



Biases in assessing the evolutionary history of the angiosperm flora of China

Abstract

In their recent paper published in *Nature* (2018, 554, 234–238), Lu et al. use phylogenetic approaches to determine the proportion of the Chinese angiosperm genera that originated during the Miocene or later, and contrast divergence times and phylogenetic dispersion between eastern and western China. One of their key conclusions is that 66% of the angiosperm genera in China originated in the Miocene or later. However, an analysis of 300 angiosperm genera shows that 139 (76.8%) of the 181 genera considered as originating in the Miocene or later by Lu et al. have fossil records before the Miocene. Thus, the evolutionary history of Chinese angiosperm flora has been substantially underestimated in Lu et al. In addition, the results of Lu et al. have been biased by using an incomplete phylogeny.

Phylogeny-based approaches are essential to understanding the origin of differences in species richness between regions. Lu et al. (2018) have recently applied phylogenetic approaches to investigate the evolutionary history of angiosperms (flowering plants) in China. While recognizing the contribution of Lu et al.'s analysis to understanding the evolutionary history of China's flora, I have concerns about the approaches used in their study, which might have substantially biased their conclusions. I comment on two major aspects below.

1 | THE EVOLUTIONARY HISTORY OF CHINESE ANGIOSPERM FLORA MIGHT HAVE BEEN SUBSTANTIALLY UNDERESTIMATED IN LU ET AL

Lu et al. state that “66% of the angiosperm genera in China did not originate until early in the Miocene epoch (23 million years ago (Mya)).” This is one of the key conclusions in their study. This conclusion was based on the observation that stem ages for about 66% of the angiosperm genera in Lu et al.'s phylogeny (i.e. “Dated_phylogeny_for_Chinese_angiosperms.tre”; <https://datadryad.org/resource/doi:10.5061/dryad.p89m3>) are younger than 23 million years. However, many of the angiosperm genera having well-preserved fossils in the Palaeogene (66–23 Mya) and the Cretaceous (145–66 Mya) have ages younger than 10 million years in Lu et al.'s phylogeny. For example, fossils of the genera of Magnolioideae (Magnoliaceae), including *Magnolia*, *Manglietia* and *Talauma*, were abundant in the Eocene and Palaeocene (66–33.9 Mya; Azuma, García-Franco, Rico-Gray, & Thien, 2001), but the earliest divergence time of the 12 genera of Magnolioideae included in Lu et al.'s phylogeny was only 8.1

million years ago. In particular, *Manglietia* originated 66 Mya in the Cretaceous (Azuma et al., 2001) but its age is only 5.6 million years in Lu et al., an underestimate of over 10 times. In general, this discrepancy is not due to the use of different taxonomies or misidentification of fossils. For example, *Sassafras*, *Ulmus* and *Zelkova* are taxonomically stable and have abundant reliable fossils in the deposits of the Eocene (56–33.9 Mya; Shiono, Kusumoto, Yasuhara, & Kubota, 2018); however, stem ages of these genera are younger than 10 million years (in the Late Miocene) in Lu et al.'s phylogeny. For angiosperm tree genera alone, over 100 genera in China have well-preserved fossils in the Palaeogene or Cretaceous, but their ages in Lu et al.'s phylogeny are younger than 23 million years and most (57%) have ages younger than 10 million years (the average age is 6.1 million years in Lu et al., but the average age is 42.8 million years based on fossils; Table S1 in Supporting Information).

Because the vast majority of the Chinese angiosperm genera do not have fossil records and because there is no comprehensive fossil database for the Chinese angiosperm genera, it is not possible at the present time to precisely determine the degree to which the estimate of Lu et al. for the proportion of the angiosperm genera in China that did not originate until early in the Miocene is biased. However, because the vast majority of the Japanese angiosperm genera also occur in China, the recently published fossil database for Japanese plants (Shiono et al., 2018) can be used to estimate the proportion of the genera in China that originated before the Miocene but were assigned to an age of the Miocene or younger in Lu et al. Of the 300 angiosperm genera that were included in Lu et al. and have well-documented fossil records in the Japanese plant fossil database (252 genera; Shiono et al., 2018) and the global database of Cenozoic plant fossils (Xing et al., 2016), 181 genera were considered as originating in the Miocene or later in Lu et al. However, 139 (76.8%) of these 181 genera have fossil records before the Miocene. There are more than 100 genera that were considered as originating in the late Miocene (~10 million years ago) or later in Lu et al. but their fossils were found in the Oligocene; many of them were also found in the Eocene, Palaeocene, and Cretaceous (Figure 1). Apparently, divergence times of many genera are substantially underestimated in Lu et al. Furthermore, the fact that a large proportion of the genera with reliable fossils have substantially younger ages in Lu et al.'s phylogeny, as partially shown in Figure 1, suggests that their phylogeny was poorly dated. As a result, the percentage of the angiosperm genera in China that did not originate until the early Miocene was substantially overestimated in Lu et al. This casts doubt on the conclusions of downstream analyses in Lu et al.

