The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2014, **113**, 29–49. With 4 figures

Calibrated chronograms, fossils, outgroup relationships, and root priors: re-examining the historical biogeography of Geraniales

KENNETH J. SYTSMA[1,*], DANIEL SPALINK[1] and BRENT BERGER[2]

[1]*Department of Botany, University of Wisconsin, Madison, WI 53706, USA*
[2]*Department of Biological Sciences, St. John's University, Queens, NY 11439, USA*

Received 26 November 2013; revised 23 February 2014; accepted for publication 24 February 2014

We re-examined the recent study by Palazzesi *et al*., (2012) published in the *Biological Journal of the Linnean Society* (107: 67–85), that presented the historical diversification of Geraniales using BEAST analysis of the plastid spacer *trnL–F* and of the non-coding nuclear ribosomal internal transcribed spacers (ITS). Their study presented a set of new fossils within the order, generated a chronogram for Geraniales and other rosid orders using fossil-based priors on five nodes, demonstrated an Eocene radiation of Geraniales (and other rosid orders), and argued for more recent (Pliocene–Pleistocene) and climate-linked diversification of genera in the five recognized families relative to previous studies. As a result of very young ages for the crown of Geraniales and other rosid orders, unusual relationships of Geraniales to other rosids, and apparent nucleotide substitution saturation of the two gene regions, we conducted a broad series of BEAST analyses that incorporated additional rosid fossil priors, used more accepted rosid ordinal topologies, or altered the placement of one fossil Geraniales prior. Our results indicate that their ages are 20–50% too young owing to a combination of (1) strong nucleotide saturation of the DNA regions starting at 65 Mya, (2) lack of a root (rosid stem) or other rosid ordinal stem fossil-based priors, (3) the inability of the two DNA regions (with alignment issues) to obtain a monophyletic Geraniales as well as reasonable relationships of Geraniales to other rosid orders, and (4) apparent issues with the nodal placement of a *Pelargonium* fossil. The Geraniales crown is much older (Campanian of the Cretaceous; 86 Mya), the posterior age distribution on all but two fossil nodes are well older than the priors, the placement of a *Pelargonium*-like fossil is more likely at the crown rather than the stem, but their models of diversification within several clades linked to climatic and orogeny appear supported. We discuss a number of the inherent issues of relaxed-clock dating and outline some 'best practice' approaches for such studies. *Biological Journal of the Linnean Society*, 2014, **113**, 29–49.

ADDITIONAL KEYWORDS: BEAST – best practices – ITS – molecular clock – Myrtales – rosids – substitution saturation – *trnL–F*.

INTRODUCTION

One of the most important and versatile additions to the repertoire of a systematist has been the development of tools that provide the ability to construct dated phylogenies or chronograms in more rigorous or objective ways (Sanderson, 1997, 2002; Drummond *et al*., 2006, 2012a). The utility of phylogenetic chronograms has not been restricted, however, to acquiring better estimates of lineage diversification times (e.g. Wikström, Savolainen & Chase, 2001; Smith & Peterson, 2002; Magallón & Castillo, 2009; Bell, Soltis & Soltis, 2010; Magallón, Hilu & Quandt, 2013). The use of time-calibrated phylogenies has now expanded outside the core systematic endeavour (Daly *et al*., 2012) to help answer a wider range of questions involving, for example, biogeography (Renner, 2005; Harris, Wen & Xiang, 2013), clade diversification (Antonelli *et al*., 2009; Drummond *et al*., 2012b), adaptive radiation (Givnish *et al*., 2009, 2014), community phylogenetics (Davis *et al*., 2005; Bytebier *et al*., 2010), and niche modelling (Evans *et al*., 2009; Töpel *et al*., 2012).

*Corresponding author. E-mail: kjsytsma@wisc.edu

As an accurate timeline is necessary within the phylogenetic framework for conducting these types of studies, considerable attention has been focused on developing 'best practices' for integrating time evidence into the phylogenetic framework. Typically this is now performed using molecular relaxed-clock methods (Drummond *et al*., 2006) that separate the time and substitution rates inherent in a phylogenetic tree, thereby permitting different substitution rates along different branches of the tree (Magallón *et al*., 2013). Many studies have examined how issues associated with tree topology or sequence evolution can compromise the performance of these relaxed-clock methods (reviewed in Magallón *et al*., 2013). For example, the use of sequence data exhibiting high levels of nucleotide substitution saturation (Hugall, Foster & Lee, 2007), or the close proximity of both long and short branches within subclades of a phylogram (Phillips, 2009; Magallón, 2010), are potential sources of error.

The fossil record is increasingly being used as the most important independent evidence to temporally calibrate node ages (Rutschmann, 2006; Yang & Rannala, 2006; Forest, 2009; Parham *et al*., 2012), although other lines of evidence, notably other geological information or secondary calibrations from other studies (reviewed in Pirie & Doyle, 2012), are still relied upon. The issues inherent in the incorporation of fossil-based information as minimum estimates of nodal age, either with point or relaxed prior dates, are still being evaluated in a broad range of studies (e.g. Sanderson, 1998; Magallón, 2004, 2010; Magallón & Sanderson, 2005; Gandolfo, Nixon & Crepet, 2008; Ho & Phillips, 2009; Inoue, Donoghue & Yang, 2009; Hedman, 2010; Ksepka *et al*., 2011; Heled & Drummond, 2012; Parham *et al*., 2012; Pirie & Doyle, 2012; Sauquet *et al*., 2012; Warnock, Yang & Donoghue, 2012; Magallón *et al*., 2013). A general conclusion of these studies is that the use of many fossil calibration points decreases the potential impact of the misplacement or incorrect age of any specific fossil (e.g. Bell *et al*., 2010; Pirie & Doyle, 2012; Sauquet *et al*., 2012).

In this regard, the recently published study by Palazzesi *et al*. (2012), in the *Biological Journal of the Linnean Society*, examining phylogenetics and biogeography in the order Geraniales via a fossil-rich, calibrated time frame, is noteworthy. The Geraniales, as then considered (Angiosperm Phylogeny Group, 2009; hereafter referred to as APG III, 2009), comprises about 900 species in 13 genera in three families (Geraniaceae, Melianthaceae, and Vivianiaceae), with a clear centre of distribution in South America and Africa but with radiations of *Geranium* and *Erodium* into the Northern Hemisphere. Previous rosid or angiosperm-wide dating studies had consistently placed the crown of the Geraniales in the Late Cretaceous (e.g. Wang *et al*., 2009 at 101–88 Mya), but options for internal fossil dating in the order were limited and confined to the Geraniaceae. The calibrated phylogenies for this family indicated Late Miocene to Pliocene for all four genera and suggested increasing aridity and the start of winter-rainfall climates as the causative driver of their diversification (Fiz *et al*., 2008; Fiz-Palacios *et al*., 2010).

Importantly, Palazzesi *et al*. (2012) provided detailed information of five new pollen fossils from the Vivianiaceae, making a total of eight Geraniales pollen fossils that they considered for fossil calibration (although they only used five). The chloroplast DNA (cpDNA) region *trn*L–*trn*F (*trnL–F*), including both the *trnL* intron and the *trnL–trnF* spacer, and the nuclear ribosomal internal transcribed spacer (ITS) region were sequenced for 57 species from all extant genera of Geraniales, along with representatives of rosid outgroup orders (Myrtales, Crossosomatales, and Fagales). Two species of *Ribes* (Saxifragaceae and Saxifragales, superrosid) were used as the ultimate outgroup for rooting purposes as they represent members of the sister group to rosids. The Geraniales phylogeny was dated with BEAST, version 1.6.1 (Drummond & Rambaut, 2007), using a Yule branching process with lognormal priors on five fossil-calibrated nodes – three in Geraniaceae and two in Vivianiaceae. No other tree or node priors were used. The key results from their study include the following. First, recognition of five families (now including Hypseocharitaceae and Francoaceae), not three, is argued based on both morphological coherence and calibrated ages of the five clades (also supported by Fiz *et al*., 2008). Second, the divergence times for major clades within the order or families are approximately 40–50% younger than those estimated by previous studies (Wikström *et al*., 2001; Linder *et al*., 2006; Fiz *et al*., 2008); they argue that the discrepancy is a result of their more accurate time frame with additional internal fossil-calibrated nodes. Third, species diversification within lineages (e.g. within South African *Melianthus* and *Greyia*) occurred recently, after a period of accelerated, major tectonic uplift and accelerated aridification in the Late Pliocene from 3.4 Mya onwards, and not in older Miocene times of minor tectonic uplift and only incipient aridification, as proposed by Linder *et al*. (2006) for the Melianthaceae.

The study by Palazzesi *et al*. (2012), on the historical diversification of the Geraniales, is important as a result of its thorough sampling across genera and families (often resorting to herbarium specimen DNAs) and thereby publishing the most detailed phylogenetic study of the order, presenting a cohesive argument for the recognition of five, rather than

three, families [as in APG III (2009)], its careful description of new fossil pollen evidence for the Vivianiaceae, and its utilization of multiple, independent calibrated nodes in the BEAST analysis. As we have been examining, for some time, the phylogenetic relationships and historical biogeography of the Myrtales and its members, the sister order to the Geraniales (APG III, 2009; Soltis *et al*., 2011), the results of Palazzesi *et al*. (2012) could have implications for our studies (Conti, Litt & Sytsma, 1996; Conti *et al*., 1997, 2002, 2004; Levin *et al*., 2003; Berry *et al*., 2004; Sytsma *et al*., 2004; Graham *et al*., 2005; Berger, 2012). We argue, however, that the methodology employed by Palazzesi *et al*. (2012) contains several significant shortcomings that may severely affect the resulting chronogram, ages of the important nodes discussed in the paper, and thus correlation to external past tectonic and climatic events. As their calibrated age results are already being cited and used in other studies (e.g. Jones *et al*., 2013), it is important that the Geraniales data set be re-examined.

Specifically here we re-analyze the two-gene region data set in BEAST but address three main concerns. First, Palazzesi *et al*. (2012) relied on only two gene regions – *trnL–F* and ITS – both known to be fast evolving (Taberlet *et al*., 1991, 2007; Baldwin, 1992; Baldwin *et al*., 1995; Buckler, Ippolito & Holtsford, 1997; Quandt *et al*., 2004; Guisinger *et al*., 2008, 2011; Poczai & Hyvönen, 2009) to uncover relationships within and among rosid orders. Therefore, there is the potential that substitution saturation may generate bias in under-representing branch lengths, especially in the older stem of the Geraniales and branches among outgroup orders in BEAST analyses, and thereby also bias divergence time estimations (Zheng *et al*., 2011). Second, we note the inability of these two gene regions to recover reasonable relationships among outgroups and to the Geraniales, and thus potentially also severely affecting estimates of temporal diversification more basal to the Geraniales. Third, Palazzesi *et al*. (2012) obtained unusually young ages for the stem root of rosids and other rosid ordinal stem nodes based on already characterized fossils, probably because of the lack of a dated prior for the root of the tree and/or the lack of calibrated nodes in deeper portions of the tree.

MATERIAL AND METHODS

Sequence alignment

We downloaded from GenBank into Geneious 6.1.6 (http://www.geneious.com/) the *trnL–F* and ITS sequences listed in table A1 of Palazzesi *et al*. (2012) representing the 67 accessions used in their reduced BEAST analysis. We aligned each data partition separately and then concatenated. Our aim was to generate alignments that produced phylogenetic trees replicating those published in Palazzesi *et al*. (2012), as we were primarily interested in the effects of adding subsequent priors (e.g. root date and topological constraints among outgroups) and not in the specifics of relationships within Geraniales per se. We explored two methods for aligning the two gene regions and subsequent phylogenetic analyses on each set of alignments. First, we followed the stated methods of Palazzesi *et al*. (2012) and simply aligned the sequences using MAFFT, version 6.814b (Katoh *et al*., 2002). As a result of sections of both gene regions appearing misaligned after MAFFT alignment, especially between species of the two main subclades in Geraniales and among outgroup taxa, we subsequently edited the alignment manually in Geneious. This was performed acknowledging the structural motifs and alignment issues known for both ITS (Buckler *et al*., 1997; Álvarez & Wendel, 2003; Poczai & Hyvönen, 2009) and the *trnL* intron and the *trnL–F* spacer (Borsch *et al*., 2003; Quandt *et al*., 2004; Taberlet *et al*., 2007) and that sole reliance on alignment programs for fast-evolving regions can introduce error (e.g. Graham *et al*., 2000). In this second alignment, we removed portions of ITS and *trnL–F* because of ambiguity in manual alignment. We initially ran concatenated alignments in Garli version 2.0 (Zwickl, 2006) under maximum-likelihood (ML) criteria, following the model of sequence evolution (GTR + Γ + G substitution) described in Palazzesi *et al*. (2012), to evaluate the impact of alignment on topology.

