Plant Science 264 (2017) 149–167

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Legume isoflavone synthase genes have evolved by whole-genome and local duplications yielding transcriptionally active paralogs

Dorota Narożna[a,1], Michał Książkiewicz[b,*,1], Łucja Przysiecka[b,c], Joanna Króliczak[a], Bogdan Wolko[b], Barbara Naganowska[b], Cezary J. Mądrzak[a]

[a] Department of Biochemistry and Biotechnology, Faculty of Agronomy and Bioengineering, Poznań University of Life Sciences, Dojazd 11, 60-632, Poznań, Poland
[b] Department of Genomics, Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479, Poznań, Poland
[c] NanoBioMedical Centre, Adam Mickiewicz University, Umultowska 85, 61-614, Poznań, Poland

ARTICLE INFO

Keywords:
Lupinus angustifolius
Isoflavone synthase
Genome duplication
Multigene family
Gene expression
Chromosomal localization

ABSTRACT

Isoflavone synthase (IFS) is the key enzyme of isoflavonoid biosynthesis. IFS genes were identified in numerous species, although their evolutionary patterns have not yet been reconstructed. To address this issue, we performed structural and functional genomic analysis.

Narrow leafed lupin, *Lupinus angustifolius* L., was used as a reference species for the genus, because it has the most developed molecular tools available. Nuclear genome BAC library clones carrying IFS homologs were localized by linkage mapping and fluorescence *in situ* hybridization in three chromosome pairs. Annotation of BAC, scaffold and transcriptome sequences confirmed the presence of three full-length IFS genes in the genome.

Microsynteny analysis and Bayesian inference provided clear evidence that IFS genes in legumes have evolved by lineage-specific whole-genome and tandem duplications. Gene expression profiling and RNA-seq data mining showed that the vast majority of legume IFS copies have maintained their transcriptional activity. *L. angustifolius* IFS homologs exhibited organ-specific expression patterns similar to those observed in other Papilionoideae. Duplicated lupin IFS homologs retained non-negligible levels of substitutions in conserved motifs, putatively due to positive selection acting during early evolution of the genus, before the whole-genome duplication. Strong purifying selection preserved newly arisen IFS duplicates from further nonsynonymous changes.

1. Introduction

Isoflavonoids constitute a large subfamily of flavonoids – plant secondary metabolites common to most of the vascular plants, playing a critical role in their growth and development [1]. These compounds, contrary to other groups of flavonoids, were for a long time presumed to be specific to the *Fabaceae*. Leguminous plants still remain the most abundant source of isoflavonoids which perform diverse biological functions [2], particularly those linked with plant-microbe interactions, both symbiotic [3,4] and pathogenic [5]. The occurrence of isoflavonoids in other plant families is sporadic [6–8]. It should be pointed out that some isoflavonoids (e.g. daidzein, genistein or coumestrol) are structurally similar to estradiol and can act as estrogens in humans and animals. They are therefore referred to as phytoestrogens [5,9]. Isoflavonoids present in genus *Lupinus* constitute structurally diverse group of natural products, presenting various potential biological or pharmacological activities [10,11]. Except differentially hydroxylated isoflavones genistein and 2′-hydroxygenistein, also isoflavones modified with isoprenyl groups (wighteone and luteone) were found in roots and leaves of white and narrow-leafed lupins [12]. All isoflavones exist in the plant tissue as the variety of *O*- or *C*-glycosylated derivatives, often acylated with malonyl group on sugar moieties. Free aglycones are released in the cells under abiotic or biotic stress [13,14]. Isoflavonoids are becoming interesting bioproducts and therefore metabolic engineering methods are being applied in order to increase their content in plant products [15–17].

Abbreviations: BAC, bacterial artificial chromosome; BLAST, basic local alignment search tool; EMBL, European Molecular Biology Laboratory; ER, endoplasmic reticulum; GA, gene expression atlas; IFS, isoflavone synthase; Ka, nonsynonymous substitution rate; Ks, synonymous substitution rate; NLL, narrow leafed lupin linkage group; SDS, sodium dodecyl sulfate; SRA, sequence read archive; SSC, saline & sodium citrate; SSPE, saline – sodium phosphate – EDTA; TA, transcriptome assembly; TE, transposable element; WGD, whole-genome duplication

