Illusion of flight? Absence, evidence and the age of winged insects

SANDRA R. SCHACHAT1,*, PAUL Z. GOLDSTEIN2, ROB DESALLE3, DEAN M. BOBO3,4, C. KEVIN BOYCE1, JONATHAN L. PAYNE1 and CONRAD C. LABANDEIRA5,6,7

1Department of Geological Sciences, Stanford University, Stanford, CA, USA
2Systematic Entomology Laboratory, USDA, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
3American Museum of Natural History, Sackler Institute for Comparative Genomics, New York, NY, USA
4Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY, USA
5Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
6Department of Entomology and Behavior, Ecology, Evolution, and Systematics Program, University of Maryland, College Park, MD, USA
7Capital Normal University, School of Life Sciences, Beijing, China

Received 31 March 2022; revised 2 October 2022; accepted for publication 3 October 2022

The earliest fossils of winged insects (Pterygota) are mid-Carboniferous (latest Mississippian, 328–324 Mya), but estimates of their age based on fossil-calibrated molecular phylogenetic studies place their origin at 440–370 Mya during the Silurian or Devonian. This discrepancy would require that winged insects evaded fossilization for at least the first ~50 Myr of their history. Here, we examine the plausibility of such a gap in the fossil record, and possible explanations for it, based on comparisons with the fossil records of other arthropod groups, the distribution of first occurrence dates of pterygote families, phylogenetically informed simulations of the fossilization of Palaeozoic insects, and re-analysis of data presented by Misof and colleagues using updated fossil calibrations under a variety of prior probability settings. We do not find support for the mechanisms previously suggested to account for such an extended gap in the pterygote fossil record, including sampling bias, preservation bias, and body size. We suggest that inference of an early origin of Pterygota long prior to their first appearance in the fossil record is probably an analytical artefact of taxon sampling and choice of fossil calibration points, possibly compounded by heterogeneity in rates of sequence evolution or speciation, including radiations or ‘bursts’ during their early history.

ADDITIONAL KEYWORDS: Carboniferous – Devonian – diversification – Palaeozoic – phylogeny.

INTRODUCTION

Winged insects (Pterygota) contain more described species than the rest of all multicellular life (Grimaldi & Engel, 2005), and ascertaining the age of pterygotes is key to reconstructing the evolution of terrestrial ecosystems. Although the pterygote fossil record extends back as far as the mid-Carboniferous (328–324 Mya), age estimates for the origin of Pterygota derived from fossil-calibrated molecular phylogenetic studies have ranged from the Late Devonian, ~370 Mya (Wheat & Wahlberg, 2013), to the Silurian, ~430 Mya (Johnson et al., 2018). These estimates highlight broader questions about the completeness of the insect fossil record, as each requires a substantial gap between the most recent common ancestor of pterygotes and their oldest known fossil representative (Fig. 1). Understanding this discrepancy requires that we evaluate not only the robustness of the insect fossil record itself but also the extent to which choice of fossil calibration points and parameters impact the reliability of divergence time estimates. In this paper, we examine the plausibility of a major gap in the fossil record of winged insects,
Discordance between age estimates derived from time-calibrated phylogenetic studies and estimates based directly on the fossil record is not uncommon (Ksepka et al., 2014; Beaulieu et al., 2015; Brown & Smith, 2018). In addition to winged insects, significant discrepancies have been noted for Metazoa, angiosperms, and mammals (Phillips & Fruciano, 2018; Budd & Mann, 2020; Ramirez-Barahona et al., 2020) as well as myriapods (Fernandez et al., 2018) and arachnids (Lozano-Fernandez et al., 2020). The fossil record of Pterygota is not nearly as complete as those of brachiopods, bivalves, gastropods, or cephalopods (Kidwell & Flessa, 1996; Foote & Sepkoski, 1999), but far more so than those of tardigrades, nematodes, and various other invertebrates (De Baets et al., 2021; Mapalo et al., 2021). Although the pterygote record is not sufficient to pinpoint the origins of most extant taxa with precision, here we examine the plausibility of such an extended gap in the fossil record for such a speciose group. The more widely distributed or ubiquitous a given group, the less likely that it will evade fossilization entirely through a given interval of geological time. Along with holometabolism, wings are considered a key innovation driving the tremendous diversification of insects (Nicholson et al., 2014), and pterygotes outnumber all other terrestrial arthropods combined among fossils from the Bashkirian onwards (323 Mya; Fig. 1). Although there is little consensus as to when Pterygota originated, the possibility that crown-group pterygotes became widespread within a few million years of their origin cannot be discounted.

The primary goal of this contribution is to evaluate the plausibility of a gap during the first 50+ Myr of pterygote evolution by examining possible explanations and the possible impacts of fossil calibrations, prior schemes, and parameterizations on re-analyses of phylogenomic data. After reviewing the insect fossil record in general terms and exploring possible explanations for the absence of pre-Carboniferous pterygote fossils, we perform four analyses. First, we use a comparative palaeontological approach to assess the likelihood of systematically overlooked pterygote fossils. Second, we use palaeontological data to directly estimate the age of Pterygota. Third, we use different fossil calibration compendia and calibration schemes to evaluate the molecular dating scheme proposed by Misof et al. (2014). Fourth and finally, we use simulations to estimate expected gap lengths in the fossil record of pterygotes.

Node-dating procedures are sensitive not only to analytical parameters, including choices of fossil
calibration points, and dataset-specific features such as taxon sampling and missing sequence data, but also to locus- and taxon-specific histories of rate heterogeneity. At issue, therefore, are the evolutionary, geological, and taphonomic explanations of the putative gap itself, and the sensitivity of its inference to choices of parameterization and fossil calibration.

BACKGROUND: INFERENTIAL EXPLANATIONS FOR AN APPARENT FOSSIL GAP

The Palaeozoic fossil record of terrestrial arthropods can be divided into two phases. First, from the appearance of the first terrestrial arthropods during the Silurian, through the Middle Mississippian (Viséan; Edgecombe, 2011; Dunlop et al., 2018), definitive fossil pterygotes are absent (Haug & Haug, 2017), and apterygote hexapod fossils appear even more rarely than myriapods and arachnids (Whalley & Jarzembowski, 1981; Shear et al., 1984; Labandeira et al., 1988; Fayers & Trewin, 2005) – the rarity of all groups driven largely by Lagerstätten effects (Edgecombe & Legg, 2013). The Late Mississippian (Serpukhovian; 330.9–323.2 Mya) marks an inflection point during which no described chilopods and only one possible arachnid is known, but multiple pterygotes appear by the end of the interval (Brauckmann & Schneider, 1996; Prokop et al., 2005). In the second phase, from the Pennsylvanian onwards, apterygote hexapods remain relatively rare while pterygote fossils outnumber those of all other terrestrial arthropods. Here we summarize aspects of the pterygote fossil record that bear on the plausibility of an extended gap between the first known winged insects and inferred ages of origin, and review explanations put forth to account for their ‘sudden’ appearance in the fossil record. These include scenarios involving rapid changes in rates of evolution and empirical bias against pterygote fossils, especially small ones.

Both the calibration and the interpretation of dated phylogenies depend not only on the inferred ages of fossils used as calibration points but on the differentiation of crown- vs. stem-groups among those fossils. ‘Crown-group’ refers to the smallest monophyletic group that includes all extant members of a given clade, their most recent common ancestor, and any extinct descendants thereof. Crown-group Hemiptera, for example, includes all extant and extinct descendants of the most recent common ancestor of Sternorrhyncha, Coelorrhyncha, Auchenorrhyncha, and Heteroptera. ‘Stem-group’ refers to extinct taxa that are phylogenetically adjacent to but outside the crown-group. Stem-group Hemiptera includes all taxa that are more closely related to Hemiptera than to any other extant order, but are not descended from the most recent common ancestor of extant Hemiptera. ‘Total-group’ encompasses the stem and crown of a given group.

The fossil record contains unambiguous stem representatives of most pterygote orders (Willmann, 1989; Ansgor, 2002; Rasnitsyn & Quicke, 2002; Grimaldi & Engel, 2005; Engel et al., 2009; Huang & Nel, 2009; Shang et al., 2011; Nel et al., 2013; Delsos et al., 2016; Pohl & Beutel, 2016) with the three possible exceptions of Zoraptera, Megaloptera, and Lepidoptera. Plausible explanations exist for each of these: Zoraptera are the least speciose and least frequently encountered of all extant winged insect orders; Megaloptera is the second-least speciose of the holometabolous orders, and its proximity to Neuroptera complicates the task of distinguishing stem Megaloptera from stem Neuroptera; and Lepidoptera have thickly scaled wings and bodies that prevent them from sinking from water surfaces to the sediments where fossilization can occur (Grimaldi & Engel, 2005). Although the fossil record contains various potential stem representatives of each of these three orders, none has been conclusively excluded from their crown-groups (Engel, 2004; Mashimo et al., 2014; Sohn et al., 2015).

