PLANT EVOLUTION AND DEVELOPMENT IN A POST-GENOMIC CONTEXT

Quentin C. B. Cronk

Large-scale gene-sequencing projects that have been undertaken in animals have involved organisms from contrasting taxonomic groups, such as worm, fly and mammal. By contrast, similar botanical projects have focused exclusively on flowering plants. This has made it difficult to carry out fundamental research on how plants have evolved from simple to complex forms — a task that has been very successful in animals. However, in the flowering plants, the many completely or partially sequenced genomes now becoming available will provide a powerful tool to investigate the details of evolution in one group of related organisms.

Plant evolution has exerted a controlling influence on the geochemical and climatic evolution of our planet. Through photosynthesis, plants have altered the atmosphere by producing oxygen and providing a sink for carbon dioxide. The colonization of the land by plants created the terrestrial biosphere. The evolution of the root by early land plants promoted soil formation from weathered rock. Finally, the enormous ecological and morphological diversity achieved by the angiosperms supports an even more staggering diversity of insect forms, and the evolution of the leaf, which has a higher concentration of protein than stems do, permits the survival of significant numbers of large complex animals. All these innovations have been driven by genetic changes. The study of plant evolution, at its grandest, is the study of how mutations in genes have affected the way in which the planet functions. In this article, I review the principal morphological landmarks that have characterized the evolution of plants and address how knowledge that has been accrued over the centuries by plant morphologists, systematists and developmental biologists can be integrated with genomic information to address the most intriguing evolutionary question of all — the nucleotide basis of development and adaptation.

Institute of Cell and Molecular Biology,
University of Edinburgh,
Kings Buildings,
Mayfield Road,
Edinburgh EH9 3JH, UK,
and Royal Botanic Garden,
20A Inverleith Row,
Edinburgh EH3 5LR, UK.
e-mail: q.cronk@rbge.org.uk

New phase of the study of evolution

Before the completion of the *Arabidopsis* genome-sequencing project, which was announced in February 2000 (REF. 1), the complexity of the genome was a strong barrier to understanding the specific relationships between genotype and evolution. This is no longer the case. With the advent of rapid sequencing methods, the study of plant evolution has shifted from the gene level to the nucleotide level. It is now possible to investigate how nucleotide changes affect phenotype through altering developmental processes and how selection acts on these phenotypes to drive evolutionary change, both past and present. The use of genomic information is, therefore, set to enhance markedly the study of the evolution of development (known as 'evo–devo'). The great endeavour for the twenty-first century will be to discover the nucleotide basis of morphological differences between organisms, so marking a third phase in the study of natural selection (BOX 1). The completion of this endeavour will require the integration of four disciplines: development, morphology, systematics and evolutionary biology. Systematics, by revealing the pattern of character change in the evolution of particular taxonomic groups, is an essential underpinning of plant evo–devo. The development of molecular systematics has been particularly important in providing accurate organism[2,3] (FIG. 1) and gene[4–6] phylogenies. This

BRYOPHYTE
Land plants in which the gametophyte generation is the larger, persistent phase. Bryophytes include the Hepaticophyta (liverworts), Anthocerotophyta (hornworts) and Bryophyta (mosses).

GAMETOPHYTE
The plant generation that has a haploid set of chromosomes and produces gametes.

SPOROPHYTE
The multicellular diploid form (in plants that undergo alternation of generations). The sporophyte results from a union of haploid gametes and meiotically produces haploid spores that grow into the gametophyte generation.

Box 1 | Paradigm shifts in the investigation of evolution by natural selection

- ***Selection at the organism level (from 1859).*** The first phase in the study of natural selection, originating with Charles Darwin and Alfred Russell Wallace, focused on the organism and on how selection operates on variation in fitness between individuals.
- ***Selection at the gene level (from 1906).*** With the rediscovery of Mendelian genetics at the turn of the twentieth century, the focus of the study of natural selection shifted from the organism to the gene level and Ronald A. Fisher, Sewell Wright and others showed mathematically the evolutionary potency with which natural selection acts on allele frequencies. This gene-level evolutionary study culminated in 'the modern synthesis' promoted by Theodosius Dobzhansky and (in plants) by (George) Ledyard Stebbins, a model that persists to this day. The modern synthesis is a theory about how evolution works at the level of genes and populations, whereas classical Darwinism was concerned mainly with individuals and species.
- ***Selection at the nucleotide level (from 1977).*** The development of sequencing methods in 1977 led eventually to a study of the gene as a collection of nucleotides rather than as a particle of inheritance. It is now possible to investigate how selection acts at the nucleotide sequence level. This further reduction will allow a much more fundamental understanding of evolution, but has yet to be synthesized with other aspects of evolutionary theory.

interaction goes both ways, because if the genetic basis of speciation and of the morphological characteristics of higher taxa can be elucidated, then the systematic task (still great, as the documentation of the world's flora is far from complete) can proceed, strengthened by interactions with these other disciplines.

On being a plant

The plant mode of life. Genomic studies have underlined the similarities between all eukaryotic life. Indeed, some fundamental aspects of basic biology, such as circadian rhythms and the control of the cell cycle, might be more conveniently studied in *Arabidopsis* because of the ease of transformation and experimentation. However, this overdue breaking down of barriers should not obscure the fact that several features are peculiar to plant biology (BOX 2).

For certain developmental problems, such as the elucidation of signal transduction pathways from external stimuli and the control of gene expression by DNA methylation, plants might be the most suitable model organisms. However, there are interesting evolutionary

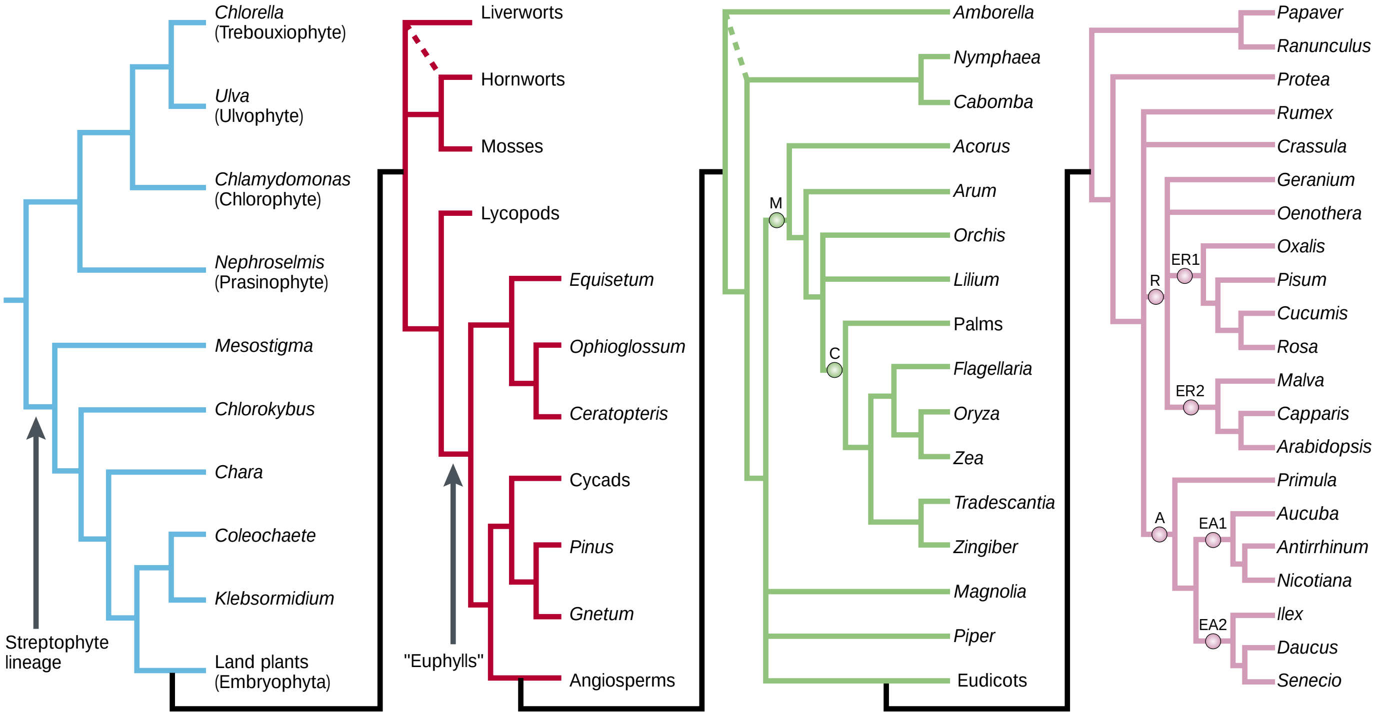


Figure 1 | **Summary of green-plant phylogeny.** Algae (blue): *Mesostigma* is an interesting basal unicellular green alga, either basal to the green algae[77], or to the streptophytes[78]. *Klebsormidium* is a filamentous alga, but *Chara* and *Coleochaete* both show some parenchymatous growth. The relationships of these three genera are still unclear. Land plants (red): hornworts, liverworts and mosses are the basal land plants but their relationships are not yet clear. There is evidence that both liverworts[79] and hornworts[80] are basal. Recent evidence shows that *Equisetum* and ferns form a monophyletic group, sister to seed plants[81]. However, the relationships between different groups of ferns and *Equisetum* are as yet poorly supported. Angiosperms (green and purple): *Amborella* and waterlilies (*Nymphaea* and *Cabomba*) are the basal angiosperms[82,83], and there is some evidence that they might form a monophyletic group[84]. *Magnolia* and *Piper* probably form a monophyletic group, with related woody basal angiosperms, that is sister to the eudicots[82]. Dashed lines indicate uncertain relationships. Eudicots: R, rosids; ER1, eurosid1; ER2, eurosid2; A, asterid; EA1, euasterid1; EA2, euasterid2; M, monocotyledons; C, commelinids.

