Morphological and molecular phylogenetic context of the angiosperms: contrasting the ‘top-down’ and ‘bottom-up’ approaches used to infer the likely characteristics of the first flowers

Richard M. Bateman1,*, Jason Hilton2 and Paula J. Rudall1
1 Jodrell Laboratory, Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AB, UK
2 School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received 4 May 2006; Accepted 13 July 2006

Abstract

Recent attempts to address the long-debated ‘origin’ of the angiosperms depend on a phylogenetic framework derived from a matrix of taxa versus characters; most assume that empirical rigour is proportional to the size of the matrix. Sequence-based genotypic approaches increase the number of characters (nucleotides and indels) in the matrix but are confined to the highly restricted spectrum of extant species, whereas morphology-based approaches increase the number of phylogenetically informative taxa (including fossils) at the expense of accessing only a restricted spectrum of phenotypic characters. The two approaches are currently delivering strongly contrasting hypotheses of relationship. Most molecular studies indicate that all extant gymnosperms form a natural group, suggesting surprisingly early divergence of the lineage that led to angiosperms, whereas morphology-only phylogenies indicate that a succession of (mostly extinct) gymnosperms preceded a later angiosperm origin. Causes of this conflict include: (i) the vast phenotypic and genotypic lacuna, largely reflecting pre-Cenozoic extinctions, that separates early-divergent living angiosperms from their closest relatives among the living gymnosperms; (ii) profound uncertainty regarding which (a) extant and (b) extinct angiosperms are most closely related to gymnosperms; and (iii) profound uncertainty regarding which (a) extant and (b) extinct gymnosperms are most closely related to angiosperms, and thus best serve as ‘outgroups’ dictating the perceived evolutionary polarity of character transitions among the early-divergent angiosperms. These factors still permit a remarkable range of contrasting, yet credible, hypotheses regarding the order of acquisition of the many phenotypic characters, reproductive and vegetative, that distinguish ‘classic’ angiospermy from ‘classic’ gymnospermy. The flower remains ill-defined and its mode (or modes) of origin remains hotly disputed; some definitions and hypotheses of evolutionary relationships preclude a role for the flower in delimiting the angiosperms. We advocate maintenance of parallel, reciprocally illuminating programmes of morphological and molecular phylogeny reconstruction, respectively supported by homology testing through additional taxa (especially fossils) and evolutionary-developmental genetic studies that explore genes potentially responsible for major phenotypic transitions.

Key words: Angiosperm, character optimization, congruence, development, evolutionary-developmental genetics, flower, fossil, gene duplication, gymnosperm, morphology, ontogeny, outgroup, phylogeny, pteridophyte, taxon sampling, tree rooting.

Introduction

‘An early hope was that the relationships indicated among species by DNA were more likely to be correct than those based on morphology; this now seems naïve.’ (Judd et al., 1999, p. 99)
The primary motivating force for preparing this book was the dramatic change in our understanding of angiosperm phylogeny during the past 10 years. Many long-standing [morphological] views of deep-level relationships were altered suddenly and substantively as a direct result of molecular analyses.’ (Solis et al., 2005, p. ix)

‘Occam’s razor? But that’s for circumcision, surely?’ (Tom Sharpe, 1995; Grantchester Grind, London: Macmillan, p. 82)

In this review we address one of the most popular discussion topics in evolutionary biology: the origin of the flower, and, by implication, the origin of the flowering plants (note that we pay much less attention to the better understood and the underlying concept is familiar to even the most lackadaisical student of botany. However, comparison of the definitions assembled in Table 1 reveals little unanimity. Four ostensibly distinct elements can be ext-

<table>
<thead>
<tr>
<th>Table 1. Representative spectrum of definitions of a flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘...a growth comprising the reproductive organs of seed plants’ (Chambers Dictionary, 1998, p. 617)</td>
</tr>
<tr>
<td>‘...an axis on which the rest of the floral organs are borne’ (Fahn, 1974, p. 425)</td>
</tr>
<tr>
<td>‘...[a] specialized reproductive shoot, consisting of an axis (receptacle) on which are inserted four different sorts of organs [sepal, petals, stamens, carpels]’ (Penguin Dictionary of Biology, 1971, p. 102)</td>
</tr>
<tr>
<td>‘...an axis (or receptacle) bearing, in its complete form, four zones of appendages that are considered the homologs of leaves’ (Bierhorst, 1971, p. 511)</td>
</tr>
<tr>
<td>‘...a highly modified shoot bearing specialized appendages (modified leaves)’ (Judd et al., 1999, p. 53)</td>
</tr>
<tr>
<td>‘...an amphi sporangiate strobilus of determinate growth and with an involucre of modified bracts’ (Arber and Parkin, 1907)</td>
</tr>
<tr>
<td>‘...the reproductive structure of Anthophyta, derived evolutionarily from a leafy shoot in which leaves have become modified into petals, sepals and into the carpels and stamens in which the gametes are formed’ (Dictionary of Biological Terms, 1995; Harlow: Longman, p. 206)</td>
</tr>
<tr>
<td>‘...a shoot beset with sporophylls: ...leaves bearing sporangia’ (Goebel, 1905, p. 469)</td>
</tr>
<tr>
<td>‘...a section of a shoot, or branch resembling a short shoot, which bears leaf organs which serve for sexual reproduction and which are transformed accordingly’ (Weberling, 1989, p. 2)</td>
</tr>
<tr>
<td>‘...a particular type of determinate sporogenous shoot ...one of its most definitive organs is the carpel: ...a megasporophyll [that is] morphologically distinct because the ovules (i.e. megasporangia) are usually enclosed within a hollow basal portion designated as the ovary’ (Foster and Gifford, 1971, pp. 593–594)</td>
</tr>
<tr>
<td>‘...a compressed, bisexual reproductive axis composed of stamens, carpels and a sterile perianth’ (Baum and Hileman, 2006)</td>
</tr>
<tr>
<td>‘...a bisexual structure with more or less recognizable carpels; that is, laminar structures largely enclosing the ovules on their upper [adaxial] surface, and with more or less recognizable stamens’ (Frohlich, 2006)</td>
</tr>
<tr>
<td>‘...a determinate axis terminating in megasporangia that are surrounded by microsporangia and are collectively subtended by at least one sterile laminar organ’ (present study)</td>
</tr>
</tbody>
</table>

What is a flower?

Given the vast tracts of published text devoted to the origin of the flower (and of the flowering plants), this is one question that could reasonably be assumed to have been unequivocally answered. Indeed, most glossaries included in textbooks and reviews omit the term ‘flower’, presumably on the assumption that the term is universally understood and the underlying concept is familiar to even the most lackadaisical student of botany. However, comparison of the definitions assembled in Table 1 reveals little unanimity. Four ostensibly distinct elements can be ext-
and reliable linear sequence along the axis towards its apex: sepals, petals, stamens, and carpels. A few definitions emphasized the function of stamens and carpels for generating gametes (male and female, respectively), and Foster and Gifford (1971) and Frohlich (2006) further identified the carpel as being the most diagnostic feature of a flower. Foster and Gifford also noted that definitions of a flower omitting reference to these structures, such as that of Goebel (1905), implicitly encompass not only all angiosperms but also all gymnosperms and even some of the more derived, reproductively complex, pteridophyte groups (thereby rendering less bizarre the aforementioned Chambers Dictionary definition: Table 1).

Crucial to the majority of definitions is an almost universally accepted interpretation of the flower as a determinate reproductive shoot with a defined number of floral organs, typically (but not universally) arranged in at least four distinct whorls: sepals, petals, stamens, and carpels. Specifically, the flower is homologized with an axis bearing sporophylls (i.e. evolutionarily modified sporangium-bearing leaves or, perhaps more conservatively, leaf-like structures). Implicit in these definitions is the dynamic concept of a transition from an earlier, profoundly different (and thus recognizable) ancestral form of reproductive organ. The identity of the ancestral group is almost universally accepted, namely the gymnosperms; an origin among the (by definition) seed-less pteridophytes requires a set of morphological transitions that seemingly is too radical for any modern botanist to seriously contemplate.

Thus, we can draw on several categories of information in order to define a flower. And, having defined a flower, surely we have by default defined a flowering plant—an angiosperm. Why then have the more informed among the land-plant morphologists appended firm cautionary notes to their preferred definitions? For example, according to Foster and Gifford (1971, p. 593), ‘As angiosperms are commonly designated the flowering plants, it might be assumed that there is rather general agreement about the scientific concept of a flower. Unfortunately, this is not the case, and the literature on floral organography, ontogeny, and structure displays widely divergent viewpoints of the fundamental nature of the flower as well as on the interpretation of its component organs (sepals, petals, stamens, and carpels). One of the basic difficulties lies in our complete [sic] ignorance of the evolutionary history of the flower...; it becomes largely a matter of conjecture whether it is justifiable to draw comparisons between modern angiospermous flowers and the spore-producing structures of other tracheophytes [vascular land-plants]. If such comparisons are attempted, it is quite possible to reach either a very broad or a very restricted concept or definition of a flower.’

Bierhorst (1971, p. 511) preceded his definition by stating that a flower is ‘a structure that, because of its various degrees of completeness, cannot adequately be defined in words’. This, and many other similar statements in the literature, refer primarily to the considerable floral variation revealed by surveys of extant angiosperms. Some species lack sepals, petals, or both. For example, in one clade (Saururaceae plus Piperaceae) of the magnoliid order Piperales, flowers are entirely perianth-less, a feature that is closely correlated (at least, in this group) with possession of indeterminate inflorescences (Remizova et al., 2005). Many other species show morphological graduation between sepals, petals, and/or stamens. Yet others contravene the requirement for bisexuality by bearing functional stamens and carpels on separate flowers, either on the same individual (monoecy) or on separate individuals (dioecy); organs of the opposite gender have either been rendered sterile or completely suppressed. For example, the much-discussed ‘primitive’ extant angiosperm Amborella (Fig. 1A, B) is commonly (though not reliably) dioecious, bearing either exclusively male flowers (lacking even sterile carpels) or female flowers that occasionally bear staminodes, sometimes with a developmental transition between stamens and carpels (Buzgo et al., 2004). Many other early-divergent extant angiosperms even lack carpel closure by tissue fusion (Endress and Igersheim, 2000a, b; Endress, 2001a) or double fertilization (Williams and Friedman, 2002). The columellar infratectum in the microspore wall of angiosperms, functionally linked to chemical recognition systems on the stigmatic surface, is also unreliably present in basally divergent extant angiosperms (Sampson, 2000; Doyle, 2006; Frohlich, 2006).

Thus, several supposedly definitive features of the flower are frequently absent from angiosperms, especially early-divergent angiosperms: these include hermaphroditism, fully closed carpels, and a distinctly whorled arrangement. Admittedly, the presence on the ovule of a second (outer) integument, and the remarkably conservative structure of the stamen (two thecae joined by a connective, each consisting of two embedded microsporangia), are more consistent features of the angiosperm flower, though they too have potential homologues in some gymnosperm groups (Doyle, 2006).

Furthermore, the flower-subtending bract, normally regarded as extra-floral, can strongly influence, or even be considered part of, the flower (e.g. Remizova et al., 2005; Buzgo et al., 2006). Bringing a palaeobotanical perspective to bear, Stewart and Rothwell (1993, p. 440) coined the phrase ‘accessory reproductive structures’ to encompass sterile organs associated with reproductive functions—not only perianth members (tepals, or petals plus sepals) but also another (less well-researched) leaf-like organ, the flower-subtending bract. Bracts (and bracteoles) share with perianth members the characteristics of exhibiting many leaf-like features but being positionally fixed with respect to either a flower or an inflorescence. In some angiosperms, the bract has become intimately integrated
into the flower; for example, *Amborella* (Fig. 1A, B) and several other early-divergent angiosperms (e.g. *Austrobaileyia, Trimenia*) exhibit a morphological continuum between bracts and perianth (Endress, 2001b; Buzgo et al., 2004). In others (e.g. Araceae, Cornaceae, Saururaceae) the inflorescence bracts are conspicuously petaloid and hence perform at least one of the functions (pollinator attraction) that are more typically performed by petals. For example, in two genera (*Houttuynia* and *Anemopsis*) of the perianth-less family Saururaceae, the inflorescence bracts form a pseudocorolla, and the entire inflorescence has a flower-like appearance (Tucker, 1981).

More significantly, many of the gymnospermous groups, both extant and extinct, which are by definition considered to lack a differentiated perianth, unequivocally possess bract-like organs. Arber and Parkin (1907) coined the terms 'anthostrobilus' for the modern angiosperm flower, and 'pro-anthostrobilus' for the type of cone manifested not only by the Mesozoic bennettite *Williamsoniella* (Fig. 1D) but also by a hypothetical group of extinct stem-group angiosperms that they termed 'Hemiangiospermae'. They noted that the bennettite cone possessed a series of sterile leaf-like organs that they interpreted as an 'undifferentiated primitive perianth'. They also insightfully defined a major subset of (derived) angiosperms as possessing 'a euanthostrobilus, of which the distinctive features are the presence of the special type of microsporophyll termed a stamen, and of closed carpels' (Arber and Parkin, 1907, p. 75). Thus, if flowers are at least partly defined not by possession of sepals and/or petals but by possession of the broader category of leaf-like accessory reproductive structures, the flower is no longer seen as a unique (and defining) attribute

![Fig. 1. (A) Female and (B) male flowers of the putative basally divergent extant angiosperm *Amborella trichopoda*. (C) Male (above) and female (below) reproductive structures of the much-discussed Jurassic fossil gymnosperm *Caytonia* (Caytoniales) and (D) hermaphrodite reproductive structure of the Jurassic fossil gymnosperm *Williamsoniella* (Bennettitales); both are candidate sister-groups to the angiosperms. (C and D reproduced, with permission, from figs 1.8 and 1.11, respectively, of Solits et al., 2005, and reprinted from Crane, 1985.) Scale bar = 100 μm.](https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/er086)

Downloaded from https://academic.oup.com/jxb/article/57/13/3471/477539 on 05 July 2024
of the angiosperms, and when using the term ‘flowering plants’ certain gymnospermous groups, especially extinct taxa such as Bennettitales, should strictly be included (cf. Crane, 1988). This conclusion is implicit in our own (inevitably imperfect) definition of a flower: a determinate axis bearing megasporangia that are surrounded by microsporophyll and are collectively subtended by at least one sterile laminar organ. In formulating this definition, the orientation of our discussion has, in practice, switched from top-down to bottom-up. We note that, although we do not necessarily exclude multiple origins of flower-like structures (see below), our definition has converged on that of Arber and Parkin (1907, p. 75), who characterized the anhstroebulus as ‘a special form of amphisporangiate cone, distinguished by the peculiar juxtaposition of the mega- and microsporangia, and by possessing a well-marked perianth’. We also note that Arber and Parkin primarily (and almost uniquely) employed a bottom-up approach in what became a benchmark study in floral evolution that remained influential throughout the ensuing century.

