



Research Article

# Molecular diversity of Protura in southern High Appalachian leaf litter

Michael Caterino<sup>‡</sup>, Ernesto Recuero<sup>‡</sup>

<sup>‡</sup> Clemson University, Clemson, United States of America

Corresponding author: Michael Caterino ([mcateri@clemson.edu](mailto:mcateri@clemson.edu))

Academic editor: Ľubomír Kováč

Received: 27 Sep 2023 | Accepted: 24 Nov 2023 | Published: 24 Nov 2023

Citation: Caterino M, Recuero E (2023) Molecular diversity of Protura in southern High Appalachian leaf litter. Biodiversity Data Journal 11: e113342. <https://doi.org/10.3897/BDJ.11.e113342>

## Abstract

The higher elevations of the southern Appalachian Mountains, U.S.A., host a rich, but little-studied fauna of Proturan hexapods. Here, we publish 117 Proturan barcode sequences from this region, estimated by automated species delimitation methods to represent 72 distinct species, whereas only nine species have previously been reported from the region. Two families, Eosentomidae and Acerentomidae, co-occur at most sampling sites, with as many as five species occurring in sympatry. Most populations exhibit very low haplotype diversity, but divergences amongst populations and amongst closely-related species are very high, a finding common to other phylogeographic studies of Proturans. Though we were unable to identify any of the barcodes to species, they form a useful, if preliminary, glimpse of southern Appalachian Proturan diversity.

## Keywords

Protura, soil biodiversity, megabarcoding, Appalachia, species delimitation

## Introduction

The Protura, or 'coneheads', are an order of minute litter and soil-inhabiting hexapods. They are poorly known and little studied, but are generally considered to be general

detritivores, though a number of species have been observed to feed more directly in mycorrhizal fungi (Sturm 1959, Pass and Szucsich 2011, Bluhm et al. 2019). The group has a relatively low diversity, with around 800 species described from all continents, but Antarctica (Galli and Rellini 2020). The North American fauna of this group was reviewed by Allen (2007), who provided an annotated catalogue and some identification keys (species of *Eosentomon* Berlese 1908), building on an earlier regional synthesis by Ewing (1940). There are very few active specialists in the group, not only in North America (e.g. EC Bernard), but worldwide (Bernard and Whittington 2021) and much taxonomic and ecological work to be done on them.

The Protura fauna of the southeastern US has not received any specific attention. Allen's (Allen 2007) checklist indicates that 12 genera and 39 species, from three different families (Eosentomidae Berlese, 1909, Acerentomidae Silvestri, 1907 and Protentomidae Mills, 1932) are known from the southeastern region generally, most from only their type localities.

In this paper, we report the results for Protura of a molecular-barcoding project focused on the arthropods occurring in leaf litter of the higher elevations of the southern Appalachian Mountains. Most elevations above 1500 m in this region host a distinctive coniferous forest type dominated by red spruce (*Picea rubens* Sarg.) and Fraser fir (*Abies fraseri* [Pursh.] Poiret). The arthropods of these sky islands have been little documented and the Protura not at all. Of the 39 species reported from somewhere in the southeastern USA, nine have been more or less explicitly reported from the southern Appalachians Mountains (Allen 2007). None of these is represented by COI barcode sequences in any public databases. Although none of the sequences we report here is identified beyond the family level, they provide a unique snapshot of the region's genetic diversity of Proturans and offer a novel perspective on ranges and distributions of monophyletic lineages in the area.

