SCIENTIFIC REPORTS

OPEN

Delimiting species of *Protaphorura* (Collembola: Onychiuridae): integrative evidence based on morphology, DNA sequences and geography

Xin Sun[1,3], Feng Zhang[2,4], Yinhuan Ding[2], Thomas W. Davies[5], Yu Li[3] & Donghui Wu[1,6,7]

Species delimitation remains a significant challenge when the diagnostic morphological characters are limited. Integrative taxonomy was applied to the genus *Protaphorura* (Collembola: Onychiuridae), which is one of most difficult soil animals to distinguish taxonomically. Three delimitation approaches (morphology, molecular markers and geography) were applied providing rigorous species validation criteria with an acceptably low error rate. Multiple molecular approaches, including distance- and evolutionary model-based methods, were used to determine species boundaries based on 144 standard barcode sequences. Twenty-two molecular putative species were consistently recovered across molecular and geographical analyses. Geographic criteria were was proved to be an efficient delimitation method for onychiurids. Further morphological examination, based on the combination of the number of pseudocelli, parapseudocelli and ventral mesothoracic chaetae, confirmed 18 taxa of 22 molecular units, with six of them described as new species. These characters were found to be of high taxonomical value. This study highlights the potential benefits of integrative taxonomy, particularly simultaneous use of molecular/geographical tools, as a powerful way of ascertaining the true diversity of the Onychiuridae. Our study also highlights that discovering new morphological characters remains central to achieving a full understanding of collembolan taxonomy.

Springtails (Collembola) are one of the dominant soil arthropods in almost all terrestrial ecosystems. Until now, there have been about 8500 species reported worldwide[1]. Having well differentiated ecomorphological life-forms and feeding guilds, springtails are considered to play important functional roles in many processes of soil ecosystems, such as carbon and nitrogen cycling, soil microstructure formation and plant litter decomposition processes[2–5]. At the same time, the structure of collembolan communities can be greatly influenced by soil acidification, nitrogen supply, global climate change and intensive farming[6,7]. When combined with recent advances in taxonomy, springtails can provide ideal model systems for many scientific fields (e.g. ecotoxicology, biogeography and ecology)[8].

The genus *Protaphorura* Absolon, 1901 (Poduromorpha: Onychiuridae) is one of the most abundant collembolan groups which includes more than 130 species[1], and is widely distributed worldwide. Species of the genus

[1]Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, 130102, China. [2]Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing, 210095, China. [3]Engineering Research Center of Chinese Ministry of Education for Edible and Medicinal Fungi, Jilin Agricultural University, Changchun, 130118, China. [4]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China. [5]Centre for Geography, Environment and Society, University of Exeter, Penryn, Cornwall, TR10 9FE, UK. [6]Key laboratory for vegetation ecology, ministry of education, Northeast Normal University, Changchun, 130117, China. [7]Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, Changchun, 130117, China. Correspondence and requests for materials should be addressed to F.Z. (email: xtmtd.zf@gmail.com) or D.W. (email: wudonghui@iga.ac.cn)

are blind and usually white as other genera of the family[2]. They are mostly euedaphic in different ecosystems, and some species are crop pests which can reduce yields and productivity[9–11].

Because of frequent variations in morphological characters, species discrimination in *Protaphorura* has been debated for nearly sixty years. The formulae of pseudocelli, which are unique epicuticular structures in Onychiuridae and Tullbergiidae, and probably the defensive structures located on head and body, are proved to be intraspecifically stable in many other genera (e.g. *Thalassaphorura*, *Onychiurus*, *Oligaphorura*, *Dimorphaphorura*) and very useful to differentiate species[12–15]. Unfortunately, the extensive variation of the pseudocelli formula in *Protaphorura* has been disputed as to whether it represents interspecific or intraspecific differences. According to Gisin[16, 17], adults of European species of the *Onychiurus armatus* group, which is the synonym of the present genus *Protaphorura*, were distinguished from each other mainly by the number of pseudocelli particularly those on the dorsal body. Several new species were reported following Gisin's definition later[18–20]. However, some authors considered many of Gisin's species as ecological or local modifications because of the inconsistency in the characters employed[21–23]. Based on Icelandic and Swedish material, Bödvarsson[24, 25] challenged the Gisin's system of species discrimination and regarded them as of intraspecific significance, because of the intermediates existing between different forms. To address these problems, Pomorski[26, 27] studied the postembryonic development of some members of the *P. armata* species group and proposed that the relative position rather than number of pseudocelli is of a high taxonomic value. Besides the formulae of pseudocelli, other diagnostic characters (e.g. the number of chaetae on prothorax, the presence/absence of additional microchaeta on abdominal terga I–III and V, the relative position of prespinal microchaetae on abdominal tergum VI) have been proposed to be useful for species identification. The number and arrangement of parapseudocelli (a similar structure as pseudocelli, but without a chitinized border), which were introduced by Pomorski and are now frequently reported, are considered by some authors to be of doubtful taxonomic value, not just in *Protaphorura*, but even in other Onychiuridae groups[28, 29]. Accordingly, the species boundaries in *Protaphorura* are often ambiguous, particularly for conspecific species reported from different populations[2, 30–33].

