Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event

Jeffrey A. Fawcett[a,b,1], Steven Maere[a,b,1], and Yves Van de Peer[a,b,2]

[a]Department of Plant Systems Biology, Flanders Institute for Biotechnology, 9052 Gent, Belgium; and [b]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium

Communicated by Marc C. E. Van Montagu, Ghent University, Ghent, Belgium, January 31, 2009 (received for review October 21, 2008)

Most flowering plants have been shown to be ancient polyploids that have undergone one or more whole genome duplications early in their evolution. Furthermore, many different plant lineages seem to have experienced an additional, more recent genome duplication. Starting from paralogous genes lying in duplicated segments or identified in large expressed sequence tag collections, we dated these youngest duplication events through penalized likelihood phylogenetic tree inference. We show that a majority of these independent genome duplications are clustered in time and seem to coincide with the Cretaceous–Tertiary (KT) boundary. The KT extinction event is the most recent mass extinction caused by one or more catastrophic events such as a massive asteroid impact and/or increased volcanic activity. These events are believed to have generated global wildfires and dust clouds that cut off sunlight during long periods of time resulting in the extinction of ≈60% of plant species, as well as a majority of animals, including dinosaurs. Recent studies suggest that polyploid species can have a higher adaptability and increased tolerance to different environmental conditions. We propose that polyploidization may have contributed to the survival and propagation of several plant lineages during or following the KT extinction event. Due to advantages such as altered gene expression leading to hybrid vigor and an increased set of genes and alleles available for selection, polyploid plants might have been better able to adapt to the drastically changed environment 65 million years ago.

angiosperms | eudicots | Cretaceous–Tertiary boundary | penalized likelihood | polyploidy

Genome-wide analyses provide overwhelming evidence that plants have undergone one or more whole genome duplications (WGD) in their evolutionary past. For instance, extensive studies of the *Arabidopsis thaliana* genome sequence unveiled the remnants of 3, possibly 4, rounds of genome duplications (1–6). At least 2 genome doublings have been proposed in poplar (*Populus trichocarpa*) (7). Legumes, represented by *Medicago* and *Lotus*, show evidence of several rounds of duplication (8, 9), whereas 3 ancestral genomes have been proposed to contribute to the *Vitis* lineage (5). One or two WGDs have also been proposed in rice (*Oryza sativa*) (10). Although the number and timing of these genome duplications are still being vividly debated, there seems to be a growing consensus that the oldest duplication events occurred early in angiosperm evolution (1, 5). As a matter of fact, it has been speculated that these older genome duplications might have been at least partly responsible for the origin and fast diversification of the angiosperms (11, 12). Strikingly, besides these older events shared by many (if not most) flowering plants, many plants whose genome (or large parts thereof) has been sequenced show evidence for an additional, independent, and more recent genome duplication. Moreover, such younger large-scale gene duplication events have been suggested for several plant species whose genome sequence is not available but large EST collections exist (13–16).

Here, we investigated the timing of these younger genome duplication events in plants by using phylogenetic approaches. Table 1 provides a list of monocot and eudicot plant species in which a large-scale duplication event has been inferred from their genome sequence or large EST collections. Although large-scale duplication events have been uncovered in many different plant lineages, determining the timing of such events is not self-evident. One of the most common methods used is to build age distributions of paralogs, where the number of duplicates is plotted against their age, inferred from the number of synonymous substitutions per synonymous site (K_S). Peaks in the distribution reflect sudden bursts in the number of new genes and are, therefore, considered evidence for large-scale gene or entire genome duplications (Fig. 1). When the rate of synonymous substitutions is known, the K_S values can be converted to absolute ages. For instance, by assuming a rate of 6.1×10^{-9} synonymous substitutions per synonymous site per year, Lynch and Conery (17) dated the youngest genome duplication in *Arabidopsis* at ≈65 million years ago (mya) (see also *SI Text*). By using the same substitution rate applied to large EST collections, Schlueter et al. (16) inferred large-scale duplications between 50 and 60 mya for several eudicots, whereas WGDs in rice and other cereals were also dated between 50 and 60 mya (see Table 1) by using a slightly different substitution rate (6.5×10^{-9}). Peaks in age distributions at similar K_S values have been found for other plant species as well (14, 15, 18, 19). However, synonymous substitution rates are often unknown for the species of interest and they are well-known to vary considerably across lineages and over time (6, 7). For instance, using the above mentioned substitution rate of the weed *Arabidopsis* to time the WGD in the tree *Populus* yielded a date between 8 and 13 mya, whereas it was later suggested that the WGD event shortly predated the split of *Populus* and *Salix*, estimated at ≈60 mya (7).

