Missing links: the genetic architecture of flower and floral diversification

Floral Genome Project Research Group (Douglas E. Soltis *et al.*)

To understand the genetic architecture of floral development, including the origin and subsequent diversification of the flower, data are needed not only for a few model organisms but also for gymnosperms, basal angiosperm lineages and early-diverging eudicots. We must link what is known about derived model plants such as *Arabidopsis*, snapdragon and maize with other angiosperms. To this end, we suggest a massive evolutionary genomics effort focused on the identification and expression patterns of floral genes and elucidation of their expression patterns in 'missing-link' taxa differing in the arrangement, number and organization of floral parts.

Douglas E. Soltis*
Dept Botany and Genetics Institute, University of Florida, Gainesville, FL 32611, USA.
*e-mail: dsoltis@botany.ufl.edu

Pamela S. Soltis
Florida Museum of Natural History and Genetics Institute, University of Florida, Gainesville, FL 32611, USA.

Victor A. Albert
Natural History Museums and Botanical Garden, University of Oslo, Sars' gate 1, N-0562, Oslo, Norway.

David G. Oppenheimer
Dept Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA.

Claude W. dePamphilis
Hong Ma
Dept Biology, Pennsylvania State University, University Park, PA 16802, USA.

Michael W. Frohlich
Herbarium, University of Michigan, Ann Arbor, MI 48109-1057, USA.

Günter Theißen
Institute for Genetics, Friedrich Schiller University, D-07743 Jena, Germany.

Variations in floral architecture are evolutionarily and economically important, affecting features such as pollination, predation and seed dispersal. However, in spite of the central role of flowers in plant reproduction, agriculture and horticulture, the processes responsible for the origin and subsequent evolution of the flower remain fundamental problems in plant biology. Major questions include the following.

- How did flowers become bisexual, given that male and female reproductive organs are in separate structures in all extant gymnosperms?
- How did the major organs (sepals, petals, stamens and carpels) originate? What genes control the number, arrangement and fusion of floral organs?
- Which effector genes generate the characteristic features of these major organs?
- Which genes control floral initiation and development throughout the angiosperms, especially in the most basal lineages, and how do they compare with those in model organisms?
- How much of the developmental machinery is common to most lineages and how much is peculiar to restricted groups of angiosperms?

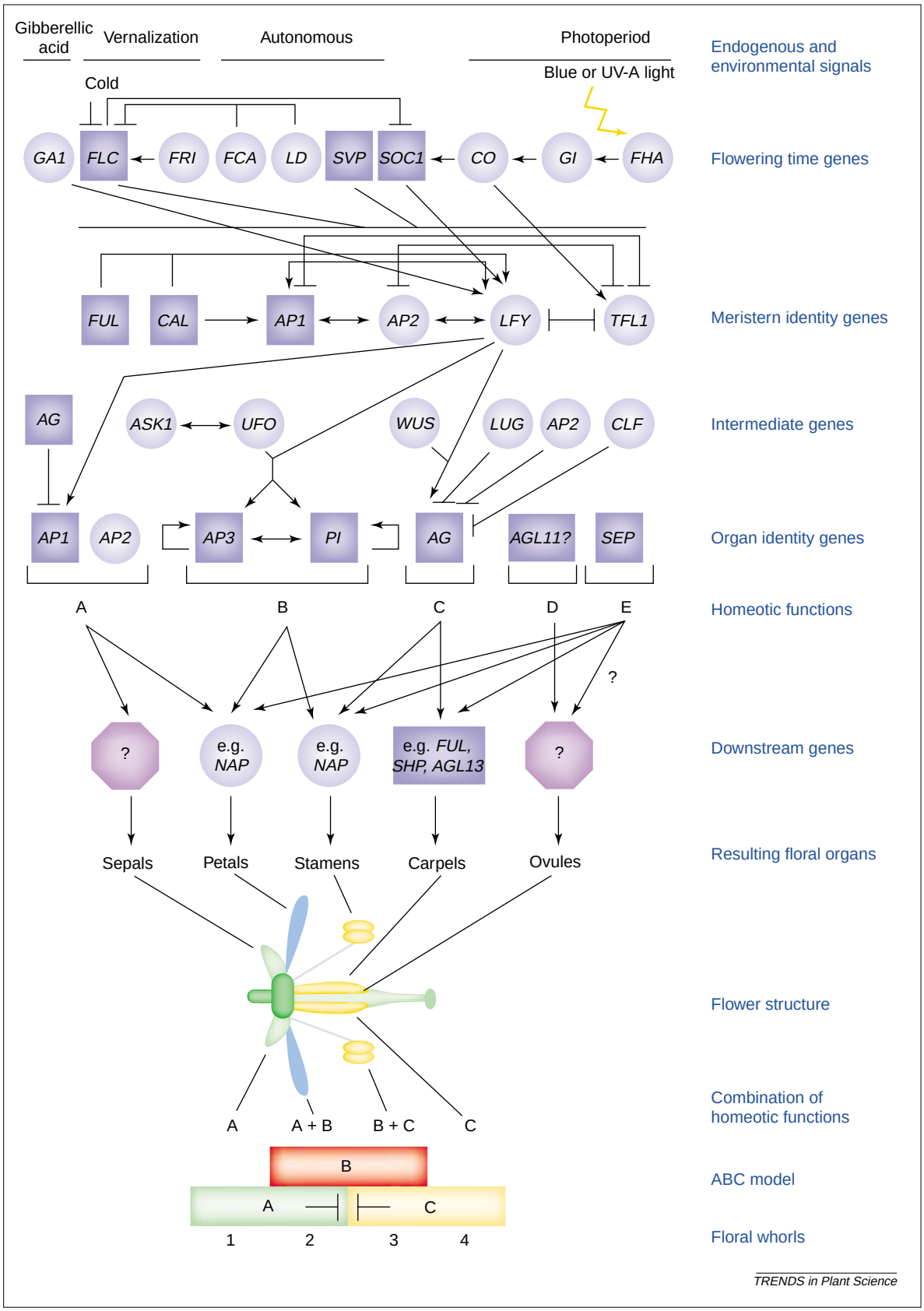
Genetic control of floral initiation and development in model plants

Floral initiation and development in the model organisms *Arabidopsis* and *Antirrhinum* are controlled by many genes from many families that exhibit a diversity of functions [1–10] (Fig. 1), including transcription factors (e.g. *FLO/LEAFY*, *HB*, *MADS* and *YABBY*) [11] and signal transduction genes (*CLAVATA1*, *CLAVATA3* and *KAPP*) [12,13]. Genes encoding bidentate ribonucleases and genes controlling cell division and chromatin structure are also important [14–16]. MADS-box genes play key roles in specifying floral organ identity and are by far the best understood. During flower development in *Arabidopsis* and *Antirrhinum*, the identity of the floral organs is specified by at least three classes of homeotic genes, including the well known A-, B- and C-function genes [1–3]. In addition, in *Arabidopsis*, *SEPALLATA 1*, *SEPALLATA 2* and *SEPALLATA 3* are also involved in the specification of petals, stamens and carpels [9,17–19]. The functional redundancy of *SEPALLATA* genes prevented the earlier discovery of their role in organ specification but it is not yet clear how widely conserved *SEPALLATA* functions are among diverse groups of angiosperms.

