Plant Mol Biol
DOI 10.1007/s11103-006-9025-y

EST database for early flower development in California poppy (*Eschscholzia californica* Cham., Papaveraceae) tags over 6000 genes from a basal eudicot

John E. Carlson · James H. Leebens-Mack · P. Kerr Wall · Laura M. Zahn · Lukas A. Mueller · Lena L. Landherr · Yi Hu · Daniel C. Ilut · Jennifer M. Arrington · Stephanie Choirean · Annette Becker · Dawn Field · Steven D. Tanksley · Hong Ma · Claude W. dePamphilis

Received: 6 April 2006 / Accepted: 24 May 2006

Abstract The Floral Genome Project (FGP) selected California poppy (*Eschscholzia californica* Cham. ssp. *Californica*) to help identify new florally-expressed genes related to floral diversity in basal eudicots. A large, non-normalized cDNA library was constructed from premeiotic and meiotic floral buds and sequenced to generate a database of 9079 high quality Expressed Sequence Tags (ESTs). These sequences clustered into 5713 unigenes, including 1414 contigs and 4299 singletons. Homologs of genes regulating many aspects of flower development were identified, including those for organ identity and development, cell and tissue differentiation, cell cycle control, and secondary metabolism. Over 5% of the transcriptome consisted of homologs to known floral gene families. Most are the first representatives of their respective gene families in basal eudicots and their conservation suggests they are important for floral development and/or function. App. 10% of the transcripts encoded transcription factors and other regulatory genes, including nine genes from the seven major lineages of the important MADS-box family of developmental regulators. Homologs of alkaloid pathway genes were also recovered, providing opportunities to explore adaptive evolution in secondary products. Furthermore, comparison of the poppy ESTs with the *Arabidopsis* genome provided support for putative *Arabidopsis* genes that previously lacked annotation. Finally, over 1800 unique sequences had no observable homology in the public databases. The California poppy EST database and library will help bridge our under-

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at http://dx.doi.org/10.1007/s11103-006-9025-y.

J. E. Carlson (✉)
The School of Forest Resources and Huck Institutes for Life Sciences, Pennsylvania State University, 323 Forest Resources Building, University Park, PA 16802, USA
e-mail: jec16@psu.edu

J. H. Leebens-Mack · P. K. Wall · L. M. Zahn · L. L. Landherr · Y. Hu. J. M. Arrington · S. Choirean · H. Ma · C. W. dePamphilis
Department of Biology, The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

Present Address:
L. M. Zahn
SCIENCE, 1200 New York Avenue NW, Washington, DC 20005, USA

L. A. Mueller · S. D. Tanksley
Department of Plant Breeding, Cornell University, Ithaca, NY 14853, USA

D. C. Ilut
Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA

J. M. Arrington
Department of Biology, Randolph-Macon Woman's College, Lynchburg, VA 24503, USA

A. Becker
Evolutionary Developmental Genetics Group, University of Bremen, 28359 Bremen, Germany

D. Field
Oxford Centre for Ecology and Hydrology, Mansfield Road Oxford OX1 3SR, UK

J. H. Leebens-Mack
Department of Plant Biology, University of Georgia, Plant Sciences Building, Athens, GA 30602-7271, USA

standing of flower initiation and development among higher eudicot and monocot model plants and provide new opportunities for comparative analysis of gene families across angiosperm species.

Keywords EST database · Flower development · California poppy · Basal eudicot

Abbreviations

ABI	Applied Biosystems
AG	*AGAMOUS* gene
AGL	*AGAMOUS*-like gene
AP	*APETALA* gene
DEF	*DEFICIENS* gene
DEPC	diethylpyrocarbonate
EF-1-α	*Elongation Factor 1-alpha gene*
ESca	*Eschscholzia californica*
EST	Expressed Sequence Tag
FIM	*FIMBRIATA* gene
FLO	*FLORICAULA* gene
GO	Gene Ontology Consortium
GLO	*GLOBOSA* gene
Ks	rate of synonymous substitutions
LFY	*LEAFY* gene
Mbp	Million base pairs
MYA	Million years ago
MYB	*myeloblastosis*-like gene
NCBI	National Center for Biotechnology Information
PCR	polymerase chain reaction
PGN	Plant Genome Network
PI	*PISTILLATA* gene
PLE	*PLENA* gene
rRNA	ribosomal RNA
RACE	Rapid Amplification of cDNA Ends
RCA	Rolling Circle Amplification
RuBisCO	ribulose-1;5-bisphosphate carboxylase; ssp., species
TAIR	The Arabidopsis Information Resource web site

Introduction

Molecular genetic analyses over the last 20 years have uncovered dozens of genes that play important regulatory roles controlling normal flower development (Zhao et al. 2001; Ma 2005). In particular, studies using two model systems, *Arabidopsis thaliana*, a rosid, and *Antirrhinum majus*, an asterid, have led to the discovery of conserved genes that specify floral meristem identity and floral organ identity. The *Arabidopsis LFY* and *Antirrhinum FLO* genes form an orthologous pair that is required for normal floral meristem identity (Coen et al. 1990; Weigel et al. 1992). In addition, characterization of homeotic mutants in these two species supported the proposal of the now well-known ABC model for floral organ identity (Coen and Meyerowitz 1991; Ma 1994; Weigel and Meyerowitz 1994; Ma and dePamphilis 2000). Molecular analysis of these floral homeotic genes revealed that most of the genes required for ABC functions encode members of the MADS-box protein family, including the *Arabidopsis* AP1, AP3, PI, and AG proteins (Ma 1994; Weigel and Meyerowitz 1994). Furthermore, genetic studies in *Arabidopsis* have identified additional genes that regulate the floral homeotic genes, that affect floral meristem size and floral organ number, that promote floral organogenesis and ovule development, and that control meiosis and pollen development (Zhao et al. 2001; Ma 2005, 2006; Zahn et al. 2006).

Molecular experiments in a number of eudicots, including petunia, tomato, and *Brassica napus*, have demonstrated conservation of function among homologs of the floral MADS-box genes important for the ABC functions in *Arabidopsis* and *Antirrhinum* (Ma 1994; Ma and dePamphilis 2000; Zahn et al. 2005a; original data from: Angenent et al. 1993 (B in petunia); Angenent et al. 1994 (C in petunia)). For example, petunia and tomato homologs of the *Arabidopsis* C-function gene *AG* have been shown to be necessary and sufficient for C function within the flower. Similarly, homologs of B-function genes *AP3* and *PI* are often critical for specifying the identity of petals and/or stamens, indicating a conservation of the B function. On the other hand, conservation of A function among species has yet to be demonstrated.

Although dozens of genes important for normal flower development have been identified from genetic studies, the number of genes with critical functions in flowers is expected to be much greater (Goldberg et al. 1993). Many (if not most) genes with critical roles remain undiscovered because of functional redundancy (e.g., Pelaz et al. 2000), early essential functions, or low levels of expression. Genomics is a productive and efficient approach to discovering and characterizing the many genes involved in complex developmental processes such as flowering. However, the genomic resources that are available for studying flowering in plants are concentrated mainly among the major crops and their experimental models (such as *Arabidopsis*). To understand the origin and subsequent diversification of flowers, data on floral development is needed from species that better encompass the breadth and depth of the plant kingdom. The Floral Genome Project has systematically targeted phylogenetically critical lineages of

angiosperms and gymnosperms that are the missing links to complete a comparative analysis of major angiosperm lineages (Soltis et al. 2002; Albert et al. 2005).

Most eudicot species are found in the two large eudicot groups, the rosids and the asterids (Soltis et al. 2000), and much of the understanding of the molecular basis of flower development is derived from the studies of the rosid *Arabidopsis* and, to a lesser degree, the asterid *Antirrhinum*. Therefore, to achieve an evolutionary understanding of flower development and diversification among eudicots, it is critical to develop genomic resources for basal eudicot species that diverged earlier than the rosids and asterids. Genes captured from basal eudicots are valuable outgroups for comparative analyses of core eudicot lineages, allowing the evolutionary differences between rosids and asterids to be polarized and interpreted. The basal eudicots themselves also contain a wide range of floral variation (Soltis et al. 2006; Endress 2004; Becker et al. 2005). We have chosen *Eschscholzia californica* Cham. ssp. *californica* (California poppy; Papaveraceae) to provide a root for genomic-scale analyses for more derived eudicot species. Papaveraceae is a member of the order Ranunculales, which roots the eudicot tree as sister to all other eudicot lineages. Previous studies have already demonstrated the value of poppies as a model for floral molecular genetics and evolution (Kramer and Irish 2000; Soltis et al. 2002), and for understanding the evolution of self-incompatibility (Thomas and Franklin-Tong 2004). In addition, poppies are also used to understand the molecular basis of alkaloid (opiate) chemistry (e.g. Kutchan 1995; Sato et al. 2001; 2005; Park et al. 2002), and latex chemistry (e.g. Decker et al. 2000; Memelink 2004; Frick et al. 2005).

Eschscholzia californica offers some advantages as a model basal eudicot to the more widely studied opium poppy (*Papaver somniferum* L.). The California poppy has a smaller genome (1,115 Mbp, approximately 6.5 times the size of *A. thaliana*) (Bennett et al. 2000; Arumuganathan et al., unpublished) than the opium poppy (3724 Mbp) (Bennett and Smith 1976), is transformable (Park and Facchini 2000a, b; Lee and Pedersen 2001; Park et al. 2002) and transgenic California poppy can be regenerated to adult plants (Park and Facchini 2000a, b). California poppy is more generally accessible to experimentation as it does not require the government research permits needed to work with opium poppy, and has a variety of floral mutants available (Becker et al. 2005; Wakelin et al. 2003). As a facultative long day plant (Nanda and Sharma 1976) it is an easily cultivated species with a generation time of only about 3 months in constant light. *E. californica* is also the most widespread poppy species in North America, ranging in natural distribution from Washington and Oregon in Northwestern USA to the Southwestern and South-Central states to Northern Mexico (Clark 1993). It has been introduced to Chile, New Zealand, Tasmania and mainland Australia, where it has become naturalized. California poppy grows in a large range of different habitats and is highly variable in structure, and life history (Cook 1962). Additionally, shoot, inflorescence and flower morphogenesis has been studied in detail (Becker et al. 2005). This offers many opportunities for studying the genetic basis of ecologically and evolutionarily important adaptations.

We report the creation of a floral EST database for *E. californica* to support evolutionary genomics studies among the eudicots, and beyond. This new database is designed to facilitate the discovery and assessment of orthology and function of florally active genes. Furthermore, it should provide tools to more broadly test current models of floral organ determination developed in model plants. Elucidating the genetics of floral development in this key lineage not only will help to answer questions regarding the origin and diversification of flowers, but should also provide opportunities to better understand the highly derived eudicot and monocot model systems and the basis of the large genetic and developmental differences that exist among them.