Estimating the divergence or duplication time of sequences in a phylogenetic tree has been the subject of much research, and various methods have been developed to account for rate variations across branches (20–22). Here, we inferred the absolute ages of the youngest WGDs in plants through a phylogenetic approach. We calculated the divergence dates of all putatively WGD-derived paralogs by using the penalized likelihood (PL) method, which accounts for rate variation between lineages through a semiparametric smoothing approach that penalizes rates that vary too much across a phylogeny (22). We show that the independent WGDs are not randomly distributed in time but instead cluster around the Cretaceous–Tertiary (KT) boundary

Author contributions: S.M. and Y.V.d.P. designed research; J.A.F. and S.M. performed research; J.A.F. and S.M. analyzed data; and J.A.F., S.M., and Y.V.d.P. wrote the paper.

The authors declare no conflict of interest.

[1]J.A.F. and S.M. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: yves.vandepeer@psb.vib-ugent.be.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900906106/DCSupplemental.

EVOLUTION

Table 1. Timing of WGD events in plants based on molecular clock rate (K_S) estimates and the construction of phylogenetic trees

Organism	K_S, mya (ref.)	Phylogenetics, previous results, mya (ref.)	Phylogenetics, this study, mya
A. thaliana	25–30 (13); 65 (17)		43
P. trichocarpa	13 (54)	60–65 (7)	48*
V. vinifera	No recent duplication	No recent duplication (5)	
M. truncatula, *L. japonicus*†, *G. max*†	>50 (8); 58 (16); 44 (16)		65
G. hirsutum	13–15 (13)		59
C. papaya	No recent duplication	No recent duplication	
S. tuberosum†, *S. lycopersicum*	50–52 (16)		69
L. sativa	>40–45 (15)		65
E. californica	Unknown		70
Musa spp.‡,	61 (18)		
O. sativa, *Sorghum bicolor*†	50–60 (16); 70 (10)		65
A. americanus	Unknown		57
P. patens	Unknown	45 (19)	

Nonreferenced dates in the phylogenetics columns are based on the constraints described in detail in Table S1.

*The inferred date of the poplar WGD through phylogenetic tree construction and penalized likelihood as described in the current study is probably underestimated due to its much slower substitution rate (see the text and *SI Text*).

†*Lotus* and *Glycine*, *S. tuberosum*, and *Sorghum* represent the same WGD events as *Medicago*, *S. lycopersicum*, and *Oryza*, respectively, but have been included because their genome sequence is known or WGD events have been previously described in these species (8, 10, 16).

‡The EST data collection for *Musa spp*. is not available in the public domain (18).

(65 mya). To explain this pattern, we argue that polyploidy may have increased the survival chances and recolonization capacity of plant lineages during and/or after the KT mass extinction.

Results and Discussion

Absolute dating of WGD paralogs was performed based on whole-genome sequence data from *A. thaliana*, *P. trichocarpa*, *Medicago truncatula*, *Vitis vinifera*, *O. sativa*, and *Physcomitrella patens*. To make sure genes from the youngest genome duplication events were dated, we first identified paralogs lying in recently duplicated segments. We also included several plant species for which a sufficient number of paralogous pairs could be identified from EST libraries, such as *Gossypium hirsutum* (cotton), *Solanum lycopersicum* (tomato), *Lactuca sativa* (lettuce), *Eschscholzia californica* (California poppy), and the basal monocot *Acorus americanus* (14, 15). Rice and *Physcomitrella* were used as outgroups when dating WGDs in eudicot and monocot species, respectively. By using the PL method, the divergence date of the paralogous genes in each phylogenetic tree was calculated relative to a reference speciation event, namely the divergence of *Vitis* and the remaining rosids (*Arabidopsis*, *Populus*, *Medicago*, and *Gossypium*) fixed at 115 mya (23, 24) when dating the WGD in rosids, and the divergence of monocots and eudicots at 145 mya (25) when dating the WGD in monocots (*Oryza* and *Acorus*; see *Methods*, *SI Text*, and Fig. S2 for detailed descriptions of the methodology).