Box 2 | Comparison of selected features of higher plants and animals

- Unlike animals, plants are characterized by an iterative mechanism of development rather than a linear one. Development continues throughout the life of the plant and proceeds by the often indeterminate, repeated iteration of modules (leaves, roots and stems). By contrast, animals have a single developmental trajectory, which is completed at maturity and ends with a fixed number of organs. The iteration of these modules is controlled by the signals that plants exchange with the environment. This continuous feedback between development and the external environment is not seen in animals, which, in general, have the power of changing their environment by motility.
- Plants are also characterized by the alternation of generations, a life cycle in which haploid and diploid generations alternate with each other. Very different developmental trajectories, which are linked to ploidy level, are thus contained in the same genome. The specific developmental paths undertaken by haploid spores and diploid zygotes indicates that radically different gene-expression patterns are possible — a phenomenon that might be linked to the changes in DNA methylation.
- Plant cells contain an extra genome that is not found in animals. The chloroplast (cp) genome is not large but it is significant, comprising much of the machinery of autotrophy. In its evolutionary history the cpDNA has exchanged genes with the nuclear DNA. Chloroplast genes interact with nuclear ones, thus contributing to cytoplasmic inheritance.
- The presence of a thick cell wall is also a distinguishing feature of plants. The elaborate cell wall of plants — a secreted lattice of carbohydrate and proteoglycans — acts like a sponge to provide an APOPLASTIC space for water and ion transport. It effectively prevents cell movement in plant development. Animal cells can slide past each other to take up programmed developmental positions. By contrast, fate information during development in plants is probably controlled more by cell–cell signalling than by developmental history (cell lineage). As a result, there is no distinction between germ line and soma in plants: the development of germ cells (male pollen and megaspore mother cells) is determined by cell position and not cell lineage.

APOPLASTIC	Pertaining to the free space of tissue; specifically the cell wall porosity and intercellular spaces.
THALLUS	A cellular expansion that forms the main body of thalloid plants, such as algae and liverworts. Thalloid plants have no roots, stems or leaves, and include liverworts, hornworts and pteridophyte gametophytes.
STOMATA	Openings in the epidermis of a plant that permit gaseous exchange with the air. In general, all land plants except liverworts have stomata in their sporophyte stage.
MONOPODIAL	Growing continuously from a single growing point (meristem).
MERISTEM	A collection of stem cells in plants — undifferentiated but determined tissue, the cells of which are capable of active cell division and subsequent differentiation into specialized and permanent tissue, such as shoots and roots.

questions posed by the unique features of plants: what was their origin? what is their biological significance? and how have they evolved? The alternation of generations, in its shifting guises, has been central to land-plant evolution[7]. In BRYOPHYTES and their early land-plant ancestors, the haploid GAMETOPHYTE was large and autotrophic, and dependent on liquid water for the motile sperms it produced to swim in. During evolution, this gametophyte became smaller, eventually being protected by a coat provided by the diploid SPOROPHYTE. In angiosperms (the flowering plants), the gametophyte generation is tiny: the pollen tube and the embryo sac (both dependent for nutrition on the sporophyte) are the vestiges of the large green THALLUS of the ancestral gametophyte.

Angiosperms have a double fertilization: one gamete fertilizes the egg, producing an embryo, and the other gamete fertilizes the two polar nuclei of the female gametophyte. The latter produces the endosperm, a triploid food body, the cells of which are destined to die in nourishing the embryo. The endosperm constitutes a ‘third organism’, which is created in addition to the gametophyte and sporophyte. Endosperm is also important in human nutrition: bread, beer, rice, pasta, popcorn and coconut (both flesh and milk) are all derived from it.

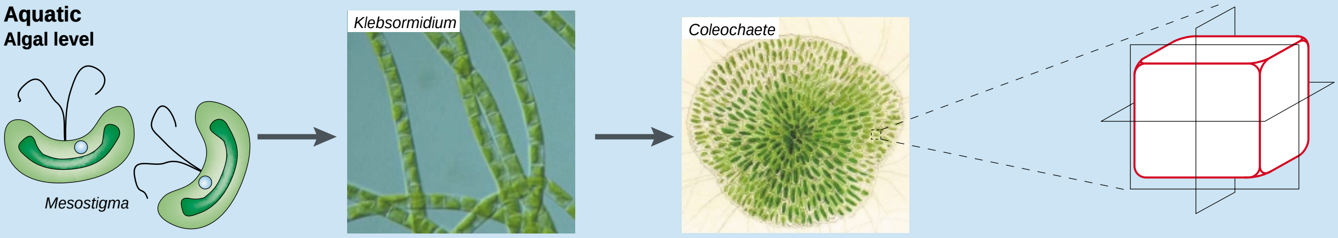
Three polycomb genes normally prevent the endosperm from forming unless fertilization of the egg occurs[8–10]. These genes, which belong to the *FERTILIZATION INDEPENDENT SEED* (*FIS*) group, are *MEDEA* (*MEA* or *FIS1*), *FIS2* and *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE* or *FIS3*). The activity of these three genes confers an important adaptive advantage on the angiosperms, as it prevents the sporophyte from producing a nutritious seed if there is no embryo to nourish. By contrast, gymnosperms produce wasteful reproductive structures even if fertilization has not occurred. In addition, because these genes are imprinted and the paternal copy is not expressed, they allow the maternal tissue, on which the embryo is effectively parasitic, to take control of the reproductive process. The discovery that the developmental trajectory of the endosperm is under the control of imprinted and imprinting genes is very interesting. Similar genes might be involved in the gametophyte versus sporophyte developmental trajectories, in which case the use of imprinting in the evolution of the endosperm might echo an older use of imprinting for alternation of generations. In this context, it could be significant that *FIE* is widely expressed in the sporophyte as well as in the gametophyte[10]. The control of parasites is an important evolutionary matter; the gametophyte, the embryo and the endosperm are all ‘parasitic’ on the sporophyte, indicating that imprinting too might be important in controlling this parasitism to be an efficient conduit for the genes of the adult sporophyte.

The rise of plants. Any evolutionary consideration of plants must begin with the algae, particularly the green algae, which are characterized by the presence of chlorophyll *b* — like the land plants they gave rise to (FIG. 1) — but are generally aquatic and relatively simple in organization. It is unfortunate, therefore, that so little genomic information exists about them. Such is the disproportionate economic importance of angiosperms that most developmental and genomic studies have been concentrated on them, to the considerable detriment of evolutionary studies — an important shortfall, as much fundamental morphological variation occurs in the algae and non-seed plants. This situation is not mirrored in animals, in which protozoa, worm, fly, fish and mouse — which form an evolutionarily graded series — have all received intense scrutiny at the genomic level.

In this review, I consider only the green algae and their land-plant descendants (chlorobionta). Other algal groups have their own evolutionary stories but, not having given rise to land plants, these lineages have produced a smaller range of morphological innovation. The first principal steps in green plant evolution were the transition from unicellular to filamentous growth (multicellularity) and the subsequent transition to parenchymatous growth, in which cell division occurs in many planes. This was followed by land colonization — one of the big events in Earth history — associated with many innovations: a cuticle to prevent water loss, pores (STOMATA) that allow gaseous exchange through the impermeable cuticle and a protective coat that permits the unicellular spores to be dispersed in the atmosphere, unharmed by ultraviolet light and desiccation. Having colonized the land, the empty terrestrial ecospace was gradually filled. This process involved more genetic and developmental innovation as plants evolved roots, lateral photosynthetic leaves, a vertical MONOPODIAL apical MERISTEM and a vascular cambium for structural support.

Box 3 | Outline summary of some important innovations of plant evolution

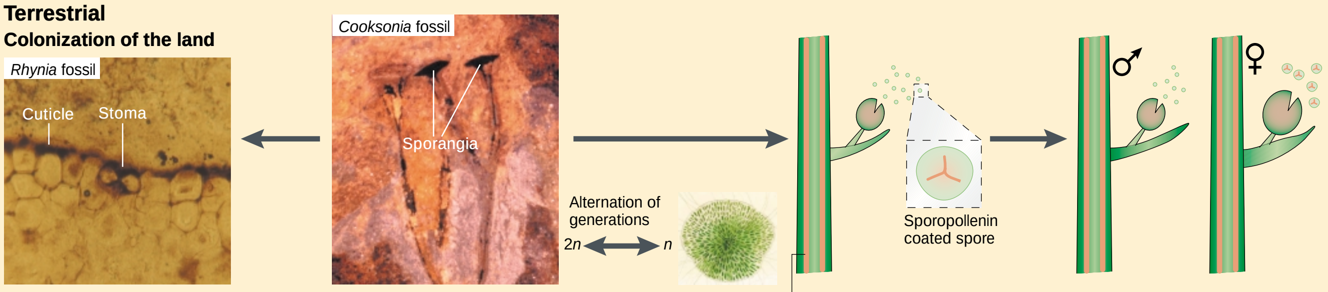
Aquatic
Algal level



Unicellular to filamentous transition. Products of mitosis do not separate, thus forming long filaments. Plane of division is always parallel to the direction of growth. Filaments can produce mats, which, by trapping evolved oxygen, can rise to the water surface (the photic zone). Evolved: several times. *Klebsormidium* image © Yuuji Tsukii, Japan.

Filamentous to parenchymatous cell division. A complex multicellular organism can form only if cell division occurs in more than one plane. This requires sophisticated control of a developmental mechanism that regulates cell division in response to external signals. Evolved: several times. Example: *Coleochaete* image © Mike Clayton, USA.

Terrestrial
Colonization of the land



Secretion of a hydrophobic chemical layer onto the surface of the plant (cuticle). Evolved: once. With the evolution of the cuticle, an alternative system for gas exchange is needed, rather than diffusion across the plant surface. Cuticularized parts of bryophtes have stomata (breathing pores), as do the pteridophytes and seed plants. Evolved: probably once.

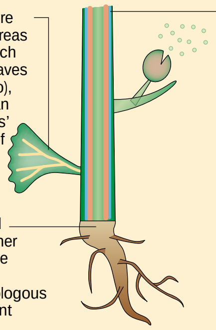
All land plants have sporopollenized spores, which is thought to protect the spore from the hostile aerial environment. The microspore cannot avoid travelling through this hostile aerial environment, as it is a principal instrument of gene flow in the population of otherwise stationary organisms. Evolved: once. *Cooksonia* and *Rhynia* images © Hans Steur, The Netherlands.

Conducting tissue. All land plants have some sort of water-conducting tissue. In mosses, this is called a hydrome. In vascular plants, this is better developed and is called xylem. Evolved: probably once.

Homospory to HETEROSPORY transition. The evolution of a large female megaspore with more resources than the small microspore represents an important evolutionary division of function, requiring a new developmental control of sex expression. Evolved: several times (at least six, probably nine), but once in the seed-plant lineage.

