Copyright © 2002 by the Genetics Society of America

Genetic Analysis, Expression and Molecular Characterization of *BoGSL-ELONG*, a Major Gene Involved in the Aliphatic Glucosinolate Pathway of Brassica Species

Genyi Li and Carlos F. Quiros[1]

Department of Vegetable Crops, University of California, Davis, California 95615

Manuscript received April 23, 2002
Accepted for publication September 17, 2002

ABSTRACT

We cloned a major aliphatic glucosinolate (GSL) gene, *BoGSL-ELONG* in *Brassica oleracea,* using the Arabidopsis sequence database. We based our work on an Arabidopsis candidate gene forming part of a gene family coding for isopropyl malate synthetase-like enzymes (IPMS). This gene is presumably responsible for synthesis of GSL possessing side chains consisting of four carbons (4C). The similarity of the Brassica homolog *IPMS-Bo* from broccoli to its Arabidopsis counterpart *IPMS-At* was on the order of 78%, both sharing the same number of exons. A nonfunctional allele of the *BoGSL-ELONG* gene from white cauliflower, based on the absence of 4C GSL in this crop, displayed a 30-bp deletion, which allowed us to develop a codominant marker for 4C-GSL. Gene expression analysis based on RT-PCR revealed a splicing site mutation in the white cauliflower allele. This resulted in a longer transcript containing intron 3, which failed to excise. Perfect cosegregation was observed for broccoli and cauliflower alleles at the *IPMS-Bo* gene and 4C-GSL content, strongly indicating that this gene indeed corresponds to *BoGSL-ELONG*. Cloning of two other major genes, *BoGSL-ALK* and *BoGSL-PRO*, is underway. The availability of these genes and *BoGSL-ELONG* is essential for the manipulation of the aliphatic GSL profile of *B. oleracea*.

GLUCOSINOLATES (GSLs) are secondary metabolites synthesized by many species in the order Capparales, including those in the family Brassicaceae. Isothiocyanates, which arise after GSL breakdown by hydrolytic action of the enzyme myrosinase, have diverse and important biological activities including carcinogen detoxification as well as inhibition of pathogenic fungal growth, among others (Rosa *et al.* 1997; Mithen *et al.* 2000; Mithen 2001). Aliphatic GSL derives from methionine (Underhill 1980), which is converted by three major enzymatic pathways, including (1) amino acid side-chain elongation, (2) synthesis of the glycone moiety, and (3) aglycone side-chain modification reactions (Haughn *et al.* 1991).

GSL studies in *Arabidopsis thaliana* (Mithen *et al.* 1995; Mithen and Campos 1996; Mithen 2001) and Brassica species (Magrath *et al.* 1994) provide evidence for the proposed biochemical pathway of these compounds. Genetic analysis indicates that aliphatic GSL synthesis is controlled by a genetic system with two distinct sets of genes, one set controlling side-chain elongation and the second set involved in controlling the modification of side-carbon chains. Carbon chain elongation is probably catalyzed by isopropyl malate synthases (IPMS; Campos de Quiros *et al.* 2000). Aliphatic GSL profiles vary considerably in *A. thaliana* and Brassica species. These compounds are synthesized in the order of methylsulfinylalkyl, alkenyl, and hydroxy types and can be grouped by the size of their side chains, which is determined by the number of carbons in those chains. In *Brassica oleracea*, the model for biosynthesis of aliphatic GSL and the genes acting in the main steps of this process are shown in Figure 1. In this model, the presence of the dominant allele for the *BoGSL-ELONG* gene will result in four-carbon (4C) GSL, whereas the presence of the dominant allele for *BoGSL-PRO* will result in three-carbon (3C) GSL. This expectation is supported by studies of the inheritance of 3C side-chain and 4C side-chain GSL in segregating populations of *B. oleracea*. Plants carrying both dominant alleles were found to produce both 3C and 4C GSL, whereas plants carrying the null alleles at both loci display only traces, if any, of aliphatic GSL (Li *et al.* 2001).

In Arabidopsis, the *GS-ELONG* locus was mapped on chromosome V (Campos de Quiros *et al.* 2000). There were two duplicated and contiguous isopropylmalate synthase-like genes (*IPMS-At1* and *IPMS-At2*), located on two flanking bacterial artificial chromosome (BAC) clones, T20O9 and MYJ24, respectively. These genes were identified as candidate genes of *GS-ELONG* (Campos de Quiros *et al.* 2000; Kroymann *et al.* 2001). We report in this article the identification of the Brassica homolog for this gene, *BoGSL-ELONG*, determining the presence of 4C GSL in *B. oleracea*. We also sequenced this gene and developed molecular markers that can be used to

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession no. AF399834.