As various studies have reiterated (Smith et al., 2006; Coiro et al., 2019; Budd & Mann, 2020), the appearance of stem- and crown-group fossils in the sequence predicted by phylogenetic reconstruction evinces the fossil record’s general reliability. Primitively wingless insects (Shear et al., 1984; Fayers & Trewin, 2005) appear in the fossil record tens of millions of years earlier than winged insects. Palaeoptera (Brauckmann & Schneider, 1996) and Polynoeoptera (Prokop et al., 2005) appear earlier than Paraneoptera (Laurentiaux, 1952). Paraneoptera are paraphyletic with respect to Holometabola (Whiting et al., 1997; Ishiwata et al., 2011; Misof et al., 2014; Johnson et al., 2018) and appear before them (Nel et al., 2007, 2013). Holometabola, in turn, appear well before the crown-groups of its component orders (Haug et al., 2015; Wolfe et al., 2016).

Obviously the fossil record is not sufficiently complete to expect that it recapitulates phylogeny with fine-tuned precision or accuracy, beyond the very broad-brush sequence of major taxa that are well represented as fossils; younger taxonomic units are expected to be less well represented as fossils, much less captured sequentially in the fossil record. Two kinds of comparisons are especially informative when utilizing the potential correspondence between fossils and phylectic sequence to evaluate the completeness of the fossil record. First are comparisons among taxa with different preservation biases. Second are comparisons of stem groups to their corresponding crown groups.
The appearance of Apterygota and Pterygota in phyletic sequence is a particularly informative example due to the disparities in preservation potential among these two groups. The Pennsylvanian–Permian fossil record contains over 5500 described occurrences of Pterygota, but fewer than 20 described occurrences of apterygote insects (Durden, 1975; Rasnitsyn et al., 2004; Rinehart et al., 2005; Engel, 2009; Staniezek et al., 2014; Mendes, 2018; Liu et al., 2021). Because apterygote insects are so much less likely to become fossilized, their presence in the fossil record during an extended interval from which pterygotes are unknown can be reasonably interpreted as true evidence of pterygote absence. Findings of apterygote insects from Early (Fayers & Trewin, 2005) and Middle Devonian (Shear et al., 1984) deposits constitute a canonical example of major groups appearing in phyletic sequence.

Some discussions of fossil gaps have centred on macroevolutionary explanations to account for whether a group for which a purported gap exists, pterygotes in this case, underwent a rapid radiation or evolutionary ‘burst’ early in their history, followed by a slowdown or stasis (Foote et al., 1999; Beck & Lee, 2014; Beaulieu et al., 2015). Although the presence of stem taxa for most major winged insect groups speaks to the strength of the fossil record, the earliest known fossil representatives of Holometabola – which appeared after several major pterygote lineages had arisen (Misof et al., 2014) – are barely 10 Myr younger than the earliest known Palaeoptera and Neoptera (Nel et al., 2013; Haug et al., 2015). This timing would suggest a rapid radiation among crown Pterygota immediately following their appearance. Such an evolutionary burst might be expected in light of the manifold advantages that wings confer, such as locating food resources, avoiding predators, regulating body temperature, finding mates, and other advantages of colonizing a three-dimensional environment (Grimaldi & Engel, 2005). Although such scenarios would be reflected in highly asymmetrical rates of speciation and/or molecular evolution, these are difficult if not impossible to parse from neontological data alone and the ‘early burst’ hypothesis is thus equally difficult to test using molecular data, even with denser sampling of taxa and loci (Tihelka et al., 2021; Craig et al., 2022).

In addition to macroevolutionary explanations, statistical and intrinsic biological explanations have been variously invoked to account for gaps in the fossil records of groups other than insects, primarily sampling bias (Gardner et al., 2019; Harper et al., 2019), preservation bias (Doyle, 2012; Dos Reis et al., 2015), and small size (Wray, 2015; Han et al., 2017; Li et al., 2019). None of these explains the lack of pterygote fossils before the Pennsylvanian. The interpretation of the fossil gap as a sampling artefact would rest on the idea that insufficient collecting effort has been dedicated to Devonian and Mississippian insects. However, a number of productive arthropod compression localities are known, such as Alken an der Mosel (Poschmann & Dunlop, 2010), Campbellton (Shear et al., 1996), Gilboa (Shear & Selden, 2001), and the Rhenish Slate Mountains (Poschmann & Dunlop, 2011), and an abundance of macerated material has been sifting (Shear et al., 1984; Labandeira et al., 1988). The case is simply that none of this material contains winged insects despite wings being the most abundantly preserved insect body part (Karr & Clapham, 2015). Similarly, the few putative fossil Pterygota that have been described from as far back as the Silurian (Agnus, 1904; Rohdendorf, 1961, 1970; Engel & Grimaldi, 2004; Gareast et al., 2012) have all subsequently been shown to belong to different arthropod groups, such as centipedes and crustaceans (Rohdendorf, 1972; Hornschemeyer et al., 2013; Haug & Haug, 2017). Again, terrestrial arthropod fossils are present, just not winged insects.

With respect to preservation bias, no mechanism has been suggested to explain how pterygotes might have been excluded from lowland habitats in which arachnids and myriapods were fossilized during the Silurian–Mississippian. Arachnids and myriapods diversified considerably during this interval and reached depositional settings where they became preserved (Edgecombe, 2011; Dunlop et al., 2018), occasionally alongside fragments of apterygote hexapods (Shear et al., 1984; Labandeira et al., 1988).

Finally, minute body size has been proposed as an explanation of how crown representatives of groups such as early Bilateria, which ostensibly diversified during the Ediacaran, could have evaded fossilization (Budd & Mann, 2020). In the case of insects, however, homologies among the primary wing veins suggest that the common ancestors of the major groups of pterygotes all had wings at least 5 mm long (Polilov, 2016). The primary wing veins have been recognized as homologous across all Pterygota (Schubnel et al., 2020), but are nearly always lost in very small insects (Polilov, 2015), suggesting that extant pterygote orders diverged from common ancestors with body lengths significantly greater than those of the minute arachnids recovered from Silurian–Mississippian deposits (Dunlop, 2019). Moreover, the two oldest known pterygote fossils represent both Palaeoptera and Neoptera but each has wings 11 mm long (Brauckmann & Schneider, 1996; Prokop et al., 2005); as noted previously (Rasnitsyn, 2003), minimum known wing lengths decrease from the Pennsylvanian through the Permian and into the Triassic (Fig. 2).
DIGGING DEEPER INTO THE FOSSIL RECORD OF HEMIPTERA

The origin of Hemiptera (true bugs, aphids, cicadas, scale insects), the most ancient and the only hemimetabolous of the ‘big five’ pterygote orders, presents an example of stem and crown groups appearing in phyletic sequence, at odds with one or more divergence time estimates that greatly pre-date first appearances in the fossil record. The earliest unequivocal hemipteran fossils – the only definitive stem-group hemipteran, followed by the first possible crown-group hemipteran (probably in the stem-group of Euhemiptera) – appear within 16 Myr in the late Carboniferous (Pennsylvanian) (Nel et al., 2013; Szwedo, 2016). During the Permian, however, hemipteran fossils become much more common. Representatives of all three non-heteropteran suborders, and five of the seven non-heteropteran infraorders, are known from this period (Szwedo, 2016). This implies that (1) Hemiptera diversified quickly and (2) diversification of crown-group Hemiptera, if not total-group Hemiptera, occurred from the Pennsylvanian onwards.

The explosive diversification of Hemiptera during the Permian has long been noted in the palaeontological literature (Wootton, 1981). Of the 1023 known pterygote fossils older than 304 Mya (pre-Gzhelian), two are Hemiptera (Laurentiaux, 1952; Nel et al., 2013). Of the 842 pterygote occurrences known from the Gzhelian (304–299 Mya), two are putative hemipterans but their affinities within the order remain unknown (Szwedo, 2016). In contrast, 531 of the 3636 known Permian occurrences of Pterygota (14.6%) have been assigned to Hemiptera, and 94% of these Hemiptera have been assigned to the total-group of an extant infraorder.