Evaluating levels of sequence and rate divergence across the phylogeny

We were interested in whether the *trnL–F* and ITS sequences used in Palazzesi *et al*. (2012) – spanning the Geraniales and the superrosids – exhibit nucleotide site substitution saturation at some point in the phylogenetic analysis spanning the Geraniales, other rosid orders, and outgroup superrosids. We were also interested in evaluating how clock-like the two gene regions evolved across the span of rosid orders as well as more locally (e.g. within the Geraniaceae). Both of these lines of evidence would provide information in setting clock priors (e.g. ucld.mean or ucld.stdev) for BEAST analyses and interpreting their results. It has been difficult to address how the amount or pattern of changes in sequence substitution rates may affect phylogenetic inference or estimation of dates in BEAST analyses, as it requires adequate modelling of rate change (Hugall *et al*., 2007; Heath, Holder & Huelsenbeck, 2012).

In lieu of more formal analyses of rate change for *trnL–F* and ITS across the phylogeny, we examined pairwise sequence divergence among the taxa relative to age. We first generated a pairwise distance matrix of uncorrected divergences in PAUP* v4.0b10 (Swofford, 2003) using the 'Distance' and 'uncorrected p' options on the concatenated data set after removal of three accessions missing either *trnL–F* or ITS. Second, we obtained a corrected distance matrix under ML using the GTR + Γ substitution model with four discrete gamma categories. We then plotted pairwise sequence distance for a given pair of species against the age of the most recent common ancestor (MRCA) for the two species using the baseline chronogram [replicating the assumptions and priors of Palazzesi *et al*. (2012)] *and* the chronogram based on additional rosid ordinal stem priors and APG III topological constraint (see the following section). Tests for a molecular clock (of all taxa or family subsets) were performed in PAUP* using the tree topology depicted in Palazzesi *et al*. (2012) or based on APG III (2009) and the model of sequence evolution described above. The likelihood of the trees with and without enforcing a molecular clock were obtained and used in a likelihood ratio test for evaluation of significant differences [degrees of freedom (d.f.) = number of tips minus two].

Relaxed-clock analyses

Following the methods in Palazzesi *et al*. (2012), we simultaneously estimated phylogenies and divergence times, on the concatenated data set, in BEAST 1.7.5 (Drummond *et al*., 2012a), using a Yule branching process, the GTR + Γ substitution model with four discrete gamma categories, and a relaxed molecular clock with a lognormal distribution of rate changes. We experimented with a range of values for the molecular clock priors, paying particular attention to 'ucld.stdev,' the parameter associated with rate heterogeneity among branches. Under a gamma distribution, we varied the 'scale' of this prior to regular intervals between 0.001 and 5, evaluating the posterior distribution relative to the prior in Tracer 1.5 (Rambaut & Drummond, 2009) of the BEAST package and how closely the resulting chronogram matched with that in Palazzesi *et al*. (2012). We replicated their Markov chain Monte Carlo sampling by combining two independent chains of 70 000 000 generations, sampling every 10 000 generations, and using a 20% burn-in.

The first BEAST analysis (referred to hereafter as the baseline analysis) incorporated the five fossil lognormal priors described by Palazzesi *et al*. (2012). Because the specifics of the lognormal priors for the fossil-calibrated nodes were not provided in the paper, we set the priors to obtain results as similar as possible to those depicted in their figure 5: *Pelargonium* stem [offset = 28.4 Mya, standard deviation (SD) = 1], *Balbisia* crown (offset = 15.79 Mya, SD = 1), *Viviania* crown (offset = 10 Mya, SD = 1), *Geranium* crown (offset = 7.248 Mya, SD = 1), and *Erodium* crown (offset = 7.246 Mya, SD = 1). We also had to add a topological constraint for the Geraniales clade in the baseline analysis, as BEAST analyses did not always recover a monophyletic Geraniales, and for the relationships within Myrtales, which was monophyletic but exhibited differing generic relationships [all other relationships within and among outgroups were recovered as depicted in Palazzesi *et al*. (2012)]. As we were trying to emulate their results in the baseline analysis, these constraints will make all analyses err on the side of their results. We also redid this baseline BEAST analysis, but only on the subset of Geraniales taxa, by removing all other rosid and outgroup taxa. Additionally, with this reduced Geraniales data set we examined the implications of placing a normally distributed prior on the crown of the Geraniales based on dates from previous rosid-wide studies, dates that Palazzesi *et al*. (2012) mention, but do not use. These crown dates range from 86 to 80 Mya (Anderson, Bremer & Friis, 2005), 98 to 84 Mya (Wikström *et al*., 2001), to 101 to 88 Mya (Wang *et al*., 2009). We thus used a normal distribution on the prior with a mean of 90 Mya and an appropriate SD (of 5.1) to generate a 95% confidence interval (CI) sampling of between 100 and 80 Mya.

We then performed a number of subsequent BEAST analyses to evaluate the impact on nodal dates of additional priors involving topological constraints among outgroup relationships and/or fossil calibrations on the stems of the rosid orders. We first examined the impact of placing three date constraints as priors on more basal nodes to the Geraniales. The date of the root (divergence of Saxifragales and all rosids; this node is referred to as either the 'superrosid crown' or the 'rosid stem') has been estimated in a number of angiosperm-wide studies employing fossils: 121–111 Mya in Wikström *et al*. (2001), 114 Mya in Magallón & Castillo (2009), 132–111 Mya in Bell *et al*. (2010), and 110 Mya in Magallón *et al*. (2013). We set a conservative uniform prior on the rosid stem of between 125 and 101 Mya. Based on the oldest known fossils attributed to the orders Fagales and Myrtales, we also placed priors on the stem nodes of these two orders. *Normapolles* pollen grains from the middle Cenomanian have clear affinities to Fagales (Friis, 1983; Sims *et al*., 1999; Schönenberger, Pedersen & Friis, 2001; reviewed in Magallón *et al*., 2013). We placed *Normapolles* at the Fagales stem node (lognormal distribution, offset = 96 Mya, SD = 0.75) following Magallón *et al*. (2013). The

earliest known fossils of Myrtales include *Esqueiria futabensis* (Takahashi, Crane & Ando, 1999) from the Upper Cretaceous (Lower Coniacian, 88.2 Mya) used by Bell *et al*. (2010) and *Myrtaceidites* (= *Syncolporites*) (Herngreen, 1975; Muller, 1981) pollen from the Upper Cretaceous (Santonian, 86 Mya) used by Sytsma *et al*. (2004). Thus, we used a conservative fossil placement at the Myrtales stem node (lognormal distribution, offset = 88.2 Mya, SD = 1.0). As these rosid fossils have been used previously at the *crowns* of either the Fagales or Myrtales in better sampled studies of these orders, their placement at the *stems* of the two orders in these analyses is conservative – these stem nodes must be older – and thus the analyses err on the side of the results of Palazzesi *et al*. (2012). We ran these BEAST analyses in two sets: with priors enforced on all nodes (three rosid ordinal nodes plus five Geraniales nodes); and without the five Geraniales priors.

Two additional sets of BEAST analyses were performed with all eight nodal priors, but with alternative topological constraints invoking relationships among the rosid orders based on two previous molecular phylogenetic analyses across angiosperms. Palazzesi *et al*. (2012) recovered an unusual ordinal topology – [((Geraniales, (Crossosomatales, Fagales)), Myrtales), Saxifragales] – in their BEAST analysis. We first utilized the ordinal topology of Chase *et al*. (1993), which was based solely on *rbcL* with the assumption that it represents a probable set of outgroup relationships when using limited gene sampling, as in Palazzesi *et al*. (2012) [((((Geraniales, Crossosomatales), Myrtales), Fagales), Saxifragales]. Second, we used the more recent APG III (2009) topology that places Myrtales sister to Geraniales [((((Geraniales, Myrtales), Crossosomatales), Fagales), Saxifragales], which was found with all three genomes (Zhu *et al*., 2007; Wang *et al*., 2009; Qiu *et al*., 2010; Soltis *et al*., 2011). Lastly, we ran two BEAST analyses with the placement of *Tricolporopollenites pelargonioides* at the crown of *Pelargonium* instead of at the stem, considering that all other fossils were used to constrain crown nodes and the ambiguity in terms of its placement within *Pelargonium* (Martin, 1973; Palazzesi *et al*., 2012). The first of these two analyses mirrored the baseline analysis with five Geraniales fossils, except that the *T. pelargonioides* fossil constraint was shifted from the stem to the crown of *Pelargonium*. The second analysis used these five Geraniales fossil priors as in the previous analysis, but included the three rosid stem priors and the APG topology constraint. Following accepted practices (Drummond *et al*., 2012a; Heath *et al*., 2012; Warnock *et al*., 2012) we also compared these results with BEAST analyses that had only priors included and all nucleotide sequences excluded. All chronograms used the geological timescale of Walker *et al*. (2013).

RESULTS

Alignment and phylogenetic analyses with *trnL–F* and ITS

The use of MAFFT in Geneious generated an alignment of both *trnL–F* and ITS that visually needed further manual alignment and that produced an ML tree (data not shown) in Garli incongruent with that depicted in either figure 4 (Vivianiaceae + Francoaceae sister to Melianthaceae) or figure 5 (Melianthaceae + Francoaceae sister to Vivianiaceae) of Palazzesi *et al*. (2012). Manual alignment and exclusion of 119 bp in *trnL–F* and of 245 bp in ITS, as a result of ambiguities in alignment across all rosids, generated a concatenated data set of 1875 bp (1241 bp of *trnL–F* and 634 bp of ITS; total excluded = 16%) (see Appendix S1). The resulting ML tree (data not shown) was almost equal to that depicted in figure 5 of Palazzesi *et al*. (2012) with respect to relationships within families of the Geraniales and among rosid orders. The two exceptions included non-monophyly of the Geraniales and specific relationships within Lythraceae (Myrtales). Geraniales comprises two well-supported, major subclades (Geraniaceae + Hypseocharitaceae vs. Melianthaceae + Francoaceae + Vivianiaceae) but whose stem branch is very short (e.g. figure 5 of Palazzesi *et al*. (2012); see the Discussion). Owing to the weak support in the branches differing between our ML tree and that used in Palazzesi *et al*. (2012), we made topological constraints (monophyly of Geraniales, within Lythraceae) to match the latter tree in all subsequent sequence and BEAST analyses.

Levels of sequence and rate divergence across the phylogeny

Molecular clock tests of the concatenated data set (with GTR + Γ + G) across the entire topology, whether using the rooted phylogram based on Palazzesi *et al*. (2012) or on APG III (2009), significantly rejected the presence of a molecular clock. Molecular clock tests on subclades within Geraniales also all rejected a molecular clock: (1) Geraniaceae + Hypseocharitaceae; (2) Melianthaceae + Francoaceae + Vivianiaceae; and (3) Geraniaceae. Uncorrected sequence divergence (data not shown) ranged from 0% (between two species of *Melianthus*) to 30%. The highest uncorrected sequence divergences of 28–30% occurred in the following: (1) Geraniaceae (*Geranium* and *Erodium*) vs. Vivianiaceae (*Balbisia* and *Viviania*); (2) Myrtales vs. *Balbisia* and *Viviania*; and (3) rosid outgroups vs. *Geranium* and *Erodium*. Corrected (with GTR + Γ + G)

Figure 1. Pairwise sequence divergence of selected representatives from each subclade in Geraniales and outgroup rosid orders versus all other species based on *trnL–F* and ITS data from Palazzesi *et al*. (2012). Corrected sequence divergence, obtained under maximum likelihood (ML) (GTR + Γ + G substitution model), is plotted against the age of the most recent common ancestor (MRCA) of each pair of species compared. (A) Ages obtained from the baseline BEAST analysis with five Geraniales fossil prior constraints replicating the analysis of Palazzesi *et al*. (2012). (B) Ages obtained from the baseline analysis with three additional rosid stem priors and using the APG III (2009) topology.

