



## REGULAR PAPER

# Comparative analysis using the draft genome sequence of California poppy (*Eschscholzia californica*) for exploring the candidate genes involved in benzyloquinoline alkaloid biosynthesis

Yasuyuki Yamada <sup>1,\*</sup>, Hideki Hirakawa,<sup>2</sup> Kentaro Hori,<sup>3</sup> Yohei Minakuchi,<sup>4</sup> Atsushi Toyoda,<sup>4</sup> Nobukazu Shitan,<sup>1</sup> and Fumihiko Sato <sup>3,5,\*</sup>

<sup>1</sup>Laboratory of Medicinal Cell Biology, Kobe Pharmaceutical University, Kobe, Japan; <sup>2</sup>Kazusa DNA Research Institute, Kisarazu, Japan; <sup>3</sup>Graduate School of Biostudies, Kyoto University, Kyoto, Japan; <sup>4</sup>National Institute for Genetics, Mishima, Japan; and <sup>5</sup>Graduate School of Science, Osaka Prefecture University, Sakai, Japan

\*Correspondence: Yasuyuki Yamada, [yyamada@kobepharm-u.ac.jp](mailto:yyamada@kobepharm-u.ac.jp); Fumihiko Sato, [fsato@lif.kyoto-u.ac.jp](mailto:fsato@lif.kyoto-u.ac.jp)

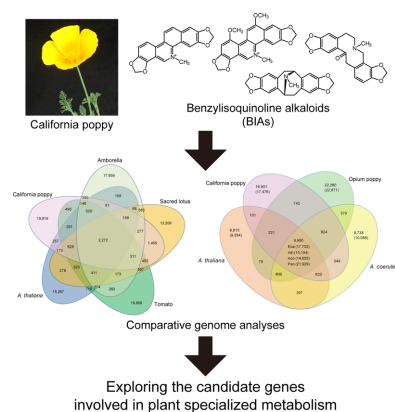
## ABSTRACT

Genome characterization of California poppy (*Eschscholzia californica* cv. "Hitoezaki"), which produces pharmaceutically important benzyloquinoline alkaloids (BIAs), was carried out using the draft genome sequence. The numbers of tRNA and rRNA genes were close to those of the other plant species tested, whereas the frequency of repetitive sequences was distinct from those species. Comparison of the predicted genes with those of *Amborella trichopoda*, *Nelumbo nucifera*, *Solanum lycopersicum*, and *Arabidopsis thaliana*, and analyses of gene ontology and Kyoto Encyclopedia of Genes and Genomes pathway indicated that the enzyme genes involved in BIA biosynthesis were highly enriched in the California poppy genome. Further comparative analysis using the genome information of *Papaver somniferum* and *Aquilegia coerulea*, both BIA-producing plants, revealed that many genes encoding BIA biosynthetic enzymes, transcription factors, transporters, and candidate proteins, possibly related to BIA biosynthesis, were specifically distributed in these plant species.

Received: 2 October 2020; Accepted: 18 November 2020

© The Author(s) 2021. Published by Oxford University Press on behalf of Japan Society for Bioscience, Biotechnology, and Agrochemistry. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Graphical Abstract



Genome characterization of California poppy provides valuable information for identifying specific genes involved in plant-specialized metabolism.

**Keywords:** *Eschscholzia californica*, California poppy, genome sequencing, comparative genome analysis

**Abbreviations:** BIA: benzylisoquinoline alkaloid; bp: base pairs; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: euKaryotic clusters of Orthologous Group; TE: transposable element

California poppy (*Eschscholzia californica*), a member of the Papaveraceae family, is native to the United States of America and Mexico and mainly found on the west coast of North America. It is a basal eudicot with a beautiful ornamental orange flower and has been used as herbal medicine owing to the pharmacological effects of specialized compounds it contains. The California poppy is not only a model plant to study floral organ development in basal eudicot (Becker, Gleissberg and Smyth 2005; Carlson et al. 2006), but also a model to study the biosynthesis of benzylisoquinoline alkaloids (BIAs), which are pharmaceutically important and chemically divergent specialized products found mainly in the Papaveraceae, Ranunculaceae, Berberidaceae, and Menispermaceae families (Kutchan 1995; Hagel and Facchini 2013; Sato 2013, 2020). Of the approximately 2500 compounds, morphine (analgesic), codeine (antitussive), magnoflorine (anti-HIV agent), and berberine (antimicrobial) are the representative BIAs. One of the major BIAs produced by the California poppy is sanguinarine, which shows antibiotic activity and has been used as a component of toothpaste. The biosynthetic pathway of sanguinarine and the transcriptional regulation mechanism by several transcription factors have been well elucidated at the molecular level (Dittrich and Kutchan 1991; Ikezawa, Iwasa and Sato 2007, 2009; Liscombe and Facchini 2007; Hagel et al. 2012; Beaudoin and Facchini 2013; Takemura et al. 2013; Yamada et al. 2015; Hori et al. 2018). However, many other specialized metabolite pathways in the California poppy, including some BIAs, such as macarpine and escholtzine biosynthesis pathways (Figure S1), and flower pigmentations have been scarcely elucidated, although there have been reports on the expressed sequence tag (EST) libraries and transcriptome data of *E. californica* (Carlson et al. 2006; Xiao et al. 2013).

To identify and characterize candidate genes involved in specialized metabolism directly from the genome sequence, we obtained the draft genome sequence (approximately 489 Mb) of the California poppy (ECA\_r1.0), which covered >95% of the whole genome (503.8 Mb). Although a biosynthetic gene cluster similar to noscapine biosynthesis in *Papaver somniferum* (opium poppy)

and steroidal glycoalkaloid biosynthesis in *Solanum lycopersicum* (tomato) and *Solanum tuberosum* (potato) (Itkin et al. 2013; Winzer et al. 2012) were not found in the California poppy genome scaffolds, we found several cytochrome P450 gene clusters and their novel function in the macarpine biosynthesis pathway (Hori et al. 2018). Generally, the evolution of specialized metabolism is considered to be related to gene duplication and diversification of key enzyme genes such as cytochrome P450. Our draft genome sequence data indicated that certain P450 family genes, for example CYP80, CYP82, and CYP719, which encode important enzymes involved in the formation of the specific chemical structure of various BIAs, are enriched in the California poppy genome compared to *Arabidopsis thaliana* (Hori et al. 2018).