The B and C functions of this model also seem to extend to maize (*Zea*) [7,20], although conservation of the A function remains to be shown and the role of B-class genes in petal initiation is difficult to discern. Nevertheless, the strong genetic and molecular evidence for the conservation of at least portions of the ABC model suggests that this is an ancient regulatory network, perhaps applicable to most angiosperms [7]. However, various modifications of the specific components of the ABC model might have occurred in different lineages of angiosperms [7,21]. For example, work on angiosperms as diverse as alfalfa (*Medicago sativa*),

Fig. 1. A simplified, preliminary depiction of the genetic hierarchy that controls flower development in the eudicot model plant *Arabidopsis*. Examples of the different types of genes within each level of the hierarchy are shown. 'Gibberellic acid', 'vernalization', 'autonomous' and 'photoperiod' refer to the different promotion pathways of floral induction. 'Intermediate genes' summarizes a functionally diverse class of genes including 'cadastral genes'. MADS-box genes are shown as squares, non-MADS-box genes as circles and genes whose sequences have not been reported as octagons. Some regulatory interactions between the genes are symbolized by arrows (activation), double arrows (synergistic interaction) or barred lines (inhibition, antagonistic interaction). Not all the known genes and interactions involved in flower development are shown. In the case of the downstream genes, just one symbol is shown for each type of floral organ, although whole cascades of many direct target genes and further downstream genes are probably activated in each organ of the flower. At the bottom of the figure, a generic flower diagram is shown with the classic 'ABC model' of floral organ identity. According to this model, floral organ identity is specified by three classes of 'floral organ identity genes' providing 'homeotic functions' A, B and C, each of which is active in two adjacent whorls. A alone specifies sepals in whorl 1; the combined activities of A and B specify petals in whorl 2; B and C specify stamens in whorl 3; and C alone specifies carpels in whorl 4. The activities A and C are mutually antagonistic, as indicated by barred lines: A prevents the activity of C in whorls 1 and 2, and C prevents the activity of A in whorls 3 and 4. Abbreviations: *AG*, *AGAMOUS*; *AGL*, *AGAMOUS*-like gene; *AP*, *APETALA*; *ASK1*, *ARABIDOPSIS SKP1*-like 1; *CAL*, *CAULIFLOWER*; *CO*, *CONSTANS*; *FLC*, *FLOWERING LOCUS C*; *FRI*, *FRIGIDA*; *FUL*, *FRUITFULL*; *GI*, *GIGANTEA*; *LD*, *LUMINIDEPENDENS*; *LFY*, *LEAFY*; *LUG*, *LEUNIG*; *NAP*, *NAC*-like, activated by *AP3/PI*; *PI*, *PISTILLATA*; *SEP*, *SEPALLATA*; *SHP*, *SHATTERPROOF*; *SOC1*, *SUPPRESSOR OF OVEREXPRESSION OF CO1*; *SVP*, *SHORT VEGETATIVE PHASE*; *UFO*, *UNUSUAL FLORAL ORGANS*; *TFL1*, *TERMINAL FLOWER1* (modified from Ref. [9]).

http://plants.trends.com 1360-1385/02/$ – see front matter PII: S1360-1385(01)02098-2



the sunflower family (Asteraceae) and poppy relatives (Papaveraceae) suggests that the functions of *APETALA3* and *PISTILLATA* as B-class organ identity genes is not rigidly conserved across all flowering plants [22–24]. As in *Zea*, only B- and C-class genes are present in gymnosperms [25,26]. Furthermore, recent studies suggest that the developmental control of floral initiation might

derive more from systems active in the male reproductive structures of gymnosperms than from the female structures, leading to the proposal of the 'mostly male' theory of flower evolutionary origins [27]. If this hypothesis is correct, the genetic skeleton that encoded the first flower can be found among the genes that control development of male structures in gymnosperms.

A complex network of interacting genes upstream of the ABC genes controls floral identity in *Arabidopsis* and *Antirrhinum*. The genes *LEAFY/FLORICAULA*, *APETALA1/SQUAMOSA*, *TERMINAL FLOWER/CENTRORADIALIS* and *AGAMOUS/PLENA* [6–8,10] regulate each other and control the expression of the ABC genes [28–30] (Fig. 1). The origin of this regulatory network and the extent of conserved gene functions and interactions must be addressed to understand how flowers initiate, develop and are modified in diverse plant lineages. In contrast with the growing understanding of ABC and upstream genes, there are few known genes that lie directly downstream of the ABC genes. These downstream effector genes control the specific features of the various floral organs, thus determining the biological functions of the flower. An increased knowledge of these genes is needed to facilitate not only a better understanding of flower development and evolution but also more effective genetic manipulation of plant reproduction through biotechnology.

Angiosperm phylogeny and missing links

Coincident with the increased understanding of the genetic architecture of floral development in model organisms has been the clarification of phylogenetic relationships across the angiosperms [31–33]. The overall framework of angiosperm phylogeny has recently crystallized, representing a dramatic advance in angiosperm systematics. These studies indicate that the model organisms *Arabidopsis*, *Antirrhinum* and maize are highly derived within the rosid, asterid and monocot clades, respectively (Fig. 2).

The evolutionary gaps between gymnosperms, maize and the eudicot model systems are enormous, particularly from the perspective of floral evolution. Although the eudicots represent ~75% of angiosperm species, most of the diversity in the arrangement and number of floral parts actually occurs in the basal angiosperm lineages such as the Nymphaeaceae (water lilies), *Amborella* (the sister to all other extant angiosperms either alone or with Nymphaeaceae [34,35]), Magnoliaceae (magnolias), Lauraceae (avocado) and Piperaceae (black pepper family) (Fig. 3). Although a perianth is typically present in basal angiosperms, clear-cut differentiation into sepals and petals is often lacking (Fig. 3). Floral organization and development are considered 'open' and highly labile in basal angiosperms [36]. By contrast, in most eudicots, the number of floral parts is low (four or five) and fixed, and floral organs are arranged in whorls, suggesting that the basic floral Bauplan became highly canalized during the early diversification of the eudicots [36–38] (Fig. 3). Thus, crucial components of the floral genetic program of derived eudicots might have evolved in the most basal lineages of angiosperms.