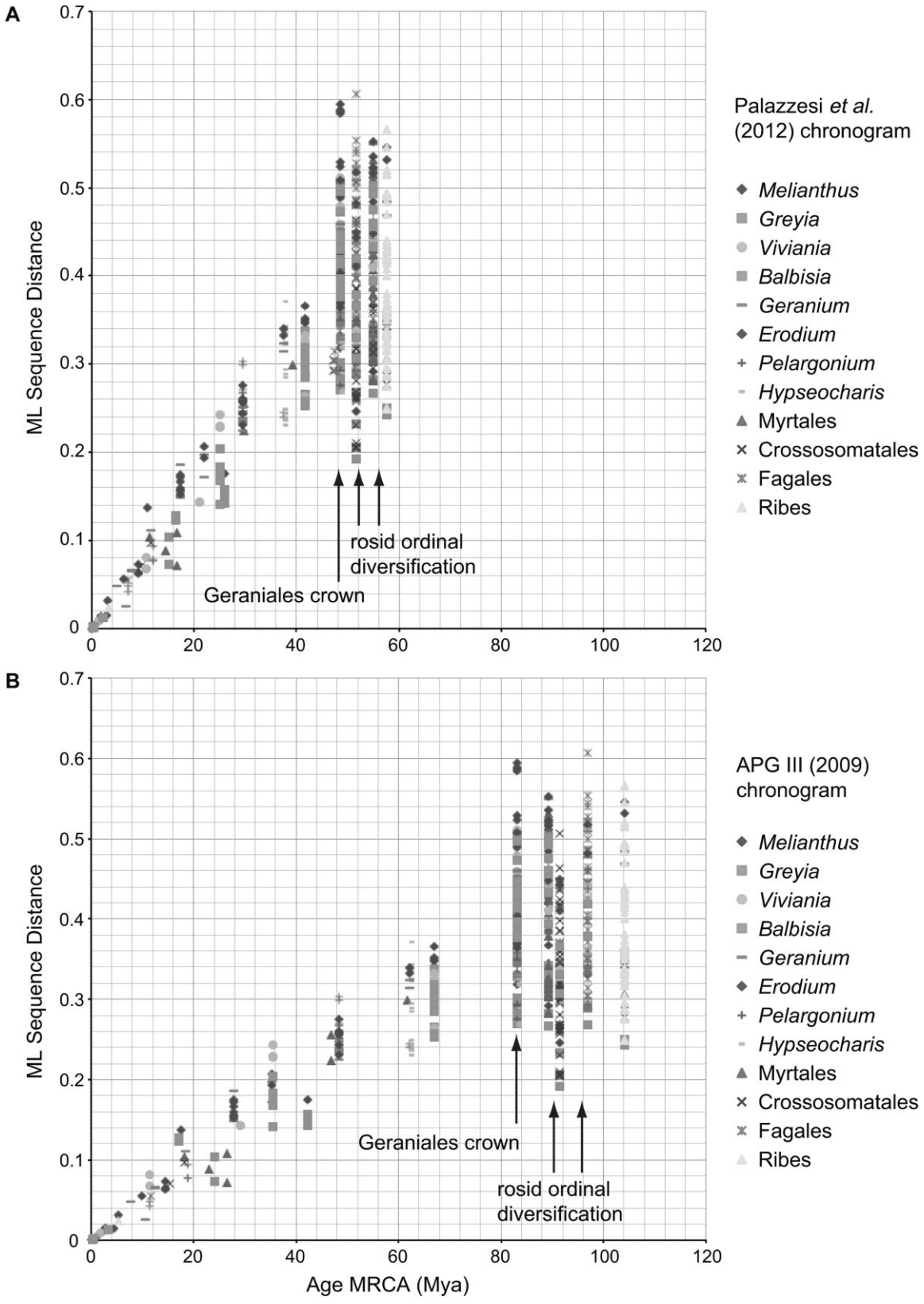
sequence divergence values ranged up to 0.62 substitutions per site (Myrtales or Fagales vs. *Geranium* or *Erodium*) (Fig. 1). The other highest ML sequence divergences (≥ 0.56 substitutions per site) are seen in the following: (1) *Geranium* and *Monsonia* (Geraniaceae) vs. *Viviania*; (2) *Erodium* vs. *Balbisia*; and (3) rosid outgroups vs. *Geranium* and *Erodium*. Thus, in addition to the expected high sequence divergence between the outgroup orders and members of Geraniales, the greatest sequence divergences occurred between Geraniaeae and Vivianiaeae, representatives of the two main subclades within the order Geraniales.

Figure 1 depicts corrected ML sequence distances based on *trnL–F* and ITS among representative species as a function of age, using age estimates from (Fig. 1A) the baseline chronogram using priors of Palazzesi *et al*. (2012), and (Fig. 1B) the APG III topology with rosid stem age prior. Representative species from eight major clades within Geraniales and from each of the outgroup rosid or superrosid orders are plotted against all other species. Sequence divergence appears to show a general correlation with age, especially with the APG III chronogram (Fig. 1B), in the time frame of diversification *within* the two subclades of Geraniales – 0–38 Mya in Figure 1A, or 0–62 Mya in Figure 1B. Large variation in sequence distance relative to time is seen, however, from the root of the tree through the diversification of the order Geraniales into two subclades – 58–42 Mya in Figure 1A or 104–66 Mya in Figure 1B. This latter time frame also includes the problematic node within one subclade representing the split of Melianthaceae + Francoaceae from Vivianiaceae, a relationship not always obtained with the same data set [see fig. 4 vs. fig. 5 in Palazzesi *et al*. (2012)]. Thus, much of the uncertainty or discordance in relationships within Geraniales and among rosid orders and in the time estimates of diversification (Fig. 2; see below) appears to be correlated with both the highest levels and the greatest variation of *trnL–F* and ITS sequence divergence.

Relaxed-clock dating analyses

The ucld.stdev posterior means ranged from *c*. 0.3 to *c*. 0.5 depending on the 'ucld.stdev scale' prior, which we set at intervals between 0.001 and 5.0 under a gamma distribution in our initial exploratory analyses. The SD of the uncorrelated lognormal relaxed clock (in log-space) is indicative of the amount of variation in rates among branches (0, the data are clock-like; > 1, the data exhibit substantial rate heterogeneity among lineages). Despite setting this prior at intervals over several orders of magnitude, all resulting posterior distributions suggest that the rates of molecular evolution for ITS and *trnL–F* do not follow a strict clock in these lineages. Rather, there is a moderate amount of heterogeneity in the rate of molecular evolution among the various branches of the chronogram. Ultimately, we selected a 'ucld.stdev scale' prior of 0.01 (resulting in a posterior of *c*. 0.4) for all subsequent analyses, as this generated chronograms most similar to that presented in Palazzesi *et al*. (2012). The BEAST chronogram derived from the baseline analysis mirroring that of Palazzesi *et al*. (2012) – five fossil priors, no root prior – is depicted in Figure 2A with estimated ages of specific nodes provided in Table 1A. The inferred ages are largely congruent with those reported by Palazzesi *et al*. (2012) (see Table 1A), with the main differences involving the crown ages of *Geranium* and *Erodium*. They found 9.8 and 8.5 Mya, respectively, for these two nodes on which they had placed mean priors of 7.2 Mya. Our baseline analysis pulled those dates older, to 11.5 and 10.9 Mya, respectively, although our 95% credibility intervals strongly overlap theirs. Removing the outgroup taxa and retaining the Geraniales topology had little impact on nodal dates (Table 1A). Adjusting the topology among outgroups to reflect the more likely Chase *et al*. (1993) or APG III (2009) relationships, but retaining the baseline analysis parameters, also had little impact on ages within the order Geraniales (Table 1B).

The simple addition of a conservative and wide prior on the rosid stem (uniform distribution from 125–101 Mya) and on the Fagales and Myrtales stems based on fossils, but maintaining all other baseline priors, had the greatest impact on the dates of the resulting chronogram (Fig. 2B, Table 1C). The rosid stem (= superrosid crown) was dated at 107.0 Mya with all other nodes approximately twofold that seen in the baseline analysis. All rosid ordinal stem lineages, including both the stem and crown of Geraniales (99.0 and 91.5 Mya, respectively), had origins in the middle Cretaceous. Removing all five Geraniales fossil


A
ML Sequence Distance
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
rosid ordinal
diversification
Geraniales crown
Palazzesi et al.
(2012) chronogram
Melianthus
Greyia
Viviania
Balbisia
Geranium
Erodium
Pelargonium
Hypseocharis
Myrtales
Crossosomatales
Fagales
Ribes
B
ML Sequence Distance
Geraniales crown
rosid ordinal
diversification
Age MRCA (Mya)
APG III (2009)
chronogram
Melianthus
Greyia
Viviania
Balbisia
Geranium
Erodium
Pelargonium
Hypseocharis
Myrtales
Crossosomatales
Fagales
Ribes

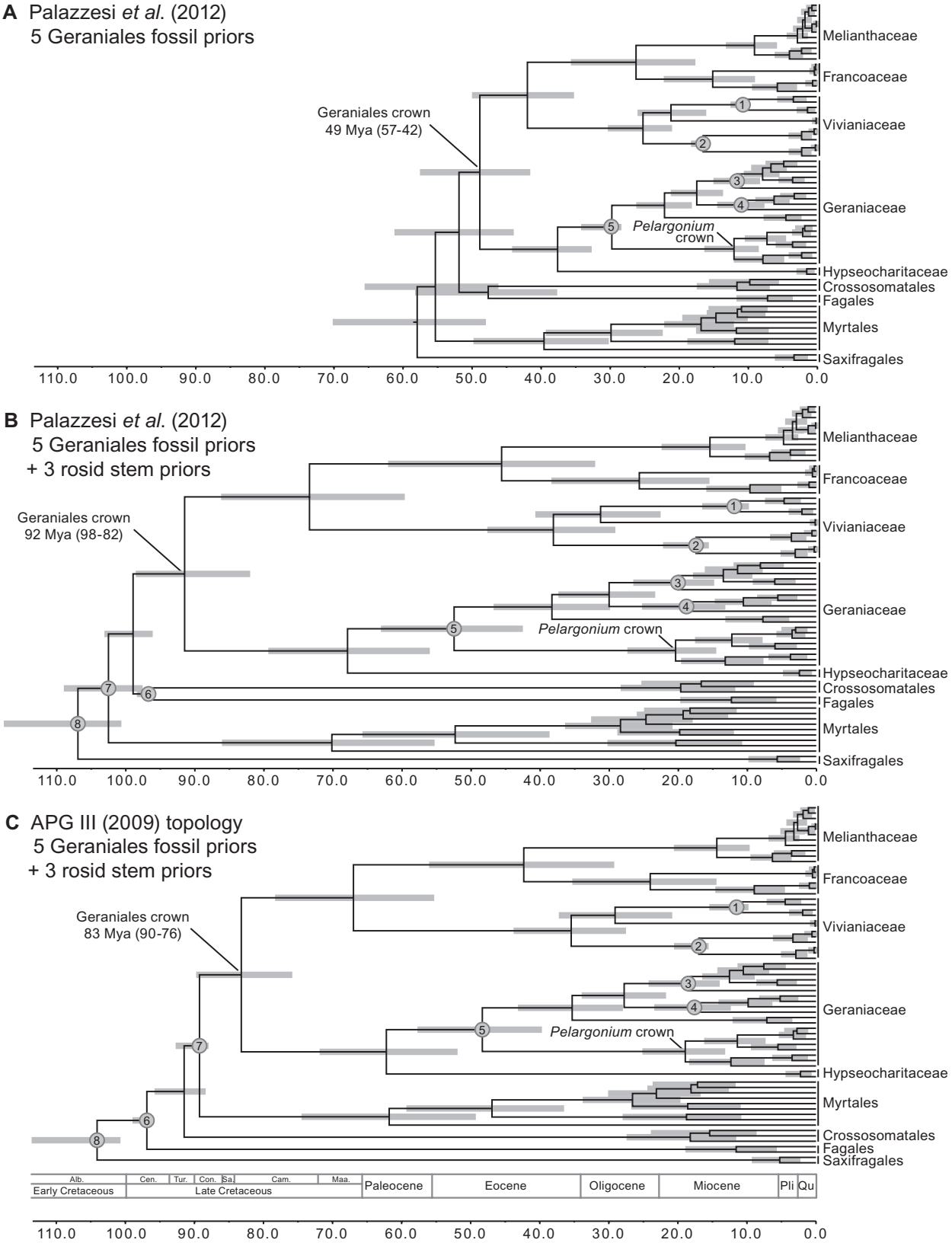

A Palazzesi *et al*. (2012)
5 Geraniales fossil priors
Geraniales crown
49 Mya (57-42)
Pelargonium
crown
Melianthaceae
Francoaceae
Vivianiaceae
Geraniaceae
Hypseocharitaceae
Crossosomatales
Fagales
Myrtales
Saxifragales
110.0 100.0 90.0 80.0 70.0 60.0 50.0 40.0 30.0 20.0 10.0 0.0
B Palazzesi *et al*. (2012)
5 Geraniales fossil priors
+ 3 rosid stem priors
Geraniales crown
92 Mya (98-82)
Pelargonium crown
Melianthaceae
Francoaceae
Vivianiaceae
Geraniaceae
Hypseocharitaceae
Crossosomatales
Fagales
Myrtales
Saxifragales
110.0 100.0 90.0 80.0 70.0 60.0 50.0 40.0 30.0 20.0 10.0 0.0
C APG III (2009) topology
5 Geraniales fossil priors
+ 3 rosid stem priors
Geraniales crown
83 Mya (90-76)
Pelargonium crown
Melianthaceae
Francoaceae
Vivianiaceae
Geraniaceae
Hypseocharitaceae
Myrtales
Crossosomatales
Fagales
Saxifragales
Alb. Cen. Tur. Con. Sa. Cam. Maa.
Early Cretaceous
Late Cretaceous
Paleocene
Eocene
Oligocene
Miocene
Pli Qu
110.0 100.0 90.0 80.0 70.0 60.0 50.0 40.0 30.0 20.0 10.0 0.0

