HOW OLD ARE SUNFLOWERS? A MOLECULAR CLOCK ANALYSIS OF KEY DIVERGENCES IN THE ORIGIN AND DIVERSIFICATION OF HELIANTHUS (ASTERACEAE)

Chase M. Mason1,*

*Department of Biology, University of Central Florida, Orlando, Florida 32816, USA

Editor: Susana Magallón

Premise of research. For many questions in evolutionary ecology, it is quite valuable to have an estimate of the span of geologic time over which a group of species has arisen and diversified. This study uses a molecular clock approach to generate a first estimate of key divergences in the evolutionary history of the genus Helianthus, which to date has lacked such an estimate.

Methodology. Divergence time analysis was performed with well-resolved phylogenies of Asteraceae and Helianthus in a stepping-stone manner using the ReTime maximum likelihood method with a variety of available macrofossil, fossil pollen, and molecular clock calibrations derived from the literature.

Pivotal results. Mean estimates from nine individual calibration scenarios time the divergence of Helianthus from the sister genus Phoebanthus at between 2.47 and 5.41 Ma. Composite calibration incorporating all literature estimates time this divergence at 3.63 Ma, with a 95% confidence interval spanning 0–8.26 Ma. Subsequent diversification within Helianthus is therefore confined to the last few million years.

Conclusions. These findings place the origin and diversification of Helianthus into the temporal context of the Pliocene expansion of open habitats across North America and the subsequent vegetation fluctuations of the Pleistocene glacial-interglacial cycles.

Keywords: dating, divergence time, Helianthus, molecular clock, sunflower.

Online enhancement: appendix figures.

Introduction

The genus Helianthus is an emerging model system in evolutionary ecology. This group of species includes broad diversification in morphology, physiology, and life history across strong environmental gradients (Heiser et al. 1969; Mason and Donovan 2015), allowing for the study of many aspects of trait evolution under adaptation to diverse habitats. The genus also contains a variety of hybridization and polyploidization events that provide excellent arenas for the study of such processes (Rieseberg 2006; Kane et al. 2009; Bock et al. 2014). The amount of genetic resources for this group is steadily increasing, with a larger number of sets of genetic markers in existence (Kane et al. 2013), several well-resolved phylogenies addressing both the diploid backbone of the genus and various polyploid clades (Bock et al. 2014; Stephens et al. 2015), an association mapping panel and associated linkage maps (Mandel et al. 2013), and the recent publication of a full reference genome for cultivated Helianthus annuus (Badouin et al. 2017).

For many questions related to understanding the evolutionary history of Helianthus, it would be very useful to have an estimate of the approximate age of the genus and the geologic periods during which this group has diversified. However, this is an herbaceous genus (i.e., poorly fossilizing relative to woody plants) and lacks diagnostically unique pollen morphology (i.e., not useful for dating from sediment deposits). As such, no fossils have been discovered for Helianthus or any close relatives other than archaeological evidence of sunflower domestication in North America within the last five millennia (e.g., Crites 1993). This makes placing the diversification of Helianthus into a geologic context difficult and limits the questions we can address with phylogenetic comparative approaches.

This study uses a molecular clock approach to generate a first set of estimates for key dates in the evolutionary history of Helianthus. To accomplish this, I exploit the existence of overlap between two well-resolved phylogenies, one of diploid Helianthus and one of a cross section of species spanning major Asteraceae tribes. By using a variety of divergence age estimates from the literature for major clades of Asteraceae and applying these to calibrate an Asteraceae phylogeny containing multiple species of Helianthus, estimates of divergence applicable to calibration of the diploid Helianthus phylogeny are generated. This stepping-stone approach allows for the estimation of divergence ages in a genus otherwise without useful calibrations.
Material and Methods

Phylogenies

Two recent phylogenies were used for divergence time analyses: a phylogeny of a cross section of representative Asteraceae and Calyceraceae spanning disparate tribes (Mandle et al. 2014) and a phylogeny of the diploid backbone of Helianthus (Stephens et al. 2015). The Asteraceae phylogeny (fig. 1a; Mandle et al. 2014) was originally created using data on 763 nuclear genes in 16 taxa using a target enrichment approach on conserved orthologous sequences, with topology and branch lengths estimated using maximum likelihood on a concatenated data set with GARLI (ver. 2.0; Zwickl 2006; Mandle et al. 2014). The tree file and concatenated alignment file were obtained from the Dryad Digital Repository (http://datadryad.org/resource/doi:10.5061/dryad.gr93t) for use in subsequent analyses.