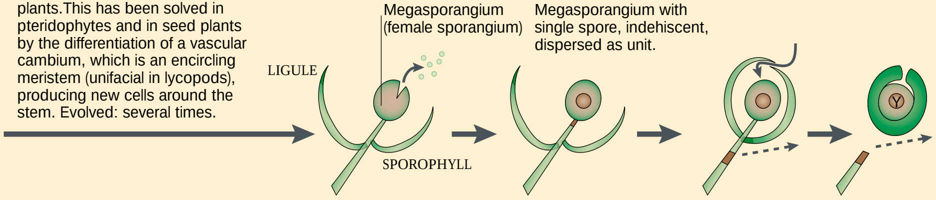
Filling the terrestrial ecospace

Leaves are a unique feature of the land flora, and whereas the first land plants of which we have fossils had no leaves (Silurian 440–410 Myr ago), by the end of the Devonian (410–360 Myr ago) 'leaves' had evolved a minimum of six times.



A means of secondarily thickening the stem after primary growth is needed to support large-stature plants. This has been solved in pteridophytes and in seed plants by the differentiation of a vascular cambium, which is an encircling meristem (unifacial in lycopods), producing new cells around the stem. Evolved: several times.

Roots might have evolved twice (in lycopods and other pteridophytes). If this is the case, lycopod roots and other roots are non-homologous and should show important differences in the genetic systems underpinning their development. This hypothesis can only be tested if more is known about the molecular genetics of root development generally, and about the molecular genetics of root development in lycopods specifically.

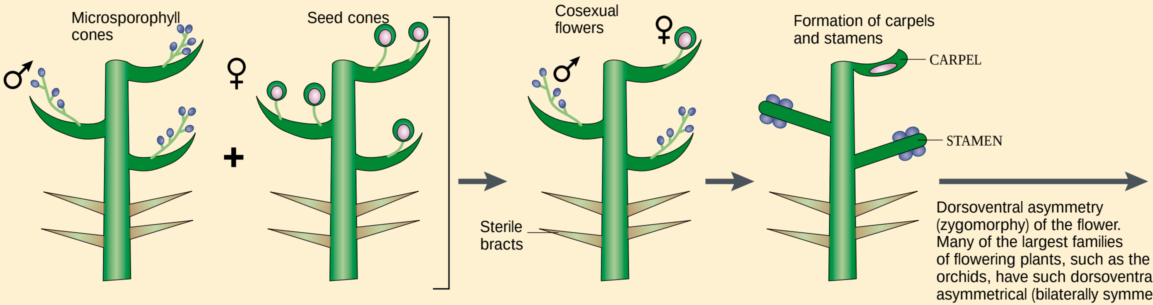


Evolution of the seed. The seed reduces the requirement for external water in which the gametes swim, by enclosing the female gametophyte in sporophyte tissue. In gymnosperms, this forms a chamber into which liquid (pollen drop) is secreted. The enclosing sporophyte tissue also provides protection when the seed is dispersed. Evolved: several times.

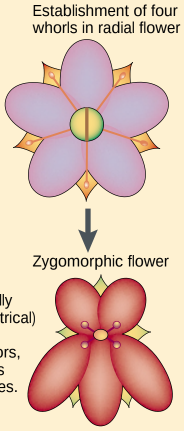
Megasporangium protected by investing bracts past which microspores have to pass.

Investing bracts form a seed coat.

Flower evolution



Production of male and female sporangia in a cosexual structure (the hermaphrodite flower). The extant gymnosperms have separate male and female cones (and this is probably true of the ancestor of gymnosperms and angiosperms), whereas the angiosperm flower is essentially a bisexual cone. Evolved: once in the ancestor of extant hermaphrodite seed plants (angiosperms), but independently in some extinct groups of gymnosperms (Example: *Bennettitales*).



Dorsoventral asymmetry (zygomorphy) of the flower. Many of the largest families of flowering plants, such as the orchids, have such dorsoventrally asymmetrical (bilaterally symmetrical) flowers, and it is possible that adaptation to specialist pollinators, aided by zygomorphy, promotes speciation. Evolved: several times.

HETEROSPORY
The condition of producing two types of spore of different sizes: megaspores (female) and microspores (male).

LIGULE
An appendage on the upper side of a grass leaf at the point where the sheath joins the blade.

SPOROPHYLL
A leaf-like organ bearing sporangia (containing spores). The sporophyll and sporangia together form the basic reproductive unit of the sporophyte generation of land plants. Ovules of seed plants are derived from sporophylls and other organs, whereas the filament of the angiosperm stamens is a sporophyll.

CARPEL
The female reproductive organ of a flower.

STAMEN
The male organ of the angiosperm flower.

SPHENOPSIDS
A group of pteridophytes called horsetails, now represented by the extant genus *Equisetum*, but formerly much more diverse with many extinct woody forms. Pteridophytes comprise vascular plants in which both the gametophyte and sporophyte are free living. Other members of the group include the extant lycophytes and ferns, and many extinct groups, such as trimerophytes.

INDETERMINATE GROWTH
Continuation of the developmental history of an organism or organ when it reaches its adult form. This is characteristic of higher plants. By contrast, determinate growth describes the cessation of the developmental history of an organism or organ when it reaches its adult form. This is characteristic of higher animals.

PHENOLOGY
The timing of periodic biological phenomena that are usually correlated with climatic conditions.

The main morphological innovations in the green plants are shown in BOX 3. It is interesting to note that in animals, multicellularity evolved only once but the land was colonized many times. This contrasts with plants, in which multicellularity occurred several times but the land was colonized only once.

It is a priority of plant evo–devo research to understand the genetic basis of the innovations which led to terrestrial colonization, but there is, as yet, no definite genetic evidence to help understand the origin of these features.

Leaf formation. The leaf was a particularly important evolutionary innovation and it arose six times independently (in mosses, liverworts, lycopods, ferns, SPHENOPSIDS and seed plants). We do not know whether similar molecular genetic processes have been recruited for each of the multiple origins. The leaves of euphyllophytes (ferns, sphenopsids and seed plants) share several important features. Most strikingly, although they are thought to have evolved from similar INDETERMINATE lateral branch systems, they are now determinate structures. Indeterminate meristematic growth is maintained by the knotted-like homeobox (KNOX) gene family. The gene *PHANTASTICA* (*PHAN*), which acts in the leaf primordium to repress KNOX gene expression, has been isolated from *Antirrhinum majus*, the snapdragon[11,12]. Homologues of this gene have now been isolated in maize (*ROUGH SHEATH2*, *RS2*)[13–15] and in *Arabidopsis* (*ASYMMETRIC LEAVES 1*, *AS1*)[16], and there is no doubt that this gene is of crucial importance for the development and evolution of the angiosperm leaf. The primary function of these genes is to downregulate the expression of meristem-promoting KNOX genes in the leaf. They, in turn, are negatively regulated by other KNOX genes, which prevent their expression in the meristem. In *Arabidopsis*, for instance, *AS1* is negatively regulated by *SHOOTMERISTEMLESS* (*STM*), which restricts *AS1* expression to the developing leaves, whereas AS1 itself downregulates the expression of *KNAT1* (*KNOTTED-like Arabidopsis thaliana 1*) and *KNAT2* in the leaf. The leaves of the *as1* mutant therefore show abnormal growth, such as leaf-lobing due to overexpression of KNOX genes. The lower leaves of *phan* mutants in *Antirrhinum* also show extra growth, which makes these genes prime candidates for specifying leaf fate and for evolving determinacy in the lateral shoot systems in ancestral euphyllophytes, such as the extinct trimerophytes. From such determinate, or at least semi-determinate, lateral branch structures, the seed-plant leaf evolved. However, seed-plant leaf evolution involved a complex series of changes (FIG. 2) and a series of genes are likely to have been recruited. For instance, leaves have evolved an important extra innovation: a distinct adaxial (upper) surface not homologous to any stem surface. Abaxial (lower) identity in stems is homologous to the abaxial surface identity of leaves and is the default state. Abaxial identity is as old as land plants and (at least in *Arabidopsis*) is specified or enhanced by *KANADI* and the YABBY gene family[17]. The establishment of the adaxial surface occurred some 200 million years after the establishment of abaxial identity. Interestingly, the upper leaves of *phan* mutants in *Antirrhinum* are radial and abaxialized, and therefore lack this crucial adaxial surface. Perhaps in the developmental context of these leaves, PHAN interacts with another (as yet poorly known) genetic pathway that specifies adaxial identity. The better elucidation of this pathway would be a great step forward in understanding leaf evolution. One component is *PHABULOSA* (*PHB*), a homeodomain leucine-zipper-encoding gene[18]. It seems that the PHB protein is only activated in the adaxial domain (probably by an adaxially expressed ligand) and the active form of the protein is then involved in a positive-feedback loop, promoting further adaxial *PHB* expression, so providing a stable specification of adaxial identity.

Evolution of reproductive mechanisms. Once the terrestrial ecospace had been colonized, selective pressures shifted towards reproductive development, as variation in reproductive development represented further unfilled niche space. Important evolution occurred in reproductive PHENOLOGY and in the fertilization, dispersal and germination of reproductive units. A repeated aspect of this was the evolution of heterospory[19], in which the co-sexual gametophyte was replaced by unisexual gametophytes with specialized features suited to their sex. To colonize drier environments, the reduction of dependence on environmental water seems to have led to the evolution of the seed, in which the gametophyte and the water-supplying and protective sporophyte tissue (NUCELLUS and INTEGUMENT) are intimately associated. Finally, the male and female SPORANGIA on different structures became associated in one co-sexual structure — the angiosperm flower.