* Corresponding author.
E-mail addresses: dorna@o2.pl (D. Narożna), mksi@igr.poznan.pl (M. Książkiewicz), lucprz@amu.edu.pl (Ł. Przysiecka), jkrol@up.poznan.pl (J. Króliczak), bwol@igr.poznan.pl (B. Wolko), bnag@igr.poznan.pl (B. Naganowska), madrzak@up.poznan.pl (C.J. Mądrzak).
[1] Dorota Narożna and Michał Książkiewicz contributed equally to this article.

http://dx.doi.org/10.1016/j.plantsci.2017.09.007
Received 27 January 2017; Received in revised form 5 September 2017; Accepted 11 September 2017
Available online 18 September 2017
0168-9452/

Isoflavone synthase (IFS) is the key enzyme responsible for isoflavonoid biosynthesis in plants. It is the first enzyme in the branch of the phenylpropanoid pathway catalyzing the conversion of common flavonoid precursors liquirtigenin and naringenin into 2,7,4′-trihydroxyisoflavanon and 2,5,7,4′-tetrahydroxyisoflavanon, respectively – the precursors of all known isoflavonoids [18]. Isoflavone synthase (2-hydroxyisoflavanone synthase) is a cytochrome P450–dependent monooxygenase [19], belonging to the CYP93C subfamily [20]. The conversion of flavanone substrates into isoflavonoid precursors, catalyzed by IFS, involves the 2-hydroxylation and aryl ring migration of substrates to yield a 2-hydroxyisoflavanone. The reaction is followed by a dehydration step catalyzed by 2-hydroxyisoflavanone dehydratase [21] to yield the isoflavone product (daidzein or genistein).

The genes encoding isoflavone synthase have been identified from a variety of legumes such as soybean, pea, alfalfa, red clover, peanut and others. In all those plant species, except red clover [22], they constitute small multigene families [20,21,23,24]. The homology between IFS genes from different genera is high, often exceeding 95% at the level of protein similarity. In angiosperms, IFS genes contain a single intron and two well-conserved exons.

Legumes (*Fabaceae*) are the third largest family of higher plants, comprising about 20 thousand species grouped into several clades, of which the largest is the subfamily Papilionoideae. Lupins belong to genistoids, one of the earliest diverging lineages of the Papilionoideae which evolved ~50–60 million years ago [25]. Lupin crops are relevant sources of protein for animal and human consumption, as well as natural fertilizer plants improving soils and increasing yields of subsequent crops. Within the genus, *Lupinus angustifolius* L. (narrow-leafed lupin) is the species with the most developed molecular resources, including linkage maps with sequence-defined molecular markers [26–28], bacterial artificial chromosome (BAC) libraries representing 6–12 × genome coverage [29,30], transcriptome assemblies for wild and domesticated lines [31] as well as a draft genome sequence [32]. As such it has been considered as a reference species for the whole genus, and subjected to comprehensive studies involving targeted gene-rich region sequencing [33], molecular cytogenetic survey [34,35], cross-genera comparative mapping and gene-based phylogenetic inference [36,37]. The possibilities of family-wide studies were greatly enhanced by the release of high-quality genome assemblies of several legume species representing the main clades of Papilionoideae: dalbergioids (*Arachis*) [38], genistoids (*Lupinus*) [32], robinioids (*Lotus*) [39], millettioids (*Cajanus, Glycine, Phaseolus, Vigna*) [40–43], and the inverted repeat-lacking clade (*Medicago* and *Cicer*) [44,45]. Recent studies provided novel evidence for the evolutionary uniqueness of the *Lupinus* genus represented by *L. angustifolius*, manifested by its early divergence in the Papilionoideae lineage and retention of triplicated gene homologs or even whole chromosome segments as remnants of ancient whole-genome duplication events [25,26,35,46].

Therefore, we aimed to reconstruct the evolution of IFS genes in the legume family utilizing genomic data from *L. angustifolius* as a case study. Here, we harnessed available genomic resources to perform structural and functional analysis of IFS genes. The *L. angustifolius* nuclear genome BAC library was screened with probes designed based on the sequences of *L. luteus* (yellow lupin) IFS cDNAs. BAC clone localization in *L. angustifolius* chromosomes was visualized using fluorescence *in situ* hybridization. IFS sequences were anchored to a linkage map of the species, using gene-based markers. IFS sequences were also retrieved from sequenced legume genomes by nucleotide and protein multiple alignment followed by protein-based hidden Markov model gene prediction. To track evolution of particular homologs, Bayesian inference of phylogeny was applied. Transcriptional activity of duplicated gene loci was revealed by gene expression analysis as well as RNA-seq and microarray data mining.

