Opinion TRENDS in Ecology and Evolution Vol.20 No.11 November 2005

Genome duplication and the origin of angiosperms

Stefanie De Bodt*, Steven Maere* and Yves Van de Peer

Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

Despite intensive research, little is known about the origin of the angiosperms and their rise to ecological dominance during the Early Cretaceous. Based on whole-genome analyses of *Arabidopsis thaliana*, there is compelling evidence that angiosperms underwent two whole-genome duplication events early during their evolutionary history. Recent studies have shown that these events were crucial for the creation of many important developmental and regulatory genes found in extant angiosperm genomes. Here, we argue that these ancient polyploidy events might have also had an important role in the origin and diversification of the angiosperms.

The rise of the angiosperms

'An abominable mystery' is how Charles Darwin referred to the rise and early diversification of the angiosperms (flowering plants), one of the greatest terrestrial radiations that has resulted in >250 000 species (Figure 1). Since then, plant biologists have been studying the evolution of the angiosperms to understand their origin and their rise to ecological dominance. However, in spite of much research and analyses of different sources of data (e.g. fossil record and phylogenetic analyses using molecular and morphological characters), the origin of the angiosperms remains unclear.

Angiosperms appear rather suddenly in the fossil record during the Jurassic [208–145 million years ago (Mya)], with no obvious ancestors for a period of 80–90 million years before their appearance. Nevertheless, the existence during the Jurassic of all known sister taxa to the angiosperms implies that the angiosperm lineage must have been established by that time [1]. However, this ancestral lineage, coined 'angiophytes', is unlikely to be equivalent to angiosperms as known from the Cretaceous (145 Mya) through to recent forms because it might have lacked many of the characteristic angiosperm features [2]. It is presumed that angiophytes went through a period of little diversification during the Late Triassic (220 Mya) and Jurassic, either because the diversity-enhancing features, such as flowers, of the crown-group angiosperms had not yet evolved in stem angiophytes or because the diversity among angiophytes was inhibited during the Jurassic by environmental conditions or biotic interactions [2].

Evidence from the history of other major clades of land plants, such as seed ferns, suggests that the characteristic features of angiosperms were acquired sequentially through time [3]. The recent transitional–combinational theory of the angiosperm origin suggests an evolution from Jurassic seed ferns through three fundamental transitions: (i) evolution of the carpel; (ii) emergence of double fertilization; and (iii) origin of the flower [4]. The extant (or modern) angiosperms did not appear until the Early Cretaceous (145–125 Mya), when the final combination of these three angiosperm features occurred, as supported by evidence from micro- and macrofossils [4]. This combination of features might have provided the opportunity for explosive evolutionary diversification, especially in response to selection from insect pollinators and herbivores, accompanied by modifications in compatibility and breeding systems.

The fossil record provides excellent evidence for this rapid diversification in floral form during the earliest phases of recorded flowering plant history [5]. Only 10–12 million years elapsed between the first fossil records (~130 Mya) and clear documentation of all of the major lines of flowering plants [1,6]. This diversification of angiosperms occurred during a period (the Aptian, 125–112 Mya; Figure 1) when their pollen and megafossils were rare components of terrestrial floras and species diversity was low [1,6]. Angiosperm fossils show a dramatic increase in diversity between the Albian (112–99.6 Mya) and the Cenomanian (99.6–93.5 Mya) at a global scale [7–10] (Figure 1).

The angiosperm radiation yielded species with new growth architectures and new ecological roles. Early angiosperms had small flowers with a limited number of parts that were probably pollinated by a variety of insect taxa but specialized for none. Accordingly, Cenomanian flowers do not yet provide strong evidence for specialization of pollination syndromes. However, by the Turonian (93.5–89.3 Mya), flowering plants had a wide variety of features that are, in extant species, closely associated with several types of specialized insect pollination and with high species diversity within angiosperm subclades. The evolution of larger seed size in many angiosperm lineages during the early Cenozoic (from 65 Mya) indicates that animal-mediated dispersal and shade-tolerant life-history strategies had become common among angiosperms by this time [2].

Corresponding author: Van de Peer, Y. (Yves.VandePeer@psb.ugent.be).
* Contributed equally.
Available online 9 August 2005