Several published time-calibrated phylogenies bear on the Hemiptera (Misof et al., 2014; Johnson et al., 2018; Montagna et al., 2019; Ye et al., 2022). Using BEAST (Drummond & Rambaut, 2007), Misof et al. (2014) inferred an Early Permian age for Hemiptera, and suggested that piercing-and-sucking feeding damage on Devonian liverworts (~380 Mya) might be attributable to hemipteroid insects, although Labandeira et al. (2013) had expressed doubts on palaeontological grounds that Hemipteroidea could have originated over 70 Myr before their first appearance in the Pennsylvanian. The divergence times inferred by Johnson et al. (2018) reconstructed with MCMCTree (Yang, 2007) are noteworthy in several respects. Their divergence estimates included an age for crown Hemiptera near the midpoint of the Devonian (419–359 Mya), with all four hemipteran suborders diverging during the Mississippian, and two of these > 20 Myr before the first appearance of winged insects in the fossil record. Their analyses further depict nine of 11 infraordinal diversifications occurring by the end of the Permian (299–252 Mya), among them the two former ‘homopteran’ infraorders (Auchenorrhyncha, including Fulgoroidea, and Sternorrhyncha), which are inferred to have radiated primarily during the Pennsylvanian – a period from which only four total-group Hemipteran fossils are known. This scenario requires either: (1) that these infraorders diversified during the Pennsylvanian, but evaded fossilization, while other Pterygota such as Palaeodictyoidea and Orthopteroidea became fossilized in considerable numbers; or (2) that these hemipteran infraorders were excluded from depositional environments by a mechanism that did not impact other winged insects. Any such mechanisms (or bias) would then also have had to disappear during the Permian to account for the prevalence of Hemiptera in the entomofauna of that period. Given that there are over 500 hemipteran fossils known from the Palaeozoic, none of which are heteropterans, and that the first fossils appeared in a phylogenetically predictable sequence, the Palaeozoic hemipteran fossil record is not deeply suspect, and it is therefore reasonable to question whether the divergence of Heteroptera from ‘Homoptera’ could have been so strongly discordant with the fossil record, occurring before the appearance even of Pterygota in the fossil record.

The timing of infraordinal diversification within the Heteroptera is especially problematic. The only Palaeozoic heteropteran fossil is dubious, and the first unequivocal heteropteran appears in the Mesozoic (Fig. 3; Szwedo, 2016). The estimates of Johnson et al. include a divergence between Heteroptera and Auchenorrhyncha in the Late Devonian, well before any wings appear in the fossil record, with the basal
heteropteran infraorders diverging by the Permian/Triassic boundary. In other words, whereas the fossil record shows explosive radiations of ‘Homoptera’ confined to the Permian following basal splits in the Pennsylvanian, Johnson and colleagues infer a Palaeozoic diversification history of Hemiptera that began almost 100 Myr earlier, and encompassed not only ‘Homoptera’ but also Heteroptera – the latter of which appear in the fossil record from the Triassic onwards.

On an infraorder-by-infraorder basis, Hemiptera do not appear in the fossil record in perfect phyletic sequence (Fig. 3). For example, Dipsocoromorpha appear before Gerrormorpha, and both Cimicomorpha and Pentatomomorpha appear before Leptopodomorpha. Aphidomorpha and Psyllodea appear before Aleyrodomorpha. However, the ubiquity of Scytiomorpha (putative stem-group Heteroptera; Shcherbakov, 2021) during the Permian, and the absence of crown-group Heteroptera during most if not all of this period, suggests that few if any divergences among crown-group Heteroptera occurred before the Triassic. Likewise, Ingrumorpha (the proposed stem-group of Coleorrhyncha; Wang et al., 2009; Ziade et al., 2011) appears well before crown-group Coleorrhyncha. The indeterminate Sternorrhynca illustrated in Figure 3 include many relatives of Aleyrodomorpha and Coccidomorpha (stem-group Sternorrhynca, Aphidiformes, and Psylliformes; Drohojowska et al., 2020) accounting for the lack of definitive Palaeozoic representatives of these infraorders. In summary, we see stem groups appearing in the fossil record before crown groups for Hemiptera, Heteroptera, Coleorrhyncha, Sternorrhyncha, and probably Euhemiptera.

EMPirical factors affecting divergence time estimates

Estimating and assigning divergence times to nodes on a phylogenetic tree is deceptively straightforward (Brown & Smith, 2018), and general interactions between prior probability settings and underlying taxon sampling, intrinsic rate heterogeneity, and missing data are difficult to evaluate based on individual cases. (Of note, our discussion of dating analyses is limited to node-dating, wherein fossils are used to date divergences, rather than tip-dating, wherein fossil taxa are scored for characters and incorporated into a phylogenetic study as terminals.) We differentiate analytical input parameters such as fossil choice, prior probability settings, taxon sampling, and the extent of missing data from underlying factors intrinsic to the data under analysis, such as the extent of rate heterogeneity among loci. Beyond analytical input parameters, rate heterogeneity is perhaps the most intractable phenomenon potentially giving rise to over-estimates of ages. Beaulieu et al. (2015) found that clade-specific rate heterogeneity can cause the age of a group to be overestimated by tens of millions of years. Building in part on these findings, Brown & Smith (2018) noted that the joint prior, generated by combining a branching process prior with the age distributions for fossil calibrations, may yield age estimates far older than those implied by the fossil record.

The choice of fossil calibrations clearly impacts divergence time estimates, in some cases perhaps more so than missing sequence data (Zheng & Wiens, 2015). Such impacts appear to be a function of phylogenetic position and the accuracy of a given fossil’s assignment to a stem vs. a crown group. Unfortunately, the choice of any given fossil calibration point in a node-dating analysis represents assumptions that are not themselves testable through that analysis. Even with partitioned analyses, an analysis based on fossil calibrations tests neither the wisdom of a given calibration point nor the accuracy of the fossil record writ large. As such, the fossil calibration points used to calibrate molecular phylogenetic studies warrant continuous scrutiny.

An additional factor complicates attempts to date the earliest divergences among winged insects: the paucity of fossils that can be used to calibrate deeper nodes in the hexapod tree. Few pre-Mississippian apterygote hexapods have been recognized (Whalley & Jarzembowski, 1981; Shear et al., 1984; Labandeira et al., 1988; Fayers & Trewin, 2005). Of these, only Rhyniella praecursor Hirst and Maulik 1926 can be confidently assigned at the level of order (Dunlop & Garwood, 2018). Stem-group Hexapoda, believed to have lived in aquatic habitats, have not been identified in the fossil record (Edgecombe et al., 2020). The pancrustacean topology recovered by Misof et al. (2014) differs from topologies that have emerged more recently (Schwentner et al., 2017; Gribet & Edgecombe, 2019; Lozano-Fernandez et al., 2019), complicating the inclusion of various potential fossil calibrations. Misof et al. calibrated their tree with two non-hexapod pancrustaceans: the Silurian ostracod Pauline avibella Siveter et al. 2013 and the Jurassic/Cretaceous cladoceran Daphnia sp. The fossil calibration compendium of Wolfe et al. (2016) and recent addenda include three non-hexapod pancrustaceans that can be used to calibrate Misof et al.’s tree as well as three non- and pre-pancrustacean fossils: the Devonian branchiopod Lepidocaris rhyniensis Scourfield & Calman 1926, the Silurian malacostracan Cinerocharis magna Briggs et al. 2004, the Ordovician/Silurian ostracod Luperisca incuba Siveter et al. 2014, the Cambrian crown-group pancrustacean Yicaris dianensis Zhang et al. 2007,
Figure 3. A timeline of hemipteran evolution. A, fossil occurrences plotted onto the time-calibrated phylogeny of Johnson et al. (2018). B, gaps between the origins of different groups, as inferred by Johnson et al. (2018), and their first appearances in the fossil record. C, ranges of relevant groups of the terrestrial biota.

**METHODS**

We conduct four analyses. The first compares occurrences of Palaeozoic Pterygota to those of arachnids, chilopods, and primitively wingless hexapods. The aim of this analysis is to evaluate the plausibility of a scenario in which Pterygota existed, but evaded fossilization, during the Silurian through Mississippian. The second analysis also uses fossil data, but only for Pterygota, to estimate the age for this group. The third analysis uses two fossil datasets to explore node-calibration procedures in the phylogenetic analysis of Misof et al. (2014): that of Wolfe et al. (2016) and that of Misof et al. itself. In these analyses, the divergence time estimates of Misof et al. (2014) were re-analysed using these two fossil datasets and a variety of parameter sets. The fourth analysis utilizes the phylogenetic topology recovered by Misof et al. (2014) to simulate the pterygote fossil record, with the aim of estimating the completeness of the fossil record by comparing various simulations to known occurrence data.

**PALAEOENTHALICAL ANALYSES**

The first two analyses, both palaeontological, were conducted with Palaeozoic data downloaded from the Paleobiology Database (PBDB; paleobiodb.org). These data were primarily compiled by Clapham et al. (2016). One insect occurrence was removed from the dataset: a roachoid whose possible age range listed in the PBDB spans 60 Myr (Martynov, 1936).