Materials and methods

Tissue source

Eschscholzia californica cv. ''Aurantica Orange'' (J.L. Hudson Seedsman) was grown from seed for 10 weeks under greenhouse conditions with 16 h light/8 h dark and app. 23°C at the Pennsylvania State University. Flower buds from 20 plants were collected in the size range of < 1.0–2.5 mm in diameter covering all of the pre-meiotic and meiotic stages of flower bud development (floral stages were as defined by Becker et al. 2005—organ initiation when the buds are below 650 μm in diameter; microsporangia initiate at stage 6 at 1 mm; male meiosis commences at stage 8 at 2.5 mm). Buds were immediately quick frozen in liquid nitrogen and stored at –80 upon harvest. Prior to RNA extraction, buds were sorted and pooled by size and weight to ensure equal approximately representation of stages of bud development.

RNA extraction

Total RNA was extracted from floral buds according to the manufacturer's protocol (http://www.ambion.com/

techlib/prot/bp_1911.pdf) for the RNAqueous®-Midi Kit (Ambion, Inc., catalog # 1911), with the following modifications. For each extraction, 200 mg of bud tissue was ground in an RNase free mortar and pestle chilled with liquid nitrogen. The frozen ground tissue was vortexed vigorously in lysis buffer that was premixed with Ambion's ''Plant Isolation Aid''® (250 uls/2000 uls) in a ratio of 8:1 to improve yields. A clarifying spin was conducted at 12,000*g* at 22°C for 10 min. Quality and purity of total RNA was determined by micro-capillary electrophoresis on the Agilent Bioanalyzer, according to manufacturers suggested protocol. RNA was precipitated using 0.1 volumes of sodium acetate and three volumes of 100% ethanol. RNA was dissolved into RNase-free (diethylpyrocarbonate (DEPC) treated) water and yields were determined by absorbance using an Eppendorf Biophotometer. Purified RNA was stored at –80°C.

mRNA isolation

Message RNA was extracted from total RNA according to the manufacturer's protocol (http://www.ambion.com/techlib/prot/bp_1916.pdf) for the Poly(A)Purist™ mRNA Purification Kit (Ambion, Inc., catalog # 1916). A total of 800 μg of total RNA (after QC) was added to each of 2 Poly(A) purist columns. After column purification, each aliquot of mRNA was ethanol precipitated using the method provided with the kit (including glycogen). The RNA was then resuspended separately in THE™ RNA storage solution (Ambion, Inc.). Resuspended mRNA was stored at –80°C. Quality Control on the mRNA, which was conducted in the same manner as for total RNA (above) confirmed that mRNA was intact and had no detectable DNA contamination. This approach yielded approx 24 μg of mRNA per gram of bud tissue.

cDNA Library construction

A directional cDNA library was constructed using the ZAP-cDNA® Synthesis Kit (Stratagene), according to manufacturer's instructions (http://www.stratagene.com/manuals/200401.pdf), except that 7 μg of mRNA was used rather than the recommended 5 μg. First strand synthesis was performed with 5-methyl dATP, producing hemimethylated cDNA, with unmethylated *Xho* I site on the primer. The mRNA was heat treated for 10 min at 65°C to relax secondary structure before annealing with the primer. *Eco*R I adapters (modified with a library-specific hexanucleotide signature sequence—CGAGCA) were ligated to the cDNA. The cDNA products were size fractionated through a drip column of Sepharose® CL-2B gel filtration medium (Stratagene), with cDNAs larger than 500 bp (mean size app. 1.6 kb) ligated into the EcoR I (5′) and Xho I (3′) sites of the Uni-ZAP® XR Lambda vector (Stratagene). The titer of the primary library was 7×10^6 total pfu. The library was amplified to a titer of 1.68×10^{11} per ml and then stored in 7% glycerol at –80°C. The pBluescript II SK(±) phagemid vector form of the library was excised from an aliquot of the lambda vector (into SOLR host cells) prior to library manipulation for DNA sequencing. The phagemid library was maintained under 100 mg/ml Ampicillin selection. Selection for white vs. blue plaques or colonies was preformed at all library titering steps, using X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactoside) with IPTG (isopropyl-beta-D-thiogalactopyranoside) for color selection. The average insert size was determined by PCR amplification with M13F/M13R primers and agarose gel electrophoresis. The cDNA library construction yielded a primary library of app. 7×10^6 total pfus and an amplified library with a titer of 1.68×10^{11} pfu per ml (with over 200 ml total volume). Analysis of the Eca01 library by PCR of 40 plasmids showed an average insert size of 1702 bp. Approximately 50,000 colonies were picked from the excised Eca01 library and stored in glycerol in 384-well plates at –80°C.

DNA Sequencing

Bacterial colonies of the excised phagemid library were picked from agar plates and automatically gridded to microtiter plates using a QPix2® robot (Genetix). Replica plates for the library were either directly entered into the sequencing queue or stored at –80°C in 8% glycerol. DNA template preparation and sequencing reactions were performed on 96-well plate format. DNA templates of cDNA inserts were prepared from overnight bacterial liquid cultures by Rolling Circle Amplification (RCA) of the pBluescript plasmids using TempliPhi™ DNA Amplification (Amersham) kits, following the manufacturer's protocol at one-quarter recommended volumes. Sequencing was conducted on RCA products using BigDye® Terminator v1.1 Cycle Sequencing Kits (Applied Biosystems) and T3 primer at one-sixth to one-quarter recommended reaction volumes. Sequencing reactions were purified using paramagnetic beads with CleanSEQ dye-terminator removal kits (Agencourt) according to the manufacturer's protocol but at one-half recommended volumes. Sequencing reaction products were automatically loaded and electrophoresed on an ABI PRISM® 3700 DNA Analyzer.

EST Informatics

The Plant Genome Network (PGN) at Cornell University provided automated processing, quality control, data archiving, unigene assembly, and library statistics. Raw sequence data trace files (chromatograms) were automatically uploaded to PGN from a project server at Penn State University on a daily basis. The first step in the EST analysis pipeline consisted of base calling using the Phred program (http://www.phrap.com/phred/), followed by removal of vector and library specific adaptor sequences, and low-quality regions, as defined by a phred score lower than 15. The trimming algorithm used was designed to extend the sequence as far as possible by performing an integration of phred quality scores over the entire sequence and selecting the sequence with the highest integrated score. Finally, any polyA tail detected in the trimmed sequence was trimmed to a maximum of 20 consecutive A's.

For quality control, the sequences were screened for contamination and sequence length and complexity. ESTs containing bacterial, rRNA, chloroplast, or mitochondrial sequence were identified by the BLASTN program of the NCBI-BLAST package (Altschul et al. 1990) and excluded from further analysis. Trimmed sequences were rejected when less than 150 bp in length and less than 96% of the base pairs were unambiguous. A sequence complexity check was also applied to identify the signature of sequencing errors specific to the ABI3700 machine. Sequences were discarded for low complexity if the same nucleotide accounted for over 60% of the sequence or the same two nucleotides accounted for 80% of the sequence. Sequences that failed any of the quality checks were marked and excluded from further analysis. To generate unigene sequence sets from the EST reads, two different clustering pipelines were used. The first pipeline was used to provide quick feedback on the sequencing process, the state of redundancy of ESTs, and gene coverage in the library, and was run each time new sequence was added for a given library. This unigene contig pipeline was based on the Phrap algorithm (http://www.phrap.com/). When sequencing for the poppy libraries was complete, a final unigene build was generated using a pipeline that was based on the cap3 assembler. This pipeline also integrated chimera screens to minimize the occurrence of chimeric unigenes. Statistics such as tentative unigene count, average trimmed sequence length, average unigene length and average GC content were calculated. To obtain preliminary functional annotation, the Arabidopsis and rice sequence databases were searched with the poppy unigene set using TBLASTX and matches with BLAST scores of e–9 or better were recorded. The analysis and unigene pipelines were tightly integrated with the database backend. The database is also the backbone of the PGN web interface.

Analysis of synonymous site divergence

To assess the history of gene duplication in California poppy, the complete poppy unigene set was analyzed in an all-against-all sequence similarity search using BLASTN. Following the procedures of Blanc and Wolfe (2004), all sequence pairs showing over 40% sequence similarity BLAST alignments over 300 bases in length were tentatively considered as paralogs. For all paralog pairs, any positions in the unigene sequences with quality scores lower than Q20 were masked and amino acid sequences were inferred using ESTscan (Iseli et al. 1999). Inferred amino acid sequences for paralog pairs were aligned using the Smith–Waterman algorithm (Smith and Waterman 1981). Nucleotide sequence alignments were then forced onto the amino acid alignments to recover codon structure. The level of synonymous divergence (K_s) between paired sequences was estimated using codeml in the PAML software package (Yang 1997) assuming that underlying codon frequencies were a function of nucleotide frequencies at each codon position (F3 × 4 model).

GO Classifications

We determined the classification of putative gene functions based on homology of translated poppy ESTs with the GenBank protein database. Pie charts were generated from GO annotations as follows: (1) ESTs were searched against the *Arabidopsis* TAIR dataset using BLASTx. Any matches with BLAST scores >1e–20 were discarded. (2) GO annotations for the best matches from TAIR were downloaded. (3) The map2slim.pl program was run with custom slims parameters for molecular function, process and component.

In situ hybridization

Fresh floral bud tissue was collected from poppy plants grown in the greenhouse and fixed as described by Lincoln et al. (2002) with the following modification—the denaturation and post-fixation steps were not performed. Expression of the *EScaGLO*, *EScaAG1*, and *EScaAGL2* genes were determined using buds from early in flower development (corresponding to stage 3 in *Arabidopsis*). For all in situ hybridization experiments, plasmid DNA was digested with restriction enzymes to remove the highly conserved MADS-box region and

probes were synthesized using T7 RNA polymerase, as directed. A control sense probe, hybridized in each experiment, was prepared from the *EScaGLO* clone by digesting the 3′ end of the clone and synthesizing the probe using T3 RNA polymerase.

Results

Eca01 Expressed sequence tag database

We conducted one-pass sequencing on 11,517 cDNA inserts. After quality control to eliminate sub-standard reads (see above), we obtained 9,079 high quality ESTs. After trimming of vector sequence and removal of low quality bases (< phred 15), the average length of ESTs used in unigene assembly was 471 bp. An assembly of overlapping, contiguous sequences yielded a total of 5,713 unigenes, of which 4,299 were singletons and 1414 were in contigs (unigene build number 5, 2005-04-14, http://pgn.cornell.edu/). The average unigene length was 593 bp (singlets averaged 464 bp), and the average GC content for the Eca01 data set was 41%, which is similar to GC contents reported for derived eudicot species such as *Arabidopsis* (Bennetzen et al. 2004). There were very few microbial, fungal, plastid or mitochondrial genes in the Eca01 EST data.