[1]*Corresponding author:* Department of Vegetable Crops, University of California, Davis, CA 95616. E-mail: cfquiros@ucdavis.edu

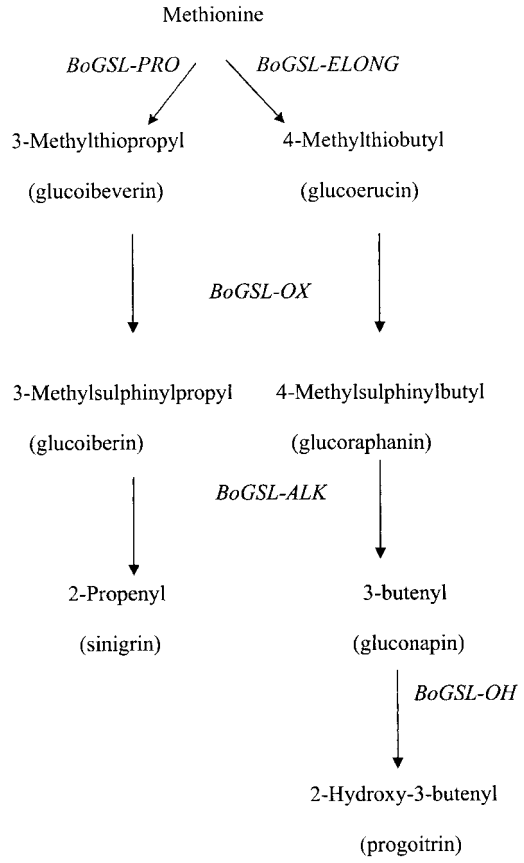

Figure 1.—Model for aliphatic glucosinolate biosynthesis in *B. oleracea*, including the inferred major genes controlling this process (Mithen *et al.* 1995; Li *et al.* 2001).

do marker-based selection to increase glucoraphanin content, an important 4C GSL, in *B. oleracea* crops, such as cauliflower, cabbage, and broccoli. This and two other genes, *BoGSL-ALK* and *BoGSL-PRO*, are essential for the manipulation of the aliphatic GSL profile of *B. oleracea*.

MATERIALS AND METHODS

Plant materials: Twenty-one commercial varieties of three different *B. oleracea* crops, broccoli, cauliflower, and collard, and four doubled-haploid (DH) lines of broccoli and cauliflower were included in this study (Table 1).

DNA amplification and library screening: Using the sequence of the *GS-ELONG* candidate genes from Arabidopsis [*IPMS-At1*(At5g23010), *IPMS-At2* (At5g23020)], a pair of primers (IPM1, 5′-GCCATCTTCGCACCCAAA-3′ and IPM2, 5′-GTGAC GGTGAACAATCTCCT-3′) was designed to amplify the corresponding region of the *B. oleracea* homolog. These primers were designed to amplify part of exon 1, exon 2, and the intervening intron between these two exons. For this purpose we used broccoli genomic DNA, extracted as reported by Li and Quiros (2001). The PCR conditions for amplification for 35 cycles were: 94°, 1 min; 56°, 1 min; 72°, 2 min. The resulting amplified DNA was confirmed to correspond to the *IPMS* genes by sequencing.

Primers IPM1 and IPM2 were then used to screen a BAC library constructed with the broccoli doubled-haploid line "Early Big-10" (Quiros *et al.* 2001) for clones harboring IPMS Brassica homologs. Two rounds of PCR were used for the library screening by 3-D pooling of the clones following the strategy of Koes *et al.* (1995).

Partial BAC sequencing was done using the SRAP protocol as described by Li and Quiros (2001). Plasmid DNA from BAC clones was prepared following the plasmid minipreparation protocol as described by Sambrook *et al.* (1989). DNA was fingerprinted using the SRAP protocol. Procedures for DNA collection from the gels and sequencing were as reported by Li and Quiros (2001).

Cosegregation analysis: An F_2 population of 450 plants, generated by crossing doubled-haploid lines from cauliflower (An-Nan-83) and broccoli (Early Big-10), was used for cosegregation analysis between 4C GSL and the *IPMS* candidate genes. This population was previously used for genetic analysis of the aliphatic GSL biosynthesis (Li *et al.* 2001). A third primer (IMP9), based on the Brassica *IPMS* homolog together with primer IPM2, was used to amplify DNA from individual plants of the F_2 population. The sequence of IPM9 was 5′-GTAGTAT TCTCAAAATCTTGT-3′. The PCR conditions were the same as those described above. The amplified products were separated using a LI-COR (Lincoln, NE) IR2 sequencer.