 0169-5347/$ - see front matter doi:10.1016/j.tree.2005.07.008

Glossary

Allopolyploidy: an allo(tetra)polyploid originates by the fusion of the genomes of two different, but closely related, species.
Autopolyploidy: a polyploid in which all the chromosomes come from the same species; i.e., a tetraploid is formed by the doubling of its own genome.
Eudicots: a monophyletic clade that is strongly supported by molecular data and by a single morphological synapomorphy, namely triaperturate pollen (pollen with three grooves). This pollen type is distinct from the uniaperturate pollen of basal dicots (other dicots that are not monophyletic with the eudicots) and monocots. The eudicot clade contains most of the flowering plants, outnumbering all other plant groups put together.
Hybrid vigour: heterosis, the phenomenon whereby hybrid individuals display characteristics that exceed even the better of the two parents. Heterosis is produced by increased heterozygosity.
K_S: the number of synonymous substitutions per synonymous site. Because synonymous substitutions do not result in amino-acid replacements, the rate of fixation of these substitutions is expected to be relatively constant in different protein-encoding genes, and to reflect the overall mutation rate. The time of duplication (T) between two sequences can then be calculated as $T = K_S/2\lambda$, where K_S is the fraction of synonymous substitutions per synonymous site, and λ is the mean rate of synonymous substitution.
Monocots: the number of cotyledons found in the embryo is the basis for distinguishing the two classes of flowering plants, and is the source of the names monocots and dicots. The cotyledons are the seed leaves produced by the embryo and serve to absorb nutrients packaged in the seed until the seedling is able to produce its first true leaves and begins photosynthesizing. The monocots comprise approximately one-quarter of all flowering plant species and include some of the largest and most familiar groups of plants, such as lilies, orchids, agaves, palms and grasses. Some monocots, such as corn, rice, wheat, and barley, are among our most important food crops. Sugar cane, pineapples, dates, bananas, and many of our familiar tropical fruits also come from monocots.
Paranome: collection of all duplicated genes in a genome.
Polyploid: a polyploid organism has more than two sets of chromosomes.
Stoichiometric quantities (of protein complex partners): quantities of protein complex partners that reflect their relative abundance in the complex.

In summary, fossils with affinities to diverse angiosperm lineages, including monocots, are all found in Early Cretaceous floras [5,8,11] and ~42 of the 94 extant orders (44%) and ~63 of the 439 extant families (14%) of flowering plants occurred for the first time during the Cretaceous [2]. However, the question remains why this was such a decisive time in the evolution of plants. Here, we discuss why whole-genome duplication events might have had a key role in the origin of angiosperms and their morphological and ecological diversification.

Of polyploids and paleopolyploids

One of the most striking features of angiosperms is that many have experienced one or more episodes of polyploidy (see Glossary) in their ancestry [12]. Apart from species that are currently polyploid, such as many crops, others are considered to have paleopolyploid genomes. In 1996, when the sequencing of the flowering plant *Arabidopsis thaliana* (Brassicaceae) genome began, this model plant, with its small genome, was not expected to be an ancient polyploid. However, five years after the release of its genome sequence [13], there is compelling evidence that the *Arabidopsis* genome, or rather that of its ancestors, has been duplicated three times (events referred to here as 1R, 2R and 3R) during the past 250 million years [14,15].

Although duplicated genes and genomes can provide the raw material for evolutionary diversification and the functional divergence of duplicated genes might offer a selective advantage to polyploids over a long time period, a beneficial effect of these duplications is assumed shortly after the duplication event. In other words, if a new genome doubling is to survive long enough to exert its long-term evolutionary effects, it must provide an immediate selective advantage that enables its establishment (Box 1; reviewed in [16]).

The fate of duplicated genes

Ancient polyploidy events might have directly influenced the increase in the number of plant species and plant complexity observed since the Early Cretaceous (Box 1; Figure 1). However, other factors, such as expansion and functional diversification of specific gene families following a polyploidy event, are likely to have been more influential and could explain, at least in part, the origin and fast diversification of angiosperm lineages. Several authors have provided evidence that gene retention after duplication is biased according not only to the function of the genes, but also to the timing and mode of the duplication events [17–19]. Blanc and Wolfe [18] studied the relationship between gene function and duplicate loss after the most recent polyploidy event (3R). Similarly, Seoighe and Gehring [19] analyzed the survivability of duplicates of various functions following 3R. Recently, Maere *et al.* [17] developed an evolutionary model that simulates the population dynamics of duplicate genes belonging to different functional categories, based on the K_S distribution of the *Arabidopsis* paranome. They took into account the three major genome-wide duplication events (1R, 2R and 3R) and a continuous mode of small-scale gene duplications (referred to as 0R). These studies all concluded that genes involved in transcriptional regulation and signal transduction have been preferentially retained following genome duplications. Similarly, developmental genes have been observed to be retained following genome duplications [17,18], particularly following the two oldest events (1R and 2R) [17] (Box 2).

However, few regulatory and developmental gene duplicates appear to have survived small-scale duplication events [17]. Their rapid loss can be explained by the fact that transcription factors and genes involved in signal transduction tend to show a high dosage effect in multicellular eukaryotes [20]. The expression of a wide range of genes regulated by these proteins show major perturbations when only one regulatory component is duplicated, rather than all components that govern a certain pathway [21,22]. Furthermore, that transcription factors and kinases are often active as protein complexes and must be present in stoichiometric quantities for their correct functioning is congruent with their high retention rate following whole-genome instead of small-scale duplication events [22,23].

