DNA Research **14**, 25–36, (2007) doi:10.1093/dnares/dsm004

Genome-wide Analyses of the Structural Gene Families Involved in the Legume-specific 5-Deoxyisoflavonoid Biosynthesis of *Lotus japonicus*

Norimoto Shimada,[1,*] Shusei Sato,[2] Tomoyoshi Akashi,[1] Yasukazu Nakamura,[2] Satoshi Tabata,[2] Shin-ichi Ayabe,[1] and Toshio Aoki[1]

Department of Applied Biological Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-8510, Japan[1] and Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan[2]

(Received 17 January 2007; revised March 11, 2007; published online 23 April 2007)

Abstract

A model legume *Lotus japonicus* (Regel) K. Larsen is one of the subjects of genome sequencing and functional genomics programs. In the course of targeted approaches to the legume genomics, we analyzed the genes encoding enzymes involved in the biosynthesis of the legume-specific 5-deoxyisoflavonoid of *L. japonicus*, which produces isoflavan phytoalexins on elicitor treatment. The paralogous biosynthetic genes were assigned as comprehensively as possible by biochemical experiments, similarity searches, comparison of the gene structures, and phylogenetic analyses. Among the 10 biosynthetic genes investigated, six comprise multigene families, and in many cases they form gene clusters in the chromosomes. Semi-quantitative reverse transcriptase–PCR analyses showed coordinate up-regulation of most of the genes during phytoalexin induction and complex accumulation patterns of the transcripts in different organs. Some paralogous genes exhibited similar expression specificities, suggesting their genetic redundancy. The molecular evolution of the biosynthetic genes is discussed. The results presented here provide reliable annotations of the genes and genetic markers for comparative and functional genomics of leguminous plants.

Key words: 5-deoxyflavonoid; gene cluster; isoflavonoid; *Lotus japonicus*; phytoalexin

1. Introduction

The last decade has seen remarkable advances in structural analysis of model plant genomes. The complete whole-genome sequence data of *Arabidopsis thaliana* and *Oryza sativa* have been published together with the predicted gene structures and annotations,[1,2] and the draft genome sequence of a model tree *Populus trichocarpa* (Torr. & Gray) was recently disclosed.[3] Two leguminous plants, *Lotus japonicus* (Regel) K. Larsen and *Medicago truncatula* Gaertn., are also subjects of genome sequencing programs.[4] The post-sequence functional genomic approaches to legumes will be important to uncover new functions of legume-specific genes and to establish the systems biology of legumes. At present, non-targeted approaches are mainly used for legume functional genomics. The EST macro-array analysis of *L. japonicus* became available recently.[5] Data mining by multivariate analyses of the integrated data of transcriptomics and metabolomics is expected to facilitate our understanding of the physiology of legumes, as is being accomplished with *A. thaliana*.[6] On the other hand, targeted approaches, that is, focusing on the genes of particular families or particular pathways, are practical to assess the functions and structures of individual genes, and thus they will complement the non-targeted ones.

The legume-specific 5-deoxyisoflavonoid pathway is one of the targets for functional legume genomics. 5-Deoxyisoflavonoids have two distinctive structural features: the B-ring attached to C-3 of the C-ring and a hydrogen attached to C-5 instead of a B-ring at C-2 and a hydroxyl group at C-5 in general flavonoids. To the best of our knowledge, no 5-deoxyisoflavonoid has been found in any plant other than the Leguminosae. About

Communicated by Kazuki Saito

* To whom correspondence should be addressed. Tel. +81 466-84-3703, Fax. +81 466-84-3353, E-mail: nshima18@brs.nihon-u.ac.jp

95% of isoflavonoids are found in legumes, and 60% of leguminous flavonoids are 5-deoxy series flavonoids.[7] Some isoflavonoids, particularly those with pterocarpan and isoflavan skeletons, are typical leguminous phytoalexins active in the defense response against phytopathogenic organisms,[8] while an isoflavone (daidzein) functions as a signal molecule toward symbiotic rhizobia to establish nitrogen-fixing root nodules in the soybean.[9]

Vestitol, an isoflavan phytoalexin of *Lotus* and other leguminous genera, is postulated to be biosynthesized from the phenylpropanoid precursor by a total of 11 enzymes (Fig. 1). The cDNAs encoding 10 of the 11 enzymes have been identified from *L. japonicus* or other leguminous plants;[10–18] the exception is 7,2′-dihydroxy-4′-*O*-methoxyisoflavanol dehydratase, which has been purified from *M. sativa*,[19] but for which no sequence data is available. We have reported the induction of vestitol accumulation by treatment with reduced glutathione (GSH)[20] and the structures and functions of paralogous genes encoding catalytically distinct chalcone isomerase (CHI) isozymes of *L. japonicus*.[21] The findings on the genes for CHI, together with those for dihydroflavonol-4-reductase[22] of the general flavonoid pathway and oxidosqualene cyclase[23] in the triterpenoid biosynthesis, suggest the functional diversification of the multigene families, due to gene duplication, followed by the accumulation of nucleotide substitutions.[21]

In the present study, we clarified the structures of the genes encoding the enzymes of the legume-specific 5-deoxyisoflavonoid pathway as comprehensively as possible, and semi-quantified the transcripts during phytoalexin induction and in several organs. The gene identification was based on a biochemical assay using heterologously expressed enzymes, comparison of the gene structures, and the phylogenetic relationship. The results offer reasonable gene annotations and genetic markers for comparative and functional genomics of legumes, as well as insights into the molecular evolution of the legume-specific biosynthetic genes.

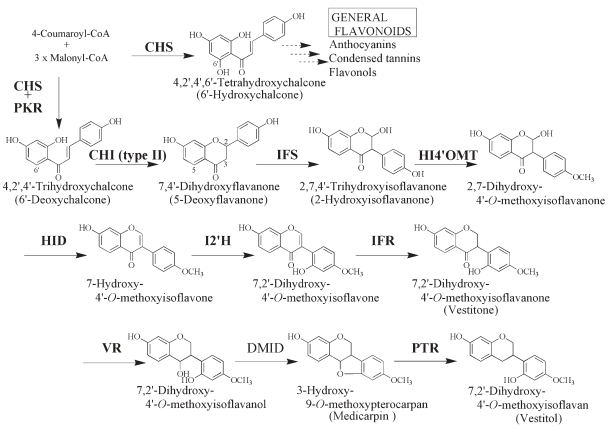

Figure 1. General and legume-specific flavonoid pathways in *L. japonicus*. Enzymes involved in vestitol biosynthesis are shown in bold. CHS, chalcone synthase; CHI, chalcone isomerase; DMID, 7,2′-dihydroxy-4′-*O*-methoxyisoflavanol dehydratase; HI4′OMT, 2,7,4′-trihydroxyisoflavanone 4′-*O*-methyltransferase; HID, 2-hydroxyisoflavanone dehydratase; I2′H, isoflavone 2′-hydroxylase; IFR, isoflavone reductase; IFS, 2-hydroxyisoflavanone synthase; PKR, polyketide reductase; PTR, pterocarpan reductase; VR, vestitone reductase.