MADS-box genes are very important for patterning the flower. The lineage leading to the chlorobionta evolved a characteristic plant class of MADS-box gene, containing a conserved MADS-box domain followed by three gene regions, I-, K- and C-, together forming the so-called MADS-IKC gene. Early seed plants contained a considerable diversity of such genes[4], which was available to be recruited to specify the complex combination of identities seen in the diverse organs that make up the flower. The evolving complement of MADS-box genes and their altering expression patterns in plant evolution have led to a wide range of floral types[20]. The expression patterns of MADS-box genes provide a powerful means of testing the homology of floral organs. In the typical EUDICOTS that have four clearly defined whorls of floral organs (SEPALS, petals, stamens and carpels), these whorls are defined by the overlapping expression domains of three classes of MADS-box gene (A, B and C). Stamen identity is determined throughout the eudicots by B + C expression, indicating a single origin of stamens and the homology of all stamens. However, although petals are determined by A + B expression in higher eudicots (*Arabidopsis* and *Antirrhinum*), B-class expression is not uniformly characteristic of petals in lower eudicots (*Ranunculus* and poppies), which indicates multiple

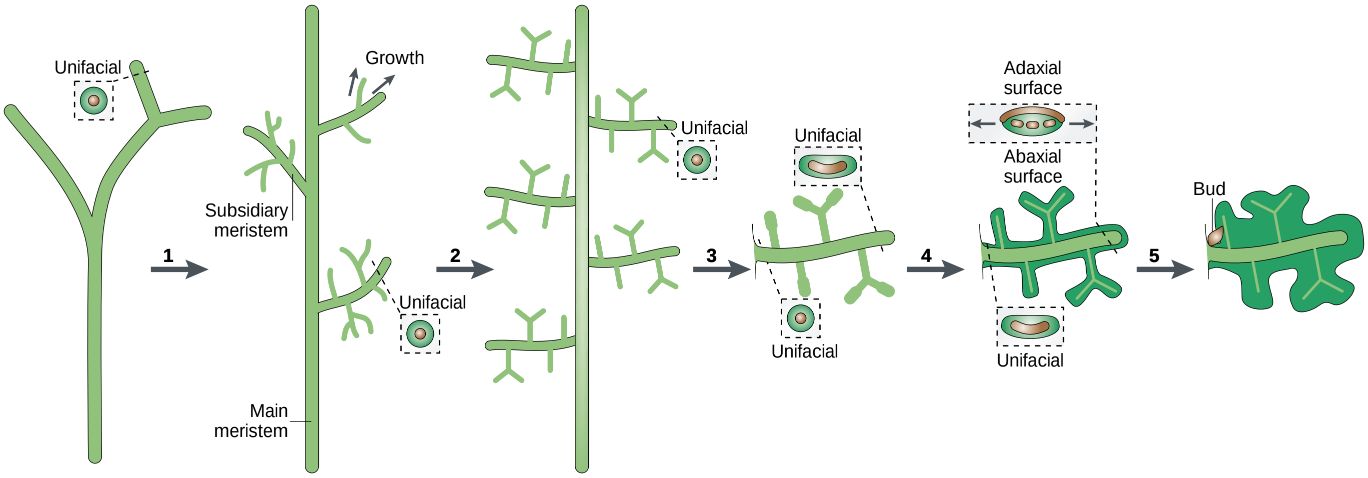


Figure 2 | **Five-step hypothesis of seed-plant leaf evolution.** Leaf evolution requires the recruitment of at least five broadly defined gene systems, each consisting of several different, but interacting, developmental genes. Recruitment of such systems in the following order would be consistent with both the observed morphology of leaves and with the fossil record: **1**, a monopodiality system — the division of meristem into main and subsidiary meristems, probably through an increase in complexity of meristem control networks; **2**, a determination system — cessation of growth at tips in lateral branch systems. Originally, this probably involved the late downregulation of KNOX genes by *PHANTASTICA* (*PH*)/*ROUGHSHEATH2* homologues, which in evolution have become expressed earlier and earlier; **3**, a dorsoventrality system — lateral branch systems of some fossils are unifacial but show internal dorsoventrality. This requires asymmetric, abaxial or adaxial (ventral or dorsal) gene expression, promoting stable adaxial or abaxial expression of key regulators, such as *PHB*, and of the YABBY gene family; **4**, a lamination system (production of a large flat sheet of tissue) — the evolution of an adaxial surface (extreme dorsoventrality) is correlated with the activity of a marginal meristem. This might have arisen through the interaction between adaxial–abaxial identity genes (such as *PHB*) and KNOX genes. As the marginal meristem (located along the margin of the leaf primordium and that forms the blade) is always active at the junction of adaxial and abaxial domains, crosstalk between the domains might constitute the stimulus for lamination; **5**, an axillary bud system — the constant association of a new meristem with the adaxial side of the leaf, which could have arisen by signalling between likely adaxial identity genes (such as the homeodomain-zip gene *REVOLUTA*) and meristem identity genes. Although for several of these functions we are in a position to suggest candidate genes, much further work needs to be carried out on the comparative evolution of the gene networks involved.

NUCELLUS AND INTEGUMENT
The nucellus is the tissue that usually makes up the greater part of the ovule of seed plants. It encloses the embryo sac. It is itself enclosed by one or two protective coats called integuments, which become the seed coat.

SPORANGIUM
A reproductive structure in plants that produces spores by meiosis; in angiosperms, the anthers are groups of four sporangia.

EUDICOTS
The largest clade of angiosperms, characterized by three symmetrically placed pollen apertures or aperture arrangements derived from this.

SEPAL
Sepals form the outer ring of modified leaves that surrounds the petals, stamens and carpels.

PERIANTH
A collective term for all the external parts of the flower: the calyx, or sepals, and the corolla, or petals.

origins of petals in the eudicots, and consequently the non-homology of the stamen-derived ranunculid petal (presumably involving the loss of B-function determination) and the sepal-derived higher eudicot petal (possibly by gain of B expression)[21]. However, studies in maize (a monocot) indicate that B-function genes might have been recruited for PERIANTH determination at an early evolutionary stage. Grasses such as maize are apetalous, but LODICULES — specialized organs found only in grasses — have traditionally been equated with the missing petals and this is supported by the expression of B-function genes[22]. Similarly, MADS-box gene-expression patterns have been used to test ideas of organ homology in the highly derived flowers of the Asteraceae (daisy family). Classically, the PAPPUS hairs, which aid wind dispersal of the dry single-seeded fruit, have been thought to be derived from sepals and, again, MADS-box gene-expression patterns confirm this view[23].

At the same time that the flower evolved, the associated female gametophytes became enclosed and protected by yet another layer of sporophyte tissue, the carpel wall. This new covering might have offered protection against the male gametophytes of the ancestral angiosperm. Pollen tubes, now having to grow through sporophyte tissue, could be recognized as 'self' and stopped. Self-incompatibility (SI) has probably arisen several times, although in many plants, as in *Nicotiana* and *Antirrhinum*, S-locus ribonucleases (S-RNases) are the main stylar agents of SI and this might represent the recruitment for SI of an ancestral RNase gene fairly early in angiosperm diversification, followed by the slow gain and loss of new alleles[24]. New alleles are likely to be strongly selected for, initially, because SI is most efficient with many alleles, as there is less false self-recognition. Conversely, SI alleles are continually lost through population bottlenecks and founder effects.

Flower shape and angiosperm radiation. Angiosperms have proved very successful in colonizing all parts of the earth, the only exception being the polar ice caps. Many physiological adaptations to extreme environments can be seen, such as succulence (both XEROPHYTIC and HALOPHYTIC). Evolution in the floral structures has been particularly important, mainly owing to adaptation to specific pollinating animals. This specialization has been promoted by the evolution of a dorsoventral axis of asymmetry across the flower (zygomorphy), allowing for complexity in flower shape[25]. It is not unreasonable to suggest that zygomorphy promotes speciation and diversification, as some of the largest plant families are predominantly zygomorphic, such as Orchidaceae (~25,000 species), Fabaceae (~20,000 species) and Asteraceae (~18,000 species). It is now known that in at least one plant — the model angiosperm *Antirrhinum* — floral asymmetry is due to the asymmetric expression of the closely related genes *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*)[26,27] (FIG. 3). These genes are members of an important family of TCP/R transcription factors[28], which are characterized by the presence of two conserved α-helix domains: the TCP domain and the R-box domain. *CYC* interacts with regulators of the cell cycle to alter growth asymmetrically across the flower[29].

Figure 3 | **Function of *CYCLOIDEA* and *DICHOTOMA* in *Antirrhinum* (snapdragon) flower morphology.** *DICHOTOMA* (*DICH*) and *CYCLOIDEA* (*CYC*) belong to the TCP/R family of transcription factors. These two closely related genes show the importance of gene duplication in evolutionary elaboration. *DICH* is expressed in the *CYC* expression domain and although it has a weaker mutant phenotype, it serves to elaborate the asymmetric morphology caused by *CYC*[26,27]. Expression of both genes is necessary to produce the typical snapdragon flower morphology (WT). (Photos courtesy of the John Innes Centre, Norwich, UK. Images reproduced with permission from REF. 26 © (1999) Excerpta Medica, Inc.)

Antirrhinum and *Linaria* plants with null mutations in *CYC* and *DICH* have a peloric (radially symmetrical) mutant phenotype[30] (FIG. 3). These mutants were studied by Linnaeus and Darwin, who were interested in their taxonomic and evolutionary significance, respectively. To Darwin, they seemed to have reverted to an ancestral state. Apparent evolutionary (as opposed to mutational) reversals to radial flowers are quite common; for instance, *Cadia* (Fabaceae) has radial flowers, although it is closely related to plants with ordinary (zygomorphic) pea-flowers. In different branches of angiosperm evolution, zygomorphy has evolved independently many times (for instance, in the pea-flowers of the Fabaceae and in the complex flowers of the orchids). It would be desirable to know whether TCP/R genes have been recruited independently in all these cases. A recent study[31] has shown that asymmetrical expression of a homologous TCP gene occurs in *Arabidopsis* and, so, an asymmetrical 'pre-pattern' pre-dates an asymmetrical phenotype. It is, therefore, easy to see how zygomorphy could have evolved many times using fundamentally the same mechanism, as an elaboration of an ancestral pre-pattern. However, at present there is no information about the developmental genetics of zygomorphy in plants outside *Antirrhinum* and *Linaria* — a gap that needs to be filled.

Plant evo–devo genes

We now have a wide range of resources at our disposal to study evo–devo in flowering plants. Is there any indication of what class of gene is likely to be important for adaptive evolution — genes which such tools will allow us to identify? John Doebley has suggested that signalling genes are unlikely to be important for morphological evolution, as the effects of mutating them will be wide-ranging and possibly deleterious. Instead, he proposed that regulatory genes (specifically transcription factors), which have a narrower scope of action, are the loci that create the variation that drives evolution. Although some artificial gain-of-function mutants in transcription factors (such as KNOX genes) are known to be pleiotropic, such genes have very precise roles in different developmental contexts and it is, therefore, likely that evolution can be narrowly partitioned to individual organs. Furthermore, there is growing evidence that altered spatio-temporal expression of regulatory genes often provides the selectable variation. These shifts in expression are caused by changes in the regulation of the gene, probably as a result of mutations in *cis*-regulatory regions[32–34]. Several authors have suggested that the *cis*-regulatory regions of transcription factors have special evolutionary significance[35]. This idea has been elegantly and fully proposed in a plant context by Doebley and Lukens[36], and illustrated by the work of the Doebley laboratory, so I shall refer to it here as the 'Doebley hypothesis'.