**Which are the (other) benchmark studies in floral evolution?**

Certain conceptual thresholds can readily be identified in the study of flowers and flowering plants. In his benchmark classification of flowering plants, Linnaeus (Linné, 1735) emphasized the significance of the number and arrangement of stamens and carpels. The possible equivalence of sepals, petals, and stamen filaments (though not explicitly the carpels) to modified leaves was greatly elaborated by Goethe (1790) in an essentialist essay partly stimulated by studying plant teratologies. However, Goethe’s ideas about simple equivalence, summarized by the famous phrase ‘all is leaf’, had no phylogenetic implications (Lönig, 1994). As noted by adherents of Zimmerman’s telome theory (Zimmerman, 1930, 1938), such as Wilson (1937), spore-bearing organs (i.e. sporangia) evolved before leaves in early land plants. By contrast, Darwin’s (1859) development of a credible evolutionary mechanism to explain (albeit in a uniformly gradualistic manner) such radical morphological transitions was followed by the harnessing of Mendelian genetics to address the control of such transitions by authors such as De Vries (1906) and later Wardlaw (1965).

These new data sources informed some increasingly sharp exchanges between proponents of the two main sets of theories competing to explain the origin of the flower. The more common euathial theory that was pioneered by Goethe (and is implicit in most of the definitions of a flower collated in Table 1) postulates derivation from an unbranched, uniaxial structure, and hence interprets the flower as a condensed sporophyll-bearing single axis with proximal microsporophylls (stamens) and distal megasporophylls (carpels) (Arber and Parkin, 1907; Arber, 1937). By contrast, the more diverse pseudanthial theories all perceive the flower as having condensed from a multiaxial structure (Wettstein, 1907; Melville, 1960; Eames, 1961; Meeuse, 1975, 1987, Stuessy, 2004). Thus, the difference between the two conflicting hypotheses relates more to the nature of individual organs than to the flower as a whole. Several authors have postulated secondary derivation of flower-like structures from inflorescences (i.e. a secondary pseudanthial origin in certain phylogenetic groups), based on ontogenetic evidence. These studies relate to both gymnosperms (Gnetales: Mundry and Stützel, 2004) and angiosperms (e.g. alismatid monocots: many authors, reviewed by Sokoloff et al., 2006). Multiple origins of flower-like structures, both within angiosperms and other seed plants, are implicit in these pseudanthial hypotheses.

Studies of flowers prompted by the various insights outlined above were, to varying degrees, comparative, and in some cases hypothesis-testing. However, they often (i) focused on limited suites of morphological characters considered a priori to be of particular importance, (ii) failed to determine whether the characters of interest were ever specified by a single genome (i.e. were observed in a single individual), and/or (iii) unjustifiably equated extreme diversity of form with increased likelihood of multiple origins of the feature in question. More importantly, they lacked a unifying conceptual framework. This crucial prerequisite has been provided, first by phylogeny reconstruction and then by evolutionary-developmental genetics, epitomized by the ABC model of floral whorl control (e.g. Coen and Meyerowitz, 1991; Irish and Kramer, 1998; Kramer and Irish, 1999; Lawton-Rauh et al., 2000; Cronk et al., 2002; Theissen et al., 2002).

**Which are the benchmark studies in reconstructing seed-plant phylogeny?**

The now ubiquitous tree motif routinely used to represent the supposed sequence of divergence of evolutionary lineages was famously employed by Darwin (1859) and elaborated by authors such as Haeckel (1894). However, only with the early cladistic works of entomologists Hennig (1966) and Brundin (1972) were we provided with the rigorous conceptual framework needed to generate such trees with a large degree of objectivity from matrices of coded taxa, each explicitly scored for competing states of a broad suite of characters. Central to this approach is the concept of the congruence test of homology (e.g. Patterson, 1988). The preferred phylogeny for the scored taxa is the one requiring fewest transitions between character-states. We can then observe in the resulting tree(s) which character-states delimit which taxonomic groups—in other words, which prior statements of homology between species have been upheld in the most-parsimonious tree(s). The
particulate nature of the numerically scored character states is preserved in the resulting trees, thereby simplifying various kinds of post hoc character analysis.

The cladistic approach was first used to compare the major groups of seed-plants by Hill and Crane (1982). With the assistance of rapidly improving computer hardware and software, it spawned an increasing number of morphological phylogenies based on parsimony analyses during the 1980s and early 1990s. If we consider the development of taxonomically broad morphological matrices for seed-plants, four main lineages can be recognized, nucleating around PR Crane (e.g. Hill and Crane, 1982; Crane, 1985, 1988), JA Doyle (e.g. Doyle and Donoghue, 1986, 1987, 1992; Doyle, 1996, 1998a, b, 2006; Hilton and Bateman, 2006), GW Rothwell (e.g. Rothwell and Serbet, 1994; Rothwell and Nixon, 2006), and DW Stevenson (e.g. Loconte and Stevenson, 1990, 1991; Nixon et al., 1994). These studies placed as closest relatives of angiosperms, in various paraphyletic combinations, several groups (most of them wholly extinct) that possess reproductive organs showing some flower-like properties, notably Gnetales, Bennettitales (Fig. 1D), Pentoxylon, Caytonia (Fig. 1C), and glossopterids. Such arrangements soon became known collectively as the anthophyte hypothesis, implying that the flower evolved only once and, hence, that all flowers are fundamentally homologous.

By 1990, these morphological analyses were competing with (and soon largely superseded by) phylogenetic studies employing as characters the sequences of nucleotides (and insertion–deletion mutations) in specific regions of the three plant genomes: nuclear-chromosomal, plastid, and mitochondrial. As the number of readily sequenced regions (and hence the number of usable characters) increased exponentially, driven by advances in sequencing technology, maximum parsimony was supplemented with more mathematically complex methods of generating trees, notably maximum likelihood and, latterly, Bayesian approaches (e.g. Page and Holmes, 1998). The current species richness and ecological pre-eminence of angiosperms have caused them to be preferentially sampled for phylogenetic studies, both morphological and molecular, consistently place the clade containing all extant angiosperm lineages (the ‘crown-group’ angiosperms: Fig. 3) on a long branch with respect to the remaining seed plants. ‘Missing links’ (i.e. stem-group) angiosperms have been postulated from the fossil record but have not hitherto obtained universal acceptance (e.g. Archaeafructus, discussed below). Furthermore, relationships among extant gymnosperms, and hence homologies among their reproductive structures, remain contentious. These issues can only be partly circumvented by adopting a top-down approach but are fundamental to any bottom-up approach.

Which kinds of phylogeny are of greatest value?

Why do tensions remain within the phylogenetic community regarding molecular versus morphological characters and extant versus extinct taxa? Over the past two decades, molecular phylogenies have in practice largely superseded morphological phylogenies. This shift of emphasis has had profound consequences, yet it is now rarely discussed. Arguments most commonly advanced against morphological rather than molecular phylogenetic analyses are:

(i) the limited number of characters available (and the existence of an asymptote of the number of phylogenetically informative character states available as taxa are successively added to a matrix: Bateman, 1992);
(ii) the high cost in time expended per unit character coded (both in terms of actual character scoring and the
‘informal apprenticeship’ that must first be undertaken in order to describe and code characters correctly; (iii) an inevitable degree of subjectivity involved in making *a priori* homology assessments (in other words, in delimiting characters before each is in turn divided into alternative character states); (iv) the supposed comparatively high level of homoplasy (this is caused by conflicting character states, which are considered to mainly reflect similar responses to similar pressures of directional or disruptive selection in lineages that are in fact only distantly related: Scotland *et al*., 2003), and; (v) the potential developmental correlation of apparently unrelated characters (e.g. pleiotropic expression of a single critical mutation).

Additional phenomena that can negatively affect both morphological and molecular matrices include:

(i) strong heterogeneity of rates of change within the study group, encouraging the most rapidly changing branches to falsely coalesce; this now thoroughly researched, but still problematic, phenomenon is termed long-branch attraction (e.g. Sanderson *et al*., 2000; Felsenstein, 2004); (ii) migration of genes (and the phenotypic characters that they underpin) between lineages through hybridization, lateral gene transfer, and organelle capture, and; (iii) the necessity of rooting the tree, typically through outgroup choice, in order to polarize characters and thereby study character evolution (another central issue of this paper: see below).
and then mapped, thereby effectively generating tautologous interpretations of evolution (Bateman, 1999).

How significant are optimization and outgroups choice?

The concept of character mapping leads naturally into discussion of another area of phylogenetics that is pivotal to the questions addressed by this paper, namely optimization. If we wish to know what the first flower looked like, but have not found it in the fossil record (such a discovery is highly improbable, given the undoubted patchiness of the fossil record of land-plants), we need to reconstruct that flower conceptually. This is achieved using combinations of characters found in species whose morphology has been carefully described and whose phylogenetic relationships have been rigorously inferred (this phrase is not oxymoronic; it is important to remember that even the most rigorously reconstructed phylogeny remains wholly inferential). Provided we have access to a matrix that thoroughly describes the morphology of all organs of the species under scrutiny, then we can reconstruct the morphology of the hypothetical ancestors that lie on each node of the cladogram. Optimization is most readily achieved by a simple logical protocol that works downward through the tree from the terminal taxa toward the outgroup node (Fig. 3) (e.g. Maddison and Maddison, 2001), though other, more complex, models can be applied, sometimes with advantageous results (Oakley and Cunningham, 2000; Polly, 2001; Webster and Purvis, 2002; Crisp and Cook, 2005).

A most-parsimonious tree, derived from a particular data matrix, yields similar information about the relationships of the coded taxa and the relative lengths of the various branches within the tree, irrespective of whether it is unrooted or rooted. However, an unrooted tree lacks polarity; it cannot be read in terms of transitions from ancestral character states (plesiomorphies) to derived character states (apomorphies). This in turn means that it is not possible to optimize character states in order to reconstruct the hypothetical ancestors occupying the nodes of the tree. In earlier (and more experimental) phases in the evolution of methods of phylogenetic analysis, several conceptually distinct approaches to rooting were explored, but most proved impractical and all appeared at least partially tautologous.

Over the last two decades a single method, termed outgroup comparison (e.g. Nixon and Carpenter, 1993), has become ubiquitous, to the point where it is routinely adopted without serious thought by almost all practising phylogeneticists. It requires the analyst to select a priori a set of comparable taxa (preferably species) that are of interest and are suspected of being an inclusive, monophyletic group (in practice, the strongest guide to perceived monophyly is a previous, and taxonomically broader,
phylogenetic analysis; thus, the outgroup method is vulnerable to accusations of logical tautology). Then, one or more additional taxa, thought to lie phylogenetically outside the chosen study group, and hence operationally termed outgroups, are chosen to simultaneously root the tree and polarize the scored characters. If multiple outgroups are chosen they constitute a test (albeit flawed) of the monophyly of the ingroup, since an outgroup taxon that is placed phylogenetically within the ingroup contradicts ingroup monophyly. Clearly, the larger the numbers of ingroup and outgroup taxa, the stronger is this test.

Outgroup choice leaves the analyst with a further quandary. In most cases, the ingroup is chosen because it is a morphologically and/or molecularly distinct aggregate of species, and so is likely to differ considerably from even its phylogenetically closest outgroups (hence, in the resulting trees, a monophyletic ingroup is likely to be subtended by a relatively long branch). These differences make character delimitation more difficult for both morphological and molecular analysts. For the morphologist, the outgroup is likely to lack structures found in the ingroup, possess structures not found in the ingroup, or the two groups may possess structures that are broadly similar yet sufficiently different that their homology cannot adequately assessed. For the molecular researcher, sequence alignment becomes more challenging for many genic regions, and base saturation becomes an ever-increasing risk. These problems of strong ingroup–outgroup divergence can, in theory, be averted by identifying one or more members of the ingroup as operational outgroups, but the analyst is then imposing a strong subjective overprint on the topology of the resulting tree(s). Certainly, the polarity of characters, and thus the optimizations that allow reconstruction of the properties of the hypothetical ancestor of the ingroup, would be strongly influenced by such a decision.

But before we contemplate implementing an optimization procedure, we must first select our preferred topology from among the very broad spectrum of land-plant phylogenies generated since 1982.

What is the best way to choose among the plethora of phylogenies?

One of our primary objectives here is to explore the relative merits of maximizing sampling of taxa versus that of characters per taxon in taxonomically broad (i.e. at least Class level) phylogenetic analyses. Morphological studies that include the best-understood fossils (i.e. conceptual whole plants reconstructed from their component parts: Chaloner, 1986; Bateman, 1992) offer the best opportunity to maximize sampling of taxa that provide strongly contrasting combinations of character states. By contrast, recent molecular trees, some of which compare entire plastid genomes (e.g. Goremykin et al., 2003, 2004, 2005), offer the most obvious means of maximizing the number of characters per taxon.

