Genomics Data 10 (2016) 4–11

Contents lists available at ScienceDirect

Genomics Data

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

Comparative analysis of transcriptomes in aerial stems and roots of *Ephedra sinica* based on high-throughput mRNA sequencing

Taketo Okada [a,*], Hironobu Takahashi [b], Yutaka Suzuki [c], Sumio Sugano [d], Masaaki Noji [b], Hiromichi Kenmoku [b], Masao Toyota [b], Shigehiko Kanaya [e], Nobuo Kawahara [f], Yoshinori Asakawa [b], Setsuko Sekita [a]

[a] Faculty of Pharmaceutical Sciences at Kagawa Campus, Tokushima Bunri University, Sanuki, Kagawa 769-2193, Japan
[b] Institute of Pharmacognosy, Tokushima Bunri University, Tokushima 770-8514, Japan
[c] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, the University of Tokyo, Kashiwa, Chiba 277-8562, Japan
[d] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, the University of Tokyo, Minato-ku, Tokyo 108-8639, Japan
[e] Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
[f] Tsukuba Division, Research Center for Medicinal Plant Resources, National Institutes of Biomedical Innovation, Health and Nutrition, Tsukuba, Ibaraki 305-0843, Japan

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form 2 August 2016
Accepted 2 August 2016
Available online 4 August 2016

Keywords:
Comparative transcriptome analysis
Ephedra sinica
High-throughput mRNA sequencing
RNA-Seq

ABSTRACT

Ephedra plants are taxonomically classified as gymnosperms, and are medicinally important as the botanical origin of crude drugs and as bioresources that contain pharmacologically active chemicals. Here we show a comparative analysis of the transcriptomes of aerial stems and roots of *Ephedra sinica* based on high-throughput mRNA sequencing by RNA-Seq. *De novo* assembly of short cDNA sequence reads generated 23,358, 13,373, and 28,579 contigs longer than 200 bases from aerial stems, roots, or both aerial stems and roots, respectively. The presumed functions encoded by these contig sequences were annotated by BLAST (blastx). Subsequently, these contigs were classified based on gene ontology slims, Enzyme Commission numbers, and the InterPro database. Furthermore, comparative gene expression analysis was performed between aerial stems and roots. These transcriptome analyses revealed differences and similarities between the transcriptomes of aerial stems and roots in *E. sinica*. Deep transcriptome sequencing of *Ephedra* should open the door to molecular biological studies based on the entire transcriptome, tissue- or organ-specific transcriptomes, or targeted genes of interest.

1. Introduction

Ephedra is one of the oldest medicinal plant genera known to mankind [1–3]. This genus belongs to the Ephedraceae family of gymnosperms, and about 50 *Ephedra* species are indigenous to areas in Asia, Europe, North Africa, and the Americas. The aerial stems of *Ephedra* plants have been utilized as a crude drug preparation known as ephedra herb (Ephedrae Herba), used mainly for treatment of bronchitis and bronchial asthma, or to induce perspiration and blood pressure elevation. Ephedra herb is particularly used in traditional Oriental medicines; it is well known as *má huáng* in traditional Chinese medicine (often abbreviated to TCM), and is frequently used in Japanese *Kampo* medicine, often as one component of a combined drug formulation. The ingredients mainly associated with the unique pharmacological and biological effects of ephedra herb are ephedrine alkaloids [*e.g.* (−)-ephedrine; (−)-*N*-methylephedrine] [1]. Since the first isolation of an ephedrine alkaloid in 1887 by Professor Nagayoshi Nagai, the founder of pharmacy in Japan, these alkaloids have been studied around the world. Ephedrine alkaloids are primarily localized in the aerial stems of several *Ephedra* species as their principal metabolites (*e.g.*, *E. sinica*, *E. intermedia*, *E. equisetina*) [4–6]. Pharmacologically, ephedrine alkaloids are a sympathomimetic agonist at α/β-adrenergic receptors, resulting in bronchodilation (β_2), enhanced cardiac rate and contractility (β_1), and peripheral vasoconstriction (α_1). The biosynthetic pathway of these alkaloids has been studied; the route primarily from L-phenylalanine has been chemically and biochemically summarized, although several of the reaction steps have been predicted in hypothetical pathways [7–16]. The underground roots of *Ephedra* plants have also been utilized as a crude drug preparation known as ephedra root (Ephedrae Radix). Interestingly, it is well known that ephedra root has hypotensive activity, which is the opposite pharmacological effect of ephedra herb. This hypotensive property is thought to be derived from several unique metabolites contained in *Ephedra* roots: ephedradines A–D [17–20]; ephedrannin A [21]; mahuannin A–D [22–24]; and feruloylhistamine [25], which were isolated by monitoring the hypotensive activity of *Ephedra* root extract. The hypotensive activities of ephedradine B and feruloylhistamine analogues have been a particular focus of pharmacological study [26,27]. In addition, maokonine [28], ephedrannin B [29], and mahuannin E [29] have also been isolated from *Ephedra* roots.

Abbreviations: EC, Enzyme Commission; Es_R, *E. sinica* roots; Es_S, *E. sinica* aerial stems; Es_SR, *E. sinica* combined aerial stems and roots; GO, gene ontology; IPR, InterPro.

* Corresponding author.
E-mail address: okadat@kph.bunri-u.ac.jp (T. Okada).

http://dx.doi.org/10.1016/j.gdata.2016.08.003

Although maokonine displays weak hypertensive activity, the primary pharmacological effect of ephedra root is still hypotensive. In this way, due to the importance of *Ephedra* plants as medicinal resources, our understanding of their biological, pharmacological, chemical, and taxonomic properties has progressed through interdisciplinary studies.

