Theor Appl Genet (2005) 110: 1203–1209
DOI 10.1007/s00122-005-1950-7

ORIGINAL PAPER

H. Matsumura · S. Watanabe · K. Harada · M. Senda
S. Akada · S. Kawasaki · E. G. Dubouzet
N. Minaka · R. Takahashi

Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean

Received: 12 June 2004 / Accepted: 3 February 2005 / Published online: 25 March 2005

Abstract Chalcone synthase (CHS), the key enzyme in the flavonoid biosynthesis pathway, is encoded by a multigene family, CHS1–CHS8 and dCHS1 in soybean. A tandem repeat of CHS1, CHS3 and CHS4, and dCHS1 that is believed to be located in the vicinity comprises the *I* locus that suppresses coloration of the seed coat. This study was conducted to determine the location of all CHS members by using PCR-based DNA markers. Primers were constructed based on varietal differences in either the nucleotide sequence of the 5′-upstream region or the first intron of two cultivars, Misuzudaizu, with a yellow seed coat (*II*), and Moshidou Gong 503, with a brown seed coat (*ii*). One hundred and fifty recombinant inbred lines that originated from a cross between these two cultivars were used for linkage mapping together with 360 markers. Linkage mapping confirmed that CHS1, CHS3, CHS4, dCHS1, and the *I* locus are located at the same position in molecular linkage group (MLG) A2. CHS5 was mapped at a distance of 0.3 cM from the gene cluster. CHS2 and CHS6 were located in the middle region of MLGs A1 and K, respectively, while CHS7 and CHS8 were found at the distal end of MLGs D1a and B1, respectively. Phylogenetic analysis indicated that CHS1, CHS3, CHS4, and CHS5 are closely related, suggesting that gene duplication may have occurred repeatedly to form the *I* locus. In addition, CHS7 and CHS8 located at the distal end and CHS2, CHS6, and CHS members around the *I* locus located around the middle of the MLG are also related. Ancient tetraploidization and repeated duplication may be responsible for the evolution of the complex genetic loci of the CHS multigene family in soybean.

Communicated by F.J. Muehlbauer

H. Matsumura
Graduate School of Life and Environmental Sciences,
University of Tsukuba, Tsukuba, Ibaraki, 305-0006, Japan

S. Watanabe · K. Harada
Graduate School of Science and Technology,
Chiba University, Matsudo, Chiba, 271-8510, Japan

M. Senda · S. Akada
Gene Research Center, Hirosaki University,
Hirosaki, 036-8561, Japan

S. Kawasaki
National Institute of Agrobiological Sciences,
Tsukuba, Ibaraki, 305-8602, Japan

E. G. Dubouzet · R. Takahashi (✉)
National Institute of Crop Science,
Tsukuba, Ibaraki, 305-8518, Japan
E-mail: masako@affrc.go.jp
Tel.: +81-29-838-8503
Fax: +81-29-838-8503

N. Minaka
National Institute for Agro-Environmental Sciences,
Tsukuba, Ibaraki, 305-8604 Japan

Present address: H. Matsumura
BioResource Center RIKEN, Tsukuba Institute,
Tsukuba, Ibaraki, 305-0074 Japan

Introduction

Flavonoids are important secondary metabolites in the plant kingdom. Individual plant species synthesize a variety of flavonoid compounds that function in providing flower pigmentation to attract pollinators, in defending plants against pathogens, in acting as signal molecules in plant-microbe interactions, and in protecting plants from UV radiation (Dooner and Robbins 1991; Shirley 1996; Dixon and Steele 1999). The biosynthetic pathway of flavonoids is well-established, and many of the structural and some of the regulatory genes have been cloned in several model plants, including maize, petunia, and snapdragon (Holton and Cornish 1995).

Chalcone synthase (CHS; EC 2.3.1.74) is the key enzyme in the flavonoid biosynthetic pathway and catalyzes the stepwise condensation of *p*-coumaroyl-coenzyme A (CoA) derived from the phenylpropanoid pathway, with the three acetate moieties derived from malonyl-CoA to give rise to the C15 flavonoid skeleton,

naringenin chalcone. Isomerization and further substitution of this central intermediate product leads to the synthesis of a variety of flavones, isoflavonoids, and anthocyanins. CHS is encoded by a single-copy gene in some plants, such as *Petroselium crispm* (Herrmann et al. 1988), and by multiple copies of the gene in others, such as *Petunia hybrida* (Koes et al. 1987). In soybean [*Glycine max* (L.) Merrill], CHS is encoded by a multigene family, CHS1–CHS8 and dCHS1 (or ICHS1, a duplicate of CHS1) (Akada and Dube 1995; Todd and Vodkin 1996; Shimizu et al. 1999; Senda et al. 2002).