2. Materials and methods

2.1. Genome structure and genetic mapping

Generation of genomic libraries of *L. japonicus* (Regel) K. Larsen accession Miyakojima MG-20, sequence strategy, and gene assignment were carried out as described elsewhere.[24] The libraries were screened for each gene by the PCR method using primer sets based on the sequences of the cDNAs isolated (Table 1). Putative genes for chalcone synthase (CHS) were in part discovered in the course of the whole-genome sequencing program of *L. japonicus*.[25] Genetic mapping was carried out with the simple sequence repeat markers found in each TAC clone.[26]

2.2. Functional characterization of paralogous genes

mRNA was isolated from 4-week-old whole plants of *L. japonicus* accession Gifu B-129 using the Straight A's mRNA isolation system (Novagen, Madison, WI, USA). Single-strand cDNAs were synthesized using Superscript II RNase H$^-$ Reverse Transcriptase (Invitrogen Corp., Carlsbad, CA, USA). To amplify the full ORF of *2-hydroxyisoflavanone synthase 2* (*IFS2*, GenBank accession no. AB279984), a primer set was designed based on the genomic sequence: 5′-AAGATCAAACAAAGCTTATG TTGGTGGAAC-3′ and 5′-GAAATCTAGAAGGTT ATTTGTGTTGTACTT-3′, which included *Hin*dIII and *Xba*I sites (underlined), respectively. The PCR product was cloned into a pT7Blue T-vector (Novagen). EST clones corresponding to *polyketide reductase 5* (*PKR5*, GenBank accession no. AV424286; synonymous with *chalcone reductase*, *CHR*) and *2-hydroxyisoflavanone dehydratase* (*HID*, GenBank accession no. AV425769) are in the collection of Kazusa DNA Research Institute. The construction of expression vectors, heterologous expression, and *in vitro* assays of IFS and HID have been described previously.[13,27] The PKR activity was tested by assaying 6′-deoxychalcone synthase activity as described previously,[28] except that the CHS source was the enzyme solution prepared from recombinant *Escherichia coli* cells expressing a cDNA for CHS obtained from licorice, instead of plant cell-free extracts.

2.3. Semi-quantitative RT–PCR

To analyze the expression patterns of each paralogous gene in GSH-treated seedlings and various organs, specific primers were designed (Table 1). mRNA isolation from *L. japonicus* Gifu B-129 and first-strand cDNA synthesis were performed as described previously.[20,22] The absence

Table 1. Primers used for screening of *L. japonicus* genome libraries and/or RT–PCR analysis

Target		Primer sequence
Gene (genome clone)	cDNA	
PKR (LjT38O12)		5′-TGGAAAGACTATAGCTCAGGTCAG-3′
		5′-TTTAAACCCAAAAGGAAAAGACAGC-3′
PKR (LjT10F22)		5′-GCAGCCAGTAAGGTTAGCACAAAAG-3′
		5′-GCAGGAACAATGAGATGAGGATG-3′
	PKR1	5′-TTCTCACCTTTGAGGAAAGGTGG-3′[a]
		5′-AGCTCTACTTTTCAGGATAAGATT-3′
	PKR2	5′-TTTTCACCTCTAAGGAAAGGAGC-3′
		5′-TAAAATAGGTTGCGACATCAAATA-3′
	PKR4	5′-AGGAATTAATTCAAGGCAAACATC-3′
	PKR5	5′-ATTGCTCTCTTTTGCTACCATCACA-3′[b]
		5′-TCAATTCAGAACGTTTACAGCTTA-3′
	PKR6	5′-CACCAAAAAAAAAGACAAAGTATTG-3′
IFS (LjT24P23)		5′-AGAAAGTGTGTACAAGAGTGTGAG-3′
		5′-AGGAACACACATGAGACTATTTGC-3′
IFS (LjT46B17)		5′-CCATATAAAAGCATTCGCAAAG-3′
		5′-AAGTTAGCATGGAAGAGAGTC-3′
	IFS1	5′-CAGAGAGGTTTTTGGAAGAAGCAGA-3′
		5′-AAACCCTCGAGCACAAAGCAACAT-3′
	IFS2	5′-CCGTCCTGAGAGGTTTTTAACTGCT-3′
		5′-GAAAACATGTAGGTTATTTGTGTTGTACTT-3′
HID (LjB01D01)		5′-GCAGGAAGGTTTGGAACTTTGTG-3′
		5′-GGATGGTAAATCTGGAAGCAATGC-3′
	HID	5′-TCAAAAGACATAGTCATCTCACAA-3′
		5′-TGTAGAATCTGCTGAAATCACCAT-3′
I2′H (LjT07D18)	*I2′H*	5′-AATGGTACCATGGATATCATCTCCTTCCTT-3′
		5′-CTCACTCTAGAAACATGTCCCCGATTCAAA-3′
IFR (LjT32H22)		5′-GTTCTGGATCCTGATGGCACCACAAGACAG-3′
		5′-TCACGAGAGGACAGGCTCGAGTAACAAACA-3′
	IFR1	5′-GAACAAGGCTGTCCACATAAGACT-3′[c]
		5′-GTTTTATTGAGCAATCTGCAGAGT-3′
	IFR2	5′-AATGGAAATAGTTTAAAGCCTGAT-3′
VR (LjT43J18)		5′-GGAAGATACAACTGCTCCCCATTC-3′
		5′-GCATCATCAAACAAGTCCTCAAGG-3′
	VR1	5′-CCTCACAAATCTACCTGGTG-3′[d]
		5′-AGCTGAGGACAAATCCATCT-3′
	VR2	5′-AGCTTAGGACAAACCCAAGG-3′
PTR (LjT44D07)		5′-TTGAAGATGAATGCTTCCTTTACTG-3′
		5′-AAACTAATAATGTTTGTTGTACACC-3′
	PTR1	5′-CTTTCCAACCCACAAAAATCAA-3′
		5′-GACACAAATATGCAGCGTGGCA-3′
	PTR2	5′-CTTTCTAGCCCTGCAAAAGCTT-3′
		5′-CCTTCACAAACGCAGCGTGAGC-3′
	PTR3	5′-CTTTCCGATCCAGCAAGATCAC-3′
		5′-CTTTCACATAGACAGCATGACT-3′
PTR (LjT37D14)		5′-GAGACAGGGCAATTCCAAAG-3′

Continued

Table 1. Continued

Target		Primer sequence
Gene (genome clone)	cDNA	
		5′-GAGGAGAAGCTGCTCAAGGA-3′
	PTR4	5′-CTCTCTGACACTGCTAAAGCTC-3′
		5′-CCTTCACGAACACCGCGTGGTT-3′
	β-tubulin	5′-AATTCTGGGAAGTCATCTGCGACGAG-3′
		5′-CTGGTGCACTGAAAGCGTAGCATTAT-3′

[a]Common primer for the amplification of *PKR1* and *PKR4* cDNAs.
[b]Common primer for the amplification of *PKR5* and *PKR6* cDNAs.
[c]Common primer for the amplification of *IFR1* and *IFR2* cDNAs.
[d]Common primer for the amplification of *VR1* and *VR2* cDNAs.

of contamination with genomic DNA in the cDNA samples was confirmed by PCR with a primer set designed to overlap one of the exon–intron junctions of the *isoflavone reductase 1* (*IFR1*) gene. The quantity of each template was adjusted to give roughly equal amplification of *β-tubulin* cDNA. RT (reverse transcriptase)–PCR was carried out with 0.5 pmol of each specific primer using *ExTaq* DNA polymerase (Takara-Bio Inc., Shiga, Japan) in a final volume of 20 μl according to the manufacturer's protocol. The products (5 μl) were separated on 1.2% (w/v) agarose gel and stained with ethidium bromide.