The genetic and genomic features of *Ephedra* species, from the viewpoint of molecular biology, have been elucidated gradually. For example, during studies of ephedrine alkaloid biosynthesis, a *pal* gene of *E. sinica* involved in the primary step of the biosynthetic pathway was cloned and characterized [14]. In a further study, mRNA in aerial stems of *E. sinica* (Es_S) was comprehensively sequenced and the gene candidates potentially involved in biosynthesis of amphetamine-type alkaloids including ephedrines were profiled [7]. Based on this study, two aromatic aminotransferases of *E. sinica* were characterized [30]. In other studies, the sequences of internal transcribed spacer 1 region of the nuclear ribosomal DNA, 18S ribosomal RNA gene, and chloroplast DNA were used to describe the taxonomy of *Ephedra* plants (*e.g.*, [31–33]). Furthermore, the chloroplast genomic sequences of *E. foeminea* was totally analyzed, and new plastid markers for phylogenetic purposes were suggested by comparison with the sequences of *E. equisetina* [34]. Thus, RNA and DNA sequences of *Ephedra* species have been effectively used for targeted studies.

In this study, the comparative analysis between two transcriptomes in Es_S and roots of *E. sinica* (Es_R) by a high-throughput mRNA sequencing using a Genome Analyzer IIx (Illumina, CA, USA) is mainly presented. The mRNAs of Es_S and Es_R were separately sequenced and the sequence data were comprehensively analyzed using bioinformatics approaches. Our comparative transcriptome analysis of Es_S and Es_R focused in particular on molecular biological annotation of *de novo* sequences and quantitation of gene expression levels. Namely, this comparative study was performed to more comprehensively understand an *Ephedra* plant as a biological system by deep transcriptome analysis.

2. Materials and methods

2.1. High-throughput mRNA sequencing

The seeds of *E. sinica* were germinated in moistened vermiculite, sand, and small stones (5:5:1) in daylight at *ca.* 25 °C/10 °C in a greenhouse, improving upon the methods previously reported by our group [14]. *E. sinica* was grown until the plan had generated aerial stems with 4–5 joints.

Es_S and Es_R were collected separately and their mRNAs were sequenced individually. Total RNAs were extracted using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and the quality of samples for high-throughput mRNA sequencing were confirmed using the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) with the Agilent RNA 6000 Pico Kit (Agilent Technologies) (Fig. S1). The sequencing samples were prepared using the mRNA-Seq Sample Preparation Kit (Illumina, CA, USA) and PE adaptors were ligated onto cDNA ends. The single read-cDNA clusters on a flow cell for sequencing were generated using cBot (Illumina). Sequencing was performed using a Genome Analyzer IIx (Illumina) with the single-read method using 36-cycle sequencing. Sequencing of each Es_S and Es_R sample was performed twice. The short sequence reads obtained from these RNA-Seq experiments were registered in the DDBJ BioProject database (PRJDB3343).

2.2. Bioinformatics analysis

The RNA-Seq reads in fastq format were assembled using the Rnnotator program [35] and contig sequences were output in fasta format. Searches by blastx query with an E-value cutoff of 1E-6, GO mapping, and annotation by EC and IPR numbers were performed for Es_S, Es_R, and combined Es_S and Es_R (Es_SR) contigs continuously using the Blast2GO program [36–38]. The method for quantitation of gene expression levels in the aerial stems and roots is summarized in Fig. 1. In this expression analysis, mapping of short sequence reads in fastq format of Es_S and Es_R to Es_SR contigs was performed using TopHat [39]. The gene expression levels in the Es_S and Es_R transcriptomes were quantified by using Cufflinks software, and the abundances of expressed genes were calculated as expected fragments per kilobase of transcript per million fragments mapped (FPKM) [40]. The differential gene expression levels of the Es_SR combined transcriptomes in Es_S and Es_R were quantified using Cuffdiff in the Cufflinks program [41]. The significance of the abundance of an expressed gene was determined by the false discovery rate < 5% (q value < 0.05).

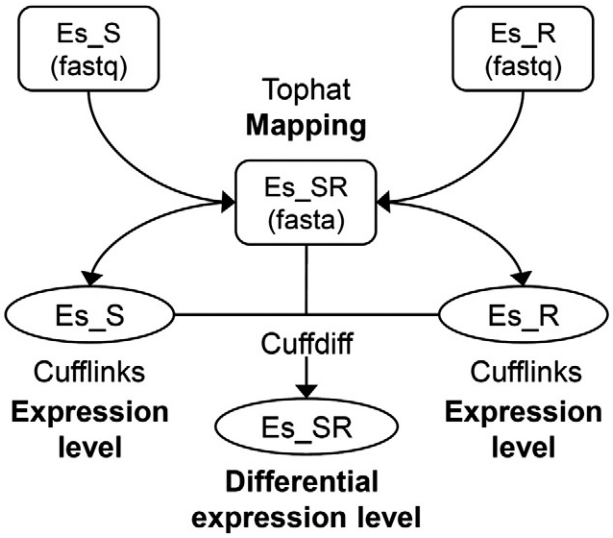

Fig. 1. Scheme for analysis of differential gene expression to compare transcriptomes of Es_S and Es_R.

3. Results

3.1. High-throughput sequencing of mRNA from Es_S and Es_R and de novo assembly

Total mRNA from both Es_S and Es_R was sequenced using a Genome Analyzer IIx (Illumina) for RNA-Seq [42,43] (Table 1). Two independent technical replicates were performed for sequencing both Es_S and Es_R. A total of 6.4×10^7 reads from Es_S and 6.3×10^7 reads from Es_R were acquired. *De novo* assembly was performed using Rnnotator software [35] and cDNA contigs were generated from Es_S, Es_R, and Es_SR. The cDNA contigs over 200 bases that we identified

Table 1
High-throughput sequencing of mRNAs from Es_S and Es_R by RNA-Seq.

Sequenced plant's part	Experiment	Length of SRS[a]	Clusters (passed filter/tile)	Total number of clusters[b]	Number of contigs (≥200 bases)	
Es_S	1st	35 bases	213,156	25,578,720	23,358	28,579[c]
	2nd		324,766	38,971,920		
	Total		537,922	64,550,640		
Es_R	1st		219,999	26,399,880	13,373	
	2nd		310,339	37,240,680		
	Total		530,338	63,640,560		

[a] Short-read sequencing.
[b] 120 Tiles/Experiment.
[c] Number of Es_SR contigs.

included a total of 23,358 contigs from Es_S, 13,373 contigs from Es_R, and 28,579 contigs from Es_SR.