A multiple-allele locus, *I*/*i*-*i*/*i*-*k*/*i*, controls the distribution of the seed-coat colors (reviewed by Palmer et al. 2004). *I* results in the complete absence of coloration over the entire seed coat, *i*-*i* restricts color to the hilum region, *i*-*k* limits color to the hilum and a saddle-shaped region surrounding it, and *i* results in a self-colored seed coat. The dominance relationship among the alleles of the *I* locus are *I* > *i*-*i* > *i*-*k* > *i*. A tandem repeat of CHS1, CHS3 and CHS4 (the CHS1-3-4 cluster) spans approximately a 10-kb region. Genomic Southern analysis revealed that the presence of the dCHS1 band is associated with the dominant *I* allele (yellow hilum) and that dCHS1 is located in the close vicinity of the CHS1-3-4 cluster (Todd and Vodkin 1996). Sequence analysis has revealed that dCHS1 is closely linked to a truncated copy of CHS3 and that this cluster (dCHS1-Δ CHS3 cluster) arose from the CHS1-3 cluster through a 1.8-kb deletion (Senda et al. 2002). Furthermore, deletions in the promoter region of dCHS1 or CHS4 restore seed-coat coloration, suggesting that the CHS1-3-4 cluster and dCHS1 comprise the inhibitor *I* locus and that the expression of CHS genes may be inhibited by a naturally occurring, homology-dependent gene silencing process that leads to a complete absence of coloration (Todd and Vodkin 1996).

In Japan, soybean cultivars with a yellow hilum are preferred to those with a brown hilum for confectionery use on the basis of their better external appearance. Cultivars with a yellow hilum generally have a gray pubescence (*II tt*), whereas cultivars with a brown hilum have a brown pubescence (*i*-*ii*-*i TT*). Chilling temperatures (about 15°C) at flowering induce brown pigmentation around the hilum and cracking of the seed coats in yellow-hilum cultivars; these phenomena are absent in brown-hilum cultivars (Sunada and Ito 1982).

To clarify the genetic basis of varietal differences in chilling tolerance, Takahashi and Asanuma (1996) and Takahashi (1997) evaluated the roles of the *T* gene (responsible for pubescence and seed-coat color) encoding flavonoid 3′-hydroxylase (Buttery and Buzzell 1973; Toda et al. 2002) and of the *I* gene using near-isogenic lines for the two loci. Independent of the genotypes at the *I* locus, the dominant *T* allele completely suppressed the development of pigmentation around the hilum region and partly suppressed seed-coat cracking. The dominant *I* allele also suppressed seed-coat pigmentation and cracking in soybeans with the *tt* genotype, although its inhibitory effect was not as obvious as that of the *T* gene. Seed coats of soybeans with a recessive allelic combination of the two loci (*i* and *t*, or *i*-*k* and *t*) were severely cracked, irrespective of environmental conditions (Stewart and Wentz 1930; Nicholas et al. 1993). Chilling stress appears to surmount the gene-silencing function of the *I* allele and, consequently, causes both pigmentation and seed-coat cracking (Takahashi 1997). Consequently, the gene-silencing function of CHS is presumed to be related to chilling tolerance, but this has not been proved.

The CHS1-3-4 cluster corresponding to the *I* locus is located in the middle region of molecular linkage group (MLG) A2 (Cregan et al. 1999), and dCHS1 is presumed to be located close by (Todd and Vodkin 1996). In addition, clusters of CHS3-4, CHS1-3-5, and CHS3-5 have been identified (Akada and Dube 1995), but their locations are as yet unknown. Further, there is no information on the genomic locations of CHS2, CHS5–CHS8.

This present study was conducted to determine the location of each CHS member using cleaved amplified polymorphic sequence (CAPS) (Konieczny and Ausubel 1993), derived cleaved amplified polymorphic sequence (dCAPS) (Neff et al. 1998) or amplicon length polymorphism (ALP) analyses to investigate the function and evolution of CHS family members.

Materials and methods

Plant materials

Recombinant inbred lines (RILs) (F_9) were developed from a cross between a Japanese soybean [*Glycine max* (L.) Merrill] cultivar, Misuzudaizu, which has a yellow seed coat, gray pubescence and white flowers (*II-ttrrw1w1*) with a Chinese forage cultivar, Moshidou Gong 503, which has a brown seed coat, brown pubescence and purple flowers (*iiTTrrW1W1*). Genomic DNA was isolated from trifoliolate leaves of the parents and the 150 RILs using the CTAB method (Murray and Thompson 1980).

