Theor Appl Genet (2015) 128:1431–1447
DOI 10.1007/s00122-015-2517-x

ORIGINAL PAPER

Genetic analysis of glucosinolate variability in broccoli florets using genome-anchored single nucleotide polymorphisms

Allan F. Brown[1] · Gad G. Yousef[1] · Robert W. Reid[2] · Kranthi K. Chebrolu[1] · Aswathy Thomas[1] · Christopher Krueger[1] · Elizabeth Jeffery[3] · Eric Jackson[4] · John A. Juvik[5]

Received: 21 November 2014 / Accepted: 11 April 2015 / Published online: 1 May 2015

Abstract

Key message **The identification of genetic factors influencing the accumulation of individual glucosinolates in broccoli florets provides novel insight into the regulation of glucosinolate levels in *Brassica* vegetables and will accelerate the development of vegetables with glucosinolate profiles tailored to promote human health.**

Abstract Quantitative trait loci analysis of glucosinolate (GSL) variability was conducted with a *B. oleracea* (broccoli) mapping population, saturated with single nucleotide polymorphism markers from a high-density array designed for rapeseed (*Brassica napus*). In 4 years of analysis, 14 QTLs were associated with the accumulation of aliphatic, indolic, or aromatic GSLs in floret tissue. The accumulation of 3-carbon aliphatic GSLs (2-propenyl and 3-methylsulfinylpropyl) was primarily associated with a single QTL on C05, but common regulation of 4-carbon aliphatic GSLs was not observed. A single locus on C09, associated with up to 40 % of the phenotypic variability of 2-hydroxy-3-butenyl GSL over multiple years, was not associated with the variability of precursor compounds. Similarly, QTLs on C02, C04, and C09 were associated with 4-methylsulfinylbutyl GSL concentration over multiple years but were not significantly associated with downstream compounds. Genome-specific SNP markers were used to identify candidate genes that co-localized to marker intervals and previously sequenced *Brassica oleracea* BAC clones containing known GSL genes (GSL-ALK, GSL-PRO, and GSL-ELONG) were aligned to the genomic sequence, providing support that at least three of our 14 QTLs likely correspond to previously identified GSL loci. The results demonstrate that previously identified loci do not fully explain GSL variation in broccoli. The identification of additional genetic factors influencing the accumulation of GSL in broccoli florets provides novel insight into the regulation of GSL levels in Brassicaceae and will accelerate development of vegetables with modified or enhanced GSL profiles.

Communicated by R. G. F. Visser.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2517-x) contains supplementary material, which is available to authorized users.

✉ Allan F. Brown
afbrown2@ncsu.edu

[1] Department of Horticultural Science, Plants for Human Health Institute, North Carolina State University, Kannapolis, NC 28081, USA

[2] Bioinformatics Services Division, University of North Carolina-Charlotte, Kannapolis, NC 28081, USA

[3] Department of Food Science and Human Nutrition, University of Illinois, Urbana, IL 61801, USA

[4] Wheat Innovation Center, General Mills Crop Biosciences, Manhattan, KS 66502, USA

[5] Department of Crop Science, University of Illinois, Urbana, IL 61801, USA

Introduction

Glucosinolates (GSLs) are a class of sulfur-containing secondary metabolites found almost exclusively in the order *Capparales* and are most commonly associated with *Brassica* vegetable, condiment, and oilseed plants. While the GSLs appear void of biological activity, the hydrolytic breakdown products of some GSLs (isothiocyanates, in particular) impact flavor, host plant resistance and mediate several positive health outcomes in mammalian

systems including inhibiting tumor initiation, promotion, and progress (Bryant et al. 2010; Davis et al. 2009; Shankar et al. 2008; Thejass and Kuttan 2006). The induction of detoxification enzymes (including glutathione-S-transferase, quinone reductase, and others) by isothiocyanates, through the Nrf2-mediated anti-oxidant response element pathway, has been well-documented in vitro and in vivo (reviewed in Zhang and Tang 2007). Evidence is also mounting that the same compounds may provide multi-faceted protection against cancer by altering endogenous cell mechanisms, including cell cycle arrest, apoptosis, histone acetylation, and mitogen-activated protein kinase signaling (Clarke et al. 2008; Jeffery and Araya 2009). Isothiocyanates differ in potency and sites of maximal bioactivity (Munday and Munday 2004), with sulforaphane (1-isothiocyanato-4-methylsulfinylbutane) derived from glucoraphanin (the principle GSL found in broccoli), the most potent (in terms of health benefits) and the most studied agent of this class.

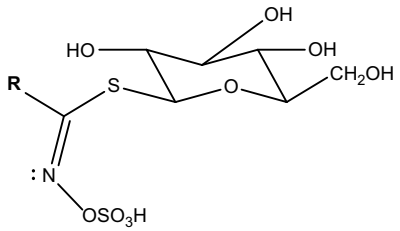
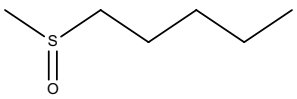
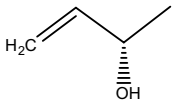
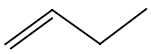
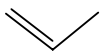
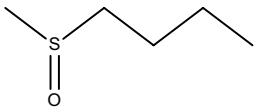
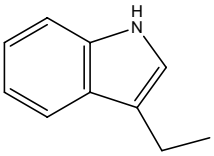
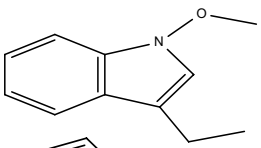
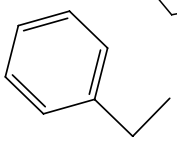
Over a 120 glucosinolates have been described, but in general fewer than a dozen are observed in any given species (Fahey et al. 2001). Among *Brassica oleracea* vegetables, these include primarily 3- and 4-carbon (3C and 4C) length aliphatic GSLs (derived from methionine), indolic GSLs (derived from tryptophan), and a single aromatic GSL (derived from phenylalanine) (Kushad et al. 1999). A list of GSLs found in *B. oleracea* vegetables, along with their structures and common names, is provided in Table 1. The biosynthesis and regulation of GSLs have been extensively studied in the related model plant *Arabidopsis* and several outstanding reviews are available (Baskar et al. 2012; Grubb and Abel 2006; Halkier and Gershenzon 2006; Sønderby et al. 2010). Briefly, GSL synthesis involves three distinct stages: the deamination and carbon-chain elongation of the precursor amino acid (methionine and phenylalanine), glucone core synthesis, and post-core modifications that can include hydroxylations, oxygenations, alkenylations, and other reactions (Fig. 1).

The GSL profile of broccoli is distinct from other vegetables of the species (cabbage, cauliflower, and others) in that glucoraphanin (4-methylsulfinylbutyl GSL) is the predominant aliphatic GSL observed (Kushad et al. 1999). A ninefold variation of glucoraphanin content (primarily genetic in nature) has been observed among broccoli accessions over multiple environments (Brown et al. 2002). Broccoli containing levels of glucoraphanin at the high end of this range have been marketed under the brand name Beneforte® and were developed through the introgression of chromosomal regions from *B. villosa* (a non-heading member of the *B. oleracea* $n = 9$ complex) (Traka et al. 2013). As this material is proprietary, it is not freely accessible for research or cultivar development. Variation in the concentration of additional aliphatic GSLs has also been observed. An evaluation of 50 broccoli accessions (Kushad et al. 1999) suggested that while most broccoli produces only trace or undetectable levels of progoitrin (2-hydroxy-3-butenyl), a significant number of accessions (~10 %) generated progoitrin in concentrations that represent half or more of the observed levels of glucoraphanin. This can be a concern, as the principle breakdown product associated with progoitrin (2-hydroxy-3-butenyl isothiocyanate) has been associated with off-flavors and under appropriate conditions can form 5-ethenyl-1,3-oxazolidine-2-thione (goitrin), an anti-nutritional compound that competes for iodine and can suppress thyroid gland function at higher concentrations (Greer and Deeney 1959).

In *Arabidopsis*, much of the structural variation of GSLs is attributable to single genes or small clusters of genes at a limited number of loci. The GSL-ELONG locus contains up to three tandem methylthioalkylmalate synthase (MAM) genes that differ in their affinity for extending short- or long-chain aliphatic GSLs and share a considerable degree of sequence similarity to the 2-isopropylmalate synthase (2-IPMS) gene family (Kroymann et al. 2001). Polymorphisms among these genes have resulted from whole or partial gene deletions, sequence exchange, and fusion between flanking genes within the cluster (Benderoth et al. 2009). Two tandem repeated MAM genes representing potential orthologs to GSL-ELONG were identified from the sequenced *B. oleracea* ('Early Big' broccoli) BAC clone 'B19N3' (Gao et al. 2005). Further sequencing of a second BAC clone ('B21F5') from the same accession identified a partial 2-IPMS gene (designated GSL-PRO) that was associated with 3C aliphatic GSLs (Gao et al. 2006).

In *Arabidopsis*, the locus GSL-AOP (or GSL-ALK/GSL-OHP) contains three tandem 2-oxoglutarate-dependent dioxygenase genes (2-ODD) designated as GSL-AOP1-3. While the function of GSL-AOP1 is currently unknown, GSL-AOP2 catalyzes the conversion of methylsulphenyl-GSL (glucoiberin and glucoraphanin) to alkenyl-GSLs (sinigrin and gluconapin). GSL-AOP3 in *Arabidopsis* is associated with the production of hydroxypropyl GSL (Kliebenstein et al. 2001), a compound not found in *Brassica* vegetables. A potential *B. oleracea* ortholog to GSL-AOP2 (GSL-ALK) has been identified from a third sequenced broccoli BAC clone ('B21H13') (Gao et al. 2004). Comparisons of GSL-ALK alleles from broccoli and collard identified a 2-bp deletion in some broccoli accessions that likely results in a non-functional enzyme due to a frame shift. It has been proposed that blocking the side chain modification pathway by introducing null alleles at this locus would lead to glucoraphanin enrichment of additional *Brassica* vegetables (Liu et al. 2012). The GSL-OH locus in *Arabidopsis*

Table 1 Glucosinolates found in *B. oleracea* vegetables: structures, common names, and molecular weight

Core structure of glucosinolate			
Class	Common name	R-side chain	MW[a]
1. Aliphatic (4C)	Glucoraphanin	4-Methylsulfinylbutyl 	357
	Progoitrin	(2R)-2-Hydroxy-3-butenyl 	309
	Gluconapin	3-Butenyl 	293
(3C)	Sinigrin	2-propenyl 	279
	Glucoiberin	3-Methylsulfinylpropyl 	399
2. Indolic	Glucobrassicin	3-Indolylmethyl 	368
	Neoglucobrassicin	N-Methoxy-3-indolylmethyl 	398
3. Aromatic	Gluconasturtiin	2-Phenethyl 	343

[a] Molecular weight for glucosinolate compounds

also contains a 2-ODD gene that encodes the enzyme responsible for the conversion of but-3-enyl (gluconapin) to 2-hydroxybut-3-enyl (progoitrin) GSL (Hansen et al. 2008). To date, a homologous gene has yet to be identified in *B. oleracea*.