On the other hand, further investigation of enzyme genes involved in the morphine biosynthesis pathway in *P. somniferum*, the same member of the Papaveraceae plant that produces morphine, showed no homologous genes for the production of morphine in the genome of California poppy (Hori et al. 2018), which cannot produce morphinan alkaloids. The California poppy genome is valuable for studying the evolution of specialized metabolite biosynthesis associated with gene diversification as well as that of basal eudicot owing to its relatively small genome size (Soltis et al. 2009).

In this report, we analyzed the structure and characteristics of the draft genome of the California poppy in comparison with the genomes of other plant species including the following: *Nelumbo nucifera* (sacred lotus) and tomato, which produce BIAs and steroidal glycoalkaloids, respectively (Friedman 2002; Nakamura et al. 2013); well-researched model plant, *A. thaliana*; and the most primitive Angiosperm, *Amborella trichopoda* (Amborella). The comparisons were made using gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. To focus on the exploration of BIA biosynthesis-related genes, we compared the California poppy genes, using orthology analysis, with those of other BIA-producing plants, specifically *Aquilegia coerulea*, a member of the Ranunculaceae plant family, and opium poppy, whose both

genome sequences have been recently revealed (Filiault et al. 2018; Guo et al. 2018). Finally, we discussed the usefulness of the California poppy genome to discover candidate genes involved in BIA biosynthesis.

## Materials and methods

### Draft genome sequencing and gene prediction of the California poppy

*E. californica* plant and genomic DNA were prepared and sequenced as described previously (Hori et al. 2018). The raw Illumina sequence reads were assembled using PLATANUS v1.2.11 (Kajitani et al. 2014) and the training sets were constructed using BRAKER v1.3 and RNA-Seq data. The protein-coding genes were predicted by Augustus v3.0.3 using the training set. Transfer RNA genes were searched using tRNAscan-SE v1.3.1 and pseudogenes were filtered out (Lowe and Chan 2016). Ribosomal RNA genes were searched using BLAST with an E-value cutoff of 1E-10 using 5.8S rRNA and 25S rRNA (accession number: X52320.1) and 18S rRNA (accession number: X16077.1) as queries.

### Detection of repetitive sequences

Repetitive sequences were detected using the following steps: first, repetitive sequences were detected using RepeatScout v1.05 (Price, Jones and Pevzner 2005), and subsequent repetitive sequences were searched against Repbase (Bao, Kojima and Kohany 2015) using RepeatMasker v4.0.6 (<http://www.repeatmasker.org>). The repetitive sequences found using RepeatMasker were classified into known repeats, while those found only using RepeatScout were classified into unique repeats in the California poppy.

### Functional classification of genes

Functional domains located in the predicted amino acid sequences were searched against the InterPro database using InterProScan v4 (<https://github.com/ebi-pf-team/interproscan>). The predicted gene sequences were classified into the euKaryotic clusters of Orthologous Group (KOG) categories and the plant GO slim categories (Ashburner et al. 2000; Tatusov et al. 2003). Comparison of orthologous gene clusters of California poppy with those of *A. coerulea*, *P. somniferum*, and *A. thaliana* was performed using OrthoFinder v2.3.1 (Emms and Kelly 2015).

### Comparison of metabolic pathways

Gene sequences obtained from genome databases (*A. trichopoda*: AmTr v1.0; *N. nucifera*: lotus marker; tomato: ITAG2.4; *A. thaliana*: Araport 11), as well as those of the predicted California poppy genes, were compared and mapped onto the KEGG reference pathways using BLAST searches against the GENES database in the KEGG. The criteria used were an E-value cutoff of 1E-10, length coverage  $\geq 25\%$ , and identity  $\geq 50\%$ , and the status of mapping was compared among the 5 plant species. Information on the programs and databases is listed in Table S1.

## Results and discussion

### Prediction of RNA-encoding genes

The draft genome sequence of the California poppy cv. “Hi-toezaki” was determined using the Illumina HiSeq 2500 systems

and assembled to 489 Mb, which is equivalent to 97% of the estimated genome size (503.8 Mb). The 41 612 protein-encoding genes predicted using Augustus showed significant sequence similarity to the registered genes, as reported in a previous manuscript (<http://eschscholzia.kazusa.or.jp>) (Hori et al. 2018). Based on the prediction of protein-encoding genes, the genome comparison was carried out as described below.

When RNA-coding genes were searched (67 genes for rRNAs, made up of 3 genes for 5.8S RNA, 24 genes for 18S RNA, and 40 genes for 25S RNA), 1343 intact genes for tRNAs were identified in the assembled genomic sequences ECA\_r1.0 (Table S2). Comparison of the number of genes for tRNAs in the genomes of other plant species indicated that the California poppy genome contains relatively more genes than other plant species such as *Amborella* (651), *sacred lotus* (960), *A. thaliana* (629), tomato (887), *Dianthus caryophyllus* L. (carnation) (1050), and *Macleaya cordata* (815). The opium poppy genome contains more tRNA genes (5467) than the California poppy genome possibly owing to its larger genome size (Table S3).

### Repetitive sequences

When the repetitive sequences were searched, the assembled California poppy genomic sequences comprised 52 683 626 bp known repeats including 64 193 bp Class I short interspersed nuclear elements (SINEs), 2 426 826 bp long interspersed nuclear elements (LINEs), and 32 671 650 bp long terminal repeat (LTR) elements (containing 24 796 251 bp *Copia* and 5 293 043 bp *Gypsy* types). The sequences also comprised 3 480 271 bp Class II DNA elements, 7 493 549 bp simple repeats, 432 877 bp Helitrons, and 2 450 250 bp low-complexity sequences. In the known repeat sequences, LTR elements were observed most frequently. However, 183 872 768 bp novel repeats were defined by *de novo* repeat finding, the sum of which corresponds to 48.4% of the assembled genomic sequences (Table 1).