The Helianthus phylogeny (fig. 1b; Stephens et al. 2015) was originally created using data on 170 nuclear genes on 38 taxa generated through target enrichment, with topology estimated through two methods: first, by maximum likelihood on a concatenated data set in RAxML (ver. 0.42.6.2012; Stamatakis 2006), and second, by maximum pseudolikelihood estimation of the species tree (MP-EST; ver. 1.4; Liu et al. 2010), a multispecies coalescent approach that estimates a consensus species tree topology from multiple gene trees and is robust to the effects of gene tree discordance from incomplete lineage sorting (Liu et al. 2010). Because coalescent methods do not produce normal branch lengths (e.g., coalescent units with fixed terminal tip lengths), branch lengths in substitution units were subsequently estimated on the fixed MP-EST topology using RAxML (Stephens et al. 2015). Tree files and aligned FASTA files were obtained from the Dryad Digital Repository (http://datadryad.org/resource/doi:10.5061/dryad.4n27n) for use in subsequent analyses. Because both the concatenation and MP-EST trees contained multiple individuals per taxa, both the tree files and FASTA data sets were pruned to contain only the single individual per species or subspecies with the longest concatenated sequence length (i.e., most complete sequence).

Fossil and Molecular Clock Calibrations

To obtain divergence time estimates for nodes in the Asteraceae phylogeny, a literature search was performed for examples of Asteraceae macrofossils or fossil pollen as well as previous molecular clock analyses that incorporate members of the Asteraceae. This yielded five relevant studies that provided age estimates applicable to four nodes in the Asteraceae phylogeny (fig. 1a; table 1). These nodes include the split between the Barnadesiaeae and the rest of the Asteraceae (fig. 1a, node A), the split between the Mutisioideae (here represented by the Mutisieae) and the rest of the Asteraceae excluding Barnadesiaeae (fig. 1a, node B), the split between the Cichorieae and the Aster oideae (fig. 1a, node C), and the most recent common ancestor of the tribes within the Aster oideae (fig. 1a, node D), here represented by the tribes Gnaphalieae, Senecioneae, Eupatoireae, and Heliantheae.

Two of the five relevant studies provided estimates of divergence directly from fossils, one from both a macrofossil and accompanying fossil pollen dated to 47.39 Ma (Barreda et al. 2012), thought to represent the minimum age of divergence of the Mutisioideae from the rest of the Asteraceae (represented by node B in fig. 1a), and another from fossil pollen dated to 16 Ma (Katinas et al. 2007), thought to represent the minimum age of emergence of tribes within the Asteroideae (represented by node D in fig. 1a). The remaining three studies used molecular clock approaches, with various fossil calibrations not directly transferable to the Asteraceae phylogeny used here (Kim et al. 2005; Bergh and Linder 2009; Barreda et al. 2015). One such study (Barreda et al. 2015) used both the fossils of Barreda et al. (2012) and newly described fossil pollen to date the divergence of Barnadesiaeae from the rest of the Asteraceae (represented by node A in fig. 1a) to between 53 and 76 Ma on the basis of five different placement scenarios, all with similar results (one additional scenario places this date much older, but the accuracy of the fossil placement in that scenario was contested, so this scenario was not included in the analyses presented here; Barreda et al. 2016; Panero 2016; Panero and Crozier 2016). Another of these studies (Bergh and Linder 2009) used Asteraceae fossil pollen linked to the divergence of Heliantheae from Tageteae along with geological calibrations based on island emergence within the genus Stoebe to derive an estimate of the divergence among the tribes of the Asteroideae (represented by node D in fig. 1a). One additional study (Kim et al. 2005) avoided using any internal Asteraceae fossils, given the relatively sparse fossil record and controversy surrounding the identity and placement of fossils at the time (something which remains today). They used a noncontroversial outgroup Coronaria to calibrate a molecular clock analysis of an Asteraceae chloroplast gene inversion data set, deriving estimates for multiple divergences, including between Barnadesiaeae and the rest of Asteraceae, between Cichorieae and the Asteroideae, and between Heliantheae and Senecioneae (represented by nodes A, C, and D, respectively, in fig. 1a). While these latter three studies represent secondary calibrations (the primary fossil evidence they employ cannot be directly transferred to the Asteraceae tree employed here), the additional information provided by them is of significant value. Both Bergh and Linder (2009) and Barreda et al. (2015) derive estimates from potentially valuable Asteraceae fossil evidence that we would otherwise not be able to include, and Kim et al. (2005) represents a set of estimates that are not subject to the identity and placement controversies that plague many internal Asteraceae fossils.