Because model organisms represent only a small portion of the phylogenetic tree, we believe that many more genes need to be identified from a more representative set of plants before we can have a thorough understanding of the control and evolution of flower development. To obtain a comprehensive understanding of the genetics of floral development, data are needed not only for a few model organisms but also for gymnosperms, basal angiosperm lineages and early-diverging eudicots. Does the current model of floral organ initiation extend to these 'missing links', which represent a much greater sample of floral diversity? Which genetic functions affecting flower and inflorescence structure first became established among these earliest branches? Which genes might play important but still undetected roles because of their complex interactions with genes that have duplicate or overlapping function? Elucidating the genetics of floral development in these key lineages should not only help to answer questions about the origin and diversification of the flower itself, but should also provide the opportunity to link what is known about derived eudicot model systems with other flowering plants. This should provide a more comprehensive picture of floral development and evolution.

An understanding of the genetic basis of floral development and modification might translate into genetic improvement of those plants whose flowers and/or fruits have economic value. These crops are not concentrated near *Arabidopsis* and *Antirrhinum* but instead are scattered across the angiosperm topology (Fig. 2). In addition to grasses, other monocot crops based on inflorescences, floral organs or fruits include pineapple, banana, vanilla, coconut and date palm. Several lineages of basal angiosperms also contain plants cultivated for their fruits, such as black pepper (*Piper nigrum*), nutmeg (*Myristica fragrans*), avocado (*Persea americana*), cherimoya (*Annona cherimola*) and star anise (*Illicum verum* and *Illicum anisatum*). Diverse eudicots yield economically important flowers or fruits, including members of several early-branching eudicot families [e.g. macadamia nut (*Macademia integrifolia*, *Macademia tetraphylla*) and poppy (*Papaver somniferum*)]. In instances of opportunistic gene flow between crops and related weeds, knowledge of how to limit or regulate sexual reproduction can be equally important byproducts of a broad understanding of genes crucial to flower development. Most core eudicot clades recognized at