Comparison of the transposable element (TE) contents of the California poppy genome with those of other plant species, such as *Amborella*, *sacred lotus*, *A. thaliana*, tomato, carnation, opium poppy, and *M. cordata* (Table S4), showed that the relative content of known TEs in the California poppy genome was lower than that of unclassified repeats. In particular, the percentage of LTR elements in the whole California poppy genome was 6.7%, which is lower than that in *Papaveraceae* plants, opium poppy (45.6%), and *M. cordata* (26.6%) genomes. Simple repeats appeared to be relatively high, similar to that of carnations. The lower content of known TEs in the assembled genomic sequences in the California poppy may be attributed to the escape of TE sequences during the process of sequence assembly, since the California poppy assemblies in this study were highly fragmented at the positions of potential repetitive sequences, as mentioned in the carnation genome analysis (Yagi et al. 2014).

### Comparison of California poppy genes with those of other plant species

To explore the genomic characteristics of the California poppy, we compared the predicted genes of California poppy, including not only metabolic pathways but also other plant processes, to those of other angiosperm plant species, namely *Amborella*, *sacred lotus*, tomato, and *A. thaliana*. The complete gene sets of these groups (26 846, 26 685, 34 725, and 27 655 protein-encoding genes, respectively), were retrieved from the AmTr v1.0, lotus marker, ITAG2.4, and Araport11 databases, respectively

Table 1. Repetitive sequences found in the California poppy draft genome

					Length occupied (bp)	Percent of whole genome (%)
Known repeats	Interspersed repeats	Class I	SINE	Total	64 193	0.01
			LINE	Total	2 426 826	0.50
			LTR elements	Copia	24 796 251	5.1
				Gypsy	5 293 043	1.1
				Total	32 671 650	6.7
		Class II	DNA elements		3 480 271	0.71
		Unclassified			69	0
		Helitrons			432 877	0.09
		Low complexity			2 450 250	0.50
		Simple repeat			7 493 549	1.5
Unique repeats	Subtotal	Unknown			26 147	0.01
					52 683 626	10.8
		Simple repeat			92 522	0.02
		Unknown			183 780 246	37.6
Total	Subtotal				183 872 768	37.6
					236 556 394	48.4

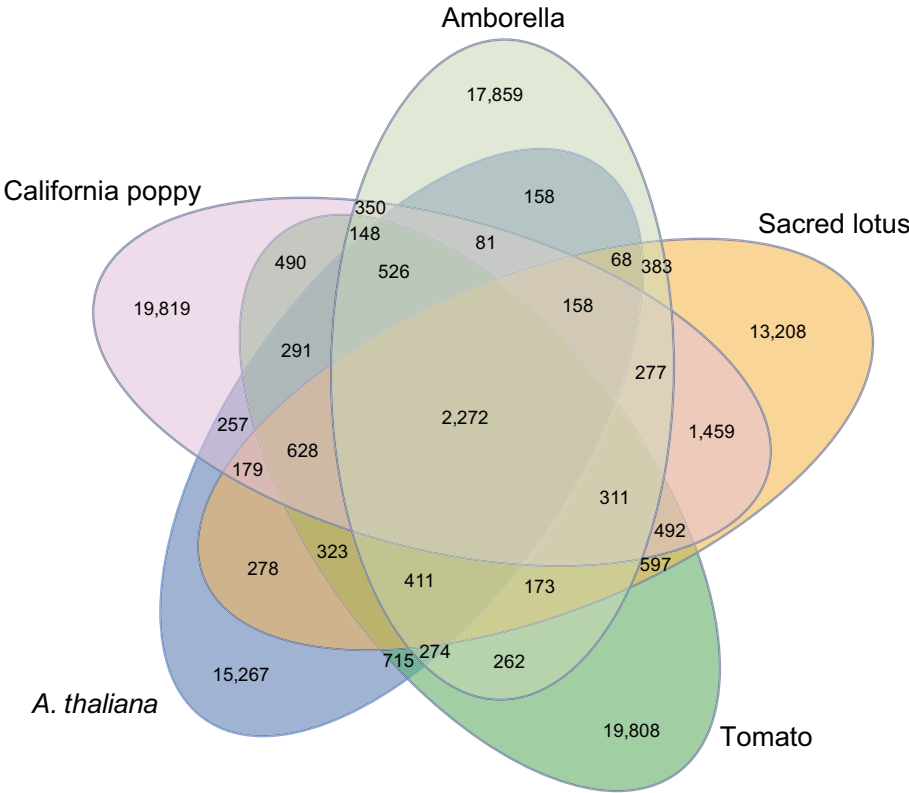


Figure 1. Classification of gene families in 5 plant species. Numbers in the Venn diagram show unique and shared gene families in each species.

(Table S1). The translated amino acid sequences of 41 612 coding sequences (CDSs), including splicing variants in the California poppy, were searched against the complete gene sets of Amborella, sacred lotus, tomato, and *A. thaliana* using BLAST with an E-value cutoff of  $1E-10$ . To create a Venn diagram of homologous gene clusters, we used the CD-HIT program (Figure 1). Only 4 genes of *N. nucifera* and 17 genes of *A. thaliana* were excluded through the CD-HIT program; however, a total of 97 522 clusters were generated (Table S5). As shown in Figure 1, 19 819 clusters (containing 29 682 California poppy genes) were unique to *E. californica*, whereas 2272 gene clusters were shared with the other

4 plant species. The highest number (1459 clusters) of shared sequence clusters with *E. californica* was found in *N. nucifera*, which is the closest species and BIA-producing plant. The number of shared gene clusters with tomato (490), Amborella (350), and *A. thaliana* (257) is similar, which is relatively consistent with their phylogenetic relationship.