Regulatory and developmental genes are thought to have been of primordial importance for the evolution of morphological complexity in plants and animals [24–26]. Overall, the three polyploidy events in the ancestors of *Arabidopsis* might have been responsible for >90% of the transcription factors, signal transducers and developmental genes created during the past ~250 million years [17]. Although this is only an approximate assessment from a single study, it is tempting to suggest that if only small-scale gene duplications had occurred in the evolutionary past of *Arabidopsis* and of angiosperms in

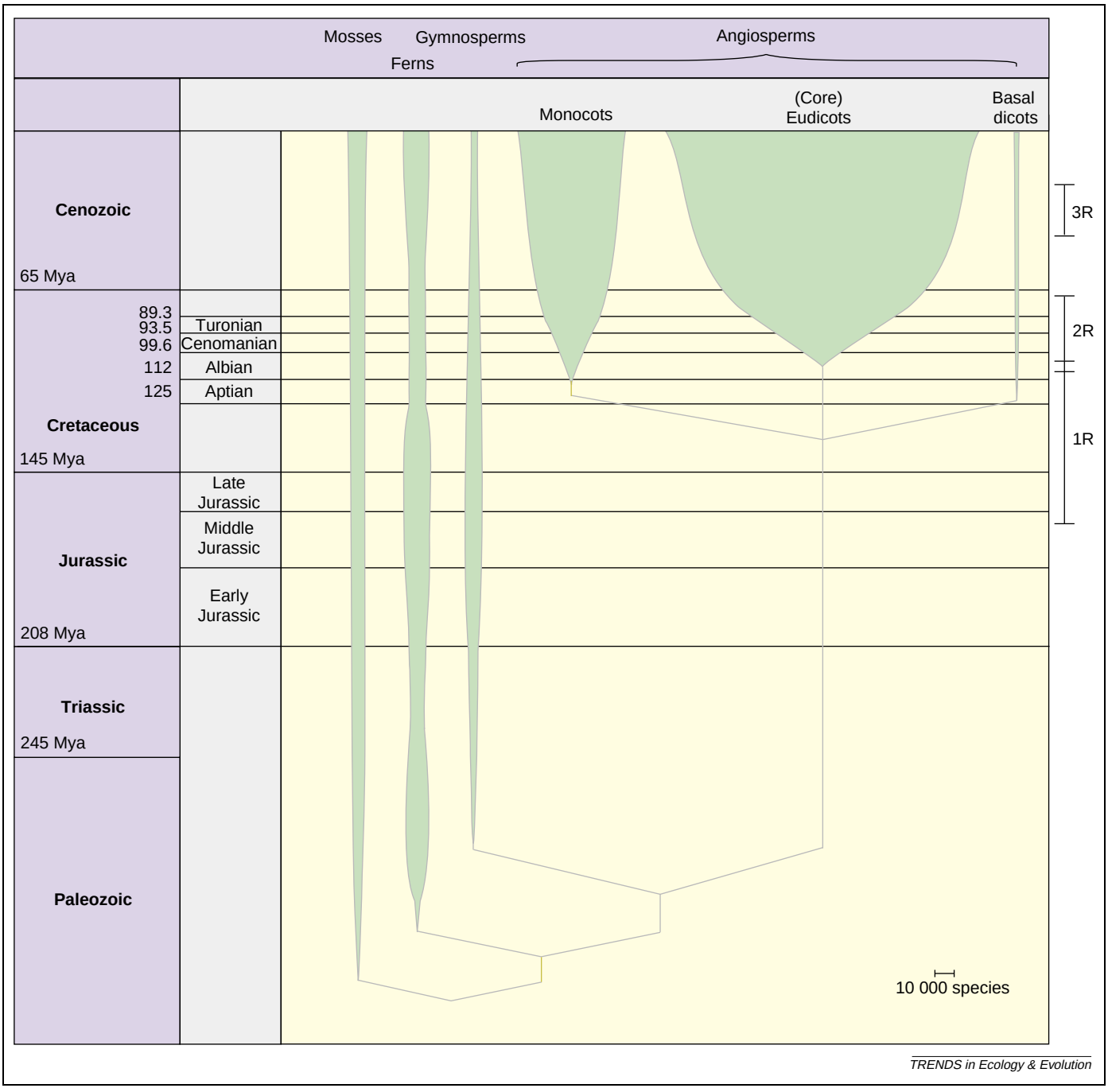

Figure 1. Simplified view of the rise and diversification of the mosses, ferns, gymnosperms and angiosperms, as inferred from fossil data [10,74]. 'Basal dicots' refers to the primitive dicotyledons, such as the magnoliids, whereas (core) eudicots refers to the monophyletic grouping of the other dicotyledonous plant families. The three whole-genome duplication events (1R, 2R and 3R), for which evidence can be found in the *Arabidopsis* genome, are also indicated.

general, the expansion of regulatory and developmental genes would have been severely hampered.