3.2. BLAST searches of contig sequences

To find amino acid sequences encoded by mRNA of *E. sinica* similar to those of other sequences, cDNA contigs longer than 200 bases from Es_S, Es_R, and Es_SR were analyzed using blastx program, which compares a nucleotide query sequence translated in all reading frames to a protein sequence database. A blastx search was performed against the public protein database Swiss-Prot, which consists of manually annotated and reviewed proteins and amino acid sequences in the UniProt Knowledgebase (UniProtKB; http://www.uniprot.org/uniprot/). As a result, 49.8% (11,643), 55.5% (7428), and 48.7% (13,925) of the Es_S, Es_R, and Es_SR contigs were annotated with known gene functions, respectively. The minimum E-values (Table S1) and the percentages of mean similarity (Table S2) distributions of the Es_SR contigs were summarized and displayed in a single figure (Fig. S2). Over 80% of the Es_SR contigs were concentrated in the ranges of E-values not over 8.67E-14 and similarity over 55%. The species of the sequences highest hits by blastx search are also statistically summarized (Table 2). Indeed, as one might expect, approximately half of the highest matches annotating the Es_SR contigs were genes from *Arabidopsis thaliana* (51.69%), and the percentages of species annotating the other contigs were <7.16%.

3.3. Classification of contigs by gene ontology

The contigs annotated by blastx search were then classified by gene ontology (GO) covering the three functional categories of molecular function, biological processes, and cellular component [44]. All GO terms annotating the gene products of these contigs were remapped using 'GO slims' [45], which are smaller and more manageable subsets of GO, to reduce the large numbers of original GO terms assigned to these contig sequences. As a result, 95.7% (11,138), 97.0% (7198), and 95.8% (13,334) of Es_S, Es_R, and Es_SR contigs, respectively, that had been annotated by blastx search could also be classified by GO terms (Table 3). Comparison of results for Es_S and Es_R contigs classified based on three GO categories are also shown in Table 3. In the transcriptome of *E. sinica*, there is little difference in the percentages of GO terms assigned to contigs of Es_S or Es_R.

Table 2
Species distribution of sequences matching Es_SR contigs by blastx search.

Species	Common name	Number of contigs	Percentage (%)
Arabidopsis thaliana	Mouse-ear cress	7198	51.69
Oryza sativa subsp. *japonica*	Rice	997	7.16
Homo sapiens	Human	594	4.27
Mus musculus	Mouse	424	3.04
Dictyostelium discoideum	Slime mold	391	2.81
Schizosaccharomyces pombe (Strain 972/ATCC 24843)	Fission yeast	234	1.68
Nicotiana tabacum	Common tobacco	141	1.01
Bos taurus	Bovine	137	0.98
Zea mays	Maize	134	0.96
Danio rerio	Zebrafish	132	0.95
Solanum lycopersicum	Tomato	126	0.9
Rattus norvegicus	Rat	124	0.89
Oryza sativa subsp. *indica*	Rice	112	0.8
Solanum tuberosum	Potato	104	0.75
Xenopus laevis	African clawed frog	100	0.72
Pinus taeda	Loblolly pine	95	0.68
Glycine max	Soybean	94	0.68
Others	–	2788	20.02

3.4. Classification of proteins and domains encoded by contigs based on enzyme commission (EC) numbers and the InterPro database

EC numbers comprehensively categorize catalytic enzymes based on the six main classes (EC 1–6) of similar enzymatic reactions [46]. In the present study, the amino acid sequences encoded by the Es_S, Es_R, and Es_SR contigs were annotated with EC numbers. As a result, EC numbers were assigned to 14.7% (3444), 18.5% (2470), and 14.2% (4053) of Es_S, Es_R, and Es_SR contigs, respectively.

The protein domains encoded by Es_S, Es_R, and Es_SR contigs were also classified using information from the InterPro (IPR) database (The European Molecular Biology Laboratory-European Bioinformatics Institute) organized by the several institutions that make up the consortium [47]. Protein domain predictions were performed using InterProScan [48]. Consequently, 77.0% (17,984), 81.0% (10,830) and 76.0% (21,732) of Es_S, Es_R, and Es_SR contigs, respectively, were characterized by IPR database. Specifically, 57.3% (10,308), 61.2% (6625), and 57.7% (12,533) of the Es_S, Es_R, and Es_SR contigs, respectively, classified by IPR database were annotated with IPR numbers.

3.5. Comparative expression analysis of transcriptomes in Es_S and Es_R based on gene functions

Differential gene expression analysis was performed using sequences of genes expressed in Es_S and Es_R to compare these transcriptomes (Fig. 1). The sequence reads from Es_S and Es_R were mapped onto Es_SR contigs using the TopHat program [39]. Subsequently, gene expression levels of Es_S and Es_R were quantified using the Cufflinks program [40], and the differential levels of gene expression in Es_S and Es_R were quantified using Cuffdiff in the Cufflinks program [41]. We found that 4.1% (1170) and 3.8% (1085) of the 28,579 contigs from Es_SR were significantly expressed in Es_S and Es_R, respectively (Fig. 2). To characterize these significantly expressed genes, the enzymatic functions of the encoded proteins were classified based using EC (Fig. 3) and IPR (Table 4) numbers annotated to contigs.