The underestimation of divergence times for many genera in Lu et al.'s study likely resulted from multiple sources of errors. The primary source appears to be insufficient use of fossil records to calibrate node ages, as noted above. Another source for the underestimation of divergence times of genera might be that the age of crown angiosperms used in Lu et al.'s study is too young. Most estimates of the crown angiosperm age fall between 150 and 250 Mya (e.g. 149–256 Mya, Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; 162–210 Mya, Magallón, 2014; 175–240 Mya, Clarke, Warnock, & Donoghue, 2011; 190–238 Mya, Murat, Armero, Pont, Klopp, & Salse, 2017; 182–257 Mya, Smith, Beaulieu, & Donoghue, 2010; 192–251 Mya, Foster et al., 2017; 158–179 Mya, Wikström, Savolainen, & Chase, 2001; 167–199 Mya, Bell, Soltis, & Soltis, 2010). Molecular clocks suggest a Triassic age for crown angiosperms (e.g. Bell et al., 2010; Smith et al., 2010; Zeng et al., 2014; Beaulieu, O'Meara, Crane, & Donoghue, 2015; Foster et al., 2017; Sauquet et al., 2012; Brown & Smith, 2018). In particular, several recent studies have demonstrated that the age of crown angiosperms is >200 million years (e.g. 203 Mya, Barba-Montoya et al., 2018; 232 Mya, Beaulieu et al., 2015; 233 Mya, Brown & Smith, 2018; 221 Mya, Foster et al., 2017; 243 Mya, Zanne et al., 2014; 232 Mya, Zeng et al., 2014). However, Lu et al. took 140 million years as the maximum angiosperm crown age when they built their phylogeny (Supplementary Table 1 of Lu et al., 2018). This age is much younger than the minimum angiosperm crown ages in most angiosperm phylogenies reported in recent studies, as noted above. Multiple lines of evidence have shown that the age of crown angiosperms is much older than 140 million years (Barba-Montoya et al., 2018). Thus, using a substantially underestimated crown angiosperm age in Lu et al. would likely have underestimated the ages of most, if not all, nodes in Lu et al.'s phylogeny. If Lu et al. had used an older age for crown angiosperms to build their phylogeny, such as those used in Zanne et al.'s (2014) and Brown & Smith, 2018's (2018) phylogenies, one would expect that divergence times of many genera in Lu et al.'s study would have increased substantially, which would lead to a much higher proportion of angiosperm genera that originated before the Miocene, compared to ~34% reported in Lu et al.

In addition, the correlation between genus ages in Lu et al.'s phylogeny and those reported in other global phylogenies is relatively low (e.g. $r = 0.579$ between genus ages in Lu et al.'s phylogeny and

those in Zanne et al.'s phylogeny; Figure S1 in Supporting Information), suggesting that using a different phylogeny in Lu et al.'s analysis might yield a substantially different conclusion. Because molecular divergence-time estimates are extremely sensitive to placements of calibrating fossils at crown versus stem nodes and to choices of methods and calibration scenarios (Magallón, 2014; Sauquet et al., 2012; Wilf & Escapa, 2015), it is inappropriate to determine the origination time of a genus, or a taxon at any rank, solely based on molecular date derived from a phylogenetic tree, as in Lu et al. (2018) and other studies (e.g. Song & Cui, 2017). Previous studies have pointed out that great caution should be exercised when using molecular dates to determine divergence times of taxa or to interpret the biological impacts of geological events (Wilf & Escapa, 2015, 2016).