All of the 9079 high quality Eca01 EST sequences have been deposited into GenBank. The entries can be easily searched using the library name Eca01. The Eca01 entries in GenBank currently extend from locus identifier CD476387 (EST eca01-38ms3-c02, dbEST Id 18508682) to CK768257 (EST eca01-3cs4-a04, dbEST Id 21648484).

The redundancy of sequences encountered by our random sequencing approached 50% when sequencing of the Eca01 library was concluded. An attempt at screening the library by hybridization with probes for high copy number sequences did not reduce the amount of re-sequencing of clones nor did it significantly improve the frequency of new genes discovered (data not shown). Overall, 27% of genes were sampled more than once and 73% of the sequences were unique. The two most highly expressed unigenes (both for elongation factor 1-alpha) were sampled 182 and 40 times (Table 1). The next 18 most highly expressed genes (Table 1) were sampled between 16 and 32 times. For these most prevalent messages, over 80% of the ESTs in the unigenes contigs were at the 5′-end of the mRNA. Because the cDNA library was constructed by oligo(dT) priming from the 3′-end of the mRNA, the prevalence of 5′ sequences indicates that full length cDNA inserts were common in the library.

Codon usage

The program GCUA: General Codon Usage Analysis (http://bioinf.may.ie/gcua/index.html; McInerney 1998) was used to generate the codon usage among the California poppy translated EST sequences. Table 2 shows the number of times each of the codons was observed in the EST data set. The relative synonymous codon usage values are also shown for the Eca01 dataset in Table 2. The codon preferences that we observed in our *E. californica* EST database were virtually identical to the frequencies of codon use in *Arabidopsis thaliana* (Nakamura et al. 2000; the Codon Use Database at http://www.kazusa.or.jp/codon/, GenBank Release 145.0, January 25 2005), and consistent with GC content of California poppy. The codon preferences for California poppy indicate a long, conserved history of codon use and bias in flowering plants. However, the relative frequency of codons and codon usage reveals a bias in codon usage in California poppy in favor of particular codons over others (e.g. much greater use of CUU for Leucine than the CUG codon).

Ks analysis

Our analysis of synonymous site divergence between paralog pairs revealed striking evidence of a genome-wide duplication event sometime in the lineage leading to California poppy. A total of 269 paralog pairs were identified using the criteria of Blanc and Wolfe (2004). A plot (Fig. 1) of the frequency of K_s values was prepared, and truncated at $K_s = 2.0$ because per site estimates of synonymous divergence are unreliable above this level due to the influence of multiple substitutions at individual sites. The frequency distribution of K_s for these pairs was bimodal with the first mode possibly representing sequencing error and allelic variation and the second mode representing a concentration of gene duplication events (Blanc and Wolfe 2004). Over 45% of the paralog pairs had K_s values between 0.35 and 0.80. Applying a per site synonymous rate calibration of 1.5×10^{-8} per year (Koch et al. 2000), this concentration of K_s values may correspond to a genome wide duplication event occurring some time between 23 and 53 million years ago.

Highly expressed genes in *E. californica* floral buds

The 20 most highly expressed poppy unigenes, shown in Table 1, accounted for 10.62% of the high quality sequence reads. The two most highly expressed genes (unigenes 337159 and 337127; http://pgn.cornell.edu/cgi-bin/unigene/unigene_info.pl?build_id=75) in the

Table 1 Twenty most highly expressed poppy unigenes in the *E. californica* database[a]

Unigene number	# ESTs (% total)	Contig length	Arabidopsis protein with best Blastx match Additional species with strong Blastx alignments	aa Identity	Blastx score
337159	182 (2.0)	1794 bp	Elongation factor 1-alpha (At5g60390; et al.)	95%	0.0
			Stevia rebaudiana, *Lycopersicon esculentum*, *S. tuberosum*	97%	0.0
337127	40 (0.44)	1839 bp	Elongation factor 1-alpha (At5g60390; et al.)	95%	0.0
			Stevia rebaudiana, *Lycopersicon esculentum*, *S. tuberosum*	97%	0.0
337778	32 (0.35)	2623 bp	Histidine kinase (At5g35750.1)	51%	2e–157
			Cytokinin receptor in *Catharanthus roseus*	54%	7e–170
338311	30 (0.33)	667 bp	No match to Arabidopsis	–	–
			No matches to GenBank (Blastn or Blastx)	–	–
337919	28 (0.31)	1817 bp	4-Coumarate:CoA ligase (At1g62940)	67%	0.0
			Nicotiana sylvestris	65%	0.0
338158	26 (0.29)	1454 bp	Glyceraldehyde-3-phosphate dehydrogenase (At3g04120)	89%	6e–162
			Daucus carota, *Oryza sativa*, *Ranunculus acris*, *Magnolia quinquepeta*	91%	2e–165
337354	26 (0.29)	868 bp	S-Adenosylmethionine synthase (SAM1) (At1g02500)	90%	7e–100
			Nicotiana tabacum, *Lycopersicon esculentum*	95%	1e–147
337185	24 (0.26)	1158 bp	ADP, ATP carrier protein 2, mitochondrial (At5g13490)	78%	1e–149
			Solanum tuberosum	83%	1e–160
338287	22 (0.24)	1125 bp	60S Ribosomal protein L3 (ARP1) (At1g43170)	88%	0.0
			Lycopersicon esculentum	90%	0.0
337531	20 (0.22)	676 bp	Chlorophyll a/b binding protein (At1g29930.1)	88%	4e–98
			Lemna gibba, *Mesembryanthemum crystallinum*, et al.	92%	1e–104
337479	20 (0.22)	1565 bp	Heat shock protein hsp70 (At3g12580)	94%	0.0
			Cucurbita maxima, *Lycopersicon esculentum*, et al.	97%	0.0
337913	20 (0.22)	1198 bp	Nucleoid DNA-binding-protein; pepsin A (At5g07030)	65%	2e–110
			Oryza sativa	58%	2e–85
337229	18 (0.20)	750 bp	60S Ribosomal protein L5 (RPL5B/ATL5) (At5g39740)	85%	7e–87
			Cucumis sativus, *Oryza sativa*, et al.	85%	6e–85
337838	18 (0.20)	1313 bp	60S Acidic ribosomal protein P0 (At2g40010)	88%	3e–130
			Glycine max, *Trifolium pratense*, *Euphorbia esula*	89%	7e–132
337631	18 (0.20)	1070 bp	Major intrinsic protein family (At2g36830)	83%	5e–101
			Aquaporin in *Vitis vinifera* and *Ricinus communis*	84%	1e–106
337619	17 (0.19)	639 bp	ADP/ATP carrier 1, mitochondrial (AAC1) (At3g08580)	70%	3e–74
			Gossypium hirsutum, *Solanum tuberosum*, et al.	74%	8e–77
338007	17 (0.19)	610 bp	No match to Arabidopsis	–	–
			No matches to GenBank (Blastn or Blastx)	–	–
337632	17 (0.19)	763 bp	S-adenosylmethionine decarboxylase (AT3g02470)	76%	5e–27
			Citrofortunella mitis x, *Malus x*, *Nicotiana tabacum*	85%	7e–34
337123	16 (0.18)	1288 bp	Tubulin beta-2/beta-3, GTP binding/GTPase (AT5g62690)	98%	0.0
			Lupinus albus, *Gossypium hirsutum*, *Oryza sativa*, et al.	97%	0.0
338237	16 (0.18)	1550 bp	Putative cytochrome P450 (At2g45580/At2g45560)	29%	1e–44
			Eschscholzia californica, *Coptis japonica*	79%	0.0

[a]Based on Eca01 unigenes obtained in build 5 (http://pgn.cornell.edu/cgi-bin/unigene/unigene_info.pl?build_id=75)

poppy floral library were comprised of 182 (2.0%) and 40 (0.44%) overlapping ESTs, respectively. These two unigenes are closely related members of the elongation factor 1-alpha gene family (EF-1-α), which is known to be involved in floral development. Both EF-1-alpha unigenes had extremely high levels (95–97%) of amino acid sequence identity to coding sequences of known plant EF-1-alpha proteins in the public databases. However, the 3′-untranslated sequences were quite distinct between the two poppy EF-1-alpha unigenes, confirming that two different elongation factor genes are expressed in developing flowers.

None of the highly expressed unigenes shown in Table 1 were from gene families known to be flower-specific. However, they did have very high levels of sequence similarity to *Arabidopsis* proteins with functions not unexpected in developing and differentiating plant tissues such as S-adenosylmethionine decarboxylase, S-adenosylmethionine synthase, molecular chaperone hsp70, chlorophyll a/b binding protein, 4-coumarate:CoA ligase, glyceraldehyde 3-phosphate dehydrogenase, histidine kinase, proteins targeted to the mitochondria, 60S ribosomal proteins, tubulin, P450s, and a major intrinsic (membrane

Table 2 Cumulative codon usage in Eca01

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	24262	(1.08)	Ser	UCU	28742	(1.74)
	UUC	20809	(0.92)		UCC	12967	(0.78)
Leu	UUA	15321	(0.89)		UCA	22431	(1.36)
	UUG	20789	(1.21)		UCG	8334	(0.50)
Tyr	UAU	16452	(1.14)	Cys	UGU	9293	(1.14)
	UAC	12510	(0.86)		UGC	7081	(0.86)
TERM′N	UAA	1448	(0.00)	TERM′N	UGA	1433	(0.00)
TERM′N	UAG	803	(0.00)	Trp	UGG	14444	(1.00)
Leu	CUU	27864	(1.63)	Pro	CCU	24252	(1.43)
	CUC	15147	(0.88)		CCC	13075	(0.77)
	CUA	13850	(0.81)		CCA	21680	(1.28)
	CUG	9834	(0.57)		CCG	8787	(0.52)
His	CAU	15248	(1.28)	Arg	CGU	9263	(0.86)
	CAC	8552	(0.72)		CGC	5821	(0.54)
Gln	CAA	22818	(1.14)		CGA	9177	(0.85)
	CAG	17220	(0.86)		CGG	5436	(0.50)
Ile	AUU	29047	(1.45)	Thr	ACU	22992	(1.58)
	AUC	17791	(0.89)		ACC	13378	(0.92)
	AUA	13437	(0.67)		ACA	16887	(1.16)
Met	AUG	24552	(1.00)		ACG	5126	(0.35)
Asn	AAU	28814	(1.19)	Ser	AGU	15949	(0.97)
	AAC	19650	(0.81)		AGC	10692	(0.65)
Lys	AAA	37860	(1.05)	Arg	AGA	19903	(1.84)
	AAG	34500	(0.95)		AGG	15398	(1.42)
Val	GUU	30783	(1.75)	Ala	GCU	36253	(1.72)
	GUC	11655	(0.66)		GCC	14252	(0.68)
	GUA	11929	(0.68)		GCA	23267	(1.10)
	GUG	16001	(0.91)		GCG	10509	(0.50)
Asp	GAU	43543	(1.47)	Gly	GGU	28257	(1.21)
	GAC	15833	(0.53)		GGC	12617	(0.54)
Glu	GAA	45304	(1.19)		GGA	31612	(1.36)
	GAG	30944	(0.81)		GGG	20761	(0.89)