## Material and methods

Samples were collected across southern Appalachia using an 8 mm mesh litter sifter with subsequent processing in Berlese funnels. Litters targeted were predominantly those of the high elevation coniferous spruce-fir forest floor, but lower elevation samples were more typically broadleaf deciduous and evergreen (*Rhododendron* L.) litters. See Fig. 1 for sampling sites and Suppl. material 2 for all sampling site data. Arthropods were collected directly into 100% ethanol and specimens sorted to unique morphotypes to which we refer as morphospecies. Prior to extraction, each specimen was photographed (<https://www.flickr.com/photos/183480085@N02/albums/72157710333048247>), then punctured. Following digestion with proteinase K, vouchers were recovered where possible (stored in 95% ethanol with a few drops of 100% propylene glycol) and the liquid fraction was purified using magnetic bead extraction. A COI minibarcode was amplified using primers BF2-BR2 (Elbrecht and Leese 2017), tagged with a unique combination of 9 bp indexes (see Meier et al. (2015)) and sequenced on either Illumina or Nanopore platforms, purifying and adding adapters following protocols specific to each. Sequences were demultiplexed and quality-checked to obtain a single barcode for each specimen. For more complete details

of laboratory and analytical procedures, please see Caterino and Recuero (2023). Recovered vouchers are housed in the Clemson University Arthropod Collection. Appendix 2 includes full label and voucher code data, as well as GenBank accession numbers.

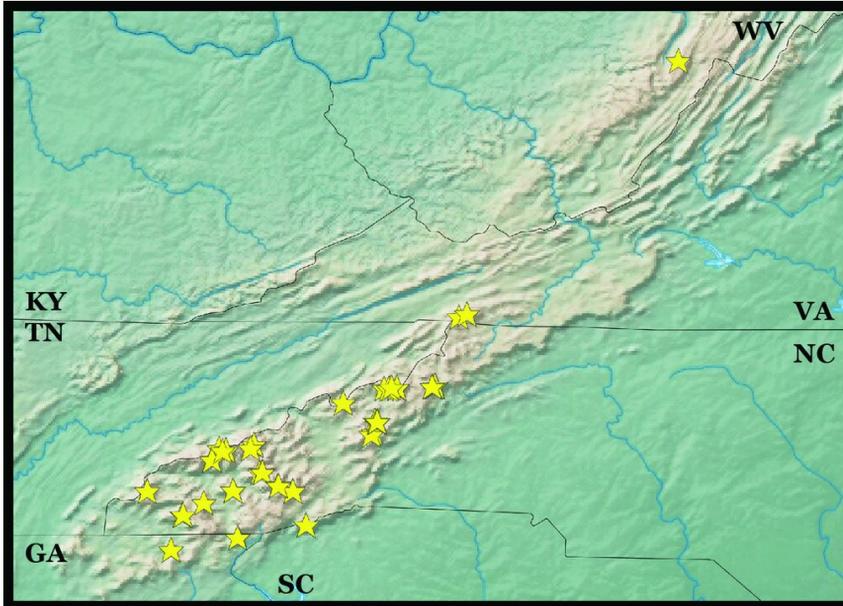


Figure 1. [doi](#)

Localities of barcoded Proturan sequences in the southern Appalachian Mountains.

To summarise similarities amongst barcoded specimens, we conducted Neighbour-joining analyses in PAUP\*, using Kimura 2-parameter distances (Kimura 1980). Preliminary hypotheses of species richness were inferred using Assemble Species by Automatic Partitioning (ASAP; Puillandre et al. (2020)), also with Kimura 2-parameter distance correction and mPTP (Kapli et al. 2017), using a tree generated with BEAST 1.10.4 (Suchard et al. 2018), using a strict molecular clock and a Yule tree prior, running for 10 million generations, sampling every 1000. As a secondary attempt to place Appalachian taxa in a broader taxonomic context, we downloaded available barcoding region sequences from GenBank and aligned those with ours, producing a matrix of 198 Proturan sequences. Relationships amongst the sampled populations were analysed with W-IQ-Tree (Trifinopoulos et al. 2016) under Maximum Likelihood criteria, using its automated model selection to choose GTR+gamma, with empirical base frequencies. All trees were rooted with a Dipluran (Campodeidae Meinert 1865) outgroup.

## Data resources

The data resources associated with this paper comprise a nexus file including all new barcode sequences (Suppl. material 1) and an Excel file (Suppl. material 2) with all associated voucher data, unique identifiers and GenBank codes.

# Results

We extracted 136 Proturan specimens from our samples for barcoding, 117 of which yielded barcode sequences (success rate 86%). Eleven voucher specimens were recovered following extraction.