In addition to key structural genes, several *Arabidopsis* regulatory genes (IQD1, SLIM1, DOF1, MYB28, MYB29, MYB76, MYB34, MYB51, and MYB122) have been identified through the use of mutation lines, transgenic over-expression, and e-QTL profiling (Celenza et al. 2005; Gigolashvili et al. 2007a, b; Hirai et al. 2007; Levy et al. 2005; Malitsky et al. 2008; Maruyama-Nakashita et al. 2006; Skirycz et al. 2006; Sønderby et al. 2007). Specificity of these regulatory genes has been observed, with some (MYB51, MYB122) associated with indolic GSL, while others are primarily associated with aliphatic GS accumulation (MYB28, MYB29, MYB76). *Arabidopsis* transgenic studies have demonstrated the up-regulation of a number of core GSL biosynthesis genes in response to these regulatory genes, but the relationship among the various transcription factors and their relative contributions to GSL accumulation remains to be clarified (Frerigmann and Gigolashvili 2014).

The objective of the present study was to utilize a *B. oleracea* sub-species *italica* (broccoli) mapping

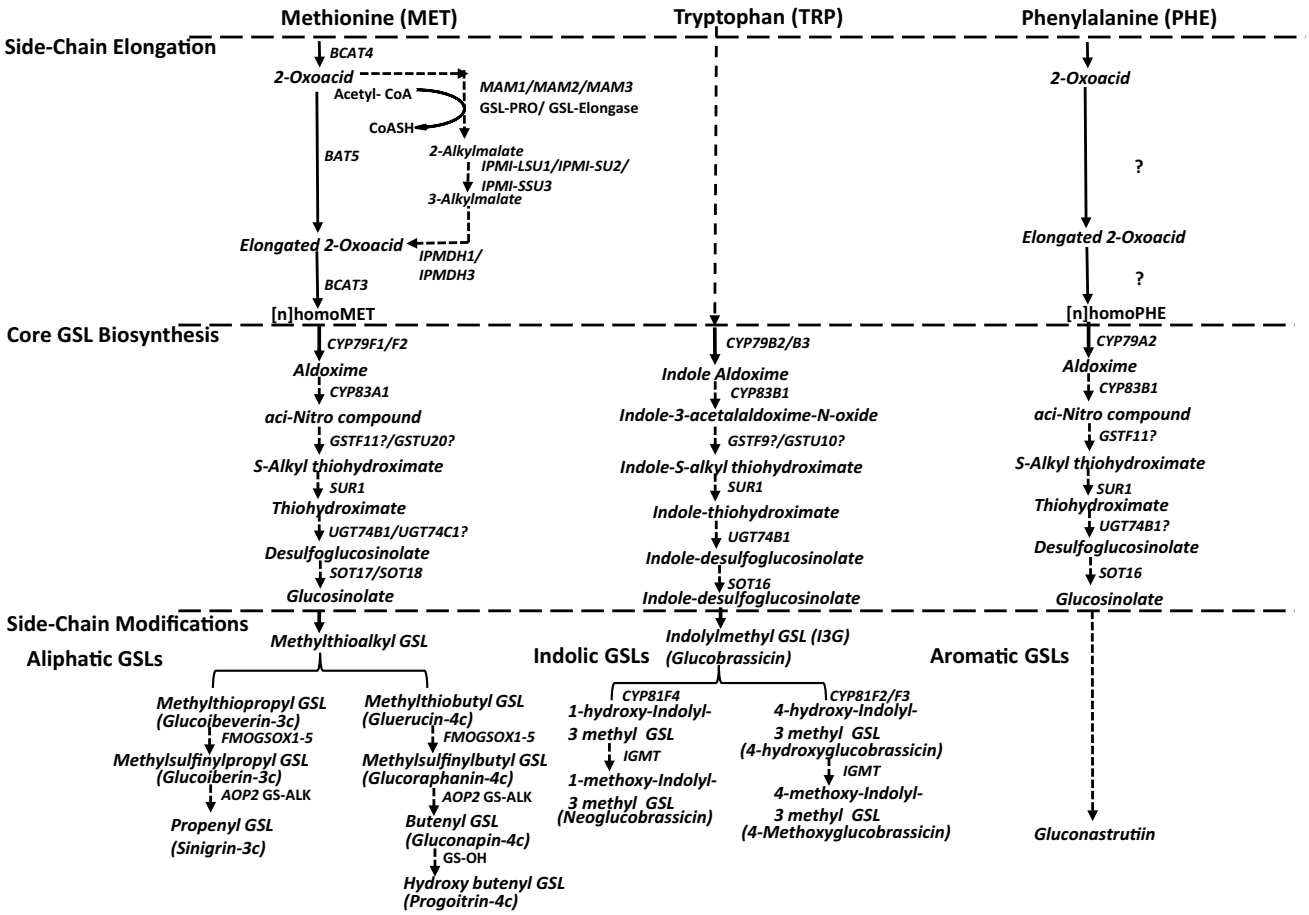


Fig. 1 Simplified scheme of glucosinolate biosynthesis in *Arabidopsis thaliana* (Baskar et al. 2012; Bender and Celenza 2009; Sønderby et al. 2010; Zandalinas et al. 2012)

population (VI-158 × BNC), recently saturated with single nucleotide polymorphism (SNP) markers from the Illumina high-density (60K iSelect format) array designed for rapeseed *(Brassica napus*, $N = 19$ AACC*)*, to identify QTL associated with qualitative and quantitative GSL variability in broccoli florets (*Brassica oleracea*, $N = 9$ CC). The parents of the population differ significantly in both the presence of specific aliphatic GSLs (progoitrin, glucoiberin, and sinigrin) and in the overall accumulation of glucoraphanin (Brown et al. 2002; Kushad et al. 1999). A second objective of the study was to utilize the linkage map associated with this population to identify putative candidate genes co-localizing with QTL. The map was constructed utilizing genome-specific markers anchored to the TO1000 *B. oleracea* reference sequence and was most recently used to identify candidate genes associated with three consistent loci impacting the accumulation of carotenoids in the same population (Brown et al. 2014). The current study includes 4 years of analysis and two distinct locations.

Materials and methods

Plant material

Two *B. oleracea* L. ssp. *italica* accessions, VI-158, a calabrese-type double haploid derived from the F_1 hybrid 'Viking' (courtesy of Dr. Mark Farnham, USDA Vegetable Lab, Charleston, SC) and 'Brocolette Neri E. Cespuglio' (BNC), a brocolette neri-type (accession PI 462209 of the USDA Plant Genetic Resource Unit Geneva, NY), were selected from a larger set of genotypes based upon their respective floret glucosinolate profiles (Brown et al. 2002; Kushad et al. 1999). A single F1 plant from a cross between the parents was bud pollinated to produce F_2 plants which were subsequently bud pollinated to produce sufficient $F_{2:3}$ family seed for genomic DNA extraction and multiple years of phenotypic and phytochemical evaluation. The phytochemical profiles of the parents, the development of population and its use in identifying genetic factors associated with both phytochemical variation and agronomic traits have been

previously described (Brown et al. 2002, 2007, 2014; Eberhardt et al. 2005; Guzman et al. 2012; Kushad et al. 1999).

Replicated $F_{2:3}$ families were grown at two locations: the University of Illinois South Farm, Urbana, Il, USA (1999 and 2000) and the Piedmont Research Station, Salisbury, NC, USA (2009 and 2010). In Illinois, approximately 24 plants of each family were sown in the greenhouse on June 10, 1999 and June 17, 2000 and transplanted into field plots at the University of Illinois South Farm after 2 weeks of acclimation on July 22, 1999 and July 26th, 2000. The soil type at Illinois was a Drummer silty clay loam. Approximately 4 weeks after transplanting, plants were side-dressed with 13N-13P-8K-17S granular fertilizer with added sulfur at a rate of 17 kg/hectare. Standard broccoli cultivation practices were followed with pesticides applied as needed. The field design was a randomized complete block with two replicates of 12 plants spaced 0.3 m apart with 0.9 m between rows. At commercial maturity, heads were harvested from each plot and packed in ice for transport to the adjacent laboratory. Samples were cut to standard size florets with equal proportions of stalk tissue and frozen in liquid nitrogen. Frozen samples were stored at −80 °C until lyophilization. Lyophilized tissue was ground to a fine powder using a coffee bean grinder. Samples were stored at −20 °C in the dark until analysis. As $F_{2:3}$ families were segregating for maturity, multiple harvests were conducted based upon uniform maximal compactness of the heads. This approach was used previously with the same population to identify consistent QTL associated with harvest date and head size (Brown et al. 2007). Proportional ground tissue from each date (weighted by the number of heads harvested on each date) was combined into a single bulked sample for analysis from each replication. In 1999, 145 $F_{2:3}$ families were evaluated but due to limited availability of seed at the time, only 87 families were replicated in 2000. In North Carolina, seedlings were transplanted to the field at Piedmont Research Station, Salisbury, NC, on Sept. 11, 2009 (136 families) and 2010 (146 families) using the same experimental design as used in Illinois. Age, condition, and spacing of the plants were the same as Illinois, but black plastic mulch and drip irrigation was added in North Carolina. Also, due to the high levels of residual sulfur present in the soil, no side-dressing was done. As in Illinois, multiple harvests were conducted at a uniform stage of compactness. No fewer than five heads were harvested from each replicate (a minimum of ten progeny total per family). Sample handling and preparation was the same at both locations.