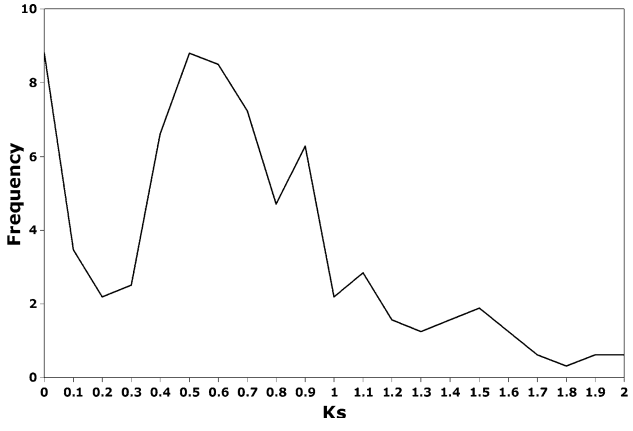

Fig. 1 Plot of synonymous divergence (K_s) between 269 paralogous gene pairs identified in the Eca01 unigene set

channel) protein. Two of the highly expressed unigenes (#338311 and #338007) were entirely novel, with neither homology to the *Arabidopsis* proteome nor to any other DNA and protein sequences in GenBank.

GO Classifications

A wide variety of genes were observed in the *E. californica* floral bud library. Classification of putative gene functions was based on homology of translated poppy ESTs with the TAIR (*Arabidopsis*) and GenBank protein databases and GO annotations were determined as described above. The GO classifications were summarized in pie chart format from three perspectives—Putative Cellular Components, Putative Biological Processes, and Putative Molecular Functions (Figs. 2–4, respectively).

Flower-related genes expressed in *E. californica* floral buds

Table 3 provides a list of unigenes from the Eca01 database that are homologous to families of genes known to be involved in flower development from *Arabidopsis* and rice. Overall, 51 known floral gene families were detected in the *E. californica* transcriptome, accounting for at least 5% of the ESTs and 345

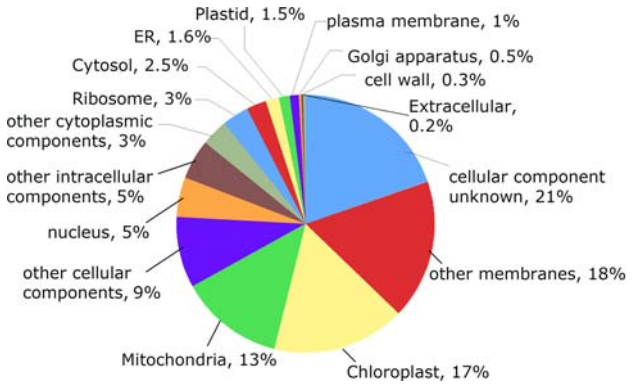

Fig. 2 Pie chart representation of GO-annotation classification of *E. californica* ESTs by putative cellular components

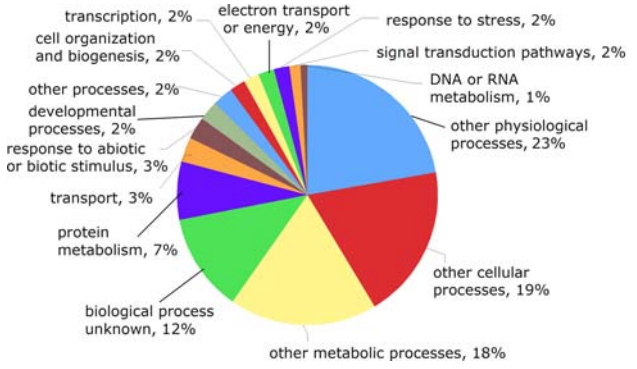

Fig. 3 Pie chart representation of GO-annotation classification of *E. californica* ESTs by putative biological processes

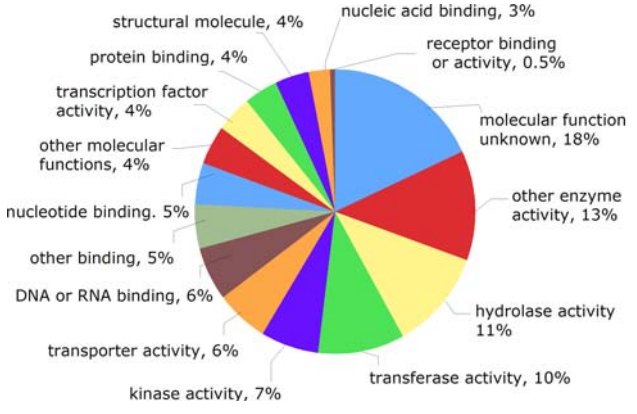

Fig. 4 Pie chart representation of GO-annotation classification of *E. californica* ESTs by putative molecular functions

of the unigenes obtained from the Eca01 floral bud library.

Floral genes (orthologs of genes known to be expressed in *Arabidopsis* flowers) were observed at levels of 0.01 to 1.23% of the sampled transcriptome. The relatively highly expressed floral genes (greater than 0.1% of the ESTs) in California poppy included those encoding CLV1-like receptor kinases, DEAD-Box RNA helicases, Tousled-like kinase, Shaggy-like kinase, ARGONAUTE, ketoacyl-CoA synthase, Auxin response factors, Chalcone synthase, GL1-like MYB, MADS-box-like proteins, NAM/NAC-domain transcription factors, the shoot apical meristem identity protein SPLAYED, HUA1, HEN4 (HUA Enhancer 4), and a Squamosa-Promoter-Binding-like protein. However, most genes were observed in relatively low abundance (<0.1% of sequenced cDNA inserts). Apparently, most floral genes are able to impart their effect on bud development with relatively low levels of expression, which is typical of regulatory factors.

Among the California poppy ESTs, nearly 300 were found to match *Arabidopsis* genes encoding transcription factors and other transcriptional regulators (Supplemental Table S1). Among these, several genes are MADS-box genes, including homologs of *AG*, *AP3*, *PI*, *AGL2*, *AGL6*, and *AGL9* (see below). In addition, homologs of *KAN2*, *SEUSS*, *SPL5* (*SQUAMOSA PROMOTER-BINDING PROTEIN LIKE 5*), *SPL9*, and the HD-ZIPIII leaf polarity gene *PHABULOSA* were also identified. These results indicate that EST analysis was successful in uncovering homologs of many of the known floral regulatory genes; furthermore, the isolation of these California poppy genes supports the hypothesis that the floral regulatory machinery is largely conserved between poppy and core eudicots, such as *Arabidopsis* and *Antirrhinum*. Other transcription factors that were identified as conserved between California poppy and *Arabidopsis* including members of many families with conserved DNA-binding domains, including AP2/ERF, ARF, AT-hook, B3, bHLH, bZIP, forkhead-domain, G-box binding factor, GRAS, HD-ZIP, HMG, LOB, Myb, NAM/NAC, PHD, TCP, WRKY, and zinc-finger. In addition, other genes conserved between poppy and *Arabidopsis* encode proteins that regulate gene expression, such as SET domain proteins (histone methylases), histone acetylase and deacetylases. The conservation of these genes between poppy and *Arabidopsis* suggests that they play important roles during flower development. This hypothesis is further supported by the observation that over 60 of the conserved *Arabidopsis* genes are preferentially expressed in the young inflorescence when compared with leaves, and 90 additional genes have higher levels of expression in the inflorescence than in leaves (Supplemental Table S1; Zhang et al. 2005).

In addition, more than 600 poppy ESTs encode homologs of *Arabidopsis* proteins that are predicted to have regulatory functions such as signal transduction and protein–protein interactions, including G proteins, receptor-like protein kinases, cytosolic

Table 3 Floral gene family sequences identified in the Eca01 database

Gene family name	Number of genes[a] in			Number in Eca01[b]		% of Eca01 "transcriptome"
	ATH	OSA	PTR	Unigenes	ESTs	
ABI1-like phosphatase	36	40	58	2	5	0.06%
alpha FARNESYLTRANSFERASE	1	1	1	1	1	0.01%
AP2	18	25	31	1	2	0.02%
ARGONAUTE	10	24	16	9	20	0.22%
AUX/IAA proteins	25	28	32	1	1	0.01%
Auxin response factors (ARFs)	23	28	44	11	13	0.14%
bHLH/MYC-type	4	3	6	2	2	0.02%
CCA-like MYB	1	0	0	1	2	0.02%
Chalcone synthase family	4	29	21	1	13	0.14%
CLV1-like receptor kinases	712	1252	1565	98	112	1.23%
CO-like Zinc finger, B-Box Zinc finger	6	4	6	1	1	0.01%
Cullin	7	9	12	2	2	0.02%
DIVARICATA-like mybs	20	19	24	1	1	0.01%
Dof zinc finger	36	30	41	1	1	0.01%
EIN3/EIL-like trans regulator	6	7	6	1	1	0.01%
EMF2-like	2	3	6	2	2	0.02%
ETR1-like	5	7	10	3	3	0.03%
EXPANSIN	30	51	37	3	4	0.04%
FCA-like RNA binding	3	3	4	1	1	0.01%
FKF1-like F-Box	4	3	5	1	1	0.01%
FPA	1	1	2	2	4	0.04%
GL1-like MYB	132	117	210	10	13	0.14%
GRAS—GAI, RGA, and SCARECROW	27	43	73	4	5	0.06%
HEN2	3	3	6	2	2	0.02%
HEN4	8	5	14	6	10	0.11%
Homeobox-leucine zipper	17	13	18	2	4	0.04%
HUA1	12	8	11	2	3	0.03%
KANADI	4	6	8	2	2	0.02%
Ketoacyl-CoA synthase	21	27	36	10	17	0.19%
LEUNIG-WD40s	2	6	6	2	2	0.02%
MADS-box	45	40	72	9	13	0.14%
MSI3-like WD-40 repeat	2	1	4	1	2	0.02%
NAM/NAC	83	85	133	9	11	0.12%
Other AP2-domain	54	51	67	5	6	0.07%
Other DEAD-Box RNA helicase	54	54	78	34	64	0.70%
Peranthia-like B-Zip	10	18	14	2	3	0.03%
Phabulosa-like	5	10	10	1	1	0.01%
Phytochrome	5	3	3	1	1	0.01%
PINOID	22	26	25	6	7	0.08%
Shaggy-like kinase	87	87	103	21	30	0.33%
Short-chain ADH	12	22	33	1	1	0.01%
SPL	16	16	31	7	9	0.10%
SPLAYED	17	20	36	8	11	0.12%
SEUSS	4	3	5	2	3	0.03%
Tousled-like kinase	167	154	207	38	46	0.51%
TSO1-like Transcription Factor	4	5	5	1	1	0.01%
WRKY TFs	52	54	88	3	3	0.03%
YABBY	5	7	13	1	1	0.01%
ZF-HD family	15	11	21	5	5	0.06%
ZF-HD family	15	11	21	5	5	0.06%
Totals	**1854**	**2473**	**3278**	**345**	**473**	5.21%