The numbers of EC numbers annotated to differentially expressed genes from Es_S and Es_R were roughly the same (219 and 229, respectively) (Fig. 3A). Genes (69 contigs) encoding EC 3 (hydrolases) were highly expressed in Es_S compared to Es_R (38 contigs) (a 1.8-fold difference) (Fig. 3A–C). In particular, genes encoding the EC 3.1.3.x enzymes (phosphoric monoester hydrolases) were characteristically expressed in Es_S. For example, for x = 2, the enzyme is acid phosphatase; if x = 4, the enzyme is phosphatidate phosphatase; if x = 11, the enzyme is fructose-bisphosphatase; if x = 37, the enzyme is sedoheptulose-bisphosphatase; and if x = 46, the enzyme is fructose-2,6-bisphosphate 2-phosphatase. EC 3.1.3.11, EC 3.1.3.37 and EC 3.1.3.46 are involved in saccharide metabolism, and EC 3.1.3.11 and EC 3.1.3.37 are related to the metabolic pathway for carbon fixation by photosynthesis in aerial parts. Moreover, the genes encoding EC 5 (isomerases) (9 contigs) were highly expressed in Es_S, including: EC 5.2.1.8, peptidylprolyl isomerase; EC 5.3.3.2, isopentenyl-diphosphate Δ-isomerase; EC 5.4.99.7, lanosterol synthase; and EC 5.4.99.8, cycloartenol synthase (Fig. 3A, D). On the other hand, genes encoding EC 1 (oxidoreductases) enzymes (108 contigs) were highly expressed in Es_R compared to Es_S (58 contigs) (a 1.9-fold difference) (Fig. 3A, E, F). The number of contigs encoding EC 1.11.1.7 (peroxidase) was particularly elevated in Es_R (4.4-fold) compared to Es_S.

IPR functional terms, which are coordinated with IPR numbers, were also assigned to Es_SR contigs, and 574 and 475 terms were annotated to the contigs of genes significantly expressed in Es_S and Es_R, respectively. Additionally, 426 and 216 terms were specifically annotated to Es_S and Es_R, respectively, and 180 terms were annotated to both Es_S and Es_R. The top-10 ranking of IPR functional terms according to the number of annotated contigs is listed in Table 4.

Table 3
Distribution of Es_S, Es_R, and Es_SR contigs annotated by GO slims.

GO functional categories	Number of Es_SR contigs	(%)	Number of Es_S contigs	(%)	Number of Es_R contigs	(%)
Cellular Component	*23,060*	*100*	*19,907*	*100*	*13,889*	*100*
Cell	1222	5.3	992	4.98	700	5.04
Cell wall	675	2.93	540	2.71	462	3.33
Cytoplasm	2142	9.29	1853	9.31	1202	8.65
Cytoskeleton	418	1.81	367	1.84	196	1.41
Cytosol	1650	7.16	1499	7.53	1068	7.69
Endoplasmic reticulum	700	3.04	604	3.03	441	3.18
Endosome	215	0.93	175	0.88	121	0.87
External encapsulating structure	3	0.01	5	0.03	1	0.01
Extracellular region	504	2.19	403	2.02	332	2.39
Extracellular space	55	0.24	53	0.27	33	0.24
Golgi apparatus	514	2.23	450	2.26	265	1.91
Intracellular	1278	5.54	1036	5.2	669	4.82
Lysosome	44	0.19	46	0.23	20	0.14
Membrane	2331	10.11	1973	9.91	1436	10.34
Mitochondrion	1324	5.74	1192	5.99	882	6.35
Nuclear envelope	120	0.52	99	0.5	75	0.54
Nucleolus	638	2.77	572	2.87	397	2.86
Nucleoplasm	569	2.47	521	2.62	290	2.09
Nucleus	2322	10.07	1997	10.03	1321	9.51
Peroxisome	227	0.98	216	1.09	189	1.36
Plasma membrane	2622	11.37	2184	10.97	1610	11.59
Plastid	2050	8.89	1855	9.32	1221	8.79
Proteinaceous extracellular matrix	10	0.04	11	0.06	4	0.03
Ribosome	328	1.42	320	1.61	287	2.07
Thylakoid	332	1.44	312	1.57	194	1.4
Vacuole	767	3.33	632	3.17	473	3.41
Molecular Function	*20,414*	*100*	*17,488*	*100*	*12,019*	*100*
Binding	2349	11.51	1987	11.36	1479	12.31
Carbohydrate binding	110	0.54	90	0.51	53	0.44
Catalytic activity	2299	11.26	1903	10.88	1458	12.13
Chromatin binding	87	0.43	89	0.51	28	0.23
DNA binding	500	2.45	438	2.5	264	2.2
Enzyme regulator activity	236	1.16	199	1.14	132	1.1
Hydrolase activity	2235	10.95	1896	10.84	1202	10
Kinase activity	1106	5.42	932	5.33	570	4.74
Lipid binding	132	0.65	102	0.58	85	0.71
Motor activity	62	0.3	55	0.31	6	0.05
Nuclease activity	127	0.62	110	0.63	57	0.47
Nucleic acid binding	167	0.82	136	0.78	76	0.63
Nucleotide binding	1830	8.96	1628	9.31	1136	9.45
Oxygen binding	57	0.28	40	0.23	34	0.28
Protein binding	4725	23.15	4146	23.71	2759	22.96
Receptor activity	199	0.97	151	0.86	103	0.86
Receptor binding	90	0.44	73	0.42	52	0.43
RNA binding	569	2.79	569	3.25	441	3.67
Sequence-specific DNA binding transcription factor activity	446	2.18	378	2.16	252	2.1
Signal transducer activity	164	0.8	141	0.81	96	0.8
Structural molecule activity	332	1.63	319	1.82	260	2.16
Transferase activity	1418	6.95	1195	6.83	770	6.41
Translation factor activity, nucleic acid binding	117	0.57	114	0.65	111	0.92
Translation regulator activity	18	0.09	19	0.11	15	0.12
Transporter activity	1039	5.09	778	4.45	580	4.83
Biological Process	*41,133*	*100*	*34,885*	*100*	*23,848*	*100*
Abscission	16	0.04	11	0.03	8	0.03
Anatomical structure morphogenesis	1358	3.3	1124	3.22	714	2.99
Behavior	113	0.27	92	0.26	60	0.25
Biological process	2	0	2	0.01	1	0
Biosynthetic process	2240	5.45	1864	5.34	1366	5.73
Carbohydrate metabolic process	837	2.03	743	2.13	574	2.41
Catabolic process	1243	3.02	1091	3.13	860	3.61
Cell communication	196	0.48	151	0.43	110	0.46
Cell cycle	793	1.93	675	1.93	383	1.61
Cell death	387	0.94	325	0.93	223	0.94
Cell differentiation	1027	2.5	834	2.39	551	2.31
Cell growth	598	1.45	493	1.41	330	1.38
Cell-cell signaling	81	0.2	71	0.2	57	0.24
Cellular component organization	2430	5.91	2113	6.06	1285	5.39
Cellular homeostasis	181	0.44	158	0.45	99	0.42
Cellular process	5016	12.19	4312	12.36	2883	12.09
Cellular protein modification process	1284	3.12	1070	3.07	673	2.82
Death	4	0.01	5	0.01	6	0.03
DNA metabolic process	422	1.03	354	1.01	184	0.77
Embryo development	848	2.06	733	2.1	461	1.93