Given the continuing uncertainties regarding relationships among the major groups of land plants (discussed below), and their potential influence on optimizations of nodes, several branch-points distant from the target node(s), we sought phylogenies that encompassed all such major groups. In the case of morphological phylogenies, no existing study convincingly stretched from the bryophytes to the more derived angiosperms (eudicots sensu APGII: Angiosperm Phylogeny Group, 2003). We therefore took the controversial step of grafting a selected most-parsimonious tree from our recent analysis of 48 lignophytes (angiosperms plus gymnosperms, plus progymnospermous pteridophytes as operational outgroups: Hilton and Bateman, 2006, fig. 10) onto the product of a recent parsimony analysis of 52 coded taxa that focused on pteridophytes but used the fossil non-vascular polysporangiophyte Aglao-phyton as outgroup and just five taxa (one progynosperm plus three primitive pteridospermous gymnosperms and the extant Pinus) as representative ‘placeholders’ for the monophyletic lignophytes (Rothwell and Nixon, 2006, fig. 3). The resulting composite phylogeny contains 95 coded taxa: 47 wholly extinct and 48 containing at least one extant species (Fig. 4).

Selecting among the many broad-brush molecular trees in order to produce Fig. 5 proved even more problematic. The controversial analysis by Goremykin et al. (2003, 2004; see also Soltis et al., 2004; Martin et al., 2005) stood out as being especially character-rich, since it was based on nearly completely sequenced plastid genomes. The trade-off is that the analysis lacked data from the other two plant genomes (nucleus and mitochondrion), and was restricted to <20 coded taxa; moreover, the taxa were selected as much for their economic importance as their likely phylogenetic significance. This left the trees open to accusations of distortion by long-branch attraction (Soltis et al., 2004; Stefanovic et al., 2004; Leebens-Mack et al., 2005), though a more recent study suggests that likelihood model mis-specification is a more probable source of systematic error in the trees (Goremykin and Hellwig, 2006). The matrix of Soltis et al. (2002) encompassed a similar number of coded taxa, albeit more evenly distributed across the extant land-plant phylogeny, and sequenced eight genic regions distributed among the plastid (rbcL, atpB, psaA, psbB), mitochondrion (mtSSU, coxI, atpA), and nucleus (18S rDNA).

At the other end of the spectrum of molecular characters per coded taxon are analyses based on a single region but with substantial taxon sampling. A good example is the 2538-taxon rbcL study of land plants by Källersjö et al. (1998, 1999), subsequently reduced to an illuminating series of analyses of various combinations of 80 taxa sampled evenly across the phylogeny of extant land plants by Rydin and Källersjö (2002). Matrices of intermediate
Fig. 4. Composite morphological phylogeny of 95 taxa (47 fossil, in boldface) obtained by grafting (at arrow) a seed-plant tree of Hilton and Bateman (2006, fig. 10) onto a pteridophyte tree of Rothwell and Nixon (2006, fig. 3a).
dimensions include the 560-taxon, three-region (\textit{rbcL}, \textit{atpB}, 18S) analysis of angiosperms by Soltis et al. (2000, 2003a), the 105-taxon five-region (\textit{atpB}, \textit{rbcL}, \textit{atp1}, \textit{matR}, 18S) analysis of basal angiosperms and gymnosperms by Qiu et al. (2000), and the subsequent 100-taxon, nine-region (\textit{atpB}, \textit{matK}, \textit{rbcL}; \textit{atp1}, \textit{matR}, \textit{mtSSU}, \textit{mtLSU}; 18S, 26S) analysis of basal angiosperms and gymnosperms by Qiu et al. (2005).

What is the best way to assess the quality of a phylogeny?

When discussing the reliability of a published phylogeny, three terms are typically employed: accuracy, resolution, and robustness. From our philosophical viewpoint, the term accuracy can legitimately be applied to the topology that reflects the relationships of the taxa analysed only in very rare cases when the analyst is attempting to reconstruct a known and wholly dichotomous genealogy engendered by mankind (e.g. Hillis \textit{et al.}, 1992; Oakley and Cunningham, 2000; \textit{contra} Rokas and Carroll, 2005). Given that we can never gain access to the ‘one true tree of life’, by definition we cannot assess its accuracy, which is an absolute rather than a relative property.

When a fully dichotomous phylogeny (i.e. one that lacks polytomous nodes) is recovered from the matrix it is often said to be well-resolved. Robustness extends this concept to determining the relative strength of support for a particular node (relationship) within the context of that particular phylogeny. It is traditionally tested using now ubiquitous data-resampling techniques to generate support values, but even strong advocates of these techniques are being obliged to acknowledge their limitations. For example: ‘With high support for their tree, researchers can become confident in incorrect topologies. Although the bootstrap method for assessing confidence was originally suggested to provide confidence intervals for tree branches (i.e. how well the data at hand represent an underlying universe of data), it is now well-recognized that this resampling method and others, such as the jack-knife, provide, at best, a representation of confidence [pertinent] only [to] the data at hand; even random data can yield high bootstrap support. Furthermore, bootstrap values decrease as the number of taxa [included] increases, making it much more likely that high bootstrap values will be obtained if few taxa are analyzed.’ (Soltis \textit{et al.}, 2004, pp. 478–479). In other words, strong statistical support for a particular topology (or, more likely, a set of equally probable topologies)—that is, for a specific hypothesis of relationships—has little bearing on the unknowable property of its accuracy. This conclusion is borne out by numerous studies wherein highly contrasting, yet reliably strongly statistically supported, hypotheses of relationship have been generated from the same substantial data matrix. Such internal contradictions are achieved either by using
contrasting tree-building algorithms or by subsampling the matrix; for example, by removing "wildcard" taxa that demonstrably destabilize the topology (an act that often passes unreported in the resulting publications: cf. Hilton and Bateman, 2006), or by omitting specific categories of characters (e.g. third bases in codons: Jeffroy et al., 2006), or both (Philippe, 2006).

If accuracy is unknowable, and robustness unreliable, what criteria are left by which to judge the credibility of a particular inferred phylogeny? Most workers, including ourselves, consider the key word to be congruence. This term was originally coined to describe the relative behaviours of comparable characters (typically morphological features) when simultaneously subjected to a parsimony analysis, which aimed to minimize incongruence among characters (termed homoplasy). But soon incongruence was also used to describe any differences between multiple topologies describing the relationships of the same range of coded taxa.

The latter usage is well illustrated by an insightful, yet infuriating, study on 'resolving incongruence' in phylogenetic relationships among eight yeast species by sequencing 106 nuclear genes (∼2% of the number present in the genome: Rokas et al., 2003). Despite the small number of taxa analysed, the majority-rule consensus trees generated from each gene collectively offered 20 alternative phylogenetic hypotheses, most rich in 'strongly supported' nodes (defined by an arbitrary threshold of a bootstrap value exceeding 70%). Over half of the gene trees required the removal of at least two of the eight species to achieve topological congruence. Moreover, none of the data partitions or statistical patterns explored among the genes allowed the relative performances of individual genes to be predicted or cogently explained. Nonetheless, 'concatenating' (combining) all 106 genes 'yielded a single tree with 100% bootstrap values at every branch', so the authors 'conclud[e]d that it accurately represents the historical relationships of these eight yeast taxa and will be referred to hereafter as their species tree [they eventually concluded that, on average, 20 genes would generate a reliable species tree]. The maximum support for a single topology regardless of the method of analysis is strongly suggestive of the power of large data sets in overcoming the incongruence present in single-gene analyses’ (Rokas et al., 2003, p. 800). We find this conclusion extraordinary; in our view, the incongruence has not been 'overcome' but rather obscured by the sheer volume of data and the failure to explain the cause(s) of the profoundly conflicting phylogenetic signals provided by the individual genes (for a re-evaluation of this data-set see the final section of this paper).

Rather, we are most interested in assessing topological congruence between trees generated from different categories of data, each scored for the same set of taxa (preferably the same species, and ideally the same representative individuals). There is certainly a strong case for comparing phylogenetic signals from the three plant genomes, and from exons and introns; these genic regions operate under strongly contrasting molecular constraints (e.g. Page and Holmes, 1998). Obtaining similar results from all (or at least most) categories of data greatly increases our confidence (in the biological, rather than the statistical, sense) that the relationships recovered are credible. But there is a strong case for analysing genic regions separately before combining them (e.g. Bateman, 1999).

Surely the most versatile category of phylogenetic data is morphology sensu lato. Although restricted in total number of characters, most of those characters are phylogenetically informative (Wortley and Scotland, 2006) and each is potentially divisible into a large number of meaningful character states. Morphology informs not only on phylogenetic relationships but also on function, and through function it directly reflects various modes of natural selection. Detailed knowledge of morphology, along with other biologically relevant data such as symbiotic partners, habitat preference and environmental tolerance, is essential to any evolutionary interpretation of the range of taxa under scrutiny. And morphology maximizes taxon sampling by allowing full integration of fossil species, which serve a vital role in filling the vast morphological and molecular lacunae that separate clades possessing extant representatives.

Thus, the congruence that we seek is similar topologies generated from the same set of taxa by highly contrasting types of data, including morphology. Where trees generated from one or more categories of data (including morphology) disagree with trees generated from the majority of categories, a causal explanation should be sought, based on our knowledge of the biological properties of each category of data. This is the simplest way to identify processes that can confound reliable reconstruction of credible phylogenies.

For molecular data matrices, such processes include lateral gene transfer between taxa (Berghthorsson et al., 2004), lateral gene transfer among the three genomes within a plant (Huang et al., 2005), hybridization (Linder and Rieseberg, 2004), organelle capture (Rieseberg and Wendel, 1993), chromosomally localized gene duplication (Zahn et al., 2005), wholesale gene duplication through polyploidy (De Bodt et al., 2005), lineage sorting (Doyle et al., 1999; Degnan and Rosenberg, 2006), and codon bias. As summarized by Frohlich (2006), codon bias describes substantial differences in the relative frequencies of synonymous codons specifying particular amino acids. It can reflect both mutational biases that affect overall GC content and selection on individual codons (Kawai and Otsuka, 2004; Liu et al., 2004; Liu and Xue, 2005; Jeffroy et al., 2006). Both factors primarily affect third codon positions in general and transitions in particular, and are especially problematic in highly expressed genes. This is because selection among codons is hypothesized to reflect the relative abundances of tRNAs that bear different anticodons but synonymously insert the same amino acid; as a codon
becomes more scarce, the synthesis of its protein product is increasingly inhibited. This argument can be used to justify removal of third positions from phylogenetic analyses of highly expressed genes, including typical plastid genes. Such process-based explanations are, in our view, a necessary pre-requisite for selectively removing certain kinds of information from a phylogenetic matrix (see below).

By contrast, morphology can be undermined by convergence toward similar function(s) that is driven by directional or disruptive selection; examples include substantial vegetative modifications associated with transitions from terrestrial to aquatic habitats and from autotrophic to heterotrophic nutrition (e.g., Bateman, 1996), and floral transitions associated with switching pollinators (e.g., Chase, 1999). Less frequently discussed but also problematic are multiple, closely juxtaposed paedomorphic transitions within a clade, where the consequent morphological simplification can under certain circumstances be misconstrued by parsimony as primitiveness rather than as secondary simplification (Bateman, 1996).

Do some nodes in the land-plant phylogeny merit particular emphasis?

In practice, it is congruence among phylogenetic studies employing contrasting data matrices that has provided a study of land-plant phylogeny with its key reference nodes delimiting major monophyletic groups. Here, we contrast some widely accepted ‘reference nodes’ with other critical areas of land-plant phylogeny that ostensibly carry much greater uncertainty. In order to effectively deploy a bottom-up approach to understanding angiosperm origins, it is necessary to begin the discussion at phylogenetic nodes well below those of greatest interest.

Working upward through a crude consensus phylogeny that is based only on extant groups of land-plants but considering both morphological and molecular analyses (Fig. 2), we immediately encounter the embryophytes, a group that is diagnosed by the presence of archegonia and antheridia and so encompasses all land plants from bryophytes onwards. Excluding the ‘bryophytes’ leaves the monophyletic tracheophytes, which possess bona fide xylem. This is followed by a basal dichotomy within the ‘pteridophytes’ between the microphyllous lycopsids and megaphyllous euphyllophytes. Excluding the ‘moniliforeans’ (‘ferns’ sensu lato) from the euphyllophytes leaves a monophyletic seed-bearing spermatophytes; this clade encompasses four gymnospermous groups (Figs 4, 5), each widely regarded as monophyletic, plus the similarly monophyletic angiosperms, which arguably are best diagnosed by their possession of a triploid endosperm (however, even this character is unreliable in early-divergent angiosperms; for example, the endosperm is diploid in Nuphar polysepalum: Williams and Friedman, 2002). Bypassing a few, mostly small, basally divergent groups to which we will presently return, 96% of extant angiosperm species fall into two groups viewed as sister clades in a minority of analyses: monocots and eudicots. Monocots must be diagnosed by a gestalt of characters, each unreliable on its own. They possess a single cotyledon and lack a vascular cambium (both characters in common with a few early-divergent angiosperms and eudicots). Other putative monocot characters, such as trimerous flowers and sulcate pollen, are actually common among early-divergent angiosperms (e.g., Rudall and Furness, 1997). Most core eudicots are well defined by the possession of tricolpate- or tricolpate-derived-pollen (i.e. pollen with three equato- apertures), which is otherwise extremely rare (Furness and Rudall, 2004).

There are four main areas of uncertain relationships in such fossil-free trees:

(i) among the three main groups of bryophytes: liverworts, hornworts, and mosses;
(ii) among the five main groups of megaphyllous ‘ferns’: Psilotaiales, Ophioglossales, Marattiales, leptosporangiates, and equisetophytes;
(iii) among the four main gymnosperm groups: cycads (Cycadaeae sister to Stangeriaceae plus the relatively diverse Zamiaceae), Ginkgo, conifers, and Getales (the relatively species-rich Ephedra sister to Gnetum plus Welwitschia), and;
(iv) among the six main groups of basally divergent angiosperms: Amborella, Nymphaeales (perhaps including the antipodean hydrophiles of the Hydatellaceae; S Graham, pers. comm., 2006), Austrobaileyales (Austrobaileyaceae plus Schisandraceae plus Trimeniaceae), Chloranthaceae, magnoliids (including Piperales/Aristolochiaceae and Laurales—the only one of these ‘basal’ groups that exceeds 100 species), and Ceratophyllum.