However, according to Maere *et al.* [17], genes involved in secondary metabolism or responses to biotic stimuli, such as pathogen attack, tend to be preserved regardless of the mode of duplication. Because plants are sessile organisms, secondary metabolite pathways, as well as genes governing responses to biotic stimuli, are crucial to the development of survival strategies against herbivores, insects, snails and plant pathogens [27]. Additionally, in angiosperms, anthocyanins and other secondary metabolites give rise to colourful and scented flowers that attract pollen- and nectar-collecting animals [27]. Thus, secondary metabolite diversification might have led to more efficient seed dispersal (compared with wind pollination, which is widespread in most seed plants) and might have provided new possibilities for reproductive isolation and the elevation of speciation rates [6]. Moreover, coevolutionary interactions between early angiosperms and, in particular, insect pollinators are thought to have been of major importance for the angiosperm radiation [28,29] (Box 3). The finding that genes involved in secondary metabolism and responses to biotic stimuli are also strongly retained following (continuously occurring)

Box 1. Polyploidy, adaptation and speciation

Genome duplication events double the amount of genetic material on which evolution can work and the functional divergence of duplicated genes is considered important for biological evolution and increases in biological complexity [33–35]. However, apart from duplicated genes adopting novel functions, polyploidy events can also contribute to evolution in other ways.

Polyploidy and hybrid vigour

Differences in phenotype between polyploids and their diploid progenitors can be caused by increased variation in dosage-regulated gene expression [45]. By having a different number of alleles at a locus, polyploid plants can differ from their diploid ancestors in overall gene expression levels. For genes having allele-dosage effects, polyploidy increases the potential variation in expression levels and can provide selective advantages [46]. For instance, changes in enzyme levels can affect many aspects of plant physiology, morphology and life history [47]. Because the combination of different genomes (or the increase in heterozygosity) can lead to hybrid vigour, the newly formed polyploid can have a selective advantage compared with closely related diploid organisms. In accordance with Spring [48], Rieseberg *et al.* [49] suggest that hybridization provides a mechanism for large and rapid adaptive transitions, made possible by the genetic variation of hundreds or thousands of genes in a single generation.

Polyploidy and sympatric speciation

Hybridization is a significant evolutionary force that creates opportunities for adaptive evolution and speciation [50–52]. It is estimated that 2–4% of the speciation events in flowering plants can be attributed to ploidy changes that have potentially broad-scale effects on gene regulation and developmental processes, effects that can produce immediate shifts in morphology, breeding system and ecological tolerances [44]. Increased cell volume is one of the most common and universal phenotypic effects of polyploidization (e.g. [53]). Changes in cell volume result in changes in surface: volume ratios, which can alter the rate of metabolic processes. Consequently, growth and developmental rates are often affected in polyploid plants.

Whereas, in general, developmental rates are lower for polyploid plants than for their diploid progenitors, larger seed sizes can have an opposite effect on the rate of (early) development and might affect the likelihood of establishing seedlings in resource-limited environments. They might also result in niche differentiation as a byproduct of polyploidization [44]. Differences in vegetative traits, such as ecological tolerance (e.g. drought), and susceptibility to arthropod infestation and fungal diseases, have been documented between polyploid and diploid plants [44,52]. In addition, several reproductive traits differ significantly, such as the initiation and duration of flowering time, fertility, self-compatibility, germination and the relative sizes and spatial relations of floral organs. Again, polyploids that persist are likely to be able to inhabit niches that are different from those of their diploid progenitors, which can then result in speciation [44].

Polyploidy and allopatric speciation

As well as its role in sympatric speciation, polyploidy can also have a role in allopatric speciation. Based on isozyme studies, Werth and Windham [54] developed a model in which the reciprocal silencing of genes in geographically separated populations promotes speciation, an idea that was revived in a model called 'divergent resolution' [55]. In this model, the loss or silencing of gene duplicates was postulated to be more important for the evolution of species diversity than was the acquisition of new functions by duplicated genes. Divergent resolution occurs in allopatric populations when different copies of duplicated genes are lost from different chromosomes, thereby creating genetic barriers to reproduction [55,56].

small-scale gene duplications might reflect the continuous interaction between plants and animals, fungi or plant pathogens imposing a constant need for adaptation. By contrast, genes involved in responses to abiotic stress, such as drought, cold and salinity, appear to have been only moderately retained after small-scale gene duplication events [17], indicating that they might have been required at more specific times in evolution, such as during major environmental changes or adaptation to new niches. Interestingly, 1R and 2R might have occurred during a period of increased tectonic activity linked to highly elevated atmospheric CO_2 levels [29] (Boxes 2,3).

The hypotheses outlined here should be approached with caution, as they are based on a few computational studies on a single flowering plant, *Arabidopsis thaliana* [17–19]. Similar analyses on other flowering plants are needed to accept or reject these hypotheses.

Darwin's abominable mystery revisited

Here, we propose that the ancient polyploidy events that occurred during early angiosperm evolution (Box 2) have created much of the genetic material found in extant angiosperms. In particular, genes involved in specific processes, such as development, transcriptional regulation and signalling, appear to have been retained in this manner [17–19]. It is therefore tempting to speculate that much of the evolutionary success of polyploidy in angiosperms, which entails a saltatory doubling of regulatory factors, is tied directly to regulatory gene evolution, both at the protein level and through the differential regulation of the regulatory genes themselves [30].

