

RESEARCH HIGHLIGHT

# A mystery unveiled

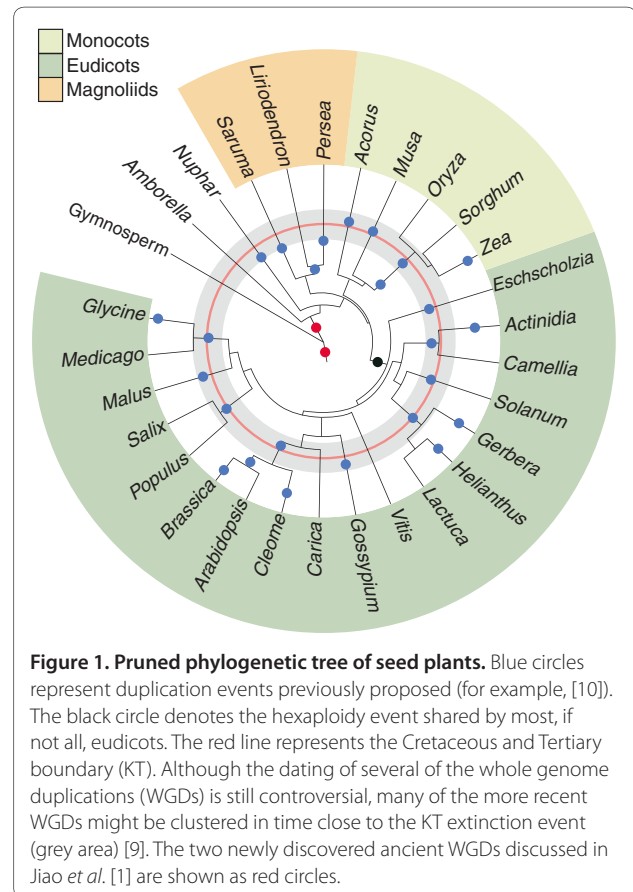
Yves Van de Peer<sup>\*1,2</sup>

## Abstract

A recent phylogenomic study has provided new evidence for two ancient whole genome duplications in plants, with potential importance for the evolution of seed and flowering plants.

The number of genome duplications uncovered in the evolutionary history of land plants seems to be steadily increasing. In a recent paper in *Nature*, Jiao *et al.* [1] provide evidence for two additional, previously unnoticed ancient whole genome duplications (WGDs) in seed plants. More precisely, the authors propose a WGD in the common ancestor of all extant angiosperms, and an even older one in the common ancestor of all extant seed plants. This means, for instance, that adding these to the well-documented and widely accepted hexaploidy event shared by most, if not all, of the eudicots, and the two more recent genome duplications after its divergence with papaya, the small genome of *Arabidopsis* should carry the traces of at least five WGDs. And although *Arabidopsis* actually might be an extreme case, it represents only one of many plant lineages that have experienced a number of nested WGDs since the origin of seed and flowering plants (Figure 1).

So why have these older events not been detected before? Plant genomes are highly dynamic and usually go through an intense phase of structural rearrangements and gene loss following duplication. Bioinformatics tools to detect within-genome colinearity can be used to find remnants of relatively recent WGDs, but, since colinearity fades with time, old(er) duplication events in plants generally can not be detected this way. In many cases, WGDs can also be detected by building age distributions of paralogs, where the number of paralogs is plotted against their age, which can be approximated by the number of synonymous substitutions per synonymous site ( $K_s$ ). A peak in such a distribution indicates a



**Figure 1. Pruned phylogenetic tree of seed plants.** Blue circles represent duplication events previously proposed (for example, [10]). The black circle denotes the hexaploidy event shared by most, if not all, eudicots. The red line represents the Cretaceous and Tertiary boundary (KT). Although the dating of several of the whole genome duplications (WGDs) is still controversial, many of the more recent WGDs might be clustered in time close to the KT extinction event (grey area) [9]. The two newly discovered ancient WGDs discussed in Jiao *et al.* [1] are shown as red circles.

burst of duplications at about the same time, and is often interpreted as a WGD event. However, due to gene loss and saturation effects,  $K_s$  age distributions become unreliable for the detection of older duplication events. To see whether they could find evidence for older WGDs in plants, Jiao *et al.* [1] used a phylogenomic approach and constructed thousands of evolutionary trees for gene families built from a collection of genes from sequenced genomes and more than 12.5 million new expressed-sequence-tag sequences from basal angiosperms and gymnosperms. A semi-parametric penalized likelihood approach was then applied to estimate the divergence time of duplicated genes or clades compared with some pre-set speciation times. Analysis of the distribution of inferred duplication times showed two clear peaks: one

\*Correspondence: yves.vandeppeer@psb.vib-ugent.be

<sup>1</sup>Department of Plant Systems Biology, VIB, Technologiepark 927, Ghent, Belgium  
Full list of author information is available at the end of the article

in the ancestor of all extant angiosperms, dated at 192 million years ago (mya), and one in the ancestor of all extant seed plants dated at 319 mya.