To further compare the distribution of genes between the California poppy and the other 4 plant species, we classified the protein-encoding genes of all 5 species into KOG categories (Table S6). The number of genes classified into KOG (excluding genes in the “Poorly characterized” category) in California poppy,

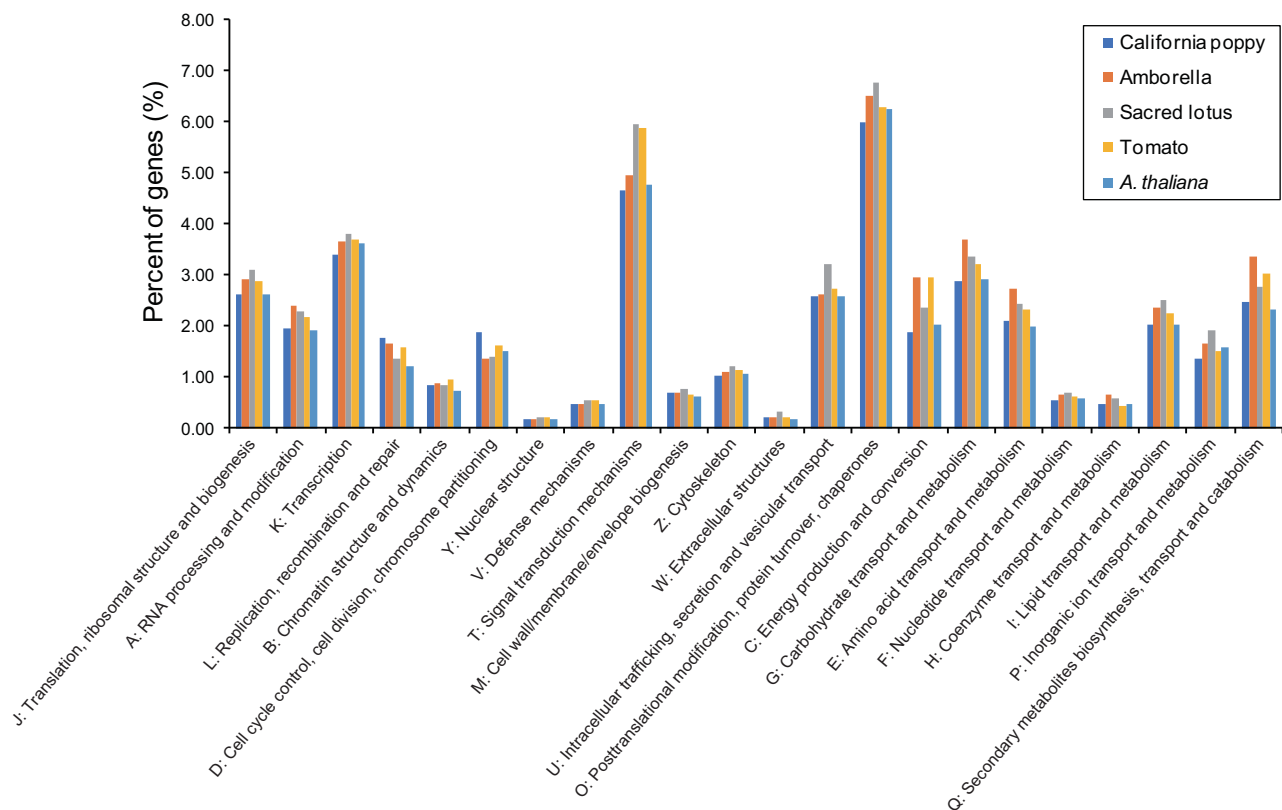


Figure 2. Gene assignment to KOG functional categories in California poppy.

Amborella, sacred lotus, tomato and *A. thaliana* was 16 596 (42.1%), 9606 (47.6%), 12 497 (48.4%), 13 955 (46.9%), and 11 547 (41.7%), respectively. Compared to the other 4 species, the California poppy genes categorized in KOG D (cell-cycle control, cell division, and chromosome partitioning) and R (general function prediction only) were relatively high (Figure 2 and Figure S2). The ratio of the genes in the other KOG categories, including KOG Q (secondary metabolites biosynthesis, transport, and catabolism), was similar among the 5 plant species.

To examine the classification of the genes based on GO slim, we assigned genes of the 5 plant species to GO terms. In the root category, the ratio of California poppy genes in the biological process (BP), cellular component (CC), and molecular function (MF) categories were 27.0%, 12.1%, and 43.2%, respectively, which were relatively close to that of Amborella genes. The distribution of gene ratio was quite similar among the 5 plant species (Figure 3a). The detailed results of GO classification showed that the genes of California poppy categorized in the BP, CC, and MF categories were also close to those of other plant species (Figure 3b–d and Table S7).