Additionally, three other fossil pollen records within Asteraceae were identified but not included in analyses because of large phylogenetic distances between the divergence represented by the fossil and the closest divergence present in the Asteraceae phylogeny (fig. 1a). These all involved divergences within Cichorieae, including a divergence between the basal Scorzoneraseae and the rest of the Cichorieae (Cichorium intybus–type pollen; Hochuli 1978; Tretesberger et al. 2013), a divergence within the Scorzoneraeae (Scorzonera hispanica–type pollen; Blackmore et al. 1986; Tretesberger et al. 2013) and a divergence among genera within the Hyoseridinae (Sonchus oleraceus–type pollen; Blackmore et al. 1986; Tretesberger et al. 2013). None of the three are sound choices for calibration of nodes within the Asteraceae phylogeny used here (fig. 1a) because this phylogeny contains only two Cichorieae species. The two possible nodes these fossils could be applied to (node C or the Lactuca–Taraxacum divergence) are either much older or
Fig. 1  Two phylogenetic trees used for RelTime analyses: a recent maximum likelihood tree of 16 select Asteraceae and Calyceraceae created with 763 nuclear genes using targeted enrichment on conserved orthologous sequences (Mandel et al. 2014) and the most recent phylogeny of diploid Helianthus created with 170 nuclear genes obtained through targeted enrichment (Stephens et al. 2015). Letters reflect the location of nodes with calibrations from fossils and previous molecular clock analyses (see table 1), and numbers reflect nodes shared between the two phylogenies and used to calibrate the Helianthus tree for RelTime analyses (see table 2).
Calibration Node Age range (Ma) Type Source
A1 A 36–42 Molecular clock Kim et al. 2005
A2 A 36–42 Molecular clock Barreda et al. 2015
B (broad) B 47.39–76 Macrofossil + pollen Barreda et al. 2012
C C 32–38 Molecular clock Kim et al. 2005
D1 (broad) D 16–76 Fossil pollen Katinas et al. 2007
D1 (narrow) D 16–26 Fossil pollen Katinas et al. 2007
D2 D 26–39 Molecular clock Kim et al. 2005
D3 D 29.6–56.6 Molecular clock Bergh and Linder 2009

Notes: Node letters refer to four divergences in the Asteraceae tree (Mandel et al. 2014) depicted in figure 1: A, the most recent common ancestor (MRCA) of Fulcaldea stuessyi and Helianthus annuus, representing the divergence between the Barnadesieae and the rest of the Asteraceae; B, the MRCA of Gerbera hybrida and H. annuus, representing the divergence between the Mutisioideae (represented by the Mutisioideae) and the rest of the Asteraceae excluding Barnadesieae; C, the MRCA of Lactuca sativa and H. annuus, representing the divergence between the Cichorieae and the Asteroideae; and D, the MRCA of Senecio vulgaris and H. annuus, representing the common ancestor of the Asteroideae. For calibration scenarios derived from previous molecular clock analyses (A1, A2, C, D2, D3), age ranges reported in the original studies were used. For calibrations derived from fossils (B, D), two age ranges were used reflecting broad and narrow estimates: one with minimum age set by the fossil and a maximum age equivalent to the oldest age of divergence between Barnadesieae and the rest of Asteraceae among our scenarios (76 Ma), and the other with minimum age set by the fossil and maximum age set to 10 Ma earlier.