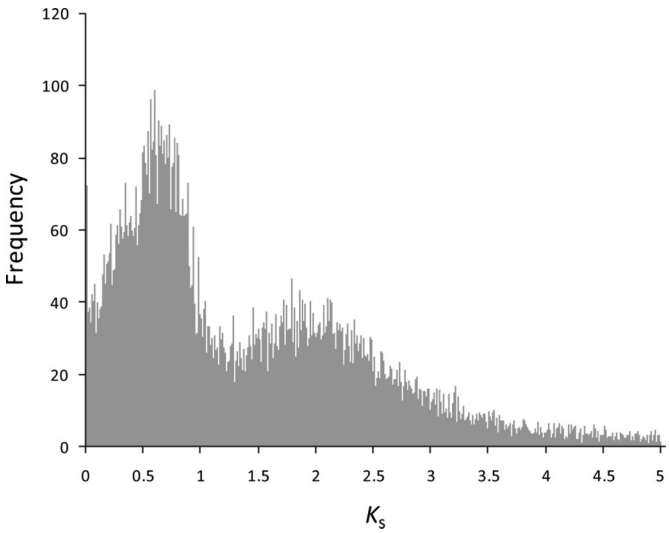

Fig. 1. Age distribution of the *Arabidopsis* paralogs based on K_S values. Age distributions for other plants can be found in *SI Text* (see Fig. S1). K_S represents the number of synonymous substitutions per site. The conspicuous peak around K_S = 0.6 originates from the youngest genome duplication in the *Arabidopsis* lineage.

Fig. 2 shows examples of age distributions of duplicated genes for *Solanum*, *Medicago*, *Oryza*, and *Gossypium* (results of the other species can be found in Fig. S3) obtained by phylogenetic tree construction and PL estimation. In Table 1, the WGD dates inferred from the absolute dating approach are shown together with the dates inferred previously with various other approaches, such as calibration of K_S distribution peaks with molecular clocks. In most cases, the dates we infer are in agreement with previous estimates. Based on the construction of phylogenetic trees, *Oryza*, *Medicago*, *Solanum*, *Gossypium*, *Lactuca*, *Eschscholzia*, and *Acorus* all show a peak at ≈57–70 mya. Younger dates were obtained for the WGD events of *Arabidopsis* and *Populus*. Our data suggest that the youngest WGD of *Arabidopsis* occurred ≈40 mya, which is more recent than the 65 or 72 mya previously proposed (17), but considerably older than the 25–30 mya proposed by Blanc and Wolfe (13), who also date the youngest WGDs in other plants such as *Medicago*, *Gossypium*, and *Solanum* much younger because a higher synonymous substitution rate was used. Although dating anchors with relatively small K_S values only provide one clear peak (Fig. S3), the recent publication of the papaya (*Carica papaya*) genome provided strong evidence for 2 WGDs in the *Arabidopsis* lineage since its divergence from papaya, ≈70 mya (6, 26). Although we did not have the means to accurately determine the age of the oldest *Arabidopsis* specific duplication, we assume that it must be after the divergence of *Arabidopsis* and *Carica* and before the WGD dated at ≈40 mya (6).

Regarding *Populus*, it is well-acknowledged that sequences of perennial species or trees such as *Populus* evolve slower than sequences of *Arabidopsis* or other weeds, and this might lead to underestimation of the age of the WGD. Indeed, the observation that the WGD is shared with *Salix* suggests that it is considerably older (60–65 mya) than the age inferred here (35 mya) (7). In addition, the underestimation of the WGD in *Populus* by PL

