PNAS RESEARCH ARTICLE | EVOLUTION

Revisiting the four Hexapoda classes: Protura as the sister group to all other hexapods

Shiyu Du[a,b], Erik Tihelka[b,c], Daoyuan Yu[d], Wan-Jun Chen[e], Yun Bu[f], Chenyang Cai[b], Michael S. Engel[g,h,i], Yun-Xia Luan[j,1], and Feng Zhang[a,1]

Affiliations are included on p. 7.

Edited by Zhonghe Zhou, Chinese Academy of Sciences, Beijing, China; received May 2, 2024; accepted August 21, 2024

Insects represent the most diverse animal group, yet previous phylogenetic analyses based on morphological and molecular data have failed to agree on the evolutionary relationships of early insects and their six-legged relatives (together constituting the clade Hexapoda). In particular, the phylogenetic positions of the three early-diverging hexapod lineages—the coneheads (Protura), springtails (Collembola), and two-pronged bristletails (Diplura)—have been debated for over a century, with alternative topologies implying drastically different scenarios of the evolution of the insect body plan and hexapod terrestrialization. We addressed this issue by sampling all hexapod orders and experimenting with a broad range of across-site compositional heterogeneous models designed to tackle ancient divergences. Our analyses support Protura as the earliest-diverging hexapod lineage ("Protura-sister") and Collembola as a sister group to Diplura, a clade corresponding to the original composition of Entognatha, and characterized by the shared possession of internal muscles in the antennal flagellum. The previously recognized 'Elliplura' hypothesis is recovered only under the site-homogeneous substitution models with partial supermatrices. Our cross-validation analysis shows that the site-heterogeneous CAT-GTR model, which recovers "Protura-sister," fits significantly better than homogeneous models. Furthermore, the morphologically unusual Protura are also supported as the earliest-diverging hexapod lineage by other lines of evidence, such as mitogenomes, comparative embryology, and sperm morphology, which produced results similar to those in this study. Our backbone phylogeny of hexapods will facilitate the exploration of the underpinnings of hexapod terrestrialization and megadiversity.

genome-scale phylogeny | Insecta | noninsect hexapods | Protura-sister

Significance

This study resolves longstanding debates over the early evolutionary relationships within Hexapoda by employing a broad range of phylogenetic models and comprehensive sampling across all hexapod orders. Our findings support the "Protura-sister" hypothesis [Protura + ((Diplura + Collembola) + Insecta)], positioning Protura as the earliest-diverging hexapod lineage, with significant implications for understanding the evolution of the insect body plan and the terrestrial adaptation of hexapods. These results, corroborated by multiple lines of evidence including mitogenomes and comparative embryology, provide a crucial framework for exploring hexapod terrestrialization and their remarkable biodiversity.

Insects represent the most prolific radiation in the animal kingdom, accounting for over half of all described metazoan species (1). Winged insects came to dominate most terrestrial ecosystems by the late Carboniferous, over 310 Mya [million years ago] (2). Partly due to their great antiquity, the origins of insect megadiversity remain elusive. Current hypotheses tie the radiation of insects to their geological age, diversification and natural extinction rates, critical anatomical innovations, ecosystem change, and/or dietary breadth (3–7). As the closest relatives of insects, the noninsect hexapods play a pivotal role in understanding the unparalleled evolutionary success of six-legged life (8, 9). These other groups comprise small-bodied, elusive terrestrial arthropods with pronounced specializations for a soil-dwelling lifestyle. Unlike insects, these noninsect hexapod clades (Protura, Collembola, and Diplura) account for <1% of animal diversity, with some 10,800 species described to date (10–12). These include the comparatively species-poor, blind, and pseudotetrapodous Protura (coneheads, who carry the forelegs raised as sensory structures giving them a tetrapodous stance), the similarly speciose Diplura (two-pronged bristletails), and the considerably more diverse Collembola (springtails) armed with a characteristic abdominal jumping apparatus that gives them their name (13). Together with insects, they constitute the clade, often classified as a superclass, Hexapoda (8, 14).

The availability of genome-scale datasets has helped settle numerous historical conundrums in insect phylogeny over the last two decades (8, 15, 16). The dawn of the phylogenomic era has confirmed the monophyly of Hexapoda and elucidated the group's closest relatives (8, 17, 18). While traditional morphological studies considered hexapods as close relatives of myriapods (19), molecular datasets have revealed that the group is nested within the "crustaceans", overwhelmingly recovering them as sister to the enigmatic clade Remipedia, which inhabits flooded coastal caves (8, 18, 20–22). These results backdate the origin of crown-group insects to the Silurian–Cambrian (8, 23, 24) and imply that hexapod diversification was preceded by a terrestrialization event (18), likely sometime during the Silurian or end-Ordovician. However, remipedes are quite distinct and possess numerous

Author contributions: S.D., Y.-X.L., and F.Z. designed research; S.D., E.T., D.Y., W.-J.C., Y.B., C.C., M.S.E., Y.-X.L., and F.Z. performed research; S.D., E.T., C.C., and F.Z. analyzed data; S.D. collected data; W.-J.C., Y.B., and Y.-X.L. transcriptome sequencing; C.C., M.S.E., Y.-X.L., and F.Z. review & edit the paper; and S.D. and E.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: yxluan@scnu.edu.cn or fzhang@njau.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2408775121/-/DCSupplemental.

Published September 19, 2024.

specializations for aquatic life, with considerable morphological differences remaining between them and modern hexapods, and likely between the common ancestor of Remipedia + Hexapoda and the eventually terrestrialization of the common ancestor of Hexapoda and prior to its subsequent cladogenetic events. The early evolution of hexapods thus remains veiled in mystery, not only because of the extreme scarcity of hexapod fossils before the Late Carboniferous (25) but also because relationships among the earliest-diverging hexapods have proven resistant to resolution, whether interrogated with morphological data (14, 26, 27), single-gene (28, 29), mitochondrial genome (30, 31), phylogenomic data (8, 17, 18, 32), as well as combined analyses (33–35). Recent studies are mostly split between favoring a clade of Protura + Collembola (the "Elliplura" hypothesis) (8), Protura + Diplura (the "Nonoculata" hypothesis) (17, 27, 36–39), or Diplura + Collembola (20, 40), and Diplura + Insecta (the "Cercophora" hypothesis) (41, 42). Earlier morphological studies have cautiously treated the noninsect hexapod clades as a single clade, "Entognatha" (14, 43), while others maintain that the noninsect hexapods form a grade to Insecta (44). Traditional morphological studies, conducted since the 19th century (45, 46), have been hampered by the rather extreme specializations of these early-diverging hexapods for life in the soil, and which has made the inference of homologous character-states challenging (41, 47). Molecular studies are complicated by the rarity and small size of many morphologically peculiar noninsect hexapod groups, which have so far been sampled only sparsely in phylogenomic studies. Moreover, the great antiquity of the divergence between the noninsect hexapods and crown-group insects represents a formidable challenge to conventional molecular phylogenetic methods, as ancient rapid divergences often induce phylogenetic artifacts such as long-branch attraction (LBA) (48, 49).

Here, we address the problem of insect origins and hexapod diversification by increasing taxon and gene sampling of overlooked groups. We sequenced the transcriptome for a proturan species, belonging to the genus *Sinentomon* (28–30), along with transcriptomes for two extremely scarce dipluran species. We employ a variety of analytical approaches to account for common sources of error in phylogenomics, interrogate the robustness of the results, and interpret them with respect to the origin of the insect body plan and hexapod terrestrialization.

