
B0233				100	100	100	100	100	76
B0488					100	100	100	100	76
B0493						100	100	100	76
B1658							100	100	76
B1670								100	76
B1668 ST5a									76

[a] Varieties corresponding to accession numbers are listed in Table 1. The two copies of this gene in B1668 are listed as *ST5a* and *ST5b*.

Table 8. Identity for nucleotide genomic (g) and coding (CDS) sequences (in %) for *GS-Alk* gene and its counterparts in *B. oleracea* and two varieties of *B. rapa* (B0493 and B0488)

Sequences	*BoGSL-Alk gDNA*	*Bra.GSALK.a-1 CDS*	*Bra.GSALK.a-1 gDNA*	*Bra.GSALK.a-2 gDNA*
BoGSL-Alk CDS (AY044425)	100	98	98	98
BoGSL-Alk gDNA		98	97	97
Bra.GSALK.a-1CDS			100	100
Bra.GSALK.a-1gDNA				100

of turnip 'Yorii spring' (B0493) and Chinese cabbage 'Matsushima' (B0488). The cDNA of this gene was only amplified from turnip young leaves.

The genomic DNA sequences of the *Bra.GSALK.a* alleles for turnip (EF611252, 1781bp) and Chinese cabbage (EF611251, 1781bp) are 100% identical, and they show high similarity (97%) with the *B. oleracea BoGSL-Alk* gene (AY044425) (Table 8). However, *BoGSL-Alk* has three exons whereas the *B. rapa* alleles have four. This is due to the splitting of exon 2 in the latter species. However, the total length of exons 2 and 3 and their intervening intron is only 1 bp longer than exon 2 in *B. oleracea* (691 bp). Additionally, the last exon is one base shorter in both alleles of *B. rapa*. The lengths of the introns are also different between the two genes (Fig. 1b) and the similarity between their encoded amino acid sequences is 97% (Table 9). The coding DNA sequence of *GS-Alk* in *A. thaliana* (780 bp) is shorter than those of both *BoGSL-Alk* and *Bra.GSALK.a*. The identity between the encoded amino acid sequences of this gene in *B. rapa* and its corresponding genes in the *A. thaliana AOP* family ranged from 36% to 64% (Table 9) *Bra.GSALK.a* showed higher amino acid sequence similarity with *AOP1* (64%) than with the other two members in *AOP* family [*AOP2 (GS-ALK)*, 45%; *AOP3 (GS-OHP)*, 36%]. However the comparison among the full-length cDNA sequences of these *A. thaliana* genes and *Bra.GSALK.a* shows that *GS-AOP2* has higher similarity (76%) with the *B. rapa* gene than *GS-AOP1* (62%) and *GS-AOP3* (36%). The reason for these differences is the presence of conserved exonic regions in 5'UTR and 3'UTR of *GS-AOP2* and *Bra.GSALK.a*.

Table 9. Identity of amino acid sequences (in %) for *GS-Alk* gene and its counterparts in *B. oleracea, B. rapa* and *A. thaliana* (AOP family)

Sequences	*Bra.GSALK.a-1*	*AOP1*	*AOP2*	*AOP3*
BoGSL-Alk	97	64	45	37
Bra.GSALK.a-1		64	45	36
AOP1			36	33
AOP2				59

Southern blot analysis suggests the existence of two to four copies of *Bra.GSALK.a* in *B. rapa* (data not shown). Therefore, other duplicate *GS-Alk* genes must be isolated before attempting to assign orthology among species, considering that these genes are also duplicated in *B. oleracea* (Li and Quiros, 2003).

Discussion

Only one copy corresponding to *BoGSL-Elong* could be amplified and cloned from *B. rapa* in our study. The sequence comparison showed that this copy has higher similarity to *MAM1,* compared to other members of *MAM* family in *A. thaliana* (Graser *et al*., 2000; Kroymann *et al*., 2001; Field *et al*., 2004; Textor *et al*., 2004, 2007; Benderoth *et al*., 2006). Therefore, it is likely that *Bra.GSELONG.a* is the ortholog for this gene, considering its high amino acid sequence similarity to *BoGSL-Elong.* Orthology between this *B. oleracea* gene and *MAM1* was previously established by Li and Quiros (2002) and Gao *et al*. (2006). It is interesting that the large size variation between the turnip and Chinese cabbage alleles for this gene is due mostly to the large deletion in intron 1 in the latter, which might have occurred during the independent domestication of these two crops (McGrath and Quiros, 1992). The Chinese cabbage accession carrying this deletion contains lower amounts of 4-carbon side chain aliphatic GSLs, which are the products of this gene, compared to the turnip accession. It is unknown whether this deletion affects gene function considering that it is in a non-translated region. On the other hand this gene is non-functional in most cauliflowers due to a splicing mutation in intron 3, resulting in a longer transcript (Li and Quiros, 2002).

