Theor Appl Genet (2000) 101:429–437

H. Campos de Quiros · R. Magrath · D. McCallum
J. Kroymann · D. Scnabelrauch · T. Mitchell-Olds
R. Mithen

α-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*

Received: 19 November 1999 / Accepted: 16 December 1999

Abstract QTL mapping of glucosinolates in a RI population derived from an F_1 hybrid between the *Arabidopsis thaliana* ecotypes Columbia and Landsberg *erecta* identified a single major QTL coincident with the *GSL-ELONG* locus which regulates side chain elongation. Physical mapping and sequencing identified two members of an isopropylmalate synthase-like gene family within the region of maximum LOD score for the QTL and the *GSL-ELONG* non-recombinant region. These genes are prime candidates for regulating glucosinolate biosynthesis.

Key words *Arabidopsis thaliana* · Glucosinolates · Isopropylmalate synthase

Introduction

Glucosinolates are thioglycosides which occur in the order Brassicales. The molecule consists of a glycone moiety and a variable aglycone side chain (Halkier and Du, 1997). Within the Brassicaceae, the major class of glucosinolates have side chains which are derived from methionine. These vary in the length of the side chain and in a variety of modifications to its structure. Methionine-derived glucosinolates are of considerable biological and economic importance due to the activity of the degradation products which are produced following tissue disruption. These products include isothiocyanates ('mustard oils'), nitriles, epithiocyanates and thiocyanates (Bones and Rossiter 1996). Degradation products from certain methionine-derived glucosinolates have antinutritional effects, such as those from 2-hydroxy-3-butenyl glucosinolate which accumulate in the seeds of canola (*Brassica napus*), whereas products from other glucosinolates have beneficial effects, such those from 4-methylsulphinylbutyl glucosinolate, which is found in broccoli and has anticarcinogenic activity in mammalian systems due to its ability to induce phase-II detoxification enzymes (Faulkner et al. 1997). Additionally, many glucosinolate degradation products mediate plant-herbivore interactions (Giamoustaris and Mithen 1995). Thus, for several reasons it is desirable to be able to modify the glucosinolate content of crop plants, either by conventional breeding or by genetic modification.

Details of the biosynthesis of methionine-derived glucosinolates are not fully understood. The side chain length varies considerably, both within and between species. For example, in *Brassica oleracea* and *Arabidopsis thaliana*, the major leaf glucosinolates have either a '3C' side chain (e.g. 3-methylthioproyl, 3-methylsulphinylpropyl, 2-propenyl glucosinolate) or a '4C' side chain (e.g. 4-methylthiobutyl, 4-methylsulphinylbutyl, 3-butenyl, 2-hydroxy-3-butenyl glucosinolate), while in the seeds of these and other species longer side chains occur. Biochemical studies have provided evidence that methionine elongation occurs prior to glucosinolate biosynthesis through the conversion of methionine to an α-keto acid followed by condensation with acetyl CoA, analogous to the biosynthesis of leucine from 2-keto-3-methylbutanoic acid ((Chisholm and Wetter 1964); Fig 1). The elongated α-keto acid then either forms the substrate for the synthesis of glucosinolates with a 3C side chain, or for further rounds of elongation through one or more further condensation reactions with acetyl CoA. In previous studies, Mendelian genes

Communicated by F. Salamini

EMBL accession numbers for new sequence data, AJ131517 – ATH131517: sequence corresponding to the region spanning the gap between MRN17 (AB006708) and MYJ24 (AB005243) (i.e. cosmid C63H5 and PCR product spanning the gap between this cosmid and MYJ24), contains *IPMS-2* from Columbia. AJ 131518 – ATH131518 contains *IPMS-1* mrRNA from Landsberg *erecta*; AJ 133892 – ATH133892 contains *IPMS-2* mRNA from Landsberg *erecta*

H. Campos de Quiros · R. Magrath · R. Mithen (✉)
John Innes Centre, Norwich Research Park, Colney Lane,
Norwich NR4 7UH, UK
Fax: +44(0)1603-450014
e-mail: Richard.Mithen@bbsrc.ac.uk

D. McCallum · J. Kroymann · D. Scnabelrauch · T. Mitchell-Olds
Max Planck Institute of Chemical Ecology,
Tatzendpromenade 1a, 07745 Jena, Germany