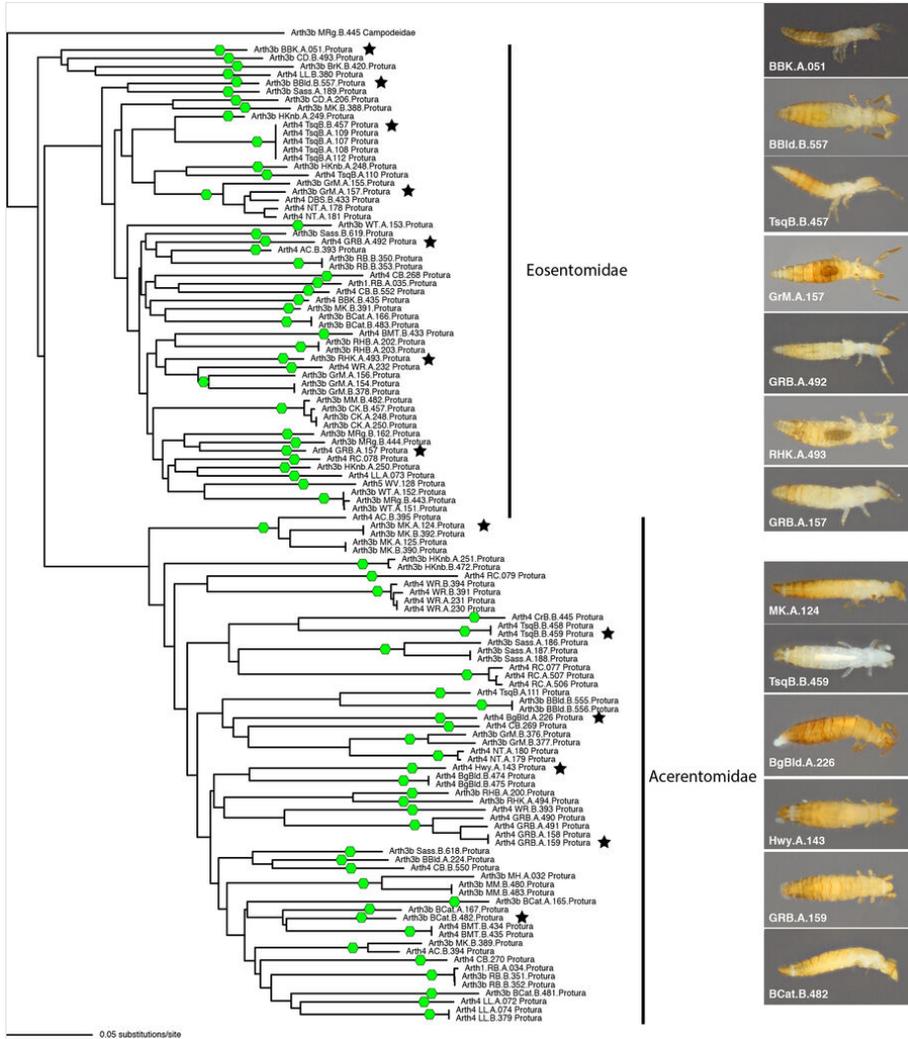


Figure 2. [doi](#)

Neighbour-joining tree of Appalachian region barcoding sequences, based on K2P distances. Green hexagons on branches indicate ASAP-inferred species level lineages. Black stars indicate specimens pictured at right.

A Neighbour-joining phylogram, showing degree of similarities amongst sequences, is shown in Fig. 2. They resolve into three main clusters, two of which apparently represent

Eosentomidae and one Acerentomidae (based on higher level comparisons not shown, but discussed further below).