Each of these four problematic groups (one bryophytic, one pteridophytic, one gymnospermous, and one angiospermous) is highly morphologically diverse.

If we parse through the land-plant phylogeny once again, this time including extinct taxa (and so confining our attention to morphological analyses), a substantially more complex story emerges (Fig. 4). The Early Devonian Aglaophyton, which uniquely possesses near-isomorphic sporophyte and gametophyte generations (e.g. Kenrick, 1994), is interpolated between the bryophytes and tracheophytes. The fossil rhiophytes form a distinctively vascularized clade that is sister to the eutracheophytes. The lycopsids are subtended by the wholly extinct Silurian-Devonian cooksonioids (probably paraphyletic) and zosterophylls (either monophyletic or paraphyletic: zosterophylls plus lycopsids=lycopophytes). The Devonian psilophytes and (arguably) the Devonian-Carboniferous stauropertids
interpolate at the base of the euphyllophytes, while Devono-Carboniferous cladoxyls provide potentially crucial information on the phylogenetic position of equisetophytes (Rothwell and Nixon, 2006). Devono-Carboniferous progymnosperms form an either monophyletic or, more likely, paraphyletic group that is interpolated between the ferns sensu lato and the spermatophytes, thereby permitting recognition of the lignophytes (progymnosperms plus spermatophytes).

Most critical of all from the viewpoint of inferring angiosperm origins is the immense diversity of gymnosperm groups that span a long and phylogenetically crucial period from the Late Devonian to the Cretaceous (see below). Three major groups of extinct Devono-Carboniferous seed-ferns (hydraspermans, medullosans, and callistophytes) together form a wholly extinct paraphyletic group of ‘core pteridosperms’, placed immediately below the divergence point of the cycads (Fig. 4). Paraphyletic or monophyletic Triasso-Jurassic peltaspers diverge above the cycads. Monophyletic Carbonifero-Permian cordaites form a sister-clade to the conifers, which now show multiple divergences of fossil taxa below the ‘crown group’. And in several studies, four derived gymnospermous groups, one Perm-Triassic (glossopterids) and three late Triassic–early Cretaceous (Pentoxylon, bennettites, Caytonia: Fig. 1C, D), each of which is resolved as sharing several attractive synapomorphies with extant angiosperms, form a paraphyletic group immediately subtending the angiosperm clade (e.g. Doyle, 2006; Hilton and Bateman, 2006).

Interestingly, it is only when considering relationships among angiosperm groups that the fossil record ceases to be central to the discussion. Much-debated fossils explored at greater length below, such as Archaefructus (Sun et al., 2002) and Sinocarpus (Leng and Friis, 2003) from the late Early Cretaceous (late Hauterivian–Early Aptian) Jehol biota (Zhou et al., 2003), were initially said to occupy phylogenetic positions earlier than the first divergence among extant angiosperm taxa (e.g. Sun et al., 1998). However, to some observers they appear more derived on closer examination, seemingly fitting comfortably within the angiosperm ‘crown group’ (Friis et al., 2003, 2006; Hilton and Bateman, 2006: see later discussion).

The node defining the crown group (Fig. 3) was described by Frohlich (2006) as constituting the ‘basal flower’, and its attributes as ‘the endpoint for any reasonable theory of flower origins.’ Certainly, on present knowledge, the hypothetical ancestor of the crown-group angiosperms can legitimately be reconstructed on the basis of extant taxa alone, in what has become the standard top-down approach (e.g. Doyle and Endress, 2000; Endress 2001b). However, we believe that fresh insights are more likely to arise from a bottom-up perspective on angiosperm origins. If so, it is a mistake to over-emphasize this particular node, which reflects the mere happenstance survival of extant lineages.

Given that some nodes delimiting major clades are widely accepted, there is an increasing temptation to constrain the results of further phylogenetic analyses to maintain those ‘known’ relationships (e.g. Doyle, 2006; Frohlich, 2006), irrespective of levels of character support within the matrix in question, and to focus attention on less well-understood regions of the phylogeny. For example, there have been no recent phylogenetic analyses, either morphological or molecular, that contradict the widely accepted monophyly of angiosperms; surely we can now take that node as read and routinely programme it into our tree-building exercises? In our view, this methodology is an excellent servant but a poor master. It is most useful as an experimental tool for exploring specific hypotheses of relationship.

For example, using similar morphological matrices, both Hilton and Bateman (2006) and Doyle (2006) tested the effect of constraining their topology to enforce monophyly of all extant gymnosperms, reflecting a topology recovered from several recent molecular analyses of extant seed-plants (Figs 2B, 5). Hilton and Bateman simply posed the question: How much less parsimonious is the morphological analysis when constrained by that one critical molecularly defined node? By contrast, Doyle elected to base the bulk of his excellent discussion of morphological character-state transitions among seed-plants on the unparsimonious topology that was constrained by the node that specified extant gymnosperm monophyly and was derived from wholly different (molecular) data matrices. Our preference is to give the characters that have been painstakingly coded within a particular matrix full freedom to express themselves. No node in land-plant phylogeny carries sufficient certainty to override the parsimony principle that is rightly fundamental to phylogeny reconstruction.

**Why is it important that optimal taxon sampling dissects long branches?**

The weakness of adopting only a neobotanical, top-down view of angiosperm origins becomes immediately apparent when one continues down the land-plant phylogeny from the node subtending the earliest divergence of extant lineages. There is no widely accepted evidence of fossil angiosperms older than the Early Cretaceous (~130 Ma), while the earliest evidence of their molecular sister-group, the gymnosperms, dates controversially from the Late Carboniferous (~290 Ma; putative primitive walchian conifers: Hernandez-Castillo et al., 2001) and certainly from the earliest Permian (~280 Ma; Cycas-like sporophylls: Gao and Thomas, 1989). The intervening period represents sufficient time not only for a great deal of morphological innovation, which tends to occur very rapidly on a geological time-scale, but also for extensive molecular change, which is widely recognized to follow
a broadly clock-like pattern (Bateman, 1999; Bateman and DiMichele, 2002; Sanderson et al., 2004). In other words, irrespective of whether one is considering molecular or morphological trees (with or without fossil taxa), the node subtending the last-diverging gymnosperms and the adjacent, relatively derived node subtending the first-diverging angiosperms are separated by an archetypal long branch (Hilton and Bateman, 2006) (branch emphasized in Fig. 3).

We have already encountered some of the problems presented by both phenotypic and genotypic long branches; their tendency to attract each other to generate a false statement of relationship, and their tendency to receive maximum bootstrap/jack-knife values that can be misinterpreted as evidence of accuracy. In addition, they can substantially weaken the effectiveness of outgroup comparison for rooting trees and polarizing characters, and pose particular challenges to a posteriori character analysis. Specifically, we cannot to infer the sequence in which the many character-state transitions supporting that branch took place, or determine whether they might be developmentally correlated. Our a priori assertions of homology of contrasting states of a character on either side of that branch will be less secure. Hypothetical ancestors reconstructed for the crucial internal nodes located at either side of the long branch will, by definition, be strongly divergent.

Fortunately, each of these problems can potentially be alleviated by improved taxon sampling. Increasing the number, and especially the phylogenetic spectrum, of taxa included in an analysis is likely to shorten branches supported by multiple character-state transitions, so that we can infer the temporal sequence in which those transitions occurred (Fig. 6). By dissociating previously positively co-occurring character-state transitions we also demonstrate that they are not directly developmentally or pleiotropically correlated, which in turn falsifies any prior hypothesis of saltational macroevolution that requires each of those character-state transitions (such profound and instantaneous heritable shifts in phenotype are the most appropriate null hypothesis for multi-transition branches according to Bateman and DiMichele, 1994, 2002).

Shortening the branch that separates a monophyletic ingroup from the operational outgroups is especially desirable (Fig. 6). We have already discussed the critical role of outgroup choice in rooting the tree and thereby polarizing character-state changes, and the reasons why the branch separating the outgroup(s) from the ingroup taxa will likely be long relative to those among the ingroup taxa. One or more added taxa that interpolate into that portion of the tree therefore have a disproportionately greater effect in improving the reliability of the resulting phylogeny, particularly by reducing the uncertainties surrounding homology assessment during the a priori character delimitation process (this statement applies equally to phenotypic and genotypic matrices).

In an ideal world, we would score every taxon that ever existed in the chosen clade. In practice, the best strategy is to sample coded taxa that are hypothesized to be distributed approximately evenly across the phylogeny under scrutiny. Where the available taxa are, by contrast, clustered in morphological and/or molecular space, increasing sampling density will simply add taxa that possess similar characteristics to those already included in the analysis, and the critical branches will not be significantly shortened; much analytical effort will thus have been expended for little interpretational reward. Beyond a certain density of sampling, this is inevitably the case in phylogenetically broad studies that analyse only extant taxa. It therefore becomes a powerful argument for pursuing analytical protocols that can allow active, and preferably full, participation of fossil taxa (Fig. 4).

What are the effects of culling taxa and/or characters?

The above argument can also be reversed by deliberately reducing the number of taxa included in a particular analysis in order to see what proportion of the original statements of relationship are retained in the reduced data matrix. One might hope that most data-rich matrices, molecular or morphological, would withstand this approach robustly, but that is rarely the case for either morphological (e.g. Hilton and Bateman, 2006; Rothwell and Nixon, 2006) or molecular (e.g. Rydin and Källersjö, 2002; Rydin et al., 2002) matrices. However, most matrices yield substantially altered relationships among the remaining taxa after the removal of relatively few taxa.

This effect is most pronounced where it is the number or identity of the operational outgroups that is altered. For example, Hilton and Bateman (2006) used in their core analysis three reconstructed progymnosperms to root their morphological analysis of extant plus extinct seed plants (Fig. 4). When only one (arguably the most derived) of the three progymnosperms, Cecropsis, was identified as outgroup, the rooting of the tree was radically altered. Specifically, the relationships of early-diverging taxa more closely resembled the results of recent molecular analyses, with angiosperms diverging below a monophyletic group of extant gymnosperms. By contrast, perceived relationships among the angiosperms, and among the extant gymnosperms, did not converge on those generated in molecular studies; rather, the fundamental polarity of evolutionary transitions was reversed. Moreover, only when Doyle (2006) constrained his similar but more angiosperm-focused morphological matrix to specify obligate monophyly of extant gymnosperms did he obtain a topology among the coded angiosperm taxa that resembled that in the most prominent molecular trees. In other words, forcing Gnetales into a sister-group relationship with extant
conifers has profound topological effects on a clade that diverges from a node several nodes distant from the taxa that actually caused the topological shift.

Moving on to molecular data, Soltis et al. (2004, fig. 3) were able to recover the much-maligned ‘monocots sister to dicots, Amborella sister to Calycanthus’ topology obtained from whole-plastid sequences by Goremykin et al. (2003, 2004) by reducing their own three-gene matrix to a similar number of taxa. However, they were also able to obtain from their data matrix their perception of the ‘correct’ topology—Amborella diverging before monocots which in turn diverge before Calycanthus—simply by adding to their analysis the orchid Oncidium. The Goremykin group responded by adding to their whole-plastid analyses data for Nymphaea (Goremykin et al., 2004), Acorus (Goremykin et al., 2005), and Phalaenopsis (Chang et al., 2006; Goremykin and Hellwig, 2006) with no consequent modification to their original topology. Indeed, a weakly supported ‘grasses-basal’ topology was also obtained by Burleigh et al. (2006) in their bootstrap consensus supertree, based on their gap-rich matrix of 69 taxa and 254 genes. Due to the large volumes of data involved, all pertinent relationships have strong statistical support. How then do we decide which statement of relationships is the more probable?

Several other taxonomically broad phylogenetic studies have explored the topological consequences of culling particular categories of character. The majority have considered the relative effects in expressed regions of sequenced genomes of first and second versus third position nucleotides in the codons that dictate the identity of the resulting amino acids. For example, the detailed analysis of the large plastid regions psaA plus psaB of 63 extant land-plants by Magallón and Sanderson (2002, fig. 3) generated substantially different topologies from parsimony analyses using (i) positions 1+2, (ii) position 3 alone, and (iii) positions 1+2+3. Analyses (ii) and (iii) placed horsetails below ferns, which were in turn improbably placed below lycopsids. Extant conifers were not monophyletic, with Gnetales the earliest to diverge. And Amborella diverged below rather than above Nymphaeales. Thus, analysis (i) would reconstruct hypothetical ancestors radically different from those of analyses (ii) and (iii) at the nodes on either side of the crucial branch separating gymnosperms from angiosperms. Adding insult to injury, maximum likelihood topologies differed significantly from parsimony topologies generated from the same matrix.

Similarly, the parsimony analysis of 31 extant vascular land-plants and 13 genic regions by Burleigh and Mathews (2004, fig. 7) revealed radically different placements between nuclear and plastid regions on the one hand, and mitochondrial regions on the other when all sites were included. However, when the most rapidly mutating sites were excluded (actually constituting two-thirds of the matrix), the three genomes yielded far more congruent topologies that placed Gnetales as sister to Pinaceae, agreed on extant gymnosperm monophyly, and placed Amborella as the earliest diverging extant angiosperm. Predictably, the authors concluded that the broad congruence of these signals from the three genomic compartments outweighed the contrasting signal when the number of characters used was maximized. Thus, the supposed primary advantage of molecular phylogenetic data—quantity of characters—was pragmatically qualified by the understandable desire to improve the perceived overall quality of signal obtained from the data. Moreover, this decision was constructively supported by an at least partial (and enhanced) understanding of the biological differences between the data types retained and those discarded.