2 | USING INCOMPLETE PHYLOGENY IN LU ET AL. WOULD HAVE OVERESTIMATED AGES OF SOME TAXA

The phylogeny used in Lu et al.'s study is inappropriate for their core analysis, which focused on mean divergence time (MDT). Lu et al. define the divergence time of a genus based on its stem node age in their phylogeny. This approach requires a complete phylogeny at the genus level. Missing genera within a family may substantially bias the estimate of divergence time for any particular genus (Figure 2). Using an incomplete phylogeny will tend to bias divergence times to older values because closer relatives are not present in the phylogeny. Because the phylogeny used in Lu et al. included only Chinese angiosperm genera, genera in Chinese families occurring only outside China are not considered in calculating most recent divergence times. When a family had only one genus in China but two or more genera in the world, Lu et al. accepted the divergence time of the family as the divergence time for the genus. About 18% of the families in Lu et al.'s study have two or more genera worldwide, only one of which occurs in China. If a genus belongs to a family that has an old divergence time but the genus only originated recently (i.e. having a short divergence time), taking the age of the family as the age of the genus might have a substantial influence on the result of an analysis. For example, the family Lecythidaceae has ~25 genera

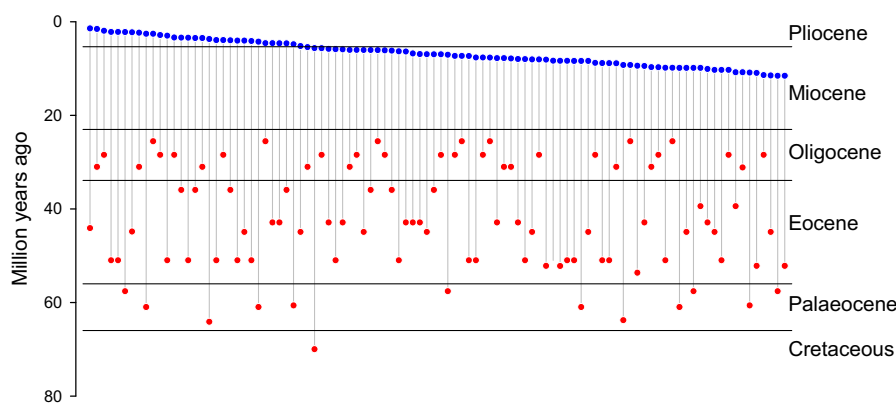


FIGURE 1 Comparison of genus ages reported by Lu et al. (blue dots) and those based on fossils (red dots) for a set of 100 exemplar genera for which genus ages reported by Lu et al. are younger than the Early Miocene (<11.63 million years ago) but reliable Oligocene or older fossils were well documented in the literature (e.g. Xing et al., 2016; Shiono et al., 2018; Appendix S1 in Supporting Information). Each pair of dots represents one genus

worldwide but only one genus (*Barringtonia*) in China, which evolved 9.7 Mya according to Zanne et al. (2014) and 10.2 Mya according to Smith and Brown (2018). Because *Barringtonia* is the only genus of Lecythidaceae in Lu et al.'s phylogeny, Lu et al. took the age of the family (104.4 million years) as the age of the genus, an overestimate of more than 10 times. Similar errors potentially occur in all families for which China does not have all genera of these families, because the bifurcations that are "lost" in Lu et al.'s phylogeny make lineage origins appear much earlier, as shown in Figure 2.

To more quantitatively assess this problem with Lu et al.'s study, I conducted two analyses, which compared divergence times derived from Lu et al.'s phylogeny with those derived from a more complete phylogeny reported by Smith and Brown (2018) for a set of the families which included fewer genera in the former than in the latter. These two phylogenies were built based on the same crown angiosperm age (i.e. 140 million years). The first analysis included 48 genera belonging to the families that each has one genus in Lu et al.'s phylogeny but two or more genera in Smith & Brown's phylogeny (Figure S2 in Supporting Information). The average divergence time of these genera is 86.8 million years in Lu et al.'s phylogeny but is 27.0 million years in Smith & Brown's phylogeny (t test, $P < 0.001$). The second analysis included 94 genera belonging to the families

that each has two or three genera in Lu et al.'s phylogeny but five or more genera in Smith & Brown's phylogeny (Figure S2 in Supporting Information). The average divergence time of these genera is 35.0 million years in Lu et al.'s phylogeny and 23.6 million years in Smith & Brown's phylogeny (t test, $P < 0.001$). These two analyses indicate that using an incomplete phylogeny in Lu et al.'s study has substantially overestimated ages of genera, which would have likely biased the main conclusion of Lu et al.'s study. This problem also occurs in other previous studies that used ages derived from a local or regional phylogeny as the ages (origination times) of taxa under investigation (e.g. Buerki, Devey, Callmender, Phillipson, & Forest, 2013; Crisp & Cook, 2013; Crisp, Cook, & Steane, 2004).