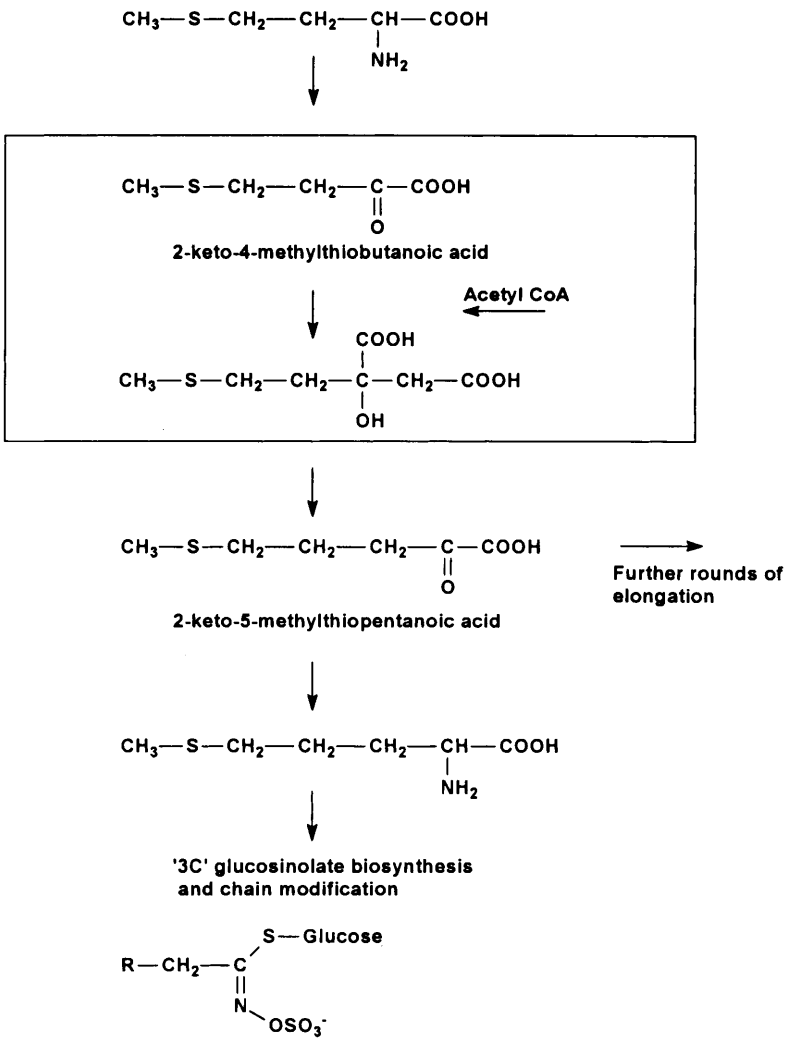

Fig. 1 Proposed pathway of methionine elongation prior to glucosinolate biosynthesis. In the proposed model, glucosinolate precursors in genotypes with the Ler *GSL-ELONG* allele undergo one round of elongation, while those with the Col allele undergo two rounds. However, in both ecotypes, smaller amounts of glucosinolates occur which result from multiple chain elongation (see Fig. 2). The condensation reaction (*boxed*) is the likely reaction catalysed by ISPM-At1 and ISPM-At2

have been positioned on chromosome 5 and chromosome 4 of *Arabidopsis thaliana* which regulate side chain elongation and modification respectively (Magrath et al. 1994; Mithen et al. 1995). In the study reported here we analysed *A. thaliana* for variation in the total amount of methionine-derived glucosinolates. We were able to show that genes of an isopropylmalate synthase (IPMS)-like gene family, which catalyse the condensation of acetyl CoA and α-keto acids, are coincident with a Mendelian gene which regulates side chain elongation and a major quantitative trait locus (QTL) which regulates the total level of methionine-derived glucosinolates in *Arabidopsis thaliana*.

Materials and Methods

Arabidopsis QTL mapping

The mapping population consisted of 60 randomly selected recombinant inbred (RI) lines from a Landsberg *erecta* (Ler)×Columbia (Cal) cross (Lister and Dean 1993). Thirty plants of each line were grown in a 8.5×8.5-cm pot under standard glasshouse conditions. Each line/pot was replicated three times and randomised. After 6 weeks growth, leaf tissue was harvested, frozen in liquid nitrogen and freeze-dried. Glucosinolates were analysed as described previously (Magrath et al. 1994). QTL analysis was undertaken with MAPMAKER/QTL. While this programme does not contain a specific algorithm for RI lines, Wang et al. (1994) and Ray et al. (1996) demonstrated its successful application to RI lines, and it has been used to map QTLs for trichome number in the same Ler X Col RI population as used in the present study (Larkin et al. 1996). In addition to MAPMAKER/QTL, single marker analysis was used to confirm the position of QTLs. ANOVA was used to compare the means of glucosinolate content of the two homozygous classes at each of 146 RFLP loci distributed across the genome. Threshold values of significance was calculated using the approach of Churchill and Doerge (1994). This involved reassigning the phenotypic data to RI lines at random and determining the maximum *t*-value. This procedure is repeated 10000 times, and the maximum *t*-value in each cycle is ranked in increasing order of significance. The 9999th and 9990th, values give the *t* values for the P=0.0001 and P=0.001 levels of significance. These values, 4.108 and 3.258, respectively, were used to assess the significance of markers.

Fine mapping of the *GSL-ELONG* locus

Thirty plants of each of 264 RI lines were grown in 8.5×8.5-cm pots and the leaves harvested and analysed for glucosinolate content after 6 weeks growth. Glucosinolates were also extracted and analysed from seeds of 100 RI lines to study the correlation of

glucosinolate profiles between seeds and leaves. In order to confirm identity of the glucosinolates, we additionally analysed selected leaf and seed extracts by LC-MS (Liquid chromatography-mass spectrometry) with atmospheric pressure chemical ionisation. Molecular ions+H^+ were obtained for all glucosinolates.

Isolation and analysis of plant DNA

Arabidopsis thaliana plant DNA was isolated and quantified as described previously (Magrath et al. 1994). Restriction digestion was performed on 20 µg of genomic DNA following the supplier's (Boehringer, Mannheim) recommendations. The resultant fragments were separated by gel electrophoresis, capillary-blotted onto nylon membrane (Hybond N+) and subsequently probed (Magrath et al. 1994).

Isolation of large insert clones and library screening

Isolation and manipulation of YAC DNA was as described by Bancroft (1999). BAC and cosmid DNA was prepared from 50 ml of overnight cultures by alkaline lysis (Birnboim and Doly 1979) followed by propan-2-ol precipitation at 4°C. Nucleic acids were recovered by centrifugation (2000 *g* for 10 min at 4°C) and the dried pellet resuspended in 2 ml TE (100 mM TRIS-HCl, pH 7.4; 10 m*M* EDTA). Subsequently, 2 ml 4 *M* lithium acetate was added prior to a 20-min incubation in 4 µl TE and centrifugation as above. Following collection of the supernatant, the DNA was precipitated with 0.2 ml 3 *M* sodium acetate and 9 ml absolute ethanol at 4°C for 1 h. The dried pellet was resuspended in 4 ml TE. This solution was extracted once with phenol/chloroform, ethanol precipitated and finally resuspended in 50 µl TE.