Gene expression analysis: We used reverse-transcription (RT)-PCR to do gene expression analysis. We designed primers located in exon 3 (5′-AAGCGATCAAAGCGGGTG-3) and exon 4 (5′-CTTCAAGCGGTGCATTCC-3′), where a splicing site change in the candidate *B. oleracea* gene IPMS *BoGLS-ELONG* occurs in white cauliflowers, as explained in the results. For RT-PCR, total RNA was prepared as described by Sambrook *et al.* (1989). Ten micrograms of RNA was used to do reverse transcription using the RT-PCR kit from Life Technologies.

Glucosinolate determination: Glucosinolate profiles in leaves were determined by high-performance liquid chromatography (HPLC) using the method described by Kraling *et al.* (1990) with some modifications. For this purpose we grind ~2 g of fresh leaves collected from 6-week-old seedlings in liquid nitrogen. Ground tissue was extracted twice with 70% methanol at 80° for 10 min. After applying the supernatant to a DEAE-Sephadex A-25 (Sigma, St. Louis) column, the glucosinolates were converted into desulfoglucosinolates with 0.5% sulfatase H-1 (Sigma) in water for 16 hr at room temperature. The desulfoglucosinolates were then eluted by adding 1.5 ml water. The resulting desulfoglucosinolates were separated by HPLC in a gradient of acetonitrile. The HPLC chromatographs were compared to the chromatograph of "Linetta," a rapeseed variety widely used as a standard for glucosinolate identification. Qualitative assessment of GSL was done visually by the presence or absence of the specific peaks. GSL content was quantified with glucotropaeolin (E.M. Science, Gibbstown, NJ) as an internal standard. Glucosinolate content was expressed as micromoles of GSL per gram of fresh leaves. We corrected the data for UV response factors for different types of glucosinolates (Wathelet *et al.* 2001).

RESULTS

Amplification of broccoli DNA with the IPMS primers produced one band displaying sequence identities of 86.3 and 85.0% with exons 1 and 2 of *IPMS-At1* and *IPMS-At2*, respectively.

In total we isolated 16 BAC clones from the broccoli library with the IPMS-designed primers. These BAC clones were divided into three putative cistronic groups according to their sequence similarity to *IPMS-At* genes

TABLE 1

List of *B. oleracea* cultivars used in this study

ID	Name	Type	Crop
B122	April	Open pollinated	White cauliflower
B130	Snow King	Open pollinated	White cauliflower
B207	White Christmas	Open pollinated	White cauliflower
B208	343 Self Blanching	Open pollinated	White cauliflower
B264	Snowball 76	Open pollinated	White cauliflower
B267	White Top	Open pollinated	White cauliflower
B272	Snow March	Open pollinated	White cauliflower
B312	Canberra	Open pollinated	White cauliflower
B1808	Snow Crown	Open pollinated	White cauliflower
B1812	White Magic	Open pollinated	White cauliflower
B1804	Guardian	F_1 hybrid	White cauliflower
B1821	Fargo—F_1 Hybrid	F_1 hybrid	White cauliflower
A1	Bai-Jiu	F_1 hybrid and DH lines	White cauliflower
A155	An-Nan Early	F_1 hybrid and DH lines	White cauliflower
B314	Cavolifiore di Sicili	Open pollinated	Purple cauliflower
B485	Violet Queen	Open pollinated	Purple cauliflower
B265	Cauliflower Purple	Open pollinated	Purple cauliflower
B10-10	Early Big	DH lines	Broccoli
B12	Li-Lu	F_1 hybrid and DH lines	Broccoli
B15	Lu-Ling	F_1 hybrid and DH lines	Broccoli
B93	Topper43-70	Open pollinated	Broccoli
B104	Georgia	Open pollinated	Collard

and to their BAC-end sequences. One of these three groups consisted of five BAC clones, B5B10, B11I7, B13D10, B19N3, and B39I16. In addition to the conserved portion of *IPMS-At* genes, all five clones had one end sequence that matched that next to the *IPMS-At* gene in Arabidopsis. Furthermore, the end sequences of B5B10, B19N3, and B39I16 were similar to Arabidopsis gene MYJ24.14, and those of B11I7 and B13D10, to gene MKD15.5. BAC clone MKD15, containing this gene, is contiguous to clone MYJ24. In total, 15 fragments of broccoli BAC B19N3 were sequenced using the SRAP protocol. After BLAST analysis, we found 1 fragment that matched Arabidosis gene MYJ24.2, which is next to *IPMS-At2*, (MYJ24.1). These results indicated that these five BAC clones contained the *IPMS-At* homolog (*IPMS-Bo*), likely matching the IPMS genes in Arabidopsis. Through direct BAC sequencing, we obtained the complete sequence of *IPMS-Bo* (GenBank accession no. AF399834). Similar to *IPMS-At1* and *IPMS-At2*, *IPMS-Bo* also contains 10 exons. Except for exons 1 and 10, all others share the same size in all three genes. At the amino acid level, *IPMS-Bo* shares 78 and 75% identity to *IPMS-At1* and *IPMS-At2*, respectively. The size of intron 1 of *IPMS-Bo* is considerably larger, being twice the size of the corresponding intron in *IPMS-At1* and four times that of *IPMS-At2*. On the basis of this analysis, *IPMS-Bo* has higher similarity to *IPMS-At1* than to *IPMS-At2* (Table 2).