Doebley *et al.*[37] traced a principal quantitative trait locus (QTL) for the difference between cultivated maize and its wild ancestor, teosinte, to the transcription factor, *TEOSINTE BRANCHED1* (*TB1*). The maize *TB1* allele confers the all-important complete APICAL DOMINANCE of maize. In their search for the differences between the maize allele and the teosinte allele, Wang *et al.*[34] found evidence of a SELECTIVE SWEEP of the 5′ flanking regions of the maize allele, indicating that the *cis*-regulatory regions of this gene might have undergone intense selection for altered expression of *TB1*. Conversely, there was no evidence for selection on the coding region of the gene (BOX 4). As directional selection under domestication can be very strong, we do not know whether this study provides a good model for the evolution of species in natural conditions. Nevertheless, it clearly shows a successful outcome of an evo–devo study in plants.

Mode and tempo in plant evolution

The term microevolution was originally introduced by Theodosius Dobzhansky in 1937 and taken up by Richard Goldschmidt in 1940. It refers to small-scale change that usually occurs at the level of species and on short timescales. By contrast, macroevolution describes evolution that occurs at or above the level of species. Although cumulative microevolutionary change might account for macroevolutionary pattern, it is not impossible that other mechanisms are involved. The existence of apparently rapid and important evolutionary transitions prompted

LODICULE
Plug or flap of tissue in the grass flower that occurs between the stamens and the bracts that enclose the flower, swelling rapidly to open the flower.

PAPPUS
A group of modified sepals often in the form of a ring of silky or bristly hairs, or scales.

XEROPHYTE
A plant adapted for growth under arid conditions.

HALOPHYTE
A salt-tolerant terrestrial plant.

APICAL DOMINANCE
The tendency for the apical meristem of a plant to be more active than its lateral or axial meristems. It is particularly evident in young trees and is due to the production of auxins (plant hormones) in the apical meristem.

SELECTIVE SWEEP
Process by which new favourable mutations become fixed so quickly that physically linked alleles also become fixed by 'hitchhiking'.

Box 4 | Evolutionary genomics – finding the signature of selection

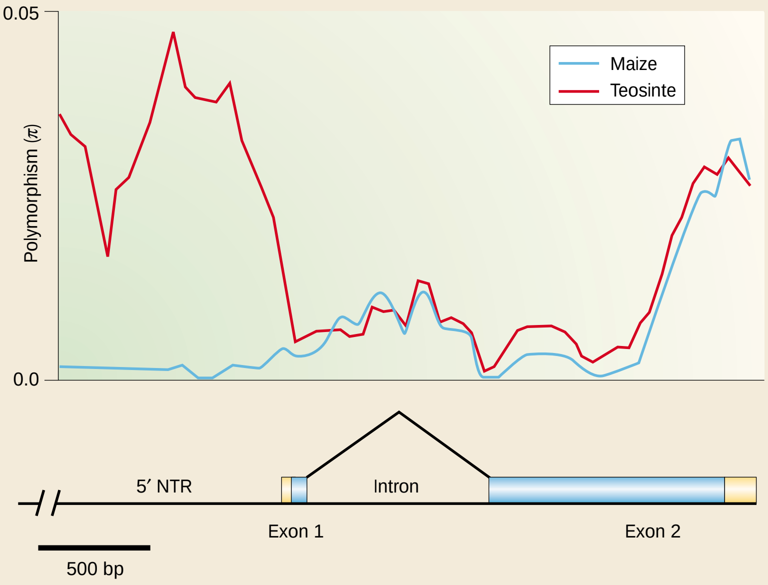
Coding regions

The raw material for quantitative phenotypic variation on which selection can act consists of nucleotide polymorphisms in populations. Recently there have been many attempts to assay this variation and to look for the molecular signature of selection[34,67–70]. In the coding region of the gene, selection can be inferred by examining the differences in synonymous and non-synonymous substitution rates, as has been done in several well-known animal examples[71,72]. Simple 'sliding window' techniques[73], which reduce noise by averaging the value from nucleotides each side of the position being analysed, are straightforward to apply.

Non-coding regions

The signature of selection is harder to find in non-coding regions, partly because no comparison can be made between synonymous and non-synonymous substitution rates and partly because the high variability of these regions makes alignment between species very difficult. However, if alignments can be made, tests of neutrality such as the HKA (Hudson, Kreitman and Aguadé) statistical test[74] can be applied and readily implemented in programs such as DnaSP, a Windows-based program that has made suitable algorithms readily available to evolutionary biologists[75]. The HKA test looks for evidence of selection by comparing fixed and polymorphic sequence differences between and within species at two loci, assuming that the neutral rate of evolution of these sequences is correlated with the levels of polymorphism in species. Where alignments are difficult, an extremely promising new computational method called 'complexity analysis' can be used (very sensitive in detecting complexity such as direct and inverted repeats in apparently random sequence)[76].

The search for the signature of selection in regulatory non-coding regions of transcription factors is particularly important because, if it is found[34], it will provide evidence for the 'Doebley hypothesis' — that morphological evolution proceeds predominantly by altering transcription factor activity. The illustration shows a sliding window analysis of the intensity of selection at the *TEOSINTE BRANCHED1* (*TB1*) locus of maize, as a result of selection under cultivation by indigenous agriculturalists in Mexico over the past 10,000 years. The 5′ non-transcribed region (NTR) of *TB1* has very low variability (polymorphism) in maize compared with its wild ancestor, teosinte, and this part of the gene has undergone a selective sweep under cultivation because of the presence of significant control regions in this area. By contrast, the transcribed part of the gene is only slightly less variable in maize than in teosinte. (Figure reproduced with permission from REF. 34 © Macmillan Magazines Ltd (1999).)

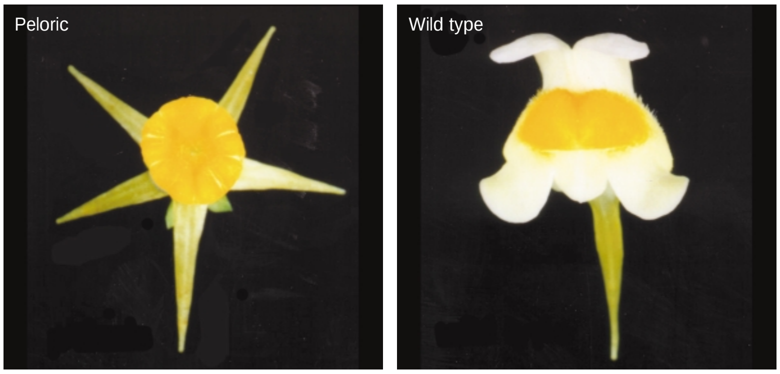


Bateman and DiMichele[38–40] to propose that a single change in a principal regulatory gene could have a large effect on morphology, and that such a phenotype could survive and come to have a fitness advantage in new, but uncompetitive, environments. This 'neoGoldschmidtian' model of evolution contrasts with the neoDarwinian view that evolutionary transitions are produced gradually by selective forces acting on many changes of small effect. NeoDarwinians have not been kind to the neoGoldschmidtian model, pointing out that a single individual (the unfortunately named 'hopeful monster') that carries the mutation of large effect will be unlikely to find an appropriate ecological niche (although epimutations might provide a bridge, see BOX 5). They also point out that, as gradual selection is sufficient to produce even marked evolutionary transitions, it is not parsimonious to put forward an explanation for which there is no need and no evidence.

However, there are some problems that neoDarwinism has not fully addressed. One of these is the problem of 'loss without vestige' (that is, loss of a structure without leaving a vestigial organ). Selection on quantitative variation in organ size is amply sufficient to remove superfluous organs. However, when the superfluous organ becomes very small, the selective advantage of further reduction becomes very low, thus explaining the numerous vestigial organs that are present in animals and plants. Good examples are the staminodes (vestigial stamens) found in most flowers with reduced numbers of stamens. However, some organs are lost completely and are never seen, not even at the earliest stages of development[41]. It is difficult, although not impossible, to see how loss without vestige could be explained in classical (neoDarwinian) microevolutionary terms, but easy to see how the deletion of a developmental pathway as a single event could produce such a result.

Examples of evolutionary change at the nucleotide level will emerge over the next few years. These will probably reveal a range of different scenarios: adaptation to particular fitness peaks is likely to have been driven by classical neoDarwinian microevolution, whereas jumps between different peaks on the adaptive landscape might have been facilitated by rare mutations of large effect. For example, *Arabidopsis* evolved as a winter annual, requiring exposure to low temperature (vernalization) to initiate flowering. However, in the warm climate of the Cape Verde Islands (off the West African coast), the plant behaves as a summer annual — a different adaptive peak. A loss-of-function mutation in the vernalization gene *FRIGIDA* (*FRI*) is responsible for this trait[42]. *FRIGIDA*, a single-copy gene in *Arabidopsis*, is predicted to encode a protein consisting of 609 amino acids. The encoded protein has dubious homology and unknown function, but seems to be essential for the vernalization process. The action of active FRI is dependent on an active *FLOWERING LOCUS C* (*FLC*) gene (a MADS-box putative transcriptional repressor) and FRI acts to increase the RNA levels of FLC. The inactivation of *FRI* function by deleting the 3′-end of the gene is a putative single-step

Box 5 | Epimutation — a plant speciality?



Natural gene variation can take many forms, as indicated by a recent study[30] on a naturally occurring mutation in *Linaria* that causes radially symmetrical flowers (see figure). These peloric flowers result from a heritable epimutation, caused by extensive methylation of the *Linaria* homologue of the *CYCLOIDEA* gene (*LCYC*). The illustration shows peloric and normal flowers. Although there have been no reports in animals of epimutations that are heritable over many generations, they do occur, albeit rarely, in plant mutants in the laboratory, and might be more important in wild plants than previously suspected. This is possibly because there is no separation between germ line and soma in plants and so plants are less efficient at re-setting the methylation status of genes by passage through the germ line. Plants use methylation as a potent means of silencing foreign, usually viral, DNA. Methylation is also part of the gene-silencing mechanism that occurs after genome duplication or after transposon activity, to silence extra gene copies or genes that contain 'foreign' transposon sequences, respectively. Although epimutations are heritable for many generations, the absence of sequence change means that they do not contribute to long-term evolution. Nevertheless, there are five ways in which methylation and epimutation are evolutionarily significant:

- Methylated DNA mutates at a higher rate compared to non-methylated DNA, so epimutations could be considered the first stage of a gene knockout process, which is completed by sequence change.
- Transposon insertion is known to be a powerful mutagenic force; through its gene-silencing effect, epimutation might aid the disruption of a gene with a nearby transposon insertion.
- Epimutations immediately expose mutant phenotypes to environmental selection, as both alleles are likely to be silenced simultaneously. Conventional loss-of-function mutants originate as rare recessive alleles that must be brought together as homozygotes before the phenotype is exposed to environmental selection.
- As duplicate genes are prone to silencing by methylation, epimutation might be a first step in returning a gene to a single locus state, the knockout being completed by subsequent sequence change.
- Epimutations are often partially reversible and the resulting phenotypes are possibly more variable and less severe than those resulting from sequence change. So, epimutation might expose loss-of-function mutations to environmental selection in a less severe way than sequence change, and by being heritable over many generations, offer added opportunity to reach habitats favourable to the mutation. It might also facilitate co-adaptation of other genes to occur, before the otherwise lethal completion of the knockout by sequence change.