**Bootstrapping analysis**

To evaluate the plausibility of a gap in the pterygote record, and specifically to explore the probability that winged insects existed during various intervals before the Pennsylvanian, we bootstrapped arachnid and chilopod occurrence data from Palaeozoic intervals before and after the appearance of Pterygota in the fossil record. Occurrence data for Palaeozoic arachnids, chilopods, and Pterygota were downloaded from the PBDB; we verified that all Silurian–Mississippian chilopod occurrences listed in Edgecombe (2011) and arachnid genera listed in Dunlop et al. (2018) were included in the PBDB. We added missing occurrences of Alllobuthus pescei Vachon and Heyler 1985, Parageralinura neerlandica Laurentiaux-Viera and Laurentiaux 1961, Gondwanarachne argentinensis Pinto and Hunicken 1980, Saccogulus selden Dunlop et al. 2006, Crussolum sp. Shear et al. 1998, and Crussolum crusserratum Shear et al. 1998. The only taxon missing from the PBDB and from our dataset is the scorpion *Paraisobuthus* sp. Kjellesvig-Waering 1986 from the Pennsylvanian Mina la Mora locality of León, Spain (Gutierrez-Marco et al., 2005), which we were unable to include due to insufficient locality and collection data. We downloaded final datasets from the PBDB on 10 February 2021. The midpoint between the minimum and maximum age for each specimen was assigned as its age.

The results presented here are primarily confined to the fossil records of insects and arachnids. Diplopods were not included in our bootstrapping analyses because they have mineralized exoskeletons (Plotnick, 1990), which confer a greater potential for preservation than those of most other terrestrial arthropods (Edgecombe & Legg, 2013). Arthropleurans were excluded from the dataset as well due to their proposed phylogenetic affinity with diplopods (Kraus & Brauckmann, 2003). *Arthropleura* appears not to have had a mineralized exoskeleton (Davies et al., 2022); the exclusion of all arthropleurans constitutes a conservative approach, as the resulting *P*-values would be lower if arthropleurans were included. Only a handful of chilopod occurrences are known from the Palaeozoic, most of which are over 400 Myr old (Shear & Edgecombe, 2010; Haug & Haug, 2017) and the rest occurring at Mazon Creek (Mundel, 1979; Shear & Edgecombe, 2010) alongside many arachnids and insects.

Specimen and measurement datasets for Pterygota were downloaded for the purpose of determining which fossil occurrences were preserved with wings. For winged insects, the only specimens considered ‘winged’ are those for which a measurement of wing length is available (including wing fragments, nymphal wing pads, and structures derived from wings such as tegmina and elytra); none of these specimens belong to secondarily wingless lineages and all represent individuals that survived long enough for wings to develop sufficiently to be recognized as such. These include isolated wings that were preserved without other body parts. We chose a conservative approach, and did not include presumptive pterygote specimens with at least one body part, but no wings, measured. Such specimens might belong to secondarily wingless lineages, represent larval or nymphal pterygotes that died before their wings developed, or be fragments of a disarticulated winged insect such as an isolated mandible or femur. Ichnofossils such as insect galls on leaves were not included. For many Pterygota, measurements were not available, so it is unknown whether those individual specimens include fossilized wings.
The bootstrapping procedure was used only on contiguous intervals that (1) pre-date the Pennsylvanian, by which point winged insects had indisputably originated, and (2) encompass the Late Mississippian (Serpukhovian Stage), because winged insects are known only from the end of this interval. The bootstrapping procedure was performed for each of the three epochs/stages of the Mississippian, the entirety of the Devonian, and the entirety of the Silurian. Because 39 of the 41 Mississippian arachnid occurrences are from the Middle Mississippian, this epoch/stage was further subdivided. (The Mississippian is the only geological period discussed here for which each epoch contains only one stage, and thus epochs and stages are interchangeable.) The ‘late-Middle’ Mississippian includes the three arachnid localities whose midpoint ages are 331.2 Mya, only 0.2 Myr below the Middle/Late Mississippian boundary; all of these localities have Late Mississippian minimum ages. The ‘mid-Middle’ Mississippian includes the four arachnid localities whose midpoint ages are less than 340 Mya. Because only one arachnid specimen has been tentatively assigned to the Late Mississippian, this interval cannot be subdivided further for additional bootstrapping analyses.

The bootstrapping procedure was conducted as follows. First, the number of arachnid and chilopod occurrences during the interval was summed. Then, the same number of arachnid and chilopod occurrences was sampled with replacement from the Pennsylvanian and Permian. This procedure was repeated for 1,000,000 iterations. The resulting P-value represents the proportion of iterations in which none of the resampled Pennsylvanian–Permian arachnids and chilopods occur at localities that also include winged insect specimens. This P-value represents the probability of recovering the known fossil record of arachnids and chilopods for a given interval without sampling a single winged insect at any of the localities where the arachnids and chilopods were found. This probability is premised on the assumptions that: (1) winged insects existed during that interval; (2) winged insects were no less abundant than arachnids and chilopods; and (3) the probability of fossilization and recovery for pterygotes, relative to other terrestrial arthropods, did not change from the Silurian–Mississippian to Pennsylvanian–Permian.

**PyRate**

Our second analysis used PyRate (Silvestro et al., 2019) to estimate the age of Pterygota. PyRate uses a Bayesian framework (reversible jump Markov chain Monte Carlo) to estimate origination and extinction rates in deep time from fossil occurrence data, and the age of the clade under consideration. We used family-level data for this analysis. We prepared ten input files, setting the seed from 1 to 10 with the set.seed() function in R. The use of ten replicates addresses the uncertainty associated with the sampling of random dates within the age range for each family. Each replicate input file was analysed separately, thus receiving a different random seed in python.

We used the -qShift option to independently estimate preservation rates for geological intervals as follows: 323.2 and 300.5 Mya mark inflection points in the rise and fall of Palaeozoic coal (Nelsen et al., 2016); 130.1 Mya, after which insect inclusions in amber become prevalent in the fossil record with the appearance of Lebanese amber (Maksoud & Azar, 2020); and 2.589 Mya marks the point after which insects preserved in un lithified and poorly lithified sediments become prevalent (Böhler, 1995; Kiselev & Nazarov, 2009). We used the -pP 1.5 0 option for the vector of preservation rates. The first value confirms the default shape prior of the gamma distribution (1.5) and the second value indicates estimation of the rate prior from the data. We analysed each replicate input file for 20,000,000 iterations, sampling every 1,000 iterations.

We read the results into R with the tracerer package (Bilderbeek & Etienne, 2018) and determined the number of iterations to discard as burn-in by iteratively calculating the effective sample size for different proportions of burn-in, using the ess function in the mcmcse package (Flegal et al., 2021). Among the results that PyRate generates is an estimate of the root age for the clade under consideration, in this case Pterygota. As discussed below, PyRate does not account for variability in the completeness of the fossil record before the appearance of the fossils under consideration. PyRate also generates estimates of origination, extinction, and diversification rates through time, which bear on the plausibility of an early burst of evolution.

**Node-dating analyses**

The analysis of Misof et al. (2014) represents the most complete phylogenomic analysis of insect orders to date that employs fossil calibrations to estimate divergence times. Misof et al. used 37 fossil calibrations spanning chelicerates, myriapods, crustaceans, and hexapods, 18 of which represent non-PTerygote arthropod lineages. We revisited the choice of fossil calibrations and prior probability assignments to examine their impacts on age estimates for Pterygota. Doubts have been raised about the use of Rhyiognatha hirsti as a calibration point (Haug & Haug, 2017). This species is known from a fossilized head capsule fragment from the Rhynie Chert, and pre-dates any definitive Pterygota by ~80 Myr. R. hirsti has been interpreted to possess...
odonate affinities, placing it within Pterygota and crown-group Dicondylia (Engel & Grimaldi, 2004), and was therefore treated as a fossil calibration for Dicondylia by Misof et al. (2014) and as a fossil calibration for Insecta by Wolfe et al. (2016). However, a recent re-examination with advanced imaging techniques suggests this specimen is a myriapod (Haug & Haug, 2017). We re-analysed the divergence time estimates of Misof et al., with and without R. hirsti, to evaluate its impact on estimated origination dates (see below) and additionally using the data of Wolfe et al. (2016). All analyses were run under the generally preferred lognormal prior probability settings and uniform (flat) priors to examine whether the estimates would be comparable, as well as under specific constraints based on defensible calibration points.

Relevant fossil calibrations from Wolfe et al. (2016) were updated as follows. Casiogrammus ichthyeros Wilson 2005 was used to calibrate the minimum age of Diplopoda following Fernández et al. (2018). The minimum age of Oligotypus huangheensis Ren et al. 2008 was updated according to the findings of Trumper et al. (2020). Palaeotaeniopteryx elegans Sharov 1961 was removed as a calibration for crown Plecoptera following Evangelista et al. (2019) and replaced with Pronemoura shii Yushuang et al. 2011. Of note, two insect calibrations in Wolfe et al. (2016) are housed in private collections, but both of the nodes calibrated with these fossils can be calibrated with other fossils from the same deposit that are housed in public museums (Mendes & Wunderlich, 2013; Chen et al., 2019). A few nodes mentioned in Wolfe et al. (2016) can be calibrated with geologically older fossils described since 2016 (e.g. Fikacek et al., 2020; Qvarnstrom et al., 2021; Yang et al., 2021). However, in order to compare our results to those of Misof et al. (2014) we chose to retain all fossils from the Wolfe et al. (2016) compendium believed to represent the relevant crown group.