much younger than the three divergences represented by each fossil, so calibration with these fossils was not attempted. Careful calibration with these distant fossils could perhaps be attempted in a Bayesian analysis with appropriate priors, but given the size of the data sets used here, a maximum likelihood approach (RelTime; Tamura et al. 2012) was favored in order to keep computation time reasonable, especially given the sensitivity analysis performed. This approach has been shown to give highly similar results to Bayesian approaches with large data sets, with orders of magnitude faster computation time (Mello et al. 2017). The use of this maximum likelihood approach precluded the inclusion of the Cichorieae fossil records, given their phylogenetic distance from available nodes, because constraining those nodes with minimum ages far from the divergences they represent would likely introduce substantial bias and uncertainty into analyses when prior probability distributions are not modeled. The sensitivity analysis performed here itself demonstrates how changing minimum or maximum ages for a given node can substantially alter divergence time estimates. Given the number of other calibrations available without phylogenetic distance issues, the Cichorieae fossils were not included.

For the four focal nodes in the Asteraceae phylogeny, there were a total of seven divergence age estimates (one to three per node; table 1), with sometimes conflicting estimates from different sources. Because there is much uncertainty inherent in assigning ages to nodes, a conservative approach was employed. First, when estimates were reported as ranges in the literature (i.e., 95% confidence intervals [CIs] from molecular clock analyses), these ranges were used in subsequent analyses as minimum and maximum ages. Second, when estimates were reported as point estimates from fossil dating analyses, two age ranges were used, reflecting broad and narrow estimates: the broad estimate, with a minimum age set by the fossil and a maximum age equivalent to the oldest age of divergence between Barnadesieae and the rest of Asteraceae among our scenarios (76 Ma), and the narrow estimate, with a minimum age set by the fossil and maximum age set to 10 Ma earlier. This approach allows for (1) a maximally conservative estimate allowing for potentially long spans between the emergence of a group and the fossilization of the member of that group that was found, accounting for poor fossilization and/or lengthy periods where a group is not widespread or numerous enough to appear in the fossil record; and (2) a less conservative estimate for comparison, one more typical of the 95% confidence ranges obtained from molecular clock analyses in Asteraceae. Additionally, given the controversial nature of the identity and placement of some Asteraceae fossils (Panero et al. 2014; Panero and Crozier 2016), a sensitivity analysis was performed by conducting analyses with each of the nine different calibration scenarios individually and comparing divergence estimates alongside a composite analysis combining calibrations. In this way, the suite of results obtained here is not strongly dependent on the validity of any one calibration.

**Divergence Time Analysis**

Because the goal of this study was to estimate the timing of the origin of Helianthus and clades within the genus, and no fossil calibrations exist for Helianthus, a stepping-stone approach was used. Primary analysis was conducted on the Asteraceae phylogeny with the above-described literature-derived calibrations, and dates derived for nodes within Helianthus were then transferred to the Helianthus phylogeny to calibrate that secondary analysis. Note that the age of Helianthus divergence from its sister genus can be inferred directly from the primary analysis with the Asteraceae phylogeny, and the secondary analysis is only needed to infer the divergence of particular clades within Helianthus.