Figure 2. Representative chronograms generated from BEAST analyses of the Geraniales on *trnL–F* and ITS data from Palazzesi *et al*. (2012). (A) Baseline chronogram with five Geraniales fossil prior constraints replicating the results of Palazzesi *et al*. (2012) (see Table 1A). (B) Baseline chronogram with three rosid stem priors added (see Table 1C). (C) Baseline chronogram with three rosid stem priors added and enforcing the APG III topology among orders (see Table 1D). The five recognized families are depicted for Geraniales; species are arranged as in figure 5 of Palazzesi *et al*. (2012). Circled numbers refer to seven fossil-based constraints (1–7) and a rosid stem prior (8) used in different analyses: 1, *Viviania* crown (10 Mya); 2, *Balbisia* crown (15.8 Mya); 3, *Geranium* crown (7.2 Mya); 4, *Erodium* crown (7.2 Mya); 5, *Pelargonium* stem (28.4 Mya); 6, Fagales stem (96 Mya); 7, Myrtales stem (88.2 Mya); and 8, rosid stem (125–101 Mya). Grey bars represent the associated credibility interval (95% highest posterior density). Alb, Albian; Cam, Campanian; Cen, Cenomanian; Con, Coniacian; Maa, Maastrichtian; Pli, Pliocene; San, Santonian; Tur, Turonian; Qu, Quaternary.

priors, and using only the rosid stem priors, pushed the root (110.1 Mya) and the rosid order stems only slightly older, but greatly increased internal nodes in Geraniales (up to 27% and 31% for *Geranium* and *Erodium* crown nodes, respectively) (Table 1C).

The combination of a rosid stem age prior and, more likely, outgroup relationships [based either on Chase *et al*. (1993) or APG III (2009)]) provide ages listed in Table 1D (for xml file, see Appendix S2; for detailed tree, see Fig. S1). Based on the presently available knowledge of rosid relationships, rosid stem node ages, and fossil calibrations provided by Palazzesi *et al*. (2012), Figure 2C depicts the best estimate of temporal diversification within Geraniales (see Table 1D). These results suggest that Geraniales, like other rosid orders, diversified in the late Albian through the Cenomanian or Turonian of the Cretaceous period. The rosid stem node is dated at 104.1 Mya and the Geraniales is an old order, with a crown diversification at 83.2 Mya. The estimated ages are considerably older than the priors used for the *Pelargonium* stem (48 Mya vs. 28 Mya) and both crown ages of *Geranium* (19 Mya vs. 7 Mya) and *Erodium* (18 Mya vs. 7 Mya). Redoing the BEAST analysis after removal of the five Geraniales fossil priors, again gives substantially larger ages for internal nodes in Geraniales (up to 23% and 26% for *Geranium* and *Erodium* crown nodes, respectively) (Table 1D).

Finally, the baseline analysis, but with the placement of the fossil *T. pelargonioides* shifted from the stem of *Pelargonium* to its crown, generated dates within the Geraniales (Table 1E) almost identical to those obtained with the baseline analysis + three rosid stem priors + APG topology constraint (Table 1D) (for detailed tree, see Fig. S2). Adding *T. pelargonioides* to the crown of *Pelargonium*, along with these latter priors, produces a chronogram with slightly older nodal dates (Table 1E) that match those obtained by removing all Geraniales fossil constraints and allowing the chronogram to be based solely on the three rosid stem priors + APG topology (Table 1D).

DISCUSSION

This detailed re-analysis of the temporal diversification within Geraniales provides a contrasting chronogram relative to Palazzesi *et al*. (2012) from which to understand the evolution and radiation of its five constituent families. Our re-analysis provides evidence that the two gene regions are most effective in terms of nucleotide substitutions within the two main subclades of Geraniales, but appear to exhibit pronounced nucleotide substitution saturation (up to 0.62 substitutions per site) and increased variation in levels of pairwise sequence distance, with respect to time, after 65 Myr of divergence. This older time frame corresponds to the separation of the two main subclades within the order and among other rosid orders sampled. We demonstrate that simply placing a reasonable and broad (conservative) uniform prior on the stems of other rosid/superrosid orders, along with the five fossil priors presented in Palazzesi *et al*. (2012), despite this apparent gene saturation, provides a time frame consistent with previously published, but older, ages. We show that this older time frame is maintained when enforcing, in BEAST, several topological constraints among the outgroup orders based on rosid relationships published over the last two decades. Lastly, we discuss how this recalibrated phylogeny of Geraniales impacts interpretations of fossil placement and geological/climatic scenarios of diversification within Geraniales.

trnL–F and ITS evolve too fast to use for dating at this ordinal level

It is readily apparent that *trnL–F* and ITS evolve so rapidly that their use across the entire taxonomic span of Geraniales presents some issues when applied to phylogenetic reconstruction and temporal dating. The first issue involves aligning the sequences among families within Geraniales and across rosid orders, and the need to omit portions of both regions where alignment became problematic. Problems inherent in using ITS and aligning its regions are well known

Table 1. Estimated minimum ages (Mya) of selected nodes within Geraniales and among rosid orders based on BEAST analyses of *trnL–F* and ITS utilizing the baseline parameters (Palazzesi *et al.*, 2012) and subsequent modifications of node or topology (Chase *et al.*, 1993; APG III, 2009) priors

BEAST analysis or previous study	*Melianthus* crown	*Pelargonium* stem (prior 28.4)	*Geranium* crown (prior 7.2)	*Erodium* crown (prior 7.2)	Geraniales crown (prior 90)	Fagales stem (prior 96)	Myrtales stem (prior 88.2)	Rosid stem (prior 101–125)
A. Baseline	2.8	29.5*	11.5*	10.9*	48.6	47.4	55.0	57.6
Palazzesi *et al.*, 2012	3.4	29.0*	9.8*	8.5*	49.5	46.5	54.0	58.0
Baseline: only Geraniales taxa	2.6	30.2*	11.5*	11.1*	51.0	–	–	–
Baseline: only Geraniales taxa + Geraniales crown prior	3.7	46.0*	16.7*	16.1*	81.6*	–	–	–
B. Baseline: Chase topology	2.8	29.6*	11.6*	10.8*	48.7	55.8	54.0	58.4
Baseline: APG topology	2.8	29.6*	11.4*	10.8*	48.9	55.0	52.2	57.6
C. Baseline + three rosid stem priors	4.8	52.5*	20.0*	18.8*	91.5	96.8*	102.6*	107.0*
Baseline + three rosid stem priors, minus five fossil priors	5.3	62.1	25.3	24.7	95.3	96.9*	105.5*	110.1*
D. Baseline + three rosid stem priors, APG III topology	**4.4**	**48.3***	**18.5***	**17.6***	**83.2**	**96.9***	**89.3***	**104.1***
Baseline + three rosid stem priors, APG III topology, minus five fossils	4.8	55.4	22.8	22.2	85.2	96.9*	89.8*	104.3*
E. Baseline, *Pelargonium* crown prior	4.5	50.6	18.9*	17.9*	77.6	75.7	88.0	92.6
Baseline + three rosid stem priors, APG III topology, *Pelargonium* crown prior	4.8	54.1	20.0*	19.1*	84.5	96.9*	89.5*	104.4*
F. Fiz *et al.* (2008)	–	45.8**	15.0	15.5	78.5	–	–	–
G. Linder *et al.* (2006)	19.7	–	–	–	–	–	–	–

Bold indicates BEAST analysis with resulting estimated dates argued to provide the best estimates with these two gene regions.

*Represent some nodes that age priors were used in a specific analysis; *Pelargonium*, *Geranium*, *Erodium* (*Viviania* and *Balbisia* are not shown) were used as five priors in baseline analysis.

**Represents one (38–47 Mya prior) of three priors used in Fiz *et al.* (2008).

(Álvarez & Wendel, 2003; Poczai & Hyvönen, 2009): sequence accuracy, high GC content, indel accumulation, and alignment between outgroup and ingroup members. The limitation of phylogenetic breadth of ITS relative to other genes has been shown in a number of publications (reviewed in Álvarez & Wendel, 2003).

A considerable amount of effort has gone into demonstrating that fast-evolving, non-coding cpDNA regions (specifically *trnT–trnL–trnF*) still appear to be effective at basal nodes within angiosperms in generating resolved phylogenetic trees that are congruent with those trees based on many, but slower-evolving, coding cpDNA regions (e.g. Borsch *et al*., 2003; Quandt *et al*., 2004; Müller, Borsch & Hilu, 2006; Worberg *et al*., 2007; Borsch & Quandt, 2009; Barniske *et al*., 2012). The *trnL–F* region sampled in the present study consists of a group I intron in *trnL* (UAA) and a transcribed spacer between *trnL* and *trnF* (GAA), both of which contain some conserved elements but also large mutational hotspots involving frequent microstructural changes as well as substitution events (Borsch & Quandt, 2009). Throughout land plants, the length and GC content of the *trnL* intron and the *trnL–F* spacer are highly variable, with the *trnL–F* spacer less conserved than the *trnL* intron (Borsch *et al*., 2003; Quandt *et al*., 2004). Alignment of *trnL–F* is difficult; we excluded 8.8% of this spacer because of uncertainty in alignment. Borsch *et al*. (2003) outlined a seven-step method for alignment of the *trnT–F* region but that can still require removal of up to 20% of the region (e.g. Borsch *et al*., 2003; Borsch & Quandt, 2009).

The second issue resides in the apparent saturation of *trnL–F* and ITS nucleotide substitutions after about 65 Myr of divergence (Fig. 1B). At this time interval the amount of uncorrected sequence divergence reaches a high of 30%, and this is seen between Geraniaceae and Vivianiaceae, representatives of the two main, and weakly supported, subclades of Geraniales. Corrected values of substitutions per site reach highs of 0.56–0.62 between these two subclades and between Geraniaceae and all outgroup orders. It may be no coincidence that Geraniaceae exhibit some of the highest sequence divergences within this data set, a majority of which is cpDNA (66%). Geraniaceae chloroplast genomes are the most rearranged of angiosperms, exhibit accelerated rates of nucleotide substitutions, and show widespread loss of genes and spacers, including the *trnT* gene just upstream of the sequenced *trnL–F* region (Guisinger *et al*., 2008, 2011). Simulation studies have suggested that substitution saturation may not occur until higher levels of sequence divergence than were previously thought (Yang, 1998). Furthermore, it was argued that data sets with uncorrected sequence divergences of 15–20%, or a pairwise sequence distance of 0.2–0.3 substitutions per site, do not necessarily exhibit saturation. Yang (1998) suggested that 30% uncorrected sequence divergence, as seen in Geraniales, should be a starting point for concerns about site saturation in fast-evolving regions. He also noted that high evolutionary rates are often associated with other serious sequence issues, namely alignment. The concatenated Geraniales data set thus exhibits both issues – difficult alignment and site saturation.

