

Evolution: The flowering of land plant evolution - whence and whither?

Philip Donoghue

School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

More than 90% of living land plants are angiosperms (flowering plants) and so it is difficult to conceive of a world without them; angiosperms are not merely decorative, effecting incalculable ecosystem services, encompassing most commercial crop species while also serving as global climate engineers [1]. But there was a world before angiosperms, though no one can quite agree when it ended because of a long-standing controversy concerning the timing of origin of the living clade of (crown) angiosperms. For over a century, repeated sampling of the fossil record has failed to find evidence of crown-angiosperms before the Cretaceous, which began about 145 million years ago, yet almost from their introduction, molecular clock analyses have estimated a much more ancient origin, perhaps more than a 100 million years earlier [2]. Two new studies, by Li *et al.* [3] and Coiro *et al.* [4], together indicate that there is no evidence of a looming détente despite improved molecular clock methods, denser and deeper sampling of plant genomes and a more refined understanding of the plant fossil record. However, a solution may lie in the nature of their disagreement.

Mismatches between molecular clock estimates and the fossil record are hardly news, following infamous controversies over the timing of diversification of mammals, birds, animals and land plants, among others. However, in most cases they have diminished in scale, both as a consequence of the development of more realistic models of rate variation and a more sober approach to the interpretation of the fossil evidence used to calibrate molecular clocks to geological time [5]. However, these advances have not had the same impact on attempts to infer the timescale of angiosperm evolution.

The new study by Li *et al.* [3] is unparalleled in scope, sampling 80 genes from the chloroplasts of 2881 species, including most families and all orders, which they use to estimate the evolutionary relationships of angiosperms (Figure 1a). Other studies have used more genes but the impact of broader taxonomic sample has been a concern. Their phylogenetic results largely corroborate convention and their evolutionary timescale estimates crown-angiosperms to have appeared in an interval extending back in time from the Early Jurassic, through the Triassic, and deep into the Permian (187-267 Ma). The staggering scale of this 42-122 million year mismatch with a fossil record starting at 130 Ma, is no surprise at all. Indeed, it is something that seasoned clock-watchers have come to expect from molecular estimates of angiosperm evolutionary history and the authors consider it high time that we take them seriously, formalizing at least a 'Jurassic Gap' interpretation of the fossil record.

Mismatches between molecular clock estimates and the fossil record should be the expectation since molecular clocks estimate the timing of genetic divergence, while the fossil record remains blind to this episode until descendent lineages have acquired morphologically distinct characteristics that can be fossilized. However, it is the scale of such mismatches that invoke the ire of palaeo folk. Palaeobotanists are united in a vociferous defense of the veracity of the angiosperm fossil record for good reason. Fossil stem-

angiosperms are known from at least as far back as the Triassic, providing evidence for the assembly of the crown-angiosperm body plan. Nevertheless, repeated claims of Jurassic crown-angiosperm macroremains have failed to withstand scrutiny [4, 6].

An absence of evidence is not evidence that crown-angiosperms were absent from Jurassic (or earlier) floras, but palaeobotanists take confidence in their interpretation of the fossil record because they rely on more than the simple presence and absence of fossils. The fossil record is not a random sample of historical biological diversity, but a decidedly non-random archive structured by local and global tectonic processes that result in the differential preservation of environments across geological time [7]. Knowledge of this structure provides for a qualified interpretation of the fossil record, one in which fossil occurrences are predictable [8]. To be sure, fossil representatives of lineages can be absent because they were present but simply not preserved and, indeed, some ancient plant lineages (e.g. bryophytes) have a truly appalling fossil record. However, the same cannot be said of angiosperms where individual plants can yield millions of pollen grains that are all-but indestructible. Absence of angiosperm pollen fossils from Jurassic strata that otherwise preserve the pollen, seeds and spores of stem-angiosperms and gymnosperms, surely tells us they are absent because crown-angiosperms had not yet evolved.

This is key point made in a recent study by a Coiro et al. [4] who provide a temporal and spatial review of the pollen fossil record through the Jurassic-Cretaceous interval, demonstrating a graduated emergence of crown-angiosperm diversity. Given the richness and consistency of the pollen fossil record, it is not surprising that the earliest crown-angiosperm pollen appears earlier than the oldest macroremains. However, the pollen record exhibits anything but the sudden emergence of crown-angiosperm diversity that the macrofossil record suggests. Rather, the evolutionary grades of pollen encountered in the early Cretaceous record match precisely the order predicted by molecular phylogenies of their living relatives, with monosulcate pollen of the early diverging ANITA-grade plants (*Amborella*, Nymphaeales, and Austrobaileyales) occurring in older strata than the tricolporate pollen grains of more derived eudicots. This correlation between fossil stratigraphic order and phylogenetic branching order strongly suggests that the fossil evidence is more signal than noise.