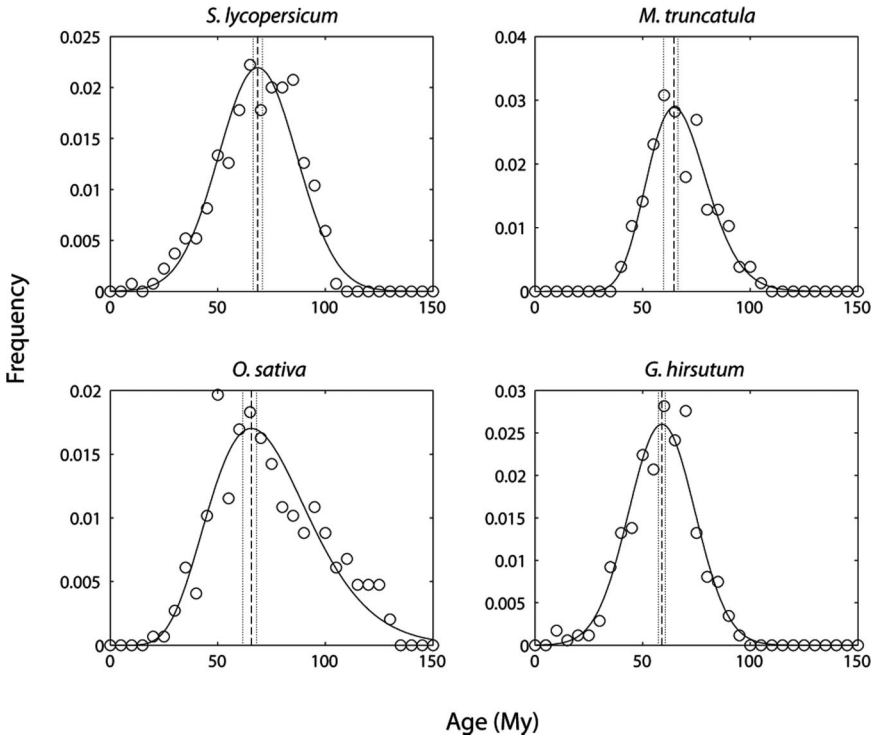

Fig. 2. ML fits of the duplicate age distributions of *S. lycopersicum* (calibration point AS120, with constraints, normal fit) (see Table S1 for an explanation of the terminology for the calibration points), *M. truncatula* (AV115, with constraints, gamma fit), *O. sativa* (AO145, with constraints, gamma fit), and *G. hirsutum* (AV115, with constraints, normal fit). The dashed line indicates the ML estimate of the distribution mode. The dotted lines delimit the corresponding 95% confidence intervals. Distributions for other plant species and constraints can be found in *SI Text*.

could be due to taxon undersampling. We investigated this possibility by adding sequences of *Manihot esculenta* (Cassava). *Manihot* and *Populus* are both members of Malpighiales and share a more common recent ancestry than *Populus* and *Medicago*. By including *Manihot*, an older age (48 mya) was obtained for the WGD in *Populus* (see Fig. S3). The distribution of absolute dates for *Physcomitrella* proved inconclusive, probably because of the very large evolutionary distance between *Physcomitrella* and the outgroup *Chlamydomonas* and the other plant species.

Both the inferred and previously estimated WGD dates were mapped onto a phylogenetic tree of monocots and eudicots (Fig. 3). The independent WGD events in several plant lineages are clearly not distributed randomly over time, which is supported by statistical analysis ($P < 0.01$; see *Methods*). Most recent WGDs in plants seem to have occurred within the same small time frame, 60–70 mya, a period in which the environment on Earth changed considerably and a large fraction of life disappeared. The KT extinction event, ≈65 mya, is known as the most recent large-scale mass extinction of animal and plant species in a geologically short period. There is now a general consensus that the KT extinction was caused by one or more catastrophic events such as a massive asteroid impact (more in particular the Chicxulub impact) and/or increased volcanic activity. These events likely generated global wildfires and global dust clouds that cut off sunlight for a period ranging from several months to several years. Probably, light levels too low for photosynthesis as well as freezing ground temperatures made seed-germination difficult and caused the extinction of many terrestrial plants (27). Indeed, recent studies focusing on changes in abundance, community structure, and taxonomic richness indicate ecological instability in plant communities and collapse, before or coincident with the peak faunal extinction at the KT boundary. Paleobotanical studies of fossil pollen, spores, and leaves from North American localities showed the disappearance of up to ≈60% of plant species (28). Other, more circumstantial evidence includes fungal proliferation at the KT boundary. Vajda and McLoughlin (29) found a fungal spike at the KT boundary in layers of coal in New Zealand implying extensive dieback of photosynthetic vegetation. The fungal peak is interpreted as representing a dramatic increase in the substrates available for nonphotosynthetic saprophytic organisms provided by global forest dieback after the Chicxulub impact. In conclusion, rapid loss of many plant species and major compositional changes in plant communities all point to global paleoecological upheaval and rapid ecosystem failure at the KT boundary (30).

The question remains why plants with double genomes would have had a greater chance of survival in this dramatically changed environment. Interestingly, Crow and Wagner (31) observed that also in vertebrates the probability of extinction seems to be reduced after polyploidization. Although many changes associated with polyploidization are likely to be disadvantageous or deleterious (32, 33), the KT polyploids apparently did have a short-term evolutionary advantage, because many of them survived and outcompeted many, if not most, of their diploid progenitors (31). There is ample evidence that gene duplication fuels long-term diversification and evolutionary success through the evolution of novel gene functions (32, 34), but the short-term advantages of polyploidy are less known. Several studies have suggested that polyploid plants can have increased tolerance to a wider range of environmental conditions compared with their diploid relatives (35, 36). Altered levels of gene expression in polyploids are probably an important factor (33). Immediate changes in gene expression can result from increased heterozygosity and dosage balance effects after genome duplication. In addition, it has been shown that polyploidization can lead to rapid epigenetic repatterning and concomitant changes in gene expression, such as tissue-specific differential expression of gene duplicates (37). Such changes in gene expression can be advantageous because they can contribute to hybrid vigor and provide