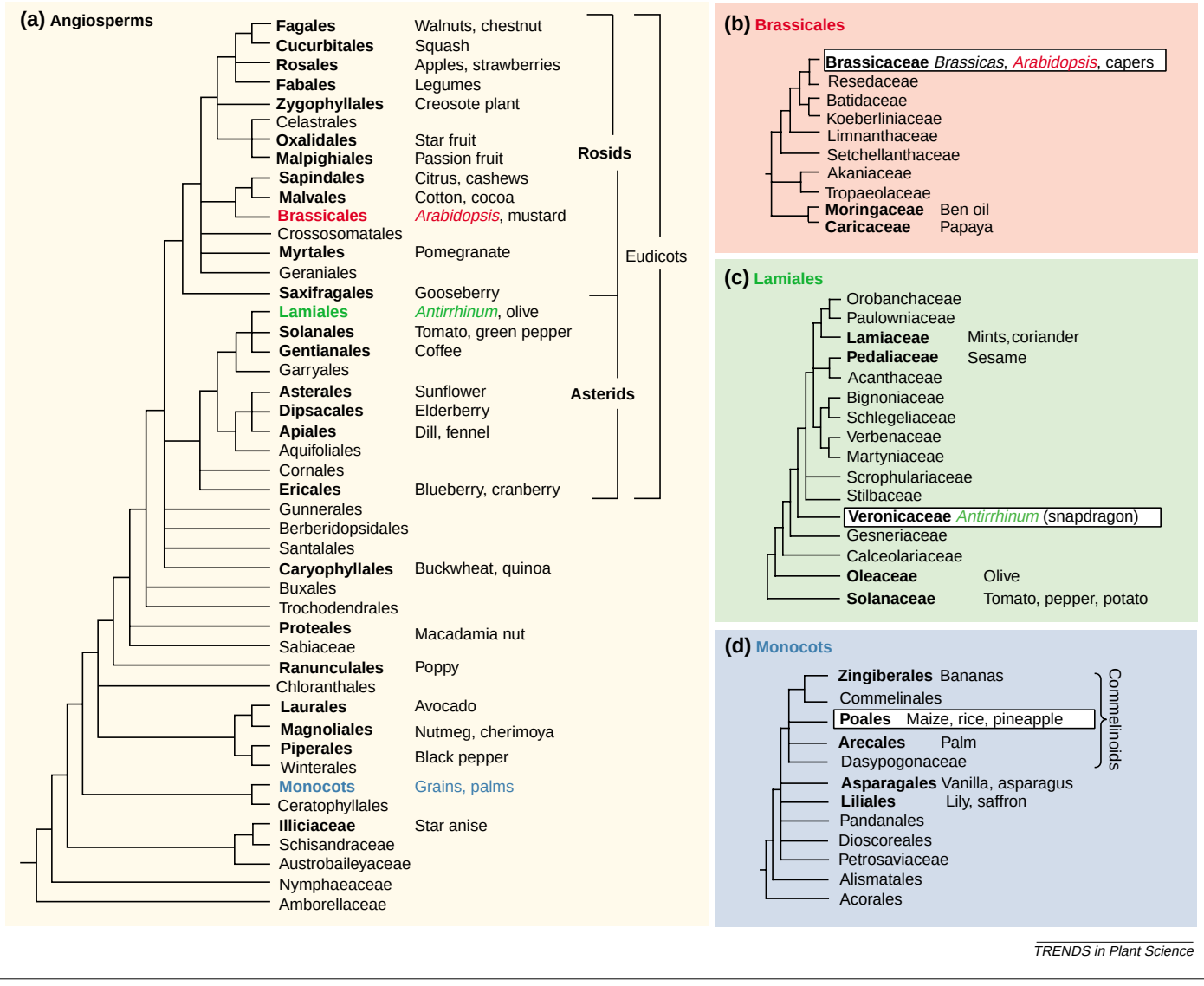


Fig. 2. The summary tree (a) is based on 567 taxa, each sequenced for three genes (~5000 bp per taxon [31,32]), with relationships between basal angiosperms updated to reflect recent analyses based on 6000 to >15 000 bp of sequence data per taxon [31,34,35,38,39]. Both *Arabidopsis* and *Antirrhinum* are members of the eudicot clade, and both appear in derived positions within well defined clades of core eudicots, the rosids and asterids, respectively. Brassicaceae (which contain *Arabidopsis*) are nested well within Brassicales [32,33]; Brassicales are, in turn, deeply nested within the rosids. An expanded summary tree (b) is provided for the Brassicales (modified from Refs 33,45). *Antirrhinum* belongs to the Scrophulariaceae (nested within the Lamiales), a family now known to be polyphyletic [46]; the subclade containing *Antirrhinum* is nested well within the asterids. An expanded summary tree (c) is provided for the Lamiales (modified from Ref. 46). Maize and other grasses (Poaceae) are also in a highly derived phylogenetic position; Poaceae are a tip clade within Poales, which are in turn nested within the commelinoid clade of the monocots (modified from Ref. 33; see also Ref. 47). A summary tree for the monocots is provided (d; modified from Ref. 33). In all figures, only well supported nodes (as measured by bootstrap or jack-knife values) are depicted. Major groups that contain plants for which some portion of the flower, fruit or inflorescence is of economic importance (but not including ornamentals) are listed in bold. Common names of one or a few such plants are provided following the clade name (e.g. Laurales – avocado). This is by no means a comprehensive list of economically important plants. Clades containing the model organisms *Antirrhinum*, *Arabidopsis* and maize are highlighted in boxes.

the ordinal level [38] contain economically important plants, including Caryophyllales, Ericales, Apiales, Asterales, Solanales, Lamiales, Myrtales, Brassicales, Malvales, Sapindales, Malpighiales, Oxalidales, Fabales, Cucurbitales, Rosales, Gentianales and Fagales (Fig. 2). When ornamentally important plants prized for their flowers are considered, the list of economically important angiosperms increases considerably. Thus, efforts to improve crop and ornamental plants also argue for an increased understanding of the genetic underpinning of floral form in a diverse, more evolutionarily representative array of angiosperms.

Advances in angiosperm phylogenetics provide the framework for selecting appropriate exemplars for a more comprehensive analysis of the genes controlling floral development, as well as for the interpretation of results in an evolutionary framework [32,40]. For example, genes with relationships that mirror organismal relationships and that are expressed in similar stages of organ development are likely to be orthologs and functional equivalents of widespread importance in angiosperms. In addition, specific evolutionary questions can be addressed in a phylogenetic context. These include the 'mostly male' hypothesis of floral origins, which predicts that a clear majority

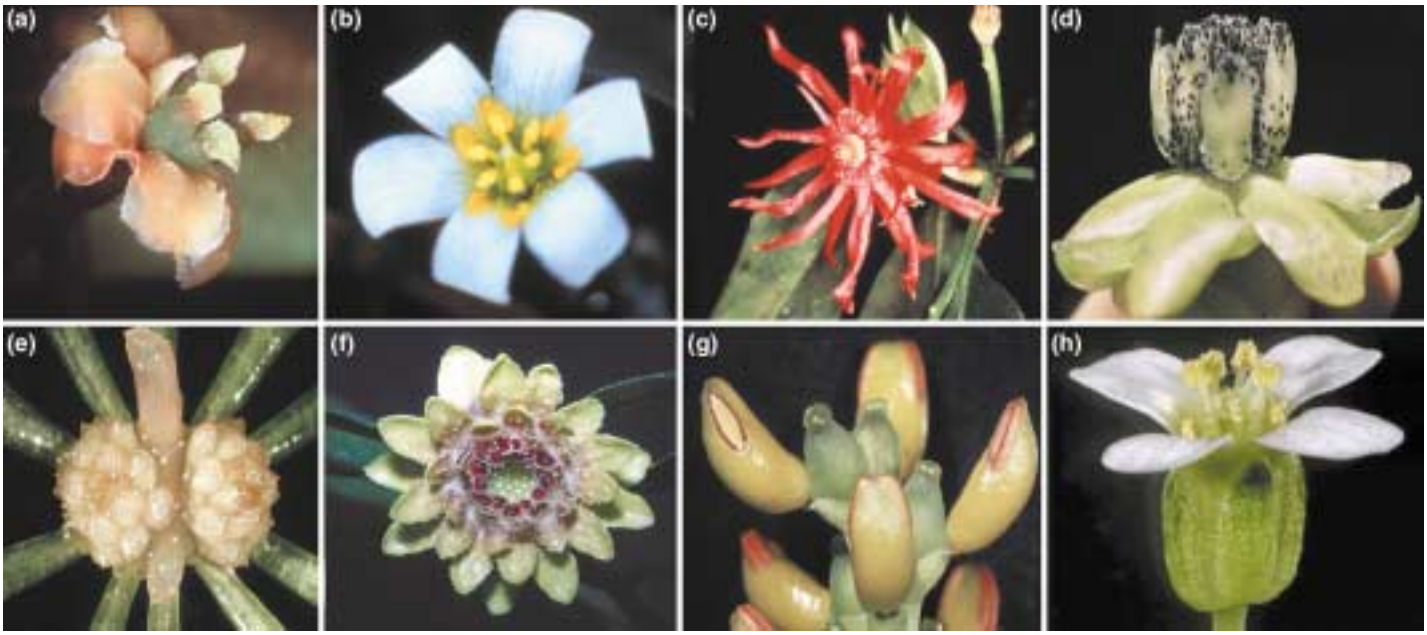


Fig. 3. Photographs of basal angiosperms (a–g) illustrating floral diversity, with *Arabidopsis* (h) shown for comparison as a core eudicot. (a) *Amborella trichopoda* (Amborellaceae), the sister to all other angiosperms (Fig. 2), has unisexual flowers (this is a pistillate flower) with a spirally arranged, undifferentiated perianth of 6–15 tepals. Pistillate flowers have four to eight free carpels. Staminate flowers possess many (12–21) laminar stamens. (Photograph courtesy of Sandra Floyd). (b) *Cabomba* sp. (Nymphaeaceae), which represents one of the earliest diverging lineages of extant angiosperms (Fig. 2), has bisexual flowers and a perianth consisting of three sepals and three petals; there are six stamens and one to three carpels. (Photograph courtesy of Ed Schnieder). (c) *Illicium* sp. (Illiciaceae), another early-diverging angiosperm lineage (Fig. 2), has bisexual flowers with an undifferentiated perianth with many tepals. The tepals are commonly arranged in several series (spiral within each series), with those of the outermost series small and sepal-like, and those of the inner series larger and more petaloid; the innermost tepals can also be reduced and transitional to the stamens. The stamens are numerous, distinct, spirally arranged and more or less laminar. Carpels are numerous and distinct. (Photograph courtesy of Douglas Soltis). (d) *Austrobaileya scandens* (Austrobaileyaceae) has bisexual flowers with numerous, undifferentiated, spirally arranged perianth parts; stamens and carpels are also spirally arranged and numerous. (Photograph courtesy of Peter Endress). (e) *Ceratophyllum demersum* (Ceratophyllaceae) has minute, unisexual flowers (two clusters of flowers are depicted). A series of scales (6–13), sometimes interpreted as a perianth, surrounds the androecium or gynoecium. Male flowers have a variable number (3–46) of stamens that are arranged either spirally or in whorls; stamens are not differentiated into filament and anther. Female flowers each possess a single carpel. (Photograph courtesy of Peter Endress). (f) *Eupomatia benettii* (Eupomatiaceae), a member of Magnoliales (Fig. 2), has distinctive bisexual flowers that do not have a perianth but are covered when young by a deciduous calyptra (interpreted as a bract). There are many (~70) spirally arranged stamens that are attached at the rim of the receptacle, the inner stamens are sterile and more or less petaloid. The carpels are also numerous and spirally arranged, and connate (fused) by their margins to form a sort of compound pistil. (Photograph courtesy of Peter Endress). (g) *Sarcandra chloranthoides* (Chloranthaceae) represents a distinct lineage of basal angiosperms (Fig. 2). This species has bisexual flowers that lack a perianth. Several flowers are shown here, each consisting of a single stamen and a single carpel. (Photograph courtesy of Peter Endress). (h) *Arabidopsis* (Brassicaceae) is a core eudicot in the large rosid clade (Fig. 2) and is provided for comparison with the basal angiosperms shown above. Flowers have well differentiated sepals and petals (four of each), as well as six stamens and two fused carpels. (Photograph courtesy of Hong Ma).