We also examined the effect of using a maximum age for Pterygota of 331.1 Mya corresponding roughly to the maximum age of the Viséan/Serpukhovian boundary at 330.9 ± 0.2 Mya (Richards, 2013), and a maximum age of 324 Mya corresponding to the age of the oldest verifiable pterygote fossils (Brauckmann & Schneider, 1996; Prokop et al., 2005), which were deposited just below the Mississippian/Pennsylvanian boundary at 323.2 ± 0.4 Mya (Richards, 2013). We rounded the age of this boundary up from 323.6 to 324 Mya because the pterygote fossils themselves slightly pre-date this boundary.

All sequence data were downloaded from https://science.sciencemag.org/content/suppl/2014/11/05/346.6210.763.DC1. We obtained 102 separate partitions from the download site. Some of these were further sub-partitioned (e.g. p1 into p1A to p1E; p3 into p3A and p3B; p7 into p7A to p7D; p14A to p14F; p36 into p36A and p36B). In the original publication the sub-partitions for these five gene regions were combined and treated as a single partition. We have treated these data likewise, but we found that treating these sub-partitions as full partitions had no significant impact on the results of our analyses. In total there were 85 gene partitions we could obtain from the data archives for the study, ranging up to 3500 amino acids in length. Each partition was treated as described in the original publication with respect to model usage (the majority used the LG amino acid substitution matrix). However, we focused our analyses on the ages of four clades: Pterygota, Holometabola, Diptera, and Neuroptera.

Calibration dates from table S8 of Misof et al. (2014) are referred to in the figures as ‘Misof’. ‘Pterygote calibration time’ refers to the 331 Mya constraint and prior-setting scheme. Additional calibration dates obtained from Wolfe et al. (2016) (see Supporting Information, Appendix S1) are referred to as Wolfe followed by the pterygote calibration time and prior setting. For instance, MisofLogN405 indicates an analysis scheme using the Misof calibration of 405 Mya and LogN (lognormal) settings (the original setting for analysis in Misof et al.). Table S2 shows the analyses used for these comparisons.

XML files were produced using BEAUti from the BEAST package (Drummond & Rambaut, 2007). Partitions were analysed separately as in the original study to facilitate reasonable run times. For the Misof LogN runs we used the same prior setting as in the original study for the MisofLogN405 run. The number of generations run in each Bayesian analysis ranged from 10 million to 30 million.

All analyses were constrained with the original Misof et al. tree. The runs were executed using a linux-based (Centos 7) computer cluster at the AMNH which consists of 1024 hyper-threaded cores over 16 compute nodes. We used BEAST v1.10.4, Beagle 3.1.2, and Oracle Java 8. Beagle was run with CPU resources only, multithreading was disabled with -beagle threading off, and scaling was set to ‘always’. The resulting .log files were combined using the LogCombiner software in the BEAST package. Individual .log files were analysed using Tracer v1.7.1 also from the Beast package. Results were graphed using Tracer v1.7.1 or Excel.

We used an uncorrelated relaxed clock with a relaxed lognormal distribution, replicating in part the analysis of Misof et al. (2014). Because of the large amount of sequence in the original dataset, partitions were analysed separately, resulting in estimates for divergence times for each partition. There are several ways that an overall estimate can be obtained,
and we chose the approach outlined by Misof et al. (2014), which is based on the median estimate from the distribution of individual partition estimates. We present the results of our analyses for Pterygota as well as those for Holometabola, Diptera, and Neuroptera as examplar nodes in the tree.

To compare density distributions of the Misof and Wolfe datasets we used Tracer v.1.7.1. We loaded the combined logfiles for each of the two datasets and graphed the density distributions of user priors against lognormal posteriors.

SIMULATING THE FOSSIL RECORD

For our fourth analysis, a simulation procedure was used to quantify Budd & Mann’s (2020) argument that the appearance of major groups in phylogenetic sequence implies that the fossil record is not so incomplete as to plausibly contain gaps of many tens of millions of years. This procedure is intended to quantify the gap lengths one can expect in the fossil record based on the sequence in which different groups appear. The tree from Misof et al. (2014) for all insects, truncated at the Permian/Triassic boundary, was imported into R.

For each of the three increases in the sampling rate mentioned below, the following procedure was repeated 5000 times. First, the branch lengths were randomly altered while constraining the tree as ultrametric, so that all terminal branches end at the Permian/Triassic boundary. A timescale was assigned to each tree by choosing a random multiplier that fulfilled two criteria: first, that crown-group pterygotes be at least 71 Myr older than the Permian/Triassic boundary, as observed in the fossil record; and second, that crown-group insects be no more than 200 Myr older than the Permian/Triassic boundary (which a direct reading of the fossil record, and all fossil-calibrated molecular phylogenetic studies that we are aware of, have shown). The random multiplier was applied to all branches of the phylogeny to retain ultrametricity.

The sim.fossils.intervals() function of the FossilSim package (Barido-Sottani et al., 2019) was used to simulate a fossil record. To account for Pterygota being less abundant – and therefore less likely to be known from the fossil record – when they first originated, the sampling rate was simulated to increase linearly per geological stage by a total of 2×, 10×, and 50× from the origin of Pterygota until the Permian/Triassic boundary. Sampling rates were specified for each geological stage. The sampling rate for any stage in which insects were present but in which pterygotes had not yet originated was set to the minimum sampling rate mentioned above: 1/2, 1/10 or 1/50 of the sampling rate for the final stage of the Permian. The sampling rate for the Late Mississippian (Serpukhovian) was set to zero to reflect the particularly depauperate record from this interval. Sampling rates did not vary by lineage. For each of the 5000 time-calibration simulations, and for each of the three protocols for increased sampling, a fossil record was simulated from the tree with the following maximum sampling rates: 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, and 10.

The above procedure was then repeated with an additional clustering procedure to mimic the Lagerstätte-driven nature of the insect fossil record. Using the iNEXT function in the iNEXT package in R (Hsieh et al., 2016), we estimated the number of fossil assemblages that we expect per number of insect fossils. These estimates were generated with the above-mentioned Palaeozoic dataset from the PBDB. Then, while running the simulation procedure, we used the base-R function kmeans to cluster the fossil occurrences into the expected number of assemblages.

Each simulation was evaluated on the basis of two criteria: (1) whether Apterygota, Palaeoptera, Polyneoptera, Hemiptera, Holometabola, and Coleoptera appear and, (2) if so, whether they appear in the correct phyletic sequence (Apterygota no later than Palaeoptera and Polyneoptera; Palaeoptera and Polyneoptera no later than Hemiptera and Holometabola; and non-coleopteran Holometabola no later than Coleoptera). We also recorded the total number of insect fossils in each simulated fossil record, as well as the temporal lag between the origin of crown-group pterygotes and their appearance, the temporal lag between the first and tenth appearances of crown-group pterygotes, and the temporal lag between the first and hundredth appearances of crown-group pterygotes.

RESULTS

BOOTSTRAPPING ANALYSIS OF FOSSIL OCCURRENCE DATA

Our bootstrapping procedure found significant differences in the taxonomic composition of the Silurian through Early Mississippian fossil record of terrestrial arthropods, as compared to the Pennsylvanian and Permian record. Regardless of whether the threshold for significance is 0.05 or 0.005, the results of the bootstrapping procedure (Table 1) are consistent with diversification of crown-group pterygotes only after the start of the Late Mississippian, 331 Mya. Because only one arachnid specimen has been tentatively assigned to the Late Mississippian, this interval cannot be subdivided further for additional bootstrapping analyses. (The trigonotarbid Anthracomartus sp. (Copeland, 1957) has been tentatively assigned to the Mabou group.
Miller & Forbes, 2001), whose Brigantian age spans the Viséan and Serpukhovian Stages with a midpoint in the Serpukhovian. Though the only option at present is to treat this occurrence of Anthracomartus sp. as belonging to the Mabou group, and to treat the Mabou group as dating to the Serpukhovian, neither of these designations can be made with any confidence.

Whereas the $P$-value for the Late Mississippian is not significant ($P = 0.72$), the $P$-values for all of the longer intervals tested are significant. The results of the bootstrapping procedure imply that winged insects were absent during the Middle Mississippian (Viséan Stage), and therefore originated during the Late Mississippian (Serpukhovian Stage) – the latest interval in which they could have plausibly arisen given the presence of both Palaeoptera and Neoptera at the Mississippian/Pennsylvanian boundary.