What are the pros and cons of monophyly?

The conventional ‘ladderized’ classification of land plants, based largely on grades rather than clades (i.e. bryophytes, pteridophytes, gymnosperms, angiosperms), was rapidly overturned by morphological phylogenies that strongly
emphasized recognition of monophyletic groups. Taxonomically well-sampled studies agreed that all major groups other than angiosperms were non-monophyletic (Figs 2A, 4). However, this conclusion has often been challenged by multi-gene molecular phylogenies using only extant taxa, which increasingly resolve as monophyletic not only extant angiosperms but also extant gymnosperms (cycads, *Ginkgo*, conifers, and, controversially, Gnetales), extant pteridophytes, and (less confidently) extant bryophytes (Figs 2B, 5). Thus, high-level classifications of land plants that emphasize monophyletic groups increasingly resemble early, pre-cladistic models, encouraging an improbable approachment between two previously disparate groups of researchers: molecular phylogeneticists and those systematic morphologists who never accepted the cladistic paradigm. Only some phylogenetic morphologists, particularly those with palaeobotanical interests like us, remain unconvinced that these events constitute unequivocal progress. What lies at the root of our continued scepticism?

We have already addressed one major set of concerns—the profoundly deleterious consequences of the vast phylogenetic lacunae that separate the major groups that possess extant representatives: bryophytes versus pteridophytes versus cycads versus *Ginkgo* versus conifers versus Gnetales versus basal angiosperms. These render less certain *a priori* homology assessment and tree rooting via outgroups, and increase the probability of generating incorrect topologies through long-branch attraction, particularly when using sequence data.

The second motivation is that a broader knowledge of the morphological diversity shown by land-plants since their putative Silurian origin not only reinforces our perception of the scale of the lacunae separating the extant ‘crown groups’ but also emphasizes the improbability that the molecular topologies specifying major-group monophyly are correct. Monophyly of extant non-lycopsid pteridophytes means that the lineage leading to lignophytes (progymnosperms plus seed-plants) diverged before the divergences among the major eusporangiate groups (including the Psilotales and Equisetales) and before that separating the eusporangiate and leptosporangiate taxa. Similarly, monophyly of extant gymnosperms requires that the lineage leading to the ‘crown group’ angiosperms diverged before all of the diverse gymnosperm lineages (e.g. Doyle, 1998a). This statement applies even to the cycads, which are widely accepted as retaining many plesiomorphic characteristics and occur in latest Pennsylvanian assemblages. This hypothesis of relationship demands a minimum of 160 my with no readily recognized fossil angiosperms, and yet more time elapsed prior to their well-defined mid-Cretaceous radiation (Friis et al., 2006). Thus, we can either accept that no extant fern lies on the lineage leading to spermatophytes and no extant gymnosperm lies on the lineage leading to angiosperms, despite the extremely low combined probability that both statements are correct, or we can explore further the possibility that sequence-based phylogenies are biased toward monophyly, presumably as a consequence of the aforementioned lacunae in taxon sampling. Certainly, we believe it is critical that debates continue regarding the pros and cons of specific topologies (and specific kinds of data).

The third set of issues concerning monophyly is central to the topic of this essay. If we perceive extant land-plants as constituting just five major taxonomic groups (in order of successive divergence: bryophytes, lycopsids, non-lycopsid pteridophytes, gymnosperms plus angiosperms; Fig. 2), we have no evidence available to reconstruct the lengthy sequences of character-state acquisitions (i.e. morphologically and molecularly long branches) between these vastly disparate groups. In other words, we cannot to address directly the most profound question in land-plant evolution. Thus, monophyly of major groups is a boon to classification but an anathema to those of us wishing to study the patterns and processes of plant evolution. Given this observation, are our concerns regarding the likelihood of major-group monophyly driven less by a dispassionate view of existing data than by biased self-interest?

**Which key characters best define the flower?**

Understandably, most developmental geneticists have come to view the *Arabidopsis* flower as archetypal, yet presumably its standard complement of 4 sepals, 4 petals, 6 (4+2) stamens and 2 carpels is derived from the classic [5+5+5+[2–5]] floral bauplan of the eudicot clade, which encompasses ~75% of extant angiosperm species. This clade is well-defined by both molecular and morphological data, and possesses some relatively consistent synapomorphies, notably the possession of tricolpate- and tricolpate-derived pollen, and also (at least in core eudicots) a ‘typical’ floral groundplan of hermaphrodite pentameric flowers showing clear differentiation of the perianth into sepals and petals, an androecium of two stamen whors, and a syngnemized gynoecium of two to five fused carpels.

Furthermore, a clear series of basally divergent extant eudicots (the non-core eudicots *sensu* Angiosperm Phylogeny Group, 2003) allows bottom-up polarization of floral character states (e.g. Soltis et al., 2003; Ronse Decraene, 2004), though even this approach is problematic. For example, Wanntorp and Ronse Decraene (2005) cautioned that the minute, dimerous and highly reduced flowers of Gunnerales (the putative sister to the core eudicots in many recent molecular phylogenies) are unlikely candidates for the ‘ancestral’ eudicot flower.

By contrast, understanding the origin of the angiosperm flower is a far more challenging task. The plethora of often contradictory theories surrounding the origin of the flower (e.g. Arber and Parkin, 1907; Wettstein, 1907; Melville, 1960; Eames, 1961; Meeuse, 1975, 1987; Hughes, 1976; Burger, 1977) suggests that a comparative top-down study
of extant angiosperms cannot provide an unequivocal picture of floral evolution. As outlined above, floral structures in early-divergent angiosperms are highly diverse and often lack definitive features, including hermaphroditism, fully closed carpels, and a distinctly whorled arrangement of organs.

Certainly, the flower is less readily definable in groups that diverged prior to the core eudicots. For example, it is problematic to distinguish between sepals and petals in the majority of monocots and early-divergent angiosperms (Endress, 2001a, b). Endress (2001a) usefully disassembled the flower into components that occur in all seed plants (ovules and microsporangia) and uniquely angiospermous elements (carpels). The carpel can be interpreted either as an ovule-bearing leaf homologue (a megasporophyll) or as a composite organ consisting of an ovule-bearing structure surrounded by a modified bract. Thus, Endress argued that the ‘real’ angiosperm innovation is post-genital carpel fusion, which led to further morphological innovations. However, he also acknowledged that in many early-divergent extant angiosperms carpel closure is caused not by bona fide fusion but rather by occlusion via ‘secretion’ (presumably of mucilage).

Top-down comparisons of extant angiosperms with putative stem-group fossils also fail to resolve the debate. The case of the fossil aquatic plant Archaeaefructus, excavated from the Yixian Formation of north-east China, nicely illustrates this point. When first discovered, Archaeaefructus was considered to be late Jurassic in age (Sun et al., 1998) and was interpreted as a basal angiosperm providing plesiomorphic morphological characters elucidating our understanding of angiosperm origins. Archaeaefructus was consequently hailed as a potential sister-taxon to all extant angiosperms, based on a morphological phylogenetic analysis (Sun et al., 2002). However, the age of the Yixian Formation was subsequently reinterpreted as late early Cretaceous (e.g. Barrett, 2000; Zhou et al., 2003), significantly younger than originally claimed. Moreover, Friis et al. (2003) reinvestigated these and other associated fossils, leading to their own morphological phylogenetic analysis. Their re-coding of the characters suggested that Archaeaefructus is better interpreted as either a magnoliid or even a basal eudicot, albeit one specialized for an aquatic habitat. Like several extant early-divergent angiosperms, Archaeaefructus lacks a perianth, and presents an array of structures of questionable homology. Friis et al. (2003) suggested that it had unisexual flowers consisting of only two stamens (in male flowers) or one or two carpels (in female flowers), rather than possessing the hermaphrodite flowers with axially separated paired carpels and stamens envisaged by Sun et al. (2002). Their reinterpretation has been supported by subsequent palaeobotanical discoveries (Ji et al., 2004; Friis et al., 2006).

This ambivalent structure is consistent with that of another aquatic wildcard taxon, Ceratophyllum, and some aquatic monocots in the monocot order Alismatales, whose origin and floral homologies have proved similarly difficult to interpret (reviewed by Sokoloff et al., 2006), as have those of the hydrophilic Hydatellaceae. However, unlike Archaeaefructus, these are extant taxa, so that their homologies could potentially be explored using evolutionary-developmental genetic techniques (see below). Less controversially, the similarly aged genus Sinocarpus is also placed as a basal eudicot (Leng and Friis, 2003) and so belongs among the ‘crown group’ angiosperms. In the morphological analysis of Hilton and Bateman (2006), adding Sinocarpus left the original topology unaltered, but it was placed within the ‘crown group’ as sister to the magnoliid family Piperaceae. By contrast (and perhaps predictably), Archaeaefructus acted as a wildcard taxon, destabilizing relationships among the basal extant groups to generate a five-taxon polytomy.

Endress (2001a) suggested that the two most prominent angiosperm floral elements, carpels and thecal stamen organization, likely arose as key innovations in as-yet undiscovered stem-group angiosperms. By contrast, features adapted to attract pollinators, such as a colourful perianth and odour- and nectar-producing structures, probably evolved more than once within the angiosperm clade. Scent production, sometimes associated with internal heat generation, is believed by some to have evolutionarily preceded colour (Thien et al., 2000). Based on fossil evidence of insect faecal remains in bennettite strobili and gymnosperm pollen in the gut contents of fossil insects, Labandeira (1997, 1998) argued that insect pollination was common before the first recognizable angiosperms appeared in the late Jurassic, at a time when foraging insects sought pollen rather than nectar. Floral diversity among extant early-divergent angiosperms probably reflects experimentation with insect mutualisms, especially with small insects such as beetles, flies, and thrips (Thien et al., 2000; Endress, 2001b). Much evidence, from both morphology and developmental genetics, indicates multiple origins of petals, either from bracts (termed bracteopetals) or from stamens (termed andropetals). Only andropetals are commonly associated with a differentiated perianth (reviewed by Kramer and Irish, 2000). Some early-divergent angiosperms lack a perianth entirely, including some Piperales and most Chloranthaceae; both these groups are well represented in the early Cretaceous fossil record (e.g. Friis et al., 2000, 2006). These taxa tend to have more generalist pollination strategies; for example, many Saururaceae (Piperales) are pollinated by a diverse range of small insects and a few by wind (Thien et al., 2000).

Why are palaeobotanists obsessed with extinct gymnosperms?

Although we could perhaps consider ourselves fortunate to have in the extant flora representatives of at least three and
arguably four major lineages of gymnosperms—cycads, *Ginkgo*, conifers, and Gnetales—these represent a modest fraction of the total morphological diversity evident among the gymnosperms if fossil groups are also considered (e.g. Stewart and Rothwell, 1993). It is therefore feasible, and even likely, that these groups plus the angiosperms arose as independent lineages from within the plexus of wholly extinct gymnosperms. Certainly, in morphological phylogenetic analyses rich in fossil taxa these extant groups rarely emerge as sisters; rather, each is sister to at least one wholly fossil lineage (reviewed by Hilton and Bateman, 2006) (Fig. 4).

Just as fossil progymnosperms are crucial for routinely inferring a single origin of seed-plants, the combined fossil and extant gymnosperms are critical for routinely inferring a single origin of angiosperms. Here, at least, there is general agreement between morphological and molecular phylogenies (Fig. 2). However, by contrast, recent molecular assertions of monophyly of extant gymnosperms profoundly alter our perception of seed-plant relationships (cf. Figs 4, 5).

Most discussions have focused on the controversial placements of the clade formed by the three extant genera of Gnetales as either sister to extant conifers or even just extant Pinaceae. However, the controversy surrounding the phylogenetic placement of Gnetales [reviewed by Burleigh and Mathews (2004), with more recent contributions from Burleigh et al. (2006), Hajibabaei et al. (2006), and D Quandt (pers. comm., 2006)] has tended to obscure the potential interest of the more plesiomorphic taxa such as *Ginkgo* and especially cycads, which on morphological phylogenetic evidence could legitimately be regarded as extant representatives of the supposedly wholly extinct pteridosperms (Hilton and Bateman, 2006). Extant cycads and *Ginkgo* offer a unique opportunity to examine the characteristics of early-diverging spermatophytes (Brenner et al., 2003). However, surprisingly few modern studies exist of these ‘living fossils’, and those that do rarely employ a bottom-up comparative approach. One notable exception is the ontogenetic study by Mundry and Stützel (2003), who reinterpreted the male sporangiophores of *Zamia* as pinnate, with synangia on reduced leaflets, and hence suggested that cycad male sporangiophores may be better homologized with the radial synangial groups observed in some Carboniferous medullosan pteridosperms than with the simple sporangiophores that characterize extant conifers.

For those of us familiar with fossil gymnosperms, the implication that extant angiosperms diverged prior to the separation of extant cycads and *Ginkgo* appears strongly incongruent, both with morphological phylogenies and with first appearances of relevant taxa in the fossil record. Admittedly, as noted by Doyle (2006), ‘the position of cycads is one of the more weakly supported aspects of molecular phylogenies (cf. Magallon and Sanderson, 2002; Solis et al., 2002)’. Nevertheless, extant gymnosperm monophyly would require that angiosperms diverged from the other seed plants by the end of the Carboniferous (~290 Ma), when the only likely candidates for ancestor in the fossil record are various pteridosperm lineages (Fig. 4). This necessitates an extraordinarily long angiosperm ghost lineage of some 160 my to reach the point in time when the first widely accepted angiosperm fossils appear, in the early Cretaceous.