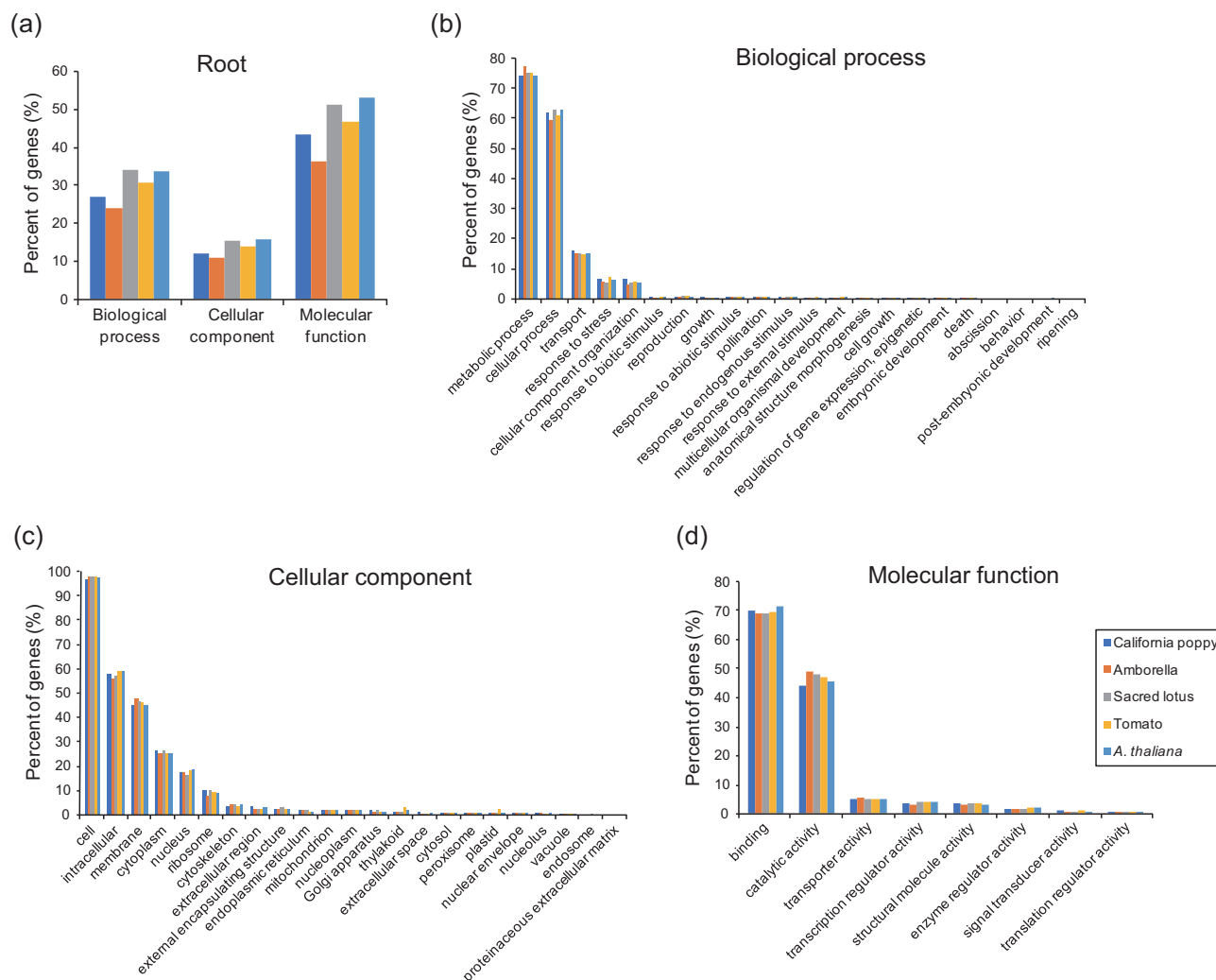
To investigate the classification of genes based on the enzymatic function of their proteins, we mapped the complete gene sets of the California poppy, Amborella, sacred lotus, tomato, and *A. thaliana* onto KEGG reference pathways. Results indicated that 1, 708, 838, 449, 614, and 633 genes of the California poppy were assigned into 5 major KEGG orthology (KO) categories, namely metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems (Figure S3a). The number of genes of the other 4 plants in

each major category was close to that of the California poppy genes. In the metabolism category, the ratio of genes belonging to the biosynthesis of other secondary metabolites in the California poppy and sacred lotus was relatively high, whereas those belonging to energy metabolism in California poppy, Amborella, and sacred lotus was clearly low compared to the other plants (Figure S3b and Table S8). A more detailed analysis revealed that the percentage of California poppy genes mapped onto “Isoquinoline alkaloid biosynthesis” was apparently higher than that of the other 4 plant species (Figure 4). In fact, many genes mapped only in the California poppy were found in the pathway maps downstream of (S)-scoulerine, which is an important intermediate of BIAs produced in the California poppy (Figure S4). Sacred lotus also produces BIAs, but has only a small number of genes mapped onto “Isoquinoline alkaloid biosynthesis” compared to the California poppy, indicating that the diversification of BIAs found only in the California poppy related to the acquisition of such specific genes during evolution.

### Orthology analysis of gene families in BIA-producing plants

Recently, genome sequences of opium poppy and *A. coerulea*, which are BIA-producing plants, have been uncovered (Filiault et al. 2018; Guo et al. 2018). To further examine the orthologous gene families, focusing on the genes involved in the biosynthesis of BIA among BIA-producing plant species, we performed orthologous gene family analysis with the OrthoFinder program using



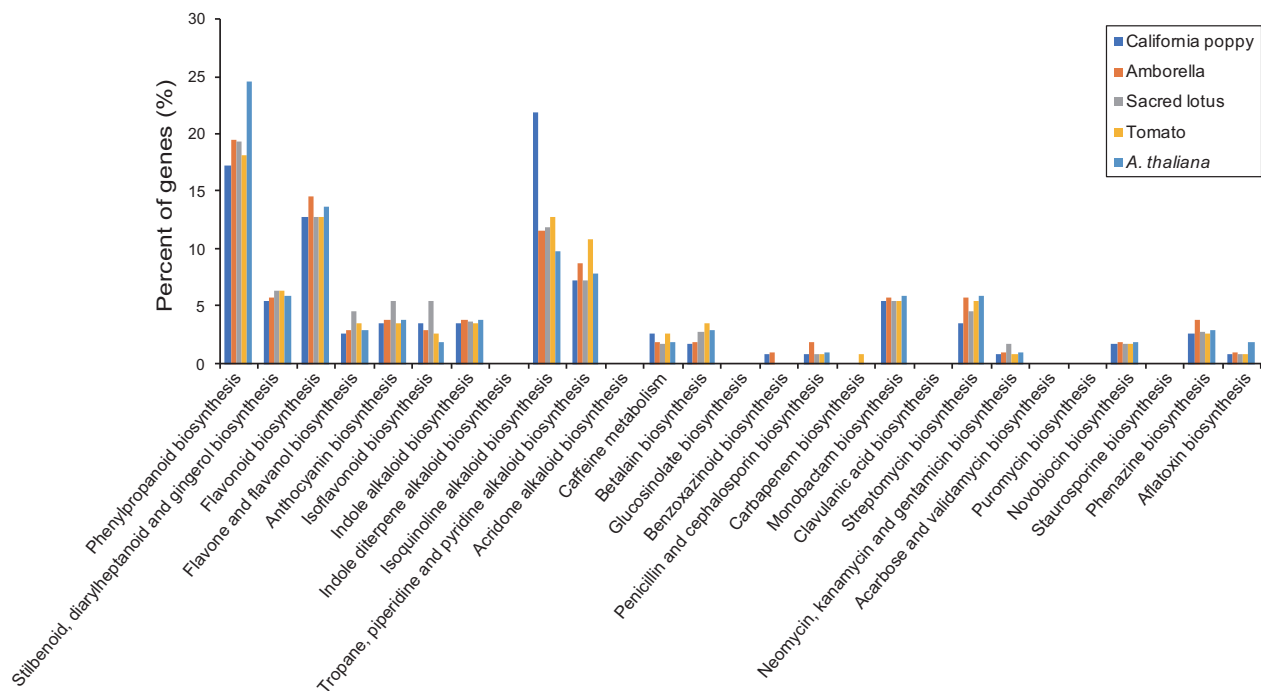


**Figure 3.** Classification of GO categories. The % of genes classified into root (a), biological process (b), cellular component (c), and molecular function (d) categories, respectively, is shown.