PyRate

Six of the ten replicate PyRate analyses reached convergence, or ‘stationarity’: i.e. estimates of the age of Pterygota remained within a narrow range of 10 Myr or less as the analysis progressed (Supporting Information, Table S1). Of these, the five with the highest effective sample sizes yielded a median age estimate of crown Pterygota between 326.09 and 326.60 Mya. The replicate with the lowest effective sample size yielded a slightly lower median age estimate for crown Pterygota, 324.96 Mya. For these six replicates, the lower bound of the 95% highest posterior density interval was above 324 Mya and the upper bound was below 330 Myr. The origination and net diversification rates, as estimated with PyRate, peaked early in the evolutionary history of Pterygota and then declined precipitously from ~317.5 to 313.4 Mya (Fig. 4).

Results of the PyRate analysis, like those of the bootstrapping analysis, suggest that Pterygota originated during the Late Mississippian – despite the two procedures using different kinds of data. The bootstrapping analysis depends entirely on the fossil records of terrestrial arthropods other than pterygotes, whereas PyRate does not account for non-terygotes in any way. PyRate accounts for the family-level affinities of fossil pterygotes, and for their post-Palaeozoic fossil records, whereas the bootstrapping analysis accounts for neither. The bootstrapping analysis utilizes the co-occurrence of fossils belonging to different taxa and incorporates all fossil occurrences, whereas PyRate ignores co-occurrence/locality data, and accounts for only the first and last appearances of each family.

Table 1. Results of the bootstrapping procedure: the interval examined, the age range of the interval, the number of arachnids and chilopods described from the interval, and the fraction of bootstrapping iterations in which the same number of arachnids and chilopods were sampled from the Pennsylvanian and Permian without ever co-occurring with winged insects

<table>
<thead>
<tr>
<th>Interval</th>
<th>Age (Mya)</th>
<th>$N$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late Mississippian</td>
<td>331–323</td>
<td>1</td>
<td>0.7182</td>
</tr>
<tr>
<td>‘late-Middle’–Late Mississippian</td>
<td>332–323</td>
<td>20</td>
<td>0.0036**</td>
</tr>
<tr>
<td>‘mid-Middle’–Late Mississippian</td>
<td>340–323</td>
<td>28</td>
<td>0.0003**</td>
</tr>
<tr>
<td>Middle–Late Mississippian</td>
<td>347–323</td>
<td>40</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Mississippian</td>
<td>359–323</td>
<td>40</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Devonian–Mississippian</td>
<td>419–323</td>
<td>131</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Silurian–Mississippian</td>
<td>444–323</td>
<td>151</td>
<td>&lt; 0.0001***</td>
</tr>
</tbody>
</table>

Asterisks denote ***$P < 0.001$ and **$P < 0.01$. 

Figure 4. Origination and net diversification rates for Pterygota at the family level as estimated with PyRate, from the Pennsylvanian to Triassic.
NODE-DATING ANALYSES

In our re-analyses of Misof et al.’s node-dating, estimates of pterygote origins vary according to the choice of fossil calibration as well as the assignment of priors (lognormal vs. uniform distribution settings) and the enforcement of constraints. Although estimates based on the calibrations of Misof et al. were consistently older than those employing the Wolfe data, the prior settings of uniform/lognormal distributions appear to impact the results of using these two datasets in opposite ways, with lognormal settings resulting in older estimates with the Wolfe et al. data and the uniform settings resulting in older estimates with the Misof et al. data. The parsed summarized node age values for Pterygota, Neuroptera, Holometabola, and Diptera can be found in ‘INSECT PUB EXCEL.xlsx’. While we focus primarily on the Pterygota node, we also summarize the Neuroptera, Holometabola, and Diptera nodes for comparison. In addition, the raw .log files contain node estimates for all nodes in the original (Misof et al., 2014) tree and these are available from the authors on request.

Misof et al.’s estimate of 405 Mya for Pterygota is recovered using their settings (Supporting Information, Figs S2, S3). The oldest divergence time estimates (412 and 420 Mya) are obtained from applying uniformly distributed priors to the Misof data when the pterygote calibration is set at 324–450 Mya and fixed at 324 Mya, respectively. Under the Misof calibrations, uniform priors result in an older estimate for the age of Pterygota relative to lognormal priors; uniform priors increase the pterygote divergence time estimate by 15 Myr when the calibration is set at 324 Mya and by 7 Myr when set at 324–450 Myr. When no calibration is assigned (n) the estimated age of Pterygota (386 Mya) represents a reduction of 19 Myr, which is 9 Myr less than the 377 Mya obtained under lognormal prior settings. When the calibration is set at 324 Mya under lognormally distributed priors, the obtained age estimate for pterygotes (386 Mya) is 34 Myr less than when uniform priors are imposed. Using the Wolfe et al. calibrations, this effect is reversed and lognormal priors result in older estimates of pterygote age than do uniform priors: 368 vs. 363 Mya (a reduction of 37–42 Myr from 405 Mya) when the pterygote calibration is removed; and 358 vs. 355 Mya (a reduction of 47–50 Myr from 405 Mya) when the pterygote calibration is set at 331 Myr. The most recent estimate from the Misof et al. data (377 Mya) comes from removing the calibration for pterygotes entirely under lognormally distributed priors.

The Wolfe et al. (2016) calibrations consistently yield more recent origin times for Pterygota than the Misof compendium (Figs 5, 6) as well as shorter gaps between the inferred origin of Pterygota and the origins of Holometabola, Diptera, and Neuroptera (Fig. 5). Under uniform prior settings, use of the Wolfe calibrations reduces the age estimate for Pterygota by 42–50 Myr regardless of whether the pterygote calibration is removed entirely or set at 331.1 Mya. Under lognormal prior settings, the estimate is reduced by 37–47 Myr whether the pterygote calibration is removed entirely or set at 331.1 Mya. Applying a strict clock (removing the need for priors) reduces the estimate by 23 Myr (382 Mya).

Following Brown & Smith (2018), we graphed the prior and posterior density distributions for the Pterygota node (Fig. 7). The posterior distributions were noticeably different from the prior distributions, suggesting that any discrepancies in the estimated age of pterygotes are not primarily attributable to insufficiency of molecular data to update the prior. This finding also holds for the other three nodes we examined: Holometabola, Diptera, and Neuroptera (Supporting Information, Figs S4, S5).

RESULTS OF SIMULATIONS

Nearly all of our simulations of the fossil record that yielded at least 1000 Palaeozoic insect fossils share four characteristics. First, in 93.99% of such simulations, the major groups of insects (Apterygota, Palaeoptera, Polyneoptera, Hemiptera, Holometabola, and Coleoptera) appear in the fossil record in the expected phylogenetic sequence (Fig. 8A). We focus here on the simulations with 5000–6000 fossils, because there are currently around 5500 fossil pterygote occurrences described from the Palaeozoic fossil record. Among these simulations, an average of 1.07 Myr passed between the origin of pterygotes and their appearance in the fossil record; an average of 4.01 Myr passed between the appearance of the first and the tenth pterygote fossils; and an average of 20.05 Myr passed between the first and hundredth pterygote fossils. This is broadly similar to what is seen in the fossil record, with fewer than ten pterygotes known from the Late Mississippian (Brauckmann & Schneider, 1996; Prokop et al., 2005) followed by various pterygote assemblages dating to the early Bashkirian such as Malanzán (Riek & Kukalova-Peck, 1984; Pinto, 1986), Ningxia (Gu et al., 2011; Su et al., 2012), and Hagen-Vorhalle (Ilger & Brauckmann, 2008; Brauckmann et al., 2010).

The fossil record, as it is currently known, has a gap of ~5 Myr between the first and tenth pterygote occurrences, and a gap of ~7.5 Myr between the first and hundredth fossil occurrences. It contains ~5500 pterygote fossils from the Palaeozoic, and the major groups appear in the sequence predicted from phylogeny. The simulation results suggest that, when all these conditions hold, there is a gap of less than
10 Myr between the origin of crown-group Pterygota and their appearance in the fossil record. (In contrast to the bootstrapping analysis discussed above, the final number of fossils is not fixed in our simulations.) These simulations suggest, however, that the fossil record is not so sparse that it fails to capture extended periods at the dawn of pterygote history (Fig. 8B–D) – especially if thousands of insect fossils are known.

**DISCUSSION**

While the estimation of phylogenetic timing relies on the suitability of fossil calibration points, the plausibility of estimates that appear to conflict with the fossil record depends on the plausibility of both their underlying assumptions and the evolutionary scenarios they require. The inference of a pterygote origin during either the Devonian (Misof et al., 2014) or Silurian (Montagna et al., 2019) requires that pterygotes evaded fossilization for at least 80 Myr, during which various arachnids, the occasional chilopod, and the occasional apterygote hexapod were preserved. It also requires that the mechanism(s) preventing pterygotes from becoming fossilized did not impede the detection of major pterygote groups in the sequence predicted by phylogenetic reconstruction.
Despite already having existed for millions of years, arachnids flourished in lowland habitats at large body sizes, for tens of millions of years. One might wonder why pterygotes, specifically might have been confined to upland habitats, numerically rare, or constrained to small body sizes while arachnids flourished in lowland habitats at large body sizes, for tens of millions of years. One might wonder why all suborders of Hemiptera, for example (Fig. 3), were unable to invade lowland habitats or increase in body size at the same time as Palaeoptera and Polyneoptera, despite already having existed for millions of years.