Results

Genomic Sequencing and Supermatrix Assembly. We sequenced the transcriptome of the proturan *Sinentomon erythranum* (SRX480876; Fig. 1*B*), a member of the rare monogeneric family Sinentomidae endemic to eastern Asia. This group was not discovered until the 1960s (50), and its phylogenetic position has stirred much controversy given the proturan's unusual head morphology and sperm ultrastructure (51–53). An analysis of two ribosomal RNA genes recovered Sinentomidae as the earliest-diverging proturan lineage (28), albeit substantial

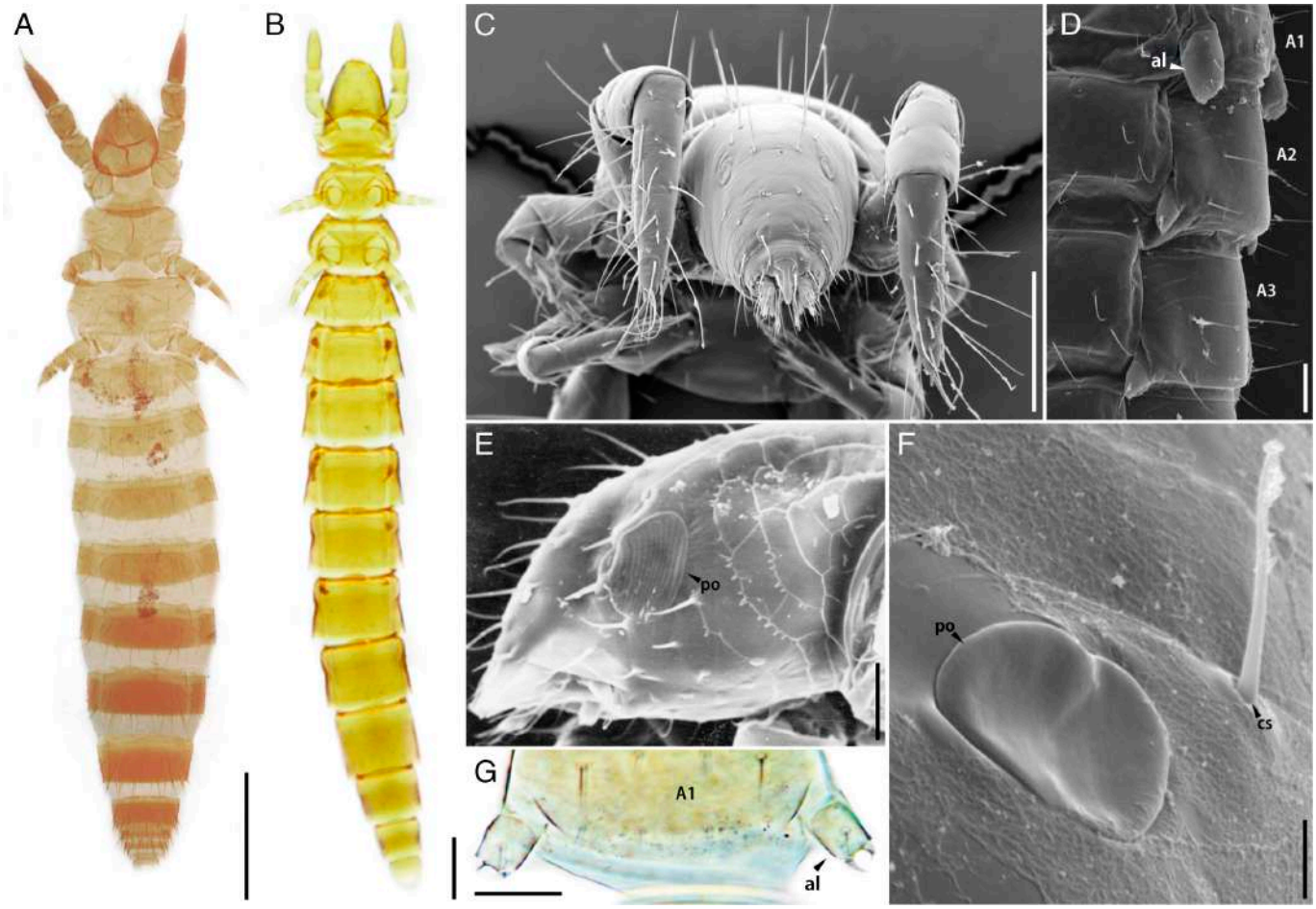

Fig. 1. Morphology of the proturans *Acerentomon microrhinus* (Acerentomidae) and *S. erythranum* (Sinentomidae). (*A*) Habitus view of *A. microrhinus* under reflected light. (*B*) Habitus view of *S. erythranum* under reflected light. (*C*) Scanning electron micrograph of *A. microrhinus* head and forelegs. (*D*) Scanning electron micrograph of the abdomen of *A. microrhinus* in lateral view. (*E*) Scanning electron micrograph of *S. erythranum* head in lateral view. (*F*) Detail of the pseudoculus of *A. microrhinus*. (*G*) Abdominal legs of *S. erythranum*. Abbreviations: A1–3: abdominal segments 1–3; al, abdominal legs; cs, cephalic seta; po, pseudoculus. [Scale bar, 5 μm (*G*); 10 μm (*F*); 20 μm (*C*–*E*); 50 μm (*A* and *B*).]

incongruence persists among studies (29, 30, 36, 54, 55). We furthermore additionally sequenced two transcriptomes belonging to the diplurans *Octostigma sinensis* (SRX3641158) and *Lepidocampa weberi* (SRX3641157), representing the superfamilies Projapygoidea and Campodeoidea, respectively. Projapygoids are a presumed evolutionary link between Campodeoidea and Japygoidea (56), but they are exceedingly rare and hard to collect for comparative studies. The interrelationships of three superfamilies and the monophyly of Diplura have been much debated. It has been suggested that diplurans may together be polyphyletic rather than a clade based on ovarian and spermatozoal characters (57, 58), albeit comparative embryological and molecular evidence so far overwhelmingly supports dipluran monophyly (8, 28, 31, 59).

We compiled genomic and transcriptomic data for 42 other hexapod species from NCBI (*SI Appendix* for details) with high Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness values plus nine aquatic "crustacean" clades (outgroups) recovered as close relatives of hexapods in previous studies (18, 20–22, 60). The inclusion of early-diverging dipluran and protura groups is particularly relevant, as previous studies have indicated that the hexapod tree is prone to LBA (16, 20, 31), which are exacerbated by limited taxon sampling (61). For detailed statistical information on each species, please refer to Dataset S1.

Our dataset comprised a total of 54 species, including nine additional outgroups. We extracted a total of 1,013 genes (loci) for subsequent phylogenetic analyses, based on the OrthoDB version 10 (62) of the Arthropoda database (n = 1,013) from BUSCO. Phylogenetic analyses were conducted using amino acid (AA) alignments to explore alternative sources of phylogenomic signal. Recognizing that multiple sequence alignments can be affected by highly divergent sites, often due to misinferred homologies and substitution saturation, which may compromise the integrity of phylogenetic tree construction (63), we implemented a rigorous preprocessing protocol. First, all alignments were curated with TAPER software to detect and mask the misaligned/aligned regions (64). To maximize the retention of phylogenetically informative sites, we further trimmed the alignments using ClipKIT, ensuring a careful balance between data preservation and error reduction (65). To address challenges related to compositional variation and consistency, we conducted the normalized Relative Compositional Frequency Variation (nRCFV) test (66, 67) along with the Stationary, Reversible, and Homogeneous (SRH) assumptions test (68). After conducting these two tests, 19 and 45 loci were removed, respectively, leaving a total of 949 loci. Paralogous genes, which are known to introduce analytical complexities (69, 70), were identified and excluded from the alignments using TreeShrink (BUSCO ids and names of the putatively spurious sequence after spurious homolog identification by using TreeShrink are listed in Dataset S2). Through the above filtering factors, an initial supermatrix, Matrix 1, was created. This supermatrix containing 455,036 sites across 949 loci (Table 1). However, upon sensitivity tests for ten possible factors (gene properties) affecting reconstruction results (the selection process and evaluation criteria for each factor and threshold are detailed in *SI Appendix*, Supplementary Appendix A), we observed variability in analysis outcomes influenced by the evolutionary rate of sequences (proxied by average pairwise identity, API) (71) and the average bipartition support (ABS, representing the strength of phylogenetic signal) (72) metrics. We created two supermatrices, Matrix 2 and Matrix 3, by applying stricter criteria on API and ABS values. Matrix 2, consisting of 267,183 sites and 475 loci, was defined by API values below 0.6 (genes with relatively higher evolutionary rate), while Matrix 3, containing 335,767 sites and 554 loci, was defined by ABS values exceeding 70 (genes with relatively stronger phylogenetic signals). To address discrepancies between concatenation and coalescent-based phylogenies, we developed Matrix 4 (296,560 sites, 586 loci). This supermatrix was derived from the full set of 1,013 loci by selectively excluding "inconsistent" genes with conflicting phylogenetic signals (i.e., those with gene-wise phylogenetic signal (ΔGLS) > 0, or gene-wise quartet scores (ΔGQS) < 0; *SI Appendix* for details) (73). Further refinement resulted in Matrix 5, which combined genes from Matrix 4 that met strict API and ABS thresholds (API < 0.6 and ABS > 70), producing a matrix of 199,586 sites across 305 loci. Finally, Matrix 6, containing 164,081 sites, was created by trimming Matrix 1 using BMGE, resulting in a refined set of 949 loci. The creation of this "minimal" supermatrix aims to minimize phylogeny reconstruction artifacts caused by compositional heterogeneity (74) while also reducing computational burden (for details of each supermatrix, Table 1). A detailed description of our filtering strategy and supermatrix generation is provided in *SI Appendix*, Supplementary Appendix A.