(continued on next page)

Table 3 (*continued*)

GO functional categories	Number of Es_SR contigs	(%)	Number of Es_S contigs	(%)	Number of Es_R contigs	(%)
Flower development	486	1.18	402	1.15	255	1.07
Fruit ripening	5	0.01	3	0.01	2	0.01
Generation of precursor metabolites and energy	379	0.92	297	0.85	315	1.32
Growth	454	1.1	399	1.14	305	1.28
Lipid metabolic process	858	2.09	753	2.16	478	2
Metabolic process	1396	3.39	1139	3.27	842	3.53
Multicellular organismal development	2010	4.89	1669	4.78	1111	4.66
Nucleobase-containing compound metabolic process	1216	2.96	1119	3.21	746	3.13
Photosynthesis	146	0.35	130	0.37	84	0.35
Pollen-pistil interaction	19	0.05	8	0.02	8	0.03
Pollination	259	0.63	217	0.62	128	0.54
Post-embryonic development	1215	2.95	1047	3	682	2.86
Protein metabolic process	710	1.73	634	1.82	493	2.07
Regulation of gene expression, epigenetic	197	0.48	163	0.47	70	0.29
Reproduction	1158	2.82	1027	2.94	639	2.68
Response to abiotic stimulus	1696	4.12	1394	4	1040	4.36
Response to biotic stimulus	1012	2.46	853	2.45	602	2.52
Response to endogenous stimulus	1266	3.08	1020	2.92	709	2.97
Response to external stimulus	419	1.02	359	1.03	243	1.02
Response to extracellular stimulus	226	0.55	193	0.55	131	0.55
Response to stress	2488	6.05	2028	5.81	1449	6.08
Secondary metabolic process	554	1.35	424	1.22	329	1.38
Signal transduction	1358	3.3	1168	3.35	744	3.12
Translation	528	1.28	535	1.53	411	1.72
Transport	1877	4.56	1574	4.51	1153	4.83
Tropism	125	0.3	109	0.31	51	0.21

4. Discussion

High-throughput mRNA sequencing by RNA-Seq technique has enabled deep transcriptome analysis of many kinds of organisms. In this study, transcripts from *E. sinica* were comprehensively sequenced and the transcriptomes of aerial stems and roots were comparatively analyzed.

Es_SR contigs longer than 200 bases totaled about 28,000, and were generated by *de novo* assembly of short sequence reads from both Es_S and Es_R (Table 1). Comparing contigs from both types of plant parts, there were 1.7-fold more Es_S contigs than Es_R contigs (23,358, and 13,373 contigs, respectively). This result suggests more active metabolism in aerial stems than in roots (*e.g.*, photosynthesis). In a blastx search against the Swiss-Prot database, ca. 50% of contigs were annotated by various encoded protein functions. BLAST results were statistically analyzed (Table 2, S1, S2, and Fig. S2) and most of these contigs could be classified using GO slims (Table 3). Interestingly, the percentages of assigned GO slims were similar between Es_S and Es_R contigs. This result suggested that although gene expression in aerial stems was relatively more active than that in roots, the overall diversity of functions expressed in each organ was very similar in a view of the broader functional categorization achieved using GO. Actually, only about 8% (Fig. 2) of genes exhibited a significant difference in expression level between Es_S and Es_R. Thus, the metabolic diversity and differences between these plant parts might be controlled by the expression of relatively few genes specific to each plant organ.

In the present study, differences in categories of expressed genes could be considered in detail using bioinformatics analysis of sequence reads (Fig. 1). The encoded protein functions of genes expressed in Es_S and Es_R were assigned to contigs according to EC and IPR numbers (Fig. 3, Table 4). For example, contigs encoding chlorophyll *a*/*b* binding proteins (IPR023329 and IPR022796) were specifically identified from among Es_S contigs (Table 4). The chlorophyll *a*/*b* binding protein is part of the light-harvesting complex, a light receptor that captures and delivers excitation energy to photosystems I and II *via* chlorophylls *a*/*b* [49,50]. This result was closely related to the result from comparing Es_S and Es_R using EC numbers, which specifically identified EC3.1.3.11 and EC3.1.3.37, which are involved in photosynthesis, in Es_S (Fig. 3B). Interestingly, the contigs encoding thiolase-like domains (IPR016038 and IPR 016039) were identified in Es_S contigs (Table 4). In the biosynthetic pathway of ephedrine alkaloids, a thiolase is presumed to catalyze the biosynthesis of benzoyl-CoA from 3-oxo-3-phenylpropionyl-CoA in a β-oxidative CoA-dependent route [7,12,14]. This assumption about the biosynthetic route agrees with the accumulation of ephedrine alkaloids in aerial stems of *Ephedra* plants.

5. Conclusions

In conclusion, the transcriptome of an *Ephedra* plant is analyzed using deep RNA-Seq and bioinformatics, focusing on a comparative analysis of gene expression in aerial stems and roots. The results of

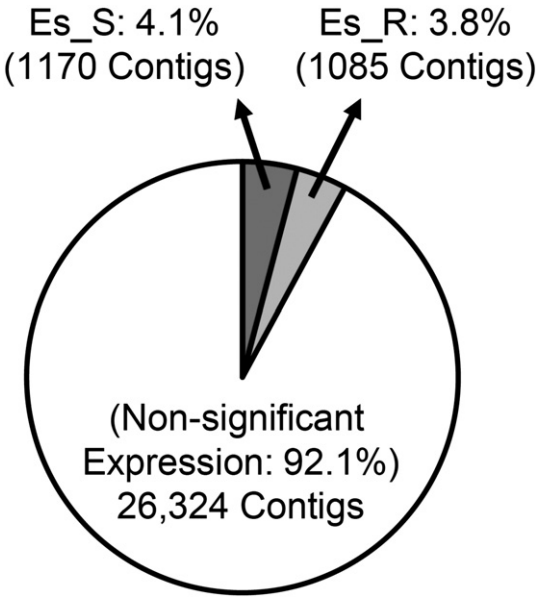

Fig. 2. Percentage of significantly expressed genes in Es_S and Es_R.