The gap between divergence dates estimated directly from the fossil record vs. from fossil-calibrated phylogenetic studies has been attributed specifically to the incompleteness of the fossil record (Barba-Montoya et al., 2018; Morris et al., 2018; Li et al., 2019; Lozano-Fernandez et al., 2020). Although the fossil record is not perfect (Holland, 2017), we are not aware of any previous efforts dedicated to quantifying precisely how long of a gap in the pterygote fossil record can be attributed to its incompleteness. Our bootstrapping analysis suggests that if Palaeozoic pterygotes were as likely to be fossilized as arachnids and chilopods, they simply were not present on the landscape during the Silurian, Devonian, and most of the Mississippian. The findings of Karr & Clapham (2015) suggest that Palaeozoic pterygotes would have been at least as readily fossilized as arachnids and chilopods: whereas pterygote heads, thoraxes, abdomens, and legs probably have preservation potentials comparable to those of equivalent structures in arachnids and chilopods, wings are by far the most common pterygote body part in the pre-amber fossil record.

The exceptional sparsity of the Late Mississippian fossil record of terrestrial arthropods in particular cautions against overinterpretation of the PyRate results. Although none of the 95% credible intervals extend as far back as the earliest 1 Myr of the Late Mississippian (329.9–330.9 Mya), the fossil record from that interval is not complete enough to preclude the possibility that crown-group pterygotes originated during that time. However, especially when considered together, the PyRate and bootstrapping analyses cast doubt upon the possibility that crown-group Pterygota would have originated during the Silurian or Devonian. Neither PyRate nor the simulation analysis make any assumptions about whether pterygotes fossilize as readily as arachnids and chilopods. The narrow 95% highest posterior density interval for the age of crown-group pterygotes obtained from the PyRate analysis falls almost entirely within the second half of the Late Mississippian (Serpukhovian). Because no fossil arachnids or chilopods are known from the first half of the Late Mississippian, there is little evidence to support the notion that crown-group pterygotes post-date the midpoint of this interval. However, because PyRate and the bootstrapping analysis rely on different data and assumptions to arrive at the same conclusion – a Late Mississippian origin of crown-group pterygotes – fossil evidence does not corroborate an origin for crown-group pterygotes older than the late-Middle to Late Mississippian. This conclusion is also supported by the simulation exercise which suggests that, because many thousands of Palaeozoic pterygotes are known from the fossil record and because the major groups appeared in the sequence suggested by their phylogenetic relationships, there
would have been a lag of only a few millions of years between the origin of crown-group pterygotes and their appearance in the fossil record.

A comparison with Lepidoptera (Fig. 9) further highlights the implausibility of a scenario in which crown Pterygota diversified, but evaded fossilization, for an extended period of their early history. Lepidoptera appear sporadically in the fossil record (Sohn et al., 2015). The Jurassic and Early Cretaceous span 101 Myr, post-dating widely accepted evidence
Estimates of the age of crown-group Pterygota from fossil-calibrated molecular phylogenies (more detail in Figure 1)

Estimates of the age of crown-group Lepidoptera from fossil-calibrated molecular phylogenies

Figure 9. Comparison of the fossil records of total-group Pterygota and total-group Monotrysia (Lepidoptera). The grey rectangles span 100 Myr each. The grey rectangle spanning much of the Silurian to Middle Mississippian represents an interval in which Pterygota were already diversifying, according to some of the studies cited in Figure 1, but no wings are known from the fossil record. The second grey rectangle spans the first 100 Myr of the known fossil record of total-group Lepidoptera, which contains only Monotrysia. Compression/impression fossil occurrences are represented by white circles, and non-compression/impression fossils are represented by black circles and are labelled. Data for Lepidoptera are from Sohn et al. (2012); Necrotauliidae and incertae sedis taxa are not shown. Estimated ages for crown-group Lepidoptera are extracted, from oldest to youngest, from Kawahara et al. (2019), Montagna et al. (2019), Rainford et al. (2014), Tong et al. (2015), Wahlberg et al. (2013), Thomas et al. (2020), and Misof et al. (2014).

The sparsity of the lepidopteran fossil record is to be expected for a few reasons. First, the scales on the wings of Lepidoptera are hydrophobic (Kim et al., 2020), causing the wings to float and thus preventing them from sinking into a depositional environment where they can be preserved as fossils. Second, the most ancient families of Lepidoptera – the Monotrysia, which comprise the entire known lepidopteran fossil record until the deposition of Myanmar amber (Sohn et al., 2015) – account for less than 2% of all lepidopteran diversity today (Nieukerken et al., 2011), with no evidence to suggest that they were previously more diverse or abundant than they are currently. In other words, even though there are plausible mechanisms that account for the poor lepidopteran fossil record, we still see an average of one compression/impression fossil lepidopteran per 5 Myr during the first ~100 Myr of their known history. In contrast, we see no fossil Pterygota during the > 100-Myr interval of the mid-Silurian to mid-Late Mississippian in which various publications (Rota-Stabelli et al., 2013; Tong et al., 2015; Johnson et al., 2018; Montagna et al., 2019) found Pterygota to have diversified.

None of the mechanisms we find that could explain the near-absence of Lepidoptera from the fossil record during their early evolutionary history apply to Pterygota or to Hemiptera. The earliest crown Pterygota possessed one of the two features credited with this group’s diversity and abundance, namely wings (Nicholson et al., 2014). In contrast, Monotrysia lack most of the features credited with the diversification and abundance of Lepidoptera: female gonopore separate from the copulatory orifice, tympana, and, in the case of the non-glossotan moths that dominate the Jurassic–Early Cretaceous record, a siphoning proboscis with intrinsic musculature, and a wing coupling mechanism (Grimaldi & Engel, 2005). The complete absence of fossil Pterygota and Hemiptera from tens of millions of years of their purported evolutionary history, therefore appears even more implausible in light of the presence of Monotrysia in a number of Jurassic and Early Cretaceous deposits (Zhang et al., 2013; Sohn et al., 2015).

Potential Causes of Discordant Age Estimates

The lack of evidence for Devonian to mid-Mississippian pterygotes raises the question of why the age estimates generated in many recent studies (illustrated in Fig. 1) are so old. We examined the choice of specific fossil calibrations, and in the course of using multiple prior probability settings we identified dataset-specific behaviours under different priors, which we differentiate from intrinsic attributes of the sequence data, such as the distribution of missing
data, and rate variation among specific clades, which we did not explore. In our analyses, eliminating the controversial calibration point of *Rhyniognatha* results in younger age estimates, as did the systemic substitution of the *Wolfe et al.* (2016) calibration data. As one might expect, the Wolfe data responded to the maximum age constraint for Pterygota by yielding more recent age estimates (Fig. 5). The Misof data counterintuitively produced older dates when a maximum age constraint was imposed for Pterygota. Moreover, the Wolfe data consistently yielded an origin of crown-group Neuroptera prior to the oldest definitive fossil representative of this group, as well as for Diptera in three out of four node-dating analyses. The Misof data, in contrast, always yielded estimates for both groups that post-date their oldest known fossil representatives.

Although substituting the Wolfe data consistently retrieved more recent ages, we noted peripherally that the impact of choosing lognormal relative to uniform prior settings was reversed under alternative sets of fossil calibrations: lognormal prior settings yielded older estimates than uniform prior settings based on the *Wolfe et al.* (2016) calibration data, but older estimates based on the original *Misof et al.* (2014) calibration data obtained from uniform prior settings. We interpret this as a function of how the priors interact with distributions of the input fossil calibrations, and specifically how the uniform prior setting may exaggerate possible disparities in the sampling of those points. The Misof calibration points are less uniformly distributed than those of Wolfe, and specifically they are more heavily concentrated away from the time period when Pterygota could have originated such that there are effectively twice as many data points in the critical range of the Wolfe dataset than in the Misof dataset. Of 36 Wolfe calibration points above 100 Mya, 20 (55%) are in the range 200–400 Mya vs. five of 22 (23%) of those of Misof; and 13 of 22 (60%) of Misof’s calibrations above 100 Mya are under 200 Mya vs. 16 of 36 (38%) of Wolfe’s (Fig. 10). Counter-intuitively, the dataset more skewed towards younger calibration points may have yielded artificially old divergence time estimates that appear even more pronounced under uniform prior settings. When *Montagna et al.* (2019) re-calibrated the *Misof et al.* phylogeny by adding eight additional mid-Triassic (240 Mya) fossil calibrations, they recovered an even greater age for Pterygota, pushing the estimate back by 27 Myr, from 406.5 to 433.5 Mya. *Montagna et al.* explained this result with the idea that insect wings originated during an interval of high pO₂ (which facilitates flight; *Dudley, 2000*), and recent reconstructions of Palaeozoic pO₂ tentatively identified a peak from ~440 to 410 Mya (*Schachat et al.*, 2018).