[a]ATH, OSA and PTR refer to the unigenes from the predicted proteomes from the whole genome sequences of *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa*, respectively

[b]Gene family assignments based on Tribes family analysis (http://www.floralgenome.org/cgi-bin/tribedb/tribe.cgi) and Eca01 unigenes from build number 5 (http://pgn.cornell.edu/cgi-bin/unigene/unigene_info.pl?build_id=75)

protein kinases, such as a homolog of the protein kinase PINOID, and phosphatase, calmodulins and related proteins, COP signalosome subunits, heat shock proteins, PPR-repeat proteins, SNF proteins, TPR repeat proteins, WD-repeat proteins, and 14-3-3 proteins (Supplemental Table S2). Although these *Arabidopsis* proteins have not been shown to regulate flower development, their similarity to the poppy ESTs strongly suggests a role in supporting normal floral formation. Moreover, ~130 of these *Arabidopsis* putative signaling/regulatory genes that share similarity with poppy sequences show preferential expression in the inflorescence over leaves (Supplemental Table S2; Zhang et al. 2005). In addition, a large number of ESTs encode proteins that are involved in ubiquitination and protein degradation (data not shown), suggesting that the control of protein turnover is important for flower development, as supported by the function of SCF complex in *Arabidopsis* flower development. Therefore, the combination of sequence comparison and expression analysis is an effective tool to uncover putative novel regulatory genes that potentially play important roles in flower development.

The California poppy ESTs also identified over 600 homologous *Arabidopsis* genes that are annotated only as ''expressed proteins'' (Supplemental Table S3). The sequence similarity between the *Arabidopsis* and California poppy genes supports the hypothesis that these ''expressed'' proteins perform functions conserved in eudicots. An examination of the available expression data in young inflorescence and leaves indicate that 126 of these Arabidopsis ''expressed proteins'' correspond to genes that have a 2 folder or greater preferential expression in the inflorescence over the leaf, with nearly 200 additional genes showing higher expression in the inflorescence than leaves (Supplemental Table S3; Zhang et al. 2005). Furthermore, 19 *Arabidopsis* genes that are annotated as ''hypothetical'' are highly similar to at least one poppy EST, suggesting that they are in fact real genes, and demonstrating the potential for a basal eudicot EST database such as Eca01 to improve the annotation of the *Arabidopsis* model genome. Finally, 1326 of the poppy ESTs were found to not match any known or predicted *Arabidopsis* gene. Among these poppy ESTs, 90 had significant BLASTx hits against GenBank, however, suggesting that they may encode functions lost from or widely diverged from *Arabidopsis*, but present in other plant species.

However, some interesting flower development genes still remain to be identified. Comparing *Eschscholzia* gene family members with *Arabidopsis*, notable genes that were not among the sequenced ESTs include a *FLORICAULA/LEAFY* homolog, members of the BEL1-like homeodomain protein subfamily, floral polarity genes from the YABBY family like *CRABS CLAW* or *INNER NO OUTER*, or bHLH genes like *SPATULA* and *INDEHISCENT*, and several florally expressed MADS-box genes. Possible reasons for the lack of homologs of these genes in the poppy EST dataset might be that they are transcribed in stages other than the ones used for library construction or their expression level might be extremely low and sequencing more clones would lead to their identification. Alternative approaches to EST sequencing, including PCR and screening of the cDNA library and a genomic BAC library with heterologous probes, are underway to identify remaining genes of interest.

Alkaloid pathway genes

The alkaloid pathway has been studied intensively in poppies due to the importance of opium that *Papaver somniferum* produces. Transcripts for all 13 genes for alkaloid biosynthesis that have been cloned and sequenced from *P. somniferum* were also observed in our Eca01 California poppy library including S-adenosyl-L-methionine:coclaurine N-methyltransferase, S-adenosyl-L-methionine:norcoclaurine 6-O-methyltransferase, S-adenosyl-L-methionine:3′-hydroxy-N-methylcoclaurine 4′-O-methyltransferase-1 and -2, (S)-N-methylcoclaurine 3′-hydroxylase (cyp80b1), berberine bridge enzyme (bbe1), NADPH-dependent codeinone reductase (cor1), tyrosine/dopa decarboxylase (genes tydc1 to tydc9), and salutaridinol 7-O-acetyltransferase (salAT). The *E. californica* unigenes with matches to alkaloid biosynthesis gene sequences and their Blast e-value scores, along with the gene family tribes from *Arabidopsis* that they belong to, are shown in Table 4. In total, they account for 59 unigenes from 62 ESTs. In addition, we also observed two other alkaloid pathway genes in our *E. californica* database that had not yet been identified in *P. somniferum*, tyrosinase and salutaridinol synthase.

Phylogenetics and expression of MADS-box genes

As mentioned above, homologs of several floral MADS-box genes were identified from the California poppy ESTs. Phylogenetic analysis (not shown) of these genes place the *Eschscholzia* MADS-box genes as homologs of known floral genes in *Arabidopsis*, including *AGAMOUS*, *AGL6*, *APETALA3*, *PISTILLATA*, *SEEDSTICK* (*AGL11*), *SEPALLATA1/2/4* (*AGL2/4/3*) and *SEPALLATA3* (*AGL9*). Detailed phylogenetic studies that placed these poppy genes within their respective subfamilies have been recently

Table 4 Alkaloid biosynthesis gene expression in *E. californica* flower buds (Eca01 database)

Putative gene assignments[a]	# ESTs	Matches to Ranunculid genes[a]			Putative gene family assignments[b]		
		Blastx score	Species	GenBank Acc'n numbers	Gene family (Arabidopsis Tribes)	# Eca01 Unigenes in tribe	Best blastx score
S-Adenosyl-L-methionine:coclaurine N-methyltransferase	2	4e–59	*P. somniferum*	AY217336/AAP45316	N-Methyltransferases	3	5.60e–59
S-Adenosyl-L-methionine:norcoclaurine 6-O-methyltransferase	4	6e–72	*P. somniferum*	AY217335	O-Methyltransferases	5	2.50e–72
S-Adenosyl-L-methionine:3′-hydroxy-N-methylcoclaurine 4′-O-methyltransferase 2	8	e–163	*P. somniferum*	AY217334/ AAP45314	O-Methyltransferase 2	5	3e–150
S-Adenosyl-L-methionine:3′-hydroxy-N-methylcoclaurine 4′-O-methyltransferase 1	1	e–150	*P. somniferum*	AY217333	O-Methyltransferase 1	5	1e–162
(S)-N-Methylcoclaurine 3′-hydroxylase	12	e–128 4e–76	*E. californica* *P. somniferum*	AF014801/AAC39453	Flavonoid 3′-hydroxylase	16	4.10e–76
Berberine bridge enzyme (bbe1)	1	1e–19 5e–19	*P. somniferum* *E. californica*	AF005655/AAC39358	Protein enzyme fad binding domain	1	4e–24
NADPH-dependent codeinone reductase (cor1)	5	3e–75	*P. somniferum*	AF108438/AAF13742	Aldo/keto oxidoreductase family	2	1e–117
Tyrosine/dopa decarboxylase (and TYDC clones)	3	9e–82	*P. somniferum*	PSU16804/AAA97535	Tyrosine decarboxylase	2	3e–45
Salutaridinol 7-O-acetyltransferase (salAT)	2	3e–17	*P. somniferum*	AF339913/AAK73661	O-acetyltransferase	1	–
Arginine decarboxylase	3	7e–27	*P. somniferum*	AF161835	Arginine decarboxylase	2	2e–103
Tyrosinase	5	None	*None*	AAO16865	Tyrosinase	4	6e–57
(*R*,*S*)-reticuline 7-O-methyltransferase	8	9e–66	*P. somniferum*	AF161835	O-Methyltransferase family 2 proteins	3	–
NADPH:ferrihemoprotein oxidoreductase	2	8e–77 3e–61	*P. somniferum* *E. californica*	PSU67185 ECU67186	NADPH-cytochrome p450 reductase	2	6e–35
(S)-Scoulerine 9-O-methyltransferase	3	1e–50	*P. somniferum*	AAQ01670	O-Methyltransferase family 2 proteins	5	5e–48
Salutaridine synthase, (*R*)-reticuline oxidase	3	None	None	AB036735	NADP-dependent Oxidoreductase	3	5e–60
Totals	**62**					**59**	

[a]Putative gene assignments based on tBlastx alignments

[b]Gene family assignments based on Tribes family analysis (http://www.floralgenome.org/cgi-bin/tribedb/tribe.cgi) and unigenes from build number 5 (http://pgn.cornell.edu/cgi-bin/unigene/unigene_info.pl?build_id=75)