Species delimitation has long been confused because of the differences among contemporary species concepts[34]. Many disciplines and corresponding species concepts, such as the widely used biological species concept, the phylogenetic species concept, and the morphological/phenetic species concept[34], could be applied to species delimitation, but no single discipline is entirely proper[35]. As a fundamental biological unit, a species needs to be defined upon some properties in a given study. The generalized "independent evolving metapopulation lineage" (Generalized Lineage Concept, GLC)[34, 36] has been broadly adopted to determine the species boundary from empirical data[37]. In recent years, an integrative approach to taxonomy has been promoted, combining different kinds of disciplines such as morphology, phylogeography, population genetics, ecology and behavioral biology[35, 37, 38]. This could increase rigor in species delimitation, and avoid failure inherent to single disciplines[35]. However, like in most other arthropod groups, the identification and delimitation of Collembola species have relied mainly on the discrimination of external morphological characteristics[8] and thus species are called as morphospecies. As there are some limitations to the use of the traditional, morphology-based method, other disciplines are often applied to solve taxonomical problems. While some data, like ecology, behavioral biology, life history and chemistry, are relatively difficult to obtain, especially in most cases of Collembola due to their small size and low research effort, DNA taxonomy has been proved particularly powerful when used in combination with traditional taxonomic approaches[39–42] and increasingly useful for the identification of cryptic species[43–48]. Meanwhile, different delimitation approaches based on distance or evolutionary models have been developed along with advances in DNA sequencing and computational technologies[37], such as Automatic Barcode Gap Discovery (ABGD)[49], the Poisson Tree Processes model (PTP)[50], and the general mixed Yule coalescent model (GMYC)[51] which were applied for the single genetic locus; the Bayesian method BPP[52] and O'Meara's heuristic method[53] for multiple loci. Every method has its own parameterization and a series of simplifying. ABGD automatically clustered sequences into candidate species based on pairwise distances by detecting barcoding gaps using a command-line version[49], and PTP tested species boundaries on non-ultrametric phylogenies by detecting significant difference in the number of substitutions between species and within species[50]. In order to strengthen the confidence in the results, applying a wide range of analyses methods to delimit species has been proposed, as any one of the assumptions implicit in using any one method could possibly be violated in a particular empirical system[37].

DNA-based methods have recently resolved a number of problems encountered when delimiting species in a number of genera of Collembola, e.g. *Orchesella*[54], *Deutonura*[55], *Lepidocyrtus* and *Pseudosinella*[56], *Megalothorax*[57], *Entomobrya*[58]. The only molecular contribution available in the genus *Protaphorura* is from the thesis of Burkhardt[59]. Distance trees of several taxa within the genus were analyzed in his work based on mitochondrial cytochrome oxidase subunits I and II (COI and COII) and the nuclear internal transcribed spacer 1 (ITS1) fragments with concordant patterns obtained from different genes. The ITS1 was less effective for species discrimination than the mitochondrial fragments. The inferences drawn from species delimitation studies should be conservative, and it will be much more convincing if the results are congruent across a wide range of species delimitation approaches applied to the data including molecular, morphological and geographical[37].

The present study aims to: (i) assess the efficiency of different data (morphology, DNA sequences, geography) and approaches to delimitating species of the genus *Protaphorura*; (ii) evaluate diagnostic utility of morphological characters currently used to distinguish species, and to identify new characters; (iii) summarize a reliable operational workflow for integrative taxonomy of the genus *Protaphorura*, as well as the Onychiuridae in general.