The errors with Lu et al.'s study discussed in this section (i.e. overestimation of genus ages) are a different source of errors from the one discussed in the previous section (i.e. underestimation of genus ages). These two sources of errors act in opposite directions. Although the source of sampling bias discussed in this section pushes ages to be too old, the fact that a large proportion of the genera with fossil records have substantially younger divergence times in Lu et al.'s study than fossil ages, as discussed above (also see Figure 1 and Supporting Information), suggests that divergence times of Chinese angiosperm genera were considerably underestimated in Lu et al.'s study.

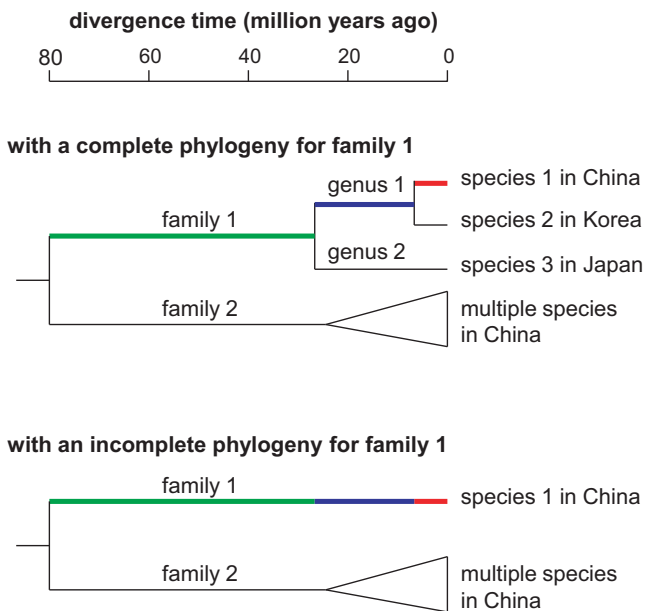


FIGURE 2 A hypothetical case showing that the approach used in Lu et al. can substantially bias the estimation of divergence time of a genus. Divergence times of genera in Lu et al. are estimated from their phylogeny, which includes only Chinese genera. Their approach uses an incomplete phylogeny. If a family has multiple genera in the world but only one genus in China, their approach would take the divergence time of the family as the divergence time of the Chinese genus. In the hypothetical case portrayed in this figure, the divergence time of genus 1 is actually ~27 Mya (the sum of red and blue lines) but it is 80 Mya based on Lu et al.'s approach (i.e. the sum of branch lengths of the red, blue, and green lines), an overestimation by a factor of 3, owing to the failure of including genus 2 in the complete phylogeny

3 | CONCLUSION

It is inappropriate to take ages of taxa, such as angiosperm genera, derived from a molecular-based phylogeny (i.e. phylogenetic ages) as the origination (divergence) times of the taxa, particularly in the case that the phylogeny is poorly dated with fossil records and incomplete, as in Lu et al. (2018). The angiosperm flora of China includes far more genera that predate the Miocene than implied by Lu et al.'s analyses. The calculation of divergence times based on phylogenies that exclude closely related genera outside of China has resulted in overestimation of genus ages, as when recent divergences have involved genera outside of China. The main conclusions of Lu et al.'s study are biased.

ACKNOWLEDGEMENT

I thank Lars Chatrou, Michael Donoghue, Peter Linder, Robert E. Ricklefs and anonymous reviewers for their helpful comments. This study was partly supported by Key Projects of the Joint Fund of the National Natural Science Foundation of China (U1802232).

Keywords

divergence time, fossil record, genus age, incomplete phylogeny, origin of angiosperm genera, phylogenetic structure

ORCID

Hong Qian  <http://orcid.org/0000-0002-1381-7496>

Hong Qian^{1,2} 

¹CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China

²Research and Collections Center, Illinois State Museum, Springfield, Illinois, 62703

Correspondence

Hong Qian, Research and Collections Center, Illinois State Museum, 1011 East Ash Street, Springfield, IL 62703