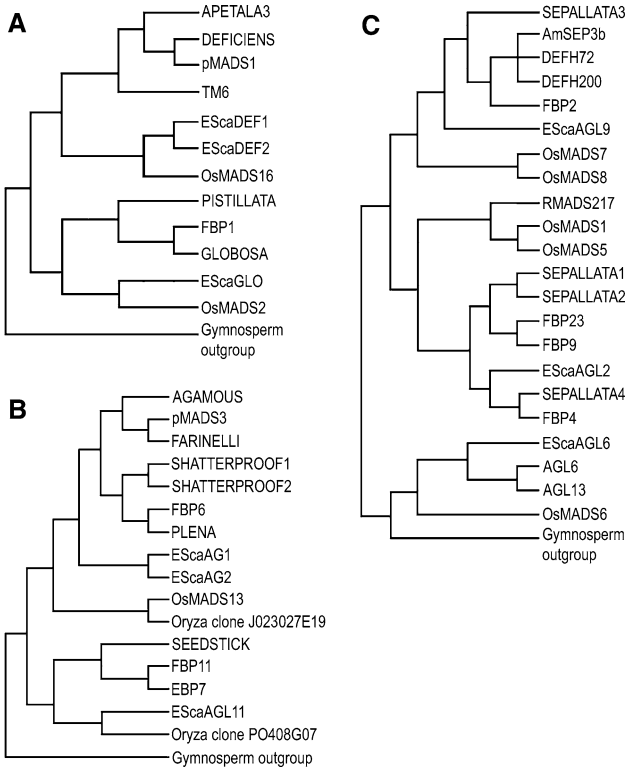

Fig. 5 **A**. Representative placement of *Antirrhinum*, *Arabidopsis*, *Eschscholzia* and *Oryza DEFICIENS/GLOBOSA* genes as placed in a larger phylogenetic context (for more details see Zahn et al. 2005a); **B**. Representative placement of *Antirrhinum*, *Arabidopsis*, *Eschscholzia* and *Oryza AGAMOUS* genes as placed in a larger phylogenetic context (for more details see Zahn et al. 2006); **C**. Representative placement of *Antirrhinum*, *Arabidopsis*, *Eschscholzia* and *Oryza SEPALLATA* genes as placed in a larger phylogenetic context (for more details see Zahn et al. 2005b)

reported (Zahn et al. 2005a, b; 2006). An illustration of the relationship between the California poppy genes and selected members of the same subfamilies is provided in Fig. 5. Two members of the *AG* subfamily, *EScaAG1* and *EScaAG2*, were found in the EST dataset and are recent duplicates that form a sister clade to other eudicot genes (Zahn et al. 2006). The third gene, *EScaAGL11*, was cloned by 3′ RACE and occupies a position basal to other eudicot members of the *SEEDSTICK* clade within the *AG* (Zahn et al. 2006). Similarly, *EScaDEF* and *EScaGLO* were placed basal to their eudicot orthologs, respectively, within the expected clades of the *DEF*/*GLO* subfamily (Zahn et al. 2005b). In addition, the *EScaAGL6* and *EScaAGL9* were basal to their counterparts from derived eudicots, and the *EScaAGL2* gene was placed in the *AGL2/3/4* clade, although its relationship with other genes is not certain (Zahn et al. 2005a).

In situ hybridization studies with sections of developing poppy floral buds reveal that early expression of *EScaGLO*, *EScaAG1* and *EScaAGL2* (Fig. 6) were similar to the respective expression patterns of their closest *Arabidopsis* homolog. The expression of *EScaGLO*, *EScaAG1* and *EScaAGL2* were detected in the floral meristem at stage 2 (as defined in Becker et al. 2005), which is comparable to stage 3 in *Arabidopsis* flower development (Smyth et al. 1990), when the *Arabidopsis* homologs are expressed. *EScaGLO* is expressed in the region of the floral meristem where petals and stamens will arise in the following developmental stages. At stage 2, *EScaAG1* is expressed in the entire floral meristem except for the sepals. These patterns are similar to those of their respective *Arabidopsis* homologs, the *PISTILLATA* and *AGAMOUS* genes (Goto and Meyerowitz 1994; Yanofsky et al. 1990). Early in stage 2 the expression of *EScaAGL2* was found in the floral meristem excluding the area where sepal primordial are about to arise, more like *SEPALLATA3* (*AGL9*) and *SEPALLATA4* (*AGL3*) in early expression and unlike the *SEPALLATA1* and *2* (*AGL2* and *AGL4*) expression in the entire floral meristem (Flanagan and Ma 1994; Savidge et al. 1995; Huang et al. 1995; Mandel et al. 1998; Ditta et al. 2004).

Discussion

Successful gene discovery in the *Eschscholzia* library

Models for flower development proposed from mutation analyses are very informative and have contributed greatly to our understanding of this important process. However, much remains to be discovered regarding the central questions of how the floral developmental program originated and diversified, and how generally applicable the information from the model systems are to floral development in other species. The approach described in this paper of generating thousands of ESTs from an early (premeiotic/meiotic) floral bud library has permitted us to identify homologs of known floral regulatory genes from model plants and to uncover potentially new floral genes and gene families in California poppy.

The value of California poppy in deciphering floral development lies in the fact that *Eschscholzia californica* is in a family of early-diverging eudicots (Papaveraceae) and its floral structure of two fused sepals, two whorls of two petals each, numerous whorls of a fixed number of stamens, and two fused carpels is appropriate for broad taxonomic comparisons of floral structure diversification.

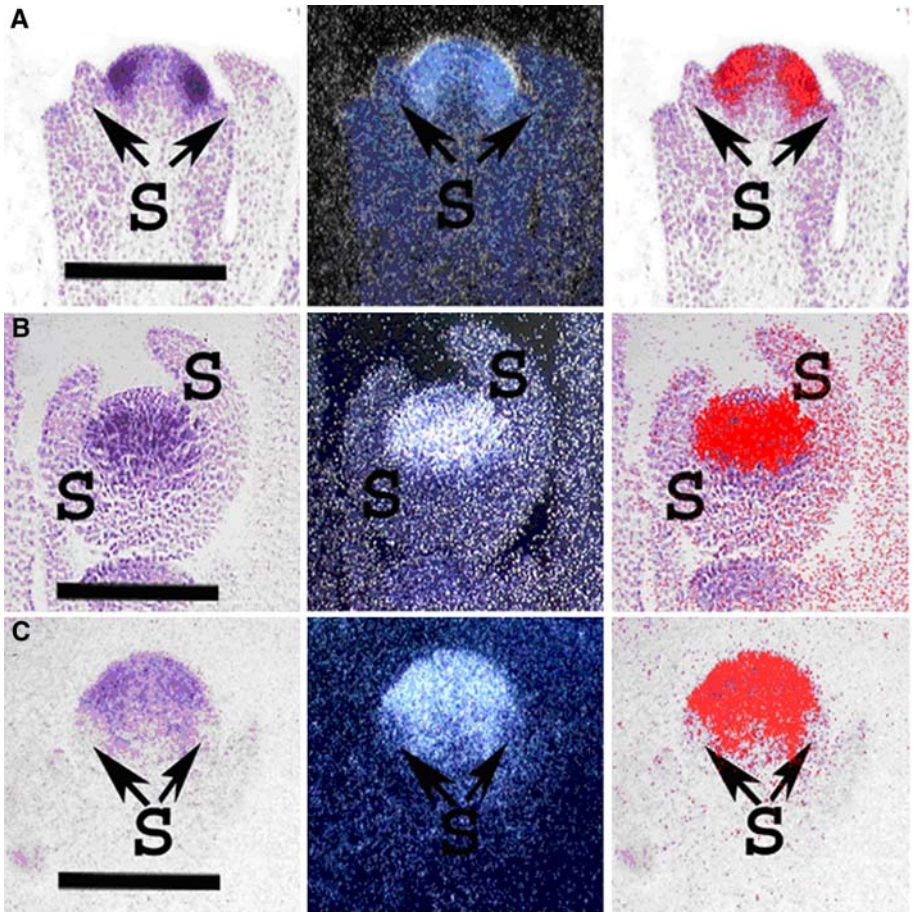

Fig. 6 Expression of *EScaGLO* (**A**), *EScaAG1* (**B**), and *EScaAGL2* (**C**) during early flower development. The expression of each gene is shown at the stage approximately corresponding to stage 3 in *Arabidopsis* at which point the sepal primordial have initiated from the floral meristem. Three images are presented of each section, a bright field image, a dark field image and a bright field image with the signal detected in the dark field image superimposed in red upon the bright field image. Sepal primordia are indicated by S or arrows. The scale bar in the bright field images denotes 0.5 mm. Abbreviations: C = carpel primordia, O = ovule primordia, P = petal primordia, S = sepal primordia, St = stamen primordia

Most of the genes detected in this study are the first representatives of their respective gene family for the Papaveraceae. Prior to this project, there were only 12 nucleotide sequence entries in GenBank for *E. californica*, which included phantastica-like *MYB* protein (*Phan*) mRNA, ATP synthase beta subunit (*atpB*) gene chloroplast gene, *RuBisCO* large subunit (*rbcL*) chloroplast gene, histone H4 mRNA, (S)-N-methylcoclaurine 3′-hydroxylase mRNA, the berberine bridge enzyme (*bbe1*) gene, NADPH:ferrihemoprotein oxidoreductase mRNA, (S)-reticuline:oxygen oxidoreductase mRNA, floricaula-like protein (*FLO*) and *SHOOTMERISTEMLESS* (*STM*) mRNA including in situ hybridization data (Busch and Gleissberg 2003; Groot et al. 2005), and a small subunit ribosomal protein (*rps11*) gene. Furthermore, prior to this project there were only 270 nucleotide sequence entries in GenBank for all species in the Papaveraceae, of which only 15 were from flowers. Thus, this EST study increased the total number of entries in GenBank for the Papaveraceae over 3,000% and increased flower-derived sequence entries for the Papaveraceae over 60-fold. Previous GenBank *Papaver* and *Chelidonium* ''floral'' sequence entries included transcripts for the MADS-box *FRUITFUL*-like gene (*PapsFL1*, *PapsFL2*, *PapnFL1*, *PapnFL2*, *CmFL1*, *CmFL2*), a Myb-related domain (pmr), *SEPALLATA3*-like genes (*PapnSEP3*), *APETALA3* homologs *(PnAP3-1*, *PnAP3-2*, *PcAP3*), two *PISTILLATA* homologs (*PnPI-1* and *PnPI-2*), and non-floral-specific gene families encoding proteins associated with cell wall biosynthesis such as pectinacetylesterase; pectin methyltransferase; pectin methylesterase; pectate lyase; polygalacturonase; xyloglucan endotransglycosylase; cellulase; beta-1,3-glucanase, and the highly expressed lignin pathway gene 4-coumarate:CoA ligase, and one homolog of the inflorescence and floral meristem maintenance gene *STM* (*CmSTM*). It is surprising that there were so few floral Papaver sequences in the public databases prior to this study given the importance of flower development in alkaloid production in poppies. However, this reveals the power of the EST approach to gene discovery relative to previous single gene (forward genetic) approaches.