Construction of CAPS and dCAPS markers

Because the coding region of the CHS gene family is reported to be quite similar (Shimizu et al. 1999), we used varietal differences in the 5′-upstream region or in the first intron for investigating linkage mapping. PCR primers were constructed to clone the 5′-upstream regions for CHS1–CHS8 based on the DNA sequences of soybean cv. Williams deposited in the GenBank database (Table 1). The PCR mixture for the CAPS and dCAPS analysis contained 1 μl of the extracted DNA, 5 pmol of primers, 10 pmol of nucleotides, and 1 U Ex *Taq* in 1× Ex *Taq* buffer supplied by the manufacturer (Takara, Japan) in a total volume of 20 μl. A 30-s denaturation at 94°C was followed by 30 cycles of 30 s

Table 1 List of target CHS genes, PCR primers for analyzing their DNA sequences, annealing temperatures, amplified region and homology with sequences deposited in the database

Accession number	Target gene	Forward primer (5′– 3′)	Reverse primer (5′– 3′)	Annealing temperature (°C)	Amplified region	Homology (%)	
						Misuzu	MG503
X56749	CHS1-3	gggagtgtatcgtatcaaca	tgtgtaactacatacggtgg	60	5′-Upstream region	99.0	99.2
X65636	CHS2	ctgcagcaaaaatcatgaat	tagatagctagctagctaat	59	5′-Upstream region	99.9	99.7
X52097	CHS4	acgaccagtttctttcatga	gtatcgcttcttaatcatcg	58	5′-Upstream region −1st intron	99.5	99.5
L07647	CHS5	gaattatttttctgggcgat	tgtgaatgaactaatgaagc	53	5′-Upstream region	99.8	98.6
L03352	CHS6	agtgtgcaaaacatctatct	aatcattgacttatcacctg	60	First intron	98.7	99.7
M98871	CHS7	gagaaatttcagcgcatgtg	tgatcatagacttgtcacct	60	First intron	99.8	99.6
AY237728	CHS8	attgtgattgaatagcgtg	aggagtgctttgaacttgtt	56	5′-Upstream region	99.9	97.6
AB052783	dCHS1	ttctcaatattcttggagat	cctaacttattagtttttgc	58	CHS3-5′-Upstream region of CHS1	99.7	–

at 94°C (denaturation), 1 min at 53–60°C(annealing) depending on the primers (Table 1), and 1 min at 72°C (extension). A final 5-min extension at 72°C completed the program. The PCR was performed in a Perkin-Elmer 9700 thermal cycler (Foster City, Calif.). The PCR products were separated on 0.8% agarose gels and visualized by ethidium bromide staining. Those PCR products with the expected molecular size were cloned into pCR2.1 vector (Invitrogen, Carlsbad, Calif.) according to the manufacturer's instructions. Sequences were determined by BigDye terminator cycle method using a genetic analyzer (ABI3100; Applied Biosystems, Foster City, Calif.), and sequence homology was evaluated with the respective sequences deposited in the database using the BLAST software (Altschul et al. 1997). CAPS or dCAPS markers were constructed based on the varietal differences in the nucleotide sequences. If varietal differences did not exist in 5′-upstream regions, first introns were cloned and sequenced to find comparable varietal differences.

CAPS and dCAPS analyses

The PCR mixture for the CAPS and dCAPS analysis contained 1 μl of the extracted DNA, 5 pmol of primers, 10 pmol of nucleotides, and 1 U of Ex *Taq* in 1× Ex *Taq* buffer in a total volume of 20 μl. A 30-s denaturation at 94°C was followed by 30 cycles of 30 s at 94°C (denaturation), 1 min at 53–60°C (annealing) depending on the primers (Table 2), and 1 min at 72°C (extension). A final 5 min extension at 72°C completed the program. The PCR products were digested with the respective restriction enzymes (Table 2). The products from the parents and 150 RILs were separated on 8% acrylamide gels using the high efficiency genome scanning (HEGS) system (Kawasaki and Murakami 2000), and the fragments were visualized by ethidium bromide staining.

Construction of ALP markers for dCHS1

Primers flanking the 1.8-kb deletion in the dCHS1-Δ CHS3 cluster were constructed to discriminate between the CHS1-3 cluster and dCHS1-Δ CHS3 cluster. The CHS1–CHS3 cluster is expected to produce a 2.6-kb band, whereas the dCHS1-Δ CHS3 cluster is expected to produce a 0.8-kb band. The PCR products were separated on 0.8% agarose gels and stained in the same manner as described for the CAPS and dCAPS analyses.