Inverse polymerase chain reaction (PCR)

Restriction fragment length polymorphism (RFLP) probes were prepared from BAC ends by restriction digestion of 5 µl DNA. Digested DNA was extracted with phenol/chloroform, and dialysised with Sepharose CL 6B sin columns (Pharmacia) into 20 µl TE. Self-ligation took place with 1 µl T4 DNA ligase (Pharmacia) and 10 µl supplied 10×buffer in a 100 µl reaction volume overnight at 12°C. An inverse (i)PCR (Ochman et al. 1988) was conducted using 3 µl of the ligation with 0.25 mM dNTPs, 10 µM primers, 1.5 U Amplitaq (Perkin Elmer) and 5 µl 10×buffer in a final volume of 50 µl. Primers used were 5´-CGACCTGCAGGCATGCAAGCTT-3´ and 5´-ACTCTAGAGGATCCCCGGGTAC-3´, right side; 5´-CCTAAATAGCTTGGCGTAATCATG-3´ and 5´-TGACACTATAGAATACTCAAGCTT-3´, left side. Cycling conditions were 94°C for 4 min, followed by 35 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 2 min and finally by 1 cycle at 72°C for 10 min. The products were resolved by means of electrophoresis and purified using the QIAquick (Qiagen) gel band purification kit.

cDNA isolation and sequence analysis

Total plant RNA was isolated with Trizol (Life Technologies) according to the manufacturer's recommendations. Poly A+ mRNA was isolated with a Dynabeads mRNA purification kit (Dynal, Norway) following the manufacturer's recommendations. A modified version of RACE (Rapid Amplification of CDNA Ends) was used to synthesise cDNAs. Specific 5'RACE primers 5´-AGCCCCTGGAGTGTTTTCTAT-3´ and 5´-TTACAACAAAAACGGCATTATCAA-3´ were used for IPMS-At1 and IPMS-At2, respectively. Specific 3´RACE primers 5´-GAGGGTAATGCTATTCACATC-3´ and 5´-TCTCAAACCCGTTGTGGAAC-3´ were used for IPMS-At1 and IPMS-At2 respectively. Each PCR contained 2 U *Taq* polymerase (Qiagen, Germany), 0.2 m*M* dNTP's, 5 pmol of each primer, template and 1×Qiagen PCR buffer in a reaction volume of 50 µl. Cycling conditions were 94°C for 3 min followed by 35 cycles of 94°C for 45 s, 53–59°C for 1 min and 72°C for 3 min and finally by 10 min at 72°C. PCR products were gel-purified (Qiagen, Germany) and cloned into pCR2.1-TOPO (Invitrogen, the Netherlands).

Cosmid 63H5 and P1 MYJ24 (see Results) are not contiguous. Genomic DNA between the ends of these DNA fragments was amplified via PCR. Reactions comprised 20 ng Columbia DNA, 10 pmol primers (5´-AATCTGTGGACTGACATCGAT-3´ and 5´-GATAAGAACCCGCACATAATG-3´), and 1 'Ready To Go' PCR bead (Amersham Pharmacia). Cycling conditions were 95°C for 5 min followed by 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 2 min and finally by 10 min at 72°C. Sequence analysis of cosmid 63H5 and the PCR product was conducted using ABI 377 Sequencer and BigDye Terminator (PE Biosystems). Sequence assembly was conducted using DNASTAR (DNASTAR, USA) software.

Results

Mapping QTLs for glucosinolate accumulation in Arabidopsis

Glucosinolates were quantified in leaves of 60 RI lines from the Lister and Dean Landsberg *erecta* (Ler)× Columbia (Cal) population within a replicated randomised experiment. Variation in glucosinolate content between the parents and RI lines was as previously described (Magrath et al. 1994; Fig. 2).There was no significant difference in the total amount of methionine derived glucosinolates in leaves between Ler and Col, but transgressive segregation was observed amongst the RI lines, with 5 lines having significantly higher levels of glucosinolates than either parent and 5 lines having significantly lower levels (Table 1). Single-marker analysis and interval mapping identified one major QTL on chromosome 5, with a maximum LOD score of 11.4. This QTL was coincident with the previously mapped Mendelian *GSL-ELONG* locus (Fig. 3; Magrath et al. 1994). Two possible minor QTLs were also identified (Table 1); one occurred on chromosome 1 between markers m213 and mi 315 (maximum LOD score=2.6) and the other on chromosome 4 between mi233 and m518 A (maximum LOD score=1.3). While these are both of questionable significance, Col alleles at these two minor QTLs resulted in higher levels of glucosinolates, providing one explanation of the genetic basis of the transgressive segregation. Single-marker analysis gave identical results to interval mapping. Markers linked to the previously mapped *GSL-ELONG* locus on chromosome 5 all gave *t*-values with $P<0.0001$ (mi433, $t=9.01$; mi90, $t=7.86$). No other marker gave a level of significant association with glucosinolates at $P<0.001$.

Physical mapping of *GSL-ELONG* and QTL-1

GSL-ELONG and QTL-1 may be the same gene, or they may represent closely linked genes. We developed a physical map around the locality of *GSL-ELONG*, including the region of maximum likelihood for the QTL,