(Photos courtesy of the John Innes Centre, Norwich, UK. Images reproduced with permission from REF. 30 © Macmillan Magazines Ltd (1999).)

mutation, which is apparently sufficient to create a summer-annual plant with a different ecology and occupying a different fitness peak in the adaptive landscape from the winter-annual. Analysis of summer-annual *Arabidopsis* plants has shown that this *FRI*-disablement has occurred at least twice. This work is important for being the first to show the molecular basis of a plant adaptive trait in the wild and for showing how, in a single step, a single mutation can produce a plant with very different ecological behaviour.

Gene duplication and redundancy in plants

The duplicated genome. Duplication is a prevalent feature of plant genomes and many genes are found in tandem arrays or in duplicated segmental clusters. The most common cause of segmental duplication in plants is probably polyploidy. Polyploidy is an important cause of plant evolution and speciation, and might be commoner in plants than in animals, perhaps because plants produce powerful spindle inhibitors such as nicotine and colchicine as a defence against herbivory. Furthermore, plants have few mechanisms to control their temperature, so are often subjected to heat and cold shocks, which might also promote polyploidy by inhibiting spindle formation.

Many plants are obvious polyploids, as judged from chromosome number and chromosome behaviour. Although *A. thaliana*, with just five chromosomes, is not an obvious polyploid, whole-genome sequencing reveals that 60% of its genome is segmentally duplicated, for which a polyploidization event is the most likely cause. Like maize[43,44], *Arabidopsis*[1] is an ancient tetraploid and further putative polyploidy events are discernable in its genome[45]. All land plants and many algae will probably have polyploid events in their ancestry. Although rarer, polyploidy in animals might also be evolutionarily significant. The duplication of homeobox (HOX) clusters between *Amphioxus* and vertebrates has revealed that two polyploidization events occurred around the time of the origin of the vertebrates. The extensive gene duplication seen in the complete human genome sequence[46] might well be an echo of these events. The widespread phenomena of polyploidization and subsequent diploidization have recently been reviewed elsewhere[47]. It is intriguing to speculate that polyploidization might have created the genomic diversity necessary for evolutionary innovation in vertebrates. The complete diploidization and chromosome number reduction that has occurred since that tetraploid vertebrate ancestor has left a very complex genomic pattern. Translocation of chromosome

EPIMUTATION
A heritable change in gene expression but not gene sequence. This usually takes place by abnormally increasing the methylation status of a gene, producing a loss-of-function phenotype. This can then be heritable for many generations, unless reset by meiosis.

FITNESS PEAK
A phenotype or part of the possible morphological variation that has high fitness.

ADAPTIVE LANDSCAPE
If all morphological variation or all possible phenotypes are considered as a landscape some will have high fitness (peaks) and others low fitness (valleys).

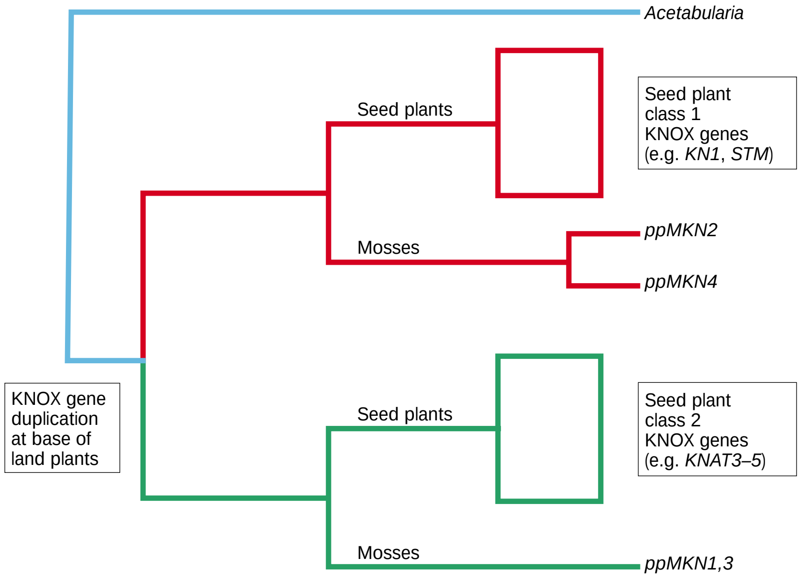


Figure 4 | **Duplication of KNOX genes.** This figure shows a highly reduced summary phylogeny of knotted-homeobox (KNOX) genes. Class 1 genes are shown in red and class 2 genes are in green. The ancient duplication of KNOX genes (present in mosses and probably in all land plants) might have been essential in the rise of the land plants. Indeed, an organized meristem (with definite structure and cell division regulated in different regions) is characteristic of land plants, and class 1 KNOX genes are known to be required for an organized meristem in angiosperms. The relative position of some key genes such as *SHOOTMERISTEMLESS* (*STM*), the *Physcomitrella patens KNOTTED* (ppMKN) family and the *KNOTTED-like Arabidopsis thaliana* (KNAT) family are shown. The divergence between the moss KNOX genes and the flowering plant genes occurred some 400 million years ago.

segments, excision of genes by unequal crossing over and the creation of pseudogenes all seem to have played their part. Viewed on a human timescale, genomes might seem to be static entities, as restructuring events are rare, but on a geological timescale they can be seen to be in a state of continual flux with genes being duplicated, destroyed and continually re-patterned on the chromosomes.

Gene duplication and evolutionary innovation. Some of these duplication events might have had great evolutionary significance. KNOX genes are homeodomain-containing plant genes that belong to the same family as *KNOTTED1* (*KN1*) of maize, a key developmental gene involved in meristem patterning. In flowering plants, KNOX genes can be divided into two classes[5,6] (FIG. 4). Class 1 includes those genes that are important in the control of the apical meristem, such as *STM*. The division into class 1 and class 2 (a division not so far found in algae, which nevertheless have knotted homeobox genes[48]) represents an ancient duplication event. Neil Ashton's laboratory[49] has recently shown that both classes occur in mosses, which puts the date of this duplication back to somewhere near the origin of the land flora (~430–475 million years ago**)**. It is interesting that class 1 KNOX genes are involved in the maintenance of the apical meristem — a characteristic of land plants — relative to the simpler construction of the algae. It would be very interesting to know whether class 1 and 2 KNOX genes are present in *Coleochaete*, a candidate extant sister group of the land plants.

A good example of how gene duplication can lead to evolutionary innovation is provided by the genes *CYC* and *DICH* in *Antirrhinum*[26] (FIG. 3). The transcription factors that these genes encode have overlapping functions in defining the difference between the large adaxial (upper) COROLLA lobes (petals) and the other corolla lobes of the flower. Single-knockout mutants of each still show differentiated upper petals, but the double mutant is striking: all petals resemble the lower ones of the wild type. *DICH* is expressed in the *CYC* domain and acts to accentuate the morphological disparity among petals produced by *CYC*. These upper petals are important in bumblebee pollination. Bumblebees are attracted by the complex dorsoventral asymmetry of the flower, and to collect pollen and nectar they must push apart the upper and lower petals using their legs and thorax. The elaborate upper petals (the result of both *DICH* and *CYC* action) are essential to the efficiency of the pollination system. In this case, specialization to bumblebee pollination seems to have provided selective pressure for the maintenance and divergence in function of two genes. An apparently analagous duplication of *CYC* homologues has occurred in the genus *Streptocarpus* in the closely related family Gesneriaceae[50], indicating that duplication in *CYC* genes might be a common occurrence.

Gene family diversity. The repeated history of gene duplication in plant evolution has led to some large multigene families. One very striking feature of the *Arabidopsis* genome is the numerous MYB transcription factors (401 according to one estimate[46]). MYB transcription factors are characterized by a conserved DNA-binding domain consisting of up to three 51–53 amino-acid imperfect repeats (R1, R2 and R3). However, most plant MYB genes contain only R2 and R3 (the so-called R1R2 MYBs). The numerous MYB genes in plants contrasts starkly with the human genome, for which a liberal prediction of the number of MYBs is 43. It will be interesting to learn what functional genomics will reveal about the role of these MYBs and it might shed some light on what is, at present, a problem in plant genomics: why evolution has favoured such a profligate number of MYB transcription factors. In maize, recent duplications of MYB transcription factors have been shown to have different spatio-temporal regulation, so MYB transcription factors might be particularly prone to assuming new expression patterns[51].

Does gene redundancy exist? The numerous gene families, coupled with the fact that knockouts of individual members often have no apparent phenotype, have led to the idea that many genes might be redundant. This has been suggested for the SHATTERPROOF (SHP) and SEPALLATA families of MADS-box transcription factors. SHATTERPROOF genes have a role in the ecologically important feature of pod shatter and therefore in seed dispersal. Both *SHP1* and *SHP2* must be knocked out to produce any detectable phenotype[52]. The double mutant has pods that do not dehisce (split open) and so the seeds are retained. *Arabidopsis* pods split at the margins of two constituent parts of the fruit: the REPLUM and

COROLLA
Whorl of floral leaves (petals) that surround the stamens. They are usually coloured and attract pollinators, and may be joined into a tube or ring, as in advanced eudicots.