Linkage mapping

Genotypes obtained from the CAPS, dCAPS and ALP analyses were combined with the genotypes of a total of 360 markers, including 177 restriction fragment length polymorphism (RFLP), 150 simple sequence repeat (SSR), 28 amplified fragment length polymorphism

Table 2 List of markers developed to map the CHS multigene family

Target gene (type of marker)	Forward primer (5′–3′)	Reverse primer (5′–3′)	Annealing temperature (°C)	PCR product length (bp)	Restriction enzyme	Digested product length (bp)	
						Misuzu	MG503
CHS1-3 (dCAPS)	gggagtgtatcgtatcaaca	gatggctgaattattcgtaaatgtct	60	331	*Xba*I	21, 313	331
CHS2 (CAPS)	ctgcagcaaaaatcatgaat	tagatagctagctagctaat	59	731	*Hpy*CH4III	98, 194, 439	292, 439
CHS4 (dCAPS)	ttatttccaacatatacctttgaatt	gtatcgcttcttaatcatcg	58	94	*Eco*RI	72, 94	94
CHS5 (CAPS)	gaattatttttctgggcgat	tgtgaatgaactaatgaagc	53	560	*Mun*I	235, 329	560
CHS6 (CAPS)	agtgtgcaaaacatctatct	aatcattgacttatcacctg	60	606	*Dra*I	178, 187, 241	187, 428
CHS7 (dCAPS)	tgttaatttggcacatctgaatt	tgatcatagacttgtcacct	60	398	*Eco*RI	398	23, 379
CHS8 (CAPS)	atttgtgattgaatagcgtg	aggagtgctttgaacttgtt	56	803	*Eco*T22I	257, 546	803
dCHS1 (ALP)	ttctcaatattcttggagat	cctaacttattagtttttgc	58	778	–	778, 2,533	2,533

Table 3 Marker name, specific sequences of selective primers, and approximate size of amplified bands in Moshidou Gong 503 (MG503) and Misuzudaizu (Misuzu) by AFLP analysis

Marker name	*Eco*RI primer	*Mse*I primer	Size of amplified bands (bp)	
			MG503	Misuzu
AFLP001	AGG	AAC	250	–
AFLP002	AGG	AAC	–	2,000
AFLP003	AGG	ATG	–	1,500
AFLP004	AGT	ACT	–	150
AFLP005	AGG	ATG	790	720
AFLP006	ACC	ACA	–	900
AFLP007	AGC	AAG	–	200

(AFLP) and five morphological markers to estimate genetic linkage using the Kosambi mapping function of MAPMAKER/EXP VER.3.0B (Lander et al. 1987). Primer sequences and band polymorphisms in the AFLP analysis are presented in Table 3. The markers were assigned to linkage groups with the criteria of LOD = 3.0 and a maximum distance of ≤ 37.2 cM.

Phylogenetic analysis

DNA sequences of CHS1–CHS8 (cv. Williams) were retrieved from the GenBank database (CHS1: x54644, CHS3: x53958, accession numbers of the other members are listed in Table 1). Phylogenetic analyses were performed based on the open reading frame sequences of CHS1–CHS8 using the maximum parsimony (MP), maximum likelihood (ML), and neighbor-joining (NJ) methods as implemented in PAUP* VER. 4 (Swofford 2002). The MODELTEST VER. 3 (Posada and Crandall 1998) was used to choose the best model of DNA sequence evolution among the ML and NJ trees.

Results and discussion

Construction of primers

Because CHS1 and CHS3 are closely located in head-to-head orientation (Todd and Vodkin 1996), we constructed a pair of primers for both CHS1 and CHS3. 5′-Upstream regions for CHS1-3, CHS2, and CHS4–CHS8 were cloned from Misuzudaizu and Moshidou Gong 503 by PCR based on the nucleotide sequences deposited in the database. Clones having inserts of the appropriate size were sequenced, and homology was ascertained. Sequence homology with those deposited in the database was more than 97%, suggesting that the expected regions were actually cloned (Table 1).