of flower genes will be closest relatives to gymnosperm male-expressed, rather than female-expressed, genes [27]. Mapping specific floral characters onto the phylogenetic trees currently available provides a series of testable hypotheses regarding the evolution of perianth differentiation, phyllotaxis and merosity [37,38]. Two alternative hypotheses for reconstructing perianth differentiation are that true sepals arose early in angiosperm evolution (assuming that single-whorled perianths are 'sepals') or that true sepals and petals might have arisen multiple times (assuming that the single-whorled perianth is not clearly specified as 'sepaloid' or 'petaloid'). A spirally arranged perianth has apparently arisen multiple times, with additional switches between spiral and whorled within single families such as Winteraceae. The trimerous floral condition (typically associated with monocots) appears to have arisen early in angiosperm evolution and might be the primary formula for basal angiosperms; a perianth of numerous parts might also have arisen on multiple occasions.

Floral genomics

We suggest that a small group of crucial 'missing link' taxa be investigated intensively through a massive effort focused on the identification and expression patterns of floral genes in models differing principally in the arrangement, number and organization of floral parts. We provide a general scheme of how such a large initiative could be undertaken and completed within a reasonable length of time (Box 1). Essentially, rather than elucidating the genetic architecture one brick at a time, as is the current approach, a functional genomics initiative would permit researchers to construct an entire wall. As a central goal of this genome initiative, many floral expressed sequence tags (ESTs) should be generated as raw data. To reduce the frequency of repeated sampling of abundant transcripts in the cDNA libraries, filter hybridization could be used to identify those clones already sequenced and allow those that do not hybridize ('cold colony') to be selected for sequencing (Box 1). Phylogenetic relationships among the cDNAs as well as between the cDNAs and known flowering genes could then be established using state-of-the-art bioinformatics and phylogeny reconstructions. We stress that this approach requires a detailed understanding of the history of

duplication and divergence for genes that play important roles in flower development and evolution, which in turn requires a careful phylogenetic analysis of full or nearly full coding regions for genes under investigation. Therefore, the generation of finished cDNA sequences is of crucial importance, including targeted isolation of specific cDNA and genomic sequences using PCR, for genes with a bearing on hypotheses of flower origins and the evolution of the floral developmental program. However, sequence analysis alone might be insufficient to identify homologs that perform equivalent functions and homologs might diverge to play different roles in different species. Therefore, genes identified phylogenetically as orthologues of floral development genes from model plants should be further characterized with respect to their expression patterns using, for example, *in situ* hybridization and microarray analysis in floral organs of different developmental stages. Because expression typically correlates well with function [6,10,41,42], adding expression information to sequence data would greatly enhance our understanding of the evolution and function of similar genes in diverse organisms (Box 1). From these data, novel insights into the evolutionary developmental genetics of flowers should be obtained. Coupling of this information with data from mutational or transformational manipulations could yield even more information on gene function.

Because of the enormous amount of data such a project would generate, it would have broad implications beyond questions about the origin and diversification of the flower. For example, increased knowledge of how evolution mediates morphological change will also provide models for those trying to engineer morphological novelty in economically important plants. Nucleotide sequencing of the entire *Arabidopsis* genome is now complete [43] and sequencing of the rice genome is nearing completion. As others have stressed [44], there are many opportunities to use this wealth of data to obtain, ultimately, a more complete understanding of the underlying genetic mechanisms that control growth and development throughout the angiosperms.

However, to achieve such a comprehensive understanding of the conservation and diversification of plant genomes, data for other angiosperms and ultimately other plant groups must be obtained to link the model taxa. A floral genome initiative would provide numerous ESTs for a diverse suite of non-model flowering plants. Through links with experimental and genome investigations of maize, rice, Solanaceae, legumes, cotton and *Arabidopsis*, and by comparing the newly generated cDNA sequences, the functions of many of these ESTs can be interpreted and compared or postulated. Such analyses would facilitate a comprehensive understanding of growth and development in angiosperms and lead to many new hypotheses regarding gene functions. Other spin-offs include the identification of a huge number of genes that could be used to provide a better understanding of angiosperm relationships and an enormous data set for use in comparative genomics and molecular evolution.

A phylogenetically based genomics effort would produce diverse genetic materials that should serve as research resources to both the plant evolutionary and the molecular developmental communities for the next decade. Equally importantly, such an effort would provide training in phylogenomics for a new generation of biologists.

Response to Baum *et al.*

We have presented a new genomic approach to understanding flower evolution and development as an addition to all the existing tools available to answer these important questions, not as a replacement of such existing tools. Clearly, genetic analysis is essential to understanding gene functions and has been successful in identifying many floral genes. Therefore, we agree with Baum *et al.* [48] that genomic and genetic methods are complementary. We support efforts using both approaches.

Baum *et al.* have two primary criticisms of the genomic approach we propose. They state that we will: (1) rely only on known floral genes in *Arabidopsis* and (2) simply do EST sequencing and hence miss genes expressed at low levels. What they criticized are indeed weaknesses of standard EST projects, but are not true for what we have proposed in our 'Missing links' paper. We reply to each of these criticisms below.