Glucosinolate extraction

Finely ground, dry samples were added to 10 mL Oak Ridge tubes (VWR, Radnor, PA), capped, and incubated on an analog dry block heater (VWR) at 95 °C for 10 min. Next, 2 mL of extraction solvent (50 % MeOH in dd water) was added to each sample tube, re-capped, and vortexed to mix tissue with solvent. Occasional vortexing was applied during sample incubation (10 min) on the heating blocks. Samples were cooled for 5 min on ice prior to adding 500 µL of benzylglucosinolate (internal standard); samples were then vortexed, and then centrifuged at 3000 rpm for 15 min at 10 °C. The supernatant was poured off into a new glass tube and saved on ice. The pellet was re-extracted using the same procedure without the addition of the internal standard and supernatants of the same sample were pooled into corresponding glass tubes. From the well-mixed supernatant, 1 mL was pipetted into a 2 mL Eppendorf tube (FisherScientific, Pittsburg, PA) and combined with 150 µL of 0.5 M lead and barium acetate solution, vortexed, then centrifuged for 3 min at 2000 rpm to allow proteins to precipitate. The supernatant from each Eppendorf tube was poured to a drained poly-prep chromatography column (Bio-Rad, Hercules, CA) containing precharged DEAE Sephadex A-25 (Sigma Chem., St. Louis). Once solution passed through the column, 3 mL of 0.02 M pyridine acetate was added, followed by 3 mL of deionized water. Glucosinolates were desulfated using S3009 sulfatase enzyme (Sigma, St. Louis, MO). To each poly-prep column, 10 units of sulfatase suspended in 500 µL deionized water were added and the columns were capped for 18 h. Desulfated GSLs were eluted from the poly-prep column by adding 3 mL of deionized water, filtered using 0.2-µm nylon syringe filters (Thermo Scientific, Rockwood, TN), and used for HPLC quantification analyses. The internal standard for GSLs (benzyl GSL) was purchased from POS Pilot Plant Crop., Saskatchewan, Canada, and was used to calculate the individual compound concentrations in floret dry tissues. Individual GSL concentrations (µmol/g floret dry tissue) were calculated in comparison to certified GSL levels in a standard rapeseed reference material (BCR 367, Commission of the European Community Bureau of References, Brussels, Belgium).

Quantification of glucosinolates

In Illinois, GSLs were analyzed with a Dionex DX500 HPLC system (Thermo Scientific, Sunnyvale, CA) consisting of a variable UV detector set at a maximum absorption wavelength of 229 nm. Injecting samples into HPLC was achieved using an AS3500 autosampler (4 °C). The separation of GSLs was achieved using LiChrospher 100 Reversed-phase C18 column, 250 mm × 4.6 mm × 5 µm (Grace Davison Discovery Science, Deerfield, IL). The mobile phase was composed of solvent A (0.1 % ammonium acetate in H2O with 0.1 % acetonitrile) and solvent B (100 % acetonitrile). The gradient system was 0, 25, 0, and

0 % of solvent B at 0, 32, 34, 36, and 40 min, respectively, with a constant flow rate of 1 mL/min.

In North Carolina, the GSL quantification was conducted using a 1200 HPLC system attached to a 6510 Q-TOF (Agilent Technologies, Santa Clara, CA). The separation of compounds was also achieved using LiChrospher 100 Reversed-phase C18 column, 250 mm × 4.6 mm × 5 µm (Grace Davison Discovery Science, Deerfield, IL). The system contained a controlled temperature autosampler (4 °C) and column compartment (20 °C). The same solvent gradient system and flow rate were applied as above. A subset of samples from North Carolina were sent to Illinois to determine if differences in analysis platforms substantially affected the quantification of GSLs and the results were comparable (data not shown).

Phenotypic data analysis

Statistical analyses were conducted with SAS software (version 9.2 for Windows; SAS Institute, Cary, NC). Means, standard deviations, and ranges were generated with Proc Univariate. Analysis of variance (ANOVA) was conducted for all traits with all factors (genotype, years, replication, genotype × year) considered random using Proc GLM procedures. The linear random model used was $y_{ijkl} = \mu + G_i + L_j + Y(L)_{jk} + R(Y)_{kl} + G \times L_{ij} + G \times L \times Y_{ijk} + \varepsilon_{(ijkl)}$, where y = response from the experimental unit, μ = overall mean, G = genotype (family), L = location, Y = year, R = replication (block), $Y(L)$ = year within location, $R(Y)$ = block within year, $G \times L$ = genotype × location interaction, $G \times L \times Y$ = genotype × location × year interaction, and ε = experimental random error. Pearson correlation coefficients (r) were generated among all compound combinations, average harvest date and average head size, using the Proc Corr. statement.

Identification of QTL associated with glucosinolates

The construction of the genetic linkage map has been previously described (Brown et al. 2014). Briefly, the map was constructed using *Brassica napus* (AACC) single nucleotide markers (Illumina *Brassica* Infinium array) and contains 547 markers with an average interval size of 1.7 cM (no intervals exceeding 7.5 cM). The map covers 429,265,051 bp of the 446,905,700-bp TO1000 reference assembly (96 %). SNP marker nomenclature includes a designation corresponding to the progenitor diploid genome ("A" = *B. rapa* or "C" = *B. oleracea*) followed by the position of the SNP as referenced by the 'Chiifu-401' or 'TO1000' genome sequence, respectively.

MapQTL® 6 (Van Oojen et al. 2002) was used to identify QTL associated with individual aliphatic, total indolic, and aromatic GSL in two locations [Illinois (1999 and 2000) and North Carolina (2009 and 2010)]. Analysis was conducted on average family values for individual GSLs in all 4 years. Preliminary QTL analysis for total aliphatic GSLs resulted in the identification of the same loci associated with the predominant aliphatic GSL in all environments (glucoraphanin), but the significance of the QTL and the percentage of phenotypic variance was dramatically reduced likely due to the divergent influence of the additional four aliphatic GSLs. Final analysis was conducted only on the individual aliphatic GSLs. Conversely, analysis of the individual indolic GSLs produced less significant results (at the same loci) and explained less of the variation than combining them into a single category of total indolic GSLs.

Single-factor analysis was performed using the MapQTL Kruskal–Wallis non-parametric test and the results were used to select markers for MQM model inclusion using default settings of the provided automated co-factor selection program. Several iterations were conducted to produce an optimal set of co-factors for analysis of each compound in each year. Non-restricted multiple-QTL mapping (MQM) was used with the default settings adjusted to a scan distance of 0.2 cM. Genome-wide threshold values (LOD = 4.0, $P < 0.03$) for declaring the presence of QTLs were estimated from 1000 permutations of each phenotypic trait using the programs provided bootstrapping algorithm. Confidence intervals were established using a 2 LOD drop off on either side of the maximum score. QTL from individual years was considered the same if confidence intervals overlapped and the magnitude and direction of the QTL effect was common between years.

Geneious® version 6.1.5 (Biomatters: http://www.geneious.com) was used to conduct protein to nucleotide Blast searches in all six possible reading frames of the TO1000 *B. oleracea* reference genome draft (Agriculture and Agri-Food Canada, Saskatoon, Canada, submitted) for putative candidate genes occurring within or adjacent to significant intervals. The alignment program Exonerate (Slater and Birney 2005) was used to determine an amino acid alignment percentage for the putative candidate genes and a 75 % identity match was used to declare a putative candidate from either sequence. A list of the sequences used as queries and results of these searches is provided in Supplement 1. *B. oleracea* genomic BACs containing previously described glucosinolate genes (Gao et al. 2004, 2005, 2006) were aligned to the TO1000 reference sequence using the programs Exonerate, Nucmer (Mummer software, version 3.22, (Kurtz et al. 2004) and BLAST (NCBI blast+, version 2.2.29, blastn, e-score cutoff <1e-90, word size = 78) to identify large, highly similar regions with high percent identity. Chromosomal regions that failed to align to the entire length of the BACs were filtered out.

Results

Analysis of glucosinolates

Means, standard deviations, and ranges of GSLs detected in the florets of broccoli $F_{2:3}$ families from four environments are presented in Table 2. Five aliphatic GSLs (glucoraphanin, gluconapin, progoitrin, glucoiberin, and sinigrin) were detected at appreciable levels among the $F_{2:3}$ families in 1999, 2009, and 2010. Only trace amounts of sinigrin were detected in 2000 and were not reported. Glucoraphanin was the predominant aliphatic GSL detected in all environments and concentrations ranged from a fourfold difference among families in Illinois (1999) to an eightfold differences among families in North Carolina (2010). Glucoraphanin concentrations were higher in Illinois and represented on average 66.5 % of the total aliphatic GSLs, while in North Carolina they represented 57 %.

Considerable variation in gluconapin and progoitrin was detected among families in all environments. Progoitrin accumulation occurs in the parent VI-158 (but not BNC) and the distribution of this compound among families suggests the segregation of single gene with ~25 % of families across environments accumulating trace or undetectable amounts of progoitrin. On average, progoitrin represented 13 % of total aliphatic GSLs in Illinois but almost 25 % of total aliphatic GSLs in North Carolina. Gluconapin accumulated in both parents and all families (across environments), but considerable variation was evident. The average concentration of gluconapin was 14.5 % of total aliphatic GSLs in Illinois and 16.5 % in North Carolina. The concentration of indolic GSLs was considerably higher in Illinois than in North Carolina, but the average concentration of the aromatic GSL (gluconasturtiin) was comparable between locations.

Significant genetic, location, and year effects were detected among families common to all 4 years of analysis (Table 3). Significant interaction terms were also detected. Correlations among individual aliphatic GSLs ranged from non-significant to moderate with the highest observed correlations of $r = 0.56$ (between the 3C aliphatics, glucoiberin, and sinigrin) and $r = 0.41$ (between the 3C and 4C aliphatic methylsulfinyl GSLs, glucoiberin, and glucoraphanin) (Table 4). Correlations among 4C aliphatic GSLs (glucoraphanin, gluconapin, and progoitrin) were not significant. Head size was not significantly correlated to most GSLs but moderate correlations to harvest dates were observed, which is consistent with our previous study (Kushad et al. 1999).