Hexapod Phylogeny. All our phylogenomic analyses recovered strong support for the monophyly of Collembola, Protura, Diplura, and Insecta, respectively (Bayesian Posterior Probabilities (BPP) = 1, SH-aLRT/UFBoot2 = 100/100, and wASTRAL bootstraps = 1; Fig. 2). A total of five weighted-ASTRAL (wASTRAL) trees, 29 Maximum Likelihood (ML) trees, and three Bayesian Inference (BI) trees were inferred from the six supermatrices (Dataset S3 and *SI Appendix*, Supplementary Appendix B) to test the effect of the substitution model on the recovered topology. Trees based on different supermatrices and inference models were congruent at most nodes (Fig. 2) but resulted in three different topological hypotheses (H1–3) about the relationships of the early-diverging hexapod clades (Fig. 3 *A*–*C*). Hypothesis 1 supported the placement of Collembola as sister to the remaining hexapods [H1: "Collembola-sister" hypothesis, i.e., Collembola + (Protura + (Diplura + Insecta))]. Under the second hypothesis [H2: "Elliplura" hypothesis, i.e., (Collembola + Protura) + (Diplura + Insecta)], Collembola and Protura formed a monophyletic group as sister group to Diplura + Insecta [= the Pleomerentoma concept of Krause and Wolfe (75)], corresponding to the "Elliplura" hypothesis (8). Protura was inferred as the sister group to the remaining three hexapod

Table 1. Summary of USCO amino acid supermatrices used for phylogenetic analyses

Supermatrix	Average missing taxa per locus (%)	No. of loci	No. of sites	Missing sites (%)	Average locus length
Matrix 1	12.5	949	455,036	35.7	479.49
Matrix 2	13.87	475	267,183	37.08	562.49
Matrix 3	12.61	554	335,767	36.36	606.08
Matrix 4	12.6	586	296,560	37.4	506.08
Matrix 5	12.96	305	199,586	38.56	654.38
Matrix 6	—	949	164,081	18.01	172.90

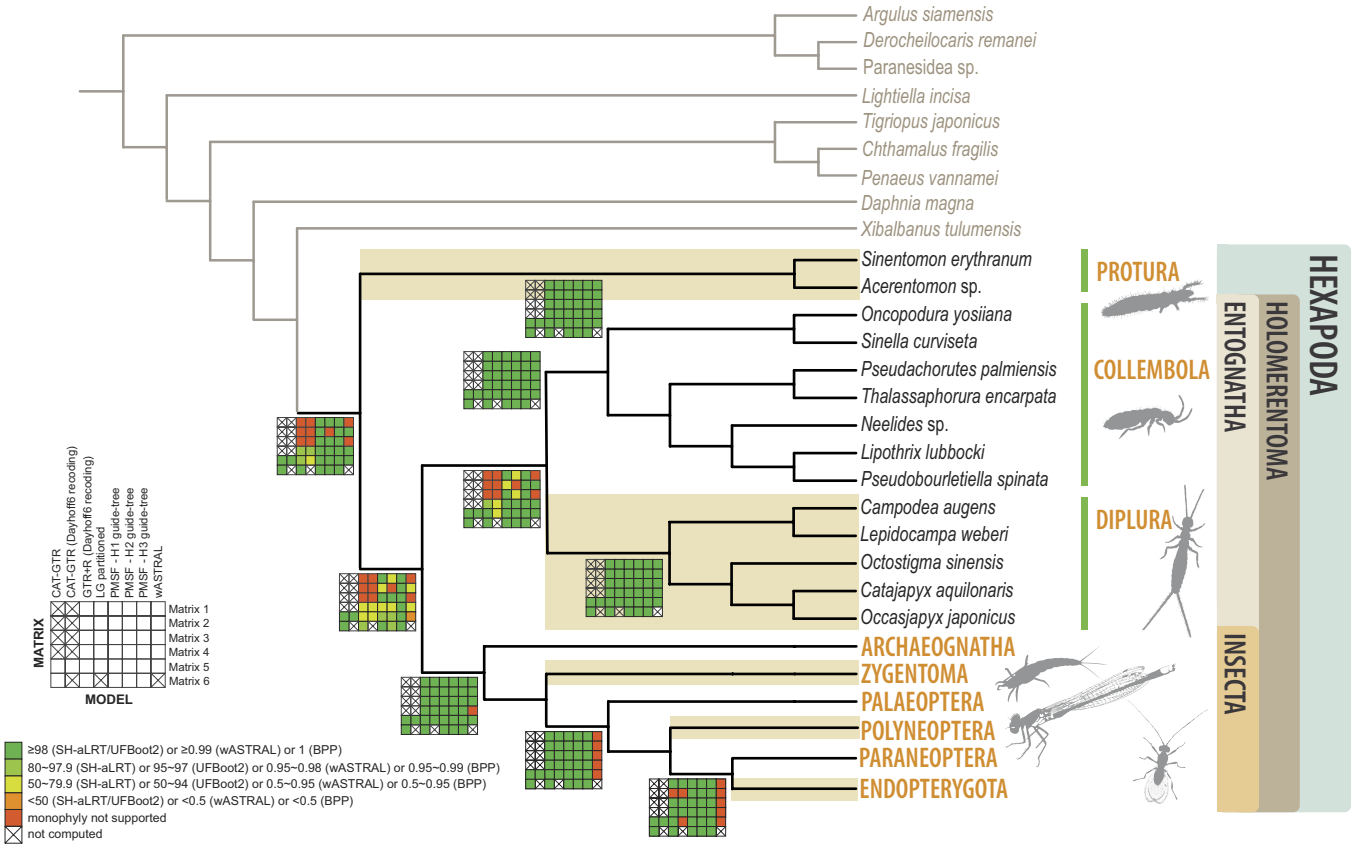

Fig. 2. Phylogeny of the noninsect hexapods. Main topology inferred from the Matrix 6 using the Bayesian across-site compositional heterogeneity CAT-GTR model implemented in PhyloBayes. Node supports from all analyses are indicated by the colored squares (The node supports of each phylogenetic tree is shown in *SI Appendix*, Supplementary Appendix B). Only the lowest support values are shown when different matrices or different models produced conflicting results. [H1.guide: Collembola + (Protura + (Diplura + Insecta)); H2.guide: (Collembola + Protura) + (Diplura + Insecta); H3.guide: Protura + ((Collembola + Diplura) + Insecta)]. Collembola + Diplura = original circumscription of Entognatha; Entognatha + Insecta = Holomerentoma (a.k.a. Holomera).

groups in the third hypothesis [H3: "Protura-sister" hypothesis, i.e., Protura + ((Collembola + Diplura) + Insecta)], or the Protura + the Holomerentoma of Prell (76).