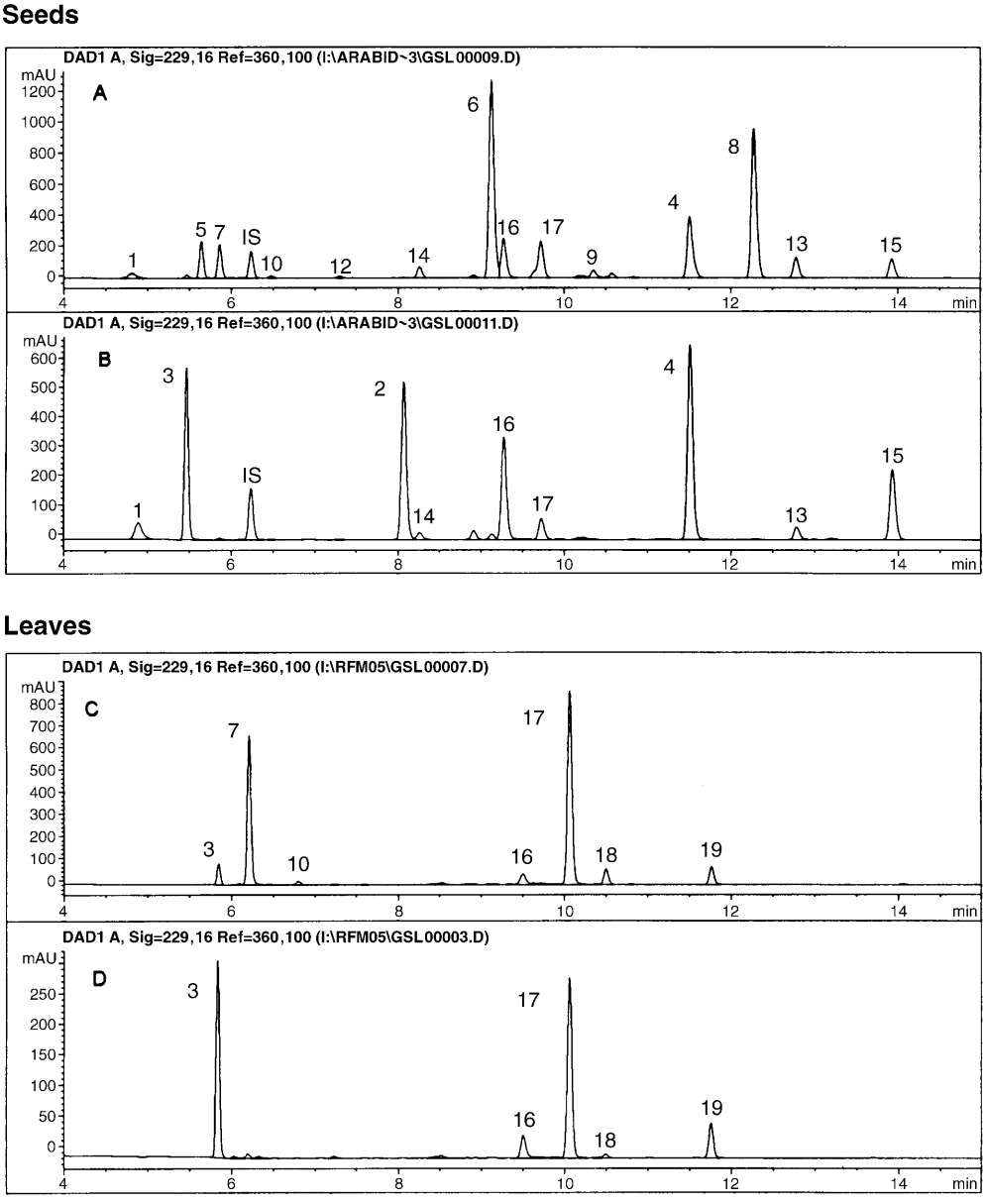

Fig. 2A–D Separation of desulphoglucosinolates occurring in the seeds and leaves of Columbia and RI line N1937. The major glucosinolates in Columbia have a 4C side chain, while those in N1937, which has a Ler allele at *GSL-ELONG* but a Col allele at *GSL-OHP*, have 3C side chain. **A** Seeds/Columbia, **B** seeds/N1937, **C** leaves/Columbia, **D** leaves/N1937. Desulphoglucosinolates were identified through LC-MS with APCI. *Peaks 1* 3-Hydroxypropyl, *2* 3-methylthiopropyl, *3* 3-methylsulphinylpropyl, *4* 3-benzoyloxypropyl, *5* 4-hydroxybutyl, *6* 4-methylthiobutyl, *7* 4-methylsulphinylbutyl, *8* 4-benzoyloxybutyl, *9* 5-methylthiopentyl, *10* 5-methylsulphinylpentyl, *12* 6-methylsulphinylhexyl, *13* 7-methylthioheptyl, *14* 7-methylsulphinylheptyl, *15* 8-methylthiooctyl, *16* 8-methylsulphinyloctyl, *17* 3-indolylmethyl, *18* 4-methoxy-3-indolylmethyl, *19* 1-methoxy-3-indolylmethyl, *IS* internal standard (2-propenyl glucosinolate)

in order to identify any candidate genes for both the QTL and *GSL-ELONG* (Fig. 3).

The *GSL-ELONG* locus was fine mapped by analysis of glucosinolates in 264 RI lines. Of these, 153 lines had 3C glucosinolates and 111 lines had 4C glucosinolates. In RI lines with predominantly 4C glucosinolates, the major glucosinolate in the leaves was 4-methylsulphinylbutyl glucosinolate, while in RI lines with predominantly 3C glucosinolates, the major glucosinolate was either 3-methylsulphinylpropyl glucosinolate or 3-hydroxypropyl glucosinolate, depending upon the genotype at the *GSL-OHP* locus (Mithen et al. 1995). The 3C or 4C leaf phenotype correlated with seed glucosinolates. However, in seeds, the major 4C glucosinolates were 4-methylthiobutyl and 4-benzoyloxybutyl glucosinolates, and the major 3C glucosinolates were 3-methylthiopropyl and 3-benzoyloxybutyl (Fig. 2). The glucosinolate profile of the seeds and leaves of Columbia were similar to that reported previously (Haughn et al. 1991).