However, accepting at face value this almost irreconcilable ghost lineage constitutes an unnecessarily nihilistic view of the implications of such phylogenies. Although the molecular placement of extant Gnetales adjacent to or within extant conifers has led to widespread triumphing of the demise of the anthophyte hypotheses (and of the implicit pseudanthial origin of the gnetalean flower), many morphological cladistic analyses (Fig. 4) still recover the remainder of the anthophyte clade, containing not only the angiosperms and the Mesozoic *Caytonia*, *Pentoxylon*, and bennettites (recently credibly reinterpreted by Rothwell and Stockey, 2002: Fig. 1C, D) but also the Carboniferous–Permian glossopterids (recently credibly reinterpreted by Doyle, 2006). Rather than agree with Frohlich’s statement that ‘the underpinnings are removed from the anthophyte theory’ (Frohlich, 2006), we prefer his view that ‘one must be careful not to discard an entire scenario if a separable component is [more correctly, ‘appears’] wrong’.

More broadly, morphological cladistic analyses provide a credible series of stem-groups to link the angiosperm crown-group to Carboniferous pteridosperms (Doyle, 2006; Hilton and Bateman, 2006). They also have the potential to elucidate how the various components of the angiosperm flower accrued, since each of these stem-groups possesses some flower-like features. In the great shell-game of angiosperm origins, it appears increasingly unlikely that the pea lies underneath the Gnetalean shell; however, the other anthophyte shells remain resolutely in play. They are a clear priority for more intensive research, preferably based on fossil assemblages that combine contrasting preservation states.

By now, it should be clear that our preferred definition of the flower is considerably more inclusive than our preferred definition of angiosperms. Defining a flower as a determinate axis terminating in megasporangia that are surrounded by microsporangia and are subtended by at least one sterile laminar organ (Table 1) effectively encompasses the reproductive structures of several anthophytes. Consequently, accepting the currently most popular phylogenetic placement of Gnetales—adjacent to or within extant conifers—would require at least two independent origins of the flower among extant taxa, and additional origins if wholly extinct taxa are also considered. Our preferred characters for delimiting the angiosperms focus on the carpel and its contents; specifically, we require both a closed (though not necessarily fused) carpel preventing direct contact between
the microspore wall and the megaspore wall, and the occurrence of double fertilization leading to formation of functional endosperm (both features are absent from Gnetales, as they are from some early-divergent angiosperms).

Are clades such as angiosperms best defined using taxa or characters?

Placing character-state transitions within the explicit context of a cladogram helps to clarify the key issue of recognizing and delimiting natural (monophyletic) groups. Three approaches are available, though in practice individual phylogeneticists rarely state which approach they prefer.

A node-based group omits all of the characters on the branch subtending the basal node of the exclusive set of coded taxa that constitute the group (Fig. 3). Thus, by definition, it can be delimited only by the taxa that it encompasses, rather than by unifying synapomorphies (consequently, this is not a practical approach to delimiting monophyletic groups). Any taxon subsequently added to the matrix that proves to be interpolated within the original subtending branch will, by definition, be excluded from that group.

Conversely, a stem-based group includes, without prejudice, all of the character-state changes on the branch subtending the basal node. If that branch is long, many ostensibly diagnostic characters are available. Any taxon subsequently added to the matrix that proves to be interpolated within the original subtending branch will, by definition, be included in that group.

The third approach bisects the subtending branch by selecting an acquisition of a single apomorphic character state as being diagnostic of that group (e.g. for angiosperms one could opt to prioritize double fertilization leading to a triploid functional endosperm; Fig. 3). This meets our requirement to readily diagnose the clade, and means that an additional taxon attaching itself directly to that branch might or might not join that clade, depending on whether or not it possessed the key diagnostic character state. It is also the only one of the three approaches to explicitly allow delimitation (via autapomorphies) of groups represented in the analysis by only one coded taxon (i.e. operationally monotypic). Of course, the requirement could be that clade membership by a taxon requires possession of not one but two or more of the synapomorphies allocated to the subtending branch (e.g. adding to the triploid functional endosperm a requirement for a proembryo that is not tiered and a secondary suspensor that is nonetheless partly derived from the primary suspensor cell, or simply the presence of a bona fide embryo sac; for further details of these characters see Doyle, 2006).

Of the three approaches, we regard specifying apomorphies as being the most logical, stable, and widely applicable. However, it does require the proponents of each analysis to strongly justify their subjective choice of diagnostic character state(s), effectively constituting a posteriori character weighting.

Do some categories of character merit particular emphasis?

Morphological characters

Ever since Linnaeus (1735) chose to prioritize floral features over vegetative features, and more precisely stamens over carpels and both over tepals, most botanical systematists have been tempted to prioritize certain selected characters (or, more precisely, in monophyletic classifications, apomorphic states of characters) for classifying plants and, latterly, for determining their phylogenetic relationships. With the exception of a widely held preference for characters readily divisible into discrete states, morphological cladistics has encouraged a more egalitarian perspective on characters; weighting, if practised at all, has generally been post hoc rather than ad hoc. Indeed, post hoc comparisons of relative levels of homoplasy in different organs have encouraged some phylogeneticists to counsel against routinely prioritizing selected organs in general and flowers in particular, instead perceiving egalitarianism among organs (and characters) as one of the greatest strengths of the cladistic method (Bateman and Simpson, 1998).

Most analysts have transferred their emphasis from morphological to molecular data when constructing phylogenies. Nonetheless, the majority still prefer one or more readily recognized phenotypic character(s) for delimiting clades in the resulting trees. Even then, tensions can emerge; for example, some of the most effective group-delimiting morphological synapomorphies are not readily recognized in fossils (these include the double fertilization and resulting triploid endosperm, and the reduced proembryo, that were highlighted in the previous section as potentially delimiting the angiosperms).

Point mutations versus large-scale mutations

Perceptions of the various processes of mutation have undergone a radical overhaul during the previous decade. Studies have overturned the long-held fundamental genetic principles that both mutation and recombination are random factors (e.g. Jin and Bennetzen, 1994; Tsunoyama et al., 2001). Moreover, typical mutation rates have been shown to be one to two orders of magnitude higher than previously thought (~2 per genome per generation: Denver et al., 2004), and to be subject to reversal to the plesiomorphic condition over a very small number of generations (Lolle et al., 2005; Weigel and Jürgens, 2005). In addition, unexpectedly high frequencies and heritabilities have been demonstrated for methylated ‘epialleles’ (e.g. Kalisz and Purugganan, 2004).
Arguing for a fundamental contrast in the nature of morphological phylogenetic and molecular phylogenetic data, Bateman (1999, p. 446) provocatively stated that ‘the vast majority of morphological character-state transitions occur during speciation events, whereas the vast majority of molecular character-state transitions occur between them’. The main exception to this rule envisaged by Bateman (1999) and Bateman and DiMichele (2002) was the minute percentage of mutations (often simple point mutations) that are both non-lethal and enable direct, substantial modification of the phenotype. These would have greatest effect when affecting key developmental genes; the mutation alters the amino acid(s) generated, which in turn alters the behaviour (or concentration) of the protein sufficiently to ultimately cause substantial modifications of the resulting phenotype. Such events are especially likely to prompt shifts in the timing (heterochrony) and/or spatial expression (heterotopy) of developmental events, and be expressed in multiple parts of the organism (i.e. pleiotropically). When fed into a molecular phylogenetic analysis such mutations disappear from view, immediately becoming needles in a nucleotide haystack (though this is not necessarily the case when they become the foci of evolutionary-developmental studies; see below).

Although the vast majority of DNA characters are point mutations, many analysts have been drawn preferentially to larger-scale mutations, involving insertions, deletions, inversions, and other genomic rearrangements (King, 1993; Rokas and Holland, 2000; Levin, 2002) that affect expressed regions of the genomes. One of the attractions of such characters (termed ‘rare genomic changes’ by Burleigh and Mathews, 2004) was the reasonable assumption of an extremely low likelihood that such mutations would show homoplasy (i.e. would have multiple origins or revert to the previous condition), though this assertion was based on the aforementioned assumptions of randomness.

One such mutation that has at least partly stood the test of time is the acquisition of the inverted repeat in the chloroplast genome on the spine of the land-plant phylogeny, which occurred between the divergence of the lycopsids and that of the remaining extant pteridophytes (Fig. 2) (Raubeson and Jansen, 1992a). This region, constituting ~20% of the plastid genome, was not only duplicated but also reversed, so that it is transcribed in the opposite direction from that of the original copy. Interestingly, the duplicated region is partly or wholly absent from all extant conifers but present in all extant Gnetales studied to date (Strauss et al., 1988; Raubeson and Jansen, 1992b; Raubeson et al., 2004), implying a Gnetales-sister topology. By contrast, the absence of functional ndh genes from the plastid genomes of both Gnetales and Pinaceae supports the Gnepine hypothesis (Chaw et al., 2000; Burleigh and Mathews, 2004). Several other genomic rearrangements serve simply to emphasize the distinctness of Gnetales from all extant conifers.

The impression gained over the last few years is that rare genomic changes are neither as rare, nor as reliably unique (i.e. immune to homoplasy), as was previously suspected. Nonetheless, this observation cannot adequately justify the rapid disappearance of these useful characters from most phylogenetic discussions.

Gene duplications and gene expression: evolutionary-developmental genetics

Another category of rare genomic change, which is proving to be less rare than previously believed, is various modes of gene duplication, which have understandably attracted much attention recently (e.g. Lynch, 2002). Naturally, much of this attention has focused on the limited number of developmental genes that underpin the ABCDE model of floral organ control (examples of A=AP1/SQUA, B=DEF/GLO, Bs=ABS, C=AG/PLE, D=AG-like, E=SEP: Theissen et al., 2002). In ‘typical’ eudicots, sepal expression of A genes, petals A+B+E, stamens B+C+E, carpels C+E(+Bs), and ovules Bs+D+E. Typical monocots (and other early-divergent angiosperms) lack a distinct A-only sepal whorl, and A-function is wholly absent from extant gymnosperms, where both male and female cones exhibit C/D expression, but males exhibit B expression whereas females exhibit Bs expression (Theissen et al., 2002; De Bodt et al., 2003; Zhang et al., 2004; Krizek and Fletcher, 2005; Hintz et al., 2006; G Theissen, pers. comm., 2006).

Current evidence suggests that B- and C/D-function gene lineages diverged prior to the node subtending extant angiosperms and extant gymnosperms, ~400–300 Ma, as did the two main clades in the C/D-function AG group (Zhang et al., 2004) and the AGL6 versus AGL2–4+AGL9 split in the SEP group (Zahn et al., 2005). Many of the remaining key divergences have been attributed to that most troublesome branch that links the divergence point of angiosperms and extant gymnosperms to the node subtending the extant angiosperms. Somewhere along this branch are placed the divergence of the C and D functions (Kramer et al., 2004), the two main clades of B-function genes, DEF/AP3 and GLO/IP1 (Theissen and Becker, 2004) (a divergence dated to 260 ± 30 Ma by Kim et al., 2005), and the two main clades within the SEP group, AGL2–4 and AGL9 (Zahn et al., 2005). Molecular clock estimates indicate that these duplications could have been contemporaneous, leading to suggestions of a crucial polyploidy event on the stem-lineage leading the extant angiosperms (De Bodt et al., 2005; Zahn et al., 2005).

These gene families show extensive evidence of duplications among the extant angiosperms (Kim et al., 2004; Kramer et al., 2004), which can often be linked to morphological shifts. For example, Litt and Irish (2003) suggested that the restriction of A-function euAP1 genes to core eudicots could indicate co-option of new mechanisms
of floral development that are correlated with the ‘fixation’ of floral structure in the eudicot clade, most notably the strong distinction between petals and sepals. Within the petaloid monocots, the exceptional differentiation among the tepals of orchids is reflected in an unusual diversity of B-class genes (Tsai et al., 2004; Xu et al., 2006). Within dipsacalean eudicots, multiple duplications in CYC-family genes have been implicated in various phylogenetic shifts in floral morphology and symmetry (Howarth and Donoghue, 2005), while the subtleties of expression of such genes among perianth members have been well-documented in various Brassicaceae (S. Zachgo, pers. comm., 2006).

Polyploidy is perhaps the easiest category of gene duplication to detect, since every nuclear gene is copied simultaneously during the whole-genome duplication process (e.g. Linder and Rieseberg, 2004; De Bodt et al., 2005). In the case of allopolyploidy, the hybridization event that accomplishes genome duplication generates a unique combination of genes, albeit at the expense of introducing into the phylogeny a reticulation event that links two previously independent lineages. For example, despite its small genome size, the eudicot model organism Arabidopsis retains evidence of no less than three duplications of the entire nuclear genome: one in the mid-Cenozoic following the main angiosperm radiation, one in the late Cretaceous during the radiation, and one broadly contemporaneous with the estimated divergence time of the monocots, basal dicots, and earliest eudicots, in the late Jurassic–early Cretaceous (De Bodt et al., 2005). Admittedly, the estimated timings of these events carry large error bars, a statement that applies equally to single-gene duplications.

Any or all of these events could have aided taxonomic diversification by allowing functional diversification within specific gene families, particularly those involved in transcriptional regulation and/or signal transduction (Blanc and Wolfe, 2004a, b; Kellogg, 2004, 2006; Maere et al., 2005); categories that include key developmental genes such as the MADS-box clusters (Zahn et al., 2005). Wholesale gene duplication remains an attractive paradigm, because it allows the subsequent accumulation of the large amount of genetic diversity necessary to explain the ‘slow fuse’ that preceded the extraordinarily rapid morphological (and thus taxonomic) diversification of the angiosperms. Fewer key developmental genes survive small-scale duplication events, as co-adapted sets of genes widely distributed on the chromosomes will become dissociated.