The most complex site-heterogeneous models, PMSF(C60) and CAT-GTR, supported the "Protura-sister" hypothesis when all supermatrices were analyzed (except for the Matrix 2, which

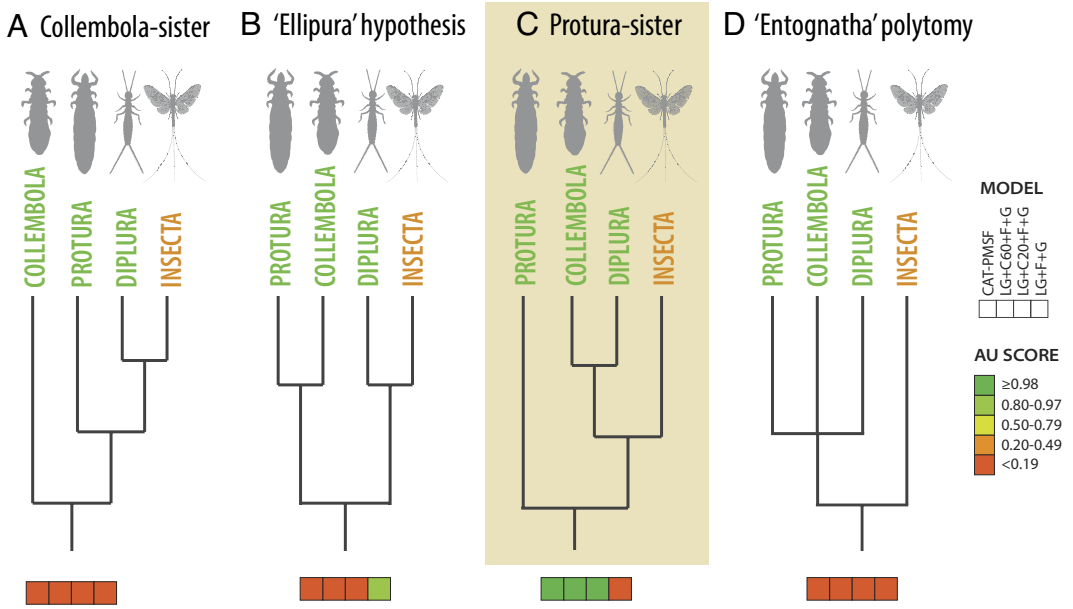

Fig. 3. Four different topological hypotheses analyzed in this study. (*A*) Collembola + (Protura + (Diplura + Insecta)) ("Collembola-sister" hypothesis). (*B*) (Collembola + Protura) + (Diplura + Insecta) (the "Elliplura" hypothesis) (8). (*C*) Protura + ((Collembola + Diplura) + Insecta) ("Protura-sister" hypothesis). (*D*) unresolved polytomy corresponding to the "Entognatha" hypothesis (14, 44). Topology test *P*-value (AU score) from Matrix 6 under CAT-PMSF, LG+C60+F+G, LG+C20+F+G, and LG+F+G models are indicated by the colored squares.

supported the "Elliplura" hypothesis (8) under the PMSF(C60) model with H2 topology as the initial guide tree). Under this topology, Protura were the sister group to Diplura + Collembola and the remaining hexapods. Our model comparison based on leave-one-out cross-validation (LOO-CV) and widely applicable information criterion (wAIC) for Matrix 6 provided high resolution on the performance of different models, offering a reliable basis for selecting the best topology (77). In our results, the site-heterogeneous CAT-GTR model fitted the dataset better than the site-homogeneous LG model (ΔCV = −35.4127 + 37.5290 = 2.1163, and ΔwAIC = −35.3988 + 37.5290 = 2.1302; Dataset S4). Partitioned analysis supported the "Elliplura" hypothesis (8) with the Matrix 1 to Matrix 3, and "Protura-sister" with the Matrix 4 and Matrix 5, while multispecies coalescent analyses recovered all three hypotheses, albeit some nodes were poorly supported (Dataset S3). When subjected to Dayhoff 6-state recoding and analyzed with GTR+R model, the results were consistent with the partitioned analyses, and the "Protura-sister" hypothesis is also supported by this analysis based on Matrix 6. In addition, gene concordance factors (gCF) and site concordance factors (sCF) were used to gain a deeper understanding of how well different genes and sites support the different hypotheses (*SI Appendix*, Supplementary Appendix B). For most branches in all three topologies, the gCF values are lower than the sCF values, suggesting that the focal sites that support these topologies are scattered across different genes.

Evaluating Alternative Hypotheses and Phylogenetic Support. We conducted several tests to discriminate between the four hypotheses (three hypotheses obtained in this study and 'Entognatha' hypothesis from the previous studies (14, 44); Fig. 3) of hexapod relationships. Topology tests were conducted on all six supermatrices with the C60+F+R model using the Approximately Unbiased (AU), Weighted Kishino–Hasegawa (WKH), and Weighted Shimodaira–Hasegawa (WSH) tests. Three supermatrices (Matrix 4 to Matrix 6) supported the "Protura-sister" hypothesis (P-value < 0.05). Three other supermatrices (Matrix 1 to Matrix 3) supported the "Elliplura" hypothesis without significance (Dataset S5). The Matrix 6 was furthermore subjected to AU test with the more computationally demanding models CAT-PMSF, LG+C60+F+G, LG+C20+F+G, and LG+F+G; these unequivocally yielded support for the "Protura-sister" hypothesis (Dataset S5 and Fig. 3).

We detected potentially confounding signals using four-cluster likelihood mapping (FcLM) analysis with all six supermatrices and evaluated which hypothesis (unrooted trees) was predominantly supported by these quartets: T1 (*SI Appendix*, Fig. S1*G*), (Collembola + Diplura) and (Protura + Insecta); T2 (*SI Appendix*, Fig. S1*H*), (Collembola + Insecta) and (Diplura + Protura); T3 (*SI Appendix*, Fig. S1*I*), (Collembola + Protura) and (Protura + Insecta). The majority of quartets placing Collembola plus Diplura were sister group to a clade comprising Protura + Insecta, i.e., unrooted topology T1 was favored, corresponding to the "Protura-sister" hypothesis, i.e., H3 (*SI Appendix*, Fig. S1*G*), with strong support (FcLM range from 64.4 to 84.6%; *SI Appendix*, Fig. S1 *A*–*F*).

In addition, we subjected every gene from the five supermatrices (except for the Matrix 6) to AU, WKH, and WSH tests to compute the distribution of gene tree supports (i.e., the gene-wise likelihood scores). In three supermatrices (Matrix 1 to Matrix 3), the "Elliplura" hypothesis had the highest number of gene tree supports, and the "Collembola-sister" hypothesis had lowest. However, the "consistent" genes (Matrix 4 and Matrix 5) lent more support to the "Protura-sister" hypothesis than the "Collembola-sister" scenario (Dataset S6 and *SI Appendix*, Fig. S2*A*). We furthermore compared the site-wise likelihood scores for the three hypothesized topologies. For Matrix 3 and Matrix 4, a distinctly greater number of sites supported "Elliplura" than any of the other two hypotheses (Dataset S7; *SI Appendix*, Fig. S2*B*). The "Elliplura" and "Protura-sister" hypotheses had a similar number of sites support in the Matrix 2 (*SI Appendix*, Fig. S2*B*).