Through the use of *Eco*RI DNA digests of genomic DNA, the *GSL-ELONG* locus was positioned between the cosmid RFLP marker pCIT718 and the YAC end yUP21F7LE. Out of 264 RI lines, there were two recombinants between the locus and pCIT718 (RI lines N4692 and N4741) and two between the locus and yUP21F7LE (N4665and N4684). YAC clone yUP6F5, which had been positioned adjacent (but not overlapping) to yUP21F7 as part of contig 7 on chromosome 5 (http://nasc.nott.ac.uk/JIC-contigs/chr5_YAC-contig7.GIF), was used as a probe into the TAMU BAC library. Twenty three clones were selected, assembled into a contig and orientated through hybridisation to pCIT718. The ends of several of these clones were isolated by iPCR and used as RFLP probes to genomic DNA from RI

Table 1 Genotypes at RFLP loci linked to QTLs which determine the level of methionine-derived glucosinlates in leaves of RI lines having either significantly greater levels or significantly lower levels of glucosinolates than either parent (ANOVA, $P<0.05$). Markers associated with QTLs fall within the region of maximum LOD score for each QTL. Probablity levels for single-marker analysis were calculated as described in the Materials and methods. QTLs 2 and 3 are of questionable significance

[a] C, Columbia; L, Landsberg

Ecotype or RI line	Methionine-derived glucosinolates (μmol g^{-1})	Allelic configuration at RFLP loci linked to QTLs		
		QTL-1 mi 433	QTL-2 g4026	QTL-3 m506
Columbia	21.0±3.40	C	C	C
Landsberg	20.4±2.82	L	L	L
High glucosinolate transgressives				
N1928	34.6±5.96	L	C	C
N1939	30.2±3.78	L	C	C
N1940	31.1±0.88	L	C	C
N1950	31.0±0.50	L	C	C
N1955	28.8±2.95	L	C	C
Low glucosinolate transgressives				
N1903	8.1±1.22	C	L	L
N1948	11.9±0.56	C	L	L
N1956	10.8±1.44	C	L	L
N1964	9.8±0.72	C	L	L
N1993	10.6±2.03	C	L	C

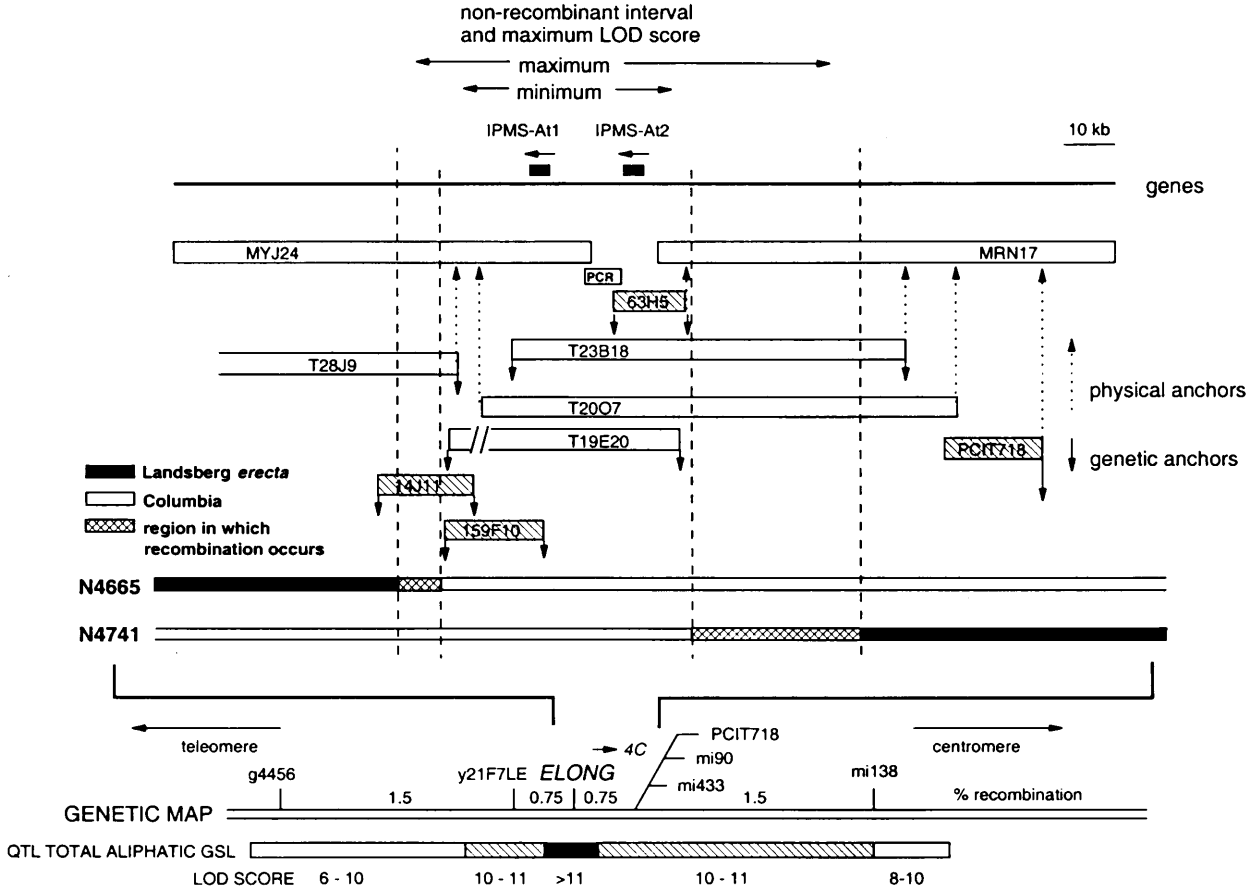

Fig. 3 Genetic and physical map in the region of QTL-1 and *GSL-ELONG*. Details of map construction are provided in the text. MYJ24 and MRN17 are sequenced P1 clones. T28J9, T23B18, T20O7, T19E20 and T19E20 are BAC clones from the TAMU library. 63H5, 14J11, 159F10 and PCIT718 are cosmid clones available from John Innes Centre. PCR is a pcr product. Cosmid 63H5 and PCR were sequenced to provide the entire nucleotide sequenced across the non-recombinant region. Genetic anchors provided by the use of BAC end sequences function as RFLP probes. Physical anchors are provided by integrating BAC end sequences to the sequenced P1 clones

lines with recombination events near to the target gene. Polymorphisms were obtained with the following BAC end probes/enzyme combinations: T23B18LE/*Xba*I, T19E20RE/*Hpa*II, T19E20LE/*Dra*I, T5K20RE/*Hpa*II, T23B18RE/*Hin*cII, T28J9RE/*Eco*RI, T28J9LE/*Hin*cIII, T21B12RE/*Sac*I, T14D2RE/*Bam*HI and T14D2LE/*Eco*RV. Both ends of T19E20 and a single end of each of clones T20O7 and T28J9 cosegregated with the phenotype. These results were consistent with the physical map based on restriction mapping (Fig. 3). The end sequence of several BACs were used to integrate the