CAPS markers for CHS2, CHS5, and CHS8 were constructed based on the varietal differences in the 5′-upstream region of the nucleotide sequences (Table 2, Fig. 1). Because no varietal differences exist in the 5′-upstream region of CHS6 and CHS7, the first intron of the gene was cloned and sequenced, and CAPS markers for CHS6 and dCAPS markers for CHS7 were constructed based on the varietal differences (Fig. 2). The dCAPS markers for CHS4 were constructed based on varietal differences in the 5′-upstream region and the first intron. Electrophoretic patterns of the respective

CHS1-3 (-1898/-1560) <*Xba*I>

MG503 GGGAGTGTATCGTATCAACA----------CTTGGCAGAGCCTCCAAACATTTACGAATAATTCAGCCATC
Misuzu GGGAGTGTATCGTATCAACA----------CTTGGCAGAGCCTCTAAACATTTACGAATAATTCAGCCATC
← 3'-TCTGTAAATGCTTATTAAGTCGGTAG-5'

CHS2 (-733/-3) <*Hpy*CH4Ⅲ>
MG503 CTGCAGCAAAAATCATGAAT-----TAATAAAACTTTTTTTTAAAT-----ATTAGCTAGCTAGCTATCTA
Misuzu CTGCAGCAAAAATCATGAAT-----TAATAAAACTGTTTTTTAAAT-----ATTAGCTAGCTAGCTATCTA

CHS4 (296/389) <*Eco*RI>
MG503 TTATTTCCAACATATACCTTTGATTTGATTAATGATATC----------GTATCGCTTCTTAATCATCG
Misuzu TTATTTCCAACATATACCTTTGATTTCATTAATGATATC----------GTATCGCTTCTTAATCATCG
5'-TTATTTCCAACATATACCTTTGAATT-3' →

CHS5 (-575/-16) <*Mun*I>
MG503 GAATTATTTTTCTGGGCGAT-----GGAGAAGCAACTGTAAAGAAG-----GCTTCATTAGTTCATTCACA
Misuzu GAATTATTTTTCTGGGCGAT-----GGAGAAGCAATTGTAAAGAAG-----GCTTCATTAGTTCATTCACA

CHS6 (237/843) <*Dra*I>
MG503 AGTGTGCAAAACATCTATCT-----ATTATTTAGACTTAAATATGT-----CAGGTGATAAGTCAATGATT
Misuzu AGTGTGCAAAACATCTATCT-----ATTGTTTAGATTTAAATATGT-----CAGGTGATAAGTCAATGATT

CHS7 (273/641) <*Eco*RI>
MG503 TGTTAATTTGGCACATCTTAATTCTGTCTTCTCGTTACA----------TGATCATAGACTTGTCACCT
Misuzu TGTTAATTTGGCACATCTTAATTTTGTCTTCTCGTTACA----------TGATCATAGACTTGTCACCT
5'-TGTTAATTTGGCACATCTGAATT-3' →

CHS8 (-833/-96) <*Eco*T22I>
MG503 ATTTGTGATTGAATAGCGTG-----AGCTGGAGATACATAAAATAA-----AACAAGTTCAAAGCACTCCT
Misuzu ATTTGTGATTGAATAGCGTG-----AGCTGGAGATGCATAAAATAA-----AACAAGTTCAAAGCACTCCT

Fig. 1 CAPS and dCAPS markers mapping the CHS multigene family. The target gene name is shown in *bold*. The *numbers in parenthesis* refer to the location of the markers in the DNA sequence where the "A" of the start codon is numbered as *1*. The identity of the restriction enzyme used is in *<brackets>*, and the target sequences of the restriction enzymes are *underlined*. The dCAPS marker primers for the polymorphic regions between Moshidou Gong 503 (*MG503*) and Misuzudaizu (*Misuzu*) are denoted *below* the gene sequences. Mismatched nucleotides incorporated for dCAPS analyses are indicated by *double underlining*

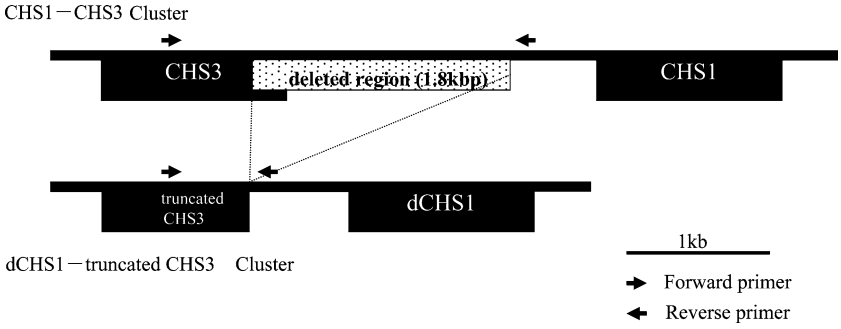

Fig. 2 Construction of the ALP marker to discriminate the dCHS1-truncated CHS3 cluster and the CHS1-3 cluster. Primers were designed to flank the 1.8-kb deletion in the dCHS1-truncated CHS3 cluster