Box 6 | Future goals in plant evo–devo

The present tendency to focus evolutionary studies on gene knock-out phenotypes is simplistic. Loss-of-function mutations with phenotypes that seem to reverse evolutionary innovations are certainly good 'candidate genes', but they might not be responsible for the evolutionary origin of that character. First, the laboratory-induced phenotype might be analogous rather than homologous (at the gene level) to the naturally occurring trait. Second, the gene might be necessary only because it is an upstream or downstream 'bolt-on' part of a regulatory network, only essential in a limited group of plants and so not the gene responsible for the original evolutionary innovation. To go beyond the siren call of candidate gene knockouts, higher-level comparative studies on whole gene networks are needed, which can only be carried out using the sophisticated genomics tools that are becoming available. To this end, there are several desirable goals for the future:

- Complete the genome sequence of at least one lower plant (probably a bryophyte), as a matter of priority. Although *Physcomitrella patens* has the best-known genetics and resources, it has a large genome size. A related species with a small genome would be preferable.
- Understand the acquisition of new gene function after gene duplication and how this leads to phenotypic innovation.
- Concentrate first on the big evolutionary innovations, that is the genetic basis of the principal features of plant evolution (see BOX 3), integrating genomics with the fossil record.
- Investigate the evolution of gene networks: particularly how the components of the complex network of genes that control meristem evolution have been assembled during land-plant evolution. Lower-plant genomics and microarray studies will be important in this regard.
- Develop new plant-targeted tools for reverse genetics and gene expression that can be used in diverse species.
- Understand how nucleotide sequence variation of *cis*-regulatory elements of transcription factors and other regions can lead to changes of expression pattern and evolutionary change.
- Integrate the environment, the phenotype and sequence variation into a coherent model of environmental selection at the nucleotide level (see BOX 1).

the valves. The SHP MADS-box genes regulate other genes that are expressed in the valve–replum boundary, ultimately causing them to peel apart. This is consistent with the usual role of MADS-box genes as important in specifying cell identity.

Data from mutant screens are problematic in assessing redundancy, as phenotypes might be subtle and only manifest themselves under stress conditions, rather than under screening conditions. The situation is complicated by the fact that 'gene redundancy' is a loosely used term, often as a proxy for 'gene similarity'. Evolutionary biology provides a stricter, and more appropriate, definition: a truly redundant gene is one whose loss has no fitness cost (that is, the loss of function is selectively neutral). A minute selection coefficient is sufficient to maintain an apparently redundant gene against the prevailing mutational load, which will otherwise turn truly redundant genes into pseudogenes. If selection is needed to maintain functional duplicate genes[53], true redundancy is inherently implausible except as a transient phenomenon immediately after duplication events.

Duplicate genes might be selected for because of a dosage effect (a larger amount of transcript) or because a gain-of-function mutation occurs before a loss-of-function mutation commences the transition to a pseudogene. The dosage effect is evident in the nuclear ribosomal DNA, which in *Arabidopsis* occurs as two tandem arrays of about 1,000 genes each, a testament to the importance of maintaining numerous ribosomes. Gain-of-function mutations are most likely to be in control regions; they can produce a slightly different spatio-temporal expression pattern, or change the environmental condition for induction. The six glutathione transferase genes in maize are expressed at different developmental stages. The existence of forms of the same enzyme with different temperature or pH optima might also be an advantage under fluctuating environmental conditions. Fixed heterozygosity due to gene duplication might promote fitness, in the same way that heterozygosity due to outbreeding does.

Single or low copy number genes. Despite the fact that extensively duplicated gene families are common, there are some genes that are consistently found as single- or double-copy only. An example is the gene *LEAFY* (*LFY*, known as *FLORICAULA* in *Antirrhinum*[54]). This gene is involved in the transition of vegetative to reproductive apices in diverse species[55] and it interacts with the MADS-box genes to pattern the flower. In both *Arabidopsis* and *Antirrhinum*, it is single copy (although in the crucifer *Jonopsidium* it is double copy[56]). In gymnosperms, it is double copy (*NEEDLY* and *Pinus radiata FLORICAULA*/*LEAFY*-like, *PRFLL*); this might be related to the fact that gymnosperms have separate male and female cones. There is some evidence that *PRFLL* might be specifically required for inducing male cones[57]. It has therefore been suggested that the single copy of *LFY* in angiosperms might be the result of gene loss and the adoption of female function by the 'male' copy of *LFY*[58], so producing a co-sexual structure, the forerunner of the angiosperm flower. This is the 'mostly male' theory of angiosperm evolution, a promising example of how potentially testable molecular hypotheses about principal events in plant evolution can be generated[58].

The absence of extensive gene duplication might simply be the default state when the maintenance of additional copies is not selected for. Duplicate copies of *LFY* might not be selectively advantageous as *LFY* acts as an integrator of several pathways. In *Arabidopsis* it is a target of both the giberellin pathway and the phytochrome signalling pathway (transducing information about the physiological state of the plant and the external seasonal environment, respectively), and integrates the two[59]. Duplicate copies might not be so precise in carrying out this integrative function.

The study of gene duplication and redundancy is a promising area in which evolutionary biology and genomics can constructively meet. More information is needed about how duplicated genes can diverge in function, thus allowing phenotypic innovation. Genes that are 'apparently redundant' provide a promising area for combining microevolution and genomics, and selection coefficients can be determined for knockouts that show no obvious phenotype. These selection experiments, however, require several generations and need to be done under conditions that are as ecologically appropriate as possible.

REPLUM
Septum dividing the ovary of crucifers (such as *Arabidopsis*) into chambers.

VALVE
Part of the ovary wall, at which splitting occurs to release the seeds.

MUTATIONAL LOAD
Negative fitness consequence of naturally occurring mutations.

Tasks for the future

The next few years will undoubtedly see a greater increase in the understanding of plant evolution than at any period in the history of the subject. Until now, plant genomics has been driven by agriculture, accounting for the strong angiosperm bias. Making best use of the new data in an evolutionary context will require a change of thinking and the generation of data that are suited to fundamental questions rather than applied ones. Evolutionary thinking has much to offer agriculture and the biotechnology industry. There are some 250,000 species of angiosperms growing under a wide range of environmental conditions. The diversity of protein, allelic and trait diversity is very high, but only fundamental understanding in a comparative framework will allow this biodiversity to be used. To this end, some tasks should receive especially high priority (BOX 6).

Particularly promising new techniques now arising include more effective specific gene silencing using double-stranded RNA interference (for example, by using hairpin intron-spliced constructs[60]), and homologous recombination techniques using the moss *Physcomitrella patens*[61–64]. The development of easy methods for transforming new species will also help[65], and large activation tagging programmes are important in that they generate dominant phenotypes that might be more significant in an evo–devo context than the knockout phenotypes seen in conventional mutagenesis screens[66]. The combination of new genomic and technical tools with evolutionary and ecological thinking is sure to yield rich dividends.

Links

DATABASE LINKS *FIS2* | *PHAN* | *RS2* | *AS1* | *STM* | *PHB* | *CYC* | *DICH* | *TB1* | *FRI* | *FLC* | *KN1* | *LFY*

FURTHER INFORMATION Theodosius Dobzhansky | Richard Goldschmidt | Alfred Russell Wallace | Ronald A. Fisher | George Ledyard Stebbins

LOWER PLANT MODELS Moss resources | C-Fern resources

GENOMICS PROJECTS The *Arabidopsis* Genome Initiative | Rice Genome Research Program | Monsanto rice

FUNCTIONAL GENOMICS AND EST PROJECTS Multinational Coordinated *Arabidopsis* 2010 Project | Maize Genome Database | Soybean Functional Genomics Program | Potato functional genomics

BIODIVERSITY Green plant phylogeny | Treebase | Tree phylogeny

SOFTWARE DnaSP

1. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
2. Bowe, L. M., Coat, G. & dePamphilis, C. W. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl Acad. Sci. USA* **97**, 4092–4097 (2000).
3. Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402–404 (1999).
One of several recent papers that have revolutionized our understanding of angiosperm phylogeny, and that have allowed ideas of gene evolution and functional evolution to be tested in an accurate phylogenetic context.
4. Becker, A. *et al.* MADS-box gene diversity in seed plants 300 million years ago. *Mol. Biol. Evol.* **17**, 1425–1434 (2000).
5. Bharathan, G., Janssen, B. J., Kellogg, E. A. & Sinha, N. Phylogenetic relationships and evolution of the KNOTTED class of plant homeodomain proteins. *Mol. Biol. Evol.* **16**, 553–563 (1999).
6. Reiser, L., Sanchez-Baracaldo, P. & Hake, S. Knots in the family tree: evolutionary relationships and functions of knox homeobox genes. *Plant Mol. Biol.* **42**, 151–166 (2000).
7. Graham, L. K. E. & Wilcox, L. W. The origin of alternation of generations in land plants: a focus on matrotrophy and hexose transport. *Phil. Trans. R. Soc. Lond.* B **55**, 757–766 (2000).
8. Spillane, C. *et al.* Interaction of the *Arabidopsis* polycomb group proteins FIE and MEA mediates their common phenotypes. *Curr. Biol.* **10**, 1535–1538 (2000).
9. Yadegari, R. *et al.* Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell* **12**, 2367–2381 (2000).
10. Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J. & Chaudhury, A. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc. Natl Acad. Sci. USA* **97**, 10637–10642 (2000).
11. Waites, R. & Hudson, A. Phantastica — a gene required for dorsoventrality of leaves in *Antirrhinum majus*. *Development* **121**, 2143–2154 (1995).
12. Waites, R., Selvadurai, H. R., Oliver, I. R. & Hudson, A. The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in *Antirrhinum*. *Cell* **93**, 779–789 (1998).
13. Timmermans, M. C. *et al.* ROUGH SHEATH2: a Myb protein that represses knox homeobox genes in maize lateral organ primordia. *Science* **284**, 151–153 (1999).
14. Tsiantis, M., Schneeberger, R., Golz, J. F., Freeling, M. & Langdale, J. A. The maize rough sheath2 gene and leaf development programs in monocot and dicot plants. *Science* **284**, 154–156 (1999).
15. Schneeberger, R., Tsiantis, M., Freeling, M. & Langdale, J. A. The rough sheath2 gene negatively regulates homeobox gene expression during maize leaf development. *Development* **125**, 2857–2865 (1998).
16. Byrne, M. E. *et al.* Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature* **408**, 967–971 (2000).
17. Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K. & Poethig, R. S. *KANADI* regulates organ polarity in *Arabidopsis*. *Nature* **411**, 706–709 (2001).
18. McConnell, J. R. *et al.* Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. *Nature* **411**, 709–713 (2001).
Shows that the gene *PHABULOSA* (*PHB*) is important in specifying adaxial identity — a key innovation in leaf evolution. A sterol/lipid-binding domain in the PHB protein acts as a receptor for an unknown adaxializing signal, which leads to the protein being actively transcribed only in the adaxial domain.
19. Bateman, R. M. & Dimichele, W. A. Heterospory — the most iterative key innovation in the evolutionary history of the plant kingdom. *Biol. Rev. Cambr. Phil. Soc.* **69**, 345–417 (1994).
20. Kramer, E. M. & Irish, V. F. Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. *Int. J. Plant Sci.* **161**, S29–S40 (2000).
21. Kramer, E. M. & Irish, V. F. Evolution of genetic mechanisms controlling petal development. *Nature* **399**, 144–148 (1999).
Examines the conservation of the ABC model of floral development from the model organisms *Arabidopsis* and *Antirrhinum* to the basal eudicots. The authors find that some aspects of the ABC model are conserved, whereas others are not.
22. Ambrose, B. A. *et al.* Molecular and genetic analyses of the silky1 gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell* **5**, 569–579 (2000).
23. Yu, D. *et al.* Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). *Plant J.* **17**, 51–62 (1999).
24. Xue, Y. B., Carpenter, R., Dickinson, H. G. & Coen, E. S. Origin of allelic diversity in *Antirrhinum* S locus RNases. *Plant Cell* **10**, 805–814 (1996).
25. Cronk, Q. C. B. & Möller, M. Genetics of floral symmetry revealed. *Trends Ecol. Evol.* **12**, 85–86 (1997).
26. Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J. & Coen, E. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* **99**, 367–376 (1999).
Shows how the closely related genes *CYCLOIDEA* (*CYC*) and *DICHOTOMA* together control the unique features of the adaxial side of the snapdragon flower. *CYC* knockout mutants abaxialize the flower, whereas a gain-of-function *CYC* mutation, BACKPETALS, which affects the *cis*-regulation of the gene, results in an adaxialized flower.
27. Luo, D., Carpenter, R., Vincent, C., Copsey, L. & Coen, E. Origin of floral asymmetry in *Antirrhinum*. *Nature* **383**, 794–799 (1996).
28. Cubas, P., Lauter, N., Doebley, J. & Coen, E. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* **18**, 215–222 (1999).
29. Gaudin, V. *et al.* The expression of D-cyclin genes defines distinct developmental zones in snapdragon apical meristems and is locally regulated by the cycloidea gene. *Plant Physiol.* **122**, 1137–1148 (2000).
30. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–161 (1999).
Shows that the phenotype of the classic *peloria* mutation, studied by Linnaeus, is due to a loss-of-function of the *Linaria CYCLOIDEA* homologue and subsequent abaxialization of the flower. The mutation is not attributable to a sequence change, but to a heritable change in the methylation status of the gene.
31. Cubas, P., Coen, E. & Martínez–Zapater, J. M. Ancient asymmetries in the evolution of flowers. *Curr. Biol.* **11**, 1050–1052 (2001).
32. Stern, D. L. A role of Ultrabithorax in morphological differences between *Drosophila* species. *Nature* **396**, 463–466 (1999).
33. Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of ovo/shaven-baby. *Proc. Natl Acad. Sci. USA* **97**, 4530–4534 (2000).
34. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
Shows that, in the evolution of maize from its wild