In addition, Jiao *et al.* [1] performed a gene ontology enrichment analysis for the genes that survived these two old WGDs, and found an overabundance of genes from several functional categories, including transferases and binding proteins, transcription factors and protein kinases. A biased retention of transcriptional and developmental regulators and proteins that form complexes is a recurrent observation following WGDs. Dosage balance effects can explain the specific retention of such genes [2,3]. When a dosage-sensitive gene is duplicated on its own, the effects of its increased dosage with respect to its interaction partners or targets would be deleterious, and the duplication would be selected against [4]. But if the gene and its interaction partners are duplicated together, as in a WGD, their relative dosage is preserved. As a matter of fact, the observation of biased gene retention following a duplication event, such as observed by Jiao *et al.* [1], is actually rather strong, albeit circumstantial, evidence that the duplication event did affect the whole genome. The consequences of these dosage balance effects is that post-WGD organisms are endowed with a collection of extra transcription factors, transporters and complexes that may not necessarily be useful immediately but that cannot be purged easily from the genome. In the long run, these non-adaptively preserved genes, which can be considered neutral byproducts or regulatory and developmental spandrels [5], may be co-opted for adaptive innovations. As previously shown [6], the total numbers of regulatory and developmental regulators added to plant genomes through WGDs is huge, and consequently WGDs may have spurred really important evolutionary innovations. In that respect, the older genome duplications described by Jiao *et al.* [1] could have been particularly important since they might have coincided with, respectively, the origin of the seed and the flower, two key developments in the evolution of land plants. For instance, the authors identified several tens of genes involved in flower developmental pathways, with at least one ancient duplication event before the divergence of monocots and dicots. Other examples include regulators of flowering time and seed germination, showing duplication patterns consistent with WGDs pre-dating the origin of seed plants and angiosperms.

Darwin was deeply bothered by what he perceived to be an abrupt origin and highly accelerated rate of diversification of flowering plants in the mid-Cretaceous period [7]. Apart from increasing the diversification potential for morphologic innovation and biological novelty, WGDs often seem to be accompanied by marked and sudden increases in species richness [8]. Although a link between WGDs and increased species diversity remains correlational

rather than causal, mechanisms such as reciprocal gene loss or gene subfunctionalization might explain how gene and genome duplications facilitate the formation of novel species. Furthermore, it has been suggested recently that WGDs might increase the chance for species to survive extinction events [9]. Therefore, the preferential retention of developmental genes and genes functioning in signaling and regulatory cascades, plus the fact that WGDs might facilitate speciation events while at the same time reducing the chance of extinction, sheds novel light on the evolution of one of the most important clades of organisms on this planet. Perhaps Darwin's 'abominable mystery' of the origin and radiation of the angiosperms might be just about to be solved.

#### Abbreviations

$K_s$ , number of synonymous substitutions per synonymous site; mya, million years ago; WGD, whole genome duplication.

#### Competing interests

The author declares that he has no competing interests.

#### Acknowledgements

I would like to thank Jeffrey A Fawcett and Xiaolian Zhu for help with Figure 1, and Steven Maere for a critical reading of the manuscript.

#### Author details

<sup>1</sup>Department of Plant Systems Biology, VIB, Technologiepark 927, Ghent, Belgium. <sup>2</sup>Department of Plant Biotechnology and Genetics, Ghent University, Technologiepark 927, Ghent, Belgium.

Published: 30 May 2011

#### References

1. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW: **Ancestral polyploidy in seed plants and angiosperms.** *Nature* 2011, **473**:97-100.
2. Papp B, Pal C, Hurst LD: **Dosage sensitivity and the evolution of gene families in yeast.** *Nature* 2003, **424**:194-197.
3. Freeling M, Thomas BC: **Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity.** *Genome Res* 2006, **16**:805-814.
4. Birchler JA, Veita RA: **The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution.** *New Phytol* 2010, **186**:54-62.
5. Freeling M: **Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition.** *Annu Rev Plant Biol* 2009, **60**:433-453.
6. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: **Modeling gene and genome duplications in eukaryotes.** *Proc Natl Acad Sci U S A* 2005, **102**:5454-5459.
7. Friedman WE: **The meaning of Darwin's 'abominable mystery'.** *Am J Bot* 2008, **96**:5-21.
8. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS: **Polyploidy and angiosperm diversification.** *Am J Bot* 2009, **96**:336-348.
9. Fawcett JA, Maere S, Van de Peer Y: **Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event.** *Proc Natl Acad Sci U S A* 2009, **106**:5737-5742.
10. Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW: **Widespread genome duplications throughout the history of flowering plants.** *Genome Res* 2006, **16**:738-749.

doi:10.1186/gb-2010-12-5-113

Cite this article as: Van de Peer Y: **A mystery unveiled.** *Genome Biology* 2011, **12**:113.