Regarding the identification of candidate floral regulatory genes, we will use conservation of sequence, and as far as possible, conservation of expression, among the species we study, as well as among these species and other species. There are many genes in *Arabidopsis* whose functions in flower development were not or have not yet been discovered by forward genetic studies because of genetic redundancy or lethality. For example, the sepaloid phenotype produced by *SEP1–3* triple-mutant plants [17] would never have been uncovered by conventional mutant screens; instead, the genes were isolated via low-stringency screening using a similar gene (*AG*) that itself was isolated by homeotic phenotype [49,50]. We expect that many genes of this type will probably be conserved in many or all angiosperms, and that we can discover them via the approach we propose. Even if for some reason some genes are no longer expressed or extant in *Arabidopsis*, if they are transcribed in other species being studied, we should be able to identify them. Importantly, however, genetic redundancy might prevent them from being discovered in any species if only a

Box 1. Conducting a genomics initiative

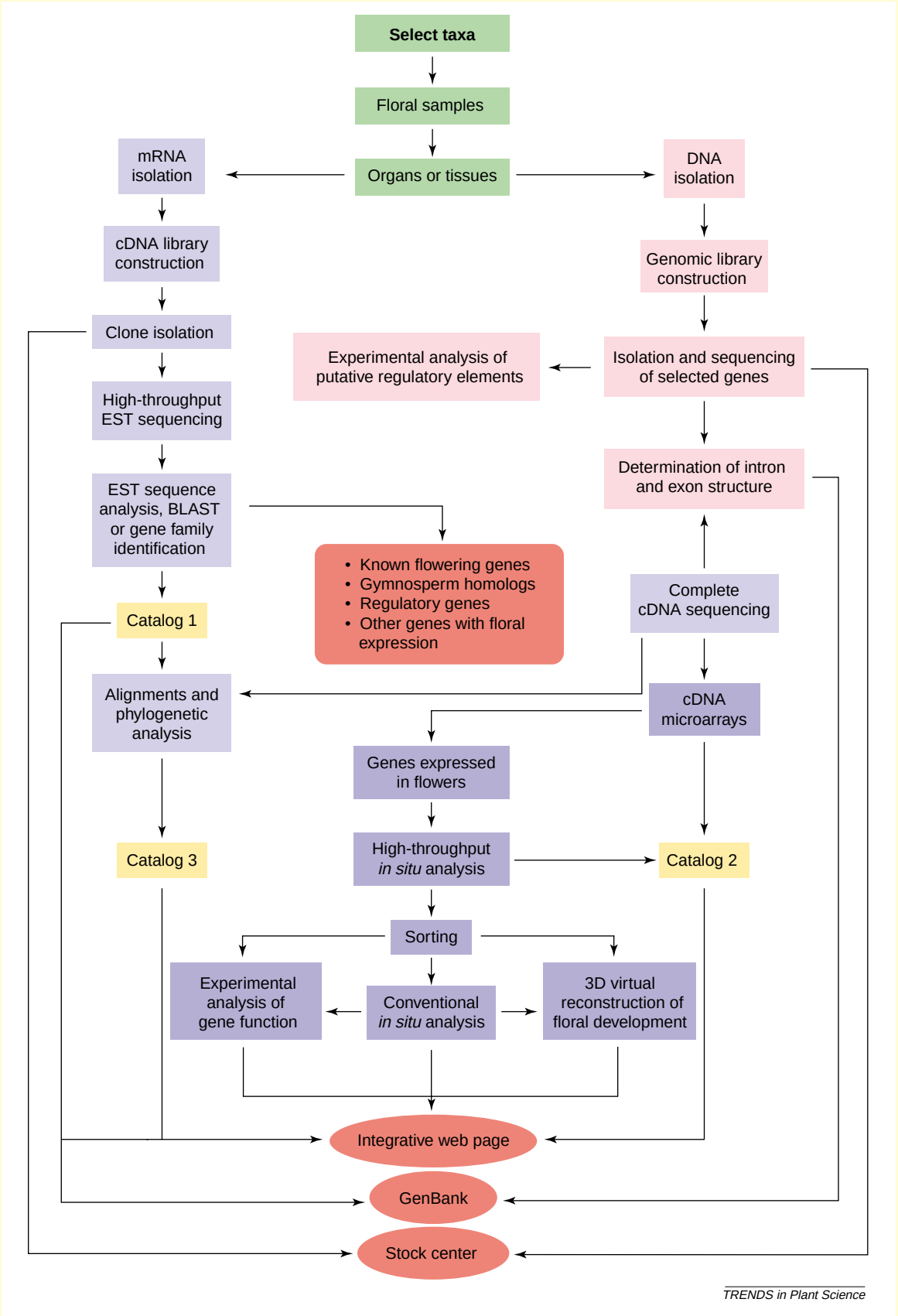


Fig. I. The proposed research scheme. Abbreviations: 3D, three-dimensional; EST, expressed sequence tag.

A general scheme for conducting a genomics initiative focused on the identification and expression patterns of floral genes in exemplar species of 'missing links' (Fig. I).

Tissue collection

The sampling strategy would be designed to identify as many of the genes expressed during the early stages of flower development as possible, because these genes will include crucial regulators. To accomplish this requires the harvest of reproductive meristems (including inflorescence and floral meristems) and floral primordia from all stages up to meiosis.

Expressed sequence tag sequencing

Expressed sequence tag (EST) sequencing will identify many genes expressed during early flower development in each species, including many homologous genes expressed in most or all of the study species. We calculate that at least 10 000 EST sequences are needed per species to detect transcripts of most genes being expressed during crucial events in flower development.

EST sequence analysis, BLAST or gene family identification

A key feature of the proposed work is that the relationships of all orthologous sequences are known in advance, because the taxa from which they are sampled would be chosen precisely because of their strongly supported phylogenetic relationships [a]. Therefore, orthologous sequences will be those that conform to the expected phylogeny, whereas paralogous sequences will be identifiable as those that show discordant patterns. Products of more recent duplications in a given species can be readily identified as they are expected to fall within clades of orthologs.

Filter screening for abundant transcripts

To reduce the frequency of repeated sampling of abundant transcripts in the cDNA libraries, filter hybridization could be used to identify already sequenced clones and allow those that do not hybridize ('cold colonies') to be selected for sequencing. One strategy is to screen lambda libraries using filter lifts from low-density library plating. Probes are generated via PCR of the sequenced inserts and then pooled for labeling and radioactive probing of filters.