Although phylogenetic reconstruction may still be possible with *trnL–F* (and ITS) at deep (e.g. Cretaceous) levels within angiosperms, as some have argued (Borsch *et al*., 2003; Borsch & Quandt, 2009), it does not imply that these regions can be effectively used for dating phylogenies at these deep levels (e.g. Zheng *et al*., 2011). It is therefore not surprising that *trnL–F* and ITS could not always recover a monophyletic Geraniales past 65 Mya, uncover reasonable relationships among rosid orders, or estimate time divergences of the most basal nodes consistent with those from any previous rosid or angiosperm relaxed-clock study. In comparison with similar studies within Myrtales, the sister order to Geraniales, only relatively slowly evolving coding regions (e.g. *rbcL*, *matK*, *ndhF*, and 18S) could be used at the ordinal level *among* families (Conti *et al*., 1996, 1997, 2002; Sytsma *et al*., 2004; Berger, 2012). In studies examining *within*-family relationships, *trnL–F* and/or ITS were either not used (Renner *et al*., 2001; Levin *et al*., 2003) or were used effectively but only in concert with other, more slowly evolving, genes (Levin *et al*., 2004; Graham *et al*., 2005; Thornhill *et al*., 2012), and/or required substantial removal of unalignable areas (Graham *et al*., 2005). Within this 65-Mya window, however, the relative ease of alignment, apparent non-saturation of nucleotide substitutions, and considerable sequence variability make these two gene regions presumably effective in uncovering phylogenetic relationships within the two main clades of Geraniales. Although differing in some details, the results of Palazzesi *et al*. (2012), within and among closely related families in Geraniales, match those obtained in other studies that often use different gene regions (Fiz *et al*., 2006, 2008; Linder *et al*., 2006; Fiz-Palacios *et al*., 2010). Thus, despite these difficulties in using *trnL–F* and ITS among orders, the data of Palazzesi *et al*. (2012) provide good evidence for recognizing five families in Geraniales, even though the relationships among the five are not all known with certainty.

The amount of sequence divergence and the rate of nucleotide substitutions are important determinants in generating accurate and supported phylogenetic relationships. In the absence of clock-like evolution, the relative rate of nucleotide substitution in

branches across a phylogram is a critical component in estimating absolute divergence times and absolute rates with relaxed-clock models (Drummond *et al*., 2006; Magallón *et al*., 2013). DNA regions (such as *trnL–F* and ITS in Geraniales and related rosids) that not only accumulate mutations at different relative rates across the phylogeny but also begin to exhibit nucleotide site saturation, would be expected to unduly influence estimates of temporal diversification using either fixed-rate or relaxed-clock methods. As reviewed by Magallón *et al*. (2013), incorrect divergence time estimates have been attributed to unaccounted substitutional saturation in the gene regions used in relaxed-clock dating (Hugall *et al*., 2007; Phillips, 2009; Brandley *et al*., 2011). This effect of substitutional saturation upon divergence time estimation becomes especially severe when few nodes have prior constraints (Magallón *et al*., 2013) or, as demonstrated in this re-analysis of Geraniales, when temporal calibrations are restricted to nodes closer to the tips. The overall effect of the combination of substitutional saturated gene regions and tip-placed fossil priors is to 'pull' dates of older nodes towards the tip and make them appear younger. Of some concern is the question of whether obtaining substitution rates from a single parametric distribution, as performed in BEAST, for example, is adequately modelling the underlying molecular evolution (personal communication, Susana Magallón). Several studies (Zheng *et al*., 2011; Dornburg *et al*., 2012; Wertheim, Fourment & Kosakovsky Pond, 2012; Magallón *et al*., 2013) have suggested, at least with current methods, that the ability to estimate correct rates is severely compromised in cases where substitution rates vary markedly across a phylogram or where rate changes occur along a single lineage (heterotachy). To what degree these issues are involved with the Geraniales data set remains to be seen.

Are root or basal rosid stem node calibration priors necessary? Simply adding rosid stem and rosid order stem fossil priors (with or without invoking the APG III (2009) topology) to the previous constraints used by Palazzesi *et al*. (2012) dramatically changed inferred ages both within Geraniales and among rosids (Fig. 1B, C; Table 1D). The age of the Geraniales stem [89 Mya, 92–88 (95% highest posterior density)] and crown (83 Mya, 89–76) disagree dramatically from that estimated (53 and 49.5 Mya, respectively) by Palazzesi *et al*. (2012), but match quite closely those suggested by previous rosid or angiosperm-wide studies cited by Palazzesi *et al*. (2012), but not further discussed. Thus, in the context of the Geraniales and the use of *trnL–F* and ITS, fossil-based priors on basal nodes and a reasonable prior on the root based on secondary calibrations were critical additions to the fossil-based priors at the more recent crown or stem of five genera.

We add to the growing set of evidence that adding 'realistic' priors to the root or basal nodes may be important components to effectively calibrate temporal diversification of lineages in relaxed-clock analyses. Similarly to our results with Geraniales, recent studies have emphasized that BEAST performs more reliably under the assumption of a basal constraint or placement of fossil priors away from the tips of the phylogeny (McCormack *et al*., 2010; Feldberg *et al*., 2013). However, the majority of studies have demonstrated that if a reasonable prior to limit the age of the root is *not* imposed, relaxed-clock methods may mistakenly estimate unrealistic *old* ages (e.g. Yang & Rannala, 2006; Benton & Donoghue, 2007; Donoghue & Benton, 2007; Ho, 2007; Hugall *et al*., 2007; Marshall, 2008; Wilkinson *et al*., 2011). Our results indicate that the lack of a reasonable root prior for Geraniales, as in Palazzesi *et al*. (2012), estimates unrealistic *young* ages – but this is probably a result of the choice of: (1) gene regions that appear site saturated over the base of the phylogeny; and (2) fossils necessitating near tip node priors.

Fossil priors in Geraniales re-examined

What does this re-analysis of temporal diversification in Geraniales, now best seen as a late Cretaceous crown radiation (89–76 Mya; Fig. 2C, Table 1D), suggest about the efficacy of the fossil priors used in Palazzesi *et al*. (2012)? The baseline BEAST analysis (Table 1A) generated date posteriors (95% highest posterior density) that overlap strongly the prior for three calibrated fossil nodes (*Viviania* and *Balbisia* crowns and *Pelargonium* stem). The posterior marginal densities for these three fossil priors in the baseline BEAST analysis, as seen in Tracer 1.5, are lognormal like the original priors (Fig. 3A, C). Our re-analysis (but not that of Palazzesi *et al*., 2012) places the 95% credibility intervals for the crown nodes of *Erodium* and *Geranium* outside (older) the prior, although they are close. The posterior marginal densities for these two fossil priors in the baseline BEAST analysis are *not* lognormal and show a more normal distribution at older ages (Fig. 3B). Use of rosid ordinal stem priors and the APG III (2009) topology strongly overrides all but the *Viviania* and *Balbisia* crown priors (Table 1D; Fig. 3). The crowns of *Geranium* and *Erodium* are estimated at 18.5 and 17.6 Mya, respectively, relative to their priors of 7.2 Mya. Despite being the oldest known fossils attributed to these clades, their young ages, relative to the ages recovered in our variously calibrated phylogenies (both utilizing and omitting them as priors), suggest that their use as crown priors may be inappropriate.

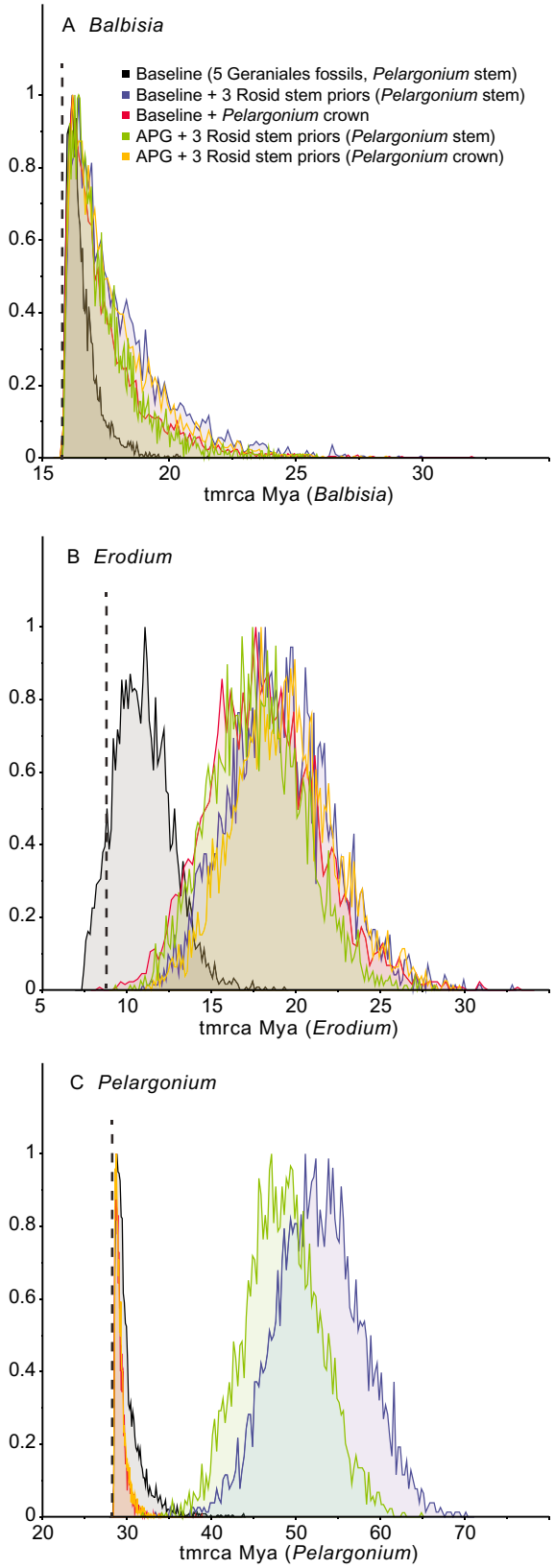


Figure 3. Bayesian posterior density plots for lognormal priors on fossil-calibrated nodes with five different BEAST analyses (see Table 1). Plots were obtained in Tracer 1.5 and are viewed as relative marginal densities. The *x* and *y* axes indicate the nodal age (in Mya) and relative density, respectively. Dashed vertical lines indicate the offset for each lognormal prior. (A) Posterior densities for the *Balbisia* fossil (prior: offset = 15.79 Mya, SD = 1) are lognormal under all BEAST analyses, reflecting the original prior. The *Balbisia* and *Viviania* (density plots not shown, but similar to that of *Balbisia*) fossils are interpreted as correctly placed at their respective crowns. (B) Posterior densities for the *Erodium* fossil (prior: offset = 7.246 Mya, SD = 1) are not lognormal and indicate that most of the normally distributed sampling at the crown is outside (older than) the prior. The *Erodium* and *Geranium* (density plots not shown, but similar to that of *Erodium*) fossils are interpreted to best calibrate nodes within their respective crown radiations. (C) Posterior densities for the *Pelargonium* fossil prior (*Tricolporopollenites pelargonioides*) – placed at either the stem or crown node – (prior: offset = 28.4 Mya, SD = 1) and providing support for correct placement (relative to other fossil calibrations) at the crown, not the stem. Marginal density is lognormal (as in the prior), as expected for the baseline analysis (black) where the prior was placed on the stem of *Pelargonium* and no older rosid priors were used. The inclusion of older rosid fossil calibrations (green, blue) indicates that the initial prior on the *Pelargonium* stem is overridden and normal distributions over much older mean dates are obtained. Placing the prior on the *Pelargonium* crown in either the baseline analysis (red) or with inclusion of rosid priors (orange) provides lognormal distributions and indicates that the prior on the crown is being sampled.

More dramatically, the stem age of *Pelargonium*, based on the fossil prior of 28.4 Mya (and recovered in the baseline analysis), is strongly not supported when additional rosid stem priors are used – the stem age increases to 48.3 Mya (40.0–57.4). The posterior marginal densities for the *Pelargonium* fossil prior with the addition of older rosid node priors are *not* lognormal and exhibit a striking shift to considerably older ages (Fig. 3C). These results indicate that the prior is being overridden at the *Pelargonium* stem. It is not clear why the fossil *T. pelargonioides* was used by Palazzesi *et al*. (2012) to place a prior on the stem, rather than the crown of *Pelargonium*, in keeping with the crown placements of the other four fossils used. No justification for this placement is given, nor is there support in the cited papers (Martin, 1973; Macphail, 1999). Indeed, when we used *T. pelargonioides* as a prior instead for the crown of *Pelargonium* in the baseline analysis with only the four other Geraniales fossils (Table 1E; see Fig. S2), the resulting dates across the Geraniales (and to

some extent for rosid order stems) are very similar to dates obtained with its placement at the stem of *Pelargonium* but adding in the rosid stem priors (Table 1D). Dramatically, the posterior marginal densities for the *Pelargonium* fossil prior placed now at the crown, with the addition of older rosid node priors, are lognormal, like the original prior (Fig. 3C). We argue that the weight of evidence indicates that *T. pelargonioides* should be used as a crown *Pelargonium* prior. These discrepancies in prior versus posterior nodal dates for *Geranium*, *Erodium*, and especially *Pelargonium*, support the observation of Magallón *et al*. (2013: 564) 'that if the signal in the data indicates an age different from the constraint, it will overcome a strong prior'. These results also demonstrate that reliance on one or few fossil priors can be misleading; it was only with the addition of more fossil priors widely distributed on the tree that issues with specific fossil priors were identified.