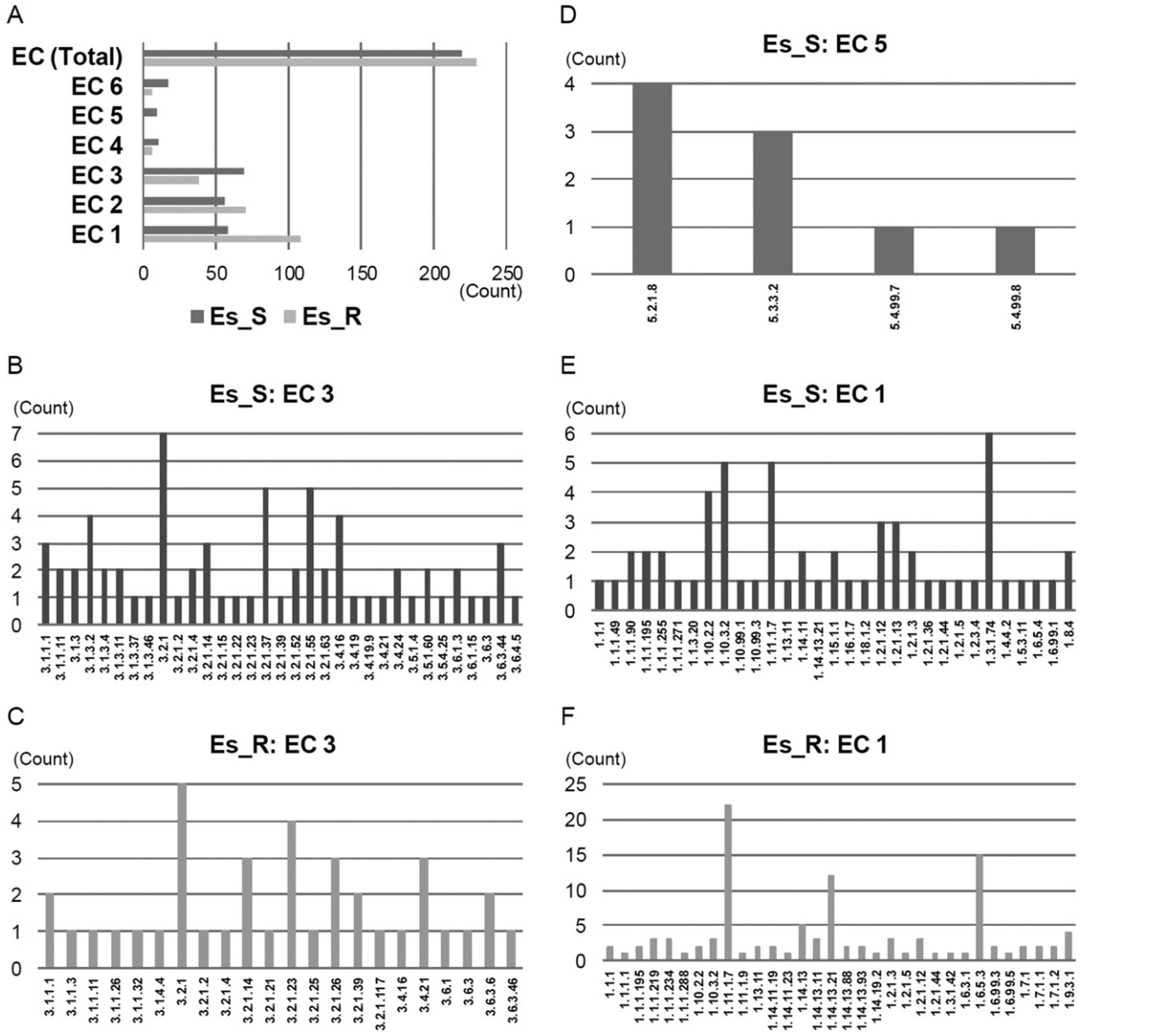

Fig. 3. Comparison of EC numbers annotated with amino acid sequences encoded by differentially expressed genes in Es_S and Es_R. A, Summary of comparison results; B–F, distribution of EC numbers (EC1, 3, and 5) according to Es_S or Es_R.

the present study will form a molecular biological basis for other research, such as evaluating various qualities of medicinal resources, distinguishing species and cultivars, and biosynthesizing specific accumulated metabolites. It is hoped that this study and further research will contribute to the useful and sustainable application and efficient cultivation of *Ephedra* plants as medicinal bioresources, and also promote their survival in their natural settings.

Transparency document

The Tranparency document associated with this article can be found, in online version.

Acknowledgements

We are grateful to Professor Si-Young Song (Faculty of Pharmaceutical Sciences at Kagawa Campus, Tokushima Bunri University, Japan) for the opportunity to demonstrate the application of high-throughput mRNA sequencing technology. We are grateful to Professor Masayuki Mikage (Graduate School of Natural Science and Technology, Kanazawa University, Japan; current affiliation, Faculty of Agriculture, Tokyo University of Agriculture, Japan) for supplying *E. sinica* seeds. We are grateful to Dr. Hisayo Sadamoto-Suzuki (Faculty of Pharmaceutical Sciences at Kagawa Campus, Tokushima Bunri University, Japan) and Mr. Satoshi Tamaki (Graduate School of Information Science, Nara Institute of Science and Technology, Japan) for useful suggestions regarding this transcriptomic study. This work was partly supported by a grant from the Ministry of Health, Labour and Welfare of Japan (H25-SOYAKU-SHITEI-006) and MEXT-Senryaku (no. S1001057).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.08.003.