complete gene sets of California poppy (41 612 protein-encoding genes), opium poppy (51 212), and *A. coerulea* (30 023), compared with *A. thaliana* (27 655) as a representative non-BIA-producing plant (Figure 5). The results showed that 16 901, 22 280, 9738, and 8815 gene cluster groups were unique to each species, respectively. These included 17 476, 22 871, 10 088, and 9394 genes, respectively, and 9980 gene cluster groups that were shared by all 4 plant species. The number of orthologous groups shared between the California poppy and either the opium poppy (743) or *A. coerulea* (349) was higher than that between the California poppy and *A. thaliana* (101). The orthologous groups shared by California poppy, opium poppy, and *A. coerulea* were the highest (824) compared to that of the other combinations including *A. thaliana*. A large number of shared genes suggests that not only 3 BIA-producing plants are closely related species, but that many genes in these groups might be diversified and involved in the biosynthesis of specific metabolites.

We extracted the orthologous groups containing genes annotated to BIA biosynthetic enzymes (see Figure S1) and listed these in Table S9. Many groups containing predicted biosynthetic enzyme genes, such as tyrosine/DOPA decarboxylase, (S)-norcoclaurine synthase, (S)-norcoclaurine 6-O-

methyltransferase, (S)-N-methylcoclaurine 3'-hydroxylase, (S)-3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase, berberine bridge enzyme, (S)-stylopine synthase, (S)-canadine synthase, (S)-tetrahydroprotoberberine N-methyltransferase, and protopine 6-monooxygenase, were shared by the California poppy, opium poppy, and *A. coerulea*. However, the orthologous groups of genes encoding (S)-cheilanthifoline synthase, (S)-reticuline-7-O-methyltransferase, dihydrosanguinarine 10-hydroxylase, and dihydrobenzophenanthridine alkaloid 10-hydroxylase, that of the gene encoding pavine N-methyltransferase, and that of the gene encoding (S)-scoulerine 9-O-methyltransferase, were shared by the California poppy-opium poppy, California poppy-*A. coerulea*, and opium poppy-*A. coerulea*, respectively. The orthologous groups found in 2 plant species contain genes encoding specific proteins involved in the branch point or downstream in the biosynthetic pathway, which suggests that these enzymes might be implicated in the diversification of specific metabolites found in each plant. A recent phylogenomic approach has also suggested that the gene-encoding enzymes involved in a common pathway of BIA biosynthesis might be present in the common ancestor (Li et al. 2020). Furthermore, the orthologous group containing a unique



**Figure 4.** Classification of genes using KEGG reference pathways. The percentage of classified genes in each pathway in the “Biosynthesis of other secondary metabolites” category is shown.

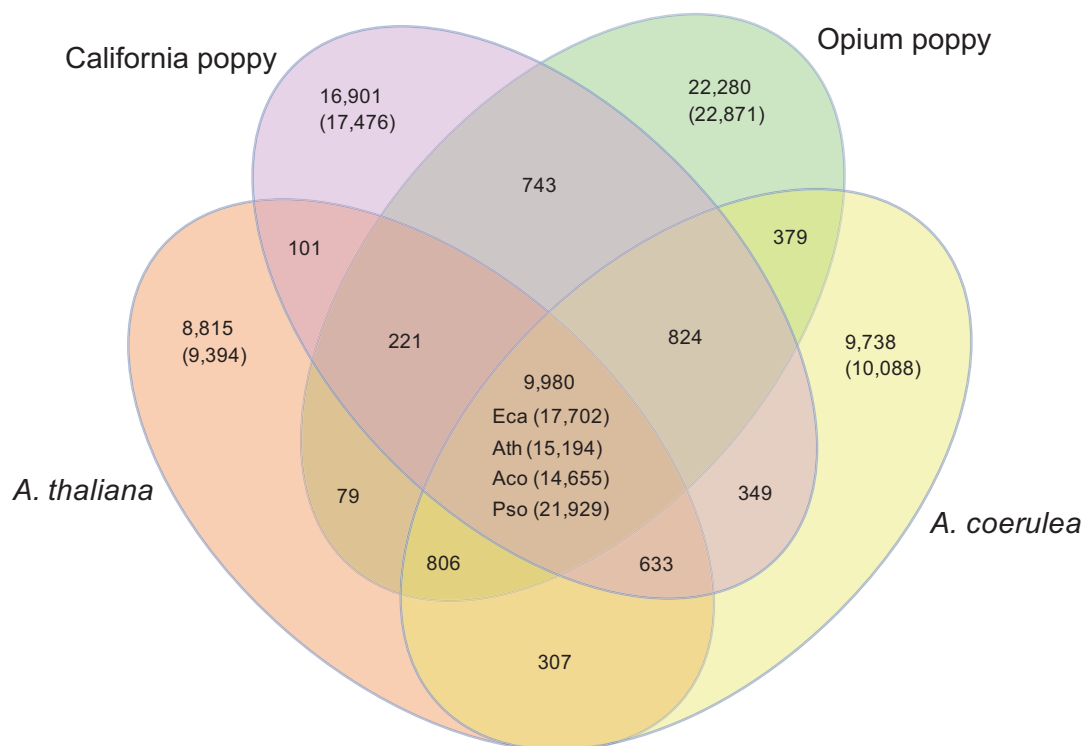
type of basic helix–loop–helix transcription factor gene to BIA biosynthesis in *Coptis japonica*, CjbHLH1, was also found in 3 BIA-producing plants (Yamada et al. 2011). This result indicates the specific distribution of CjbHLH1 homologs in BIA-producing plants, as previously reported (Yamada, Koyama and Sato 2011).