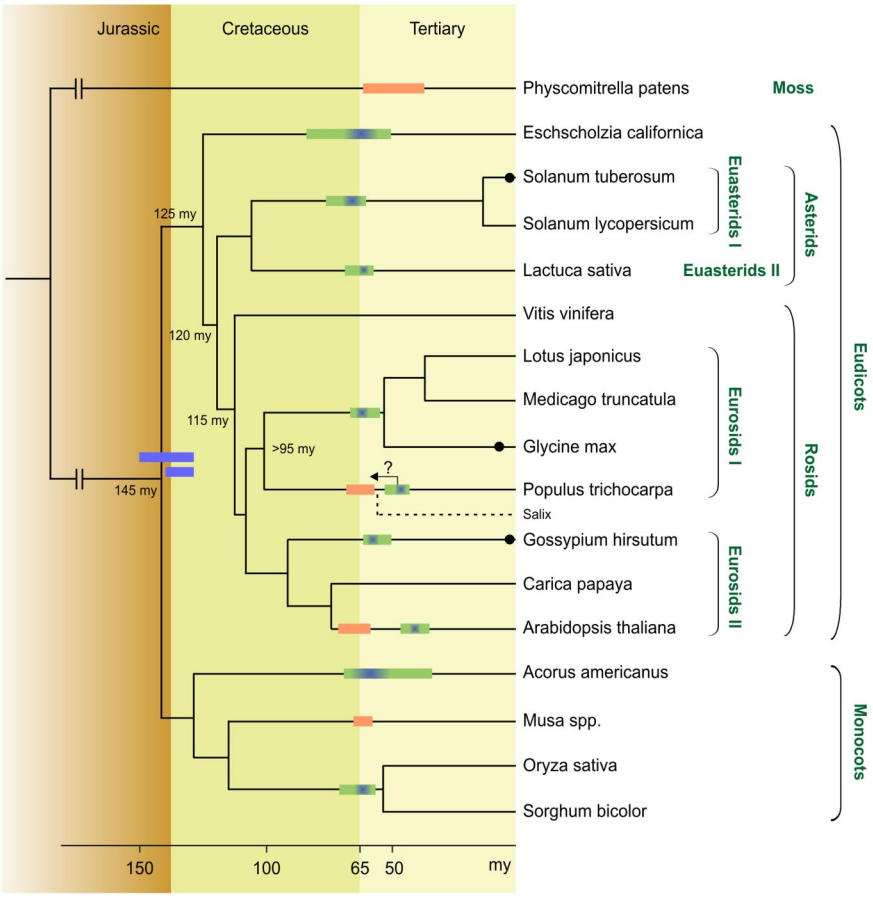

Fig. 3. Phylogenetic tree of flowering plants (eudicots and monocots) for which the genome sequence has been determined or for which large EST collections are available. WGDs are indicated by green bars depicting the union of their 95% age confidence intervals calculated with various constraints (see Table S1). The dark green portions of the bars are centered on the best age estimates (see Table 1). Orange bars are WGD age estimates from literature. The WGD in poplar [here estimated by including *Manihot* (see Table S1)] has most probably occurred before its divergence of *Salix*, although dating by K_S values and phylogenetic means suggest a younger date, probably due to the slower evolutionary rate in trees (see text and *SI Text* for details). Blue bars denote the hexaploid nature of the ancestral eudicot genome (5, 26). The black dots indicate very recent polyploidy events, ≈1–2 mya in *G. hirsutum*, <10 mya in *Solanum tuberosum*, and 10–15 mya in *Glycine max*. The resulting tetraploids have not or only partially diploidized so far.

variation that might allow fast adaptation to novel conditions (38–40). Expression pattern instability might offer polyploids a broader phenotypic range compared with their diploid progenitors. Alterations or partitioning of parental gene expression in polyploid cotton have been shown to occur in response to abiotic stimuli (41). By partitioning ancestral expression patterns in response to environmental stresses, duplicated genes can become subfunctionalized (42) and thus undergo separate processes of genetic evolution. Differential expression of these duplicated genes could imply that a new hybrid individual might be better adapted for survival in a different ecological niche. Polyploidy is also known to facilitate self-fertilization and the formation of asexually-reproducing (apomictic) species (43), which might have been advantageous in a time when sexual mates were scarce.