Proliferation of transcription factor genes, followed by the recruitment of these genes as upstream or downstream regulators in developmental pathways, is believed to have contributed substantially to the evolution of morphological diversity in animals and plants [24–26]. Furthermore, transcriptional regulators, activated by signal transducers, can act as key switches in plant development [24,31,32]. Subject to diverse selective pressures, these extra genes might have subsequently evolved novel functions (e.g. [33]) that resulted in major changes in biological complexity [34,35]. The first genome duplication event appears to have occurred during the Late Jurassic or Early Cretaceous, a period during which the angiosperms originated and rose to ecological dominance [36–38] (Figure 1; Box 2). The high retention of duplicated genes underlying transcriptional regulatory networks and development after 1R suggests that this event provided the additional genetic material to produce the observed increase in plant complexity. For instance, co-sexual flowers originated, and major developmental changes in leaf morphology occurred [6,39]. The occurrence of 1R before the monocot–eudicot split appears to correlate well with this decisive moment in plant evolution [15] (Box 2).

According to the dating described in Box 2, 2 R occurred ~66–109 Mya, probably after the divergence of monocots and eudicots [15], in a time period when flowering plants developed a wide variety of specialized insect pollination strategies and where species diversity within the angiosperm subclades increased rapidly [28]. The magnitude of

Box 2. Inferring the age of genome duplications

The timing of 1R, 2R, and 3R in the evolutionary past of *Arabidopsis thaliana* is still debated. Based on phylogenetic analysis, Bowers *et al.* [15] concluded that 3R occurred 14.5–86 Mya, after the divergence of Brassicales and Malvales, but before the divergence of *Arabidopsis* and *Brassica*. Blanc *et al.* [41] narrowed this to 24–40 Mya. However, the age of the older genome duplications is less clear. Arguably, 2R occurred after the divergence of monocots and eudicots, whereas 1R appears to predate the monocot–eudicot divergence but could not be positioned relative to the angiosperm–gymnosperm divergence owing to the sparseness of appropriate sequence data [15,37]. Based on simulation of the K_S distribution of the *Arabidopsis* paranome, Maere *et al.* [17] recently proposed K_S values for 3R (K_S=0.7–0.8), 2R (K_S=2.0–2.1) and 1R (K_S=3.1–3.2).

One way of relating these K_S estimates to time is to use molecular clocks, or synonymous substitution rate estimates, such as that of Koch *et al.* [57] (1.5×10^{-8} ss site^{-1} year^{-1}) or Lynch and Conery [58] (6.1×10^{-9} ss site^{-1} year^{-1}) (grey lines in Figure I). However, there are theoretical and empirical concerns about the accuracy of molecular clocks [59–62], exemplified by the large difference in the clock rate estimates of Koch *et al.* [57] and Lynch and Conery [58]. Some of the major issues are rate heterogeneity in and between lineages caused by evolutionary factors (e.g. generation time), difficulties in interpreting the fossil data used to calibrate the clock, and rate variation among genes, even at synonymous sites. Because rates of nuclear gene evolution vary widely within an organism [63,64], sampling only a few genes in molecular clock calibrations can be risky.

To overcome at least some of these limitations, we could use the K_S estimate for 3R (K_S~0.75) to calibrate the clock. By avoiding inter-lineage comparison, and by the many duplicates involved, the K_S estimates for large-scale duplications should be more representative of the average rate of synonymous substitution in *Arabidopsis* and its ancestors. Because 3R is the most recent genome duplication event, it is the most obvious choice. There is still a large population of retained duplicates from 3R compared with 1R and 2R, and 3R can be more accurately positioned phylogenetically. Unfortunately, no accurate fossil calibration point for 3R is currently available and the 3R age intervals proposed by Blanc *et al.* [41] and Bowers *et al.* [15] are themselves dependent on a molecular clock [65]. A more-accurate dating of the 3R event would require a better phylogenetic positioning of the 3R event on the tree of life (i.e. a better sampling of the Brassicales) and an adequate fossil calibration point. Here, we used the Blanc *et al.* [41] interval (3R occurred 24–40 Mya) as a best guess and extrapolated the resulting relationship to obtain estimated age intervals for 1R and 2R. The 1R interval (101–168 Mya) overlaps with the occurrence of the first *bona fide* angiosperm fossils, approximately 130 Mya [4], whereas the 2R interval (66–109 Mya) is centred around the Turonian (89.3–93.5 Mya). Interestingly, the fossil record shows a dramatic modification of angiosperm flowers at about this time [28,66]. Additionally, the suggested time interval for 2R roughly corresponds to the time period during which the core eudicots emerged and radiated [9,67].