ASAP analyses suggests these 117 sequences to correspond to 72 distinct species, 38 in the Eosentomidae and 34 in the Acerentomidae and a threshold distance of 9.9%. Most studied localities with presence of Protura harbour two of these units ( $n = 10$ ), but we have found localities with one ( $n = 7$ ), three ( $n = 8$ ), four ( $n = 4$ ) and up to five ( $n = 2$ ) approximately species-level units (Fig. 3). According to our data, most delimited units have a limited range and are exclusive to a single population and only five of them, two Eosentomidae and three Acerentomidae, are present in more than one sampled locality (three localities at most). The results of mPTP analyses, with 49 delimited species, are more conservative (Suppl. material 3); despite having high genetic distances, divergent haplotypes are often lumped into single hypothetical species. Considering mPTP species, we observe a few widely distributed species, some of them occurring across much of the southern Appalachian Mountains. A more detailed study of morphology might help distinguish between those hypotheses. In the absence of other evidence, these results must be considered as mere approximations; however, given the high distance threshold considered for ASAP results, it is likely that the number of species is closer to our higher estimates.

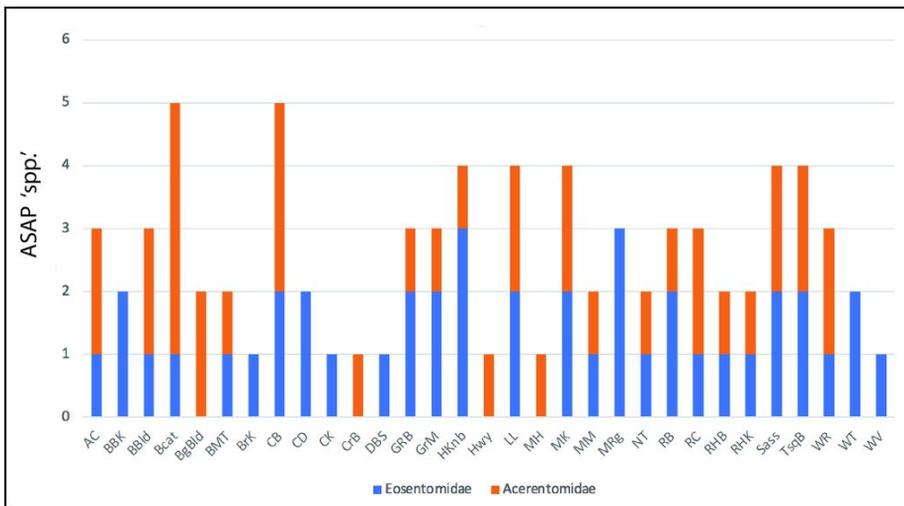


Figure 3. [doi](#)

Number of ASAP-estimated species by sampling site in each represented family.

## Discussion

Diversity of high Appalachian Protura appears to be high, with more than 70 inferred species in two different families over a relatively limited spatial sampling. Nearly all of these putative species are restricted to single sampling sites, with a few exceptions discussed further below. Sequences hypothesised to be conspecific were frequently identical, even

from different local samples and dates. Divergences from nearest neighbours, however, were invariably high, 8-10% uncorrected distance or higher. Clusters generally reflect geography, with nearest neighbours very often from nearby peaks or ranges (e.g. CD.A. 206 and MK.B.388, from Clingmans Dome and Mt. Kephart, separated by ~ 4 km on either side of the Newfound Gap in the Smokies). Similar patterns of very low intrapopulation variation in combination with high interpopulation and interspecific distances have been reported in other species-level studies of the group (Resch et al. 2014), suggesting some interesting population level dynamics.

Nearly all localities are furthermore represented by multiple distinct species and lineages, indicating a complex fauna with a long history of mixing. Big Cataloochee and Cowee Bald samples contain five species, while Huckleberry Knob, Tusquitee Bald, Sassafras Mt. and Mt. Kephart each have four. Thus, local faunas of Protura may be reasonably rich. We must also acknowledge the likelihood that most of our sites are likely still undersampled; particularly, by selecting only a single specimen per morphospecies per population, we may have overlooked morphologically similar or even cryptic species that may co-exist in the same locality.