2.4. Construction of phylogenetic tree

Neighbor-joining trees were generated from the results of 1000 bootstrap replicates using the CLUSTAL W program[29] of the DNA Data Bank of Japan (Shizuoka, Japan). The phylogenetic trees were displayed by NJplot software.[30]

A *CHS* phylogenetic tree was constructed based on the coding sequences of *CHS*s and plant polyketide synthases, i.e. *acridone synthase*, *bibenzyl synthase*, and *stilbene synthase*. GenBank accession numbers for the nucleotide sequences of polyketide synthases are *Antirrhinum majus CHS* (X03710), *Arabidopsis thaliana CHS* (M20308), *Camellia sinensis CHS1* (D26593), *Cicer arietinum CHS* (AJ012690), *Daucus carota CHS2* (AJ006780), *Gerbera hybrida CHS1* (Z38096), *Glycine max CHS2* (X65636), *G. max CHS3* (X53958), *G. max CHS4* (X52097), *G. max CHS5* (L07647), *G. max CHS6* (L03352), *G. max CHS7* (M98871), *G. max CHS8* (AY237728), *Malus x domestica CHS1* (DQ026297), *Ipomoea purpurea CHSA* (U15946), *I. purpurea CHSB* (U15947), *Lycopersicon esculentum CHS1* (X55194), *L. esculentum CHS2* (X55195), *Medicago sativa CHS1* (L02901), *M. sativa CHS2* (L02902), *M. sativa CHS4* (L02903), *M. sativa CHS8* (L02904), *M. sativa CHS9* (L02905), *M. sativa CHS12-1* (U01021), *Perilla frutescens CHS* (AB002815), *Petunia x hybrida CHS* (AF233638), *Phalaenopsis* sp. *bibenzyl synthase* (X79903), *Pinus strobus CHS* (AJ004800), *P. sylvestris bibenzyl synthase* (X60753), *P. strobus stilbene synthase* (Z46914), *Pisum sativum CHS2* (X63334), *Pueraria lobata CHS* (D10223), *Ruta graveolens acridone synthase* (Z34088), *Trifolium subterraneum CHS1* (M91193), *T. subterraneum CHS2* (M91194), *T. subterraneum CHS3* (L24515), *T. subterraneum CHS5* (L24517), *T. subterraneum CHS6* (M91195), *Vitis stilbene synthase* (S63221), and *Vitis vinifera CHS* (X75969).

A phylogenetic tree of *PKR* cDNAs was constructed based on the 384 bp region from the initiation codon of *aldo/keto reductase* (*AKR*) and *PKR/CHR*, because full-length ORF sequences of EST clones were unavailable. GenBank accession numbers for the nucleotide sequences used for the construction of a *PKR* phylogenetic tree are *A. thaliana AKR* (NM104687), *Fragaria x ananassa AKR* (AY703448), *G. max PKR* (X55730), *G. max PKR*-like (BG882535), *Glycyrrhiza echinata PKR* (D83718), *Glycyrrhiza glabra PKR2* (D86559), *Hydrangea macrophylla AKR* (AY382665), *M. sativa CHR* (X82368), *M. truncatula PKR*-like (TC100399), *Papaver somniferum codeinone reductase* (AF108435), *P. lobata CHR* (AF462632), *Oryza sativa AKR* (XM462652), and *Sesbania rostrata CHR* (AJ223291).

A phylogenetic tree of the genes for short-chain dehydrogenase/reductase (SDR) was constructed based on full-length ORFs of *IFR*, *leucoanthocyanidin reductase*, *phenylcoumaran benzylic ether reductase* (*PCBER*), *pinoresinol–lariciresinol reductase* (*PLR*), and *pterocarpan reductase* (*PTR*). GenBank accession numbers for the nucleotide sequences of these are *C. arietinum IFR* (X60755), *Desmodium uncinatum leucoanthocyanidin reductase* (AJ550154), *Forsythia x intermedia PLR* (U81158), *G. max IFR* (AJ003245), *M. sativa IFR* (X58078), *M. truncatula IFR* (AF277052), *Pinus taeda PCBER* (AF242490), *P. sativum IFR* (S72472), *Populus trichocarpa PCBER* (AJ132262), *Tsuga heterophylla PLR1* (AF242501), *T. heterophylla PLR2* (AF242502), *Thuja plicata PLR1* (AF242503), *T. plicata PLR2* (AF242504), *T. plicata PLR3* (AF242505), and *T. plicata PLR4* (AF242506).

Table 2. Summary of genes involved in vestitol biosynthesis of *L. japonicus*

Gene	Genome clone (accession number)	Location (initiation → termination codon)	cDNA accession	Map position	Biochemical identification
CHS1	LjT09H02 (AP006710)	6693 → 8287	BP033951[a]	Chr. I, 51.8cM	
CHS2	LjT30A24 (AP009237)	47359 ← 48662	BP034657[a]	Chr. I, 26.1cM	
CHS3	LjT30A24 (AP009237)	51392 → 52763	AV767686[a]	Chr. I, 26.1cM	
CHS4	LjT01O03 (AP006702)	14266 ←15581		Chr. II, 61.7cM	
CHS5	LjT01O03 (AP006702)	9231 → 10524		Chr. II, 61.7cM	
CHS6	LjT01O03 (AP006702)	5437 ←6749		Chr. II, 61.7cM	
CHS7	LjT08P04 (AP006701)	8189 → 9571	BP076978[a]	Chr. II, 62.1cM	
	LjT41D19 (AP006709)	47678 → 49060			
CHS8	LjT08P04 (AP006701)	3147 → 4437		Chr. II, 62.1cM	
	LjT41D19 (AP006709)	42636 → 43926			
CHS9 (pseudogene)	LjT41D19 (AP006709)	32690 → 34006		Chr. II, 62.1cM	
CHS10	LjT41D19 (AP006709)	26974 → 28272		Chr. II, 62.1cM	
CHS11	LjT10F20 (AP007302)	17404 ← 18703	BP083372[a]	Chr. II, 62.1cM	
CHS12	LjT09I23 (AP006706)	8878 ← 10177	BP051174[a]	Chr. III, 13.3 cM	
CHS13	LjT03B03 (AP004528)	8283 → 9582		Chr. VI, 27.6 cM	
PKR1	LjT38O12 (AP009072)	80096 → 81290	AB263016	Chr. I (B-129)	Shimada et al.[31]
				Chr.II (MG-20, 72.5 cM)	
PKR2	LjT38O12 (AP009072)	68792 → 69986		Chr. I (B-129)	
				Chr.II (MG-20, 72.5 cM)	
PKR3 (pseudogene)	LjT38O12 (AP009072)	55902 → 62367		Chr. I (B-129)	
				Chr.II (MG-20, 72.5 cM)	
PKR4	LjT38O12 (AP009072)	23657 ← 24849	AW428662[a]	Chr. I (B-129)	
				Chr.II (MG-20, 72.5 cM)	
PKR5	LjT10F22a (AP009238)	54908 ← 56168	AV424286[a]	n.d.	This work
PKR6	LjT10F22a (AP009238)	62291 → 63545	AV407669[a]	n.d.	
CHI1	LjT47K21 (AP004250)	17264 → 18700	AB054801	Chr. V, 1.2 cM	Shimada et al.[21]
CHI2	LjT47K21 (AP004250)	26521 ← 28478	AB054802	Chr. V, 1.2 cM	Shimada et al.[21]
CHI3	LjT47K21 (AP004250)	14079 → 15609	AB073787	Chr. V, 1.2 cM	Shimada et al.[21]
CHI4	LjT47K21 (AP004250)	24721 → 25965		Chr. V, 1.2 cM	
IFS1	LjT24P23B (AP009063)	31597 ← 33467	AB024931	n.d.	Shimada et al.[20]
IFS2	LjT46B17B (AP009070)	53978 → 55659	AB279984	Chr. IV, 11.6 cM	This work
IFS3 (pseudogene)	LjT46B17C (AP009071)	2325 → 3979		Chr. IV, 11.6 cM	
HI4′OMT	LjT24P23C (AP009064)	13643 → 15476	AB091686	n.d.	Akashi et al.[11]
HID	LjB01D01 (AP009065)	80214 → 81203	AV425769[a]	n.d.	This work
I2′H	LjT07D18 (AP009066)	77648 ← 80381	AB025016	Chr. IV, 5.6 cM	Shimada et al.[20]
IFR1	LjT32H22B (AP009068)	4305 ← 5851	BP081283[a]	Chr. II, 56.1 cM	
IFR2	LjT32H22B (AP009068)	9850 ← 11396	AB265595	Chr. II, 56.1 cM	
VR1	LjB20B09 (AP009074)	27385 ← 29733	BP073948[a]	Chr. I, 23.3 cM	
VR2	LjB20B09 (AP009074)	21816 ← 23749	BP051977[a]	Chr. I, 23.3 cM	

Continued

Table 2. Continued

Gene	Genome clone (accession number)	Location (initiation → termination codon)	cDNA accession	Map position	Biochemical identification
VR3 (pseudogene)	LjT43J08 (AP009073)	124819 ←Not found		Chr. I, 23.3 cM	
PTR1[b]	LjT44D07 (AP009075)		AB265589	Chr. III, 75.6 cM	Akashi et al.[12]
PTR2[b]	LjT44D07 (AP009075)		AB265590	Chr. III, 75.6 cM	Akashi et al.[12]
PTR3[b]	LjT44D07 (AP009075)		AB265591	Chr. III, 75.6 cM	Akashi et al.[12]
PTR4	LjT37D14 (AP009195)	58138 ← 59854	AB265592	Chr. I, 17.7 cM	Akashi et al.[12]

n.d., not determined.