Pattern and timing of diversification in Geraniales re-examined

Finally, how does this revised time estimate of diversification within Geraniales shape our understanding of how intercontinental distribution patterns formed, the role of progressively intensive aridification in the diversification of South African clades, and the impact of Andean orogeny on South American clades? Despite the older dates in our re-analysis, the three striking intercontinental disjunctions between Africa and South America within Geraniales are still best explained by long-distance dispersal, as argued by Fiz *et al*. (2008) and Palazzesi *et al*. (2012) (see chronograms and dates in Figures 2C and 4, Fig. S1). The split of the South American family Hypseocharitaceae and largely African-basal Geraniaceae (see Fiz *et al*., 2008) dates to 62 Mya (72–52) in the Paleocene, versus 57 Mya by Fiz *et al*. (2008) and 37 Mya by Palazzesi *et al*. (2012). Likewise, the separation of the South American Vivianiaceae from South African Melianthaceae + Francoaceae (South Africa and South America) also occurred in the Paleocene at 67 Mya (78–56), versus 45 Mya as reported by Palazzesi *et al*. (2012). The intercontinental disjunction within Francoaceae is dated at 27 Mya (35–15), versus 19 Mya, as proposed by Bell *et al*. (2010), and 11 Mya, as proposed by Palazzesi *et al*. (2012). The actual direction of dispersal (between South America and South Africa) for these three disjunct pairs is uncertain (without a larger biogeographical analysis of related rosids) as the two large subclades within Geraniales both exhibit the South America + Africa pattern. In any case, these oldest possible dispersal events postdate the separation of the two continents already by 110 Mya in southern areas and by 90 Mya in northern areas (Pittman *et al*., 1993) – see discussion and dates for other South American and African disjuncts – Givnish *et al*. (2000, 2004, 2011), Renner (2004), Sytsma *et al*. (2004).

Our re-analysis (Table 1F; Fig. 2C; Fig. S1) provides considerably older ages for *Geranium*, *Erodium*, and *Pelargonium* (Geraniaceae) than seen in Palazzesi *et al*. (2012). Our data indicate that the Geraniaceae arose 48 Mya (57–39) as compared with 29 Mya in Palazzesi *et al*. (2012). The older phylogenetic and biogeographical study of Geraniaceae (Fiz *et al*., 2008), using *rbcL* and different programs and calibration points, gives a range for the crown of Geraniaceae at 47–38 Mya, dates that overlap strongly with our 95% credibility intervals. The temporal diversification model of Geraniaceae proposed by Fiz *et al*. (2008), in contrast to that in Palazzesi *et al*. (2012), is thus largely consistent with our dates.

Despite the increased ages seen in much of the Geraniales chronogram (Figs 2C and 4) relative to the baseline chronogram (Fig. 2A), two models of diversification proposed by Palazzesi *et al*. (2012) are supported by our analysis. These include the diversification within Melianthaceae and *Greyia* (Francoaceae) relative to periods of intensification of aridity in South Africa, and diversification within Vivianiaceae and *Francoa/Tetilla* (Francoaceae) relative to Andean orogeny. This is a result of the near similarity of dates in the most recent branches (< 10 Mya) seen in all chronograms for the discussed radiations of *Melianthus*, *Greyia*, *Viviania*, *Balbisia*, *Francoa*, and *Tetilla*. Figure 4 depicts the temporal diversification within this clade of three families relative to timing of aridification in South Africa and uplift of the Andes based on our re-analysis (derived from Fig. 2C, Fig. S1). Linder *et al*. (2006) had previously generated a chronogram of Melianthaceae based on a single secondary calibration point using the divergence times of *Bersama* and *Greyia* from Wikström *et al*. (2001). Using this calibration method they argued that the largely South African *Melianthus* had diverged in the early Miocene around 20 Mya followed by separation of eastern and western clades by 15 Mya in the middle Miocene, the latter coincident with – and probably correlated to – the first onset of aridification in South Africa (see Fig. 4). This aridification in a previously (early Miocene) warmer and more mesic landscape was largely a result of glaciation in Antarctica, a decrease of sea surface temperature, a strengthening of the south Atlantic high pressure cell, and the establishment of a winter-rainfall climate along the west coast of South Africa (Linder *et al*., 2006). Our calibrated phylogeny supports the finding of Palazzesi *et al*. (2012) that these dates are too old (Table 1A, D, G). Importantly,

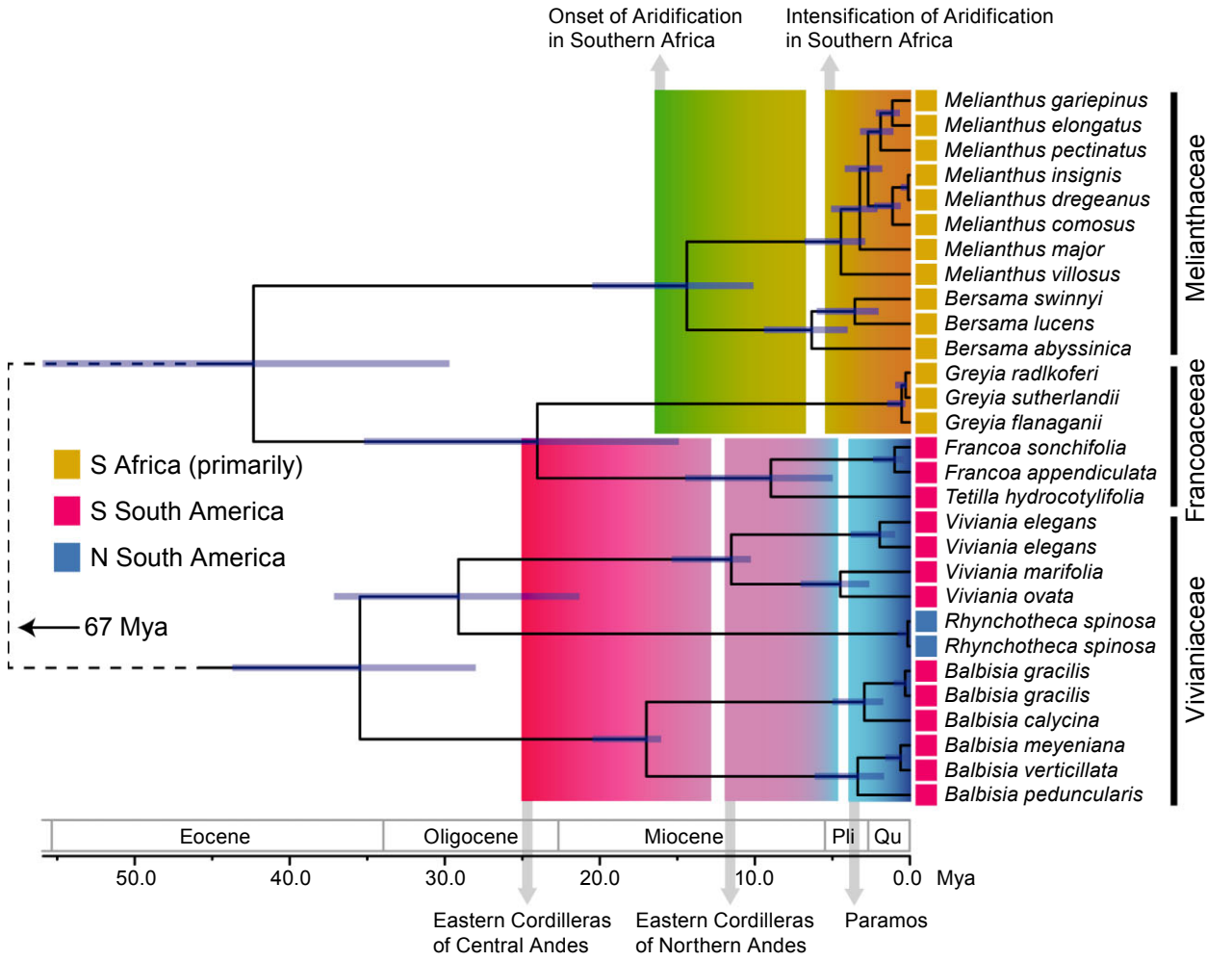


Figure 4. Chronogram of the Melianthaceae, Francoaceae, and Vivianiaceae derived from the baseline chronogram with three rosid stem priors added and enforcing the APG III topology among orders (see Fig. 2C, Fig. S1). Species are scored for geographical location. Time periods for aridification in South Africa are obtained from Linder *et al*. (2006) and Palazzesi *et al*. (2012). Time periods for the uplift of the central and northern regions of the Andes are based on Antonelli *et al*. (2009), Graham (2009), and Chaves, Weir & Smith (2011). Blue horizontal bars represent the associated credibility interval (95% highest posterior density). Pli, Pliocene; Qu, Quaternary.

it indicates that Melianthaceae as a family first diversified during the middle Miocene at 14.4 Mya (20–10) during this first onset of aridification in South Africa, and that the subsequent and rapid diversification of *Melianthus* and *Bersama* (as well as *Greyia* in Francoaceae) occurred during the Pliocene and Quaternary intensification of aridity across South Africa (Fig. 4).

Lastly, our chronogram (see Fig. 4) supports the conclusions of Palazzesi *et al*. (2012) concerning the temporal pattern of diversification in the two South American clades of Vivianiaceae (*Viviania, Balbisia*, and *Rhynchotheca*) and Francoaceae (*Francoa/Tetilla*). These genera, with the exception of *Rhynchotheca*, are restricted to southern South America, mainly west of the Andes with some species east of the Andes. *Rhynchotheca* is the only central to the northern Andean genus of Geraniales, largely confined to 3000–3500 m elevation scrub forest or cloud forest margins. Our analysis places the crown of Vivianiaceae at 35 Mya (44–28) near the Eocene/Oligocene border, well before the rise of the eastern cordilleras of the central and (later) northern Andes (see Fig. 4). *Rhynchotheca* diverged from the *Viviania* lineage shortly thereafter (*c*. 29 Mya). Our dates, and the presence of fossil material for all three genera of Vivianiaceae in the southernmost regions of South America that predates the main uplift of the central and northern Andes, support the scenario outlined by Palazzesi *et al*. (2012). They proposed that the

Vivianiaceae arose and diversified in southernmost South America, migrated north as aridification intensified to the north, but displayed niche conservatism as they largely tracked similar climatic conditions to the present, despite the opportunities of moving into new climatic and ecologic niches afforded by the orogeny of the Andes.

The exception within Vivianiaceae to this niche conservatism is *Rhynchotheca*. It is of Oligocene origin (29 Mya), certainly arose in lower elevations of southern South America as a result of the presence of fossil pollen in Patagonia at 16 Mya (Palazzesi *et al*., 2012), but has tracked into northern Andean high-elevation communities (Peru, Ecuador) as these regions recently uplifted (Fig. 4). Thus, at least two now exclusively higher Andean lineages, *Hypseocharis* – sister to Geraniaceae (Fig. 2C) and restricted to high alpine regions of Bolivia/Argentina – and *Rhynchotheca* (Fig. 4), first diverged long before Andean orogeny provided the elevations and climatic conditions that they now inhabit. *Hypseocharis* and *Rhynchotheca* separated from their sister clades 62 and 29 Mya, respectively, well before the central Andes had reached a third (20 Mya) or a half (10–15 Mya) of their current height (Graham, 2009; Hoorn *et al*., 2010). *Rhynchotheca* at least represents another example of the radiation of central or northern Andean taxa from southern lineages, a pattern seen also in genera of Mutisieae and Liabeae of Asteraceae (Hershkovitz *et al*., 2006; Soejima *et al*., 2008), *Lepechinia* of Lamiaceae (Drew & Sytsma, 2013), *Heliotropium* of Heliotropiaceae (Luebert, Hilger & Weigend, 2011), *Fuchsia* of Onagraceae (Berry *et al*., 2004), and *Puya* of Bromeliaceae (Jabaily & Sytsma, 2010, 2013).