To confirm that candidate gene *IPMS-Bo* corresponded to the *BoGSL-ELONG*, we amplified with primers IPM9 and IPM2 the parental lines of the segregating F_2 population resulting from crossing cauliflower and broccoli. Using these primers, we successfully developed a codominant marker, which detected a 30-bp deletion in intron 1 in cauliflower. Among 383 plants of the F_2 population, 89 plants lacked 4C GSL and all these plants were homozygous for the smaller-size cauliflower marker. All plants with 4C GSL carried at least one broccoli allele. Therefore, there was complete cosegregation between 4C GSL content and the IPMS-based marker (Figure 2).

Initially, our genetic analysis was mainly focused on the parental lines and their derived segregating population. In this preliminary survey we had observed that white cauliflower varieties did not have 4C GSL. To confirm this observation, we extended our GSL survey to the varieties and doubled-haploid lines listed in Table 1. Among the varieties, there were 15 white cauliflowers, 3 purple cauliflowers, four broccolis, and a collard (Table 1). The glucosinolate composition of this material is presented in Table 3. All white cauliflower varieties had phenotype BoGLS-ELONG-/BoGSL-PRO$^+$/BoGSL-ALK$^+$ since the alkyl GSLs they contained were exclusively 3C GSLs (glucoiberin and sinigrin or only glucoiberin). All 3 purple cauliflower varieties contained either the 4C GSL (glucoraphanin) or both 4C and 3C GSL (glucoiberin and glucoraphanin). Accessions B314 and B485 containing 3C and 4C GSL had phenotype BoGSL-ELONG$^+$/GSL-PRO$^+$/GSL-ALK$^-$ whereas B265 had phenotype BoGSL-ELONG$^+$/BoGSL-PRO$^-$/BoGSL-ALK$^-$. On the other hand, the broccoli varieties had exclusively

TABLE 2

Exon and intron sizes (base pairs) of two *IPMS-At* genes of Arabidopsis and homologous alleles (*IPMS-Bo*) from broccoli and cauliflower

Homolog	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7	Exon 8	Exon 9	Exon 10
IPMS-Bo (broccoli)	378	260	214	81	93	87	101	34	75	141
IPMS-Bo (cauliflower)	378	260	404(214 + 109 + 81)[a]		93	87	101	34	75	141
IPMS-At1	444	260	214	81	93	87	101	34	75	132
IPMS-At2	444	260	214	81	93	87	101	34	75	123

Homolog	Intron 1	Intron 2	Intron 3	Intron 4	Intron 5	Intron 6	Intron 7	Intron 8	Intron 9	Total size
IPMS-Bo (broccoli)	1384	469	99	187	83	94	118	125	179	4202
IPMS-Bo (cauliflower)	1354	469	0	187	83	94	118	125	179	4182
IPMS-At1	677	501	130	90	86	204	92	102	194	3597
IPMS-At2	299	397	110	495	107	92	105	166	149	3932

[a] Fusion of exon 3, intron 3, and exon 4.

4C GSL, glucoraphanin, that is, phenotype BoGSL-ELONG$^+$/GSL-PRO^{-+}/GSL-ALK$^-$, whereas in the collard variety, sinigrin (3C) and progoitrin (4C) were the predominant GSLs (>90% of total aliphatic GSL). Therefore the phenotype of this crop was BoGSL-ELONG$^+$/BoGSL-PRO$^-$/BoGSL-ALK$^+$.

GSL analysis was also performed in two white cauliflower doubled-haploid line populations, in two broccoli doubled-haploid line populations, and in their original F_1 hybrid parental varieties. The same results were obtained as those described above, where white cauliflower had only 3C glucosinolates and broccoli 4C glucosinolates. Noteworthy is the fact that sinigrin segregated among the cauliflower doubled-haploid lines, indicating that the desaturation gene *BoGSL-ALK* was heterozygous in both F_1 hybrid parental varieties of these lines. The broccoli DH lines had only glucoraphanin, indicating that *BoGSL-ALK* was null in this material.