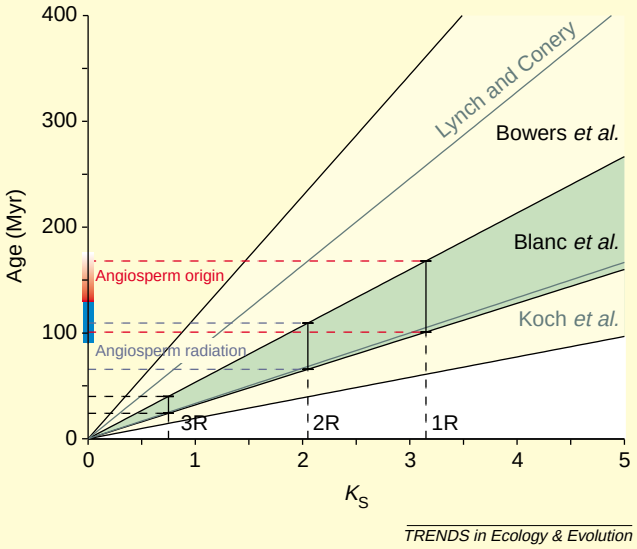

Figure I. Estimated ages for 1R, 2R and 3R, based on K_S data for *Arabidopsis thaliana* [17]. The grey lines depict the relationship between time and K_S, based on the molecular clock estimates of Koch *et al.* [57] and Lynch and Conery [58]. The beige and green areas represent the K_S versus time window extrapolated from the 3R age interval proposed by Bowers *et al.* [15] and Blanc *et al.* [41] respectively, using K_S=0.75 as the average K_S value for 3R duplicates [17].

these events was phrased aptly by Crepet [28]: 'Had Darwin witnessed such a pattern, he might have been even more astonished by the rapid ascension of the flowering plants.' Would he have considered it a second abominable mystery? 2R might have provided, at least in eudicots, the raw material on which evolution could act to produce the observed specialized floral phenotypes with insects providing the necessary selective forces [6]. In support of these theories, Zahn *et al.* [40] recently provided evidence that, on two separate occasions, several duplication events have occurred simultaneously in certain subfamilies of MADS-box transcription factors, known to be important in floral development. Both of these duplication bursts, one supposedly close to the origin of angiosperms, and the other in the ancestor of the core eudicots, might be linked to genome duplication events [40]. Given the role of MADS-box genes in flower development, these gene duplications, possibly linked to 1R and 2R, might have had an important role in the morphological invention of the flower and subsequent diversification of floral forms [40].

Developmentally related genes were found to be less strongly, although significantly, retained after the youngest genome duplication event [17,18], dated before the divergence of *Arabidopsis* and *Brassica* and after the divergence of Brassicales and Malvales [15,41], possibly close to the emergence of the crucifers [41]. Compared to 1R and 2R, gene retention following 3R is lower and less biased towards functional class [17]. This does not imply that the youngest genome duplication did not contribute to plant evolution, but it might have done so in a different and less dramatic way. For instance, the youngest genome duplication in *Arabidopsis* has created many duplicates [14,15] that could be divergently resolved (Box 1), and such genes might have had a prominent role in the radiation of the crucifer family, which comprises >3500 species [42].

Conclusion

It has been suggested that large-scale gene duplication or whole-genome duplication events can be associated with important evolutionary transitions, major leaps in development, and/or adaptive radiations of species [34,35,43,44]. So far, evidence linking major duplication events with evolution or biological innovation has been scarce. However, because *Arabidopsis* has undergone

Box 3. Dinosaurs, bugs and CO_2

Alternative scenarios have been put forward to explain the origin of the angiosperms and their subsequent tempo and pattern of diversification, most of which could be compatible with the genome duplication theory proposed here. For instance, some authors (e.g. [68]) have speculated that the origin and diversification of angiosperms was mediated by changes in the browsing behavior of dinosaurs (reviewed in [29]). Moreover, the ecological radiation of angiosperms has been associated with the evolution of complex jaw mechanisms, such as pleurokinesis (sideway chewing movement) among ornithischian dinosaurs such as ceratopsians and stegosaurs. However, others have refuted the idea of dinosaurs being causative agents in the origin of angiosperms [29].

Insect pollinators are thought to have had a much larger role in angiosperm diversification. During the Cretaceous, the development of specialized floral features, indicating a high level of pollinator specificity [66], seems to coincide well with the origin and diversification of many insect clades, such as lepidopterans, hymenopterans and dipterans. It is now widely accepted that coevolution between early angiosperms and anthophilous insects has been of major importance for the evolution of angiosperms [69]. Furthermore, insect herbivores such as aphids and beetles might have also had an important role in many of the adaptive radiations that account for the present diversity of flowering plants (and also beetles) [70,71].