Finished sequencing

Finished sequences should provide an important source for examining the evolution of key genes and gene families for flower development, and for tracing the origins of these genes in gymnosperms. 5′ EST sequences will often reach into coding regions, allowing initial assignments of protein homology. However, the approach we propose requires a detailed understanding of the history of duplication and divergence for genes that play important roles in flower development and evolution. This requires a careful phylogenetic analysis of full- or nearly full coding regions for genes under investigation. Therefore, the generation of finished cDNA sequences is crucial for genes bearing on hypotheses of flower origins and the evolution of the floral developmental program.

Expression studies

Expression analysis will be crucial for: (1) testing the conservation of expression patterns for key genes in early flower or inflorescence development in model organisms; (2) generating hypotheses for the function of homologs of angiosperm flower genes in gymnosperm relatives; (3) generating new hypotheses of function based on expression patterns; and (4) building a global classification of sets of genes with specific expression patterns during early flower development. Studies of existing floral regulatory genes and other genes indicate that expression patterns often correlate well with gene function [b–e]. To gain additional insights into the possible functions of newly identified floral genes, the expression patterns of these genes should be characterized at three complementary levels. Information on global expression profiles could be obtained using microarray analysis. This approach should facilitate the analysis of many cloned genes and yield data on temporal expression patterns and spatial expression patterns. A second approach would be to determine the spatial expression patterns of many genes (at least several hundred genes per species) using a newly developed high-throughput PCR-based *in situ* expression analysis [f]. The third approach is conventional *in situ* RNA hybridization. Because this is a more laborious method, it should be used on a selected subset of genes.

Informatics

A genomics initiative of this scope will require a large informatics component. A comprehensive data pipeline will be needed to provide a method to transmit, store and efficiently retrieve information. The large amount of data collected will have to be assembled into a form that can be efficiently retrieved; this can be achieved using catalogs. Catalog 1 ~(gene family identification), catalog 2 (phylogenetic analysis) and catalog 3 (gene expression) will be searchable and have appropriate links to each other.

Three-dimensional virtual reconstruction of floral development

The culmination of this project should be a 'Virtual Flower' web site, an interactive and visual database of floral genetics, development, morphology and evolution.

References

a Soltis, D.E. *et al.* (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL* and *atpB* sequences. *Bot. J. Linn. Soc.* 133, 381–461

b Ma, H. (1998) To be, or not to be, a flower – control of floral meristem identity. *Trends Genet.* 14, 26–32

c Zhao, D. *et al.* Genetic control of reproductive meristems. *Meristematic Tissues in Plant Growth and Development* (*Annu. Plant Rev.*) (Vol. 6) (McManus, M. and Veit, B., eds), Sheffield Academic Press, Sheffield, UK (in press)

d Liljegren, S.J. *et al.* (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770

e Sommer, H. *et al.* (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

f Koltai, H. and Bird, D.M. (2000) High throughput cellular localization of specific plant mRNAs by liquid-phase *in situ* reverse transcription–polymerase chain reaction of tissue sections. *Plant Phys.* 123, 1203–1212

forward genetic approach is used. Hence, we will rely on much more than the known *Arabidopsis* genes, or even known regulatory gene families to identify new floral genes.

Baum *et al.* also imply that we will simply do EST sequencing, and therefore miss low-expression level genes. However, we disagree with this critical evaluation because we plan to do cold-colony selection. This approach will reduce the frequency of repeated sampling of abundant transcripts in the cDNA libraries (Box 1). We have also designed a multi-tiered strategy for sequencing and identifying genes that will include bioinformatic analyses, some of which will identify hypothetical orthologs and paralogs, and others might be expected to group genes by co-expression (Box 1). Although we agree that we will miss some genes just from EST and selected cDNA sequencing, we will employ an added tool – PCR based on information from some of the species that we will study to isolate those genes missing from species among our selected taxa. Even then, we will still miss some genes, but our approach should be a much faster way to identify new genes associated with flower development than the forward genetic approaches are.

Another advantage of our approach is that expression patterns will be important: for example, if a tubulin gene is important in only one of the species we study, we will identify it. This is because our informatics approach will pick up such genes through the use of hierarchical expression profiling and other analysis, regardless of what type of sequences that they might have.

Although forward genetics is not biased by prior knowledge of gene functions, it is biased by genetic redundancy and lethality, both of which have hindered the discovery of gene functions in model plant species. Forward genetics is also a game of chance; the identification of all non-redundant genes involved in a process requires a great amount of effort that will entail isolating mutations in the same genes many times over. This is why genomics has greatly enhanced *Arabidopsis* research, and this is why such a genomics-based effort will provide tremendous insights into floral genes in basal angiosperms.

In short, we feel that Baum *et al.* have misinterpreted our proposed floral genome approaches, and have overlooked the disadvantages of forward genetic approaches. Nevertheless, we do agree that forward and reverse genetic approaches are important and complementary to our genomic approaches. Therefore, additional model plant systems in which both forward and reverse genetic experiments can be conducted should greatly facilitate testing of hypotheses generated from the new floral-genomic information we anticipate.

Acknowledgements
Our work was supported in part by a Fulbright Distinguished Professor award to D.E.S. and P.S.S. We thank Sandra Floyd, Ed Schneider and Peter Endress for providing photographs and images of flowers. The authors are a subset of a larger Floral Genome Project Research Group.