Whereas relaxed-clock methods are increasingly able to accommodate evolutionary rates that vary among lineages and among regions of the genome (‘site’ and ‘lineage’ effects), temporal variability in evolutionary rates (‘epoch effects’ *sensu Lee & Ho, 2016*) remains far more difficult if not impossible to detect, much less control for (*Ho, 2020*). Epoch effects are especially intractable because they do not necessarily cause patterns of genomic diversity that violate the null predictions based on assumptions of rate homogeneity (*Lee & Ho, 2016*).

*Beaulieu et al.* (2015) suggest that the inability to accommodate rate heterogeneity accurately has resulted in overestimates of divergence times among angiosperms, specifically by assuming rate homogeneity for groups that underwent high rates of diversification and molecular evolution early in their evolutionary history followed by a decrease in both rates — something plausible but inherently untestable. *Budd & Mann* (2020) suggested that hyperdiverse groups may be inherently vulnerable to age overestimates based on certain analyses. Drawing on insight from birth–death modelling, they suggested that, because many groups become extinct within tens to hundreds of millions of years of their origin, an early burst of diversification may account for the persistence of many hyperdiverse groups throughout the Phanerozoic (*Phillimore & Price, 2008*). ( Whereas high rates of molecular evolution and high rates of net speciation are distinct phenomena, various authors have found them to be correlated; *Duchene et al.*, 2017; *Hua & Bromham, 2017; Bromham, 2020*.)

In the absence of molecular or palaeontological tests for epoch effects, the increased rates of molecular evolution that may accompany heightened rates of net diversification may predispose current methods for using fossil-calibrated phylogenetic data to overestimate ages of groups that experienced an early burst of evolution. Consequently, despite the increasing evidence of the ubiquity of initial bursts of diversification (*McPeek, 2008; Phillimore & Price, 2008*).
CONCLUSIONS

Winged insects are hardly the only group for which age estimates drastically pre-date their appearance in the fossil record. However, the discrepancy between the fossil record and various recent fossil-calibrated phylogenetic studies is particularly noteworthy in the case of winged insects for at least two reasons. First is the length of the purported gap in their fossil record: 100 Myr is nearly a third of the duration of the known pterygote fossil record. Second, pterygotes overwhelm the fossil record of terrestrial arthropods as soon as they appear. One might expect the first winged insects to have been rare on the landscape briefly, but an extended fossil gap is less plausible given the transformative nature of insect flight, their high abundance following their appearance in the fossil record, and the available fossil data from related insect and arthropod groups. Because the potential explanations for winged insects’ absence from the early- to mid-Palaeozoic fossil record (minute body size, exclusion from lowland habitats) are no less applicable to other terrestrial arthropods, we suggest that the absence of Pterygota from Silurian–Mississippian deposits that contain arachnids, centipedes, Entognatha, and Apterygota constitutes meaningful evidence of their actual absence prior to the Carboniferous. The phylogenetic recalibration exercises presented here suggest that the discordance among age estimates for Pterygota may become less severe as fossil calibration datasets improve.

ACKNOWLEDGEMENTS

We thank Russell Garwood for comments that greatly improved the manuscript. This contribution was made possible by contributors to the Paleobiology Database. We thank Joanna Wolfe, Andrew Leslie, Mike D’Antonio, and Pulkit Singh for valuable discussion. S.R.S. was funded by the Chair’s Fellowship, Harriet Benson Fellowship, and Coleman F. Fung Graduate Research Fellowship at Stanford University. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA; USDA is an equal opportunity provider and employer. We have no conflicts of interest to declare.

DATA AVAILABILITY

All data used as input for our analyses are publicly available from the Paleobiology Database (paleobiodb.org) and from the supplemental information of Misof et al. (2014). The raw log files from the BEAST and PyRate analyses are available upon request.

REFERENCES


Riek EF, Kukalová-Peck J. 1984. A new interpretation of dragonfly wing venation based upon Early Upper
Carboniferous fossils from Argentina (Insecta: Odonatoidea) and basic character states in pterygote wings. Canadian Journal of Zoology 62: 1150–1166.


A zero in the third column indicates that tree priors without a calibration date were used for the pterygote node.

In the second column the prior setting for the analysis and the third column the calibration date for the pterygote node.

### List of analyses undertaken in this study

The first column gives the name of the fossil dataset, the second column the prior setting for the analysis and the third column the calibration date for the pterygote node.

- **Initial age** is the first sampled root age estimated for Pterygota, sampled when running PyRate in python.
- **Seed (R)** is the value entered into the set.seed function in R while producing the input dataset.
- **Seed (p)** is the random seed used while entering into the set.seed function in R while producing the input dataset.
- **Seed (p)** is the value entered into the set.seed function in R while producing the input dataset.

### Results of the PyRate analyses, which were run on ten replicate input datasets

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site. See additional information for full details.

---

**SUPPORTING INFORMATION**

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

**Table S1.** Results of the PyRate analyses, which were run on ten replicate input datasets. ‘Seed (R)’ is the value entered into the set.seed function in R while producing the input dataset. ‘Seed (p)’ is the random seed used when running PyRate in python. ‘Initial age’ is the first sampled root age estimated for Pterygota, sampled after 1000/20 000 000 iterations. We note whether each PyRate run reached convergence. For those that did, we note the proportion of iterations discarded as burn-in, the effective sample size (ESS) and the estimated age of Pterygota: the median value and upper and lower limits of the 95% highest posterior density interval.

**Table S2.** List of analyses undertaken in this study. The first column gives the name of the fossil dataset, the second column the prior setting for the analysis and the third column the calibration date for the pterygote node. A zero in the third column indicates that tree priors without a calibration date were used for the pterygote node.
Table S3. Age estimates for the origin of crown Pterygota derived from different calibration schemes using the Misof et al. (2014) dataset. The ‘Difference’ column denotes the difference between each age estimate and the age estimate published by Misof et al. (2014).

Figure S1. Estimates for the age of Pterygota (Mya) from the six PyRate analyses that reached convergence. The interval from 324 to 330 Mya is shaded in blue.

Figure S2. Age estimates for Pterygota based on 11 combinations of datasets and priors. Abbreviations: M = Misof et al. age estimate priors; W = Wolfe age estimate priors; lgn = lognormal prior settings (with uncorrelated relaxed clock); unif = uniform prior settings (with uncorrelated relaxed clock); 324, 331 and 324–450 refer to Mya settings for pterygote priors; strict = strict molecular clock setting.

Figure S3. Scatter plots of median values for the 85 gene regions in the study. Abbreviations: M = Misof et al. age estimate priors; W = Wolfe age estimate priors; lgn = lognormal prior settings (with uncorrelated relaxed clock); unif = uniform prior settings (with uncorrelated relaxed clock); 324, 331 and 324–450 refer to Mya settings for pterygote priors; strict = strict molecular clock setting.

Figure S4. Density distributions for user-defined lognormal priors, and lognormal posterior distributions for Misof datasets with pterygote calibration and without Pterygota calibration for the Diptera, Neuroptera, Holometabola and Pterygota nodes. Density is on the y-axis and divergence time is on the x-axis in all panels.

Figure S5. Density distributions for user-defined lognormal priors, and lognormal posterior distributions for Wolfe datasets with pterygote calibration and without Pterygota calibration for the Diptera, Neuroptera, Holometabola and Pterygota nodes. Density is on the y-axis and divergence time is on the x-axis in all panels.

Figure S6. Results of another version of the fossil simulation procedure, for which the initial sampling rate is 50% of the final sampling rate at the Permian/Triassic boundary, and in which occurrences were not clustered into Lagerstätten.

Figure S7. Results of another version of the fossil simulation procedure, for which the initial sampling rate is 50% of the final sampling rate at the Permian/Triassic boundary, and in which occurrences were clustered into Lagerstätten.

Figure S8. Results of another version of the fossil simulation procedure, for which the initial sampling rate is 10% of the final sampling rate at the Permian/Triassic boundary, and in which occurrences were clustered into Lagerstätten.

Figure S9. Results of another version of the fossil simulation procedure, for which the initial sampling rate is 2% of the final sampling rate at the Permian/Triassic boundary, and in which occurrences were not clustered into Lagerstätten.

Figure S10. Results of another version of the fossil simulation procedure, for which the initial sampling rate is 2% of the final sampling rate at the Permian/Triassic boundary, and in which occurrences were clustered into Lagerstätten.

Appendix S1. The fossils from the Wolfe et al. (2016) compendium used in our analyses.