To explore candidate genes related to BIA biosynthesis, we selected representative orthologous groups with functional annotation shared by 3 BIA-producing plant species. As shown in Table S10, the genes encoding not only BIA biosynthesis-related proteins (highlighted in yellow) but also uncharacterized enzymes, including dehydrogenases, cytochrome P450s, and methyltransferases, were found in the orthologous groups. Furthermore, the orthologous groups containing predicted transcription factor-encoding genes, for example bHLHs, MYBs, and AP2/ERF/ethylene responsive factors (AP2/ERFs), and predicted transporter-encoding genes, such as ABC transporters, multidrug and toxic compound extrusion (MATE) transporters, and purine permeases, were also found. It has been reported that AP2/ERF transcription factors belonging to the group IX subfamily regulate the biosynthesis of other alkaloids such as nicotine, indole alkaloids, and glycoalkaloids (van der Fits and Memelink 2000; Shoji, Kajikawa and Hashimoto 2010; Cárdenas et al. 2016). The orthologous group, OG0011259, contains predicted group IX AP2/ERF genes, including 1 California poppy gene. In *C. japonica*, B-type ABC transporters, CjABCB1 and CjABCB2, and a MATE transporter, CjMATE1, are involved in the translocation and compartmentation of berberine (Shitan et al. 2003; Shitan et al. 2013; Takanashi et al. 2017). The orthologous groups, OG0008001 and OG0012836, contain B-type ABC transporter genes that are homologous to CjABCB1, and the orthologous groups, OG0001473 and OG0007280, contain MATE-type transporter genes that are homologous to CjMATE1. Recently, novel BIA uptake purine permeases (BUPs) have been isolated from the opium poppy and their functions have been characterized (Dastmalchi et al. 2019). Interestingly, BUP1 gene belongs to the orthologous group,

OG000544, which contains 1 California poppy gene. The other group, OG000703, which contains 10 California poppy genes, also includes predicted purine permease genes. These orthologous genes might be involved in the biosynthesis and diversification of specific metabolites. The functional characterization of these uncharacterized enzymes and transcription factor genes is ongoing (Yamada et al. 2020).

## Conclusions

*E. californica* (California poppy), a member of the Papaveraceae plant species, is a useful model plant to study both floral organ development and evolution in basal eudicot, and the biosynthesis of BIAs including pharmaceutically important chemicals. The elucidation of draft genome sequences of the California poppy facilitates these studies. To characterize the features of the California poppy genome, we performed comparative analyses using several plant species. Our repetitive sequence analysis indicated that LTR elements were observed less frequently in the California poppy than in other plants. When we performed a functional classification of predicted protein-encoding genes with KOG and GO categories, there was no apparent difference among the California poppy, Amborella, sacred lotus, tomato, and *A. thaliana* genes. However, further classification of predicted genes using the KEGG reference pathway revealed that the ratio of California poppy genes categorized in “Biosynthesis of other secondary metabolites,” particularly “Isoquinoline alkaloid biosynthesis” was higher than that of other plant genes. Orthology analysis using recently revealed genome sequences of BIA-producing plants, *P. somniferum* and *A. coerulea*, showed that a large number of orthologous groups shared by 3 BIA-producing plants were found. Some of these contain not only BIA biosynthetic enzyme genes and a unique-type of bHLH transcription factor gene, but also various gene encoding enzymes,



**Figure 5.** Classification of gene families in California poppy, opium poppy, *A. coerulea*, and *A. thaliana*. Numbers in the Venn diagram show unique and shared gene cluster groups, and those in parentheses show the number of genes in each species.

e.g. methyltransferases and P450 proteins, transcription factors (bHLH, MYB, and AP2/ERF proteins), and transporters (B-type ABC proteins), MATEs, and purine permeases; therefore, these genes might be involved in BIA biosynthesis. This study also suggests that the specific distribution of metabolism in limited plant species might be attributed to gene diversification during evolution. Our findings obtained from comparative genome analyses provide a useful resource for identifying and characterizing specific genes involved in specific metabolism in future studies.

### Supplementary material

Supplementary material is available at [Bioscience, Biotechnology, and Biochemistry](#) online.

### Data availability

The genome and gene sequences of the California poppy are available from the *Eschscholzia* Genome DataBase (<http://eschscholzia.kazusa.or.jp>). In the database, BLAST searches against the genome and gene sequences and keyword searches against the annotation are also available.

### Author contribution

Y.Y. and F.S. conceived and designed the research. H.H., Y.M, and A.T. conducted draft genome sequencing of the California poppy. H.H. performed comparative analysis using programs and acquired original datasets. Y.Y., H.H., and F.S. analyzed the data, and Y.Y. and F.S. wrote the manuscript. N.S. and F.S. discussed the results and contributed to the improvement of the manuscript. All authors have reviewed the manuscript.

### Funding

This research was supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) (Grant-in-Aid for Scientific Research (S) 26221201 to F.S.). The draft genome sequencing of the California poppy was supported by MEXT KAKENHI (No. 221S0002).

### Disclosure statement

No potential conflict of interest was reported by the authors.

### References

- Ashburner M, Ball CA, Blake JA et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25-9.
- Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 2015;**6**:11.
- Beaudoin GAW, Facchini PJ. Isolation and characterization of a cDNA encoding (S)-cis-N-methylstylophine 14-hydroxylase from opium poppy, a key enzyme in sanguinarine biosynthesis. *Biochem Biophys Res Commun* 2013;**431**:597-603.
- Becker A, Gleissberg S, Smyth DR. Floral and vegetative morphogenesis in California poppy (*Eschscholzia californica* Cham.). *Int J Plant Sci* 2005;**166**:537-55.
- Cárdenas PD, Sonawane PD, Pollier J et al. GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat Commun* 2016;**7**:10654.
- Carlson JE, Leebens-Mack JH, Wall PK et al. EST database for early flower development in California poppy (*Eschscholzia californica* Cham., Papaveraceae) tags over 6,000 genes from a basal eudicot. *Plant Mol Biol* 2006;**62**:351-69.