ancestor (teosinte) by selection under cultivation, very strong selection has been applied to the *cis*-regulatory regions of the *TEOSINTE BRANCHED1* gene.

35. Carroll, S. B. Developmental regulatory mechanisms in the evolution of insect diversity. *Development* **21**, S7–S223 (1994).
36. Doebley, J. & Lukens, L. Transcriptional regulation and the evolution of plant form. *Plant Cell* **10**, 1075–1082 (1998).
37. Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
38. Bateman, R. M. & DiMichele, W. A. in *Shape and Form in Plants and Fungi* (eds Ingram, D. S. & Hudson, A.) 63–102 (Academic, London, 1994).
39. Bateman, R. M. *et al.* Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. *Annu. Rev. Ecol. Syst.* **29**, 263–292 (1998).
40. DiMichele, W. A. & Bateman, R. M. Plant paleoecology and evolutionary inference: two examples from the Paleozoic. *Rev. Palaeobot. Palynol.* **90**, 223–247 (1996).
41. Tucker, S. C. The ontogenetic basis for missing petals in *Crudia* (Leguminosae: Caesalpinioideae: Detarieae). *Int. J. Plant Sci.* **162**, 83–89 (2001).
42. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347 (2000).
A pioneer paper that shows the molecular basis of naturally occurring, putatively adaptive variation — in this case, flowering time. The authors show how populations with a different flowering time result from a deletion at the *FRIGIDA* locus.
43. Gaut, B. & Doebley, J. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl Acad. Sci. USA* **94**, 6809–6814 (1997).
44. Gaut, B. S. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome Res.* **11**, 55–66 (2001).
45. Vision, T. J., Brown, D. G. & Tanksley, S. D. The origins of genomic duplications in *Arabidopsis*. *Science* **290**, 2114–2117 (2000).
46. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
47. Wolfe, K. H. Yesterday's polyploids and the mystery of diploidization. *Nature Rev. Genet.* **2**, 333–341 (2001).
48. Serikawa, K. A. & Mandoli, D. F. *Aaknox1*, a kn1-like homeobox gene in *Acetabularia acetabulum*, undergoes developmentally regulated subcellular localization. *Plant Mol. Biol.* **41**, 785–793 (1999).
49. Champagne, C. E. M. & Ashton, N. W. Ancestry of plant KNOX genes revealed by bryophyte (*Physcomitrella patens*) homologs. *New Phytol.* **150**, 23–36 (2001).
Shows that the fundamental duplication of KNOX (knotted homeobox) genes into two classes is characteristic of all land plants but not of an alga, indicating that this duplication might have been a key step in the evolution of land flora.
50. Citerne, H. L., Möller, M. & Cronk, Q. C. B. Diversity of cycloidea-like genes in Gesneriaceae in relation to floral symmetry. *Ann. Bot.* **86**, 167–176 (2000).
51. Zhang, P. F., Chopra, S. & Peterson, T. A. Segmental gene duplication generated differentially expressed myb-homologous genes in maize. *Plant Cell* **12**, 2311–2322 (2000).
52. Liljegren, S. J. *et al.* SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766–770 (2000).
53. Shimeld, S. M. Gene function, gene networks and the fate of duplicated genes. *Semin. Cell. Dev. Biol.* **10**, 549–553 (1999).
54. Coen, E. S. *et al.* Floricaula — a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322 (1990).
55. Rottmann, W. H. *et al.* Diverse effects of overexpression of LEAFY and PTLF, a poplar (*Populus*) homolog of LEAFY/FLORICAULA, in transgenic poplar and *Arabidopsis*. *Plant J.* **22**, 235–245 (2000).
56. Shu, G. P. *et al.* LEAFY and the evolution of rosette flowering in violet cress (*Jonopsidium acaule*, Brassicaceae). *Am. J. Bot.* **87**, 634–641 (2000).
57. Mouradov, A. *et al.* Molecular control of early cone development in *Pinus radiata. Protoplasma* **208**, 3–12 (1999).
58. Frohlich, M. W. & Parker, D. S. The mostly male theory of flower evolutionary origins: from genes to fossils. *Syst. Bot*. **25**, 155–170 (2000).
Significant as the first paper to put forward a potentially testable molecular hypothesis for the origin of the hermaphrodite angiosperm flower from unisexual gymnosperm reproductive structures. This might have occurred by integrating male and female cone development genes under the control of a single angiosperm gene called *LEAFY*.
59. Blazquez, M. A. & Weigel, D. Integration of floral inductive signals in *Arabidopsis*. *Nature* **404**, 889–892 (2000).
Shows how the giberellin pathway and the photoperiod pathway (both of which can induce flowering) integrate on separate *cis*-regulatory elements of the same downstream target gene: *LEAFY*, a gene that promotes the transition from vegetative to floral meristems.
60. Smith, N. A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* **407**, 319–320 (2000).
61. Schaefer, D. G. Gene targeting in *Physcomitrella patens*. *Curr. Opin. Plant Biol.* **4**, 143–150 (2001).
62. Wood, A. J., Oliver, M. J. & Cove, D. J. Bryophytes as model systems. *Bryologist* **103**, 128–133 (2000).
63. Reski, R. Molecular genetics of *Physcomitrella*. *Planta* **208**, 301–309 (1999).
64. Hofmann, A. H. *et al.* A specific member of the Cab multigene family can be efficiently targeted and disrupted in the moss *Physcomitrella patens*. *Mol. Gen. Genet.* **261**, 92–99 (1999).
65. Trieu, A. T. *et al.* Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant J.* **22**, 531–541 (2000).
66. Weigel, D. *et al.* Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013 (2000).
67. Purugganan, M. D. The molecular population genetics of regulatory genes. *Mol. Ecol.* **9**, 1451–1461 (2000).
68. Purugganan, M. D., Boyles, A. L. & Suddith, J. I. Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics* **155**, 855–862 (2000).
69. Purugganan, M. & Suddith, J. Molecular population genetics of the *Arabidopsis* cauliflower regulatory gene; non-neutral evolution and naturally occurring variation in floral homeotic function. *Proc. Natl Acad. Sci. USA* **95**, 8130–8134 (1998).
70. Lukens, L. & Doebley, J. Molecular evolution of the teosinte branched gene among maize and related grasses. *Mol. Biol. Evol.* **18**, 627–638 (2001).
71. Yang, Z. & Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**, 409–418 (1998).
72. Sutton, K. & Wilkinson, M. Rapid evolution of a homeodomain: evidence for positive selection. *J. Mol. Evol.* **45**, 579–588 (1997).
73. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and non-symonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
74. Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
75. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).
76. Chuzhanova, N. A., Krawczak, M., Nemytikova, L. A., Gusev, V. D. & Cooper, D. N. Promoter shuffling has occurred during the evolution of the vertebrate growth hormone gene. *Gene* **254**, 1–2 (2000).
77. Lemieux, C., Otis, C. & Turmel, M. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* **403**, 649–652 (2000).
78. Qiu, Y. L. & Lee, J. Transition to a land flora: a molecular phylogenetic perspective. *J. Phycol.* **36**, 799–802 (2000).
79. Qiu, Y. L., Cho, Y., Cox, C. J. & Palmer, J. D. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394**, 671–674 (1998).
80. Nickrent, D. L., Parkinson, C. L., Palmer, J. D. & Duff, R. J. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* **17**, 1885–1895 (2000).
81. Pryer, K. M. *et al.* Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**, 618–622 (2001).
82. Qiu, Y. L. *et al.* The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
83. Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402–404 (1999).
84. Barkman, T. J. *et al.* Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl Acad. Sci. USA* **97**, 13166–13171 (2000).

Acknowledgements

I thank R. Bateman and two anonymous referees for constructive comments on the manuscript.