Barrett and Willis [29] suggest that the diversification of angiosperms might have been the direct result of a major geological event that led to increased levels of atmospheric CO_2. During the past 500 million years, there have been five periods of increased tectonic activity, the most recent of which occurred between 120 Mya and 80 Mya, at the time of the final break-up of Pangaea. Such periods of increased activity are associated with significant changes in atmospheric CO_2 levels owing to increased volcanic activity [29,72]. In turn, elevated CO_2 levels could increase the ability of plants to colonize drier and/or more nutrient-poor sites, thereby increasing the range of niches available for plant growth and, thus, the opportunities for the emergence of novel plant forms [29]. Apart from the colonization of new habitats, increased levels of CO_2 could also have an effect on the rate of turnover of plant generations [73]. If generation turnover rates increase, the rate of speciation is also expected to increase. Therefore, as stated by Barrett and Willis [29], higher levels of CO_2 might have increased the evolutionary and selective pressures on plants owing to the induction of (abiotic) stress from which plants cannot escape; evolving to cope with the changing environment might have been the only option for plants.

several well-documented rounds of genome duplication, it is an ideal model system with which to study the effect of such events. As has been shown in recent studies, the increase in the number of genes that are important in development, transcriptional regulation and signalling has been mainly the result of large-scale gene duplications; these genes might not have been retained in small-scale duplication events [17]. Given that such genes are considered important for introducing phenotypic variation and increase in biological complexity, linking ancient polyploidy events with decisive moments in evolution becomes less speculative and the origin and evolution of angiosperms perhaps less of a mystery.

Acknowledgements

S.D.B. thanks the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT) for a predoctoral fellowship. S.M. is a Research Fellow of the Fund for Scientific Research (FWO). The authors also acknowledge the constructive comments of three anonymous referees.

References

1 Doyle, J.A. and Donoghue, M.J. (1993) Phylogenies and angiosperm diversification. *Paleobiology* 19, 141–167
2 Wing, S.L. and Boucher, L.D. (1998) Ecological aspects of the cretaceous flowering plant radiation. *Annu. Rev. Earth Planet Sci.* 26, 379–421
3 Friis, E.M. *et al*. (2005) When Earth started blooming: insights from the fossil record. *Curr. Opin. Plant Biol.* 8, 5–12
4 Stuessy, T.F. (2004) A transitional–combinatorial theory for the origin of angiosperms. *Taxon* 53, 3–16
5 Friedman, W.E. *et al*. (2004) The evolution of plant development. *Am. J. Bot.* 91, 1726–1741
6 Crane, P.R. *et al*. (1995) The origin and early diversification of angiosperms. *Nature* 374, 27–33
7 Lidgard, S. and Crane, P.R. (1990) Angiosperm diversification and Cretaceous floristic trends; a comparison of palynofloras and leaf macrofloras. *Paleobiology* 16, 77–93
8 Crane, P.R. and Lidgard, S. (1989) Angiosperm diversification and paleolatitudinal gradients in Cretaceous floristic diversity. *Science* 246, 675–678
9 Magallón-Puebla, S. *et al*. (1999) Phylogenetic pattern, diversity and diversification of eudicots. *Ann. Mo. Bot. Gard.* 86, 297–372
10 Crane, P.R. *et al*. (2004) Fossils and plant phylogeny. *Am. J. Bot.* 91, 1683–1699
11 Friis, E.M. *et al*. (2000) Reproductive structure and organization of basal angiosperms from the Early Cretaceous (Barremian or Aptian) of Western Portugal. *Int. J. Plant Sci.* 161, S169–S182
12 Tate, J.A. *et al*. (2005) Polyploidy in plants. In *The Evolution of the Genome* (Gregory, T., ed.), pp. 371–426, Elsevier, Academic Press
13 The *Arabidopsis* Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
14 Simillion, C. *et al*. (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13627–13632
15 Bowers, J.E. *et al*. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438
16 Van de Peer, Y. and Meyer, A. (2005) Large-scale gene and ancient genome duplications. In *The Evolution of the Genome* (Gregory, T., ed.), pp. 330–368, Elsevier, Academic Press
17 Maere, S. *et al*. (2005) Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5454–5459
18 Blanc, G. and Wolfe, K.H. (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16, 1679–1691
19 Seoighe, C. and Gehring, C. (2004) Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* 20, 461–464
20 Birchler, J.A. *et al*. (2001) Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev. Biol.* 234, 275–288
21 Amoutzias, G.D. *et al*. (2004) Convergent networks by single-gene duplications in higher eukaryotes. *EMBO Rep.* 5, 274–279
22 Papp, B. *et al*. (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197
23 Krylov, D.M. *et al*. (2003) Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13, 2229–2235
24 Kellogg, E.A. (2004) Evolution of developmental traits. *Curr. Opin. Plant Biol.* 7, 92–98
25 Carroll, S.B. (2000) Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101, 577–580
26 Gibson, G. and Honeycutt, E. (2002) The evolution of developmental regulatory pathways. *Curr. Opin. Genet. Dev.* 12, 695–700
27 Chen, F. *et al*. (2003) Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell* 15, 481–494
28 Crepet, W.L. (2000) Progress in understanding angiosperm history, success, and relationships: Darwin's abominably 'perplexing phenomenon'. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12939–12941
29 Barrett, P.M. and Willis, K.J. (2001) Did dinosaurs invent flowers? Dinosaur–angiosperm coevolution revisited. *Biol. Rev. Camb. Philos. Soc.* 76, 411–447