With the sequence of candidate gene *BoGSL-ELONG* in hand, we proceeded to determine whether this gene was expressed in the *B. oleracea* varieties. All white cauliflower varieties tested were considered phenotypically as BoGSL-ELONG$^-$ since they lacked 4C GSL and therefore were expected to carry the null allele for this gene. On the other hand, the broccoli varieties tested had 4C GSL, thus being BoGSL-ELONG$^+$. When we performed RT-PCR with cDNA from broccoli and white cauliflower with the primers on the basis of the sequence of *BoGSL-ELONG*, we detected a polymorphism resulting in bands of two different sizes for each crop type. After sequencing both bands, we found that a mutation in the white cauliflower allele caused a splicing site change (intron 3 failed to excise), resulting in a larger-size cDNA (Table 2, Figure 3). The larger-size cDNA band cosegregated with absence of 4C GSL in the F_2 population of broccoli × cauliflower (Figure 4). This allele was present in all white cauliflower varieties and DH lines we tested.

DISCUSSION

Now that the sequence of the Arabidopsis genome is available, many genes have been annotated at a high rate of speed in this species due to the extensive input and effort from many laboratories throughout the world. To transfer this information to crops for their improvement, finding Arabidopsis homologs of genes of economic importance has become a research priority. Arabidopsis gene sequences are now being used to clone useful genes controlling important agronomic traits in crop plants. This is illustrated by our work where we successfully used the Arabidopsis candidate gene *GS-ELONG* to clone and characterize its corresponding homolog in *B. oleracea*. Through cosegregation analysis and gene expression, we further confirmed that the *IPMS-At* genes are indeed the best candidate genes for 4C GSL synthesis, although final confirmation still will have to come in the future by complementary transformation. Our strategy was based on the conservation of

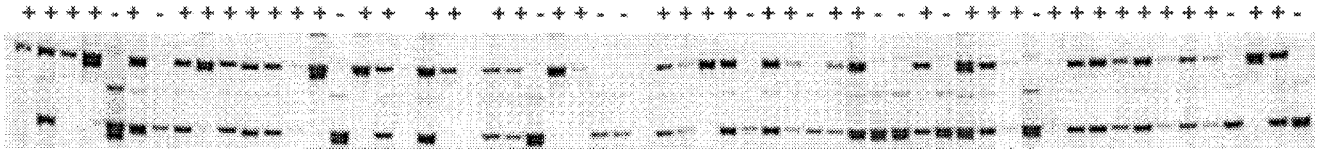

FIGURE 2.—Broccoli × cauliflower F_2 population segregating for codominant marker (intense bands): top band from broccoli and bottom band from cauliflower. Change in band size is due to 30-bp deletion in cauliflower haplotype. + indicates presence of 4C GSL, and − indicates absence of 4C GSL.

TABLE 3

Concentration of glucosinolates (micromoles per gram of fresh leaf) for the Brassica varieties and doubled-haploid lines used in this study

	Alkyl				Indolyl				
ID	GI	PR	SI	GR	HGB	GB	MGB	NGB	Total
Varieties									
B122	0.06	0	0	0.18	0.06	1.08	0.06	0.12	1.55
B130	0.06	0	0	0.09	0.03	0.97	0.07	0.23	1.44
B207	0.14	0	0	0.12	0.05	0.72	0.07	0.10	1.19
B208	0.06	0	0	0.09	0.03	0.84	0.07	0.10	1.19
B264	0.06	0	0	0.17	0.03	1.16	0.06	0.10	1.58
B267	0.08	0	0	0.08	0.01	0.60	0.04	0.11	0.92
B272	0.09	0	0	0.09	0.02	0.44	0.05	0.05	0.74
B312	0.08	0	0	0.12	0.07	0.71	0.09	0.07	1.14
B1808	0.09	0	0	0.26	0.04	0.8	0.05	0.15	1.45
B1812	0.09	0	0	0.09	0.02	0.45	0.04	0.08	0.77
B1804	0.08	0	0	0.12	0.05	0.62	0.04	0.12	1.03
B1821	0.09	0	0	0.05	0.02	0.6	0.05	0.06	0.85
B314	0.06	0	0.03	0	0.06	1.60	0.10	0.04	1.89
B485	0.05	0	0.15	0	0.04	1.10	0.10	0.03	1.47
B265	0	0	0.06	0	0.03	1.12	0.09	0.01	1.31
B10-10	0	0	0.09	0	0.03	0.68	0.08	0.10	0.98
B93	0	0	0.06	0	0	0.46	0.04	0.01	0.56
B104	0.05	0.67	0.06	1.10	0.04	0.01	0.01	0	2.03
DH lines									
A1F1	0.27	0	0	0.04	0	0.25	0.02	0	0.58
A1-75	0.16	0	0	0	0	0.09	0.02	0	0.26
A1-94	0.07	0	0	0.03	0	0.09	0.02	0	0.21
A1-104	0.12	0	0	0.02	0	0.21	0.02	0	0.37
A1-143	0.17	0	0	0.03	0	0.28	0.02	0.02	0.52
A1-262	0.11	0	0	0.05	0	0.32	0.04	0	0.52
A1-402	0.11	0	0	0	0	0.30	0.02	0	0.43
A1-393	0.04	0	0	0.02	0	0.12	0.02	0	0.20
A1-409	0.12	0	0	0	0	0.33	0.04	0.01	0.50
A1-431	0.05	0	0	0.02	0	0.16	0.02	0.01	0.25
A1-464	0.21	0	0	0.06	0	0.48	0.03	0.02	0.81