[a]EST clones.

[b]Positions of *PTR1-3* genes have not been determined due to the high number of repeated sequences in LjT44D07.

3. Results

3.1. Assignment of legume-specific 5-deoxyisoflavonoid genes

The structural genes involved in vestitol biosynthesis except for *CHS* genes were found in the genomic clones isolated by PCR-based screening of *L. japonicus* genome libraries.[24] Some of the *CHS* genes had already been reported,[25] and, in the present study, the other paralogous genes were discovered in the genomic sequences obtained in the course of the whole-genome sequencing program of this plant.[24] The chromosomal localization and accession numbers of all genes identified are summarized in Table 2. Those that have been functionally characterized before are *PKR1*, *CHI1*, *CHI3*, *IFS1*, *2,7,4′-trihydroxyisoflavanone 4′-O-methyltransferase* (*HI4′OMT*), *isoflavone 2′-hydroxylase* (*I2′H*), and *PTR1-PTR4*.[11–13,20,31] The genes for PKR5, HID, and another IFS (IFS2) were functionally assigned in this study by *in vitro* assays using the heterologously expressed proteins (Supplementary Fig. S1). The other paralogous genes and putative genes for CHS, IFR, and vestitone reductase (VR) were assigned based on the sequence similarity to those identified from other plants, the resemblance of exon/intron structures, and the phylogenetic relationship of the predicted coding sequences as described in what follows. *IFS1*, *I2′H*, and *IFR1* correspond to the previously reported cDNAs, *LjCYP-1*, *LjCYP-2*, and *LjIFR*, respectively.[20]

3.2. Structures of isoflavonoid biosynthetic genes

3.2.1. CHS Structural analysis of putative *CHS* genes of *L. japonicus* revealed 13 paralogs, named *CHS1–CHS13* (Fig. 2A). Two clusters, composed of three (*CHS4*, *CHS5*, and *CHS6*) and five (*CHS7*, *CHS8*, *CHS9*, *CHS10*, and *CHS11*) genes, respectively, are found on chromosome II. *CHS9* should be a pseudogene because the initiation codon (ATG) is changed to ATA by a transition. The other five *CHS* genes, *CHS1* to *CHS3*, *CHS12* and *CHS13*, are separately located on chromosome I, III, and VI, respectively. The structure of the *CHS* genes is highly conserved: all the *CHS* genes are composed of two exons, i.e. the first exon (178 bp) and second exon (992 bp), except for *CHS1*, which has a second exon of 998 bp.

3.2.2. PKR Four *PKR* genes, *PKR1–PKR4*, compose a cluster in a 50 kb region of a genomic clone LjT38O12 mapped in the south end of chromosome II in accession Miyakojima MG-20 and chromosome I in accession Gifu B-129 (Fig. 2B), which is known as the reciprocal translocated region.[26] They are composed of three exons, and the lengths of the first (323 bp) and second (244 bp) exons are conserved. The first exon of *PKR3* contains ca. 5.3 kb sequence in which a terminal-repeat retrotransposon in miniature-like sequence is found, and thus *PKR3* should be a pseudogene. The coding sequences of *PKR1*, *PKR2*, and *PKR4* are more than 90% identical to each other. Two additional *PKR* genes, *PKR5*, and *PKR6*, are found in LjT10F22a, whose chromosome location is unknown. *PKR5* and *PKR6* are composed of four exons, and the corresponding exons of the two genes are identical in length.

3.2.3. IFS The *IFS1* gene encoding CYP93C17[20] is found in LjT24P23, and *IFS2* and *IFS3* are located in LjT46B17 (Fig. 2C). The map position of LjT24P23 has not been clarified because no linkage marker was found in LjT24P23.

The multiple alignments of the *IFS* sequences revealed insertion/deletion (I/D) polymorphisms among the *IFS* genes. Fig. 3 shows the alignments of a partial nucleotide (A) and amino acid (B) sequences of IFS from *L. japonicus* (*IFS1–IFS3*) together with two genes for functionally identified IFS, CYP93C2 of *Glycyrrhiza echinata*[18] and CYP93C1v2 of *Glycine max.*[17] The cladistic relationships of the *IFS* genes were elucidated using the I/D polymorphisms (Fig. 3C), demonstrating that the *IFS3* of *L. japonicus* is the most derived. In *IFS3*, a frame shift is caused by a 4-bp insertion at 1217, and a non-sense codon (TGA) is formed next to His427 (asterisk in Fig. 3B). The heme-binding amino acid residue, which is critical for the activity of cytochrome P450, was