Discussion

Molecular and Morphological Congruence. As with many other ancient radiations (16), molecular phylogenetic studies have found it challenging to elucidate the relationships of the noninsect hexapod clades, which may have diverged as early as the Cambrian–Silurian (8, 78), although a Cambrian origin is less likely given that arthropods had not yet ventured onto land. Expanding the taxon sampling of noninsect hexapods, including sequencing the transcriptome of the enigmatic *Sinentomon*, *Octostigma*, and *Lepidocampa*, enabled us to explore various sources of phylogenomic signal and mitigate common artifacts at the base of the hexapod tree of life, which has been plagued by topological uncertainty (9, 16). We recovered three alternative topologies, corresponding to long-standing competing hypotheses regarding insect origins (8, 38, 79, 80) (Dataset S3 and Fig. 3 *A*–*C*). Under the GTR+R (with Dayhoff6 recoding alignments) and partitioned ML models, the "Elliplura" hypothesis was supported, as in Misof et al. (8), along with the "Protura-sister" hypothesis. The multispecies coalescent analyses recovered all three topologies (Dataset S3). The site-heterogeneous PMSF(C60) model, as well as CAT-GTR model, provide support for Protura as the first-diverging lineage of hexapods. Additionally, based on Matrix 2, the PMSF analysis, with H2 topology as the initial guide tree, supports the "Elliplura" hypothesis (Dataset S3). The question is then why similar analyses give different results and how we should interpret variation in results obtained from different analyses. The first important insights pertain to model fit. Cross-validation (CV) is a reliable approach to assessing the fit of models to the data (77). The general idea is to split the dataset into two subsets, using one subset for training the model and then evaluating the fit of the model over the remaining subset. In the context of Bayesian inference, a natural procedure to implement CV is to average the validation likelihood over the training posterior distribution. The resulting score is then log transformed and averaged over multiple random splits of the original dataset into training and validation sets. In LOO-CV of PhyloBayes, each observation is taken in turn and set aside for validation, using the n − 1 remaining observations to train the model (77). We used LOO-CV to compare site-heterogeneous model (CAT-GTR) and the site-homogeneous model (LG) on the Matrix 6. Our analysis revealed that CAT-GTR provided a better fit to the dataset compared to LG (Dataset S4). Therefore, LOO-CV supports the hypothesis that the compositionally site-heterogeneous model CAT-GTR provides a better fit than the site-homogeneous models with LG. Other topologies were supported by less well-fitting models, and by partitioned analyses, the latter of which has been shown to fit empirical data significantly less than approaches that consider heterogeneity at the site level, in most cases (16). The second insight pertains to topology tests. We compared the four topologies on all supermatrices under C60+F+R model using the AU, WKH, and WSH tests. Three supermatrices supported the "Protura-sister" hypothesis with strong confidence, and the others supported the "Elliplura" hypothesis with no confidence. When topology tests were conducted using the site-heterogeneous models CAT-PMSF, LG+C60+F+G, and LG+C20+F+G, the "Protura-sister" hypothesis was supported unequivocally. Meanwhile, the FcLM analyses also supported the "Protura-sister" hypothesis. Additionally, we further found that evolutionary rates of genes and "inconsistent" genes can influence the tree reconstruction results. For instance, when using Matrix

2, Matrix 4, or Matrix 5, the "Protura-sister" hypothesis can be recovered even without employing complex site-heterogeneous models. Lower API values, i.e., higher evolutionary rate, indicate lower sequence conservation and more variable sites, which in turn provide stronger phylogenetic signals. Therefore, it is evident that selecting appropriate loci is crucial when addressing the issues in our study. These analyses suggest that we could recover "Protura-sister" hypothesis over the much broader substitution model and topology test.

Proturans have long been considered morphologically divergent hexapods, leading some early authors to argue that they may not be related to hexapods at all (81). The status of proturans as the earliest-diverging hexapods is further supported by a suite of morphological characters shared with myriapods and crustaceans. In proturans, the first three abdominal segments retain segmented or unsegmented vestigial appendages (Fig. 1 *D* and *G*: al) (82), a plesiomorphy shared with most myriapods and crustaceans where all trunk segments are equipped with a pair of segmented limbs (83). These abdominal appendages have been reduced to unsegmented stubs or have been lost altogether in most hexapods (84). A further plesiomorphic character proturans share with myriapods and crustaceans (85), but not other hexapods, is their anamorphic postembryonic development (anamorphic development may be plesiomorphic), but it is highly variable in groups like myriapods, where epimorphic development is common (e.g., Scolopendromorpha, Geophilomorpha). That is, proturans emerge from the egg with nine abdominal segments and add a segment with the first molt and two more segments with the second molt, which results in 12 segments in the adult abdomen, including a distinct telson. The proturan embryonic membrane possesses the ability to differentiate into the dorsal body wall, a feature shared with aquatic "crustaceans" and myriapods, but not with other hexapods (86). A further potential plesiomorphy of proturans may be the single claw (pretarsus) on each leg, while other hexapods have paired pretarsal claws (87). Proturans have no antennae, and they walk on four legs with the front two repurposed as raised sensory structures, which diverges strongly from other hexapods (88), but assuredly a unique apomorphy for their unusual life history. They have no eyes, just pseudoculi, whose homology remains uncertain, and which probably only senses light without forming images (Fig. 1 *E* and *F*: po) (89), another autapomorphy for the clade. Flagellate spermatozoa in proturans have a variable axonemal pattern, but a common, distinctive feature is the absence of central microtubules (90). Proturans moreover possess a simplified or absent tracheal system unlike any other hexapods (91); when tracheae are present at all, they are present as only two pairs of spiracles on the thorax (92), a remarkable apomorphic reduction of the tracheal system unique to the lineage.

Characters supporting a Collembola + Diplura as a clade, i.e., the original circumscription for Entognatha, which did not include Protura (93), are fewer but include a similar process of blastokinesis (94), and each antennal division with intrinsic musculature, whereas in the Insecta only the antennal scape possesses intrinsic muscles (27). A close relationship between the two groups is moreover supported by some analyses of mitochondrial protein-coding genes (59) and genomic datasets under heterogeneous models (20). We herein propose to return Entognatha to its original conception by removing Protura and returning it to a clade with only Collembola and Diplura. This narrowed Entognatha thus restored is a well-founded monophyletic group sister to Insecta.

Implications for Hexapod Terrestrialization. The terrestrialization of hexapods, the most cryptic episode of the clade's evolutionary history, has long remained shrouded in mystery, but equally attracted interest due to its importance for delimiting the ground plan of the ancestral hexapod. Unfortunately, there is no fossil evidence supporting the idea that Protura is the earliest-diverging lineage of hexapods. Consequently, utilizing molecular data, especially the omics data, has remained the only effective means to address this uncertainty. The resolution of proturans as the earliest-diverging hexapods enables the reconstruction of the last hexapod ancestor and permits the sequence of character evolution to be traced and homologies to be established. A lasting contention in understanding hexapod terrestrialization is whether adaptations for life on land were acquired in a stepwise fashion or whether the last common ancestor of Hexapoda already possessed a complex respiratory, reproductive, and sensory systems (95, 96). In other words, were there a series of serial stem groups progressively more "terrestrialized" and progressively closer to crown-Hexapoda? Some molecular and morphological studies over the past decade have argued that given their unusual organ systems, some proturan characters of the reproductive and respiratory systems may not be homologous with other hexapods and instead represent an independent ancient lineage (41, 95, 97). Despite the highly unusual morphology of the proturans, evidently sculpted by their soil habitats, several inferences can be drawn.

Our results suggest that the last common ancestor of the hexapods was terrestrial, contrary to some earlier hypotheses that suggested possible aquatic or semiaquatic modes of life in early hexapods (52, 98). The hexapod colonization of land was facilitated by a sequence of morphological innovations. Crucial to survival on land is respiration in air, which in insects is facilitated by a well-developed tracheal system. The tracheal system is present in some proturans, collembolans, and all diplurans, leading to debates about its homology (30, 95). In light of the present results, it is plausible that the tracheal system was present in the ancestor of Diplura + Collembola and may have been present in the ancestor of Hexapoda, with subsequent apomorphic reduction in Protura as evidenced by the vestigial tracheal elements sometimes present, although the latter remains somewhat contentious but is a more parsimonious interpretation of the pattern of character-states relative to Remipedia and other "crustaceans". Terrestrial arthropods, namely insects, myriapods, and arachnids, share osmoregulation and excretion facilitated by the Malpighian tubule system. The "Protura-sister" hypothesis implies that the Malpighian tubules were secondarily reduced to small papillae in Protura and Diplura, while they have been secondarily lost altogether in Collembola (99). All in all, our results imply that the last common ancestor of Hexapoda already possessed many of the key adaptations for life on land.