(*continued*)

gene sequence and gene order along the chromosome between Arabidopsis and Brassicas. At the nucleotide level, we (Quiros *et al.* 2001) have found ~80–90% identity levels in the coding region of both species. The linear arrangement of genes is fairly well conserved, although rearrangements are common (Quiros *et al.* 2001). Taking as a starting point the Arabidopsis candidate gene for 4C-GSL synthesis, we picked a fairly large number of clones containing sequences matching the *IPMS-At* genes. This was due to the fact that IPMS is actually a gene family consisting of four members in Arabidopsis (Campos de Quiros *et al.* 2000; Kroymann *et al.* 2001). In spite of this complication, we were able to identify the clones harboring the *GS-ELONG* homolog. After sequencing the Brassica homolog *IPMS-Bo*, we successfully developed a codominant marker, which made it easier to do cosegregation analysis. The complete linkage of this marker with the presence of 4C GSL in the F_2 population strongly indicated that candidate gene *IPMS-Bo* is likely to correspond to *BoGSL-ELONG*. Further confirmation was provided by the cDNA marker on the basis of the splicing mutation, which cosegregated with 4C GSL content in the same F_2 population. By sequence comparison, we found that *IPMS-Bo* is more similar to *IPMS-At1* than to *IPMS-At2* at the translation and structural levels. In Arabidosis, *IPMS-At1* was identified as functional whereas *IPMS-At2* was nonfunctional in a heterologous expression system using *Escherichia coli* (Kroymann *et al.* 2001). However, complementary transformation with different constructs may provide final evidence for this assessment. For the first time, we used differential gene expression to match phenotypic differences in GSL biosynthesis. The splicing mutation of the *IPMS-Bo* allele present in all white cauliflower varieties and breeding lines that we tested associated perfectly with the absence of 4C GSL. Therefore, it should be possible to easily introduce a functional *BoGSL-ELONG* allele into white cauliflower to create new types contain-

TABLE 3

(Continued)

	Alkyl				Indolyl				
ID	GI	PR	SI	GR	HGB	GB	MGB	NGB	Total
A155F1	0.56	0	0	0.05	0	0.47	0.05	0.01	1.13
A155-7	0.32	0	0	0.08	0.01	0.36	0.05	0	0.82
A155-13	0.05	0	0	0.04	0	0.50	0.05	0	0.64
A155-18	0.36	0	0	0.13	0	0.26	0.06	0.01	0.83
A155-30	1.06	0	0	0.17	0.02	0.60	0.06	0.06	1.97
A155-38	0.96	0	0	0	0	0.22	0.04	0.01	1.23
A155-43	0.58	0	0	0	0.01	0.41	0.07	0.03	1.09
A155-49	0.26	0	0	0	0.01	0.56	0.02	0	0.81
A155-56	0.59	0	0	0	0.01	0.58	0.07	0.07	1.32
A155-78	0.20	0	0	0.08	0.01	0.56	0.08	0	0.93
B12F1	0	0	0.11	0	0	0.23	0.03	0.02	0.38
B12-3	0	0	0.06	0	0	0.17	0.01	0.05	0.30
B12-6	0	0	0.09	0	0	0.22	0.04	0.03	0.37
B12-9	0	0	0.08	0	0	0.22	0.03	0.15	0.48
B12-10	0	0	0.06	0	0	0.19	0.03	0.05	0.33
B12-15	0	0	0.06	0	0	0.5	0.04	0.03	0.60
B12-22	0	0	0.10	0	0	0.28	0.03	0.09	0.50
B12-36	0	0	0.06	0	0	0.15	0.03	0.14	0.38
B12-34	0	0	0.11	0	0	0.12	0.03	0.09	0.34
B15F1	0	0	0.06	0	0	0.10	0.02	0.01	0.20
B15-121	0	0	0.02	0	0	0.08	0.02	0.03	0.14
B15-156	0	0	0.05	0	0	0.26	0.03	0	0.34
B15-200	0	0	0.04	0	0	0.11	0.01	0.03	0.19
B15-223	0	0	0.04	0	0	0.10	0.03	0.01	0.17
B15-254	0	0	0.08	0	0	0.20	0.03	0.03	0.34
B15-272	0	0	0.04	0	0	0.13	0.04	0.02	0.22
B15-411	0	0	0.10	0	0	0.11	0.03	0.02	0.26
B15-436	0	0	0.15	0	0	0.13	0.02	0.02	0.31
B15-501	0	0	0.05	0	0	0.11	0.02	0.01	0.18