Cleverly, Coiro and colleagues explore the predictions of molecular clock analyses like the one conducted by Li et al. [3], inferring the nature of ancestral pollen based on the characteristics of living lineages and a molecular timescale of angiosperm evolution. Their results require the presence of ANITA-grade pollen deep in the Triassic and eudicot-grade pollen from the latest Jurassic (Figure 1b); these expectations are not met (Figure 1c). Mapping fossil occurrences onto palaeogeographic reconstructions they demonstrate an early Cretaceous expansion to global distribution of crown-angiosperm pollen. Perhaps most importantly, they demonstrate that the early endemic distribution and earlier absence of crown-angiosperm pollen is not merely a consequence of non-preservation since there are rich records of gymnosperm and stem-angiosperm pollen, seeds, and spores, extending deep into the geological record. These records are important since they demonstrate that the environmental, ecological and preservational conditions were ripe for the preservation of crown-angiosperms; they are absent because they had not yet evolved.

Coiro and colleagues accept that there remains scope for a Late Jurassic origin of crown-angiosperms – that the fossil record might even require it – but this still leaves yawning mismatch with molecular timescales that infer a fully Jurassic Gap. The discordance likely arises from three main factors: (i) the perfunctory nature of most molecular clock analyses; (ii) the way in which the results of molecular clock analyses are interpreted; and (iii) the way in which fossil evidence is interpreted to calibrate molecular clock analyses.

Molecular clock methods are now so parameter rich that it is possible to obtain just about any desired result. Most molecular analyses follow a simple pipeline that does not explore the impact of rate model, gene sample, partition strategy, competing phylogenetic hypotheses, or the degree to which fossil calibrations approximate the true time of divergence. Each of these factors (and more) can have a profound impact on the ensuing evolutionary timescale [10] and very often there is no objective criterion to choose among them. For this reason, it is more appropriate to integrate the results of experiments that explore parameter space [e.g. 11] which may lead to results that are less precise than we might want - but more likely to yield the accuracy we need.

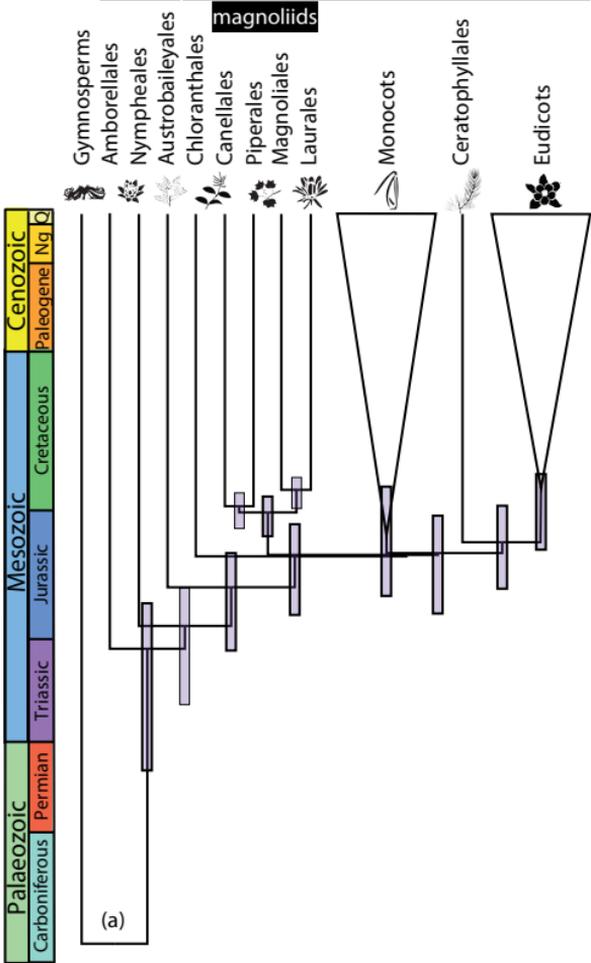
The results of molecular clock analyses are probabilistic, estimating a distribution of clade ages that are commonly summarized by a point estimate, such as a mean or median. However, simulation studies indicate that point estimates are usually wrong while the span of the distribution of age estimates usually encompasses the true clade age [12]. Molecular clock analyses have recovered estimates for the crown age of angiosperms that encompass the late Jurassic interval that Coiro and colleagues predict as the origin time of crown-angiosperms [e.g. 9]. In this sense, the perception of mismatch between molecular clock estimates and the fossil record is more an issue of interpretation rather than fundamental incongruence.