- Dastmalchi M, Chang L, Chen R et al. Purine permease-type benzylisoquinoline alkaloid transporters in opium poppy. *Plant Physiol* 2019;**181**:916-33.
- Dittrich H, Kutchan TM. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. *Proc Natl Acad Sci* 1991;**88**:9969-73.
- Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;**16**:157.
- Filialt DL, Ballerini ES, Mandáková T et al. The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *Elife* 2018;**7**:e36426.
- Friedman M. Tomato glycoalkaloids: role in the plant and in the diet. *J Agric Food Chem* 2002;**50**:5751-80.
- Guo L, Winzer T, Yang X et al. The opium poppy genome and morphinan production. *Science* 2018;**362**:343-7.
- Hagel JM, Beaudoin GA, Fossati E et al. Characterization of a flavoprotein oxidase from opium poppy catalyzing the final steps in sanguinarine and papaverine biosynthesis. *J Biol Chem* 2012;**287**:42972-83.
- Hagel JM, Facchini PJ. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol* 2013;**54**:647-72.
- Hori K, Yamada Y, Purwanto R et al. Mining of the uncharacterized cytochrome P450 genes involved in alkaloid biosynthesis in California poppy using a draft genome sequence. *Plant Cell Physiol* 2018;**59**:222-33.
- Ikezawa N, Iwasa K, Sato F. Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylophine biosynthesis in *Eschscholzia californica*. *FEBS J* 2007;**274**:1019-35.
- Ikezawa N, Iwasa K, Sato F. CYP719A subfamily of cytochrome P450 oxygenases and isoquinoline alkaloid biosynthesis in *Eschscholzia californica*. *Plant Cell Rep* 2009;**28**:123-33.
- Itkin M, Heinig U, Tzfadia O et al. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 2013;**341**:175-9.
- Kajitani R, Toshimoto K, Noguchi H et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* 2014;**24**:1384-95.
- Kutchan TM. Alkaloid biosynthesis-The basis for metabolic engineering of medicinal plants. *Plant Cell* 1995;**7**:1059-70.
- Li Y, Winzer T, He Z et al. Over 100 million years of enzyme evolution underpinning the production of morphine in the Papaveraceae family of flowering plants. *Plant Commun* 2020;**1**:100029.
- Liscombe DK, Facchini PJ. Molecular cloning and characterization of tetrahydroprotoberberine cis-N-methyltransferase, an enzyme involved in alkaloid biosynthesis in opium poppy. *J Biol Chem* 2007;**282**:14741-51.
- Lowe TM, Chan PP. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 2016;**44**:W54-57.
- Nakamura S, Nakashima S, Tanabe G et al. Alkaloid constituents from flower buds and leaves of sacred lotus (*Nelumbo nucifera*, Nymphaeaceae) with melanogenesis inhibitory activity in B16 melanoma cells. *Bioorg Med Chem* 2013;**21**:779-87.
- Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics* 2005;**21**:i351-8.
- Sato F. Characterization of plant functions using cultured plant cells, and biotechnological applications. *Biosci Biotechnol Biochem* 2013;**77**:1-9.
- Sato F. *Plant Alkaloid Engineering*. Liu H-W and Begley TP ed. Natural Products III Oxford: Elsevier, 2020, 700-55.
- Shitan N, Bazin I, Dan K et al. Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proc Natl Acad Sci USA* 2003;**100**:751-6.
- Shitan N, Dalmas F, Dan K et al. Characterization of *Coptis japonica* CjABCB2, an ATP-binding cassette protein involved in alkaloid transport. *Phytochemistry* 2013;**91**:109-16.
- Shoji T, Kajikawa M, Hashimoto T. Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* 2010;**22**:3390-409.
- Soltis DE, Albert VA, Leebens-Mack J et al. Polyploidy and angiosperm diversification. *Am J Bot* 2009;**96**:336-48.
- Takanashi K, Yamada Y, Sasaki T et al. A multidrug and toxic compound extrusion transporter mediates berberine accumulation into vacuoles in *Coptis japonica*. *Phytochemistry* 2017;**138**:76-82.
- Takemura T, Ikezawa N, Iwasa K et al. Molecular cloning and characterization of a cytochrome P450 in sanguinarine biosynthesis from *Eschscholzia californica* cells. *Phytochemistry* 2013;**91**:100-8.
- Tatusov RL, Fedorova ND, Jackson JD et al. The COG database: an updated version includes eukaryotes. *BMC Bioinform* 2003;**4**:41.
- van der Fits L, Memelink J. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* 2000;**289**:295-7.
- Winzer T, Gazda V, He Z et al. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 2012;**336**:1704-8.
- Xiao M, Zhang Y, Chen X et al. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J Biotechnol* 2013;**166**:122-34.
- Yagi M, Kosugi S, Hirakawa H et al. Sequence analysis of the genome of carnation (*Dianthus caryophyllus* L.). *DNA Res* 2014;**21**:231-41.
- Yamada Y, Kokabu Y, Chaki K et al. Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in *Coptis japonica*. *Plant Cell Physiol* 2011;**52**:1131-41.
- Yamada Y, Koyama T, Sato F. Basic helix-loop-helix transcription factors and regulation of alkaloid biosynthesis. *Plant Signaling Behavior* 2011;**6**:1627-30.
- Yamada Y, Motomura Y, Sato F. CjBHLH1 homologs regulate sanguinarine biosynthesis in *Eschscholzia californica* cells. *Plant Cell Physiol* 2015;**56**:1019-30.
- Yamada Y, Nishida S, Shitan N et al. Genome-wide identification of AP2/ERF transcription factor-encoding genes in California poppy (*Eschscholzia californica*) and their expression profiles in response to methyl jasmonate. *Sci Rep* 2020;**10**:18066.