Increased phenotypic variability, heterosis effects, mutational robustness, subfunctionalization, and changed reproductive modes all have the potential to allow the polyploid to survive environmental conditions that are unfavorable to the diploid progenitors (44). Evidence in support of this hypothesis comes from the study of present-day polyploids. An example of the rapid adaptation of polyploids to new and extreme niches has been provided by the study of the arctic flora. Phylogenetic analyses indicate that arctic allopolyploids are abundant and have been particularly efficient in invading newly deglaciated areas during periods of dramatic climate change (45). Even more recently, novel man-made habitats, such as industrial wastelands, have been successfully colonized by newly formed allopolyploids such as *Senecio eboracensis* (York groundsel), first discovered in the late 1970s (46). As a matter of fact, it has been long known that many invasive plants are polyploids (47, 48) and that polyploid plants are often able to exploit habitats that their diploid progenitors were unable to (36, 49). In a drastically changed environment such as on Earth after the cataclysmic events that occurred ≈65 mya, where many plants became extinct, it is likely that competition was seriously reduced and new niches suddenly became available for occupation, and therefore invasive species with double genomes might have spread rapidly (50).

In conclusion, we have shown that most WGDs so far identified in different plant families during the last 110 my or so have occurred at about the same time, or at least in a very short time frame. Furthermore, these WGD events seem to have occurred at a time corresponding with the KT extinction event, which wiped out a considerable fraction of life on Earth. It should be noted that absolute dating is not a trivial task and there are several caveats that could result in some of the ages being over- or underestimated. Nevertheless, we believe that the significant clustering of WGD events around the KT boundary is unlikely to change (see *SI Text* for further discussion regarding the methodology). The 10-15 my interval in which the PL method applied here places most of the paleopolyploidy events might seem, at first sight, still rather wide. This wide range could be

because the extinction event, as well as the propagation of polyploids, covered a longer period. It has been suggested that a global climate change had already been set in motion well before the asteroid impact, e.g., as the result of increased volcanic activity (51), whereas the effects after the impact probably lasted a considerable period. Alternatively, the spread of the WGDs could be caused by the uncertainty in dating. As stated above, different species evolve at different rates and a 10- to 15-my interval falls, we believe, perfectly within expectations considering differences in substitution rates in different species over several tens of millions of years (*SI Text*). Therefore, we would like to put forward the hypothesis that plants with double genomes around the KT boundary had a selective advantage compared with their diploid progenitors. Indeed, because of the putative advantages of (allo)polyploidy, such as altered gene expression leading to hybrid vigor and an increased set of genes and alleles available for selection, polyploid plants might have been better adapted to the changed environment after the asteroid impact and as such may have outcompeted many of their diploid progenitors (see also ref. 31). The relatively low incidence of paleopolyploidy over the last 110 my compared to present-day rates of polyploid formation and the fact that a considerable fraction of the polyploids that survived seem to be clustered in time, might be explained by the fact that polyploidy is usually an evolutionary dead end and mostly disadvantageous, unlike what has been suggested by others (52). Thus, it can be argued that polyploids can survive only if the circumstances are right. A cataclysmic event or environmental upheaval, such as what happened around the KT boundary, might have given the polyploids an evolutionary advantage over their diploid progenitors that allowed them to become established and proliferate.

Methods

Details regarding data resources, the identification of paralogs created by large-scale duplication events, the construction of orthologous gene families, and the inference of phylogenetic trees can be found in *SI Text*.

Age Distributions. For each gene family that could be considered (Table S1 and *SI Text*), the divergence time of the 2 paralogs was estimated for each of the 100 bootstrapped samples by using the PL method as described in *SI Text* (see also ref. 53). A statistical test from Torsten Eriksson's package (www.bergianska.se/index_forskning.php) was performed to test whether the 100 dates estimated for each bootstrapped sample conformed to a normal distribution. The mean age of the inferred dates was taken as the age of the duplication event for each family if it passed the statistical test. An age distribution was inferred with the estimated ages of all of the duplication events for each given species (Fig. 2 and Fig. S3). Some trees failed the cross validation procedure, or it was not possible to estimate the age of the duplication node. Furthermore, only families for which >70 of 100 dates were calculated were included for further analysis. Different calibration points and constraints were used to date duplication events (see *SI Text*).