QTL associated with glucosinolate variability

QTL analysis identified 14 loci (designated GSL01-14) associated with GSL variability at or above the LOD

Table 2 Means, standard deviations, and range of glucosinolates in the florets of parents, F_1, and $F_{2:3}$ families in the broccoli population (VI-158 × BNC) over 4 years of analysis

Glucosinolate[a]	Parental line			$F_{2:3}$ population	
	VI-158	BNC	F_1	Mean ± SD	Range
Glucoraphanin					
IL 1999	9.5[b]	16.9	14.5	16.5 ± 6.9	6.8–29.6
IL 2000	11.9	14.9	11.7	16.9 ± 6.9	5.6–31.6
NC 2009	10.8	11.8	–	12.2 ± 4.2	3.4–28.8
NC 2010	11.3	11.7	9.6	9.6 ± 3.2	2.8–20.9
Progoitrin					
IL 1999	10.4	0.1	2.4	3.4 ± 2.9	0.0–10.4
IL 2000	8.7	0.5	3.0	3.0 ± 2.7	0.3–14.4
NC 2009	3.6	0.1	–	5.7 ± 4.8	0.0–26.1
NC 2010	3.4	0.1	1.7	3.7 ± 2.9	0.0–16.8
Gluconapin					
IL 1999	5.0	3.6	6.0	4.1 ± 1.4	2.0–8.5
IL 2000	4.8	5.4	6.8	4.3 ± 2.1	1.0–9.1
NC 2009	2.6	2.0	–	2.7 ± 0.6	1.5–5.0
NC 2010	2.7	3.2	3.6	2.9 ± 0.6	1.3–5.8
Sinigrin					
IL 1999	1.2	–	0.3	0.3 ± 0.3	0.0–1.6
IL 2000	–	–	–	–	–
NC 2009	0.4	–	–	0.3 ± 0.3	0.0–1.4
NC 2010	0.4	–	0.3	0.4 ± 0.3	0.0–1.4
Glucoiberin					
IL 1999	1.4	0.7	1.5	1.0 ± 1.0	0.0–4.8
IL 2000	1.4	0.7	1.5	1.3 ± 1.0	0.0–4.0
NC 2009	0.4	0.4	–	0.5 ± 0.6	0.0–4.5
NC 2010	0.4	0.3	0.3	0.4 ± 0.4	0.0–2.9
Total aliphatic					
IL 1999	27.6	21.3	24.7	25.0 ± 8.4	10.6–44.8
IL 2000	26.8	21.5	22.9	25.6 ± 8.4	10.5–42.0
NC 2009	17.8	14.3	–	21.4 ± 7.2	7.3–45.7
NC 2010	18.2	15.2	15.5	16.9 ± 4.9	6.5–32.2
Glucobrassicin					
IL 1999	9.4	8.2	7.0	5.0 ± 2.0	0.1–10.3
IL 2000	7.2	7.4	5.5	7.2 ± 3.5	3.1–14.9
NC 2009	6.7	4.4	–	4.3 ± 1.5	1.5–9.4
NC 2010	6.3	4.0	2.6	3.8 ± 1.5	1.0–8.7
Neoglucobrassicin					
IL 1999	1.4	1.8	1.2	2.5 ± 1.8	0.4–7.4
IL 2000	6.3	1.2	0.9	6.4 ± 5.0	1.1–14.7
NC 2009	1.3	1.0	–	1.1 ± 0.8	0.2–9.1
NC 2010	1.1	0.9	0.7	0.8 ± 0.5	0.1–3.8
Total indolic					
IL 1999	10.8	10.0	8.2	7.5 ± 2.9	1.3–17.0
IL 2000	13.4	8.6	6.4	13.6 ± 5.2	5.6–24.6
NC 2009	8.0	5.4	–	5.3 ± 1.9	1.9–14.6
NC 2010	7.5	4.9	3.3	4.6 ± 1.7	1.4–9.9

Table 2 continued

Glucosinolate[a]	Parental line			$F_{2:3}$ population	
	VI-158	BNC	F_1	Mean ± SD	Range
Gluconasturtiin					
IL 1999	0.7	1.0	0.5	1.2 ± 1.1	0.4–2.4
IL 2000	1.5	1.1	1.0	1.8 ± 1.2	0.7–3.4
NC 2009	1.7	1.5	–	1.4 ± 0.4	0.6–2.4
NC 2010	1.7	1.6	1.4	1.2 ± 0.4	0.4–2.5

[a] Glucosinolates reported for 1999 and 2000 were measured in Illinois location (IL), while for 2009 and 2010 measured in North Carolina location (NC)

[b] Glucosinolate values are expressed as μmol/g freeze dried tissue

threshold of 4.0 on seven chromosomes (Table 5; Fig. 2). Eleven of the identified QTLs were detected within the same map intervals in multiple years. Variation in 3C aliphatic GSLs (glucoiberin and sinigrin) was associated with a single QTL (GSL08) on C05 that accounted for up to 56 % of the phenotypic variation associated with glucoiberin over 4 years and 24 % of the phenotypic variation of sinigrin over 3 years. The QTL was flanked by the markers Bn-C05-07607063 and Bn-C05-09973851. BLAST searches of the associated interval in the TO1000 sequence identified multiple GSL candidates, including a putative copy of 2-IPMS (2-isopropylmalate synthase) at 8,846,276 bp. The allele enhancing the 3C aliphatics

Table 3 Analysis of variance of individual and total glucosinolates of 67 $F_{2:3}$ families of the broccoli population (BNC × VI158) evaluated in Illinois (1999 and 2000) and North Carolina (2009 and 2010)

Source	Glucoraphanin	Progoitrin	Gluconapin	Sinigrin	Glucoiberin	Total aliphatics	Glucobrassicin	Neoglucobrassicin	Total indols	Gluconasturtiin
Family	119**,[a]	51**	10**	0.2**	3.6**	271**	31**	27**	55**	1.5**
Location	13397**	62**	1001**	6.9**	121.8**	21884**	1920**	2965**	6594**	42.6**
Year (location)	1005**	246**	7**	1.8**	0.3	1983**	14	511**	1831**	1.6**
Rep (year)	28	1.0	1.0	0.0	0.3	31	33**	121**	485**	0.8**
Family × location	35**	14**	6**	0.2**	1.4**	77**	11**	20**	35**	0.6**
Family × location × year	19**	10**	7**	0.1**	0.3**	43**	12**	11	28**	0.4**
Error	11	4	1.0	0.0	0.1	18	4	8	15	0.2
R^2	92	87	93	91	93	92	88	83	86	85
CV[b]	20	22	22	28	29	16	31	31	24	26

[a] Significant at $P < 0.01$

[b] Coefficient of variation

Table 4 Pearson's correlation coefficients between individual glucosinolate measured in the $F_{2:3}$ broccoli population

	Progoitrin	Gluconapin	Sinigrin	Glucoiberin	Total aliphatics	Glucobrassicin	Neoglucobrassicin	Total indols	Gluconasturtiin
Glucoraphanin	0.192	0.181	0.32*,[a]	0.41**	0.84**	−0.13	0.22	0.08	0.37**
Progoitrin		0.29*	0.6**	0.10	0.65**	−0.20	0.06	−0.06	0.43**
Gluconapin			0.39**	0.07	0.46**	0.14	0.18	0.23	0.48**
Sinigrin				0.56**	0.66**	−0.11	0.34**	0.20	0.57**
Glucoiberin					0.46**	−0.01	0.23	0.18	0.41**
Total aliphatics						−0.15	0.24	0.10	0.59**
Glucobrassicin							0.03	0.69**	−0.09
Neoglucobrassicin								0.73**	0.56**
Total indolic									0.36**

[a] * and ** are significant at $P < 0.05$ and $P < 0.01$, respectively

(glucoiberin and sinigrin) was contributed by VI-158 and functioned in a predominately additive manner.

Unlike the 3C aliphatic GSLs, no single locus was detected in any year that was associated with more than one 4C aliphatic GSL. Three QTLs on C02, C04, and C09 (designated GSL03, GSL07, and GSL14, respectively) were associated with glucoraphanin in multiple years. GSL03 explained up to 18.5 % of the phenotypic variation in 3 years of analysis and was flanked by the markers C02-49900179 and C02-51859664. BLAST searches of the TO1000 genomic sequence identified three putative GSL candidate genes in this interval including a MYB 28-3 transcription factor (50,898,605 bp) and at least one potential MAM gene (50,470,745 bp). GSL03 was also significantly associated with the 3C aliphatic glucoiberin, in 2010 (LOD = 4.5), but fell short of the genome-wide significance rate for the same compound in 2009 and 1999 (LOD = 3.5 and 2.5, respectively). Glucoraphanin and glucoiberin concentrations were increased by the same parental allele contributed by BNC.

GSL07 on C04 explained up to 12 % of glucoraphanin variation in 3 years of analysis and was flanked by the markers C04-04379394 and Bn-A04-13116453. A BLAST search of this region did not readily identify a putative GSL candidate. The allele at this locus enhancing glucoraphanin concentration was also contributed by BNC. A third QTL (GSL14) impacting glucoraphanin concentration was identified on C09 but was only observed in the 2 years of analysis from North Carolina (2009, 2010). The interval containing this locus was flanked by the markers Bn-C9-p20106801 and Bn-C9-p39911768 and a BLAST search of this region identified a potential ortholog to *Arabidopsis* Bile acid: sodium symporter family protein (BAT5) at 29,099,176 bp. The allele from BNC at this locus was associated with lower levels of glucoraphanin.

A single QTL (GSL12) on C09 (distinct from GSL14) was significantly associated over multiple years with up to 39 % of the phenotypic variation in progoitrin. The allele resulting in increased progoitrin concentration was contributed by the parent VI-158 and functioned in an additive manner. No associations were detected between this locus and concentrations of gluconapin or glucoraphanin. A BLAST search of the TO1000 sequence representing the interval flanked by the markers Bn-A09-00100554 (essentially, the end of C09) and Bn-C09-01859507 identified four 2-ODD genes between 1,386,905 and 1,412,577 bp. Two of these genes occurred in tandem (1,408,993 and 1,412,577 bp).