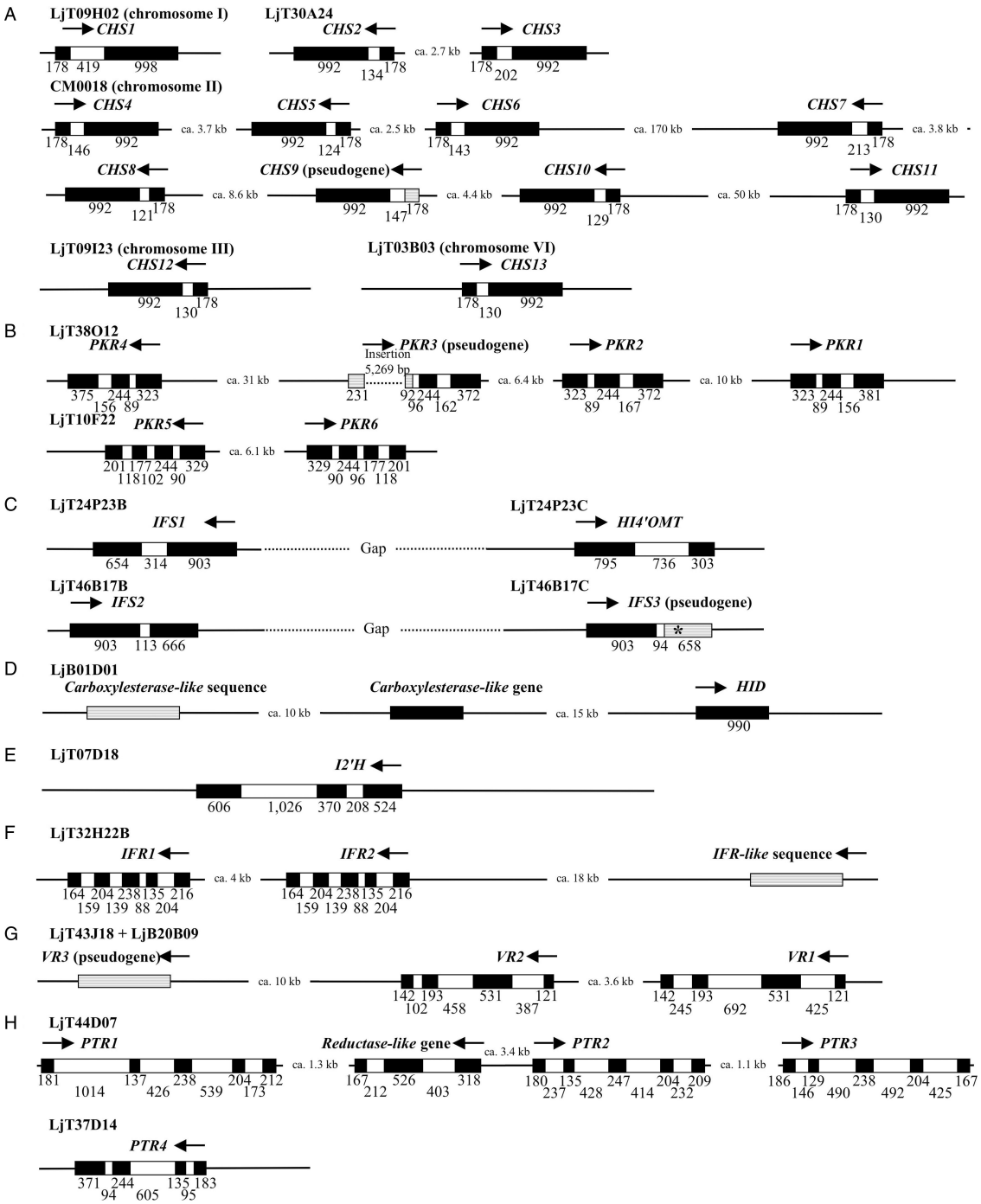

Figure 2. Structures of genes encoding enzymes involved in the biosynthesis of 5-deoxyisoflavonids in *L. japonicus* genome. The figures show only the structural features and relative lengths of exon (black boxes) and intron (white boxes) sequences of each gene, not distances between the genes.

predicted to be Cys445 in IFS1 and Cys449 in IFS2, but is lacking in *IFS3*. *IFS3* is thus considered to be a pseudogene. On the other hand, the 3- or 12-base deletions found in *IFS1*, *IFS2*, and *CYP93C2* at the other sites cause no frameshift, and the amino acid residues encoded by deletion sites are probably unessential.

3.2.4. HI4′OMT, HID, and I2′H HI4′OMT catalyzes the reaction step next to IFS. The *HI4′OMT* gene is found in the same clone as *IFS1* (LjT24P23), although the detailed organization of the genes has not been clarified (Fig. 2C). It is noteworthy that a cluster contains non-homologous genes encoding different enzymes of sequential reactions in the same pathway. Similar gene organizations and the implications for the evolution of metabolic diversity have also been reported.[32,33]

The genes for HID and I2′H (CYP81E6) are located in LjB01D01 (chromosome location unknown) and

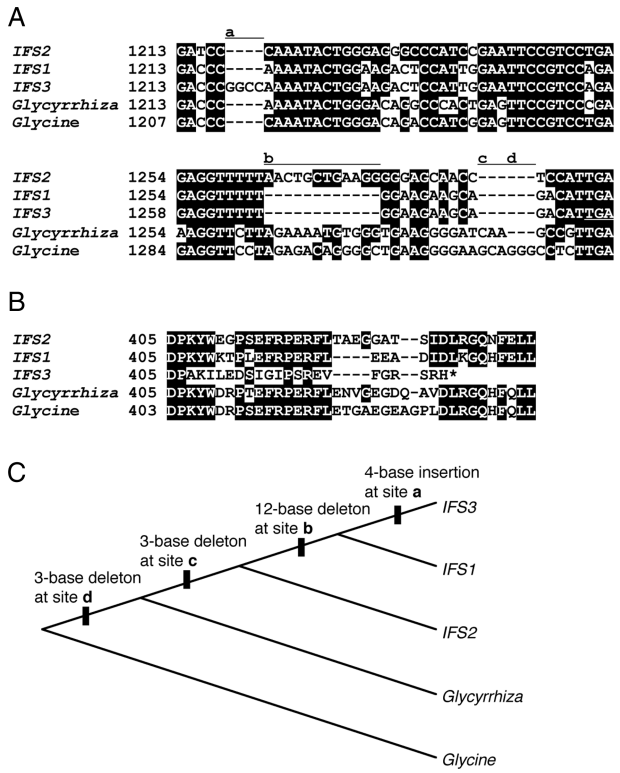

Figure 3. The I/D polymorphisms in *IFS* genes of *L. japonicus* (*IFS1–IFS3*), *Glycyrrhiza echinata* (Genbank accession no. AB023636) and *Glycine max* (Genbank accession no. AF135484). Multiple alignments of the coding sequences from 1213 to 1292 of *IFS2* (**A**) and deduced amino acid sequences of this region (**B**), and the cladistic relationship of the *IFS* genes based on the I/D polymorphism (**C**) are shown. Identical nucleotides and residues are shown in reverse type, and dashes represent gaps for alignment purposes. The I/D sites are shown in bold lower-case letters (**a**–**d**). The 4-bp insertion at site **a** in *IFS3* causes a frameshift, which results in the formation of a nonsense (TGA) codon shown by the underline in (A) and the asterisk in (B). On the other hand, the 3- or 12-base deletions at the other sites (**b**–**d**) cause no frameshift. The synapomorphic character state is indicated at each internode of (C).

LjT07D18 on chromosome IV, respectively (Fig. 2D and E). The deduced amino acid sequence of *L. japonicus* HID showed 75.1% identity to *G. echinata* HID and possesses both an oxyanion hole and catalytic triad, which are characteristic structures of the active site in the carboxylesterase family.[13] Further inspection of LjB01D01 revealed two *carboxylesterase*-like genes in the adjacent region. The deduced amino acid sequence of one of these genes is only 52.3% identical to *L. japonicus* HID, and thus it is not expected that this protein has HID activity. The other one is possibly a pseudogene, because no complete ORF was predicted.

DNA gel blot analyses that detected homologous sequences with >80% identity suggest that *L. japonicus* has single copies of *HI4′OMT* and *HID* (Supplementary Fig. S2) as well as *I2′H.*[20]

3.2.5. IFR and VR Putative *IFR* genes (*IFR1* and *IFR2*) are located in a ca. 7 kb region of LjT32H22 mapped on chromosome IV (Fig. 2F). The structures of the two genes are highly similar: all the exons and introns are exactly the same in length, and the nucleotide sequences between the initiation and termination codons, including both exons and introns, are 99.7% identical. On the other hand, the identities of the 5′- and 3′-untranslated regions are less than 50%. Another *IFR*-like sequence was found ca. 18 kb distant from *IFR2* but is not thought to be IFR because of the low sequence identity (ca. 60%).

Two putative *VR* genes (*VR1* and *VR2*) are within ca. 17 kb in a contig comprised of LjT43J18 and LjB20B09 mapped on chromosome I (Fig. 2G). Another *VR*-like gene (*VR3*) is at ca. 10 kb from *VR2*, but it is judged to be a pseudogene because it lacks an initiation codon.

3.2.6. PTR Four genes are shown to encode PTR.[12] Three of these, *PTR1–PTR3*, are located within a 15 kb region in LjT44D07 on chromosome III, and a *reductase*-like gene is also found in the same clone (Fig. 2H). Because of the high number of repeated sequences in LjT44D07, the nucleotide sequence of the clone has been deposited in the database as a draft sequence. The positions of *PTR1*, *PTR2*, and *PTR3* genes described in Table 2 are based on the draft sequence. *PTR4* gene is located in LjT37D14 on chromosome I. The *PTR1–PTR3* genes are composed of five exons, but *PTR4* has four. Their exon lengths are not conserved (Fig. 2H). The identities of the deduced amino acid sequences of the four PTRs are 60–67%.