Overall, known floral gene families accounted for over 5% of all of the ESTs obtained from the *E. californica* floral bud library (Table 3). The ability to develop ''electronic Northern'' results from such

informatic analyses of the sequence data demonstrates the value of deep sequencing in non-normalized and non-subtracted libraries. Expression of floral-specific genes at 5% of the transcriptome is within previously observed ranges for early stages of reproductive bud development. The database for the Massively Parallel Signature Sequencing (MPSS) project for gene expression in *Arabidopsis thaliana* (http://mpss.udel.edu/at/?) shows transcripts in developing floral buds for individual floral genes such as *YABBY*, *AGAMOUS*, *SEPALLATA*, *APETALA*, etc., at levels of 0–400 transcripts per million sequence tags. Although the MPSS data is certainly an underestimate, it does suggest that the sum of 473 ESTs from a total of 9,079 sequence reads (or 100–12,300 transcripts per million tags for the 51 floral gene families detected) in *E. californica* floral buds (Table 3), is within expectation. This result again documents the value of the EST approach with non-normalized, non-selected libraries for discovering genes. This also demonstrates how important expression is of members of these floral gene families at pre-meiotic stages of flower bud development. In addition to homologs of known floral regulators, the poppy ESTs also uncovered numerous putative regulatory genes that are conserved in the *Arabidopsis* genome and expressed in the *Arabidopsis* flower. Not all known floral genes were detected among the 5713 poppy unigenes, however, which is not unexpected with an EST approach. Those floral genes for which we are still searching in poppy include the *FLORICAULA/LEAFY*, *BEL*1-like, *YABBY*, and *bHLH* gene families. Perhaps these genes were expressed at such early stages of floral bud development that they were by default underrepresented in the Eca01 cDNA library because of the small amount of RNA contributed by the earliest stages of bud development relative to the total by weight of all buds collected.

A number of poppy ESTs were observed that match *Arabidopsis* genes that are annotated to encode ''expressed proteins'' or ''hypothetical proteins''. These conserved poppy EST sequences provide support that the previously hypothetical genes in *Arabidopsis* do indeed encode proteins. These results demonstrate that the poppy ESTs will be of value for understanding genes in *Arabidopsis* as well as in poppy. Furthermore, the fact that *Arabidopsis* and poppy have genes with highly similar sequences are suggest that these genes have been conserved during the evolutionary history of the eudicots, even though the *Arabidopsis* homologs of the poppy ESTs range from genes with well understood functions from genetic studies to genes that are predicted from the genomic sequences without any experimental support.

The wide array of transcripts in the floral bud cDNA library reveals the diversity of cellular functions that are necessary for organ initiation and development, from protein translation and transport machinery to cell wall biosynthesis and intermediary metabolism. The summary of GO classification results (Figs. 2–4) using different approaches to categorizing functions shows just how extensive are the types of gene families present in our data set. Thus, even though this project targeted genes involved in floral bud developmental, our results show that the approach of deep EST sequencing of a non-normalized, non-selected library will also reveal many genes important to the elucidation of other cellular and developmental processes in basal plants.

We investigated the suitability of *E. californica* and of our Eca01 database for discovering genes for alkaloid biosynthesis. We found that BLAST alignment searches of the Eca01 database to *P. somniferum* alkaloid gene sequences produced strong matches among our *E. californica* unigenes. Sequences for all previously characterized *P. somniferum* alkaloid genes were identified in the *E. californica* unigene set, as well as two alkaloid biosynthesis genes not yet characterized from *P. somniferum*. From 1 to 10 *E. californica* transcripts were detected per alkaloid gene. Other genes in the alkaloid biosynthesis pathway remain to be discovered, however, even though the stages of flower buds used to create the Eca01 library were at a much earlier stage than peak alkaloid production occurs in *P. somniferum* seed pod development, our results show that the library may be a good source for additional alkaloid genes with deeper sequencing. This result also indicates that the California poppy may be a good alternate model system to opium poppy for studying alkaloid biosynthesis, to avoid the limitations of working with a legally controlled plant.

The two most highly expressed genes in the poppy floral library, comprising a total of 2.44% of the good sequence reads, are members of the *EF-1-α* (translation elongation factor 1-alpha) gene family. High levels of protein translation must be important in most developmental processes, especially in early stages of tissue differentiation such as floral bud initiation and growth. *EF-1-α* genes have been identified in *Arabidopsis* and other plants and appear to be highly regulated in expression during meristem development (Trémousaygue et al. 1999; 2003), and thus are recognized as an important component of floral development. The fact that two *EF-1-α* genes are being transcribed at very high levels during floral development suggests that some redundancy may occur in *E. californica* to insure that adequate levels of *EF-1-α* transcripts are obtained when required. Alternatively,

expression of the two *EF-1-α* genes may be differentially regulated, but not discernable by electron northern analysis because of the multiple cell types and developmental stages present in the floral buds used to prepare the cDNA library. Microarray and in situ hybridization analyses will be used in future experiments by the Floral Genome Project to determine cell and tissue specific expression patterns of the gene families identified in the *E. californica* EST database such as *EF-1-α*. The sequence and expression data will be compared with data from other species to derive a consensus set of floral regulators.

Gene expression—MADS-box genes are similar in expression to their *Arabidopsis* homologs

Floral organ identity in angiosperms seems to be controlled by three conserved genetic functions that act in a combinatorial manner (Coen and Meyerowitz 1991). The ABC model, which describes the role of these functions in floral development, proposes that sepal identity is controlled by A-function, petals by A- and B-functions, stamens by B- and C-functions, and carpels by C-function (Coen and Meyerowitz 1991). Furthermore, E function genes also are required for floral organ identity in all whorls of the flower (Pelaz et al. 2000; Ditta et al. 2004). In *Antirrhinum*, floral organ identity genes include *DEFICIENS* (*DEF*) (Sommer et al. 1990) and *GLOBOSA* (*GLO*) (Tröbner et al. 1992), both required for petal and stamen identity (B-function), and *PLENA* (*PLE*) (Bradley et al. 1993), required for stamen and carpel formation (C-function). Cadastral genes establish the expression boundaries of the organ identity genes (reviewed by Weigel and Meyerowitz 1994; Zhao et al. 2001). Mutations in these genes, e.g. *FIMBRIATA* (*FIM*), have a dual effect by altering the whorl patterning as well as the organ identity boundaries (Simon et al. 1994). The floral organ identity genes of *Antirrhinum DEF*, *GLO* and *PLE*, and the meristem identity gene, *SQUA*, are members of the MADS-box family, coding for transcription factors (Schwarz-Sommer et al. 1990).

In *E. californica* we recovered MADS-box genes representing members of the *AGAMOUS* (*PLENA*), *AGL6*, *DEFICIENS*/*GLOBOSA* and *SEPALLATA* subfamilies. Members of each of these subfamilies have been demonstrated to be required for the specification of floral organ identity in *Arabidopsis* and other angiosperms (reviewed in Becker et al. 2003). As an example of functional studies now possible with the new *E. californica* EST data, we conducted in situ hybridizations with cDNA probes for a *PISTILLATA* homolog, *EScaGLO*, an *AGAMOUS* homolog, *EScaAG1*, and a *SEPALLATA* homolog, *EScaAGL2*. The expression in early bud development shown in Fig. 6 for these three representative MADS-box genes from *E. californica* demonstrates the conservation of expression patterns to those of known for *Arabidopsis* genes at the time when sepal primordia separate from the rest of the floral meristem. Recent reports also indicate that expression of *EScaAGL9* and *EScaGLO* are very similar to those of the *Arabidopsis SEP3* (*AGL9*) and *PI* genes at multiple flower developmental stages (Zahn et al. 2005a, b).

Genome organization in California poppy

The study of K_s values for paralogous gene pairs (Fig. 1) revealed a striking concentration of duplicated genes in the Eca01 unigene set. The K_s values and large number of duplicated genes indicate that *Eschscholzia californica* poppy underwent a relatively recent genome duplication or polyploidization event. The EST data from this study suggests a genome wide duplication event occurring some time between 23 and 53 million years ago, long after the Ranuncules split from other lineages of the eudicots, estimated at 120 MYA (Schneider et al. 2004). The new and unexpected information on genome duplication will greatly inform the use of *E. californica* for evolutionary studies.

Even though there is good evidence for a genome wide duplication event, many transcription factors from *E. californica* seem not to appear in paralogous pairs. This may in part be attributed to an insufficient number of sequenced clones. However, we conducted Southern hybridization on genomic DNA for *EScaAGL11* and two *YABBY* genes (data not shown), and our results clearly indicated that these are single-copy genes lacking a paralogous copy. These missing partners might illustrate that a) only part of the genome underwent a duplication event or b) that duplicates were lost in many developmental genes. Following the latter line of reasoning, it seems that selective pressure on protein evolution persists for a long time following a speciation event but not after gene duplication (Castillo-Davis 2005). Such relaxed selection on a paralogous gene pair could easily result in the loss of one partner.

We chose *Eschscholzia californica* for floral EST studies to provide a root for genomic-scale analyses for more derived eudicot species. The Papaveraceae family is a member of the Ranunculales order, which roots the eudicot tree as sister to all other eudicot lineages. The phylogenetic position of *Eschscholzia californica* as an early branching dicot together with the possibility of genetic manipulation to study gene function enhances the value of California poppy as model system

for molecular studies. At present widely used model plants like the higher eudicot *A. thaliana* and the Poaceae rice are both morphologically highly derived species, which only poorly represent the variation observed in angiosperm flower development. A better understanding of the molecular biology of *E. californica* flower development will also help to bridge the gap between the morphological and developmental differences of model species like *A. thaliana* and rice. The *E. californica* EST collection provides a resource for further research on the molecular basis of flower development and on special features of poppy such as alkaloid biosynthesis.

The PGN—public access database

A relational database, the Plant Genome Network (PGN), was developed for public access to the raw sequence data, the unigene sets, the library statistics and the annotations. The unigene assemblies and individual sequences can be queried and viewed and trace files can be downloaded individually or in bulk. PGN also provides the tools necessary for the storage, retrieval and annotation of plant ESTs and also houses sequence databases for other taxa involved in the Floral Genome Project's study of flower evolution. In addition, PGN was designed to grow into a general plant EST data warehouse to provide a stable web address to EST sequencing projects that are not able to create their own data analysis and web interface infrastructure. Data can be submitted to PGN using an interactive data submission system. PGN can be found on the world wide web at http://pgn.cornell.edu.