Email: hqian@museum.state.il.us

REFERENCES

- Azuma, H., García-Franco, J. G., Rico-Gray, V., & Thien, L. B. (2001). Molecular phylogeny of the Magnoliaceae: The biogeography of tropical and temperate disjunctions. *American Journal of Botany*, *88*, 2275–2285. <https://doi.org/10.2307/3558389>
- Barba-Montoya, J., Reis, M. D., Schneider, H., Donoghue, P. C. J., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytologist*, *218*, 819–834. <https://doi.org/10.1111/nph.15011>
- Beaulieu, J. M., O'Meara, B. C., Crane, P. R., & Donoghue, M. J. (2015). Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Systematic Biology*, *64*, 869–878. <https://doi.org/10.1093/sysbio/syv027>
- Bell, C. D., Soltis, D. E., & Soltis, P. S. (2010). The age and diversification of the angiosperms re-visited. *American Journal of Botany*, *97*, 1296–1303. <https://doi.org/10.3732/ajb.0900346>
- Brown, J. W., & Smith, S. A. (2018). The past sure is tense: On interpreting phylogenetic divergence time estimates. *Systematic Biology*, *67*, 340–353. <https://doi.org/10.1093/sysbio/syx074>
- Buerki, S., Devey, D. S., Callmander, M. W., Phillipson, P. B., & Forest, F. (2013). Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society*, *171*, 304–329. <https://doi.org/10.1111/boj.12008>
- Clarke, J. T., Warnock, R. C. M., & Donoghue, P. C. J. (2011). Establishing a time-scale for plant evolution. *New Phytologist*, *192*, 266–301. <https://doi.org/10.1111/j.1469-8137.2011.03794.x>
- Crisp, M. D., & Cook, L. G. (2013). How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 303–324. <https://doi.org/10.1146/annurev-ecolsys-110512-135910>
- Crisp, M. C., Cook, L., & Steane, D. (2004). Radiation of the Australian flora: What can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*, 1551–1571. <https://doi.org/10.1098/rstb.2004.1528>
- Foster, S. P., Sauquet, H., van der Merve, M., McPherson, H., Rossetto, M., & Ho, S. Y. W. (2017). Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Systematic Biology*, *66*, 338–351.
- Lu, L.-M., Mao, L.-F., Yang, T., Ye, J.-F., Liu, B., Li, H.-L., & Chen, Z.-D. (2018). Evolutionary history of the angiosperm flora of China. *Nature*, *554*, 234–238. <https://doi.org/10.1038/nature25485>
- Magallón, S. (2014). A review of the effect of relaxed clock method, long branches, genes, and calibrations in the estimation of angiosperm age. *Botanical Sciences*, *92*, 1–22. <https://doi.org/10.17129/botsoci.37>
- Murat, F., Armero, A., Pont, C., Klopp, C., & Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics*, *49*, 490–496. <https://doi.org/10.1038/ng.3813>
- Sauquet, H., Ho, S. Y. W., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., ... Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of Nothofagus (Fagales). *Systematic Biology*, *61*, 289–313. <https://doi.org/10.1093/sysbio/syr116>
- Shiono, T., Kusumoto, B., Yasuhara, M., & Kubota, Y. (2018). Roles of climate niche conservatism and range dynamics in woody plant diversity patterns through the Cenozoic. *Global Ecology and Biogeography*, *27*, 865–874. <https://doi.org/10.1111/geb.12755>
- Smith, S., Beaulieu, J., & Donoghue, M. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 5897–5902. <https://doi.org/10.1073/pnas.100125107>
- Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*, 302–314. <https://doi.org/10.1002/ajb2.1019>
- Song, J., & Cui, B.-K. (2017). Phylogeny, divergence time and historical biogeography of Laetiporus (Basidiomycota, Polyporales). *BMC Evolutionary Biology*, *17*, 102. <https://doi.org/10.1186/s12862-017-0948-5>
- Wikström, N., Savolainen, V., & Chase, M. W. (2001). Evolution of the angiosperms: Calibrating the family tree. *Proceedings of the Royal Society of London B*, *268*, 2211–2220. <https://doi.org/10.1098/rspb.2001.1782>
- Wilf, P., & Escapa, I. H. (2015). Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. *New Phytologist*, *207*, 283–290. <https://doi.org/10.1111/nph.13114>
- Wilf, P., & Escapa, I. H. (2016). Molecular dates require geologic testing. *New Phytologist*, *209*, 1359–1362. <https://doi.org/10.1111/nph.13831>
- Xing, Y., Gandolfo, M. A., Onstein, R. E., Cantrill, D. J., Jacobs, B. F., Jordan, G. J., & Linder, H. P. (2016). Testing the biases in the rich Cenozoic angiosperm macrofossil record. *International Journal of Plant Sciences*, *177*, 371–388. <https://doi.org/10.1086/685388>
- Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., & Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*, 89–92. <https://doi.org/10.1038/nature12872>
- Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., & Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications*, *5*, 4956. <https://doi.org/10.1038/ncomms5956>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.