A classic study of the potential significance of duplication in nuclear gene families was conducted by Frohlich and Parker (2000) (see also Frohlich, 2002, 2006), who sequenced the low-copy nuclear gene LEAFY (FLO) in a phylogenetically well-chosen set of place-holder gymnosperm and angiosperm taxa, plus two pteridophyte outgroups. The resulting trees were consistent with extant gymnosperm monophyly, but also revealed the presence of two copies of LEAFY in most of the gymnosperms studied (the exception was the Gnetalean Gnetum). The copy that is absent from pteridophytes, angiosperms, and Gnetum was named Needly. Assuming extant gymnosperm monophyly, the most parsimonious explanation of the phylogenetic distribution of copies of LEAFY s.l. would be a single duplication event, which generated Needly, occurring between the divergence of the angiosperms from the gymnosperms and the crown-group node subtending the extant gymnosperms; this was in turn followed by a loss of Needly from Gnetum. In other words, Needly never occurred in the angiosperm lineage. However, the molecular phylogenies, including the amino acid-based phylogeny derived from LEAFY itself, suggest that the duplication of LEAFY preceded the divergence of extant angiosperms and extant gymnosperms; gymnosperm LEAFY is depicted as sister to angiosperm LEAFY rather than to gymnosperm Needly. Acceptance of this topology requires a less parsimonious explanation of the gene duplication itself. The original duplication is now perceived as having occurred below the divergence point of the angiosperms, with the Needly copy suffering parallel losses (strictly, deactivations) in Gnetum and, more importantly, early in the independent life of the angiosperm lineage.

It was this interpretation that, in turn, prompted the ‘mostly male’ theory of angiosperm origins, arguing that the deactivation in angiosperms of Needly, which specifies femaleness in gymnosperms, allowed a switch from ancestral monococious cones to bisexual flowers by immediate (i.e. saltational sensu Bateman and DiMichele, 2002) ectopic (broadly, heterotopic) expression of ovules on a fundamentally male shoot. This intriguing hypothesis is eminently testable through further evolutionary-developmental study (Frohlich, 2006). However, we note that there is no obvious method of demonstrating that the Needly copy was indeed present in the shared ancestor of angiosperms and extant gymnosperms; this key element of the hypothesis remains an inference because of the growing perception of extant angiosperms and extant gymnosperms as monophyletic sister-groups.

Baum and Hileman (2006) recently built, on groundwork laid by Theissen and coworkers (Theissen et al., 2002; Theissen and Becker, 2004), an elegant evolutionary model designed to explain the determinate growth and bisexuality of the flower, and the origin of petals. It boils down to a sequence of four key steps:

(i) Evolution of a bisexual axis via a gynomonoecious intermediate, through homeotic conversion of distal microsporophylls (stamens) into megasporophylls (carpels) within the pollen cone. This was explained by differences in maximal expression levels of B- and C-class organ identity genes and competition among their gene products. Specifically, the B-class MADS-box genes that, along with C-class genes, are...
necessary for expressing maleness generate proteins that must operate in combination with E-class SEPALLATA genes. Femaleness requires the presence of C-class expression but the absence of B-class expression. Any concentration of C-function proteins in the apex of a male cone could complex excessive amounts of SEopalata protein so that little remains to interact with the B-function proteins, thereby allowing the terminal portion of the cone to cross the critical threshold into femaleness.

(ii) Evolution of floral axis compression and determinacy. Here, C-class genes became negative regulators of the meristem maintenance gene WUSCHEL, which maintains the indeterminacy of the apical meristem. As a result of the ensuing switch to determinacy, apical growth ceases and the pattern of initiation of organ primordial grades from a helix to a spiral, allowing their relative positions to become a critical factor in determining their respective developmental fates.

(iii) Evolution of petaloid perianth by sterilization of outer stamens. This is postulated to have occurred when WUSCHEL was co-opted as co-regulator of C-class genes, causing B-gene expression to extend closer to the axial apex than C-gene expression and thus creating a zone where expression of petals (effectively hybrid organs) was feasible.

(iv) Evolution of the classic perianth of core eudicots. The perianth became strongly dimorphic and relatively invariant, as B-class function became increasingly dependent on UFO co-regulation.

We echo Frohlich (2006) in recognizing that both of these hypotheses of the origin of the angiosperm flower are elegant and testable. However, they also emphasize the fact that the increasing numbers of gene duplication studies are revealing ever-more complex patterns. The term ‘single copy nuclear gene’ has been replaced by ‘low copy nuclear gene’ as it has become increasingly clear that both duplications and losses occur more commonly in most developmentally expressed gene families than was previously supposed (Page, 2000; Barker and Pagel, 2005; Frohlich, 2006). As well as potentially undermining the ensuing phylogenetic analyses, this perception of increasing complexity also increases the risk that true sister-genes (i.e. those derived from the most recent duplication event in the lineage) will be overlooked among the plethora of their relatives.

On the other hand, plants generally do not appear to support more than two copies of LEAFY. Some polyploids in Brassicaceae possess two copies but others rapidly lose the second copy, due to a low inherent probability of one copy being co-opted for a novel function. This could reflect either drift (Baum et al., 2005) or the unusual gradational (rather than on–off) mode of expression of LEAFY, which allows the possibility of major shifts in phenotype through epigenetic threshold effects, and permits strong selection against possession of two copies (Albert et al., 2002; Frohlich, 2006). Similarly, considerable pleiotropy is evident in LEAFY expression in eudicots. In addition to influencing the initiation and timing of the transition from vegetative to reproductive growth, LEAFY is required to render pea leaves compound but not those of tomato (reviewed by Frohlich, 2006). Within angiosperms, similarly profound phylogenetic transitions in the location and effect of expression have been observed in every well-studied family of key genes.

Gene duplications as erstwhile tools for assessing error in molecular phylogenies

One of the best-documented examples of gene duplication involves the phytochrome (PHY) gene family, which occurs as three to five copies in most angiosperms. PHYA and PHYC are monophyletic families that have subtly different roles in generating photoreceptors that influence seed germination and seedling development in most angiosperms, whereas only one related lineage occurs in extant gymnosperms, suggesting that a duplication event occurred immediately below the angiosperm crown-group node (Mathews and Sharrock, 1997; Mathews and Donoghue, 1999). The trees simultaneously generated from PHYA and PHYC by Mathews and Donoghue (2000, fig. 1) are more congruent in both topology and comparative branch lengths than those generated in most other studies of duplicate genes, but even here only slightly over half of the nodes are shared by the two topologies. Thus, even a relatively congruent result contains a large proportion of conflicting statements of relationship, in the sense that the contrasting statements cannot both reflect the species tree (though it is of course perfectly possible, and indeed likely, that neither gene tree accurately reflects the species tree).

We have not come across any studies that use the subsequent history of duplicated genes to estimate the minimum level of congruence shown by the subsequent mutational histories of the duplicates. It seems to us that these cases could be exploited far more effectively, applying to duplicate genes broadly the same logic that is applied to studies of twin children in biomedical and psychological research.

Gene duplications as erstwhile tools for cutting long branches

Cases such as the PHY duplication can, in theory, free sequence-based analyses from the tyranny of outgroup rooting, especially in cases (the majority) where the chosen outgroups are strongly divergent and thus prone to long-branch attraction. Instead, they allow duplicate gene rooting of phylogenetic trees, using one or other of two distinct philosophies. As described by Mathews and Donoghue...
Mathews and Donoghue (2000, p. S51) that ‘in effect, be more frequent than was previously believed. Accumulating data are showing each of these processes to duplication event from those that occurred subsequently. Not allow us to directly separate events on that branch (other a gene duplication (or deactivation) event on a branch does by inserting an additional node marking a divergence event a taxon to an analysis can cut a morphological long branch analogous to the way adding a molecular long branch, analogous to the way adding. Frohlich and Parker’s LEAFY study described above, sequence parsimony has effectively been prioritized over minimizing duplication/loss events). However, it is difficult to compare duplicated copies of the gene with unduplicated copies. One suggested solution, termed ‘uninode coding’ (Simmons et al., 2000), effectively reconstructs hypothetical ancestors at internal nodes of the gene tree that preclude the inferred duplication event (note that it is the ancestral gene sequence that is being reconstructed, not the ancestral morphology). Unfortunately, this approach is particularly questionable in cases where concerted evolution, lineage sorting, lateral transfer, and/or extensive duplication or losses are suspected between the duplicated gene lineages (Mathews and Donoghue, 2000; Simmons et al., 2000). Moreover, accumulating data are showing each of these processes to be more frequent than was previously believed.

Even more relevant to this essay is the assertion by Mathews and Donoghue (2000, p. S51) that ‘in effect, a duplication occurring along the branch to the ingroup would bisect [our italics] the long branch connecting the ingroup with outgroups’. As summarized by Frohlich (2006), ‘long branches can be cut by gene duplications (Mathews and Donoghue, 2000), and this may be the only way to cut long branches of crucial importance in understanding plant evolution. A lineage of organisms that persists over a long geologic interval, and does not generate any surviving sister groups during that interval, cannot have its long branch cut by additional taxon sampling of unduplicated genes. But if a gene duplication event occurs during this interval, and if both gene lineages survive to the present, then the long branch is cut at the point in time when the duplication event occurred.’ A similar argument could perhaps be made for the timing of the loss (or, more accurately, deactivation) of particular copies of key developmental genes, such as the independent losses of the Needly copy of LEAFY postulated in the angiosperm and gnetalean lineages by Frohlich and Parker (2000).

However, does this method truly ‘cut’ or ‘bisect’ a molecular long branch, analogous to the way adding a taxon to an analysis can cut a morphological long branch by inserting an additional node marking a divergence event in the species tree (Fig. 6)? Our view is that these two approaches are not truly analogous, in the sense that placing a gene duplication (or deactivation) event on a branch does not allow us to directly separate events on that branch (other than the gene duplication itself) into those that preceded the duplication event from those that occurred subsequently. The branch is not truly cut, and certainly not bisected, but rather it simply acquires an additional synapomorphy. The attractive concept of cutting the branch might arguably be rescued, at least in part, by Frohlich’s rationale (Frohlich, 2006) that ‘the long branch is cut at the point in time when the gene duplication occurred’. He predicted that many gene duplications would prove beneficial, thereby stimulating, and so immediately antedating, evolutionary radiations (cf. Bateman, 1999; Kellogg, 2006). This in turn should mean that, based on the somewhat unreliable ticking of the molecular clock, little sequence divergence should separate the duplication event from subsequent species divergences. The products are conventionally termed paralogous and orthologous gene copies, respectively; a better terminology may be in-paralogue for the products of duplication within the clade of interest and out-paralogue for duplication events that preclude the crown-group node (cf. Remm et al., 2001; O’Brien et al., 2005; Kellogg, 2006).

However, this hypothesis assumes not only that the duplication will prove immediately beneficial to the evolutionary lineage but also that a representative percentage of the products of the ensuing radiation will benefit sufficiently to survive to the present day. This assumption may have validity when, for example, it is applied to relatively recent radiations within the eudicot angiosperms, but when compared with the much greater high-level morphological diversity evident in the fossil record, the frequency of survivorship to the present of lineages from deeper radiations appears decidedly poor. Moreover, the currently preferred molecular topologies, which encapsulate monophyly of extant non-lycopsid pteridophytes and monophyly of extant gymnosperms, mean that very few of the long branches in the topology are capable of yielding critical information on angiosperm origins even if they are somehow ‘cut’. Specifically, the emphasis is placed on just two branches separating three nodes: the divergence point of extant non-lycophyte pteridophytes from extant seed plants, that of extant gymnosperms from angiosperms, and that located at the base of extant angiosperms (Fig. 2b).

Post hoc homology assessment

We conclude that a phylogeny provides a valuable guide to evolutionary-developmental studies, by focusing attention on those taxa best placed phylogenetically to allow comparative studies of key phenotypic characters and the genes that control them (e.g. Albert et al., 2005). For their part, the results of such studies constitute a simultaneous homology test of corresponding features of the organism and its genome. This is an especially valuable post hoc homology test, in that it is not dependent on the congruence test inherent in simultaneously pitting the phylogenetic signal from each character against every other character when the phylogeny is being algorithmically constructed (Patterson,
et al. Have recently become known as DNA bar-coding (e.g. Goremykin et al., 2003, 2004; Martin et al., 2005; Soltis et al., 2005)—an approach that, in its most minimalistic form oriented toward sequence-based specimen identification, has recently become known as DNA bar-coding (e.g. Tautz et al., 2003; Savolainen et al., 2005). Despite our reservations concerning some recent optimistic predictions regarding the likely effectiveness of this approach, we choose to view this debate constructively, as evidence that molecular phylogenetic analyses are becoming more sophisticated and more biologically informed. In particular, morphologists will be encouraged that, in molecular circles, data quality is making a welcome resurgence against simplistic arguments that over-emphasize data quantity.

In the course of this debate, Martin et al. (2005, pp. 208–209) argued forcefully that both character-rich (here termed ‘genomic’) and relatively taxon-rich (here termed ‘phenomic’) research programmes should progress in parallel. We heartily agree, but would further argue that a third (currently relatively poorly resourced) phylogenetic research programme—one that maximizes taxonomic sampling (‘morphologic’) remains essential, not just in the taxonomically broad studies discussed here but at all levels in the phylogenetic–taxonomic hierarchy. Only better taxon sampling can improve our understanding of the sequence of acquisition of the functional phenotypic character states that are the raw material of adaptation, or permit improved reconstructions of the hypothetical ancestors that occupy key nodes in the phylogeny. And only morphological analyses make full use of reconstructed fossil plants, benefiting from their unique combinations of characters that bridge evolutionary lacunae and their implicit statements regarding the timing of specific evolutionary events. Such dating allows estimation of the relative extents in contrasting topologies of gaps (‘ghost lineages’) in the fossil record (e.g. Doyle, 1998a), and permits calibration of molecular clocks (e.g. Won and Renner, 2003; Sanderson et al., 2004). Techniques for reconstructing hypothetical ancestors merit further exploration. Although parsimony-based optimizations remain attractive for their simplicity, there is merit in exploring likelihood-type techniques that, for example, weight towards the more common state observed in the vicinity of the node to be reconstructed (cf. Frohlich, 2006).