Only a few species range beyond single sites. One putative species occurs not only on one of Grandfather Mountain's summits (Calloway Peak), but also on its lower slopes (Daniel Boone Scout and Nuwati Trails – though in another lineage that spans these sites, these form a more divergent sister species pair). A species in the Black Mountains was collected both on the summit of Mount Mitchell and on Celo Knob, 10 km apart. Another Black Mountains species was found on both Mount Mitchell and Mt. Hallback (2 km apart). A species was shared by Mt. Rogers and Whitetop, in southwestern Virginia, spanning 6 km. To some degree, these results will have been impacted by the density of sampling and it is conceivable that more species sharing across localities would be suggested if intervening localities were better sampled. However, it should not be surprising for such minute, relatively immobile and environmentally sensitive animals to be highly isolated and, consequently, highly diverse in such varied topography and environments.

There are limited clear biogeographic signals in these sequences and it is likely that many of these broader distributions predate many of the environmental fluctuations of the Pleistocene. Any larger cluster includes representatives from both sides of the Asheville Depression, typically observed as a major biogeographic breakpoint in the region (e.g. Browne and Ferree 2007, Hedin et al. 2015, Caterino and Langton-Myers 2019, Caterino in press). However, resolution that might reveal more significant biogeographic relationships is also limited by deep divergences and saturated signal in such a short mitochondrial fragment. More extensive sequencing of a similar set of taxa might be expected to show more consistent geographic signal.

Our sequences apparently represent the families Eosentomidae and Acerentomidae, as indicated in Fig. 2. This is based on nesting within clades of identified sequences of both, predominantly Asian species (reported in Bu and Wu 2012, Bu and Bai 2013, Shrubovych et al. 2014a, Bu et al. 2017, Qian et al. 2018, Carapelli et al. 2019) and a few Neotropical Acerentomidae (*Andinentulus* Tuxen, 1984; Shrubovych et al. (2014b)). Though most are

quite distant from southern Appalachian sequences, one of seven (unidentified) Eosentomidae sequences from eastern Canada available in the BOLD database resolves within our Appalachian species, as sister to a sequence from Whitetop Mt. in Virginia and one Asian *Eosentomon nivoculum* Yin 1981 sequence is resolved as sister to a rather divergent clade of sequences from the Great Smoky, Plott Balsam and Great Balsam Mts. (including CD.B.493, BBK.A.051, BrK.B.420 and LL.B.380).

This study represents the first view, limited though it is, of Proturan diversity in the southern Appalachian Mountains. As in many groups, this biodiversity hotspot appears to host a wealth of Proturan species, at least twice as much (and likely several times more) species richness than is yet reported. One challenge with a barcode-based approach, such as we report here, will be corresponding morphological work. Only a small number of voucher specimens (11) were recoverable following extraction. Between their small size and near transparency after they are digested with proteinase K, it is very difficult to avoid pipetting or otherwise overlooking the specimens themselves. It is also apparent that a morphospecies-based presorting is a very poor approximation to species. Where we compare pre-extraction photographs of putative DNA-based conspecifics, we see little consistency, with sorting frequently misled by varied degrees of sclerotisation, distension and other taxonomically meaningless artefacts. Bridging the gap to forge an integrative taxonomy of Proturans will require much more careful approaches (such as detailed in Böhm et al. (2011), Resch et al. (2014)). However, as a sorely neglected and likely megadiverse lineage, the results should repay the efforts.

## Acknowledgements

This study was funded by the U.S. National Science Foundation (Award DEB-1916263 to MSC) and the Clemson University Experiment Station (SC-1700596 to MSC). We also acknowledge the support of the John and Suzanne Morse Endowment for Arthropod Biodiversity. For permissions and assistance with fieldwork, we are grateful to the North Carolina State Parks, Great Smoky Mountains National Park, Blue Ridge Parkway National Park, Monica Martin, Frank Etzler, Curt Harden, Patricia Wooden, Adam Haberski, Roy Kucuk, Laura Vásquez-Vélez, Laary Cushman, Paul Marek, Michael Ferro and Will Kuhn. Mary Atieh, Caroline Dukes, Caroline McCluskey, Grace Holliday, Grace Arnold, Hannah Skinner, Alejandra Carranza and Anthony Villanueva provided valuable assistance in the lab. Thank you to Ernest C. Bernard and Yun Bu for their helpful reviews. This paper represents Technical Contribution No. **7231** of the Clemson University Experiment Station.