Four QTLs (designated GSL01, GSL02, GSL04, and GSL06) were associated with variation in gluconapin. In 3 years of analysis, GSL04 was identified in the interval on C03 flanked by the markers Bn-C03-p05308401 and Bn-C03-p06163365 and accounted for up to 35 % of the phenotypic variation associated with gluconapin. A BLAST search of TO1000 genome using *Arabidopsis* protein sequences for *AOP2*/*AOP3* revealed a highly significant hit to an annotated 2-OGD gene in the same interval (5,961,505–5,962,801 bp) (Fig. 2; Table 5). Three QTLs (GSL05, GSL09, and GSL10) were associated with indolic GSL accumulation in 2009 and 2010. GSL09 had the largest effect explaining up to 20 % of the phenotypic variation in 2009. Alleles increasing indolic GSL concentration were contributed by both parents (GSL05 from VI-158, GSL09 and GSL10 from BNC). BLAST searches of the TO1000 genome failed to identify candidate genes in these regions with homology to previously identified genes associated with indolic GSL biosynthesis or regulation. A single QTL (GSL11), associated with the aromatic GSL gluconasturtiin in 2009 and 2010, was identified on C07 in the interval flanked by the markers Bn-C07-p42734095 and Bn-C07-p43802746. BLAST searches of this region identified a putative copy of BAT5 (similar to GSL14) in the adjacent interval at 42,549,931 bp. The allele enhancing the concentration of gluconasturtiin was contributed by VI-158.

Alignment of previously sequenced *B. oleracea* BAC clones (containing known GSL genes) to the TO1000 genomic sequence suggested the most likely position for BAC 'B21H13' (GSL-ALK) was on C09 from 1,366,584 to 1,498,031 bp, while the most likely position for BAC 'B21F5' (GSL-PRO) was on C05 from 8,765,483 to 8,850,857 bp (Fig. 2). The most likely position for BAC 'B19N3' (GSL-ELONG) in the TO1000 genome occurs on C02 from 50,444,237 to 50,538,201 bp.

Discussion

Functional analysis of natural variation has been an extremely useful tool for the identification and cloning of several genes involved in GSL biosynthesis and regulation in *Arabidopsis*. Unfortunately, most commercial broccoli is derived from a limited number of Calabrese-type cultivars with a relatively narrow genetic base, representing only a small portion of the usable variability within the *italica* gene pool (Gray 1982). Crosses between broccoli and other varietal types of *B. oleracea,* such as cabbage (*capitata*) and cauliflower (*botrytis*), have produced mapping populations that have proven useful in identifying genetic factors associated with qualitative GSL variability [e.g., determination of aliphatic side chain length and the alkenylation of aliphatic GSLs (Gao et al. 2004, 2005, 2006)], but have not provided insight into the considerable quantitative variation of glucoraphanin that is observed in the florets of different broccoli accessions (Brown et al. 2002; Kushad et al. 1999).

Table 5 QTL chromosomal map location, flanking SNP markers, variations explained by QTL, and candidate genes in broccoli genome associated with individual glucosinolates isolated for two locations; Illinois (1999 and 2010) and North Carolina (2009 and 2010)

QTL/Env[a]	Compound	Chr.[b]	LOD[c]	Pos.[d]	M1[e]	M2[f]	μ_A[g]	μ_H[h]	μ_B[i]	P%[j]	A[k]	D[l]	Putative candidate[m]
Gsl-01													
NC 2009	Gluconapin	C02	10.2	19.3	Bn-C2-p01750947	Bn-C2-p02374956	2.3	2.5	3.0	18	−0.4	−0.2	
NC 2010	Gluconapin	C02	4.3	19.3	Bn-C2-p01750947	Bn-C2-p02374956	2.8	2.2	2.5	7.3	0.1	−0.4	
Gsl-02													
NC 2010	Gluconapin	C02	5.56	33.3	Bn-C2-p03744551	Bn-C2-p04050029	2.3	3.0	2.9	9.7	−0.3	0.3	
Gsl-03													
NC 2009	Glucoraphanin	C02	8.4	77.3	Bn-C2-p49900179	Bn-C2-p51859664	9.1	12.3	14.1	18.5	−2.5	0.7	Myb28/ELONG
NC 2010	Glucoraphanin	C02	6.9	76.3	Bn-C2-p49900179	Bn-C2-p51859664	7.3	9.1	10.4	12.3	−1.6	0.3	Myb28/ELONG
IL 1999	Glucoraphanin	C02	7.5	79.0	Bn-C2-p49900179	Bn-C2-p51859664	12.2	17.5	17.4	17.8	−2.6	2.7	Myb28/ELONG
NC 2009	Glucoiberin	C02	3.5[ns]	77.3	Bn-C2-p49900179	Bn-C2-p51859664	0.5	0.7	0.78	5.1	−0.2	0.1	Myb28/ELONG
NC 2010	Glucoiberin	C02	4.4	78.0	Bn-C2-p49900179	Bn-C2-p51859664	0.4	0.5	0.72	7.2	−0.2	−0.1	Myb28/ELONG
IL 1999	Glucoiberin	C02	2.5[ns]	78.0	Bn-C2-p49900179	Bn-C2-p51859664	0.5	0.8	0.94	4.2	−0.2	0.1	Myb28/ELONG
Gsl-04													
NC 2009	Gluconapin	C03	9.8	30.5	Bn-C3-p05308401	Bn-C3-p06163365	2.3	2.8	3.1	17.2	−0.4	0.1	ALK
NC 2010	Gluconapin	C03	7.0	30.5	Bn-C3-p05308401	Bn-C3-p06163365	2.4	2.9	2.9	12.5	−0.2	0.3	ALK
IL 1999	Gluconapin	C03	4.5	30.1	Bn-C3-p05308401	Bn-C3-p06163365	4.1	4.8	5.3	9.6	−0.6	0.1	ALK
Gsl-05													
NC 2009	Indolic GSL	C04	5.5	7.7	Bn-C4-p01165349	Bn-C4-p01549566	6.1	6.3	4.9	7.4	0.6	0.8	
NC 2010	Indolic GSL	C04	5.9	4.9	Bn-C4-p00544296	Bn-C4-p01267536	6.0	5.2	4.2	7.5	0.9	0.1	
Gsl-06													
NC 2010	Gluconapin	C04	4.9	16.6	Bn-C4-p02943481	Bn-C4-p03809977	2.9	2.6	2.4	8.5	0.3	−0.1	
Gsl-07													
NC 2009	Glucoraphanin	C04	5.7	54.3	Bn-C4-p43279394	Bn-A04-13116453	9.6	12.2	14.3	12.1	−2.3	0.3	
NC 2010	Glucoraphanin	C04	4.9	55.4	Bn-C4-p45767999	Bn-A04-13116453	7.8	9.9	10.6	8.6	−1.4	0.7	
IL 1999	Glucoraphanin	C04	4.1	50.0	Bn-C4-p37595206	Bn-A04-8178053	12.4	15.7	17.2	9.1	−2.4	0.9	
Gsl-08													
NC 2009	Glucoiberin	C05	20.0	54.0	Bn-C5-p07607063	Bn-C5-p09973851	1.1	0.6	0.2	41.4	0.5	0.0	PRO
NC 2010	Glucoiberin	C05	17.5	54.0	Bn-C5-p07607063	Bn-C5-p09973851	1.0	0.5	0.2	34.9	0.4	−0.1	PRO
IL 1999	Glucoiberin	C05	18.4	53.1	Bn-C5-p07607063	Bn-C5-p09973851	1.5	0.6	0.1	43.1	0.8	−0.1	PRO
IL 2000	Glucoiberin	C05	19.1	53.1	Bn-C5-p07607063	Bn-C5-p09973851	2.6	1.4	0.3	56	1.2	−0.1	PRO
NC 2009	Sinigrin	C05	5.2	53.1	Bn-C5-p07607063	Bn-C5-p09973851	0.5	0.4	0.2	16.1	0.1	0.1	PRO
NC 2010	Sinigrin	C05	7.7	53.1	Bn-C5-p07607063	Bn-C5-p09973851	0.5	0.4	0.2	23.8	0.1	0.1	PRO
IL 1999	Sinigrin	C05	5.7	53.1	Bn-C5-p07607063	Bn-C5-p09973851	0.5	0.1	0.0	16.8	0.3	−0.1	PRO

Table 5 continued

QTL/Env[a]	Compound	Chr.[b]	LOD[c]	Pos.[d]	M1[e]	M2[f]	μ_A[g]	μ_H[h]	μ_B[i]	P%[j]	A[k]	D[l]	Putative candidate[m]
Gsl-09													
NC 2009	Indolic GSL	C06	12.7	43.1	Bn-C6-p06570259	Bn-A07-17353806	4.2	5.4	6.7	19.7	−1.3	−0.1	
NC 2010	Indolic GSL	C06	10.5	43.1	Bn-C6-p06570259	Bn-A07-17353806	4.1	5.2	6.1	14.5	−1.0	0.1	
Gsl-10													
NC 2009	Indolic GSL	C07	7.42	17.2	Bn-A06-24068524	Bn-C7-p30487360	4.8	6.0	6.2	10.3	−0.7	0.5	
NC 2010	Indolic GSL	C07	5.03	17.2	Bn-A06-24068524	Bn-C7-p30487360	4.5	4.6	5.7	6.3	−0.6	−0.6	
Gsl-11													
NC 2009	G_nasturtiin	C07	15.9	53.4	Bn-C7-p42734095	Bn-C7-p43802746	1.9	1.5	1.3	31.8	0.3	−0.1	BAT5
NC 2010	G_nasturtiin	C07	8.03	53.5	Bn-C7-p42734095	Bn-C7-p43802746	1.4	1.2	1.0	12.8	0.2	−0.1	BAT5
Gsl-12													
NC 2009	Progoitrin	C09	16.7	1.0	Bn-A09-00100554	Bn-C9-p01869607	10.3	5.6	0.9	36.6	4.7	−0.1	OH
NC 2010	Progoitrin	C09	18.8	2.0	Bn-A09-00100554	Bn-C9-p01869607	6.6	3.5	0.6	39.4	3.0	−0.1	OH
IL 1999	Progoitrin	C09	8.0	0.0	Bn-A09-00100554	Bn-C9-p01573330	3.9	2.9	0.2	26.1	1.9	0.9	OH
Gsl-13													
NC 2010	Gluconapin	C09	9.7	43.8	Bn-C9-p08945466	Bn-A09-04950180	3.0	2.3	2.2	18.3	0.4	−0.3	
Gsl-14													
NC 2009	Glucoraphanin	C09	4.2	56.6	Bn-C9-p20106801	Bn-C9-p39911768	13.5	13.2	10.3	8.7	1.6	1.3	BAT5
NC 2010	Glucoraphanin	C09	7.7	56.6	Bn-C9-p20106801	Bn-C9-p39911768	11.0	9.6	7.4	14	1.8	0.4	BAT5

ns not significant at genome-wide LOD

[a] QTL and environment (IL 1999 and IL 2000) and North Carolina location (NC 2009 and NC 2010)