All divergence time analyses were performed using RelTime (Tamura et al. 2012) to estimate local clock rates and divergence times with the maximum likelihood analytical method in MEGA7 (Kumar et al. 2016). For each analysis, a general time reversible (GTR) nucleotide substitution model was used, with a discrete Gamma distribution (+G) with eight categories used to model evolutionary rate differences among sites. The
rate variation model allowed some sites to be invariant (+1), with 32.1% of sites invariant for the Asteraceae data set (Mandel et al. 2014) and 41.2% of sites invariant in the Helianthus data set (Stephens et al. 2015). All sites with less than 50% coverage (e.g., missing data, alignment gaps) were eliminated, for a total of 74,257 sites used for analysis in the Asteraceae data set and 74,960 sites used for analysis in the Helianthus data set.

First, divergence time analysis was performed on the Asteraceae phylogeny independently for each of the nine individual calibration scenarios derived from the literature (table 1) as well as in a composite calibration using the minimum and maximum ages for each node across literature sources. Ninety-five percent CIs were estimated for each node in the Asteraceae phylogeny in each analysis. All calibrated time trees generated from these analyses are presented in figures A1–A5 (figs. A1–A11 are available online).

Second, divergence time analysis was performed on both the concatenation and MP-EST phylogenies of Helianthus using age estimates for the five nodes overlapping between the Asteraceae and Helianthus trees (fig. 1; table 2). Three sets of age ranges were used for RelTime analysis, reflecting (1) the range of minimum and maximum age estimates for each node obtained on the Asteraceae tree under the nine scenarios used, (2) the pooled 95% CI of the mean of the age estimates obtained for each node, and (3) the 95% CI around node ages obtained under the composite calibration (table 2). All calibrated time trees generated from these analyses are presented in figures A6–A11.

**Results**

**Asteraceae Analysis**

Divergence time analysis of the Asteraceae tree yielded estimates for the mean divergence between the Barnadesieae and the rest of the Asteraceae (node A) ranging from as early as 84.29 Ma (the D1 broad calibration; fig. A4) to as late as 38.48 Ma (the D1 narrow calibration; fig. A4). The composite calibration, however, places the mean divergence between these two groups at 56.52 Ma, with a 95% CI spanning 47.39–76.00 Ma (fig. A5).

For the divergence between Phoebanthus and Helianthus (our focal divergence), the nine individual calibration scenarios yielded mean estimates ranging from 2.47 to 5.41 Ma (again from the narrow and broad D1 calibrations; table 2; figs. 2, A4). Estimates derived from the other calibration scenarios cluster within this range, with a mean of 3.92 Ma and a pooled 95% CI spanning 3.73–12.83 Ma (table 2; fig. 2). The composite calibration places the mean divergence estimate very closely nearby at 3.63 Ma, with a 95% CI spanning from 0 to 8.26 Ma (table 2; figs. 2, A5). For all nodes within Helianthus, mean estimates derived from the nine individual calibration scenarios all fall within 4.5 Ma before present, while estimates from the composite calibration are all span between 1.54 and 3.01 Ma (table 2). For divergences within Helianthus, CIs for individual calibration scenarios often include the present day (0 Ma), but a substantial number do not, and the pooled CI across the nine individual calibration scenarios includes only 0 Ma for the divergence between the sister species Helianthus annuus and Helianthus argophyllus (table 2).

**Helianthus Analysis**

Divergence time analysis on the two Helianthus trees yielded very similar results for a given calibration, despite slight differences in topology (figs. A6–A11). The pooled CI calibration generated the oldest estimates, with the basal Helianthus porteri diverging from other Helianthus 5.19 Ma and only one divergence less than 1 Ma before present (between subspecies of Helianthus debilis; fig. A9). The range of estimates calibration generated the youngest estimates, with H. porteri diverging from