References

1 Bowman, J.L. *et al.* (1989) Genes directing flower development in *Arabidopsis*. *Plant Cell* 1, 37–52
2 Schwarz-Sommer, Z. *et al.* (1990) Genetic control of flower development in *Antirrhinum majus*. *Science* 250, 931–936
3 Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37
4 Lee, I. *et al.* (1997) A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. *Curr. Biol.* 7, 95–104
5 Riechmann, J.L. and Meyerowitz, E.M. (1997) MADS domain proteins in plant development. *Biol. Chem.* 378, 1079–1101
6 Ma, H. (1998) To be, or not to be, a flower – control of floral meristem identity. *Trends Genet.* 14, 26–32
7 Ma, H. and dePamphilis, C. (2000) The ABCs of floral evolution. *Cell* 101, 5–8
8 Theißen, G. *et al.* (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* 42, 115–149
9 Theißen, G. (2001) Flower development, genetics of. In *Encyclopedia of Genetics* (Brenner, S. and Miller, J.H., eds), Academic Press
10 Zhao, D. *et al.* (2001) Genetic control of reproductive meristems. In *Meristematic Tissues in Plant Growth and Development* (*Annu. Plant Rev.*) Vol. 6 (McManus, M. and Veit, B., eds), pp. 89–141, Sheffield Academic Press, Sheffield, UK
11 Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.* 3, 423–434
12 Trotochaud, A.E. *et al.* (1999) The *CLAVATA1* receptor-like kinase requires *CLAVATA3* for its assembly into a signaling complex that includes *KAPP* and a Rho-related protein. *Plant Cell* 11, 393–405
13 Kardailsky, I. *et al.* (1999) Activation tagging of the floral inducer *FT*. *Science* 286, 1962–1965
14 Jacobsen, S.J. *et al.* (1999) Disruption of an RNA helicase/RNAse III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* 126, 5231–5243
15 Song, J.Y. *et al.* (1997) Regulation of meristem organization and cell division by *TSO1*, an *Arabidopsis* gene with cysteine-rich repeats. *Development* 127, 2207–2217
16 Bernstein, E. *et al.* (2001) Role for bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366
17 Pelaz, S. *et al.* (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405, 200–203
18 Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529
19 Theißen, G. and Saedler, H. (2001) Floral quartets. *Nature* 409, 469–471
20 Ambrose, B.A. *et al.* (2000) Molecular and genetic analysis of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell* 5, 569–579
21 Baum, D.A. (1998) The evolution of plant development. *Curr. Opin. Plant Biol.* 1, 79–86
22 Kramer, E.M. and Irish, V.F. (1999) Evolution of genetic mechanisms controlling petal development. *Nature* 399, 144–148
23 Yu, D. *et al.* (1999) Organ identity genes and modified patterns of flower development in *Gerbera hybrida*. *Plant J.* 17, 51–62
24 Heard, J. and Dunn, K. (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5273–5277
25 Rutledge, R. *et al.* (1998) Characterization of an *AGAMOUS* homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant J.* 15, 625–634
26 Winter, K.U. *et al.* (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7342–7347
27 Frohlich, M.W. and Parker, D.S. (2000) The mostly male theory of flower evolutionary origins: from genes to fossils. *Syst. Bot.* 25, 155–170
28 Parcy, F. *et al.* (1998) A genetic framework for floral patterning. *Nature* 395, 561–566
29 Busch, M.A. *et al.* (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science* 285, 585–587
30 Liljegren, S.J. *et al.* (1999) Interactions among *APETALA1*, *LEAFY* and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell* 11, 1007–1018
31 Qiu, Y-L. *et al.* (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402, 404–407
32 Soltis, P.S. *et al.* (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402, 402–404
33 Soltis, D.E. *et al.* (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL* and *atpB* sequences. *Bot. J. Linn. Soc.* 133, 381–461

34 Soltis, P.S. *et al.* (2000) Basal lineages of angiosperms: relationships and implications for floral evolution. *Int. J. Plant Sci.* 161, S97–S107
35 Barkman, T.J. *et al.* (2000) Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13166–13171
36 Endress, P.K. (1994) Floral structure and evolution of primitive angiosperms: recent advances. *Plant Syst. Evol.* 192, 79–97
37 Albert, V.A. *et al.* (1998) Ontogenetic systematics, molecular developmental genetics, and the angiosperm petal. In *Molecular Systematics of Plants II* (Soltis, D.E. *et al.*, eds), pp. 349–374, Kluwer
38 Zanis, M. *et al.* Phylogeny and perianth evolution in basal angiosperms. *Ann. MO Bot. Gard.* (in press)
39 Angiosperm Phylogeny Group (1998) An ordinal classification for the families of flowering plants. *Ann. MO Bot. Gard.* 85, 531–553
40 Kuzoff, R.K. and Gasser, C.S. (2000) Recent progress in reconstructing angiosperm phylogeny. *Trends Plant Sci.* 5, 330–336
41 Liljegren, S.J. *et al.* (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770
42 Sommer, H. *et al.* (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
43 The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
44 Somerville, C. and Somerville, S. (1999) Plant functional genomics. *Science* 285, 380–383
45 Rodman, J.E. *et al.* (1998) Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *Am. J. Bot.* 87, 997–1006
46 Olmstead, R.G. *et al.* (2001) Disintegration of the Scrophulariaceae. *Am. J. Bot.* 88, 348–361
47 Grass Phylogeny Working Group (2000) A phylogeny of the grass family (Poaceae), as inferred from eight character sets. In *Monocots: Systematics and Evolution* (Jacobs, S.W.L. and Everett, J.E., eds), pp. 3–7, CSIRO Publishing, Victoria, Australia
48 Baum, D.A. *et al.* (2002) Response: Missing links: the genetic architecture of flower and floral diversification. *Trends Plant Sci.* 7, 31–34
49 Ma, H. *et al.* (1991) *AGL1–AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* 5, 484–495
50 Mandel, M.A. and Yanofsky, M.F. (1998) The *Arabidopsis* AGL9 MADS box gene is expressed in young flower primordia. *Sex. Plant Reprod.* 11, 22–28

Response: Missing links: the genetic architecture of flower and floral diversification

David A. Baum, John Doebley, Vivian F. Irish and Elena M. Kramer

The genomic approach to understanding how evolution has generated the extraordinary diversity of flowers is to assemble a floral EST database for several missing-link taxa and then use gene phylogenies and expression data to identify genes that are important in flower evolution. However, such a genomic approach is likely to miss important genes that are not members of gene families that control flower development in *Arabidopsis*, and can overlook genes that are not expressed, or are weakly expressed, at their site of action. Therefore we propose complementary genetic approaches in which a few phylogenetically well distributed species are developed as model systems and floral differentiation among closely related species is studied using functional approaches.

There are numerous intellectual and economic reasons for striving to understand the genetics of floral evolution, not least of which is the need to explain the astonishing morphological diversity of the flowers around us. Given recent advances in developmental genetics, phylogenetics and genomics, this is an opportune moment to consider the kinds of strategies that are available for such research. Douglas Soltis and colleagues [1] outline a program to identify and characterize the expression of floral genes in divergent lineages of angiosperms ('missing links'). Here we discuss the importance of complementing their genomic strategy with comparative genetic approaches. Specifically, we advocate in-depth genetic analysis of a few missing links combined with genetic analysis of floral divergence among closely related species.

'…by identifying conserved mechanisms underlying angiosperm flower development and by studying similar processes in gymnosperms, we could learn about the origin of flowers and their component organs.'

Central questions

The starting point for any discussion of research strategy has to be a clear statement of what one wishes to learn. Although our objectives are similar to those articulated by Soltis *et al.* [1], we find it useful to distinguish questions of genetic conservation from questions of phenotypic divergence.

Question 1 (conservation)

What genes and developmental mechanisms are common to the major lineages of angiosperms? This is significant because widely conserved genetic interactions constitute the tablet upon which floral diversification has been written. Furthermore, by identifying conserved mechanisms underlying

http://plants.trends.com 1360-1385/02/$ – see front matter PII: S1360-1385(01)02181-1