[b] Chromosome number

[c] Logarithm of the odds

[d] Most likely position for QTL on chromosome

[e] Front SNP flanking marker for QTL

[f] Back SNP flanking marker for QTL

[g] Mean for homozygous QTL marker allele from parent 1 (VI-158)

[h] Mean for heterozygous QTL marker allele

[i] Mean for homozygous QTL marker allele from parent 2 (BNC)

[j] Percent of phenotypic variation explained by QTL

[k] Additive gene effect

[l] Dominance gene effect

[m] Candidate glucosinolate genes identified within marker interval

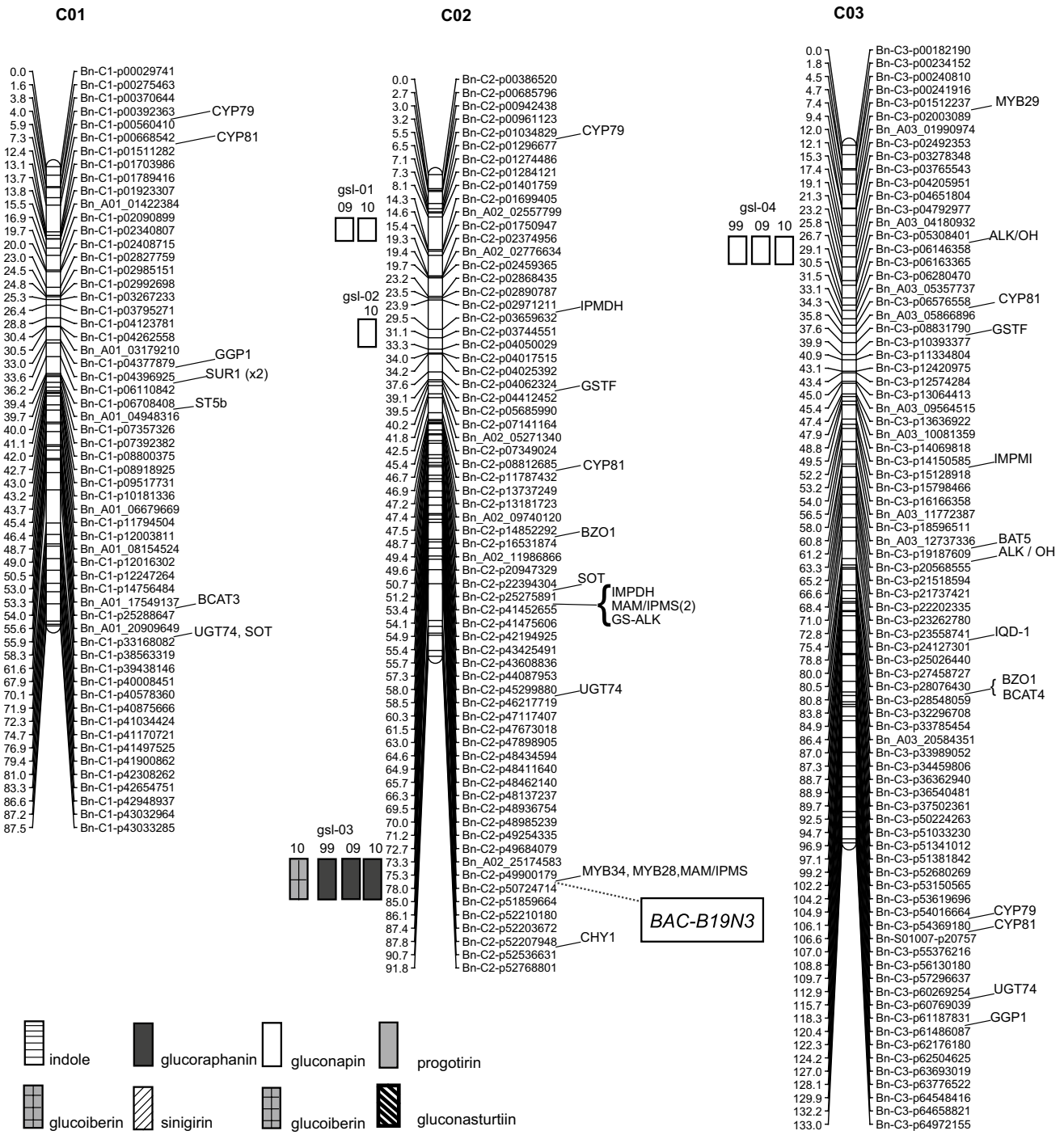


Fig. 2 Quantitative trait loci associated with glucosinolate variability in the F2:3 broccoli population (VI158 × BNC), location of BLAST results of candidate genes using known *Arabidopsis* protein sequences, and alignment of previously sequenced *B. oleracea* BAC clones to the TO1000 genomic sequence

The *B. oleracea* L. *italica* mapping population, VI-58 × BNC, was developed to study the observed qualitative and quantitative differences in GSL profiles between the parents. Unlike most broccoli accessions, VI-158 produces high levels of progoitrin and moderate amounts of glucoiberin and sinigrin. Conversely, BNC produces higher levels of glucoraphanin. Both parents produce comparable levels of gluconapin. The glucoraphanin levels of BNC reported here are somewhat misleading, as it is a landrace accession and not genetically fixed. The original BNC seed

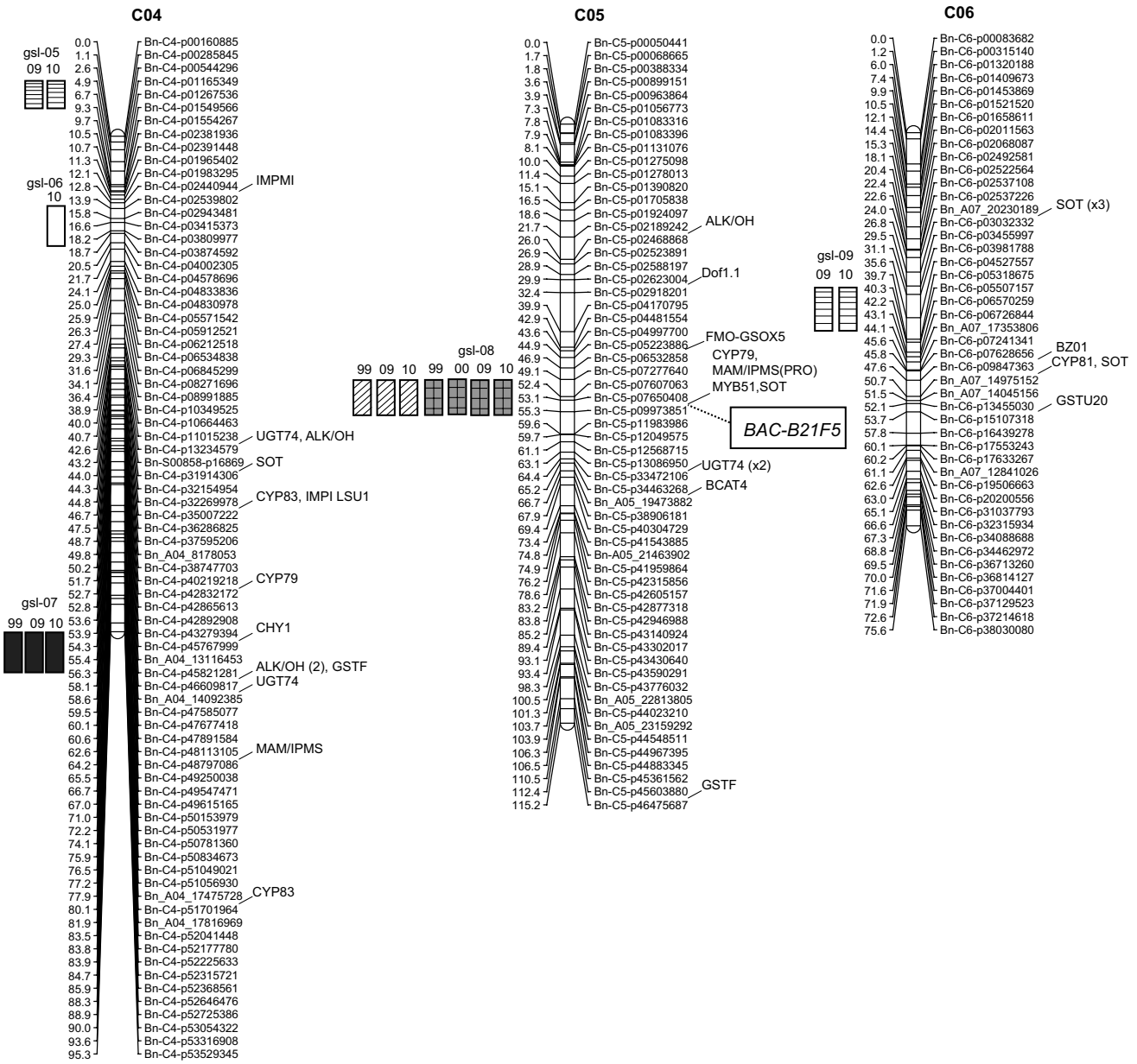


Fig. 2 continued

(obtained in limited quantities) has been self-pollinated for replicated trials and the resultant progeny show considerable variation in GSL content (data not shown). The population, however, was created by self-pollinating a single F_1 plant that fortuitously provided the principle genetic factors originating from the heterogeneous landrace.

Initial GSL analysis of this population in Illinois (1999 and 2000) suffered from a restricted number of families in 2000 (86 $F_{2:3}$ families) due to inadequate amounts of seed and a sparse genetic linkage map (Brown et al. 2007). The recent saturation of the map with SNP markers anchored to the TO1000 rapid cycling *Brassica* genomic sequence and the inclusion of two additional years of data from North Carolina has allowed for a thorough and powerful analysis of both quantitative and qualitative GSL variability in this population. Data from 2000 were included in this study but it should be noted that only one QTL (GSL08) was significant in that year and was supported by the results of the additional 3 years.