<table>
<thead>
<tr>
<th>Table 2</th>
<th>RelTime Estimates of Divergences for Helianthus Species in the Asteraceae Tree (Ma before Present)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2.50 (.82–4.43)</td>
</tr>
<tr>
<td>A2</td>
<td>4.14 (.36–9.23)</td>
</tr>
<tr>
<td>B (broad)</td>
<td>4.56 (.09–11.08)</td>
</tr>
<tr>
<td>B (narrow)</td>
<td>3.87 (1.39–6.80)</td>
</tr>
<tr>
<td>C</td>
<td>3.45 (1.90–5.24)</td>
</tr>
<tr>
<td>D1 (broad)</td>
<td>5.41 (.00–28.44)</td>
</tr>
<tr>
<td>D1 (narrow)</td>
<td>2.47 (.34–5.55)</td>
</tr>
<tr>
<td>D2</td>
<td>3.82 (93–7.77)</td>
</tr>
<tr>
<td>D3</td>
<td>5.07 (.00–13.72)</td>
</tr>
<tr>
<td>Range of estimates</td>
<td>2.47–5.41</td>
</tr>
<tr>
<td>Mean of estimates + pooled CI</td>
<td>3.92 (.37–12.83)</td>
</tr>
<tr>
<td>Composite calibration</td>
<td>3.63 (.00–8.25)</td>
</tr>
</tbody>
</table>

Note. RelTime mean estimates with 95% confidence intervals (CIs) for nodes in the Asteraceae tree (Mandel et al. 2014) are presented (Ma before present). Letters represent the nine individual calibration scenarios described in table 1. Three additional derived summaries were created and used for subsequent analysis with the Helianthus tree (Stephens et al. 2015): a simple range of the estimates derived from the nine individual calibration scenarios, the mean of the nine individual calibration scenario estimates with pooled 95% CI, and a composite calibration made using the minimum and maximum ages used for each node across sources (table 1; A = 36–76 Ma; B = 47.39–76 Ma; C = 32–38 Ma; D = 16–56.6 Ma).
Fig. 2  RelTime estimates of divergence between *Helianthus* and the sister genus *Phoebanthus* using nine different individual calibration scenarios (table 1) on the Asteraceae tree (fig. 1). Diamonds indicate the mean date of divergence, while gray bars reflect the 95% confidence interval (CI) around each estimate (see table 2). Dark gray bars reflect derived summaries, including the range of estimates, the mean of estimates with a pooled 95% CI, and a composite calibration made using the minimum and maximum ages used for each node across sources (see table 2).
other *Helianthus* only 2.82 Ma, although still with only three divergences less than 1 Ma before present (between subspecies of *H. debilis* and *Helianthus niveus* and sister species *Helianthus angustifolius* and *Helianthus florianus*; fig. A11). The composite calibration generated intermediate estimates, although all three calibrations estimate the major divergences among the three major clades of *Helianthus* (fig. 1) at between means of approximately 2.33 and 4.55 Ma (figs. A9, A11). However, CIs for both the composite calibration and the pooled CI calibration were quite large (on the order of several million years for deeper nodes), making precise inferences of divergence times unreliable (figs. A6, A8). The range of estimates calibration generated much smaller CIs, most under 1.5 Ma (fig. A10), by virtue of the smaller ranges used for calibration (table 2).

**Discussion**

First and foremost, it is important to note the uncertainty inherent in the methods used in these analyses and molecular clock dating methods generally (Donoghue and Benton 2007; Pulvériro and Nichols 2007; Tamura et al. 2012; Duchêne et al. 2014; Mello et al. 2017). The approach used here attempts to be conservative and preserve uncertainty in the dating estimates derived, but it is important not to interpret these results as more than a first approximation of the timeline of the origin and radiation of *Helianthus*.