Calculation of Maximum Likelihood (ML) WGD Age Estimates and 95% Confidence Intervals. Normal and gamma distributions were fitted to the age distribution of the duplication events for all species by using ML estimation routines in Matlab. The ML WGD age estimate was taken to be the mode (= mean) of the fitted normal distribution or the mode $\hat{m} = (\hat{a} - 1)\hat{b}$ of the fitted gamma distribution with ML parameters $\hat{a}$ and $\hat{b}$, whichever had the largest likelihood. For the normal fits, the 95% confidence interval on the mode is given by the Matlab normfit function. For the gamma fits, the 95% confidence interval on the mode of the distribution was obtained through numerical integration of the probability density function of the mode:

$$P(m) = \int_{-\infty}^{\infty} \frac{1}{|x|} \frac{\sqrt{C_{aa}C_{bb}}}{2\pi\sqrt{C_{aa}C_{bb} - C_{ab}{}^2}} \exp \left\{ - \frac{C_{aa}C_{bb}}{2(C_{aa}C_{bb} - C_{ab}{}^2)} \times \left[\frac{(x - (\hat{a} - 1))^2}{C_{aa}} + \frac{(m/x - \hat{b})^2}{C_{bb}} - \frac{2C_{ab}(x - (\hat{a} - 1))(m/x - \hat{b})}{C_{aa}C_{bb}} \right] \right\} dx \quad [1]$$

where C is the covariance matrix of the fit. When multiple sets of calibrations and constraints were used to date a particular WGD, the confidence intervals displayed in Fig. 3 are the union of the confidence intervals obtained for the different fits. In most cases, the ages calculated with or without the constraints did not change much (see Table S1 and Fig. S3).

Clustering of WGD Events in Time. To assess whether there has been a significant grouping of WGD events in time, we used the median distance between the WGD events as an indicator of the degree of clustering of the events. The median distance is calculated over all pairs of WGDs; smaller median distances indicate a tighter clustering of events in time. We used the mean distance and the mean and median waiting times between successive WGD events as alternative test statistics and obtained similar results. We excluded WGD events that have occurred <10 mya because it is not clear whether they will stand the test of time. Indeed, new polyploids are continuously being formed among extant plants, but most of these will probably not survive for longer than a few million years. The upper boundary of WGD ages was set to 110 mya, roughly coinciding with the *Arabidopsis–Vitis* split, because the number and timing of events in the monocot and eudicot lineages before this time is presently unclear. Random sampling was used to estimate the probability that a median distance lower than or equal to the one inferred from the calculated WGD age estimates occurs when randomly distributing the WGDs over a 100-my timespan. Assuming that the background probability of a WGD occurring at a certain point in time is proportional to the number of species present at that time, the ages of the WGDs in each of a million random samples were drawn from a distribution on the interval [10–110 mya] that reflects the evolution of the number of species over time (see Fig. S4*A*; only the species shown in Fig. 3 are taken into account). We tested several combinations of age estimates from Table S1. In all cases, the observed median distance was significantly lower than expected at random ($P < 0.01$). Taking into account only our phylogenetic date estimates from Table 1, without the *Musa* and *Physcomitrella* branches on Fig. 3, yields $P = 0.0012$ (see Fig. S4; median distance = 8.90, average median distance in random samples = 29.26). When we replaced the anomalously low date for *Populus* with a more realistic date of 62.5 mya, which is based on fossil data, and included dates from other studies of *Physcomitrella* (45 mya), *Musa* (61 mya), and the more recent WGD in the *Glycine max* lineage (14 mya), we obtained $P = 3.58 \times 10^{-4}$, (median distance = 10.00, average median distance in random samples = 29.10). Expansion of the timeframe to 150 mya and inclusion of one or two older WGDs between 110 mya and 150 mya yields P values of ≈0.01.

ACKNOWLEDGMENTS. We thank Todd Vision and Grigoris Amoutzias for helpful comments on the manuscript, Lieven Sterck and Klaas Vandepoele for technical assistance, and 4 reviewers for their suggestions and enthusiasm. S.M. is a Postdoctoral Fellow of the Research Foundation Flanders (Belgium). This work was supported by European Union EU-FP6 Food Safety and Quality Contract FOOD-CT-2006-016214 and by Belgian Federal Science Policy Office Grant IUAP P6/25 (BioMaGNet).

1. Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.
2. Maere S, et al. (2005) Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* 102:5454–5459.
3. Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. *Science* 290:2114–2117.
4. Simillion C, et al. (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 99:13627–13632.
5. Jaillon O, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467.
6. Tang H, et al. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res* 18:1944–1954.
7. Tuskan GA, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604.
8. Cannon SB, et al. (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proc Natl Acad Sci USA* 103:18026.