The accumulation of 3C aliphatic GSLs in this population was associated primarily with a single locus on C05 and was observed in all four environments. Our alignment of the sequenced BAC clone 'B21F5' and BLAST searches of the TO1000 genome suggested that the most likely

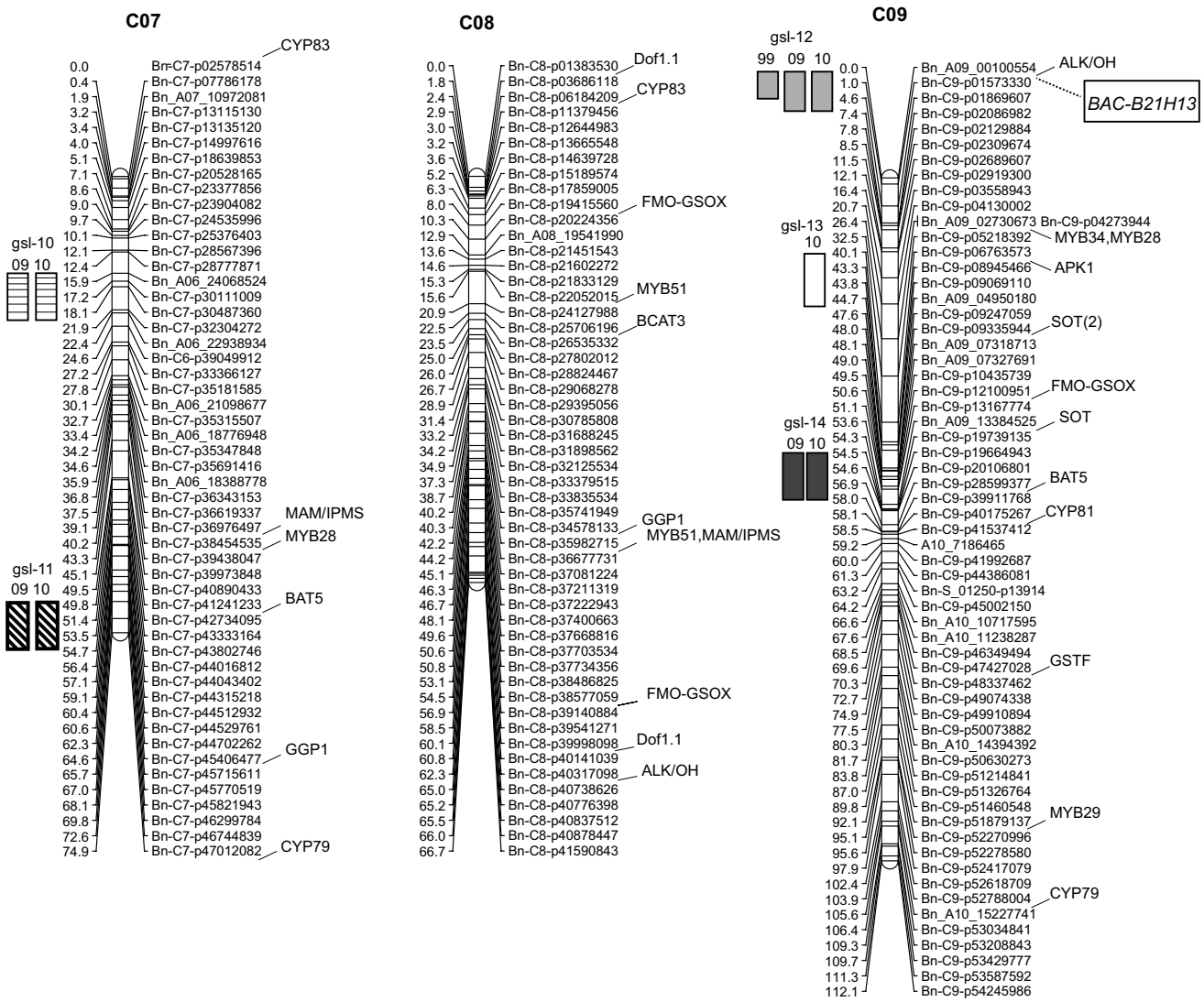


Fig. 2 continued

candidate is the previously described IPMS gene that has been designated 'GSL-PRO' (Gao et al. 2006). A moderate but significant correlation ($r = 0.56$) between glucoiberin and sinigrin across environments also provides support for the hypothesis of common regulation for these 3C aliphatic GSLs. What was perhaps most surprising, however, was the lack of corresponding evidence for common regulation of the 4C aliphatic GSLs. Despite the co-localization of previously described GSL genes (MAM, ALK, MYB-28) to multiple QTL intervals and the alignment of previously sequenced BAC clones containing known aliphatic GSL genes, the analysis did not detect a single locus that was associated with the variation of more than one 4C aliphatic GSL in any year of analysis. Correlations between glucoraphanin concentrations and downstream 4C aliphatic GSLs (gluconapin and progoitrin) were also not significant ($r = 0.19$ and 0.18, respectively). Given the size of the population we acknowledge that the analysis may not have identified all genetic factors associated with the accumulation of individual GSLs, but given the magnitude of the phenotypic variation associated with identified QTL (Table 5) and the proportion of total variation attributed to genetic variation (Table 3), these potentially unidentified QTLs likely represent relatively limited sources of variation.

It has been suggested that glucoraphanin accumulation in broccoli is associated with a non-functional GSL-ALK locus (Li and Quiros 2003) and that one strategy for enhancing glucoraphanin concentrations in *Brassica* vegetables such as cauliflower or cabbage would be to block the side chain modification pathway downstream of glucoraphanin (Liu et al. 2012). As all families in this study (and both parents) accumulated varying amounts of gluconapin, it is likely that at least one functional copy

of GSL-ALK is present in this population. The alignment of the BAC clone 'B21H13' to the distal end of C09 suggests that this is the most likely position of the previously described GSL-ALK locus (Gao et al. 2004). Interestingly, this locus was not associated with variability of gluconapin or glucoraphanin but was significantly associated with up to 39 % of the variability of progoitrin. One possible explanation for what we have observed is that this region potentially contains a functional copy of GSL-ALK (from both parents) and is also the location of GSL-OH activity in *B. oleracea*. Our BLAST searches of the TO1000 genomic sequence identified four putative 2-ODD genes in this region (<1,869,607 bp) but further work will be required to determine if the observed variability in progoitrin is attributable to the functional ALK allele from VI-158 (pleiotropy) or to one of the tandem occurring 2-OGD genes (linkage). It is also interesting to note that while gluconapin was not associated with the predicted location of ALK (BAC clone 'B21H13') on C09, it was significantly associated in 3 years with a region of C03 (5,308,401–6,163,365 bp) that includes a second putative ALK candidate in the TO1000 genomic sequence between 5,961,505 and 5,962,801 bp. Given the extensive genome duplication that has been observed in *B. oleracea* (Parkin et al. 2003), it is not unreasonable to speculate that both parents contributed functional ALK genes on C09, and that the landrace BNC contributed a second partially redundant copy of ALK on C03, not present in VI-158.

The relationship between the accumulation of progoitrin and glucoraphanin was surprising. We could find no evidence to suggest that reducing progoitrin would lead to a subsequent increase in glucoraphanin. To illustrate this further, we compared the mean glucoraphanin concentration of the 20 $F_{2:3}$ families with highest progoitrin concentration in North Carolina (mean = 10.3 µmol/g progoitrin) with the 20 families accumulating the lowest concentration of progoitrin (mean = 0.2 µmol/g progoitrin) and found that the mean glucoraphanin concentration was not significantly different between the two groups (11.8 and 12.1 µmol/g glucoraphanin, respectively) (Fig. 3).

The alignment of the BAC 'B19N3' to the TO1000 genomic sequence provides support that the interval flanking GSL03 on C02 (associated with glucoraphanin) likely corresponds to the previously identified GSL-ELONG location (Gao et al. 2005). BLAST searches of the corresponding interval of the TO1000 genomic sequence identified at least one full length MAM gene and the only MYB-28 TF that could be identified on C02. Due to the presence of this TF, it is likely that this locus also corresponds to the segment of C02 introgressed from *B. villosa* into F_1 broccoli varieties commercialized under that brand name Beneforte® (Traka et al. 2013), but the uninformative nature of the molecular markers used in the previous

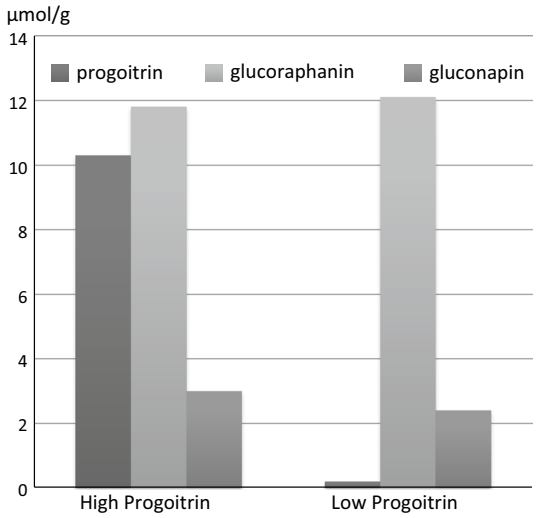


Fig. 3 Comparison of the 20 highest (mean = 10.3 µmol/g) and 20 lowest (mean = 0.2 µmol/g) progoitrin-producing families in the $F_{2:3}$ broccoli population VI158 × BNC and their respective average glucoraphanin and gluconapin concentrations

study makes comparisons inconclusive. The concentrations of glucoraphanin in Beneforte® reported by that study (20–30 µmol/g) are comparable to the concentrations we have observed in our highest accumulating families over four environments (20.9–31.6 µmol/g). MYB-28 and MAM are less than 500 kb apart in the TO1000 genome and further work will be required to validate if either (or potentially both) is responsible for the higher levels of observed glucoraphanin. It should be noted, however, that this locus is not significantly associated with gluconapin or progoitrin (which would be expected from GSL-ELONG) but is associated with the accumulation of both 3- and 4-C methylsulfinyl-GSLs (glucoraphanin and glucoiberin) which would not be expected solely as a function of side chain elongation. Additionally, it should be noted that both parents contained relatively high levels of 4-C aliphatic GSLs which suggests they both possess functional copies of GSL-ELONG.