Dates obtained from the Asteraceae analysis appear consistent with geologic evidence relating to early Asteraceae history. Inference from the breakdown of Gondwana and the uplift of the Southern Andes suggests that Asteraceae diverged from the sister genus Calyceraceae sometime between 50 and 90 Ma before present (Funk et al. 2005), so the estimates obtained here for the subsequent divergence between the basal Barnadesieae clade and the rest of the Asteraceae (e.g., composite calibration with a mean of 56.52 Ma and a 95% CI of 47.39–76.00 Ma) are consistent with that evidence. This bodes well for the validity of the other estimates derived from the Asteraceae analysis. Additionally, another recent molecular clock analysis focusing on whole-genome duplications across the family places the divergence between Barnadesieae and the rest of the Asteraceae at between 52.2 and 72.1 Ma (Huang et al. 2016), consistent with the results found here. That study also found generally consistent estimates at other nodes shared with the analyses performed here, including placing the divergence between the Cardueae and Asteroideae at between 42.3 and 54.8 Ma (as compared with between 32.0 and 56.5 Ma for the 95% CI of the results of the composite calibration here). Other node estimates differ somewhat, for instance, the divergence among Asteroideae tribes is dated to between 36.1 and 46.6 Ma (Huang et al. 2016), while here our composite calibration places that divergence at between 16 and 38 Ma on the basis of the constraints of calibrations at C and D nodes (though estimates for two-thirds of individual calibrations have 95% CIs overlapping the range of estimates from Huang et al. 2016). Some of the differences between the results found here and those of Huang et al. (2016) may arise from the far larger number of taxa included in that study as well as the large number of non-Asteraceae fossil calibrations (eight) that they were able to include, given the larger taxonomic scope of their data set. Regardless, taken as a whole, the general concordance of results for these deeper nodes in Asteraceae suggest that the range of estimates found here for *Helianthus* are likely to be reliable.

Working from CIs, the majority of calibration scenarios place the divergence between *Helianthus* and the sister genus *Phebanthus* at sometime between the mid-Miocene and the Pleistocene (fig. 2). Mean estimates across scenarios are spread throughout the Pliocene, with the range of estimates extending just a few thousand years into both the Miocene and the Pleistocene (fig. 2). Both the mean of estimates and the composite calibration place this divergence in the Zanclean age of the Pliocene (fig. 2). Simulation work on divergence time analysis suggests that deeper node calibrations tend to give more accurate estimates and that using multiple calibrations give more accurate estimates than single calibrations (Duchêne et al. 2014), so estimates from the node A scenarios and the composite calibration should perhaps be given more subjective weight than some of the other individual calibration scenarios. Regardless, most of the various calibrations agree rather well with one another in placing the *Phebanthus-Helianthus* divergence.

In recent years, paleoclimate and vegetation reconstructions have become much more well developed and allow for useful insights in comparison to the divergence estimates obtained here. Interestingly, historical temperature reconstructions during this period show that global average surface temperature declined significantly between the late Miocene and mid-Pliocene (fig. 3; Hansen et al. 2013). This cooling trend resulted in a concomitant drying trend across North America, and paleobiome vegetation reconstructions from the Torotoro (fig. 3; Forrest et al. 2015) and Piacenzian (fig. 3; Salzmann et al. 2008) show a major reduction of closed forest habitat between the late Miocene and mid-Pliocene and an increase in open shrubland, grassland, and desert habitats across the continent. This pattern largely continued until the present day, with further cooling and drying through the Pleistocene (Hansen et al. 2013). Modern members of the genus *Helianthus* are largely heliophytes, with most species occupying high-light open-vegetation environments, such as deserts, grasslands, herbaceous wetlands, rock outcrops, and pine savannas (Heiser et al. 1969; Mason and Donovan 2015). An increase in the availability of open habitats occurring at the same time as the origin and subsequent diversification of *Helianthus* fits well with what we know of the ecology of the genus today, and assuming reasonable accuracy in divergence estimates, the climatic shifts during the late Miocene and Pliocene may have generated ecological opportunities for *Helianthus* to diversify across the continent.