Ultimately, it is naive to perceive molecular timescales and the fossil record as competing and adversarial. The diverse family of molecular clock methods attempt in their different ways, to integrate fossil constraints on the age of living clades with phylogenetic hypotheses, molecular distances, and evolutionary models. To be sure, molecular timescales are hypotheses and they must be tested, such as through the discovery of incompatible fossil evidence. However, rejection in such circumstances merely indicates that their underlying fossil evidence was incorrect, inspiring new analyses based on improved fossil sampling - rather than inspiring despair with the entire enterprise [e.g. 13]. Indeed, fossil calibrations are the soft underbelly of most molecular clock methods since they require fossils to inform not only the minimum age of living clades, but also constrain their maximum age. Inference of a 'Jurassic Gap' in the fossil record of crown-angiosperm evolutionary history is rooted largely in the loose maximum constraints on the age of this clade. An improved evolutionary timescale for crown-angiosperms requires more precise (but no less accurate) fossil constraints. These may not be achievable for many of the infamous molecular clock-fossil record mismatches, but the uniquely rich and dense fossil record of seed plants may hold the key. Coiro and colleagues provide objective evidence for tightening these constraints and their implementation will doubtless lead to a dramatically diminished Jurassic Gap.

Figure 1: Comparison of (a) molecular timescale of angiosperm evolution (summarized from Li et al. [3]) to (b) the pollen record it predicts, and (c) the known pollen record; (b-c after Coiro et al. [4]). Note that fading colours and grey scale of the pollen in (b) reflects the uncertainty associated with the estimates for the timing of origin of crown-angiosperms and eudicots in (a).

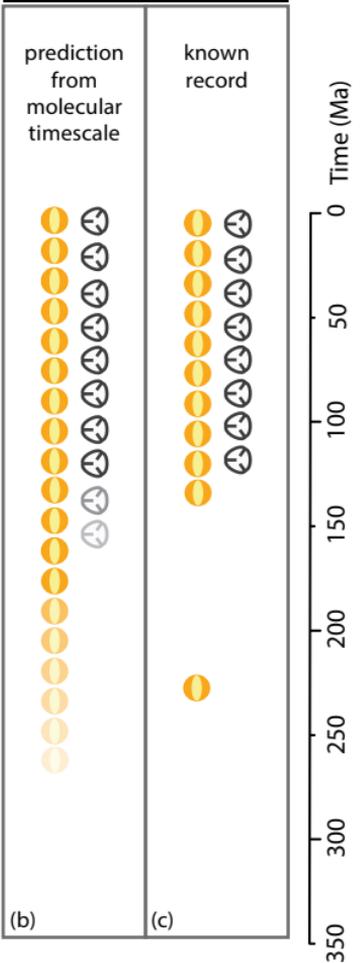
1. Boyce, C.K., and Lee, J.-E. (2017). Plant evolution and climate over geological timescales. *Annual Review of Earth and Planetary Sciences* 45, 61-87.
2. Ramshaw, J.A.M., Richards, M., Richards, D.L., Brown, R.H., Boulter, D., Thompson, E.W., and Meatyard, B.T. (1972). Time of origin of flowering plants determined by using amino-acid sequence data of cytochrome-c. *New Phytologist* 71, 773-779.
3. Abadi, S., Azouri, D., Pupko, T., and Mayrose, I. (2019). Model selection may not be a mandatory step for phylogeny reconstruction. *Nature communications* 10, 934.
4. Coiro, M., Doyle, J.A., and Hilton, J. (2019). How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *The New phytologist* 223, 83-99.
5. Donoghue, P.C., and Yang, Z. (2016). The evolution of methods for establishing evolutionary timescales. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371, 20160020.
6. Herendeen, P.S., Friis, E.M., Pedersen, K.R., and Crane, P.R. (2017). Palaeobotanical redux: revisiting the age of the angiosperms. *Nat Plants* 3, 17015.
7. Holland, S.M. (2016). The non-uniformity of fossil preservation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371.
8. Patzkowsky, M.E., and Holland, S.M. (2012). *Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space*, (Chicago: Chicago University Press).
9. Barba-Montoya, J., Dos Reis, M., Schneider, H., Donoghue, P.C.J., and Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *The New phytologist* 218, 819-834.
10. Warnock, R.C.M., Yang, Z., and Donoghue, P.C.J. (2012). Exploring uncertainty in the calibration of the molecular clock. *Biology Letters* 8, 156-159.
11. Betts, H.C., Puttick, M.N., Clark, J.W., Williams, T.A., Donoghue, P.C.J., and Pisani, D. (2018). Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology & Evolution* 2, 1556-1562.
12. Warnock, R.C.M., Yang, Z., and Donoghue, P.C.J. (2017). Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proceedings. Biological sciences / The Royal Society* 284.
13. Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., and Halanych, K.M. (2015). The Evolution of Annelids Reveals Two Adaptive Routes to the Interstitial Realm. *Current biology : CB* 25, 1993-1999.

crown-angiosperms
ANITA Mesangiospermae



(a)

crown angiosperm pollen fossil record



(b)

(c)

Time (Ma)