GI, Glucoiberin, 3C [3-methylsulfinylpropyl, UV response factor (RF) 1.22]; PR, progoitrin, 4C (2-hydroxy-3-butenyl, RF 1.01); GR, glucoraphanin, 4C (4-methylsulfinylbutyl, RF 0.89); SI, sinigrin, 3C (2-propenyl, RF 1.00); HGB, 4-hydroxy-glucobrassicin (hydroxy-indolyl-3-methyl, RF 0.23); GB, glucobrassicin (indolyl-3-methyl, RF 0.29); MGB, 4-methoxy-glucobrassicin (4-methoxy-indolyl-3-methyl, RF 0.25); NGB, neoglucobrassicin (1-methoxy-indolyl-3-methyl, RF 0.20).

ing glucoraphanin, a 4C GSL, which releases an isothiocyanate possessing the ability to upregulate carcinogen detoxification enzymes in mammalian cells (Mithen 2001). The purple cauliflower varieties, which are classified as cauliflowers, are actually intermediate types more closely resembling broccoli in inflorescence type (Smith and King 2000). Therefore, it was not surprising to find that they contain glucoraphanin, which is a 4C GSL typical of broccoli (Kushad *et al.* 1999). Smith and King (2000) provided a historic scenario on the origin of broccoli and cauliflower, postulating that cauliflower derived from broccoli via Sicilian purple cauliflower is an intermediate in a two-step process. This conclusion was reached according to inflorescence morphology and presence of specific alleles determining this trait, such as those at gene *BoCAL-a* and *BoAP1* loci. The broccoli-

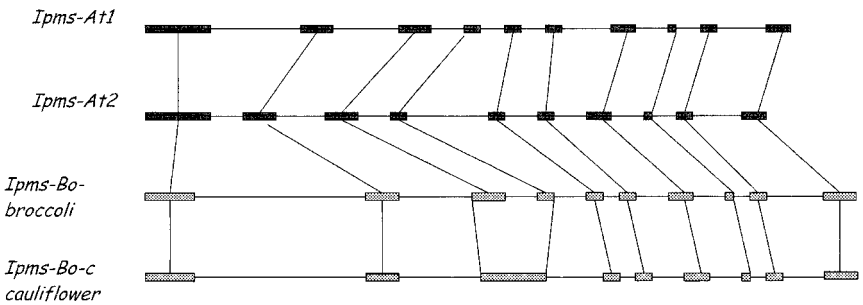

Figure 3.—Diagrammatic representation at approximate scale of *IPMS* genes in Arabidopsis and the two Brassica alleles homologous to these genes. Boxes represent exons. In the cauliflower allele exons 3 and 4 are fused due to a splicing mutation (diagram based on Kroymann *et al.* 2001).

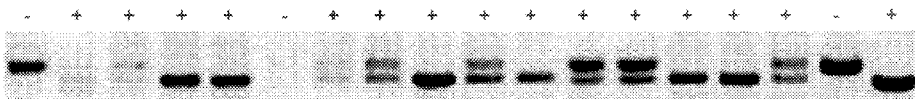

FIGURE 4.—Sample of segregation in a broccoli × cauliflower F_2 population for cDNA bands of two sizes resulting from the splicing site mutation illustrated in Figure 3. The larger band (top) corresponds to the mutant allele from white cauliflower. + indicates presence of 4C GSL, and − indicates absence of 4C GSL.

to-cauliflower inflorescence transformation, discounting color, occurred by two successive mutations in these genes, the first originating purple cauliflower and the second white cauliflower. Putting this in the GSL gene context, most broccoli with phenotype BoGSL-ELONG$^+$/BOGSL-PRO$^-$/BoGSL-ALK$^-$ will require a single gain-of-function mutation at gene *BoGSL-ALK* to produce the purple cauliflower phenotype BoGSL-ELONG$^+$/BOGSL-PRO$^-$/BoGSL-ALK$^+$. The passage from purple cauliflower to white cauliflower of phenotype BoGSL-ELONG$^-$/BoGSL-PRO$^+$/BoGSL-ALK$^+$ will require two additional mutations, a loss-of-function mutation at *BoGSL-ELONG* and a gain-of-function mutation at *BoGSL-PRO*, to result in 3C GSL synthesis. Unless there are some other varieties of purple cauliflower that we have not tested, which do not have the *BoGSL-ALK*$^+$ allele, a simpler alternative is that broccoli and white cauliflower had independent origins of domestication and purple cauliflower is the result of the hybridization of these two crops.