3.3. Phylogenetic analyses

CHS and PKR are members of plant polyketide synthase and aldo/keto reductase superfamilies, respectively. IFR and PTR are in the SDR family, which includes two enzymes of the lignan pathway, PCBER and PLR. The phylogenetic relationships of the orthologous and paralogous genes related to *CHS*, *PKR*, *IFR*, and *PTR* were investigated based on the coding sequences (Fig. 4).

Most of the established and putative *CHS* genes of *L. japonicus* and other leguminous *CHS* genes constituted a monophyletic group (Fig. 4A). The topology of the legume-specific clade was consistent with the phylogeny of Papilionoideae.[34] It is thus likely that the duplication processes that generated the paralogous *CHS* genes in *L. japonicus* occurred after the divergence of the legume clades. The phylogenetic tree also suggests the existence of a non-leguminous type of CHS, *CHS1*, of *L. japonicus* (Fig. 4A), but the origin and actual catalytic function of the CHS1 protein are unknown.

The *PKR* genes of legumes fell into two subclades (Fig. 4B). *PKR1*, *PKR2*, and *PKR4* of *L. japonicus* were grouped with *PKR/CHR* genes of other legumes previously identified, whereas *PKR5* and *PKR6* were in another branch with several genes of *G. max* and *M.*

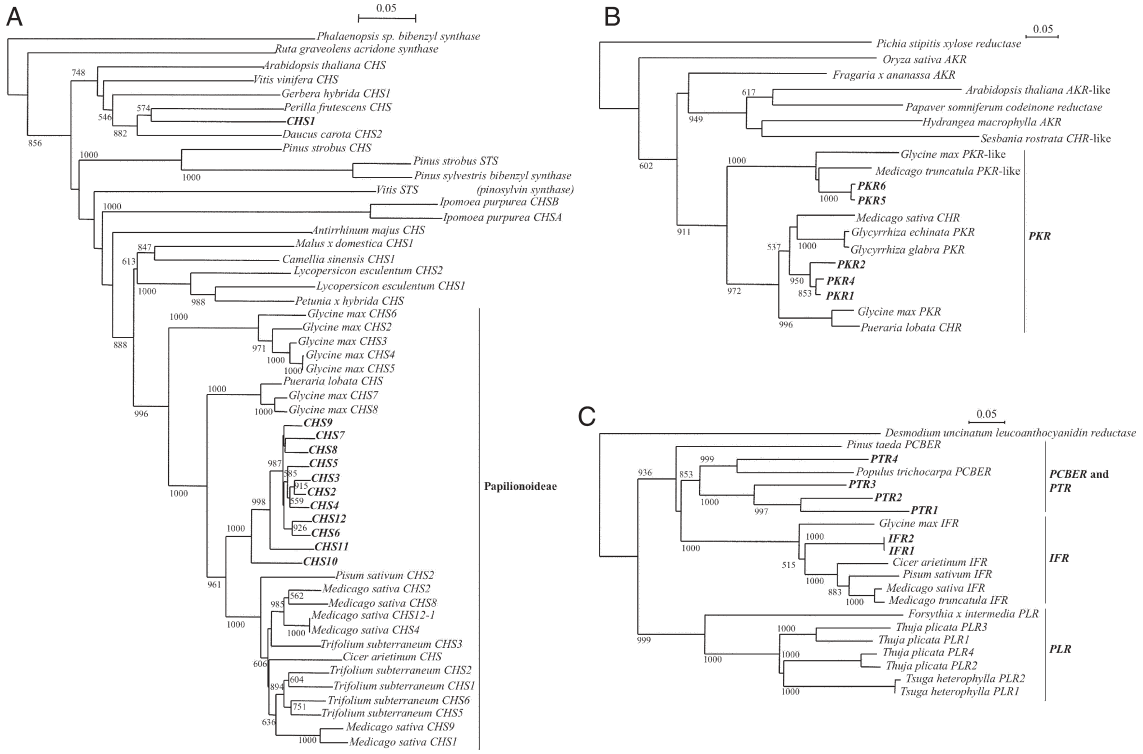

Figure 4. Phylogenetic trees of CHSs (**A**), PKRs (**B**), and the SDR family including IFRs and PTRs (**C**), constructed by the neighbor-joining method. The bootstrap replicates were 1000 (values over 500 are given at the nodes). *Phalaenopsis* sp. *bibenzyl synthase*, *P. stipitis xylose reductase*, and *D. uncinatum leucoanthocyanidin reductase* were used as an outgroup to root the trees of CHSs, PKRs, and the SDR family, respectively.

truncatula annotated as '*CHR*-like'. Because PKR5 showed catalytic activity, as described earlier, it is possible that other *CHR*-like genes in this clade may encode active PKR. In the evolutionary processes, the ancestral *PKR* was possibly duplicated once before the speciation of Papilionoideae, and the subsequent duplication caused the paralogous genes in the two subclades.

A recent study showed that four reductases of *L. japonicus*, which are significantly similar to IFR, PCBER, and PLR, have PTR activity.[12] The phylogenetic tree of these reductases showed that the *PTR* genes of *L. japonicus* (*PTR1*–*PTR4*) constitute a monophyletic clade with the *PCBER* of *Populus trichocarpa* (Fig. 4C). The putative *IFR*s of *L. japonicus* were likely to encode active IFR proteins because they were positioned in a monophyletic clade with other functionally characterized *IFR* genes. Both PCBER and PLR are involved in lignan biosynthesis, but the PTR/PCBER clade is closer to the IFR clade than the PLR clade. No orthologous gene similar to *PCBER* or *PLR* has so far been found in *L. japonicus*.

3.4. Expression of vestitol biosynthesis genes

We previously showed that treatment of *L. japonicus* seedlings with GSH induced the production of vestitol and the accumulation of transcripts from type II *CHI*s, *CHI1* and *CHI3*, and *I2′H* genes.[20,21] To examine the expression of each paralogous gene in GSH-treated seedlings comprehensively, semi-quantitative RT–PCR analysis was performed. Among the 17 genes examined, the transcripts of 11 genes, i.e. *PKR1*, *PKR2*, *PKR6*, *IFS1*, *IFS2*, *HI4′OMT*, *HID*, *IFR1*, *IFR2*, *VR2* and *PTR3*, were markedly elevated 10 h after GSH treatment (Fig. 5A). A slight increase of *PKR4* and *VR1* transcripts during 10 h of elicitation and their decrease in 20 h after GSH treatment were also obvious. In contrast, those of *PKR5*, *PTR1*, *PTR2*, and *PTR4* were constitutively expressed and not substantially affected by GSH.

The RT–PCR analysis also showed that transcripts of the genes were detectable in almost all tissues investigated. The expression patterns were rather complex, but some genes clearly showed organ specificity (Fig. 5B). Characteristically, no transcripts of any *PKR* paralogs, *HID*, *I2′H*, *IFR1*, *VR2*, and *PTR2* genes in flowers were observed, and neither the transcripts of *PTR* paralogs in nodules. The *PKR* transcripts were accumulated in both aerial (stems and leaves) and underground (roots and nodules) parts, but the expression level of *PKR* paralogs was higher in underground parts. In contrast, the expression of all *PTR* genes was mainly observed in aerial parts, and only the *PTR4* transcript was detected