Little or no variation was observed for the different alleles of the three glycone formation genes in the three species except for the exon fusion in the cultivated

species for *GS-SUR1* and the duplication observed for *Bra.ST5* in turnip tops 'Rapa 60 Giorni'. *GS-Alk* in *A. thaliana* belongs to the *AOP* gene family, corresponding to *AOP2* which is responsible for the conversion of methylsufinylalkyl to alkenyl GSL (Mithen *et al*., 1995; Kliebenstein *et al*., 2001a,b; Kroymann *et al*., 2003). The ortholog for this gene, *BoGSL-Alk,* has been reported for *B. oleracea* (Li and Quiros, 2003). Southern blot analysis in *B. rapa* indicated that, similar to *A. thaliana* and *B. oleracea*, the *Bra.GSALK.a* genes has multiple copies in this species. The high DNA and amino acid sequences similarity of both Chinese cabbage and turnip with *B. oleracea BoGSL-Alk,* suggests that this gene and *Bra.GSALK.a* are orthologs. The GSL profile of the Chinese cabbage accessions, judged by the absence of 4-methylsulphinylbutyl GSL (glucoraphanin) (Yang, 2007) indicates that the *Bra.GSALK.a* alleles sampled are functional. Therefore the structural changes resulting in the splitting of exon 2 in *B. rapa* does not significantly affect the function of this gene. In broccoli, the function of this gene is abolished by a 2 bp deletion in exon 2 causing a frameshift that reduces the product of this gene (Li and Quiros, 2003). Regarding the conservation and differences of these five genes among *B. rapa*, *B. oleracea* and *A. thaliana*, the conclusions based on the data are as follows: 1. Exonic numbers, sizes and sequences are highly conserved among the three species for the *GS-Elong, GS-Alk*, *GS-CYP83B1* and *GS-ST5a* orthologs; 2. Splice site mutations cause the different exonic numbers and sizes among the three species for the *GS-Alk* and *GS-SUR1* orthologs, but the sequences of the exonic regions are still highly conserved for them; 3. Differences in the intronic regions are the main cause of differences among these five genes from the three species, except *GS-ST5a* that lacks introns. 4. Cloning of the major genes involved in the glucosinolate biosynthesis pathway of *B. rapa* and their sequence analysis provide a useful basis for further functional analysis of these candidate genes in the future. This method could be extended to characterize other genes in other organisms which do not have extensive genome sequence information available.

Acknowledgements

The research was supported by USDA-NRI FAFS grant (2005-35301-15886 «Cloning and characterization of the major glucosinolate genes involved in the aliphatic glucosinolate pathway of Brassica crops»). We acknowledge the technical assistance of Mr. Vincent D'Antonio. We are also grateful to Dr. David Neale and Dr. Andrew Walker for critical reading of the manuscript and helpful discussion and to Mr J. Erron Haggard for proofreading it.

References

BENDEROTH M., TEXTOR S., WINDSOR A.J., MITCHELL-OLDS T., GERSHENSON J., KROYMANN J., 2006. Positive selection driving diversification in plant secondary metabolism. Proc Natl Acad Sci USA 103, 9118-9123.

CHEN S., GLAWISCHNIG E., JORGENSEN K., NAUR P., JORGENSEN B., OLSEN C.E., HANSEN C.H., RASMUSSEN H., PICKETT J.A., HALKIER B.A., 2003. CYP79F1 and CYP79F2 have aliphatic glucosinolates in *Arabidopsis*. Plant J 33, 923-937.

FIELD B., CARDON G., TRAKA M., BOTTERMAN J., VANCANNEYT G., MITHEN R., 2004. Glucosinolate and amino acid biosynthesis in *Arabidopsis*. Plant Physiol 135, 828-839.

GAO M., LI G., YANG B., MCCOMBIE W.R., QUIROS C.F., 2004. Comparative analysis of a *Brassica* BAC clone containing several major aliphatic glucosinolate genes with its corresponding *Arabidopsis* sequence. Genome 47, 666-679.

GAO M., LI G., POTTER D., MCCOMBIE W.R., QUIROS C.F., 2006. Comparative analysis of methylthioalkylmalate synthase (MAM) gene family and flanking DNA sequences in *Brassica oleracea* and *Arabidopsis thaliana*. Plant Cell Rept 25, 592-598.

GIAMOUSTARIS A., MITHEN R., 1996. Genetics of aliphatic glucosinolates. 4. Side-chain modification in *Brassica oleracea*. Theor Appl Genet 93, 1006-1010.

GRASER G., SCHNEIDER B., OLDHAM N.J., GERSHENSON J., 2000. The methionine chain elongation pathway in the biosynthesis of glucosinolates in *Eruca sativa* (Brassicaceae). Arch Biochem Biophys 378, 411-419.