Cloning of the *BoGSL-ELONG* gene opens new avenues for Brassica breeding. This is the second major gene cloned in the aliphatic GSL pathway of Brassica. The chain modification gene *BoGSL-ALK* has been already cloned in this species (G. LI and C. F. QUIROS, unpublished data). With these two genes in hand, it might be possible to utilize other varieties or wild forms of *B. oleracea* with high glucoraphanin content or high level of GSL in general to maximize the content of glucoraphanin or specific GSL of interest in any Brassica crop and not only in purple cauliflower and broccoli. The use of wild species has been already explored by FAULKNER *et al.* (1997). The tools are now in place to improve Brassica crops for type and content of GSL by either marker-assisted selection or genetic transformation.

We are indebted to Vincent D'Antonio, Bo Yang, Muquiang Gao, and Pierina Benavente for technical assistance. Our thanks also to Drs. Elizabeth Earle and Peter McVetty for critical reading of the manuscript and Steffen Abel for providing access to his HPLC. This work was supported by United States Department of Agriculture Institute of Food and Agricultural Sciences grant "Development of Genomic Tools for *Brassica oleracea*."

LITERATURE CITED

CAMPOS DE QUIROS, H., R. MAGRATH, D. MCCALLUM, J. KROYMANN, D. SCHNABELRAUCH *et al.*, 2000 α-Keto acid elongation and glucosinolate biosynthesis in Arabidopsis thaliana. Theor. Appl. Genet. **101:** 429–437.

FAULKNER, K., R. MITHEN and G. WILLIAMSON, 1997 Selective increase of the potential anticarcinogenic 4-methylsinphinylbutyl glucosinolate in broccoli. Carcinogenesis **19:** 605–609.

HAUGHN, G. W., L. DAVIN, M. GIBLIN and E. W. UNDERHILL, 1991 Biochemical genetics of plant secondary metabolites in *Arabidopsis thaliana*. The glucosinolates. Plant Physiol. **97:** 217–226.

KOES, R., E. SOUER, A. VAN HOUWELINGEN, L. MUR, C. SPELT *et al.*, 1995 Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. Proc. Natl. Acad. Sci. USA **92:** 8149–8153.

KRALING, K., G. ROBBELEN, W. THIES, M. HERRMANN and M. R. AHMADI, 1990 Variation of seed glucosinolates in lines of *Brassica napus*. Plant Breed. **105:** 33–39.

KROYMANN, J., S. TEXTOR, J. G. TOKUHISA, K. L. FALK, S. BARTRAM *et al.*, 2001 A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol. **127:** 1077–1088.

KUSHAD, M. M., A. F. BROWN, A. C. KURILICH, J. A. JUVIK, B. P. KLEIN *et al.*, 1999 Variation of glucosinolates in vegetable crops of *Brassica oleracea*. J. Agric. Food Chem. **47:** 1541–1548.

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

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

MAGRATH, R., M. MORGEN, F. BANO, I. PARKIN, A. SHARPE *et al.*, 1994 Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. Heredity **72:** 290–299.

MITHEN, R., 2001 Glucosinolates—biochemistry, genetics and biological activity. Plant Growth Reg. **34:** 91–103.

MITHEN, R., and H. CAMPOS, 1996 Genetic variation of aliphatic glucosinolates in Arabidopsis thaliana and prospects for map based cloning. Entomol. Exp. Appl. **80:** 202–205.

MITHEN, R. F., J. CLARKE, C. LISTER and C. DEAN, 1995 Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. Heredity **74:** 210–215.

MITHEN, R. F., M. DEKKER, R. VERKERK, R. RABOT and I. T. JOHNSON, 2000 The nutritional significance, biosynthesis and bioavailability of glucosinolates in human foods. J. Sci. Food Agric. **80:** 967–984.

QUIROS, C. F., F. GRELLET, J. SADOWSKI, G. LI, T. SUZUKI *et al.*, 2001 Arabidopsis and Brassica comparative genomics: sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. Genetics **157:** 1321–1330.

ROSA, E. A., R. K. HEANEY, G. R. FENWICH and C. A. M. PORTAS, 1997 Glucosinolates in crop plants. Hort. Rev. **19:** 99–215.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SMITH, L. B., and G. J. KING, 2000 The distribution of BoCAL-a alleles in Brassica oleracea is consistent with a genetic model for curd development and domestication of the cauliflower. Mol. Breed. **6:** 603–613.

UNDERHILL, E. W., 1980 Glucosinolates, pp. 493–511 in *Encyclopedia of Plant Physiology (New Series)*, Vol. 8. Springer Verlag, Berlin.

WATHELET, J-P., R. IORI, N. MABON, S. CALMIERI, O. LEONI *et al.*, 2001 Determination of the response factors of several desulfoglucosinolates used for quantitative analysis of *Brassicaceae*. International Rapeseed Congress Technical Meeting, June 5–7, Poznan, Poland.

Communicating editor: C. S. GASSER