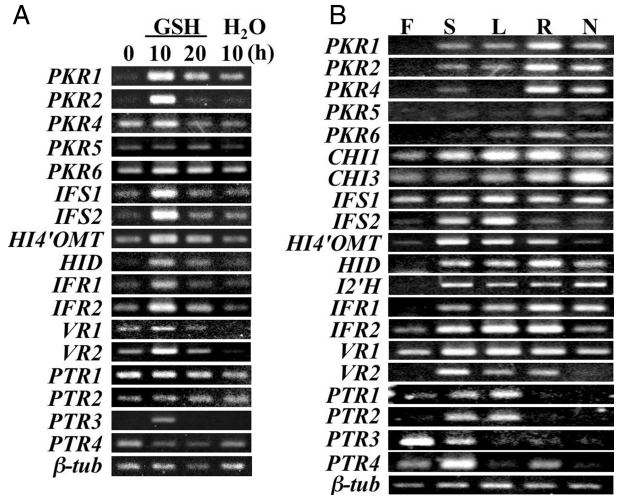

Figure 5. Accumulation of transcripts encoded by the vestitol biosynthetic genes in *L. japonicus*. mRNA was isolated from 4-day-old seedlings at 10 and 20 h after the GSH treatment or at 10 h after H_2O treatment (**A**) or from flowers (F), stems (S), leaves (L), roots (R), and nodules (N) of 8-week-old *L. japonicus* plant (**B**). Amplification of *β-tubulin* served as a control to adjust the amount of PCR template DNA.

in roots. Also, the expression of some paralogous genes was different among the organs: e.g. *IFS1* was expressed in all organs, whereas *IFS2* was highly expressed in leaves and weakly in underground parts; and the expression of *VR1* was ubiquitous in all tissues investigated, but expression of *VR2* was limited to leaves, roots, and stems. On the other hand, the expression patterns of paralogous genes, such as *CHI1* and *CHI3* and *PKR1* and *PKR2*, were almost the same, suggesting their functional redundancy.

4. Discussion

Structural analysis of the genes involved in the legume-specific 5-deoxyisoflavonoid biosynthesis of *L. japonicus* revealed that six genes of the 10 investigated compose multigene families and form gene clusters in many cases. The origin of multigene families is attributed to gene duplication. The paralogous genes of *L. japonicus* involved in the flavonoid and triterpenoid biosynthesis were suggested to have acquired new functions different from the ancestral types as the result of accumulated nucleotide substitutions during their molecular evolution.[21–23] The gene cluster of *HID* and *carboxylesterase*-like genes (Fig. 2D) also implies the evolution of *HID* through local gene duplication of an ancestral *carboxylesterase*-like gene. These findings support the idea that gene duplication is a major driving force for the evolution of novel metabolic pathways.[35–37] Considering that enzymes with new functions might have been established through such evolutionary processes, we expected that candidates for enzymes and genes could be narrowed down by the combination of the predicted reaction mode of the target enzyme and the phylogenetic relationship. Actually, the genes for PTR were found in the SDR family, which includes PCBER, IFR, and PLR.[12] The phylogenetic tree implies that *PTR* genes should have been derived from the ancestral *PCBER* genes (Fig. 4C).

We present a comprehensive expression analysis of the genes involved in 5-deoxyisoflavonoid biosynthesis. Together with the previous finding on the GSH-induced transcription of type II *CHI* and *I2′H* genes,[20,21] all single genes and most of the paralogous genes encoding the biosynthetic enzymes as far as identified were shown to be synchronously up-regulated by the GSH treatment (Fig. 5A), which has been thought to mimic the signal of defense response, guaranteeing the elicitor responsive synthesis of vestitol. The coordinate increase of the transcripts suggests a common set of transcription factors involved in the regulation of this pathway, and the GSH-treatment will serve as a potential experimental system to analyze the transcriptional regulation of the 5-deoxyisoflavonoid genes. Although no significant *cis* elements common to the GSH-dependent genes has been predicted by *in silico* analysis, they will be obtained by further experiments of promoter analysis and used to identify the corresponding transcription factors. An interesting observation is the constitutive expression of *PTR* paralogs, except for *PTR3*, which was apparently up-regulated on GSH treatment. The role of PTR3 in vestitol biosynthesis has been considered to be insignificant, because it possesses low specific activity and no enantiospecificity.[12] If the final step of vestitol synthesis is performed by enantiospecific PTR1 and PTR2 with high specific activity, this step is out of the regulation asserted by elicitor and not a rate-limiting step of phytoalexin biosynthesis.

The semiquantitative RT–PCR analyses also revealed the organ specific expression of each paralogous gene (Fig. 5B). Weak or no expression of *PTR* paralogs and the high expression levels of the other up-stream genes in underground parts imply that 5-deoxyisoflavonoids may be produced in symbiotic roots and nodules. Functions of flavonoids as positive and negative regulators in nodule development have been discussed.[38] The observation that transcripts of some genes, i.e. *PKR* paralogs and *HID*, were not detected in flowers may conform the notion that the normal *L. japonicus* flower tissue producing flavonols and anthocyanins does not synthesize 5-deoxyisoflavonoids. However, the expression of genes may not simply reflect the flavonoid compositions, and more detailed and comprehensive examination of the transcripts and metabolites of the 5-deoxyisoflavonid pathway in specific tissues and cells and during nodule organogenesis would be expected in the future.

On the other hand, the nearly identical expression patterns of some paralogous genes, such as *PKR1* and *PKR2*, and *CHI*s (Fig. 5B), strongly suggested that these paralogous genes are functionally redundant. They are likely to have been maintained during the legume evolution under selection pressure. A previously presented genetic model on the molecular evolution of genes showed that genetic redundancy is common in some cases, contrary to the widespread view that a redundant gene would be abolished by accumulated deleterious mutations.[39] According to the model, redundant genes can be stable, provided they have pleiotropic functions or they are expressed in specific spatiotemporal patterns. Our expression analysis showed that some duplicated genes, e.g. *PKR4*, *IFS2*, *VR2*, *PTR3*, and *PTR4*, have acquired different expression patterns (Fig. 5A and B). *IFS* paralogs of soybean were also reported to show different expression patterns in different tissues, and in response to nodulation and defense signal. Alternatively, one of the two completely redundant genes will become extinct, but it may take a long time, provided that the mutation rates of the two genes are not very different.[39]

The fact that the legume-specific 5-deoxyisoflavonoid pathway is at present widely distributed among Papilionoideae strongly suggests its substantial contributions to the fitness of leguminous plants, that is, 5-deoxyisoflavonoids have significant ecological and physiological functions. The roles of the individual biosynthetic genes will be clarified in part by dissection of the spatiotemporal patterns of gene expression and the distribution of 5-deoxyisoflavonoids, namely, by transcriptomics and metabolomics. In general, these '-omics' approaches are profoundly dependent on gene annotations. The results of this study provide the correct annotations for the legume-specific 5-deoxyisoflavonoid genes. On the other hand, reverse genetic approaches such as RNA interference or TILLING will enable us to test the significance of a gene, a biosynthetic step, or a pathway. RNA interference is applicable to redundant genes if the consensus sequences common to all the paralogous genes are known;[40,41] single genes, e.g. *HI4′OMT*, *HID*, and *I2′H* of *L. japonicus*, can be the targets for knocking out by TILLING. The present study will also offer useful information for these functional genomic studies of leguminous plants.

Acknowledgements: This work was supported by the 21st Century COE program 'Bioresource Utilization Based on Microbial Symbiotic Systems' of the Ministry of Education, Culture, Sports, Science and the project 'Development of Fundamental Technologies for Controlling the Material Production Process of Plants' of the New Energy and Industrial Technology Development Organization. This work was conducted at the Life Science Research Center (Shonan Campus, Nihon University). We wish to thank Satoru Sawai (Nihon University) for technical assistance.