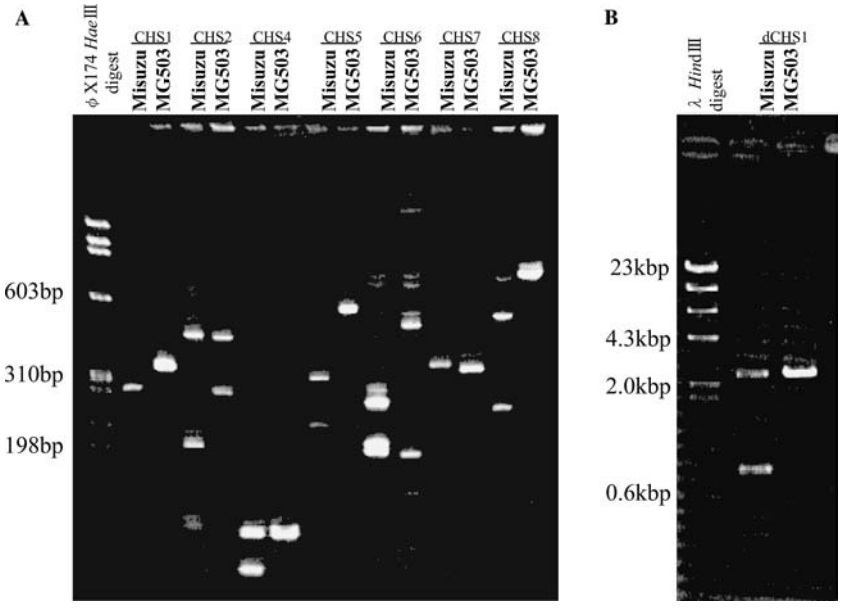

Fig. 3 Electrophoretic pattern of the CHS multigene family. **A** CAPS and dCAPS analysis using 8% polyacrylamide gel electrophoresis for Misuzudaizu (*Misuzu*) and Moshidou Gong 503 (*MG503*). **B** ALP analysis to detect polymorphism in dCHS1 by 0.8% agarose gel electrophoresis. The primers for CHS4 produced a 94-bp band that consisted of a mixture of CHS4 and CHS5 fragments. This 'mixed' CHS4 band in Misuzudaizu produced a 94-bp and a 72-bp band following digestion with *Eco*RI

PCR products are shown in Fig. 3. The primers for CHS4 produced a 94-bp band that consisted of a mixture of CHS4 and CHS5 band fragments because the DNA sequences of the amplified regions from CHS4 and CHS5 were quite similar. The 'mixed' CHS4 band of Misuzudaizu was digested with *Eco*RI, producing 94 bp of an undigested CHS5 band and a 72-bp CHS4 fragment (Fig. 3). The *Eco*RI site was absent from the CHS4

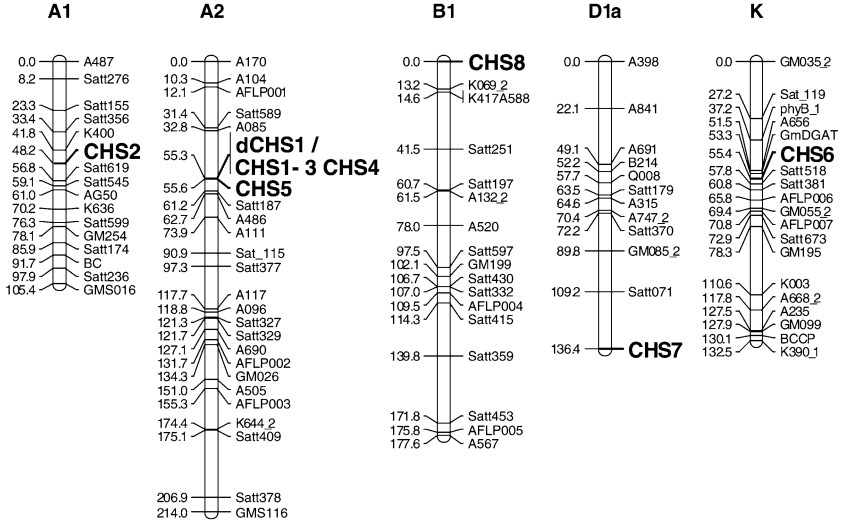

Fig. 4 Linkage map of the CHS multigene family in soybean. Only linkage groups containing CHS genes are exhibited. The name of each linkage group is indicated at the *top*. Distances between markers are shown on the *left* of each linkage group. CHS genes and the *I* locus are shown in *bold*. The *prefixes Satt*, *Sat*, or *AG* indicate SSR markers, the *AFLP prefix* indicates AFLP markers, *other prefixes* indicate RFLP markers