BLAST searches of the TO1000 genomic interval containing GSL14 on C09 flanked by markers Bn-20106801 to BN-39911768 (containing the third QTL associated with glucoraphanin concentration) identified a putative ortholog to bile acid:sodium symporter family protein 5 (BAT5). Knock-out mutants of BAT5 in *Arabidopsis* have shown decreased aliphatic GSL concentrations and it has been suggested that BAT5 plays a role in translocation across the chloroplast membrane to the site of side chain elongation (Sawada et al. 2009). Interestingly, a second ortholog to BAT5 was also identified on C07 near GSL-11 which was significantly associated with variation of the aromatic GSL gluconasturtiin. Phenylalanine-derived GSLs such as

gluconasturtiin also undergo an elongation step homologous to aliphatic GSLs, but to our knowledge the enzymatic steps associated with this have yet to be elucidated. BLAST searches of the TO1000 genomic interval containing GSL07 (associated with glucoraphanin and flanked by markers Bn-C4-p43279394 and Bn-A04-13116453) did not readily reveal a GSL candidate gene, but did reveal the presence of multiple ATP-binding cassette (ABC) transporters which can function in the transport across cellular membranes. The transport of GSLs for enzymatic reactions and the eventual sequestering within plant vacuoles is not well understood in either *Arabidopsis* or *Brassica*, but could reflect potential regulatory mechanisms for maintaining levels of specific GSLs within the plant cell.

Given the health benefits associated with glucoraphanin, understanding how this compound is accumulated in broccoli, cabbage, and other cruciferous vegetables is essential. The plant material and genomic resources utilized in this study have provided additional support for previously identified loci associated with GSL regulation. This work has also identified additional genetic factors suggesting that altering, enhancing, or regulating individual 4C aliphatic GSLs in *Brassica* vegetables may be more complex than previously assumed. This finding is consistent with recent genome-wide association studies in *Arabidopsis thaliana* that also suggest the involvement of previously unidentified loci (Chan et al. 2011).

Author contribution statement Dr. Allan Brown conducted QTL analysis from all locations, conducted glucosinolate analysis in Illinois and is the primary author. Dr. Gad Yousef conducted glucosinolate analysis in North Carolina, and statistical analysis of phenotypic data. Dr. Robert Reid, Dr. Kranthi Chebrolu, Ms. Aswathy Thomas, and Mr. Christopher Krueger conducted bioinformatics analysis, identification of candidate sequences from the genomic sequence of B. oleracea, and alignment of genomic sequences. Dr. Eric Jackson contributed to analysis of SNP markers. Dr. Elizabeth Jeffery and Dr. John Juvik provided funding for analysis in Illinois and contributed to the writing.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Standards The experiment conducted complies with the laws of the United States.

References

Baskar V, Gururani M, Yu J, Park S (2012) Engineering glucosinolates in plants: current knowledge and potential uses. Appl Biochem Biotechnol 168:1694–1717

Bender J, Celenza J (2009) Indolic glucosinolates at the crossroads of tryptophan metabolism. Phytochem Rev 8:25–37

Benderoth M, Pfalz M, Kroymann J (2009) Methylthioalkylmalate synthases: genetics, ecology and evolution. Phytochem Rev 8:255–268

Brown AF, Yousef GG, Jeffery EH, Klein PB, Wallig MA, Kushad MM, Juvik JA (2002) Glucosinolate profiles in broccoli: variation in levels and implications in breeding for cancer chemoprotection. J Am Soc Hort Sci 127:807

Brown AF, Jeffery EH, Juvik JA (2007) A polymerase chain reaction-based linkage map of broccoli and identification of quantitative trait loci associated with harvest date and head weight. J Am Soc Hort Sci 132:507–513

Brown A, Yousef G, Chebrolu K, Byrd R, Everhart K, Thomas A, Reid R, Parkin IP, Sharpe A, Oliver R, Guzman I, Jackson E (2014) High-density single nucleotide polymorphism (SNP) array mapping in *Brassica oleracea*: identification of QTL associated with carotenoid variation in broccoli florets. Theor Appl Genet 127:1–14

Bryant C, Kumar S, Chamala S, Shah J, Pal J, Haider M, Seward S, Qazi A, Morris R, Semaan A, Shammas M, Steffes C, Potti R, Prasad M, Weaver D, Batchu R (2010) Sulforaphane induces cell cycle arrest by protecting RB-E2F-1 complex in epithelial ovarian cancer cells. Mol Cancer 9:47

Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR, Normanly J, Bender J (2005) The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiol 137:253–262

Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol 9:e1001125

Clarke JD, Dashwood RH, Ho E (2008) Multi-targeted prevention of cancer by sulforaphane. Cancer Lett 269:291–304

Davis R, Singh KP, Kurzrock R, Shankar S (2009) Sulforaphane inhibits angiogenesis through activation of FOXO transcription factors. Oncol Rep 22:1473–1478

Eberhardt MV, Kobira K, Keck A-S, Juvik JA, Jeffery EH (2005) Correlation analyses of phytochemical composition, chemical, and cellular measures of antioxidant activity of broccoli (*Brassica oleracea* L. Var. *italica*). J Agric Food Chem 53:7421–7431

Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56:5–51

Frerigmann H, Gigolashvili T (2014) MYB34, MYB51 and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Mol Plant 7:814–828

Gao M, Li G, Yang B, McCombie WR, Quiros CF (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, McCombie WR, Quiros C (2005) Comparative analysis of a transposon-rich *Brassica oleracea* BAC clone with its corresponding sequence in *A. thaliana*. Theor Appl Genet 111:949–955

Gao M, Li G, Potter D, McCombie WR, Quiros C (2006) Comparative analysis of methylthioalkylmalate synthase (MAM) gene family and flanking DNA sequences in *Brassica oleracea* and *Arabidopsis thaliana*. Plant Cell Rep 25:592–598

Gigolashvili T, Berger B, Mock H-P, Müller C, Weisshaar B, Flügge U-I (2007a) The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J 50:886–901

Gigolashvili T, Yatusevich R, Berger B, Müller C, Flügge U-I (2007b) The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J 51:247–261

Gray AR (1982) Taxonomy and evolution of broccoli (*Brassica oleracea* var. *italica*). Econ Bot 36:397–410

Greer MA, Deeney JM (1959) Antithyroid activity elicited by the ingestion of pure progoitrin, a naturally occurring thioglycoside of the turnip family. J Clin Invest 38:1465–1474

Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. Trends Plant Sci 11:89–100

Guzman I, Yousef GG, Brown AF (2012) Simultaneous extraction and quantitation of carotenoids, chlorophylls, and tocopherols in *Brassica* vegetables. J Agric Food Chem 60:7238–7244

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Ann Rev Plant Biol 57:303–333

Hansen BG, Kerwin RE, Ober JA, Lambrix VM, Mitchell-Olds T, Gershenzon J, Halkier BA, Kliebenstein DJ (2008) A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in *Arabidopsis*. Plant Physiol 148:2096–2108

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci 104:6478–6483

Jeffery E, Araya M (2009) Physiological effects of broccoli consumption. Phytochem Rev 8:283–298

Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics 159:359–370

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon 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

Kurtz S, Phillippy A, Delcher A, Smoot M, Shumway M, Antonescu C, Salzberg S (2004) Versatile and open software for comparing large genomes. Genome Biol 5:R12

Kushad M, Brown AF, Kurilich A, Juvik JA, Klein PP, Wallig M, Jeffery E (1999) Variation of glucosinolates in vegetable crops of *Brassica oleracea*. J Agric Food chem 47:1541–1548

Levy M, Wang Q, Kaspi R, Parrella MP, Abel S (2005) *Arabidopsis* IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. Plant J 43:79–96

Li G, Quiros CF (2003) In planta side-chain glucosinolate modification in *Arabidopsis* by introduction of dioxygenase *Brassica* homolog BoGSL-ALK. Theor Appl Genet 106:1116–1121

Liu Z, Hirani A, McVetty PE, Daayf F, Quiros C, Li G (2012) Reducing progoitrin and enriching glucoraphanin in *Braasica napus* seeds through silencing of the GSL-ALK gene family. Plant Mol Biol 79:179–189

Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A (2008) The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators. Plant Physiol 148:2021–2049

Maruyama-Nakashita A, Nakamura Y, Tohge T, Saito K, Takahashi H (2006) *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. Plant Cell 18:3235–3251

Munday R, Munday CM (2004) Induction of phase II detoxification enzymes in rats by plant-derived isothiocyanates: comparison of allyl isothiocyanate with sulforaphane and related compounds. J Agric Food Chem 52:1867–1871

Parkin IAP, Sharpe AG, Lydiate DJ (2003) Patterns of genome duplication within the *Brassica napus* genome. Genome 46:291–303

Sawada Y, Akiyama K, Sakata A, Kuwahara A, Otsuki H, Sakurai T, Saito K, Hirai MY (2009) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. Plant Cell Physiol 50:37–47

Shankar S, Ganapathy S, Srivastava RK (2008) Sulforaphane enhances the therapeutic potential of trail in prostate cancer orthotopic model through regulation of apoptosis, metastasis, and angiogenesis. Clin Cancer Res 14:6855–6866

Skirycz A, Reichelt M, Burow M, Birkemeyer C, Rolcik J, Kopka J, Zanor MI, Gershenzon J, Strnad M, Szopa J, Mueller-Roeber B, Witt I (2006) DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis*. Plant J 47:10–24

Slater G, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31

Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ (2007) A systems biology approach ldentifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. PLoS One 2:e1322

Sønderby IE, Geu-Flores F, Halkier BA (2010) Biosynthesis of glucosinolates—gene discovery and beyond. Trends Plant Sci 15:283–290

Thejass P, Kuttan G (2006) Antimetastatic activity of sulforaphane. Life Sci 78:3043–3050

Traka MH, Saha S, Huseby S, Kopriva S, Walley PG, Barker GC, Moore J, Mero G, van den Bosch F, Constant H, Kelly L, Schepers H, Boddupalli S, Mithen RF (2013) Genetic regulation of glucoraphanin accumulation in Beneforté® broccoli. New Phytol 198:1085–1095

Van Oojen JW, Boer MP, Jansen RC, Maliepaard C (2002) MapQTL 4.0, software for the calculation of QTL position on genetic maps. Plant Res Int Wagening 198:1085–1095

Zandalinas SI, Vives-Peris V, Gómez-Cadenas A, Arbona V (2012) A fast and precise method to identify indolic glucosinolates and camalexin in plants by combining mass spectrometric and biological information. J Agric Food Chem 60:8648–8658

Zhang Y, Tang L (2007) Discovery and development of sulforaphane as a cancer chemopreventive phytochemical1. Acta Pharm Sinica 28:1343–1354