2. Materials and methods

2.1. Biological material

Seeds of *L. angustifolius* cv. Sonet were obtained from the Breeding Station Wiatrowo (Poznań Plant Breeders LTD. Tulce, Poland). Three-day-old seedlings were used for DNA isolation. Plant tissues for RNA isolation were sampled from plants grown in a growth chamber under controlled conditions (temperature 18 °C, photoperiod 12/12 h). Shoot apical meristems, stems, and leaves were collected from 5 day-old plants. Root nodules were isolated from greenhouse grown plants 28 days after infection with *Bradyrhizobium japonicum* (*lupinus*) UPP331 [47].

A nuclear genome BAC library of *L. angustifolius* cv. Sonet, constructed in pIndigoBAC5 HindIII-Cloning Ready vector [30], was used as a source of BAC clones for molecular and cytogenetic analyses.

Seeds of the *L. angustifolius* mapping population, developed as 89 F_8 recombinant inbred lines from the cross combination 83A:476 × P27255 [48], were kindly provided by Dr. Hua'an Yang from the Department of Agriculture and Food, Western Australia.

2.2. Hybridization probe and BAC library screening

BAC clones containing sequences of IFS genes were selected by screening the nuclear BAC library of *L. angustifolius*. Primers for the amplification of IFS fragments were designed based on the nucleotide sequences of *L. luteus* IFS cDNAs determined earlier in our laboratory (LlIFS1 – FJ539089 and LlIFS2 – FJ539090). Primers Ll-IFSR and Ll-IFSF (Supplementary Table 1) were designed to amplify fragments of exon II of IFS genes. Polymerase chain reactions using Applied Biosystems GenAmp PCR system 2400 were performed in a reaction mixture (total volume 25 µl) containing: 1 × PCR buffer, 100 µM dNTP (each) and 1U Taq DNA polymerase. The amplification consisted of the following steps: 94 °C, 2 min; followed by 30 cycles of: 94 °C, 30s; 54 °C, 30s; 72 °C, 1 min, and final extension: 72 °C, 4 min.

Hybridization probes were labeled with P32 dATP (MP Biomedicals) using PCR with annealing temperature of 54 °C. The probe was purified using Montage PCR Centrifugal Filter Devices (Merck Millipore, Darmstadt, Germany), denatured at 94 °C for 5 min and incubated on ice. Hybridization of the labeled probe with three macroarrays containing the whole nuclear genome BAC library of *L. angustifolius* cv. Sonet [30] was conducted in 5 × SSPE buffer with 0.5% SDS at 50 °C overnight. Filters were washed for 15 min at 50 °C in solutions of increasing stringency (5 × SSC + 0.5% SDS; 2.5 × SSC + 0.25% SDS; 1 × SSC + 0.1% SDS; 0.5 × SSC + 0.05% SDS). Blots were exposed to BAS-MS 2340 imaging plates (Fujifilm, Tokyo, Japan) for 48 h and analyzed using the FLA-5100 phosphorimager (Fujifilm).

2.3. Verification and sequencing of selected BAC clones

DNA from clones yielding positive hybridization signals was isolated using the PhasePrep BAC DNA Kit (Sigma-Aldrich, St. Louis, USA). The presence of IFS encoding fragments was verified by PCR using Ll-IFSeieF/R primers (Supplementary Table 1) amplifying the fragment consisting of part of exon I, the intron, and part of exon II of IFS genes.

Positively verified BAC clones were delivered to Genomed (Warsaw, Poland) for sequencing. GS FLX TITANIUM 454 DNA Sequencing (Roche 454 Life Sciences, Branford, USA) was performed with tagged BAC DNA samples. The 1/8 picotiterplate (PTP) standard shotgun reads for the set of 3 BACs were performed with read length up to 400 nucleotides. Assuming even distribution of 454 reads between the different tagged samples, the planned sequencing scheme was equivalent to approximately 20 × coverage of 454 reads for the sequence of single BAC clones. Sequences were assembled by Genomed using CLC Genomic Workbench (v7.0.4) software under default parameters [35]. Ambiguous sites were resolved by PCR amplification and sequencing of products amplified from a particular BAC clone.