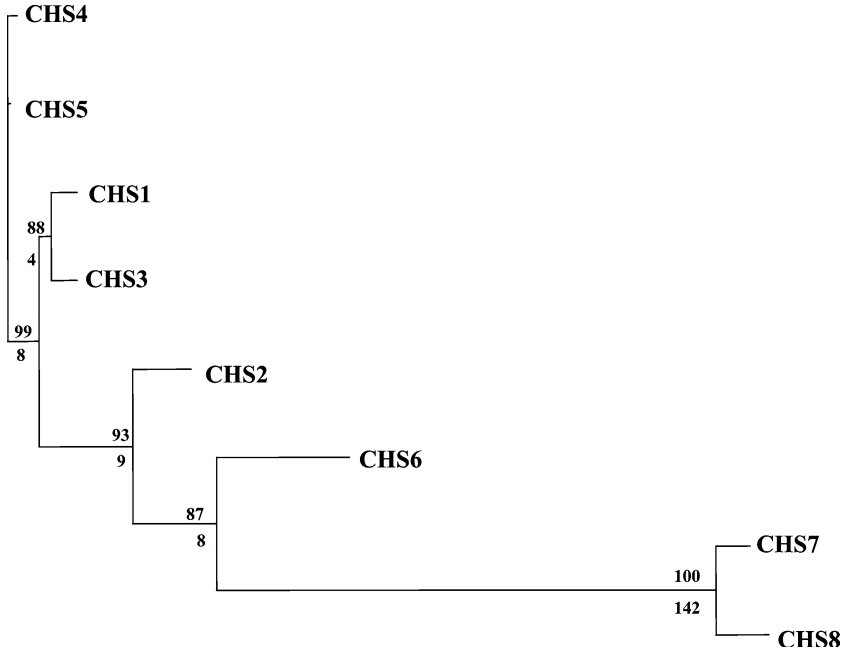

Fig. 5 The most parsimonious tree of the CHS multigene family in soybean. A branch-and-bound search for the best tree was conducted. *Numbers above the branches* are bootstrap probabilities (10,000 replicates), and those *below the branches* are decay indices for assessing the confidence levels of the tree. Calculations were performed using PAUP* 4.0 (Swofford 2002)

gene of Moshidou Gong 503. The dCHS1 primers produced an approximately 780-bp band present in Misuzudaizu with a yellow seed coat (*I*). As expected, this band was absent in Moshidou Gong503 with a brown seed coat (*i*).

Linkage mapping

One hundred and fifty RILs were used for linkage mapping with a total of 360 markers. The molecular linkage map spanned 2,664 cM, covering approximately 90% of the entire genome of soybean. Linkage mapping elucidated that CHS1, CHS3, CHS4, dCHS1, and the *I* locus were mapped at the same location on MLG A2 (Fig. 4). This observation is consistent with results reported by Todd and Vodkin (1996) that CHS1, CHS3 and CHS4 spans about a 10-kb region corresponding to the *I* locus and that genomic Southern bands corresponding to dCHS1 co-segregate with seed coat color . Among the 150 RILs tested, only one RIL exhibited recombination between the CHS gene cluster (CHS1-3, CHS4, dCHS1 and the *I* locus) and CHS5, suggesting that CHS5 is closely linked (0.3 cM distance) with the gene cluster, although the exact physical distance between the CHS1-3-4 cluster and CHS5 remains to be determined. Akada and Dube (1995) have also reported the presence of CHS3-5 and CHS1-3-5 clusters. Any possible association of the CHS5 mapped in this report with these gene clusters has yet to be determined. Linkage analysis revealed that CHS2 and CHS6 are located in the middle region of MLGs A1 and K, respectively, and that CHS7 and CHS8 are located at the distal end of MLGs D1a and B1, respectively (Fig. 4).

Phylogenetic analysis

The MP statistical method revealed a single most parsimonious tree (Fig. 5). Its bootstrap values and decay indices indicated that each clade of this MP tree had a high confidence level. The other methods (ML and NJ) gave similar tree shapes, including the same clades CHS7 + CHS8. However, their bootstrap confidence levels were much lower than those of the MP tree. The phylogenetic tree revealed that CHS members mapped around the *I* locus, CHS1, CHS3, CHS4, and CHS5 were closely related, thereby suggesting that gene duplication may have occurred repeatedly to form the suppressor *I* locus. Furthermore, CHS7 and CHS8, which were located to the distal end, and CHS2, CHS6, and the CHS members around the *I* locus that were located around the middle of the MLG are also related. Evolutionary studies suggest that soybean is an ancient tetraploid that later became diploidized (Hadley and Hymowitz 1973). Ancient tetraploidization may partly account for the fact that closely related CHS genes locate at similar chromosomal positions. Shoemaker et al. (1996), using RFLP markers, found that large portions of the soybean genome have undergone duplication, with more than one round of duplication having occurred, and genes conferring quantitative differences in seed composition showed concordance with the homologous regions. These investigators reported a large area of duplication between MLGs A1 and A2 that is consistent with the duplication between the CHS2 and CHS members around the *I* locus. Tetraploidization and repeated duplication may be responsible for the evolution of the complex genetic loci of the CHS multigene family in soybean.