CONCLUSIONS

There is growing evidence that lack of priors, on or near the base of a phylogenetic tree, can dramatically influence the chronogram. This typically has been seen in pushing the root date to what appears to be unreasonably old. However, we demonstrate here the opposite effect, in which the more basal nodes are pulled to unreasonably young dates. In the case of the Geraniales, the combination of fossil constraints near the tip and the use of two gene regions appearing already saturated in nucleotide substitutions, generates dates across the phylogeny 20–50% younger. As argued by Magallón *et al*. (2013), the use of temporal constraints in relaxed-clock dating and the estimation of branch rates is a double-edged sword. If used and interpreted correctly, these constraints are invaluable for estimating divergence times and rates. If used and interpreted incorrectly, these constraints can generate errors in age and rate estimates. When only few constraints are used, they are positioned only near the tips, or the root or basal nodes lack constraints, these errors can easily propagate across the phylogeny, as demonstrated in Geraniales in the present study.

Temporal constraints are key to unlocking the historical diversification of lineages. As demonstrated by Yang & Rannala (2006), the inclusion of an almost infinite amount of sequence data cannot override the impact provided by fossil calibrations. We highlight four 'best-practice' approaches (Benton & Donoghue, 2007; Hugall *et al*., 2007; Ho & Phillips, 2009; Heled & Drummond, 2012; Parham *et al*., 2012; Yang & Rannala, 2012) based on this re-analysis of the Geraniales data set.

First, the use of multiple fossil calibrations is essential, as was carefully carried out by Palazzesi *et al*. (2012).

Second, having priors on nodes spread across the phylogeny may be crucial, as we demonstrate here with Geraniales and rosid outgroups – this being critical when using sequence data reaching nucleotide substitution saturation and exhibiting rate heterogeneity among lineages.

Third, reliance on 'secondary calibrations', especially for the root or other basal nodes, should not be necessarily viewed with scepticism. Providing a broad prior on the rosid stem, based on congruence across multiple, fossil-calibrated studies, was essential to obtain realistic dates of diversification for Geraniales and other rosid orders. Moreover, the additions of these priors suggested alternative placements (younger nodes) for several of the Geraniales fossils used. As noted by Magallón *et al*. (2013), relaxed clocks permit increased flexibility in implementing calibrations using a variety of temporal information in the form of fossils, geological events, or dates derived from independent studies. All these approaches are, of course, potentially fraught with danger. Fossil placement can be questionable, as we show is probably the case with *T. pelargonioides*. Even the temporal dates of presumably well-characterized geological events can be questioned, for example the closure of the Isthmus of Panama (Bacon *et al*., 2013). And, of course, secondary calibrations are only as good as the study that generated them. Indeed, our interest in re-analyzing the Geraniales data set was motivated by our desire to obtain meaningful nodal dates for our ongoing studies of temporal diversification across the order Myrtales, sister to the Geraniales.

Fourth, we urge, as common practice in studies of temporal diversification, that archiving (e.g. GenBank, TreeBASE, Dryad; see Drew *et al*., 2013) should not be restricted to just sequence data, alignments and phylogenetic trees (although journals differ in their requirements for even these (e.g. only GenBank information was available for this

Geraniales study). Owing to the often-limited cursory discussion of nodal calibrations and other parameters in the published methodological section, we recommend that the xml files of BEAST (or alternative programs) be made available or archived in open source depositories. This archiving would explicitly detail the calibrations of all prior distributions utilized in the analysis, is essential to assess the appropriateness of these parameters, and offers the possibility of reproducing the results.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers who provided helpful comments. We acknowledge the support of Susana Magallón who provided references and information about issues with calibration of phylogenetic trees and the assistance of Bret Larget concerning issues with sequence divergence in a maximum-likelihood context. We give credit to Sarah Friedrich for assistance with the figures. Finally, we thank members of the Systematics Section in the Botany Department, University of Wisconsin, for detailed discussion and comments of the manuscript.

REFERENCES

Álvarez I, Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29:** 417–434.

Anderson CL, Bremer K, Friis EM. 2005. Dating phylogenetically basal Eudicots using *rbcL* sequences and multiple fossil reference points. *American Journal of Botany* **92:** 1737–1748.

Angiosperm Phylogeny Group. 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161:** 105–121.

Antonelli A, Nylander JAA, Persson C, Sanmartín I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 9749–9754.

Bacon CD, Mora A, Wagner WL, Jaramillo CA. 2013. Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical Journal of the Linnean Society* **171:** 287–300.

Baldwin BG. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* **1:** 3–16.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82:** 247–277.

Barniske A-M, Borsch T, Müller K, Krug M, Worberg A, Neinhuis C, Quandt D. 2012. Phylogenetics of early branching eudicots: comparing phylogenetic signal across plastid introns, spacers, and genes. *Journal of Systematics and Evolution* **50:** 85–108.

Bell CD, Soltis DE, Soltis PS. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97:** 1296–1303.

Benton MJ, Donoghue PCJ. 2007. Paleontological evidence to date the tree of life. *Molecular Biology and Evolution* **24:** 26–53.

Berger B. 2012. Myrtales: molecules, mangroves and *Metrosideros*. Unpublished D. Phil. Thesis, University of Wisconsin, Madison.

Berry PE, Hahn WJ, Sytsma KJ, Hall JC, Mast A. 2004. Phylogenetic relationships and biogeography of *Fuchsia* (Onagraceae) based on noncoding nuclear and chloroplast DNA data. *American Journal of Botany* **91:** 601–614.

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* **16:** 558–576.

Borsch T, Quandt D. 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution* **282:** 169–199.

Brandley MC, Wang Y, Guo X, Nieto Montes De Oca A, Feriaortíz M, Hikida T, Ota H. 2011. Accommodating heterogeneous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic Biology* **60:** 3–15.

Buckler ES, Ippolito A, Holtsford TP. 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* **145:** 821–832.

Bytebier B, Antonelli A, Bellstedt DU, Linder HP. 2010. Estimating the age of fire in the Cape flora of South Africa from an orchid phylogeny. *Proceedings of The Royal Society B-Biological Sciences* **278:** 188–195.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michael HJ, Kress WJ, Karol KA, Clark WD, Hedrén M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW Jr, Barrett SCH, Dayanandan S, Albert VA. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* **80:** 528–580.

Chaves JA, Weir JT, Smith TB. 2011. Diversification in *Adeloymia* hummingbirds follows Andean uplift. *Molecular Ecology* **20:** 4564–4576.

Conti E, Eriksson T, Schönenberger J, Sytsma KJ, Baum DA. 2002. Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution* **56:** 1931–1942.

Conti E, Litt A, Sytsma KJ. 1996. Circumscription of Myrtales and their relationships to other rosids: evidence from *rbcL* sequence data. *American Journal of Botany* **83:** 221–233.

Conti E, Litt A, Wilson PG, Graham SA, Briggs BG, Johnson LAS, Sytsma KJ. 1997. Interfamilial relationships in Myrtales: molecular phylogeny and patterns of morphological evolution. *Systematic Botany* **22:** 629–647.

Conti E, Rutschmann F, Eriksson T, Sytsma KJ, Baum DA. 2004. Calibration of molecular clocks and the biogeographic history of Crypteroniaceae: a reply to Moyle. *Evolution* **58:** 1874–1876.

Daly M, Herendeen PS, Guralnick RP, Westneat MW, McDade L. 2012. Systematics Agenda 2020: the mission evolves. *Systematic Biology* **61:** 549–552.

Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ. 2005. Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. *American Naturalist* **165:** 36–65.

Donoghue PCJ, Benton MJ. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution* **22:** 424–431.

Dornburg A, Brandley MC, McGowen MR, Near TJ. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution* **29:** 721–736.

Drew BT, Gazis R. Cabezas P, Swithers KS, Deng J, Rodriguez R, Katz LA, Crandall KA, Hibbett DS, Soltis DE. 2013. Lost branches on the Tree of Life. *PLoS Biology* **11:** e1001636.

Drew BT, Sytsma KJ. 2013. The South American radiation of *Lepechinia* (Lamiaceae): phylogenetics, divergence times and evolution of dioecy. *Botanical Journal of the Linnean Society* **171:** 171–190.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4:** 699–710.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012a. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

Drummond CS, Eastwood RJ, Miotto STS, Hughes CE. 2012b. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Systematic Biology* **61:** 443–460.

Evans MEK, Smith SA, Flynn RS, Donoghue MJ. 2009. Climate, niche evolution, and diversification of the 'bird-cage' evening primroses (*Oenothera*, sections *Anogra* and *Kleinia*). *The American Naturalist* **173:** 225–240.

Feldberg K, Heinrichs J, Schmidt AR, Váňa J, Schneider H. 2013. Exploring the impact of fossil constraints on the divergence time estimates of derived liverworts. *Plant Systematics and Evolution* **299:** 585–601.

Fiz O, Vargas P, Alarcón M, Aedo C, García JL, Aldasoro JJ. 2008. Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. *Systematic Botany* **33:** 326–342.

Fiz O, Vargas P, Alarcón M, Aldasoro JJ. 2006. Phylogenetic relationships and evolution in *Erodium* (Geraniaceae) based on *trnL-trnF* sequences. *Systematic Botany* **31:** 739–763.

Fiz-Palacios O, Vargas P, Vila R, Papadopulos AST, Aldasoro JJ. 2010. The uneven phylogeny and biogeography of *Erodium* (Geraniaceae): radiations in the Mediterranean and recent recurrent intercontinental colonization. *Annals of Botany* **106:** 871–884.

Forest F. 2009. Calibrating the Tree of Life: fossils, molecules and evolutionary timescales. *Annals of Botany* **104:** 789–794.

Friis EM. 1983. Upper Cretaceous (Senonian) floral structures of juglandalean affinity containing *Normapolles* pollen. *Review of Palaeobotany and Palynology* **39:** 161–188.

Gandolfo MA, Nixon KC, Crepet WL. 2008. Selection of fossils for calibration of molecular dating models. *Annals of the Missouri Botanical Garden* **95:** 34–42.

Givnish T, Millam K, Evans T, Hall J, Chris Pires J, Berry P, Sytsma K. 2004. Ancient vicariance or recent long-distance dispersal? Inferences about phylogeny and South American-African disjunctions in Rapateaceae and Bromeliaceae based on *ndhF* sequence data. *International Journal of Plant Sciences* **165:** 35–54.

Givnish TJ, Barfuss MHJ, Ee BV, Riina R, Schulte K, Horres R, Gonsiska PA, Jabaily RS, Crayn DM, Smith JAC, Winter K, Brown GK, Evans TM, Holst BK, Luther H, Till W, Zizka G, Berry GE, Sytsma KJ. 2011. Phylogeny, adaptive radiation, and historical biogeography in Bromeliaceae: insights from an eight-locus plastid phylogeny. *American Journal of Botany* **98:** 872–895.

Givnish TJ, Barfuss MHJ, Van Ee B, Riina R, Schulte K, Horres R, Gonsiska PA, Jabaily RS, Crayn DM, Smith JAC, Winter K, Brown GK, Evans TM, Holst BK, Luther H, Till W, Zizka G, Berry PE, Sytsma KJ. 2014. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Molecular Phylogenetics and Evolution* **71:** 55–78.

Givnish TJ, Evans TM, Zjhra ML, Patterson TB, Berry PE, Sytsma KJ. 2000. Molecular evolution, adaptive radiation, and geographic diversification in the amphiatlantic family Rapateaceae: evidence from *ndhF* sequences and morphology. *Evolution* **54:** 1915–1937.

Givnish TJ, Millam KC, Mast AR, Paterson TB, Theim TJ, Hipp AL, Henss JM, Smith JF, Wood KR, Sytsma KJ. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society of London-B-Biological Sciences Proceedings* **276:** 407–416.

Graham A. 2009. The Andes: a geological overview from a biological perspective. *Annals of the Missouri Botanical Garden* **96:** 371–385.

Graham SA, Hall J, Sytsma K, Shi S. 2005. Phylogenetic analysis of the Lythraceae based on four gene regions and

morphology. *International Journal of Plant Sciences* **166:** 995–1017.

Graham SW, Reeves PA, Burns ACE, Olmstead RG. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* **161:** S83–S96.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2008. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences* **105:** 18424–18429.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* **28:** 583–600.

Harris AJ, Wen J, Xiang Q-YJ. 2013. Inferring the biogeographic origins of inter-continental disjunct endemics using a Bayes-DIVA approach. *Journal of Systematics and Evolution* **51:** 117–133.

Heath TA, Holder MT, Huelsenbeck JP. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. *Molecular Biology and Evolution* **29:** 939–955.

Hedman MM. 2010. Constraints on clade ages from fossil outgroups. *Paleobiology* **36:** 16–31.

Heled J, Drummond AJ. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology* **61:** 138–149.