30 Grotewold, E. (2005) Plant metabolic diversity: a regulatory perspective. *Trends Plant Sci.* 10, 57–62
31 Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell* 10, 1075–1082
32 Tautz, D. (2000) Evolution of transcriptional regulation. *Curr. Opin. Genet. Dev.* 10, 575–579
33 Taylor, J.S. and Raes, J. (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 9, 615–643
34 Aburomia, R. *et al*. (2003) Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J. Struct. Funct. Genom.* 3, 45–52
35 Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag
36 Chaw, S.M. *et al*. (2004) Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* 58, 424–441
37 Chapman, B.A. *et al*. (2004) A comparative phylogenetic approach for dating whole-genome duplication events. *Bioinformatics* 20, 180–185
38 Davies, T.J. *et al*. (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1904–1909
39 Cronk, Q.C. (2001) Plant evolution and development in a post-genomic context. *Nat. Rev. Genet.* 2, 607–619
40 Zahn, L.M. *et al*. (2005) The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223
41 Blanc, G. *et al*. (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* 13, 137–144
42 Cronquist, A. (1981) *An Integrated System of Classification of Flowering Plants*, Columbia University Press
43 Wendel, J.F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249
44 Otto, S.P. and Whitton, J. (2000) Polyploid incidence and evolution. *Annu. Rev. Genet* 34, 401–437
45 Guo, M. *et al*. (1996) Dosage effects on gene expression in a maize ploidy series. *Genetics* 142, 1349–1355
46 Osborn, T.C. *et al*. (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19, 141–147
47 Thompson, J.N. *et al*. (1997) Plant polyploidy and insect/plant interactions. *Am. Nat.* 150, 730–743
48 Spring, J. (2003) Major transitions in evolution by genome fusions: from prokaryotes to eukaryotes, metazoans, bilaterians and vertebrates. *J. Struct. Funct. Genom.* 3, 19–25
49 Rieseberg, L.H. *et al*. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216
50 Ehrendorfer, F. (1980) Polyploidy and distribution. In *Polyploidy – Biological relevance* (Lewis, W.H., ed.), pp. 45–60, Plenum Press
51 Arnold, M.L. (1997) *Natural Hybridization and Evolution*, Oxford University Press
52 Ramsey, J. and Schemske, D.W. (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501
53 Masterston, J. (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264, 421–423
54 Werth, C.R. and Windham, M.D. (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* 137, 515–526
55 Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473
56 Taylor, J.S. *et al*. (2001) Genome duplication, divergent resolution and speciation. *Trends Genet.* 17, 299–301
57 Koch, M.A. *et al*. (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498
58 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
59 Soltis, P.S. *et al*. (2002) Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4430–4435
60 Strauss, E. (1999) Can mitochondrial clocks keep time? *Science* 283, 1435, 1437–1438
61 Senchina, D.S. *et al*. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol. Biol. Evol.* 20, 633–643
62 Sanderson, M.J. *et al*. (2004) Molecular evidence on plant divergence times. *Am. J. Bot.* 91, 1656–1665
63 Gaut, B. (1998) Molecular clocks and nucleotide substitution rates in higher plants. In *Evolutionary Biology* (Hecht, M.K., ed.), pp. 93–120, Plenum Press
64 Zhang, L. *et al*. (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in Arabidopsis thaliana. *Mol. Biol. Evol.* 19, 1464–1473
65 Yang, Y.W. *et al*. (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *J. Mol. Evol.* 48, 597–604
66 Crepet, W. (1996) Timing in the evolution of derived floral characters: Upper Cretaceous (Turonian) taxa with tricolpate and tricolpate-derived pollen. *Rev. Palaeobot. Palynol.* 90, 339–359
67 Wikström, N. *et al*. (2001) Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. London B Biol. Sci.* 268, 2211–2220
68 Bakker, R.T. (1978) Dinosaur feeding behaviour and the origin of flowering plants. *Nature* 274, 661–663
69 Grimaldi, D.A. (1999) The co-radiations of pollinating insects and angiosperms in the Cretaceous. *Ann. Mo. Bot. Gard.* 86, 373–406
70 Farrell, B.D. (1998) 'Inordinate fondness' explained: why are there so many beetles? *Science* 281, 555–559
71 Wilf, P. *et al*. (2000) Timing the radiations of leaf beetles: hispines on gingers from the latest Cretaceous to Recent. *Science* 289, 291–294
72 Sheridan, R.E. (1997) Pulsation tectonics as a control on the dispersal and assembly of supercontinents. *J. Geodynam.* 23, 173–196
73 Phillips, O.L. and Gentry, A.H. (1994) Increasing turnover through time in tropical forests. *Science* 263, 954–958
74 Friis, E.M. *et al*. (2004) Araceae from the Early Cretaceous of Portugal: evidence on the emergence of monocotyledons. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16565–16570