Acknowledgements We thank Dr. Joseph G. Dubouzet (National Institute of Crop Science) for his critical reading of the manuscript. This study was partially supported by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Development of DNA Marker-aided Selection Technology for Plants and Animals).

References

Akada S, Dube SK (1995) Organization of soybean chalcone synthase gene clusters and characterization of a new member of the family. Plant Mol Biol 29:189–199

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Buttery BR, Buzzell RI (1973) Varietal differences in leaf flavonoids of soybeans. Crop Sci 13:103–106

Cregan PB, Jarvic T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic map of the soybean genome. Crop Sci 39:1464–1490

Dixon RA, Steele CL (1999) Flavonoids and isoflavonoids-a gold mine for metabolic engineering. Trends Plant Sci 4:394–400

Dooner HK, Robbins TP (1991) Genetic and developmental control of anthocyanin biosynthesis. Annu Rev Genet 25:173–199

Hadley HH, Hymowitz T (1973) Speciation and cytogenetics. In: Caldwell BE (ed) Soybeans: improvement, production, and uses, 1st edn. Agron Monogr 16. ASA, CSSA, SSSA, Madison, pp 97–116

Herrmann A, Schulz W, Harlbrock K (1988) Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. Mol Gen Genet 212:93–98

Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell 7:1071–1083

Kawasaki S, Murakami Y (2000) Genome analysis of *Lotus japonicus*. J Plant Res 113:497–506

Koes RE, Spelt CE, Mol JNM, Gerats AGM (1987) The chalcone synthase multigene family of *Petunia hybrida* (V30): sequence homology, chromosomal localization and evolutionary aspects. Plant Mol Biol 10:375–385

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Lander ES, Green P, Abrahamson J, Barlow A, Day MJ, Lincoln SE, Newberg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage map of experimental and natural populations. Genomics 1:174–181

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J 14:387–392

Nicholas CD, Lindstrom JT, Vodkin LO (1993) Variation of proline rich cell wall proteins in soybean lines with anthocyanin mutations. Plant Mol Biol 21:145–156

Palmer RG, Pfeiffer TW, Buss GR, Kilen TC (2004) Qualitative genetics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. Agron Monogr 16. ASA, CSSA, SSSA, Madison, pp 137–233

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Senda M, Jumonji A, Yumoto S, Ishikawa R, Harada T, Niizeki M, Akada S (2002) Analysis of the duplicated *CHS1* gene related to the suppression of the seed coat pigmentation in yellow soybeans. Theor Appl Genet 104:1086–1091

Shimizu T, Akada S, Senda M, Ishikawa R, Harada T, Niizeki M, Dube SK (1999) Enhanced expression and differential inducibility of soybean chalcone synthase gene by supplemental UV-B in dark grown seedlings. Plant Mol Biol 39:785–795

Shirley BW (1996) Flavonoid biosynthesis: new functions for an old pathway. Trends Plant Sci 1:377–382

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olsen T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338

Stewart RT, Wentz JB (1930) A defective seed-coat character in soybeans. J Am Soc Agron 22:658–662

Sunada K, Ito T (1982) Soybean grain quality as affected by low temperature treatments in plants (color of hilum, seed coat cracking) (in Japanese). Rep Hokkaido branch, Crop Sci Soc Jpn and Hokkaido branch. Jpn Soc Breed 22:34

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and Other Methods). Version 4.0b10. Computer software. Sinauer Associates Sunderland MA, USA

Takahashi R (1997) Association of soybean genes *I* and *T* with low-temperature induced seed coat deterioration. Crop Sci 37:1755–1759

Takahashi R, Asanuma S (1996) Association of *T* gene with chilling tolerance in soybean. Crop Sci 36:559–562

Toda K, Yang D, Yamanaka N, Watanabe S, Harada K, Takahashi R (2002) A single-base deletion in soybean flavonoid 3′-hydroxylase gene is associated with gray pubescence color. Plant Mol Biol 50:187–196

Todd JJ, Vodkin LO (1996) Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell 8:687–699