## Funding program

Division of Environmental Biology

## Grant title

Leaf litter arthropods of High Appalachia

## Conflicts of interest

The authors have declared that no competing interests exist.

## References

- Allen R (2007) Studies on the North American Protura 1: catalogue and atlas of the Protura of North America; description of new species; key to the species of *Eosentomon*. Proceedings of the Academy of Natural Sciences of Philadelphia 156 (1): 97-116. [https://doi.org/10.1635/0097-3157\(2007\)156\[97:sotnap\]2.0.co;2](https://doi.org/10.1635/0097-3157(2007)156[97:sotnap]2.0.co;2)
- Bernard E, Whittington A (2021) Papers and New Species of Minor Insect Orders Published in Zootaxa, 2001–2020. Zootaxa 4979 (1): 232-235. <https://doi.org/10.11646/zootaxa.4979.1.26>
- Bluhm S, Potapov A, Shrubovych J, Ammerschubert S, Polle A, Scheu S (2019) Protura are unique: first evidence of specialized feeding on ectomycorrhizal fungi in soil invertebrates. BMC Ecology 19 (1). <https://doi.org/10.1186/s12898-019-0227-y>
- Böhm A, Bartel D, Szucsich NU, Pass G (2011) Confocal imaging of the exo- and endoskeleton of Protura after non-destructive DNA extraction. Soil Organisms 83: 335-345.
- Browne R, Ferree P (2007) Genetic structure of Southern Appalachian “Sky Island” populations of the Southern Red-backed Vole (*Myodes gapperi*). Journal of Mammalogy 88 (3): 759-768. <https://doi.org/10.1644/06-mamm-a-049r1.1>
- Bu Y, Wu DH (2012) Revision of Chinese *Yamatentomon* (Protura: Acerentomata: Acerentomidae), with description of one new species, redescription of *Yamatentomon yamato*, and key to world species. Florida Entomologist 95 (4): 839-847. <https://doi.org/10.1653/024.095.0404>
- Bu Y, Bai Y (2013) *Hesperentomon yangi* sp. n. from Jiangsu Province, Eastern China, with analyses of DNA barcodes (Protura, Acerentomata, Hesperentomidae). ZooKeys 338: 29-37. <https://doi.org/10.3897/zookeys.338.6099>
- Bu Y, Qian C, Luan Y (2017) Three newly recorded species of Acerentomata (Hexapoda: Protura) from China, with analysis of DNA barcodes. Entomotaxonomia 39: 1-14. <https://doi.org/10.11680/entomotax.2017001>
- Carapelli A, Bu Y, Chen W, Nardi F, Leo C, Frati F, Luan Y (2019) Going deeper into high and low phylogenetic relationships of Protura. Genes 10 (4): 1-25. <https://doi.org/10.3390/genes10040292>
- Caterino MS, Langton-Myers SS (2019) Intraspecific diversity and phylogeography in southern Appalachian *Dasycerus carolinensis* (Coleoptera: Staphylinidae: Dasycerinae). Insect Systematics and Diversity 3 (6): 1-12. <https://doi.org/10.1093/isd/ixz022>
- Caterino MS, Recuero E (2023) Shedding light on dark taxa in high Appalachian leaf litter - assessing patterns of endemism using large scale, voucher-based barcoding. Insect Conservation and Diversity. <https://doi.org/10.1111/icad.12697>
- Caterino MS (in press) A new, flightless species of *Medon* (Staphylinidae: Paederinae) from high Appalachia, with intraspecific phylogeographic analysis and description of its associated larva. The Coleopterists Bulletin.

- Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science* 5: 1-11. <https://doi.org/10.3389/fenvs.2017.00011>
- Ewing HE (1940) The Protura of North America. *Annals of the Entomological Society of America* 33 (3): 495-551. <https://doi.org/10.1093/aesa/33.3.495>
- Galli L, Rellini I (2020) The geographic distribution of Protura (Arthropoda: Hexapoda): A review. *Biogeographia* 35: 51-69. <https://doi.org/10.21426/B635048595>
- Hedin M, Carlson D, Coyle F (2015) Sky island diversification meets the multispecies coalescent - divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Molecular Ecology* 24 (13): 3467-3484. <https://doi.org/10.1111/mec.13248>
- Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T (2017) Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* 33 (11): 1630-1638. <https://doi.org/10.1093/bioinformatics/btx025>
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16 (2): 111-120. <https://doi.org/10.1007/bf01731581>
- Meier R, Wong W, Srivathsan A, Foo M (2015) \$1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics* 32 (1): 100-110. <https://doi.org/10.1111/cla.12115>
- Pass G, Szucsich NU (2011) 100 years of research on the Protura: many secrets still retained. *Soil Organisms* 83: 309-334.
- Puillandre N, Brouillet S, Achaz G (2020) ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources* 21 (2): 609-620. <https://doi.org/10.1111/1755-0998.13281>
- Qian CY, Bu Y, Luan YX (2018) DNA barcoding and an updated key to the genus *Hesperentomon* (Protura: Acerentomata: Hesperentomidae), with a new species from Northwest China. *Zootaxa* 4462: 523-534. <https://doi.org/10.11646/zootaxa.4462.4.5>
- Resch MC, Shrubovych J, Bartel D, Szucsich N, Timelthaler G, Bu Y, Walzl M, Pass G (2014) Where taxonomy based on subtle morphological differences is perfectly mirrored by huge genetic distances: DNA barcoding in Protura (Hexapoda). *PLOS One* 9 (3). <https://doi.org/10.1371/journal.pone.0090653>
- Shrubovych J, Schneider C, D'Haese C (2014a) Two new species of *Acerentulus* Berlese 1908 (Protura: Acerentomata: Acerentomidae) with their barcode sequences and a key to the *cunhai* group. *Annales de la Société entomologique de France (N.S.)* 50 (2): 129-140. <https://doi.org/10.1080/00379271.2014.934036>
- Shrubovych J, Schneider C, D'haese CA (2014b) Revision of the genus *Andinentulus* (Protura: Acerentomidae: Berberentulinae), with a key to South and Central American Acerentomidae species. *Annals of the Entomological Society of America* 107 (3): 567-574. <https://doi.org/10.1603/an13151>
- Sturm H (1959) Die Nahrung der Proturen. *Die Naturwissenschaften* 46 (2): 90-91. <https://doi.org/10.1007/bf00599134>
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4 (1): vey016. <https://doi.org/10.1093/ve/vey016>

- Trifinopoulos J, Nguyen L, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* 44 <https://doi.org/10.1093/nar/gkw256>

## Supplementary materials

### Suppl. material 1: Cytochrome Oxidase I barcode region sequences for Appalachian Protura [doi](#)

**Authors:** Caterino MS, Recuero E

**Data type:** phylogenetic

**Brief description:** A total of 117 partial COI sequences of Protura from the southern Appalachians, in nexus format.

[Download file](#) (51.95 kb)

### Suppl. material 2: Collecting and voucher information for Proturan barcode sequences [doi](#)

**Authors:** Caterino MS, Recuero E

**Data type:** occurrence

**Brief description:** An Excel spreadsheet containing specimen collecting data (locality, date, lat/lon), voucher codes, DNA extraction codes and GenBank accession numbers for all sequences reported.

[Download file](#) (19.44 kb)

### Suppl. material 3: Species delimitation by mPTP. Tree generated by BEAST. [doi](#)

**Authors:** Caterino MS, Recuero E

**Data type:** phylogenetic hypothesis

**Brief description:** Bayesian topology generated by BEAST for mPTP species delimitation analysis. Clades in red represent collections of specimens united as distinct putative species hypotheses, while terminal branches in green indicate putative species represented by single OTUs.

[Download file](#) (22.59 kb)