Proturans are notable among the other early-diverging hexapod linages for their highly specialized mode of life. An almost obligate ectomycorrhizal feeding (100), or at least strong preference for fungal feeding (98, 101, 102), has been documented across proturans in laboratory and field experiments. All proturans possess specialized sucking mouthparts and appear to feed on hyphal cytoplasm. Meanwhile, diplurans and collembolans are generalist omnivores, feeding on roots, fungal hyphae, spores, and decaying organic matter or may sometimes be predaceous (12, 103, 104). The recovery of proturans as the earliest-diverging hexapod clade poses the possibility that specialized ectomycorrhizal mycelia represent a plesiomorphic condition also present in the last hexapod ancestor. Fossil mycorrhizal fungi are known from the Early Devonian (102), putative terrestrial fungi fossils occur from the Silurian onward (105), and molecular clock studies suggest that they may have been present as early as the Cambrian (103). In any case, early terrestrial ecosystems before the development of tall woody vegetation were likely dominated by microbial mats and fungi (106), highlighting a possible food source for early hexapods that may have facilitated their invasion of land. Further understanding of the ancestral hexapod body plan relies on a better understanding of the extant

early-diverging hexapod lineages such as Protura and Diplura, for which many character systems are poorly known. Resolution of relationships among noninsect hexapods will further facilitate ground plan comparisons with other arthropod lineages and the reinterpretation of controversial fossils (107) that may help trace the transition of marine pancrustaceans to the terrestrial realm and ultimately the earliest of fossil hexapods.

Materials and Methods

Transcriptomes from two rare diplurans (*O. sinensis* and *L. weberi*) and one proturan (*S. erythranum*) were sequenced. Forty-five other hexapod species and nine crustacean outgroups were downloaded from NCBI to deepen our understanding of phylogenomic relationships among hexapods (Dataset S1). Universal single-copy orthologs (USCOs) were extracted from all samples, aligned multiple sequences, trimmed poorly aligned regions, filtered loci, and used to construct six distinct supermatrices. We employed a variety of phylogenetic methods to tackle common systematic errors, including multispecies coalescent and concatenation-based models. Our study further utilized multiple analytical methods such as topology tests, four-cluster likelihood mapping analyses, site-wise and gene-wise likelihood analyses, and model comparison, all aimed at accurately delineating relationships among noninsect hexapods. See also *SI Appendix* for details.

Data, Materials, and Software Availability. All data and materials are available from the public repository on GitHub (108). The custom scripts, based on Du et al. (109), can also be found on GitHub (https://github.com/xtmtd/Phylogenomics/tree/main/basal_hexapods/scripts). All supplementary appendices are available on GitHub (https://github.com/xtmtd/Phylogenomics/tree/main/basal_hexapods/Supplementary_material) (108). All matrices and tree files are available on GitHub (https://github.com/xtmtd/Phylogenomics/tree/main/basal_hexapods/matrices and https://github.com/xtmtd/Phylogenomics/tree/main/basal_hexapods/treefiles) (108). NCBI accession numbers are provided in Dataset S1.

ACKNOWLEDGMENTS. We extend our gratitude to three anonymous reviewers whose insightful comments have added substantial value to our work. The research contributions of F.Z. were supported by the National Natural Science Foundation of China (32270470, 31970434) and the National Science and Technology Fundamental Resources Investigation Program of China (2018FY100300). Y.-X.L.'s research efforts were underwritten by the National Natural Science Foundation of China (32170425, 31970438). Financial support of C.C. was provided by the National Natural Science Foundation of China (42222201). The work of M.S.E. received financial backing from CONCYTEC via the PROCIENCIA program, specifically through the "Interinstitutional Alliances for Doctorate Programs" initiative (contract PE501084299-2023-PROCIENCIA-BM).

Author affiliations: [a]Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China; [b]State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, and Centre for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Nanjing 210008, China; [c]Department of Earth Sciences, University of Cambridge, Cambridge CB2 1TN, United Kingdom; [d]Department of Ecology, College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China; [e]Mammoth (Shenzhen) Education Technology Co. Ltd, Shenzhen 518000, China; [f]Natural History Research Center, Shanghai Natural History Museum, Shanghai Science & Technology Museum, Shanghai 200041, China; [g]Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024; [h]Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima 15081, Perú; [i]Departamento de Entomología, Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Lima 15081, Perú; and [j]Guangdong Provincial Key Laboratory of Insect Development Biology and Applied Technology, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China