Table 4
IPR numbers assigned to Es_SR contigs of genes significantly expressed in Es_S and Es_R.

Plant organ	Ranking	IPR number	Number of contigs	Annotation
Es_S specific	1	IPR001763	7	Rhodanese-like domain (D)
		IPR005150		Cellulose synthase (F)
		IPR008030		NmrA-like domain (D)
		IPR013026		Tetratricopeptide repeat-containing domain (D)
	5	IPR013601	6	FAE1/Type III polyketide synthase-like protein (D)
		IPR016038		Thiolase-like, subgroup (D)
		IPR016039		Thiolase-like (D)
		IPR023329		Chlorophyll *a*/*b* binding protein domain (D)
	9	IPR001305	5	Heat shock protein DnaJ, cysteine-rich domain (D)
		IPR002937		Amine oxidase (D)
		IPR005746		Thioredoxin (F)
		IPR013766		Thioredoxin domain (D)
		IPR022796		Chlorophyll A-B binding protein (F)
Es_R specific	1	IPR001461	13	Aspartic peptidase (F)
		IPR021109		Aspartic peptidase domain (D)
	3	IPR004158	7	Protein of unknown function DUF247, plant (F)
		IPR010987		Glutathione *S*-transferase, C-terminal-like (D)
	5	IPR001480	6	Bulb-type lectin domain (D)
		IPR004045		Glutathione *S*-transferase, N-terminal (D)
		IPR004046		Glutathione *S*-transferase, C-terminal (D)
	8	IPR001750	5	NADH:ubiquinone/plastoquinone oxidoreductase (D)
		IPR003445		Cation transporter (F)
		IPR006094		FAD linked oxidase, N-terminal (D)
		IPR016166		FAD-binding, type 2 (D)
Es_S and Es_R	1	IPR001128	50	Cytochrome P450 (F)
	2	IPR002213	27	UDP-glucuronosyl/UDP-glucosyltransferase (F)
	3	IPR002401	26	Cytochrome P450, E-class, group I (F)
		IPR016040		NAD(P)-binding domain (D)
	5	IPR011009	19	Protein kinase-like domain (D)
	6	IPR023213	18	Chloramphenicol acetyltransferase-like domain (D)
	7	IPR000719	17	Protein kinase domain (D)
		IPR003480		Transferase (F)
		IPR017972		Cytochrome P450, conserved site (S)
	10	IPR017853	16	Glycoside hydrolase, superfamily (D)

D, Domain; F, Family; S, Conserved site. (It should be noted that IPR numbers are revised occasionally upon InterPro database updates.)