HALKIER B.A., GERSHENSON J., 2006. Biology and biochemistry of glucosinolates. Ann Rev Plant Biol 57, 303-333.

HALL C., McCALLUM D., PRESCOTT A., MITHEN R., 2001. Biochemical genetics of glucosinolate modification in *Arabidopsis* and *Brassica*. Theor Appl Genet 102, 369-374.

HANSEN C.H., WITTSTOCK U., OLSEN C.E., HICK A.J., PICKETT J.A., HALKIER B.A., 2001. Cytochrome P450 CYP79F1 from *Arabidopsis* catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. J Biol Chem 276, 11078-11085.

HANSEN C.H., DU J.C., NAUR P., OLSEN C.E., AXELSEN K.B., HICK A.J., PICKETT J.A., HALKIER B.A., 2003. CYP83B1 is the oxime-metabolizing enzyme in the glucosinolate pathway in *Arbidopsis*. J Biol Chem 276, 24790-24796.

KLIEBENSTEIN D.J., GERSHENSON J., MITCHELL-OLDS T., 2001a. Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics 159, 359-370.

KLIEBENSTEIN D.J., LAMBRIX V.M., REICHELT M., GERSHENSON J., MITCHELL-OLDS T., 2001b. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell 13, 681-693.

KROYMANN J., TEXTOR S., TOKUHISA J.G., FALK K.L., BARTRAM S., GERSHENSON J., MITCHELL-OLDS T., 2001. A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127, 1077-1088.

KROYMANN J., DONNERHACKE S., SCHNABELRAUCH D., MITCHELL-OLDS T., 2003. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc Natl Acad Sci USA 100, 14587-14592.

LI G., QUIROS C.F., 2002. Genetic analysis, expression and molecular characterization of *BoGSL-ELONG*, a major gene involved in the aliphatic glucosinolate pathway of *Brassica* species. Genetics 162, 1937-1943.

LI G., QUIROS C.F., 2003. In planta side-chain glucosinolate modification in *Arabidopsis* by introduction of dioxygenase *Brassica* homolog *BoGSL-ALK*. Theor Appl Genet 106, 1116-1121.

LI G., RIAZ A., GOYAL S., ABEL S., QUIROS C.F., 2001. Inheritance of three major genes involved in the synthesis of aliphatic glucosinolates in *Brassica oleracea*. J Amer Soc Hort Sci 126, 427-431.

McGRATH J.M., QUIROS C.F., 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theor Appl Genet 83,783-790.

MITHEN R., CLARKE J., LISTER C., DEAN C., 1995. Genetics of aliphatic glucosinolates. 3. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. Heredity 74, 210-215.

MIKKELSEN M.D., NAUR P., HALKIER B.A., 2004. *Arabidopsis* mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. Plant J 3, 770-777.

NAUR P., PETERSEN B.L., MIKKELSEN M.D., BAK S., RASMUSSEN H., OLSEN C.E., HALKIER B.A., 2003. CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in *Arabidopsis*. Plant Physiol 133, 63-72.

ØSTERGAARD L., KING G.J., 2008. Standardized gene nomenclature for the Brassica genus. Plant Methods 4, 10.

PARKIN I., MAGRATH R., KEITH D., SHARPE A., MITHEN R., LYDIATE D., 1994. Genetics of aliphatic glucosinolates. II. Hydroxylation of alkenyl glucosinolates in *Brassica napus*. Heredity 72, 594-598.

PIOTROWSKI M., SCHEMENEWITZ A., LOPUKHINA A., MULLER A., JANOWITZ T., WEILER E.W., OECKING C., 2004. Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. J Biol Chem 279, 50717-50725.

REINTANZ B., LEHNEN M., REICHELT M., GERSHENSON J., KOWALCZYK M., SANDBERG G., GODDE M., UHL R., PALME K., 2001. Bus, a bushy *Arabidopsis* CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. Plant Cell 13, 351-367.

SAGHAI-MAROOF M.A., SOLIMAN K.M., JORGENSEN R.A., ALLARD R.W., 1984. Ribosomal DNA spacer length polymorphism in barley. Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci 81, 8014-8019.

TEXTOR S., BARTRAM S., KROYMANN J., FALK K.L., HICK A., PICKETT J.A., GERSHENSON J., 2004. Biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*: recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle. Planta 218,1026-1035.

TEXTOR S., DE KRAKER J.W., HAUSE B., GERSHENSON J., TOKUHISA J.G., 2007. MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in Arabidopsis. Plant Physiol 144, 60-71.

WITTSTOCK U., HALKIER B.A., 2002. Glucosinolate research in the *Arabidopsis* era. Trend Plant Sci 7, 263-270.

YANG B., 2007. Glucosinolate and comparative genomics in *Brassica rapa*. University of California, Davis, Ph. D. Dissertation.