1. R. M. May, Biological diversity: How many species are there? *Nature* **324**, 514–515 (1986).
2. M. S. Engel, Insect evolution. *Curr. Biol.* **25**, R868–R872 (2015).
3. P. J. Mayhew, Why are there so many insect species? Perspectives from fossils and phylogenies. *Biol. Rev.* **82**, 425–454 (2007).
4. P. J. Mayhew, Explaining global insect species richness: Lessons from a decade of macroevolutionary entomology. *Entomol. Exp. Appl.* **166**, 225–250 (2018).
5. S. B. Carroll, Chance and necessity: The evolution of morphological complexity and diversity. *Nature* **409**, 1102–1109 (2001).
6. J. J. Wiens, R. T. Lapoint, N. K. Whiteman, Herbivory increases diversification across insect clades. *Nat. Commun.* **6**, 1–7 (2015).
7. D. B. Nicholson, A. J. Ross, P. J. Mayhew, Fossil evidence for key innovations in the evolution of insect diversity. *Proc. Roy. Soc. B.* **281**, 20141823 (2014).
8. B. Misof *et al.*, Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
9. R. G. Beutel, M. I. Yavorskaya, Y. Mashimo, M. Fukui, K. Meusemann, The phylogeny of Hexapoda (Arthropoda) and the evolution of megadiversity. *Proc. Arthropod. Embryol. Soc. Jpn.* **51**, 1–15 (2017).
10. A. Szeptycki, Catalog of the world Protura. *Acta Zool. Cracov. Ser. B* **50**, 1–210 (2007).
11. A. Potapov *et al.*, Towards a global synthesis of Collembola knowledge – challenges and potential solutions. *Soil Organisms* **92**, 161–188 (2020).
12. A. Sendra, A. Jiménez-Valverde, J. Selfa, A. S. P. S. Reboleira, Diversity, ecology, distribution and biogeography of Diplura. *Insect Conserv. Divers.* **14**, 415–425 (2021).
13. D. Grimaldi, M. S. Engel, *Evolution of the Insects* (Cambridge University Press, ed. 1, 2005).
14. W. Hennig *Die Stammesgeschichte der Insekten* (Waldemar Kramer, ed. 1, 1969).
15. M. D. Trautwein, B. M. Wiegmann, R. Beutel, K. M. Kjer, D. K. Yeates, Advances in insect phylogeny at the dawn of the postgenomic era. *Annu. Rev. Entomol.* **57**, 449–468 (2012).
16. E. Tihelka *et al.*, The evolution of insect biodiversity. *Curr. Biol.* **31**, R1299–R1311 (2021).
17. K. Meusemann *et al.*, A phylogenomic approach to resolve the arthropod tree of life. *Mol. Biol. Evol.* **27**, 2451–2464 (2010).
18. B. M. von Reumont *et al.*, Pancrustacean phylogeny in the light of new phylogenomic data: Support for Remipedia as the possible sister group of Hexapoda. *Mol. Biol. Evol.* **29**, 1031–1045 (2012).
19. M. J. Telford, R. H. Thomas, Demise of the atelocerata? *Nature* **376**, 123 (1995).
20. J. Lozano-Fernandez *et al.*, Pancrustacean evolution illuminated by taxon-rich genomic-scale data sets with an expanded remipede sampling. *Genome Biol. Evol.* **11**, 2055–2070 (2019).
21. M. Schwentner, D. J. Combosch, J. P. Nelson, G. Giribet, A phylogenomic solution to the origin of insects by resolving crustacean–hexapod relationships. *Curr. Biol.* **27**, 1818–1824 (2017).
22. J. P. Bernot *et al.*, Major revisions in Pancrustacean phylogeny and evidence of sensitivity to taxon sampling. *Mol. Biol. Evol.* **40**, msad175 (2023).
23. R. J. Howard *et al.*, The Ediacaran origin of Ecdysozoa: Integrating fossil and phylogenomic data. *J. Geol. Soc.* **179**, jgs2021-107 (2022).
24. G. Giribet, G. D. Edgecombe, The phylogeny and evolutionary history of arthropods. *Curr. Biol.* **29**, R592–R602 (2019).
25. W. A. Shear, An insect to fill the gap. *Nature* **488**, 34–35 (2012).
26. N. P. Kristensen, Phylogeny of insect orders. *Annu. Rev. Entomol.* **26**, 135–157 (1981).
27. C. Bitsch, J. Bitsch, Phylogenetic relationships of basal hexapods among the mandibulate arthropods: A cladistic analysis based on comparative morphological characters. *Zool. Scripta* **33**, 511–550 (2004).
28. Y.-X. Luan, J. M. Mallatt, R. Xie, Y. Yang, W. Yin, The phylogenetic positions of three basal-hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. *Mol. Biol. Evol.* **22**, 1579–1592 (2005).
29. Y. Gao, Y. Bu, Y.-X. Luan, Phylogenetic relationships of basal hexapods reconstructed from nearly complete 18S and 28S rRNA gene sequences. *Zool. Sci.* **25**, 1139–1145 (2008).
30. A. Carapelli *et al.*, Going deeper into high and low phylogenetic relationships of Protura. *Genes (Basel)* **10**, 292 (2019).
31. W.-J. Chen *et al.*, The mitochondrial genome of *Sinentomon erythranum* (Arthropoda: Hexapoda: Protura): An example of highly divergent evolution. *BMC Evol. Biol.* **11**, 246 (2011).
32. E. Dell'Ampio *et al.*, Decisive data sets in phylogenomics: Lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol. Biol. Evol.* **31**, 239–249 (2014).
33. J. Bitsch, C. Bitsch, T. Bourgoin, C. D'Haese, The phylogenetic position of early hexapod lineages: Morphological data contradict molecular data. *Syst. Entomol.* **29**, 433–440 (2004).
34. G. Giribet, G. D. Edgecombe, J. M. Carpenter, C. A. D'Haese, W. C. Wheeler, Is Ellipura monophyletic? A combined analysis of basal hexapod relationships with emphasis on the origin of insects. *Org. Divers. Evol.* **4**, 319–340 (2004).
35. W. C. Wheeler, M. Whiting, Q. D. Wheeler, J. M. Carpenter, The Phylogeny of the Extant Hexapod Orders. *Cladistics* **17**, 113–169 (2001).
36. B. M. von Reumont *et al.*, Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. *BMC Evol. Biol.* **9**, 119 (2009).
37. J. Mallatt, C. W. Craig, M. J. Yoder, Nearly complete rRNA genes assembled from across the metazoan animals: Effects of more taxa, a structure-based alignment, and paired-sites evolutionary models on phylogeny reconstruction. *Mol. Phylogenet. Evol.* **55**, 1–17 (2010).
38. E. Dell'Ampio *et al.*, Testing for misleading effects in the phylogenetic reconstruction of ancient lineages of hexapods: Influence of character dependence and character choice in analyses of 28S rRNA sequences. *Zool. Scripta* **38**, 155–170 (2009).
39. J. L. Rainford, M. Hofreiter, D. B. Nicholson, P. J. Mayhew, Phylogenetic distribution of extant richness suggests metamorphosis is a key innovation driving diversification in insects. *PLoS ONE* **9**, e109085 (2014).
40. G. Sasaki, K. Ishiwata, R. Machida, T. Miyata, Z.-H. Su, Molecular phylogenetic analyses support the monophyly of Hexapoda and suggest the paraphyly of Entognatha. *BMC Evol. Biol.* **13**, 236 (2013).
41. J. Kukalová-Peck "Fossil History and the evolution of Hexapod Structures" in *The Insects of Australia, a Textbook for Students and Researchers*, I. D. Naumann, Ed. (Melbourne Univ. Press, Melbourne, ed. 2, 1991), pp. 141–179.
42. R. G. Beutel, M. I. Yavorskaya, Y. Mashimo, M. Fukui, K. Meusemann, The Phylogeny of Hexapoda (Arthropoda) and the Evolution of Megadiversity. *Proc. Arthropod. Embryol. Soc. Jpn.* **51**, 1–15 (2017).
43. J. C. Regier *et al.*, Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
44. H. Letsch, S. Simon, Insect phylogenomics: New insights on the relationships of lower neopteran orders (Polyneoptera). *Syst. Entomol.* **38**, 783–793 (2013).
45. A. D. Imms, The ancestry of insects. *Nature* **139**, 399–400 (1937).
46. S. J. Lubbock *Monograph of the Collembola and Thysanura* (Ray Society, 1873).