Supplementary Data: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

References

1. International Rice Genome Sequencing Project. 2005, The map-based sequence of the rice genome, *Nature*, **436**, 793–800.
2. Arabidopsis Genome Initiative. 2000, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature*, **408**, 796–815.
3. Tuskan, G. A., DiFazio, S., Jansson, S., et al. 2006, The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray), *Science*, **313**, 1596–1604.
4. Udvardi, M. K., Tabata, S., Parniske, M. and Stougaard, J. 2005, *Lotus japonicus*: Legume research in the fast lane, *Trends Plant Sci.*, **10**, 222–228.
5. Kouchi, H., Shimomura, K., Hata, S., et al. 2004, Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*, *DNA Res.*, **11**, 263–274.
6. Hirai, M. Y., Yano, M., Goodenowe, D. B., et al. 2004, Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*, *Proc. Natl Acad. Sci. USA*, **101**, 10205–10210.
7. Hegnauer, R. and Gpayer-Barkmeijer, R. J. 1993, Relevance of seed polysaccharides and flavonoids for the classification of the leguminosae: A chemotaxonomic approach, *Phytochemistry*, **34**, 3–16.
8. Harborne, J. B. 1999, The comparative biochemistry of phytoalexin induction in plants, *Biochem. Syst. Ecol.*, **27**, 335–367.
9. Kosslak, R. M., Bookland, R., Barkei, J., Paaren, H. E. and Appelbaum, E. R. 1987, Induction of *Bradyrhizobium japonicum* common nod genes by isoflavones isolated from *Glycine max*, *Proc. Natl Acad. Sci. USA*, **84**, 7428–7432.
10. Aoki, T., Akashi, T. and Ayabe, S. 2000, Flavonoids of leguminous plants: Structure, biological activity, and biosynthesis, *J. Plant Res.*, **113**, 475–488.
11. Akashi, T., Sawada, Y., Shimada, N., Sakurai, N., Aoki, T. and Ayabe, S. 2003, cDNA cloning and biochemical characterization of *S*-adenosyl-L-methionine: 2,7,4′-trihydroxyisoflavanone 4′-*O*-methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway, *Plant Cell Physiol.*, **44**, 103–112.
12. Akashi, T., Koshimizu, S., Aoki, T. and Ayabe, S. 2006, Identification of cDNAs encoding pterocarpan reductase involved in isoflavan phytoalexin biosynthesis in *Lotus japonicus* by EST mining, *FEBS Lett.*, **580**, 5666–5670.
13. Akashi, T., Aoki, T. and Ayabe, S. 2005, Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis, *Plant Physiol.*, **137**, 882–891.
14. Paiva, N. L., Edwards, R., Sun, Y. J., Hrazdina, G. and Dixon, R. A. 1991, Stress responses in alfalfa (*Medicago sativa* L.). 11. Molecular cloning and expression of alfalfa isoflavone reductase, a key enzyme of isoflavonoid phytoalexin biosynthesis, *Plant Mol. Biol.*, **17**, 653–667.

15. Paiva, N. L., Sun, Y. J., Dixon, R. A., Vanetten, H. D. and Hrazdina, G. 1994, Molecular cloning of isoflavone reductase from pea (*Pisum sativum* L—evidence for a 3*R*-isoflavanone intermediate in (+)-pisatin biosynthesis, *Arch. Biochem. Biophys.*, **312**, 501–510.
16. Guo, L. N. and Paiva, N. L. 1995, Molecular cloning and expression of alfalfa (*Medicago sativa* L.) vestitone reductase, the penultimate enzyme in medicarpin biosynthesis, *Arch. Biochem. Biophys.*, **320**, 353–360.
17. Steele, C. L., Gijzen, M., Qutob, D. and Dixon, R. A. 1999, Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean, *Arch. Biochem. Biophys.*, **367**, 146–150.
18. Akashi, T., Aoki, T. and Ayabe, S. 1999, Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice, *Plant Physiol.*, **121**, 821–828.
19. Guo, L. N., Dixon, R. A. and Paiva, N. L. 1994, The pterocarpan synthase of alfalfa: association and co-induction of vestitone reductase and 7,2′-dihydroxy-4′-methoxyisoflavanol (DMI) dehydratase, the two final enzymes in medicarpin biosynthesis, *FEBS Lett.*, **356**, 221–225.
20. Shimada, N., Akashi, T., Aoki, T. and Ayabe, S. 2000, Induction of isoflavonoid pathway in the model legume *Lotus japonicus*: molecular characterization of enzymes involved in phytoalexin biosynthesis, *Plant Sci.*, **160**, 37–47.
21. Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S. and Ayabe, S. 2003, A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*, *Plant Physiol.*, **131**, 941–951.
22. Shimada, N., Sasaki, R., Sato, S., et al. 2005, A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the *Lotus japonicus* genome, *J. Exp. Bot.*, **56**, 2573–2585.
23. Sawai, S., Shindo, T., Sato, S., et al. 2006, Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*, *Plant Sci.*, **170**, 247–257.
24. Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S. 2001, Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 Mb regions of the genome, *DNA Res.*, **8**, 311–318.
25. Kato, T., Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E. and Tabata, S. 2003, Structural analysis of a *Lotus japonicus* genome. V. Sequence features and mapping of sixty-four TAC clones which cover the 6.4 Mb regions of the genome, *DNA Res.*, **10**, 277–285.
26. Hayashi, M., Miyahara, A., Sato, S., et al. 2001, Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F-2 population, *DNA Res.*, **8**, 301–310.
27. Akashi, T., Fukuchi-Mizutani, M., Aoki, T., et al. 1999, Molecular cloning and biochemical characterization of a novel cytochrome P450, flavone synthase II, that catalyzes direct conversion of flavanones to flavones, *Plant Cell Physiol.*, **40**, 1182–1186.
28. Akashi, T., Saito, N., Hirota, H. and Ayabe, S. 1997, Anthocyanin-producing dandelion callus as a chalcone synthase source in recombinant polyketide reductase assay, *Phytochemistry*, **46**, 283–287.
29. Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, **22**, 4673–4680.
30. Perriere, G. and Gouy, M. 1996, WWW-query: an on-line retrieval system for biological sequence banks, *Biochimie*, **78**, 364–369.
31. Shimada, N., Nakatsuka, T., Nishihara, M., Yamamura, S., Ayabe, S. and Aoki, T. 2006, Isolation and characterization of a cDNA encoding polyketide reductase in *Lotus japonicus*, *Plant Biotechnol.*, **23**, 509–513.
32. Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R. and Osbourn, A. 2004, A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants, *Proc. Natl Acad. Sci. USA*, **101**, 8233–8238.
33. Frey, M., Chomet, P., Glawischnig, E., et al. 1997, Analysis of a chemical plant defense mechanism in grasses, *Science*, **277**, 696–699.
34. Wojciechowski, M. F., Lavin, M. and Sanderson, M. J. 2004, A phylogeny of legumes (Legumenosae) based on analyses of the plastid *matK* gene resolves many well-supported subclades within the family, *Am. J. Bot.*, **91**, 1846–1862.
35. Kliebenstein, D. J., Lambrix, V. M., Reichelt, M., Gershenzon, J. and Mitchell-Olds, T. 2001, Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis, *Plant Cell*, **13**, 681–693.
36. Ober, D. 2005, Seeing double: gene duplication and diversification in plant secondary metabolism, *Trends Plant Sci.*, **10**, 444–449.
37. Seitz, C., Eder, C., Deiml, B., Kellner, S., Martens, S. and Forkmann, G. 2006, Cloning, functional identification and sequence analysis of flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase cDNAs reveals independent evolution of flavonoid 3′,5′-hydroxylase in the Asteraceae family, *Plant Mol. Biol.*, **61**, 365–381.
38. Novak, K., Lisa, L. and Skrdleta, V. 2004, Rhizobial nod gene-inducing activity in pea nodulation mutants: dissociation of nodulation and flavonoid response, *Physiol. Plantarum*, **120**, 546–555.
39. Nowak, M. A., Boerlijst, M. C., Cooke, J. and Smith, J. M. 1997, Evolution of genetic redundancy, *Nature*, **388**, 167–171.
40. Subramanian, S., Graham, M. Y., Yu, O. and Graham, T. L. 2005, RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*, *Plant Physiol.*, **137**, 1345–1353.
41. Subramanian, S., Stacey, G. and Yu, O. 2006, Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*, *Plant J.*, **48**, 261–273.