9. Sato S, et al. (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Res* 15:227–239.
10. Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903–9908.
11. De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. *Trends Ecol Evol* 20:591–597.
12. Soltis DE, Bell CD, Kim S, Soltis PS (2008) Origin and early evolution of angiosperms. *Ann NY Acad Sci* 1133:3–25.
13. Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678.
14. Cui L, et al. (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res* 16:738–749.
15. Barker MS, et al. (2008) Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol* 25:2445–2455.
16. Schlueter JA, et al. (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47:868–876.
17. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
18. Lescot M, et al. (2008) Insights into the *Musa* genome: Syntenic relationships to rice and between *Musa* species. *BMC Genomics* 9:58.
19. Rensing SA, et al. (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol Biol* 7:130.
20. Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol* 18:352–361.
21. Rannala B, Yang Z (2007) Inferring speciation times under an episodic molecular clock. *Syst Biol* 56:453–466.
22. Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol Biol Evol* 19:101–109.
23. Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: Calibrating the family tree. *Proc Biol Sci* 268:2211–2220.
24. Schneider H, et al. (2004) Ferns diversified in the shadow of angiosperms. *Nature* 428:553–557.
25. Chaw S-M, Chang C-C, Chen H-L, Li W-H (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58:424–441.
26. Tang H, et al. (2008) Synteny and collinearity in plant genomes. *Science* 320:486–488.
27. Vajda V, Raine JI, Hollis CJ (2001) Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. *Science* 294:1700–1702.
28. Wilf P, Johnson KR (2004) Land plant extinction at the end of the Cretaceous: A quantitative analysis of the North Dakota megafloral record. *Paleobiology* 30:347–368.
29. Vajda V, McLoughlin S (2004) Fungal proliferation at the Cretaceous-Tertiary boundary. *Science* 303:1489.
30. McElwain JC, Punyasena SW (2007) Mass extinction events and the plant fossil record. *Trends Ecol Evol* 22:548–557.
31. Crow KD, Wagner GP (2006) What is the role of genome duplication in the evolution of complexity and diversity? *Mol Biol Evol* 23:887–892.
32. Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annu Rev Genet* 34:401–437.
33. Comai L (2005) The advantages and disadvantages of being polyploid. *Nat Rev Genet* 6:836–846.
34. Taylor JS, Raes J (2004) Duplication and divergence: The evolution of new genes and old ideas. *Annu Rev Genet* 38:615–643.
35. Levin DA (1983) Polyploidy and novelty in flowering plants. *Am Nat* 122:1–25.
36. Thompson JD, Lumaret R (1992) The evolutionary dynamics of polyploid plants—Origins, establishment and persistence. *Trends Ecol Evol* 7:302–307.
37. Adams KL, Wendel JF (2005) Novel patterns of gene expression in polyploid plants. *Trends Genet* 21:539–543.
38. Hegarty MJ, et al. (2008) Changes to gene expression associated with hybrid speciation in plants: Further insights from transcriptomic studies in *Senecio*. *Philos Trans R Soc London Ser B* 363:3055–3069.
39. Rieseberg LH, et al. (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129:149–165.
40. Rieseberg LH, et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216.
41. Liu Z, Adams KL (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr Biol* 17:1669–1674.
42. Force A, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
43. Bicknell RA, Koltunow AM (2004) Understanding apomixis: Recent advances and remaining conundrums. *Plant Cell* 16:S228–S245.
44. Hegarty M, Hiscock S (2007) Polyploidy: Doubling up for evolutionary success. *Curr Biol* 17:R927–R929.
45. Brochmann C, et al. (2004) Polyploidy in arctic plants. *Biol J Linn Soc* 82:521–536.
46. Lowe AJ, Abbott RJ (2003) A new British species, *Senecio eboracensis* (Asteraceae), another hybrid derivative of *S. vulgaris* L. and *S. squalidus* L. *Watsonia* 24:375–388.
47. Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc Natl Acad Sci USA* 97:7043–7050.
48. Pandit MK, Tan HTW, Bisht MS (2006) Polyploidy in invasive plant species of Singapore. *Bot J Linn Soc* 151:395–403.
49. Soltis DE, Soltis PS, Tate JA (2003) Advances in the study of polyploidy since plant speciation. *New Phytol* 161:173–191.
50. Prentis PJ, et al. (2008) Adaptive evolution in invasive species. *Trends Plants Sci* 13:288–294.
51. Keller G, et al. (2004) Chicxulub impact predates the K-T boundary mass extinction. *Proc Natl Acad Sci USA* 101:3753–3758.
52. Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* 131:452–462.
53. Sanderson MJ, Doyle JA (2001) Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. *Am J Bot* 88:1499–1516.
54. Sterck L, et al. (2005) EST data suggest that poplar is an ancient polyploid. *New Phytol* 167:165–170.