References

[1] E.A. Abourashed, A.T. El-Alfy, I.A. Khan, L. Walker, Ephedra in perspective – a current review. Phytother. Res. 17 (2003) 703–712.
[2] S. Caveney, D.A. Charlet, H. Freitag, M. Maier-Stolte, A.N. Starratt, New observations on the secondary chemistry of world *Ephedra* (Ephedraceae). Am. J. Bot. 88 (2001) 1199–1208.
[3] C. Rydin, K.R. Pedersen, E.M. Friis, On the evolutionary history of *Ephedra*: cretaceous fossils and extant molecules. Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 16571–16576.
[4] A.Y. Leung, S. Foster, Encyclopedia of Common Natural Ingredients Used in Food, Drugs, and Cosmetics. second ed. John Wiley & Sons, New York, 1996.
[5] Y.M. Liu, S.J. Sheu, S.H. Chiou, H.C. Chang, Y.P. Chen, A comparative study of commercial samples of Ephedrae herba. Planta Med. 59 (1993) 376–378.
[6] Ministry of Health, Labour and Welfare of Japan, The Japanese Pharmacopoeia. 16th ed., 2011 Japan.
[7] R.A. Groves, J.M. Hagel, Y. Zhang, K. Kilpatrick, A. Levy, F. Marsolais, E. Lewinsohn, C.W. Sensen, P.J. Facchini, Transcriptome profiling of khat (*Catha edulis*) and *Ephedra sinica* reveals gene candidates potentially involved in amphetamine-type alkaloid biosynthesis. PLoS One 10 (2015), e0119701.
[8] G. Grue-Sørensen, I.D. Spenser, Biosynthesis of ephedrine. J. Am. Chem. Soc. 110 (1988) 3714–3715.
[9] G. Grue-Sørensen, I.D. Spenser, The biosynthesis of ephedrine. Can. J. Chem. 67 (1989) 998–1009.
[10] G. Grue-Sørensen, I.D. Spenser, Biosynthesis of the *Ephedra* alkaloids: evolution of the C_6-C_3 skeleton. J. Am. Chem. Soc. 115 (1993) 2052–2054.
[11] G. Grue-Sørensen, I.D. Spenser, Biosynthetic route to the *Ephedra* alkaloids. J. Am. Chem. Soc. 116 (1994) 6195–6200.
[12] J.M. Hagel, R. Krizevski, F. Marsolais, E. Lewinsohn, P.J. Facchini, Biosynthesis of amphetamine analogs in plants. Trends Plant Sci. 17 (2012) 404–412.
[13] R. Krizevski, E. Bar, O.R. Shalit, A. Levy, J.M. Hagel, K. Kilpatrick, F. Marsolais, P.J. Facchini, S. Ben-Shabat, Y. Sitrit, E. Lewinsohn, Benzaldehyde is a precursor of phenylpropylamino alkaloids as revealed by targeted metabolic profiling and comparative biochemical analyses in *Ephedra* spp. Phytochemistry 81 (2012) 71–79.
[14] T. Okada, M. Mikage, S. Sekita, Molecular characterization of the phenylalanine ammonia-lyase from *Ephedra sinica*. Biol. Pharm. Bull. 31 (2008) 2194–2199.
[15] K. Yamasaki, U. Sankawa, S. Shibata, Biosynthesis of ephedrine in *Ephedra*, participation of C_6–C_1 unit. Tetrahedron Lett. 10 (1969) 4099–4102.
[16] K. Yamasaki, T. Tamaki, S. Uzawa, U. Sankawa, S. Shibata, Participation of C_6–C_1 unit in the biosynthesis of ephedrine in *Ephedra*. Phytochemistry 12 (1973) 2877–2882.
[17] M. Tamada, K. Endo, H. Hikino, C. Kabuto, Structure of ephedradine A, a hypotensive principle of *Ephedra* roots. Tetrahedron Lett. 20 (1979) 873–876.
[18] M. Tamada, K. Endo, H. Hikino, Structure of ephedradine B, a hypotensive principle of *Ephedra* roots. Heterocycles 12 (1979) 783–786.
[19] C. Konno, M. Tamada, K. Endo, H. Hikino, Structure of ephedradine C, a hypotensive principle of *Ephedra* roots. Heterocycles 14 (1980) 295–298.
[20] H. Hikino, M. Ogata, C. Konno, Structure of ephedradine D, a hypotensive principle of *Ephedra* roots. Heterocycles 17 (1982) 155–158.
[21] H. Hikino, M. Takahashi, C. Konno, Structure of ephedrannin A, a hypotensive principle of *Ephedra* roots. Tetrahedron Lett. 23 (1982) 673–676.
[22] H. Hikino, N. Shimoyama, Y. Kasahara, M. Takahashi, C. Konno, Structures of mahuannin A and B, hypotensive principles of *Ephedra* roots. Heterocycles 19 (1982) 1381–1384.
[23] Y. Kasahara, N. Shimoyama, C. Konno, H. Hikino, Structure of mahuannin C, a hypotensive principle of *Ephedra* roots. Heterocycles 20 (1983) 1741–1744.
[24] Y. Kasahara, H. Hikino, Structure of mahuannin D, a hypotensive principle of *Ephedra* roots. Heterocycles 20 (1983) 1953–1956.
[25] H. Hikino, M. Ogata, C. Konno, Structure of feruloylhistamine, a hypotensive principle of *Ephedra* roots. Planta Med. 48 (1983) 108–110.
[26] H. Hikino, K. Ogata, C. Konno, S. Sato, Hypotensive actions of ephedradines, macrocyclic spermine alkaloids of *Ephedra* roots. Planta Med. 48 (1983) 290–293.
[27] H. Hikino, Y. Kiso, M. Ogata, C. Konno, K. Aisaka, H. Kubota, N. Hirose, T. Ishihara, Pharmacological actions of analogues of feruloylhistamine, an imidazole alkaloid of *Ephedra* roots. Planta Med. 50 (1984) 478–480.
[28] M. Tamada, K. Endo, H. Hikino, Maokonine, hypertensive principle of *Ephedra* roots. Planta Med. 34 (1978) 291–293.
[29] H. Tao, L. Wang, Z. Cui, D. Zhao, Y. Liu, Dimeric proanthocyanidins from the roots of *Ephedra sinica*. Planta Med. 74 (2008) 1823–1825.
[30] K. Kilpatrick, A. Pajak, J.M. Hagel, M.W. Sumarah, E. Lewinsohn, P.J. Facchini, F. Marsolais, Characterization of aromatic aminotransferases from *Ephedra sinica* Stapf. Amino Acids 48 (2016) 1209–1220.
[31] A. Inoko, N. Kakiuchi, M. Yoshimitsu, S. Cai, M. Mikage, *Ephedra* resource in Sichuan and Yunnan provinces 2007. Biol. Pharm. Bull. 32 (2009) 1621–1623.
[32] Y. Kitani, S. Zhu, J. Batkhuu, C. Sanchir, K. Komatsu, Genetic diversity of *Ephedra* plants in Mongolia inferred from internal transcribed spacer sequence of nuclear ribosomal DNA. Biol. Pharm. Bull. 34 (2011) 717–726.
[33] A.L. Qin, M.M. Wang, Y.Z. Cun, F.S. Yang, S.S. Wang, J.H. Ran, X.Q. Wang, Phylogeographic evidence for a link of species divergence of *Ephedra* in the Qinghai-Tibetan plateau and adjacent regions to the Miocene Asian aridification. PLoS One 8 (2013), e56243.
[34] C. Hou, N. Wikström, C. Rydin, The chloroplast genome of *Ephedra foeminea* (Ephedraceae, Gnetales), an entomophilous gymnosperm endemic to the Mediterranean area. Mitochondrial DNA (2015), http://dx.doi.org/10.3109/19401736.2015.1122768.

[35] J. Martin, V.M. Bruno, Z. Fang, X. Meng, M. Blow, T. Zhang, G. Sherlock, M. Snyder, Z. Wang, Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11 (2010) 663.
[36] A. Conesa, S. Götz, Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008 (2008) 619832.
[37] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (2005) 3674–3676.
[38] S. Götz, J.M. García-Gómez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36 (2008) 3420–3435.
[39] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25 (2009) 1105–1111.
[40] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28 (2010) 511–515.
[41] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. Nat. Protoc. 7 (2012) 562–578.
[42] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (2009) 57–63.
[43] B.T. Wilhelm, J.R. Landry, RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48 (2009) 249–257.
[44] The Gene Ontology Consortium, Gene Ontology annotations and resources. Nucleic Acids Res. 41 (2013) D530–D535.
[45] M.J. Davis, M.S. Sehgal, M.A. Ragan, Automatic, context-specific generation of Gene Ontology slims. BMC Biochem. 11 (2010) 498.
[46] E.C. Webb, The nomenclature committee of the international union of biochemistry and molecular biology. Enzyme Nomenclature, 1992, Academic Press, California, 1992.
[47] S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T.K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, et al., InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40 (2012) D306–D312.
[48] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier. Nucleic Acids Res. 33 (2005) W116–W120.
[49] J.F. Allen, Protein phosphorylation in regulation of photosynthesis. Biochim. Biophys. Acta 1098 (1992) 275–335.
[50] E.M. Aro, I. Ohad, Redox regulation of thylakoid protein phosphorylation. Antioxid. Redox Signal. 5 (2003) 55–67.