Acknowledgements We thank Michael Kosco and Yoshita Oza for assistance with preparations for the in situ hybridization experiments and A. Omeis for plant growth and care. We thank Sheila Plock for assistance with manipulations and curation of the cDNA library, and Marlin Druckenmiller in the Schatz Center for Tree Molecular Genetics at Penn State for assistance in high throughput sequencing. We also thank William Farmerie for his suggestions on sequencing procedures. The Floral Genome Project was supported by a grant to C. dePamphilis and co-PIs from the NSF Plant Genome Research Program (DBI-0115684), A. Becker was supported by a fellowship from the German Research Foundation (DFG, BE 2547/2-1) and the PGN database hosted by Cornell University was supported by NSF grants DBI-9872617 and DBI-0115684.

References

Albert VA, Soltis DE, Carlson JE, Farmerie WG, Wall PK, Ilut DC, Mueller LA, Landherr LL, Hu Y, Buzgo M, Kim S, Yoo M-J, Frohlich MW, Perl-Treves R, Schlarbaum S, Bliss B, Tanksley S, Oppenheimer DG, Soltis PS, Ma H, dePamphilis CW, Leebens-Mack JH (2005) Floral gene resources from basal angiosperms for comparative genomics research. BMC Plant Biol 5:5

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Angenent GC, Franken J, Busscher M, Colombo L, van Tunen AJ (1993) Petal and stamen formation in petunia is regulated by the homeotic gene fpg1. Plant J 4:101–112

Angenent GC, Franken J, Busscher M, Weiss D, van Tunen AJ (1994) Co-suppression of the petunia homeotic gene fbp2 affects the identity of the generative meristem. Plant J 5:33–44

Becker A, Saedler H, Theissen G (2003) Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. Dev Genes Evol 213:567–572

Becker S, Gleissbergy S, Smyth DR (2005) Floral and vegetative morphogenesis in California poppy (*Eschscholzia californica* Cham.). Int J Plant Sci 166:537–555

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Phil Trans Royal Soc London B 274:227–274

Bennett MD, Bhandol P, Leitch IJ (2000) Nuclear DNA amounts in angiosperms and their modern uses—807 new estimates. Annals Bot 86:859–909

Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opion Plant Biol 7:732–736

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Bradley D, Carpenter R, Sommer H, Hartley N, Coen E (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *PLENA* locus of *Antirrhinum*. Cell 72:85–95

Busch A, Gleissberg S (2003) EcFLO, a FLORICAULA-like gene from *Eschscholzia californica* is expressed during organogenesis at the vegetative shoot apex. Planta 217:841–848

Castillo-Davis CI. (2005) The evolution of noncoding DNA: how much junk, how much func? Trends Genet 21:533–536

Clark C (1993) Papaveraceae (poppy family). In: Hickman JC (ed) The jepson manual, University of California Press, Berkeley, CA

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353:31–37

Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R (1990) Floricaula—a homeotic gene required for flower development in *Antirrhinum majus*. Cell 63:1311–1322

Cook SA (1962) Genetic system, variation, and adaptation in *Eschscholzia californica.* Evolution 16:278–299

Decker G, Wanner G, Zenk MH, Lottspeich F (2000) Characterization of proteins in latex of the opium poppy (*Papaver somniferum*) using two-dimensional gel electrophoresis and microsequencing. Electrophoresis 21:3500–3516

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Curr Biol 14:1935–1940

Endress PK (2004) Structure and relationships of basal relictual angiosperms. Aus Syst Bot 17:343–366

Flanagan CA, Ma H (1994) Spatially and temporally regulated expression of the MADs-Box gene *AGL2* in wild-type and mutant *Arabidopsis* flowers. Plant Mol Biol 26:581–595

Frick S, Kramell R, Schmidt J, Fist AJ, Kutchan TM (2005) Comparative qualitative and quantitative determination of alkaloids in narcotic and condiment *Papaver somniferum* cultivars. J Nat Prod 68:666–673

Goldberg RB, Beals TP, Sanders PM (1993) Anther development: basic principles and practical applications. Plant Cell 5:1217–1229

Goto K, Meyerowitz EM (1994) Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes Dev 8:1548–1560

Groot EP, Sinha N, Gleissberg S (2005) Expression patterns of *STM*-like *KNOX* and Histone *H4* genes in shoot development of the dissected-leaved basal eudicot plants *Chelidonium majus* and *Eschscholzia californica* (Papaveraceae). Plant Mol Biol 58:317–331

Huang H, Tudor M, Weiss CA, Hu Y, Ma H (1995) The *Arabidopsis* MADs-Box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol Biol 28:549–567

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 1999:138–148

Kramer EM, Irish VF (1999) Evolution of genetic mechanisms controlling petal development. Nature 399:144–148

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis, Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498.

Kutchan TM (1995) Alkaloid biosynthesis-the basis for metabolic engineering of medicinal plants. Plant Cell 7:1059–1070

Lee J, Pedersen H (2001) Stable genetic transformation of *Eschscholzia californica* expressing synthetic green fluorescent proteins. Biotechnol Prog 17:247–251

Lincoln C, Long J, Meyerowitz E (2002) http//www.its.caltech.edu/~plantlab/protocols/insitu.htm, accessed August 27, 2004

Ma H (1994) The unfolding drama of flower development: recent results from genetic and molecular analyses. Genes Dev 8:745–756

Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Biol 56:393–434

Ma H (2006) A molecular portrait of Arabidopsis meiosis. In: The Arabidopsis Book. Somerville CR, Meyerowitz EM, Dangl J, Stitt M (eds), American Society of Plant Biologists, Rockville, MD, doi/10.1199/tab.0009, http://www.aspb.org/publications/arabidopsis/

Ma H, dePamphilis C (2000) The ABCs of floral evolution. Cell 101:5–8

McInerney JO (1998) GCUA: general codon usage analysis. Bioinformatics 14:372–373

Mandel MA, Yanofsky MF (1998) The *Arabidopsis AGL9* MADS box gene is expressed in young flower primordial. Sex Plant Reprod 11:22–28

Memelink J (2004) Putting the opium in poppy to sleep. Nature Biotech 22:1526–1527

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28:292

Nanda KK, Sharma R (1976) Effects of gibberellic acid and cyclic 39,59-adenosine monophosphate on the flowering of *Eschscholtzia californica* Cham, a qualitative long day plant. Plant Cell Physiol 17:1093–1095

Park S-U, Facchini PJ (2000a) *Agrobacterium rhizogenes*-mediated transformation of opium poppy, *Papaver somniferum* L., and California poppy, *Eschscholzia californica* Cham., root cultures. J Exp Bot 51:1005–1016

Park S-U, Facchini PJ (2000b) *Agrobacterium*-mediated genetic transformation of California poppy, *Eschscholzia californica* Cham., via somatic embryogenesis. Plant Cell Rep 19:1006–1012

Park S-U, Min, Yu M, Facchini PJ (2002) Antisense RNA-mediated suppression of benzophenanthridine alkaloid biosynthesis in transgenic cell cultures of California poppy. Plant Physiol 128:696–706

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature 405:200–203

Sato F (2005) RNAi and functional genomics. Plant Biotech 22:431–442

Sato F, Hashimoto T, Hachiya A, Tamura K, Choi K-B, Morishige T, Fujimoto H, Yamada Y (2001) Metabolic engineering of plant alkaloid biosynthesis. Proc Natl Acad Sci USA 98:367–372

Savidge B, Rounsley SD, Yanofsky MF (1995) Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. Plant Cell 7:721–733

Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallon S, Lupia R (2004) Ferns diversified in the shadow of angiosperms. Nature 428:553–557

Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H (1990) Genetic control of flower development in *Antirrhinum majus*. Science 250:931–936

Simon R, Carpenter R, Doyle S, Coen E (1994) Fimbriata controls flower development by mediating between meristem and organ identity genes. Cell 78:99–107

Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in *Arabidopsis*. Plant Cell 2:755–767

Soltis D, Soltis P, Albert V, Oppenheimer D, Frohlich MW, dePamphilis CW, Ma H, Theissen G (2002) Missing links: the genetic architecture of flower and floral diversification. Trends Plant Sci 7:22–31

Soltis PS, Soltis DE, Kim S, Chanderbali A, Buzgo M (2006) Modifications of the ABC model based on analyses of basal angiosperms. In: Soltis DE, Soltis PS, Leebens-Mack JH (eds) Developmental genetics of the flower. Advances in botanical research series. Elsevier Limited, London. In press

Soltis PS, Soltis DE, Zanis MJ, Kim S (2000) Basal lineages of angiosperms: Relationships and implications for floral evolution. Int J Plant Sci 161:S97–S107

Sommer H, Beltran J-P, Huijser P, Pape H, Lonnig W-E, Saedler H, Schwarz-Sommer Z (1990) *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. EMBO J 9:605–613

Thomas SG, Franklin-Tong VE (2004) Self-incompatibility triggers programmed cell death in *Papaver pollen*. Nature 429:305–309

Trémousaygue D, Garnier L, Bardet C, Dabos P, Hervé C, Lescure B (2003) Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in *Arabidopsis thaliana* cycling cells. Plant J 33:957–966

Tremousaygue D, Manevski A, Bardet C, Lescure N, Lescure B (1999) Plant interstitial telomere motifs participate in the control of gene expression in root meristems. Plant J 20:553–562

Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig W-E, Saedler H, Sommer H, Schwarz-Sommer Z (1992) *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in

the control of *Antirrhinum* floral organogenesis. EMBO J 11:4693–4704

Wakelin AM, Lister CE, Conner AJ (2003) Inheritance and biochemistry of pollen pigmentation in California poppy (*Eschscholzia californica* Cham.). Int J Plant Sci 164:867–875

Weigel D, Meyerowitz EM (1994) The ABCs of floral homeotic genes. Cell 78:203–209

Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell 69:843–859

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. Nature 346:35–43

Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE, dePamphilis CW, Ma H (2005a) The evolution of the *SEPALLATA* subfamily of MADS-box genes: a pre-angiosperm origin with multiple duplications throughout angiosperm history. Genetics 169:2209–2223

Zahn LM, Leebens-Mack J, Arrington JM, Hu Y, Landherr L, dePamphilis CW, Becker A, Theissen G, Ma H (2006) Conservation and divergence in the *AGAMOUS* subfamily of MADS-Box genes: evidence of independent sub- and neofunctionalization events. Evol Dev 8:30–45

Zahn LM, Leebens-Mack J, DePamphilis CW, Ma H, Theissen G (2005b) To B or not to B a flower: The role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. J Hered 96:225–240

Zhang X, Feng B, Zhang Q, Zhang D, Altman N, Ma H (2005) Genome-wide expression profiling and identification of gene activities during early flower development in Arabidopsis. Plant Mol Biol 58:401–419

Zhao D, Yu Q, Chen C, Ma H (2001) Genetic control of reproductive meristems. In: McManus MT, Veit B (eds) Meristematic tissues in plant growth and development. Academic Press, Sheffield, pp 89–142