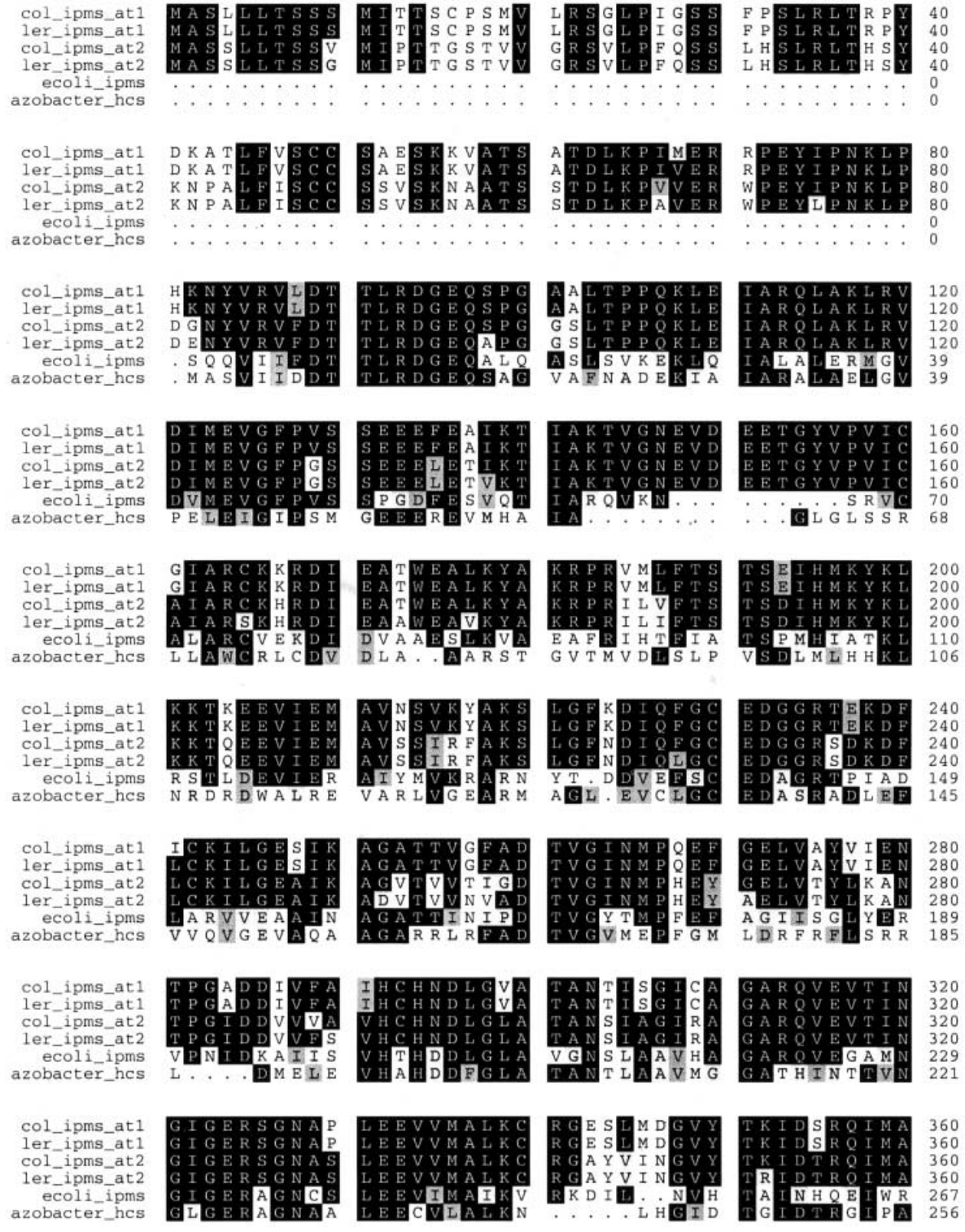

Fig. 4 Sequence homology between predicted amino acid sequences of the Col and Ler alleles of IPMS-At1 and IPMS-At2, and IPMS from *E. coli* and homocitrate synthase from *Azobacter vinelandii* (Swissprot accession number P09151 and P05342, respectively). Black shading indicates high regions of homology

physical map to P1 clones which had been sequenced as part of the international Arabidopsis genome programme. The non-recombinant interval spanned 2 sequenced clones, MYJ24 and MRN17, and a non sequenced gap between the clones, estimated to be approximately 12 kb (Sato et al. 1997). Two cosmids (63H5 and 159F10) from a Columbia library were selected by means of hybridsation with BAC T19E20 that