Within the genus, CIs from the analysis of the *Helianthus* tree are large and highly overlapping for all but the range of estimates calibration (figs. A6–A11). With such large CIs, mean divergence estimates are largely uninterpretable, other than to observe that species diversification likely took place from the late Pliocene through the Pleistocene until the present day. Diversification through the lengthy glacial-interglacial cycles of the Pleistocene is interesting, because these climate cycles are known to have generated vegetation shifts that would have repeatedly mimicked the initial opening up of closed habitats between the late Miocene and mid-Pliocene (Axelrod 1985; Williams et al. 2004; Anderson 2006; fig. 3). This phenomenon may perhaps have also contributed to the widespread reticulate evolution observed in *Helianthus* (Rieseberg 1991; Timme et al. 2007) because repeated vegetation shifts could have conveniently brought
Fig. 3  Paleobiome and historical temperature reconstructions relative to *Helianthus-Phoebanthus* divergence. **Top**, paleobiome vegetation reconstructions of North America during the Late Miocene (adapted from Forrest et al. 2015) and the Middle Pliocene (adapted from Salzmann et al. 2008), two well-modeled periods that happen to overlap with the confidence intervals (CIs) of many of the *Helianthus-Phoebanthus* divergence estimates (fig. 2). Note in particular the increasing abundance of open habitats between the Tortonian and Piacenzian. **Bottom**, reconstruction of global average surface temperature (adapted from Hansen et al. 2013) overlaid with both the Tortonian (green) and Piacenzian (orange) and the three derived summary estimates for *Helianthus-Phoebanthus* divergence. Thick solid line reflects a 500,000-yr moving average fit over the more fine-scale temperature estimates contained by the gray area (adapted from Hansen et al. 2013).
species in and out of contact with one another again and again. This so-called Pleistocene pump hypothesis has been invoked in relation to the diversification of other groups whose habitats would have experienced repeated subdivision and reuniﬁcation with glacial-interglacial cycles, including Amazonian rain forest ﬂora and fauna (Haffer 1987, 2008), alpine ground beetles and butterflies (Schoville and Roderick 2009; Schoville et al. 2012), island rodents and beetles (Papadopoulou and Knowles 2015; Wainer and Robert 2016), freshwater ﬁshes (April et al. 2013), montane grassland plants (Fernandes 2016), and members of the Cape Flora of South Africa (Midgley and Roberts 2001).

One other source of dating information is relevant to our infrageneric estimates: there have been a number of dating analyses made within the three extensively studied stable homoploid hybrid species Helianthus anomalus, Helianthus deserticola, and Helianthus paradoxus. These three species all arose through independent hybridizations of the same two widespread parent species, Helianthus annuus and Helianthus petiolaris (Rieseberg 2006). Using coalescent approaches with microsatellite marker data, H. anomalus is estimated to be between approximately 144,000 and 116,000 yr old (Schwarzbach and Rieseberg 2002), H. deserticola is estimated to be between approximately 170,000 and 63,000 yr old (Gross et al. 2003), and H. paradoxus is estimated to be between approximately 208,000 and 75,000 yr old (Welch and Rieseberg 2002). These dates provide an independent source of information on the divergence between the two parent species as well as divergences between parent species and their respective sister taxa, both of which must have occurred sufﬁciently before the formation of the three hybrids. On the basis of results from the Asteraceae tree, the range of estimates for the divergence between H. annuus and Helianthus niveus (which is sister to H. petiolaris) spans 2.82–1.29 Ma, well before any of the hybrid origin estimates. Additionally, the range of estimates for the divergence between H. annuus and its sister taxa Helianthus argophyllus spans 2.30–1.05 Ma, also compatible with the hybrid origin estimates and indicating that the divergence time estimates obtained here within Helianthus are reasonable.

Overall, results of divergence time analysis indicate that the genus Helianthus and the major clades it contains are likely no more than a few million years old. This suggests that the diversiﬁcation of these species may be best viewed in light of the Pliocene expansion of open habitats across North America and continued ﬂuctuations during the glacial-interglacial cycles of the Pleistocene. Placing the origin of Helianthus in the context of geologic time will hopefully inform further study of the evolutionary history of this group and additionally allow for the formulation of new hypotheses relating to drivers of Helianthus diversiﬁcation.

Acknowledgments

I wish to thank Eric Goolsby, Vicki Funk, Susana Magallón, and an anonymous reviewer for helpful comments on the manuscript.

Literature Cited


Fernandes GW 2016 Ecology and conservation of mountain top grasslands in Brazil. Springer, Switzerland.