Herngreen GFW. 1975. An Upper Senonian pollen assemblage of borehole 3-PIA-10-AL state of Alagoas, Brazil. *Pollen Spores* **17:** 93–140.

Hershkovitz MA, Arroyo MT, Bell C, Hinojosa LF. 2006. Phylogeny of *Chaetanthera* (Asteraceae: Mutisieae) reveals both ancient and recent origins of the high elevation lineages. *Molecular Phylogenetics and Evolution* **41:** 594–605.

Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology* **38:** 409–414.

Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58:** 367–380.

Hoorn C, Wesselingh FP, Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, Jaramillo C, Riff D, Negri FR, Hooghiemstra H, Lundberg J, Stadler T, Särkinen T, Antonelli A. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330:** 927–931.

Hugall AF, Foster R, Lee MSY. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology* **56:** 543–563.

Inoue J, Donoghue PCJ, Yang Z. 2009. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology* **59:** 74–89.

Jabaily RS, Sytsma KJ. 2010. Phylogenetics of *Puya* (Bromeliaceae): placement, major lineages, and evolution of Chilean species. *American Journal of Botany* **97:** 337–356.

Jabaily RS, Sytsma KJ. 2013. Historical biogeography and life history evolution of Andean *Puya* (Bromeliaceae). *Botanical Journal of the Linnean Society* **171:** 201–224.

Jones CS, Martinez-Cabrera HI, Nicotra AB, Mocko K, Marais EM, Schlichting CD. 2013. Phylogenetic influences on leaf trait integration in *Pelargonium* (Geraniaceae): convergence, divergence, and historical adaptation to a rapidly changing climate. *American Journal of Botany* **100:** 1306–1321.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30:** 3059–3066.

Ksepka DT, Benton MJ, Carrano MT, Gandolfo MA, Head JJ, Hermsen EJ, Joyce WG, Lamm KS, Patane JSL, Phillips MJ, Polly PD, Van Tuinen M, Ware JL, Warnock RCM, Parham JF. 2011. Synthesizing and databasing fossil calibrations: divergence dating and beyond. *Biology Letters* **7:** 801–803.

Levin RA, Wagner WL, Hoch PC, Hahn WJ, Rodriguez A, Baum DA, Katinas L, Zimmer EA, Sytsma KJ. 2004. Paraphyly in tribe Onagreae: insights into phylogenetic relationships of Onagraceae based on nuclear and chloroplast sequence data. *Systematic Botany* **29:** 147–164.

Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytsma KJ. 2003. Family-level relationships of Onagraceae based on chloroplast *rbcL* and *ndhF* data. *American Journal of Botany* **90:** 107–115.

Linder HP, Dlamini T, Henning J, Verboom GA. 2006. The evolutionary history of *Melianthus* (Melianthaceae). *American Journal of Botany* **93:** 1052–1064.

Luebert F, Hilger HH, Weigend M. 2011. Diversification in the Andes: age and origins of South American *Heliotropium* lineages (Heliotropiaceae, Boraginales). *Molecular Phylogenetics and Evolution* **61:** 90–102.

Macphail MK. 1999. Palynostratigraphy of the Murray Basin, inland Southeastern Australia. *Palynology* **23:** 197–240.

Magallón S. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. *International Journal of Plant Sciences* **165:** 7–21.

Magallón S. 2010. Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology* **59:** 384–399.

Magallón S, Castillo A. 2009. Angiosperm diversification through time. *American Journal of Botany* **96:** 349–365.

Magallón S, Hilu KW, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* **100:** 556–573.

Magallón S, Sanderson MJ. 2005. Angiosperm divergence times: the effect of genes, codon positions and time constraints. *Evolution* **59:** 1653–1670.

Marshall CR. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist* **171:** 726–742.

Martin HA. 1973. Upper Tertiary palynology in southern New South Wales. *Geological Survey of Australia Special Publication* **4:** 35–54.

McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL. 2010. Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution* **65:** 184–202.

Muller J. 1981. Fossil pollen records of extant angiosperms. *Botanical Review* **47:** 1–142.

Müller KF, Borsch T, Hilu KW. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. *Molecular Phylogenetics and Evolution* **41:** 99–117.

Palazzesi L, Gottschling M, Barreda V, Weigend M. 2012. First Miocene fossils of Vivianiaceae shed new light on phylogeny, divergence times, and historical biogeography of Geraniales. *Biological Journal of the Linnean Society* **107:** 67–85.

Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, Inoue JG, Irmis RB, Joyce WG, Ksepka DT, Patane JSL, Smith ND, Tarver JE, Van Tuinen M, Yang Z, Angielczyk KD, Greenwood JM, Hipsley CA, Jacobs L, Makovicky PJ, Muller J, Smith KT, Theodor JM, Warnock RCM, Benton MJ. 2012. Best practices for justifying fossil calibrations. *Systematic Biology* **61:** 346–359.

Phillips MJ. 2009. Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene* **441:** 132–140.

Pirie MD, Doyle JA. 2012. Dating clades with fossils and molecules: the case of Annonaceae. *Botanical Journal of the Linnean Society* **169:** 84–116.

Pittman III WC, Cande S, LaBrecque J, Pindell J. 1993. Fragmentation of Gondwana: the separation of Africa from South America. In: Goldblatt P, ed. *Biological relationships between Africa and South America*. New Haven, CT: Yale University Press, 15–34.

Poczai P, Hyvönen J. 2009. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Molecular Biology Reports* **37:** 1897–1912.

Qiu YL, Lin L, Wang B, Xue BY, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* **48:** 391–425.

Quandt D, Müller K, Stech M, Hilu KW, Frey W, Frahm J-P, Borsch T. 2004. Molecular evolution of the chloroplast *trnL-F* region in land plants. *Monographs in Systematic Botany from the Missouri Botanical Garden* **98:** 13–37.

Rambaut A, Drummond AJ. 2009. Tracer version 1.5. Computer program and documentation distributed by the authors, Available at: http://beast.bio.ed.ac.uk/Tracer

Renner S. 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* **165:** 23–33.

Renner SS. 2005. Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science* **10:** 550–558.

Renner SS, Clausing G, Meyer K. 2001. Historical biogeography of Melastomataceae: the roles of Tertiary migration and long-distance dispersal. *American Journal of Botany* **88:** 1290–1300.

Rutschmann F. 2006. Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. *Diversity and Distributions* **12:** 35–48.

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14:** 1218–1231.

Sanderson MJ. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? In: Soltis P, Soltis D, Doyle J, eds. *Plant molecular systematics*, 2nd edn. London: Chapman and Hall, 242–264.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19:** 101–109.

Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, Lee DM, Murphy DJ, Sniderman JMK, 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.

Schönenberger J, Pedersen KR, Friis EM. 2001. *Normapolles* flowers of fagalean affinity from the Late Cretaceous of Portugal. *Plant Systematics and Evolution* **226:** 205–230.

Sims HJ, Herendeen PS, Lupia R, Christopher RA, Crane PR. 1999. Fossil flowers with *Normapolles* pollen from the Late Cretaceous of southeastern North America. *Review of Palaeobotany and Palynology* **106:** 131–151.

Smith AB, Peterson KJ. 2002. Dating the time of origin of major clades: molecular clocks and the fossil record. *Annual Review of Earth and Planetary Sciences* **30:** 65–88.

Soejima A, Wen J, Zapata M, Dillon MO. 2008. Phylogeny and putative hybridization in the subtribe Paranepheliinae (Liabeae, Asteraceae), implications for classification, biogeography, and Andean orogeny. *Journal of Systematics and Evolution* **46:** 375–390.

Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z, Rushworth CA, Gitzendanner MA, Sytsma KJ, Qiu YL, Hilu KW, Davis CC, Sanderson MJ, Beaman RS, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98:** 704–730.

Swofford DL. 2003. *PAUP* phylogenetic analysis using parsimony (*and other methods), Version 4.0b10*. Sunderland, MA: Sinauer Associates.

Sytsma KJ, Litt A, Zjhra M, Pires JC, Nepokroeff M, Conti E, Walker J, Wilson P. 2004. Clades, clocks, and continents: historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the Southern Hemisphere. *International Journal of Plant Sciences* **165:** 85–105.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C,

Willerslev E. 2007. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35:** e14.

Taberlet P, Gielly L, Patou G, Bouvet J. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105–1109.

Takahashi M, Crane PR, Ando H. 1999. *Esgueiria futabensis* sp. nov., a new angiosperm flower from the Upper Cretaceous (Lower Coniacian) of northeastern Honshu, Japan. *Paleontological Research* **3:** 81–87.

Thornhill AH, Popple LW, Carter RJ, Ho SYW, Crisp MD. 2012. Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies? A comparative study using Myrtaceae. *Molecular Phylogenetics and Evolution* **63:** 15–27.

Töpel M, Antonelli A, Yesson C, Eriksen B. 2012. Past climate change and plant evolution in Western North America: a case study in Rosaceae. *PLoS ONE* **7:** e50358.

Walker JD, Geissman JW, Bowring SA, Babcock LE. 2013. *The Geological Society of America Geologic Time Scale: GSA Bulletin*. doi:10.1130/B30712.1.

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 3853–3858.

Warnock RCM, Yang Z, Donoghue PCJ. 2012. Exploring uncertainty in the calibration of the molecular clock. *Biology Letters* **8:** 156–159.

Wertheim JO, Fourment M, Kosakovsky Pond SL. 2012. Inconsistencies in estimating the age of HIV-1 subtype due to heterotachy. *Molecular Biology and Evolution* **29:** 451–456.

Wikström N, Savolainen V, Chase M. 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London-B-Biological Sciences* **268:** 2211–2220.

Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavaré S. 2011. Dating primate divergences through an integrated analysis of paleontological and molecular data. *Systematic Biology* **60:** 16–31.

Worberg A, Quandt D, Barniske A-M, Löhne C, Hilu KW, Borsch T. 2007. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Organisms Diversity & Evolution* **7:** 55–77.

Yang Z. 1998. On the best evolutionary rate for phylogenetic analysis. *Systematic Biology* **47:** 125–133.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Systematic Biology* **23:** 212–226.

Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics* **13:** 303–314.

Zheng Y, Peng R, Kuro-o M, Zeng X. 2011. Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (Order Caudata). *Molecular Biology and Evolution* **28:** 2521–2535.

Zhu XY, Chase MW, Qiu YL, Kong HZ, Dilcher DL, Li JH, Chen ZD. 2007. Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evolutionary Biology* **71:** 217.

Zwickl D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Unpublished D. Phil. Thesis, University of Texas, Austin.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Enlarged BEAST chronogram of Figure 2C: baseline chronogram as in Palazzesi *et al*. (2012), but with three rosid stem priors added and enforcing the APG III topology among orders. Circled numbers refer to seven fossil-based constraints (1–7) and a rosid stem prior (8) used: 1 = *Viviania* crown (10 Mya), 2 = *Balbisia* crown (15.8 Mya), 3 = *Geranium* crown (7.2 Mya), 4 = *Erodium* crown (7.2 Mya), 5 = *Pelargonium* stem (28.4 Mya), 6 = Fagales stem (96 Mya), 7 = Myrtales stem (88.2 Mya), 8 = rosid stem (125-101 Mya). Grey bars represent the associated credibility interval (95% highest posterior density). Abrreviations: Q, Quaternary; Pl, Pliocene; Oligo, Oligocene; Paleo, Paleocene.

Figure S2. BEAST baseline chronogram following Palazzesi *et al*. (2012), but shifting the *Pelargonium* fossil constraint from stem to crown (see Table 1E). Circled numbers refer to five fossil-based constraints (1–7) and a rosid stem prior (8) used: 1 = *Viviania* crown (10 Mya), 2 = *Balbisia* crown (15.8 Mya), 3 = *Geranium* crown (7.2 Mya), 4 = *Erodium* crown (7.2 Mya), 5 = *Pelargonium* crown (28.4 Mya). Grey bars represent the associated credibility interval (95% highest posterior density). Abrreviations: Q, Quaternary; Pl, Pliocene; Oligo, Oligocene; Paleo, Paleocene.

Appendix S1. Nexus file of the re-aligned ITS and *trnL–F* data set of Palazzesi *et al*. (2012). Two gene regions are interleaved and deleted base pairs are indicated by command.

Appendix S2. Representative xml file generated by BEAUti for BEAST analysis. The file was used in the analysis with all five Geraniales fossil priors of Palazzesi *et al*. (2012), two additional fossil priors for the stems of Myrtales and Fagales, and a root prior for the rosid-stem node. This analysis generated trees seen in Figure 2C and Figure S1.