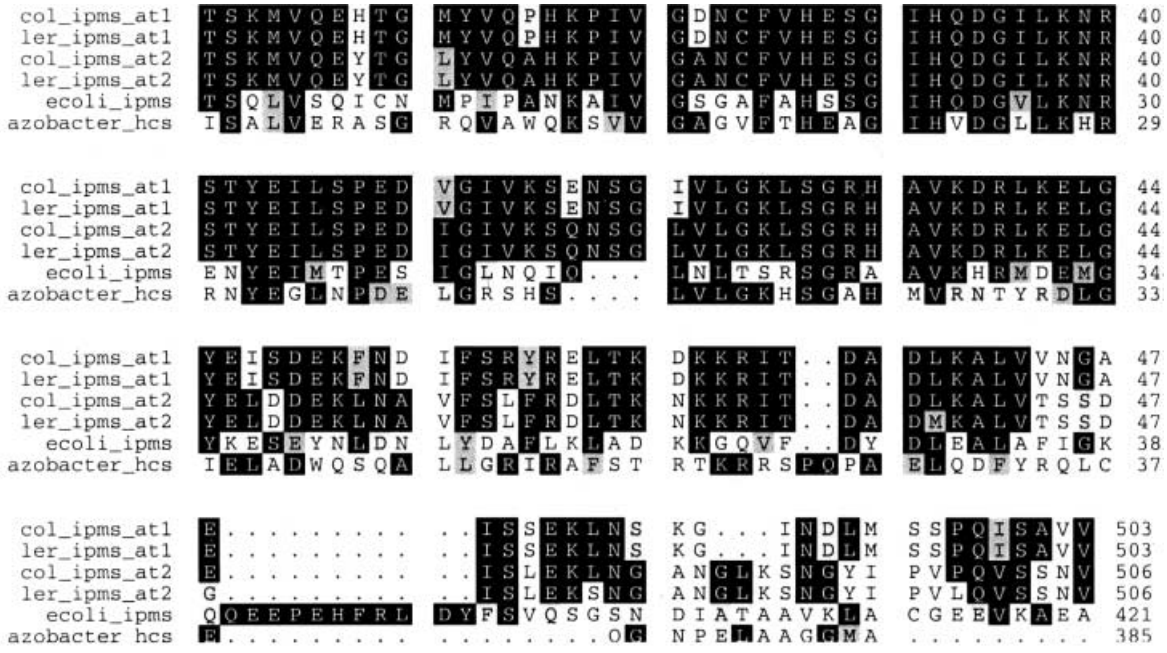

Fig. 4

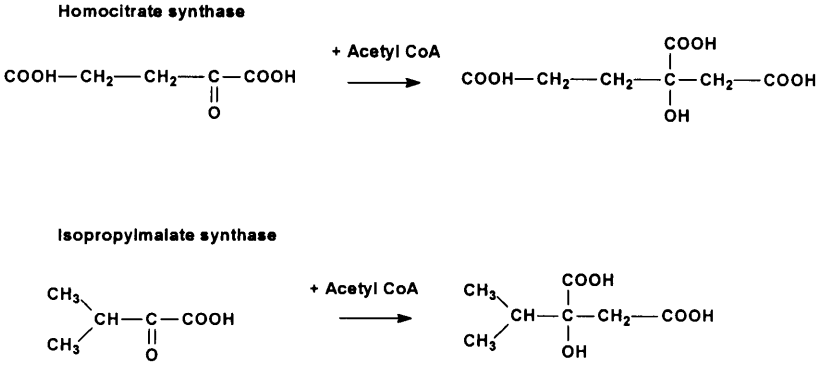

Fig. 5 Condensation reactions between α-keto acids and acetyl CoA catalysed by homocitrate synthase and isopropylmalate synthase. These reactions are similar to the second step in methionine elongation (Fig. 1)

were located partially in the interval between the P1 clones. As expected, when these two cosmids were used as probes on to genomic DNA they identified cosegregating RFLPs. Cosmid 63H5 was sequenced and shown to be 17.2 kb long. The gap between the end of MYJ24 and 63H5 was bridged by the PCR, and the resulting 1.2-kb fragment was sequenced. Thus, the sequence for the entire non-recombinant region and the region of maximum LOD score was obtained.

Identification of a candidate gene family

Within the non-recombinant region, there were two similar genes that were strong candidates for involvement in glucosinolate biosynthesis. Both of these had high homology to isopropylmalate synthase and to homocitrate synthase from nitrogen fixing bacteria (Fig. 4), with the exception of an additional 80 bp putative chloroplast targeting sequence, identified by the high frequency of serine and theronine residues and homology to other chloroplast targeting sequences. These two genes were designated IPMS-At1 and IPMS-At2. Isopropylmalate synthase and homocitrate synthase catalyse the condensation of acetyl CoA with α-keto acids (Evans et al. 1991; Fig 5). The genes share several highly conserved domains, such as the ...DTTLRDHEQ..... motif near the 5' end, which are thought to provide the active sites for the keto acid condensation reaction (Evans et al. 1991). A α-keto acid/acetyl CoA condensation reaction is the second step in the suggested elongation reaction of methionine (Fig. 1). No other genes within the non-recombinant interval were obvious candidates for involvement in glucosinolate biosynthesis and chain elongation. Details of these genes are available at www.kazusa.org.jp.

Discussion

In this study, transgressive segregation for glucosinolate accumulation in *A. thaliana* enabled an important QTL to be located on chromosome 5 that was coincident with the Mendelian *GSL-ELONG* locus. Transgressive segregation for glucosinolate content has also been reported in

B. napus in which alleles at one QTL in the 'low' glucosinolate cultivar Stellar contributed to a 'high' glucosinolate phenotype when expressed in a different genetic background (Toroser et al. 1995). The manner by which alleles at these different QTLs in both *A. thaliana* and *B. napus* interact is not known: they may act independently from each other and regulate different fluxes within the biosynthetic pathway so that the resultant phenotype is a sum of these fluxes; alternatively, there may be more complex epistatic interactions occurring, in which combinations of alleles regulate the flux through a single point in the pathway.

Members of a IPMS-like gene family have been identified as candidate genes to regulate glucosinolate side chain elongation and the total level of methionine-derived glucosinolates in *Arabidopsis thaliana* via the condensation of an α-keto acid derived from methionine and acetyl CoA. This is consistent with biochemical studies which demonstrated that the elongation of methionine prior to glucosinolate biosynthesis is analogous to leucine biosynthesis (Chisholm and Wetter 1964), which requires isopropylmalate synthase (Strassmann and Ceci 1963). It is also consistent with genetic studies on *Brassica* and *Arabidopsis* which suggest that the genes which reduce methionine glucosinolates have no effect on indolyl glucosinolate content and are therefore likely to be involved in the supply of methionine elongated homologues rather than the subsequent conversion of amino acids to glucosinolates (Toroser et al. 1995). In *Brassica napus*, an association between genes regulating side chain elongation and the total amount of methionine-derived glucosinolates has also been reported (Toroser et al. 1995).