47. Y. Mashimo, R. G. Beutel, R. Dallai, C.-Y. Lee, R. Machida, Embryonic development of Zoraptera with special reference to external morphology, and its phylogenetic implications (Insecta). *J. Morphol.* **275**, 295–312 (2014).
48. L. Salichos, A. Rokas, Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**, 327–331 (2013).
49. L. S. Jermiin, S. Y. W. Ho, F. Ababneh, J. Robinson, A. W. D. Larkum, The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.* **53**, 638–643 (2004).
50. W. Y. Yin, Studies on Chinese Protura. II., A new family of the suborder Eosentomoidea. *Acta Entomol. Sin.* **14**, 186–195 (1965).
51. R. Dallai, W. Y. Yin, Sperm structure of *Sinentomon* (Protura) and phylogenetic considerations. *Pedobiologia* **25**, 313–316 (1983).
52. J. François, R. Dallai, W. Y. Yin, Cephalic anatomy of *Sinentomon erythranum* Yin (Protura: Sinentomidae). *Int. J. Insect Morphol. Embryol.* **21**, 199–213 (1992).
53. S. L. Tuxen, Systematical position of *Sinentomon* (Insecta, Protura). *Bull. Natl. Mus. Nat. Sci., Ser. A, Zool.* **3**, 25–36 (1977).
54. Q. Xie, X. Tian, Y. Qin, W. Bu, Phylogenetic comparison of local length plasticity of the small subunit of nuclear rDNAs among all Hexapoda orders and the impact of hyper-length-variation on alignment. *Mol. Phylogenet. Evol.* **50**, 310–316 (2009).
55. E. Dell'Ampio, N. U. Szucsich, G. Pass, Protura and molecular phylogenetics: Status quo of a young love. *Soil Organisms* **83**, 347–358 (2011).
56. J. Rusek, Octostigma herbivora n. gen. & sp. Diplura: Projapygoidea: Octostigmatidae n. fam.) injuring plant roots in the Tonga Islands. New Zeal. *J. Zool.* **9**, 25–32 (1982).
57. P. Štys, J. Zrzavý, F. Weyda, Phylogeny of the Hexapoda and ovarian metamerism. *Biol. Rev.* **68**, 365–379 (1993).
58. B. G. M. Jamieson, *The Ultrastructure and Phylogeny of Insect Spermatozoa* (Cambridge University Press, 1987).
59. W.-J. Chen, M. Koch, J. M. Mallatt, Y.-X. Luan, Comparative analysis of mitochondrial genomes in Diplura (Hexapoda, Arthropoda): Taxon sampling is crucial for phylogenetic inferences. *Genome Biol. Evol.* **6**, 105–120 (2014).
60. M. Schwentner, S. Richter, D. C. Rogers, G. Giribet, Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: Highlighting the strength of taxon-specific matrices in phylogenomics. *Proc. Roy. Soc. B* **285**, 20181524 (2018).
61. A. R. Nabhan, I. N. Sarkar, The impact of taxon sampling on phylogenetic inference: A review of two decades of controversy. *Brief. Bioinform.* **13**, 122–134 (2012).
62. E. V. Kriventseva *et al.*, OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
63. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* **56**, 564–577 (2007).
64. C. Zhang, Y. Zhao, E. L. Braun, S. Mirarab, TAPER: Pinpointing errors in multiple sequence alignments despite varying rates of evolution. *Methods Ecol. Evol.* **12**, 2145–2158 (2021).
65. J. L. Steenwyk, T. J. Buida, Y. Li, X. X. Shen, A. Rokas, ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, 1–17 (2020).
66. M. J. Phillips, D. Penny, The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* **28**, 171–185 (2003).
67. J. F. Fleming, T. H. Struck, nRCFV: A new, dataset-size-independent metric to quantify compositional heterogeneity in nucleotide and amino acid datasets. *BMC Bioinform.* **24**, 145 (2023).
68. S. Naser-Khdour, B. Quang Minh, R. Lanfear, Assessing confidence in root placement on phylogenies: An empirical study using nonreversible models for mammals. *Syst. Biol.* **71**, 959–972 (2021).
69. X.-X. Shen *et al.*, Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1–13 (2018).
70. U. Mai, S. Mirarab, TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19**, 272 (2018).
71. M.-Y. Chen, D. Liang, P. Zhang, Phylogenomic resolution of the phylogeny of laurasiatherian mammals: Exploring phylogenetic signals within coding and noncoding sequences. *Genome Biol. Evol.* **9**, 1998–2012 (2017).
72. L. Salichos, A. Rokas, Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**, 327–331 (2013).
73. X. X. Shen, J. L. Steenwyk, A. Rokas, Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Syst. Biol.* **70**, 997–1014 (2021).
74. A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
75. A. Krausse, M. Wolff, Eine Uebersicht über die bisher aufgestellten fossilen und rezenten Insecte. *Archiv für Naturgeschichte, Abteilung A* **85**, 151–171 (1919).
76. H. Prell, Beiträge zur Kenntnis der Proturen. II. Anamerentoma und Holomerentoma, eine neue Einteilung der Hexapoden. *Zoologisches Anzeiger* **39**, 357–365 (1912).
77. N. Lartillot, Identifying the best approximating model in Bayesian phylogenetics: Bayes factors, cross-validation or wAIC? *Syst. Biol.* **72**, 616–638 (2023).
78. S. Klopfstein, The age of insects and the revival of the minimum age tree. *Austral Entomol.* **60**, 138–146 (2021).
79. S. Simon, H. Hadrys, A comparative analysis of complete mitochondrial genomes among Hexapoda. *Mol. Phylogenet. Evol.* **69**, 393–403 (2013).
80. S. Koenemann, F. R. Schram, T. M. Iliffe, Trunk segmentation patterns in Remipedia. *Crustaceana* **79**, 607–631 (2006).
81. A. Berlese, Nuovi Acerentomidi. *Redia* **5**, 16–19 (1908).
82. A. G. Sharov, Basic Arthropodan Stock, with Special Reference to Insects (Pergamon Press, 1966).
83. M. T. Neiber *et al.*, Global biodiversity and phylogenetic evaluation of Remipedia (Crustacea). *PLoS ONE* **6**, e19627 (2011).
84. D. A. Grimaldi, 400 million years on six legs: On the origin and early evolution of Hexapoda. *Arthropod Struct. Devel.* **39**, 191–203 (2010).
85. S. Koenemann *et al.*, Post-embryonic development of remipede crustaceans. *Evol. Dev.* **9**, 117–121 (2007).
86. R. Machida, Evidence from embryology for reconstructing the relationships of hexapod basal clades. *Arthropod Struct. Devel.* **64**, 95–104 (2005).
87. H. B. Boudreaux, *Arthropod Phylogeny with Special Reference to Insects* (Wiley-Interscience, 1979).
88. R. Dallai, M. Gottardo, R. G. Beutel, Structure and evolution of insect sperm: New interpretations in the age of phylogenomics. *Annu. Rev. Entomol.* **61**, 1–23 (2016).
89. C. Bedini, P. Tongiorgi, The fine structure of the pseudoculus of acerentomids Protura (insecta Apterygota). *Monitore Zoologico Italiano– Ital. J. Zoo.* **5**, 25–38 (1971).
90. W. Yin, L. Xue, Comparative spermatology of Protura and its significance on proturan systematics. *Scientia Sinica (Series B)* **36**, 575–586 (1993).
91. R. G. Beutel, F. Friedrich, X.-K. Yang, S.-Q. Ge, *Insect Morphology and Phylogeny: A Textbook for Students of Entomology* (Walter de Gruyter, 2013).
92. W. Y. Yin, A new idea on phylogeny of Protura with approach to its origin and systematic position. *Sci. Sin. Ser. B.* **27**, 149–160 (1984).
93. R. von Stummer-Traunfels, Vergleichende Untersuchungen über die Mundwerkzeuge der Thysanura und Collembolen. Sitzungsberichte der kaiserlichen Akademie der Wissenschaften: Mathematisch-naturwissenschaftliche Classe. *Abtheilung* **I**, **100**, 216–234 (1891).
94. Y. Ikeda, R. Machida, Embryogenesis of the dipluran *Lepidocampa weberi* Oudemans (Hexapoda, Diplura, Campodeidae): External morphology. *J. Morphol.* **237**, 101–115 (1998).
95. K. Dittrich, B. Wipfler, A review of the hexapod tracheal system with a focus on the apterygote groups. *Arthropod Struct. Dev.* **63**, 101072 (2021).
96. R. G. Beutel, F. Friedrich, E. P. Economo, Patterns of morphological simplification and innovation in the megadiverse Holometabola (Insecta). *Cladistics* **38**, 227–245 (2022).
97. R. Dallai *et al.*, The spermatogenesis and sperm structure of Acerentomon microrhinus (Protura, Hexapoda) with considerations on the phylogenetic position of the taxon. *Zoomorphol* 129, 61–80 (2010).
98. A. Malmström, T. Persson, Responses of Collembola and Protura to tree girdling–some support for ectomycorrhizal feeding. *Soil Organisms* **83**, 279–285 (2011).
99. P. Farina, S. Bedini, B. Conti, Multiple functions of Malpighian tubules in insects: A review. *Insects* **13**, 1001 (2022).
100. S. L. Bluhm *et al.*, Protura are unique: First evidence of specialized feeding on ectomycorrhizal fungi in soil invertebrates. *BMC Ecol.* **19**, 10 (2019).
101. R. Machida, I. Takahashi, Rearing technique for proturans (Hexapoda: Protura). *Pedobiologia* **48**, 227–229 (2004).
102. L. Galli *et al.*, A synopsis of the ecology of Protura (Arthropoda: Hexapoda). *Rev. Suisse Zool.* **126**, 155–164 (2020).
103. T. Bilde, J. A. Axelsen, S. Toft, The value of Collembola from agricultural soils as food for a generalist predator. *J. Appl. Ecol.* **37**, 672–683 (2000).
104. K. Endlweber, L. Ruess, S. Scheu, Collembola switch diet in presence of plant roots thereby functioning as herbivores. *Soil Biol. Biochem.* **41**, 1151–1154 (2009).
105. M. R. Smith, Cord-forming Palaeozoic fungi in terrestrial assemblages. *Bot. J. Linn. Soc.* **180**, 452–460 (2016).
106. J. Lozano-Fernandez *et al.*, A Cambrian–Ordovician terrestrialization of arachnids. *Front. Genet.* **11**, 182 (2020).
107. F. Haas, D. Waloszek, R. Hartenberger, *Devonohexapodus bocksbergensis*, a new marine hexapod from the Lower Devonian Hunsrück Slates, and the origin of Atelocerata and Hexapoda. *Org. Divers. Evol.* **3**, 39–54 (2003).
108. S. Du, F. Zhang, Data from "Phylogeny of the hexapod classes: Protura as the sister group to all other hexapods." GitHub. https://github.com/xtmtd/Phylogenomics/tree/main/basal_hexapods. Deposited 8 January 2024.
109. S. Du *et al.*, Construction of a phylogenetic matrix: Scripts and guidelines for phylogenomics. *Zool. Syst.* **48**, 107–116 (2023).