We would argue that the most significant advances of the last decade in understanding seed-plant relationships have not been made as a consequence of the volume of available data per se. Rather, they are the result of what has been learnt about the biological constraints on, and evolutionary implications of, each of the various categories of data that has been applied to phylogenetic problems. This trend is likely to continue through the next decade. Hopefully, the present fashion for combining different categories of data in increasingly data-rich simultaneous analyses without first analysing them separately will fade, since such analyses preclude deeper understanding of the properties of each individual data category (admittedly, few taxonomically broad studies have combined morphological and molecular data in simultaneous analyses: Nandi et al., 1998; Doyle and Endress, 2000; Pryer et al., 2001; Rothwell and Nixon, 2006). Moreover, we are more persuaded by the weight of evidence implicit in congruent topologies obtained from different categories of data experiencing different constraints. When meaningful discrepancies have been identified, we can then seek the underlying cause, and our fundamental understanding of the data is genuinely enhanced (e.g. Jeffroy et al., 2006).

Note that progressing the three phylogenetic approaches in parallel is not synonymous with progressing them in isolation. Pragmatically, better co-ordination of taxonomic sampling would be helpful, to allow several different categories of data to be gathered not only from the same higher taxon, but from the same reference individual of a single species. Then we can at least be certain that the data reflect, either directly or indirectly, the same integrated genome(s).

Even more important are the manifold opportunities for reciprocal illumination between contrasting data streams (e.g. Crane et al., 2004). For example, we are confident that whole-genome sequencing will provide invaluable yardsticks for process-based interpretations. They will help to determine the frequency of lateral gene transfer, both between the three genomes within a species and among the same category of genome between species. Such studies will usefully inform others that use a more restricted range of candidate genes, helping to optimize choices of (i) region(s) to be sequenced, (ii) approaches to filtering the resulting data, and (iii) methods of building the trees. Increasingly sophisticated chromosome mapping will
inform not only phylogeny reconstruction but also evolutionary-developmental genetics, which will act as an external homology test linking studies of candidate genes and morphogenesis (Kellogg, 2006). We concur with Frohlich’s (2006) prediction that complementation studies will demonstrate more subtlety in exploring differences in gene function than do the currently prevalent over-expression studies, and that paralogues will be more effectively compared if the range of transformable model organisms can be expanded to include species that have phylogenetic placements closer to the key divergences and that permit reciprocal transformation (Baum, 2002). Most importantly, we hope that it will become more widely accepted that studies intended to improve our understanding of the relationships of taxa without simultaneously improving our understanding of character evolution have relatively little utility.

One prediction we make with confidence is that the number of concepts broadly encompassed by the much-used term ‘homology’ will continue to increase. Concepts applied to morphology have frequently been contrasted with those applied to sequence alignment, but now we also face the challenges posed by deeper knowledge of the control of phenotype by genotype. These have long been confounded to a degree by pleiotropy and epigenetic phenomena. However, we can now demonstrate, with some confidence, cases of ‘homocracy’, where organs demonstrably share orthologous gene expression but nonetheless are inferred via phylogenies to have independent origins (e.g. K Geuten, pers. comm., 2006).

Our knowledge of molecular evolution (and the associated rapidly growing areas of bioinformatics) is insufficient to allow us to gaze insightfully deep into our crystal ball. Nonetheless, we are confident that integrating new categories of data into various aspects of phylogeny reconstruction will be one of the major challenges of the next decade. Taxonomically ever-wider surveys of expressed sequence tags offer one obvious source of data that may prove challenging to code (e.g. Kellogg, 2006), while researchers wishing to link gene phylogenies to precise functions of particular gene copies in particular species will be obliged to pay more attention to the interactions of genes in functional networks (e.g. Barker and Pagel, 2005), operating within the expanding realms of epigenetics and proteomics. Improved tests are needed for recognizing, and differentiating contrasting modes of, selection at the molecular level, along with more sophisticated modelling of the ‘ancestral’ and ‘derived’ proteins.

Returning to more traditional data sources, we suspect that the preferred methods for analysing morphological data on the one hand and molecular data on the other, which have in practice diverged substantially during the previous decade, will continue to diverge during the next decade. Parsimony will remain the preferred technique for building morphological trees (see also Rothwell and Nixon, 2006).

Both a priori homology statements and a posteriori character evolution will continue to be assessed by the analyst on a character by character basis, which permits intimate knowledge of the data. By contrast, sequence-based trees will routinely be generated from vast tracts of data via mathematical models (currently likelihood and/or Bayesian; Palmer et al., 2004). Increasingly complex algorithms and automated computerized routines will determine both prior homology assessment (i.e. alignment) and subsequent character analysis (e.g. first and second versus third positions in codons; also, sorting genomic regions—and codon positions within those regions—into contrasting mutation rate categories). In both morphological and molecular analyses, confidence estimates of nodes will increasingly utilize measures such as the decay index that, unlike resampling techniques such as the bootstrap, jackknife, and posterior probability, do not require absolute upper bounds to their values (Frohlich, 2006).

Indeed, there is a downside to rapid increases in the number of characters available, which in practice prevent the analyst from exploring the behaviour and significance of individual characters, either implicitly or using more explicit methods such as sequential character removal (Davis et al., 1993; Frohlich, 2006). Rather, specific characters routinely become lost in a statistical mélange, blurring reciprocal illumination between characters and topologies and complicating attempts to determine the functions of, and factors that control, particular phenotypic features. Certainly, when seeking to define the evasive threshold beyond which a set of gene trees can be viewed as delivering the desired species tree, we would favour use of process-based explanations of incongruence rather than a statistically based justification of perceived congruence.

Fortunately, when molecular studies address deeper issues, seeking not only inferred relationships but also the genetic changes that permit phenotypic shifts, a more intimate relationship between analyst and data is restored. Moreover, the ever-expanding volumes of sequence data available, increasingly including whole genomes (Margulies et al., 2005; Kellogg, 2006; Service, 2006), will encourage analysts to re-code those bases as amino acids. The consequent radical reduction in number of phylogenetically informative characters is more than compensated for by the greatly increased number of potential states per character (as individual characters, amino acids behave far more like morphological characters than do nucleotides).

Pluralistic approaches that are carefully integrated to address the same suite of high-profile hypotheses, among them the origin of the flower, should reap rich rewards. However, even then, a healthy level of critical evaluation should be applied to the resulting evolutionary trees. In an enjoyably provocative essay, Frohlich (2006) argued that one major problem compromising our understanding of angiosperm origins ‘is the deadening hand of wonderful, classic research. This phrasing is not oxymoronic. Superb
studies sometimes become enthroned as classics of scientific research, to the point that no one looks at the subject again for decades, although analytic tools improve and scientific questions change. I think that this has been the case in the study of the morphology of living plants.' We would make two observations in the context of this statement. Firstly, most genuinely classic morphological studies date from the nineteenth and early twentieth centuries; modern studies are far more likely to cite more recent syntheses that paraphrase, and frequently misrepresent, the earlier primary literature. In short, several recent 'discoveries' have in truth been rediscoveries. Secondly, we perceive the main advances achieved in understanding the origins of flowers over the last two decades as being more explicit means of stating and exploring hypotheses—hypotheses of relationship, hypotheses of homology, and hypotheses of character-state transitions. The botanical community has erected an invaluable framework of overlapping, and potentially integrated, analytical approaches that offers the promise of better answers to some of botany's longest-standing questions.

However, we would also argue strongly that, for the many reasons elucidated above, scepticism should be applied to the conclusions of any study that strongly advocates one particular hypothesis of relationship over its competitors, irrespective of the nature of the underlying data. Rather than attempting to 'resolve incongruence' between data sets (Rokas et al., 2003; Martin et al., 2005; Soltis et al., 2005; Jeffroy et al., 2006) (cf. Figs 4, 5), it might be better to explore its causes while heeding the cautionary note that it sounds.

A particularly apt molecular case study is the exemplary re-analysis by Jeffroy et al. (2006) of the 106-gene matrix of Rokas et al. (2003) for eight yeast species (discussed above in the context of assessing the quality of a phylogeny). Jeffroy et al. cogently argued that genome-scale matrices are likely to progressively overcome topological incongruence caused by (i) stochastic errors and (ii) violations of the assumption of orthology by processes such as gene duplication, lateral gene transfer, and lineage sorting. (Admittedly, in our view, the 'buffering' of these violations by sheer volume of data is no substitute for their accurate identification and subsequent amelioration.) Rather, Jeffroy et al. identified as most problematic systematic errors caused by heterogeneity of nucleotide composition, rate variation across lineages, and/or within-site rate variation. Together, these factors can lead to incorrect but highly statistically supported trees (Phillips et al., 2004; Philippe et al., 2005; H Philippe, pers. comm., 2006).

Most of the among-gene incongruence reported for the yeast data by Rokas et al. (2003) was attributed by Jeffroy et al. (2006) to stochastic error. However, they identified strong incongruence separating trees generated by maximum parsimony and those generated using Bayesian inference that increased when the 106 genes were concatenated. Most of this systematic incongruence was attributable to nucleotide compositional bias, which had a much stronger effect (especially under parsimony) when the matrix was coded as nucleotides rather than amino acids. This was due largely to a combination of mutational bias and mutational saturation at third-codon positions, primarily affecting transitions rather than transversions. Thus, Jeffroy et al. ultimately viewed the vast number of characters available in phylogenomic data-sets primarily as facilitating the progressive elimination of much of the initial data while still retaining sufficient information to yield a resolved and statistically robust tree. Discarding third positions, and coding sequence data as amino acids (preferably for analysis by Bayesian or likelihood methods), was the treatment recommended for large data-sets that had been accumulated to explore deep phylogenetic divergences. Increasing taxon sampling was especially recommended as a means of better detecting multiple substitutions, combined with probabilistic methods that minimize the harmful effects of model mis-specification (Steel, 2005). Thus, new statistical tests are being developed that seek to identify and quantify such mis-specification (Goremykin and Hellwig, 2006). Jeffroy et al. (2006, p. 230) understandably concluded by advocating that the phylogenetic community should place greater research 'emphasis on the development and refinement of objective methods aimed at detecting and removing the part of the data containing a high level of non-phylogenetic signal', an argument reinforced by H Philippe (pers. comm., 2006).

The alternative approach to character culling is to retain the more problematic components of the data matrix but attempt to compensate for their negative effects. One example is the analysis by Kawai and Otsuka (2004) of \(rbcL\) sequences for 13 representative extant land-plants (sadly excluding cycads and Ginkgo) plus a charophyte outgroup. Rather than discarding characters when tree-building they differentially treated contrasting categories of base position in order to accommodate third-position bias attributed to selection, via a factor termed the 'empirical relative base change probability'. This protocol resulted in identical NJ (neighbour-joining) and UPGMA (unweighted pair group method with arithmetic means) trees that yielded expected relationships (Kawai and Otsuka, 2004, fig. 4) with one notable exception: Gnetales were placed firmly below the divergence of extant conifers and angiosperms, and were calculated to have diverged from other seed plants in the earliest Permian. This topology undermines not only the anthophyte hypothesis but also its currently widely favoured replacement, the monophyletic extant gymnosperms hypothesis. Interestingly, this placement was also obtained by D Quandt and co-workers (pers. comm., 2006), based on the highly conserved P8 region of the plastid \(trnL\) intron, which suggested retention in Gnetales of the
ancestral splicing system found in bryophytes and ferns. Once again, the best defence available to proponents of these hypotheses against the Kawai and Otsuka topology may be the sparse taxon sampling evident in the study. We also caution that, contrary to frequent assertions, our inability to access the ‘one true tree’ means that topological conflicts cannot genuinely be ‘resolved’, they can only be explored.

More generally, it is clear that an experimental approach to cladistics remains highly desirable—i.e. an approach that tests the relative partia of particular combinations of taxa, characters, and topologies (e.g. Magallón and Sanderson, 2002; Rydin and Källersjö, 2002; Rydin et al., 2002; Doyle, 2006; Frohlich, 2006; Hilton and Bateman, 2006; Jeffroy et al., 2006; Rothwell and Nixon, 2006). A top-down, sequence-dominated, group-oriented perspective presently holds sway. It has proved fairly effective at exploring the eudicots, a relatively recently evolved group that consequently today retains much of the taxonomic diversity that it has possessed throughout its entire history. However, travelling backward through time to the base of the eudicots carries us only a modest percentage of the phylogenetic distance back to the earliest angiosperm. During the remainder of our top-down journey, it becomes less certain that the flowers encountered are single composite organs. The long-standing debates regarding euathial versus pseudanthial origins, most of which assume a single origin of the flower, become simplistic, suggesting that homologies are better tested (and the flower is better defined) at the lower hierarchical level of individual categories of floral organ (Endress, 2001a; Friis et al., 2006).

We believe that it is equally important to pursue a bottom-up approach to exploring the origin(s) of the angiosperms. Increasingly popular molecular hypotheses implying monophyletic extent gymnosperms and monophyletic extent non-lycopsid pteridophytes mean than any such approach must begin no later than the bryophytes, (Wellman and Grey, 2000). Patchy representation of each of these three groups in the extant flora means that the bottom-up approach requires utilization of comparative data not only from derived flower-bearing seed plants such as angiosperms and Gnetales but also from ‘living fossils’ such as cycads and Ginkgo, and from the many pertinent groups of bona fide fossils. In this context, we applaud the clarity and modernity of the classic discussion of angiosperm origins featured in Fig. 1. JH acknowledges financial support from the Nuffield Foundation (NAL/00883/G) and the Royal Society.

Acknowledgements

We are grateful to David Baum, Jim Doyle, Gar Rothwell and, especially, Mike Frohlich for offering previews of excellent manuscripts that are frequently cited in this essay, and Günter Theissen for a constructive review. We also thank Yohan Pillon for providing the material, and Mathew Box for capturing the SEM images, of the Amborella flowers featured in Fig. 1. JH acknowledges financial support from the Nuffield Foundation (NAL/00883/G) and the Royal Society.

References


De Vries H. 1906. Species and varieties: their origin by mutation, 2nd edn. Chicago, IL: Open Court.