The glucosinolate profile of leaves and seeds of Columbia reported in this study are similar to those reported by Haughn et al. (1991). In this previous study a mutant was obtained in which there is an apparent shift in the profile from predominantly 4 C to 3 C glucosinolates. It is possible that the mutation was in one of the two IPMS-At genes described in the current study.

IPMS is known to be an important regulatory gene in leucine biosynthesis (Umbarger 1997). This enzyme cannot only catalyse the condensation of 2-keto-3-methylbutanoic acid with acetyl CoA but also the series of elongated homologues, 2-ketopropanoic acid, 2-ketobutanoic acid and 2-ketopentanoic acid (Umbarger 1997). Moreover, Howell et al. (1998) showed that an enzyme from a single gene in *Methanoccous jannaschii* (MJ0503) can catalyse the condensation of acetyl CoA with a series of elongated α-keto acids, in the synthesis of the alkyl portion of mercaptoheptanoic acid. In the proposed scheme of glucosinolate biosynthesis, the IPMS-like genes would, in a similar manner, catalyse the condensation of 2-keto-4-methylthiobutanoic acid, 2-keto-5-meththiopentanic acid, 2-keto-6-methylthiohexanoic acid, and so on, up to 2-keto-9-nonanoic acid to provide elongated forms of methionine for subsequent synthesis of chain-elongated glucosinolates. The difference between the Ler *GSL-ELONG* allele and the Col *GSL-ELONG* allele is that the former would primarily only catalyse the condensation of 2-keto-4-methylbutanoic acid with acetyl CoA to result in 3C glucosinolates, while the latter would catalyse the condensation of both 2-keto-4-methylthiobutanoic acid and 2-keto-5-methylpentanoic acid to result in 4C glucosinolates. Studies to test this hypothesis requires functional analysis via in vitro expression of the IPMS-like genes.

These IPMS-like genes may have dual specificity for both 2-keto-3-methylbutanoic acid (leucine biosynthesis) and 2-keto-4-methylthiobutanoic acid and longer homologues (glucosinolate biosynthesis), or these genes may have a higher degree of specifity so that different members of this gene family function in primary and secondary metabolism, respectively. These questions need to be resolved via functional analysis of ISPM-At1 and ISPM-At2, ideally via transgenic approaches in Arabidopsis.

Acknowledgements The authors thank the Biotechnology and Biological Sciences Research Council for funding.

References

Bancroft I (2000) Physical mapping: YACs, BACs, cosmids and nucleotide sequences. In: Wilson Z (ed) Arabidopsis: a practical approach. Oxford University Press, Oxford, pp 199–234

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513

Bones AM, Rossiter JT (1996) The myrosinase-glucosinolate system, its organisation and biochemistry. Physiol Plant 97: 194–208

Chisholm MD, Wetter LR (1964) Biosynthesis of mustard oil glucosides. IV. The administration of methionine-^{14}C and related compounds to horseradish. Can J Biochem 42:1033–1040

Churchill MD, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Evans DJ, Jones, R, Woodley PR, Wilborn JR, Robson RL (1991) Nucleotide sequence and genetic analysis of *the Azotobacter chroococcum* nifUSVWZM gene cluster, including a new gene (*nifP*) which encodes a serine acetyltransferase. J Bacteriol 173:5457–5469

Faulkner K, Mithen RF, Williamson G (1997) Selective increase of the potential anticarcinogen 4-methylsulphinylbutyl glucosinolate in broccoli. Carcinogenesis 19:605–609

Giamoustaris A, Mithen RF (1995) The effect of modifying the glucosinolate content of leaves of oilseed rape (Brassica napus spp oleifera) on its interaction with specialist and generalist pests. Ann Appl Biol 126:347–363

Halkier BA, Du L (1997) The biosynthesis of glucosinolates. Trends Plant Sci 2:425–431

Haughn GW, Davin L, Giblin M, Underhill EW (1991) Biochemical genetics of plant secondary metabolism in *Arabidopsis thaliana*. The glucosinolates. Plant Physiol 97:217–226

Howell DM, Harich K, Xu H, White RH (1998) α-Ketoacid elongation reactions involved in the biosynthesis of coenzyme B (7-mercaptoheptonoyl threonine phosphate) in methanogenic Archaea. Biochemistry 37:10108–10117

Larkin J, Young G, Prigge M, Marks M (1996) The control of trichome spacing and numbers in Arabidopsis. Development 122:997–1005

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers. Plant J 4:745–750

Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, Lister C, Dean C, Lydiate D, Mithen RF (1994) Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. Heredity 72:290–199

Mithen RF, Clarke J, Lister C, Dean C (1995) Genetics of aliphatic glucosinolates. III. Side chain modifications in *Arabidopsis thaliana*. Heredity 74:210–215

Ochman H, Gerber AS, Hart DL (1988) Genetic application of inverse polymerase chain reaction. Genetics 120:621–623

Ray J, Yu L, MsCouch S, Champoux M, Wang G, Nguyen H (1996) Mappig quantitative trait loci associated with root penetration ability in rice. Theor Appl Genet 92:627–636

Sato S, Kotani H, Nakamura Y, Keneko T, Asamizu E, Fukami M, Miyajima N, Tabata S (1997) Structural analysis of Arabidopsis thaliana chromosome 5.I. Sequence features of the 1.6 Mb covered by phyically assigned twenty P1 clones. DNA Res 4:215–230

Strassman M, Ceci LN (1963) Enzymic formation of α-isopropylmalic acid, an intermediate in leucine biosynthesis. J Biol Chem 238:2445–2452

Toroser D, Thormann CE, Osborn TC, Mithen RF (1995) RFLP mapping of quantitative trait loci controlling seed aliphatic glucosinolate content in oilseed rape (*Brassica napus*). Theor Appl Genet 91:802–808

Umbarger HE (1997) Biosynthesis of branched-chain amino acids. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella* cellular and molecular biology, vol 1, 2nd edn. ASM Press, Washington D.C., pp 422–456

Wang G, Mackill D, Bonman J, McCouch S, Champoux M, Nelson R (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistance rice cultivar. Theor Appl Genet 136:1421–1434