Results

Species delimitation. In total, 108 individuals (75% of all barcoded specimens) were examined for morphological diagnostic characters after DNA extraction (Supplementary Table S3). The remaining 36 specimens broke during the extraction or characters were not visible in slide mounted specimens. Ten morphospecies were recognized based on traditional morphological diagnostic characters, including five individuals with either reduced
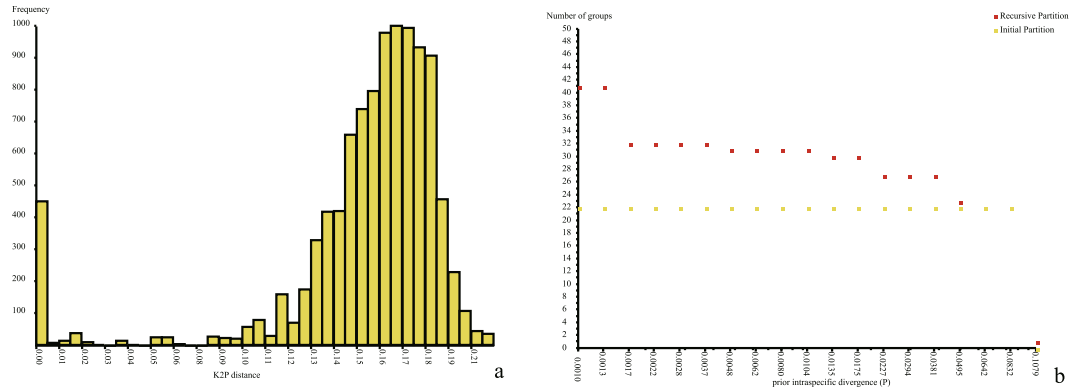
Figure 1. ABGD species delimitation. (**a**) Frequency histogram of K2P pairwise distances. (**b**) Partitions under different prior intraspecific divergences.
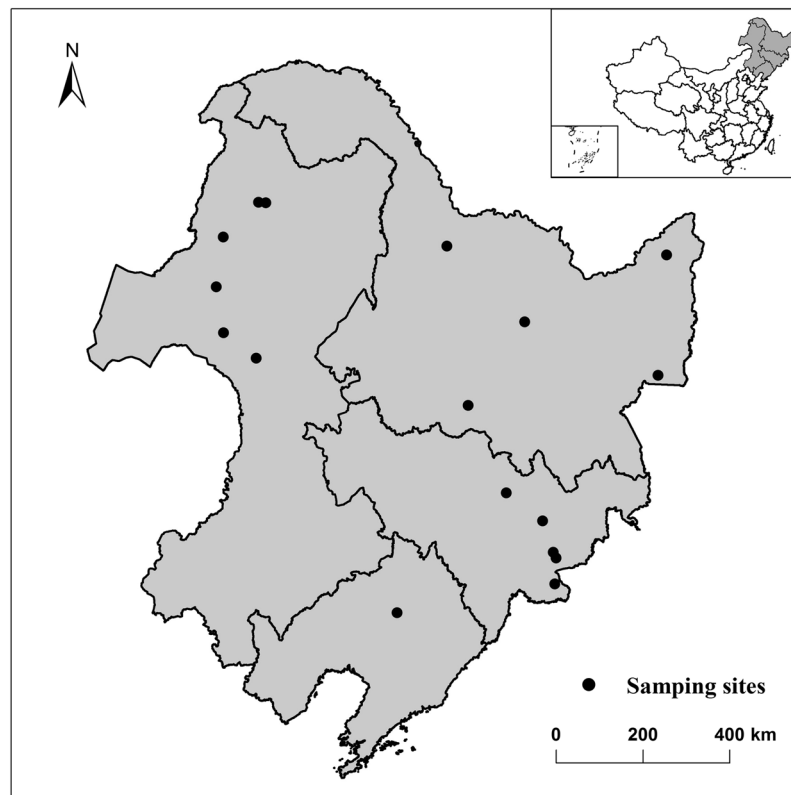

Figure 2. The distribution map of collection sites. The map is generated using software ArcGIS 10.3 (http://www.esri.com/software/arcgis) based on the geospatial data from the National Geomatics Center of China.

or supplementary pseudocelli developed in one side of the body (Supplementary Tables S3 and S6). Results of examination of other minor additional characters were listed in Supplementary Table S6.

ABGD generated 22 initial/primary partitions and 22–41 recursive/secondary partitions while prior intraspecific divergences (P) varied from 0.001 to 0.1079 (Fig. 1b). Rough "barcoding gaps" were observed at K2P distance of around 0.03, 0.045, 0.08 (Fig. 1a), corresponding to plateaus of 27, 23, 22 MOTUs (molecular operational taxonomical units) (Fig. 1b). Primary and secondary partitions were identical at 22 MOTUs (P > 0.064). Because primary partitions are stable on a wider range of prior values and are usually close to the hypothetical species number described by taxonomists[49], the value 22 was finally selected as the result of ABGD (Fig. 3).

For the single phylogenetic tree, bPTP.py estimated 26–40 (mean 30.98) putative species with a best supported partition of 30 species. With multiple bootstrap trees as input, 25–36 (mean 29.15) species were estimated by PTP.py, 27 in highest bootstrap supported partition. All clusters under both PTP.py and bPTP.py models formed monophyletic clades with high node bootstrap supports (>0.75, Fig. 3).
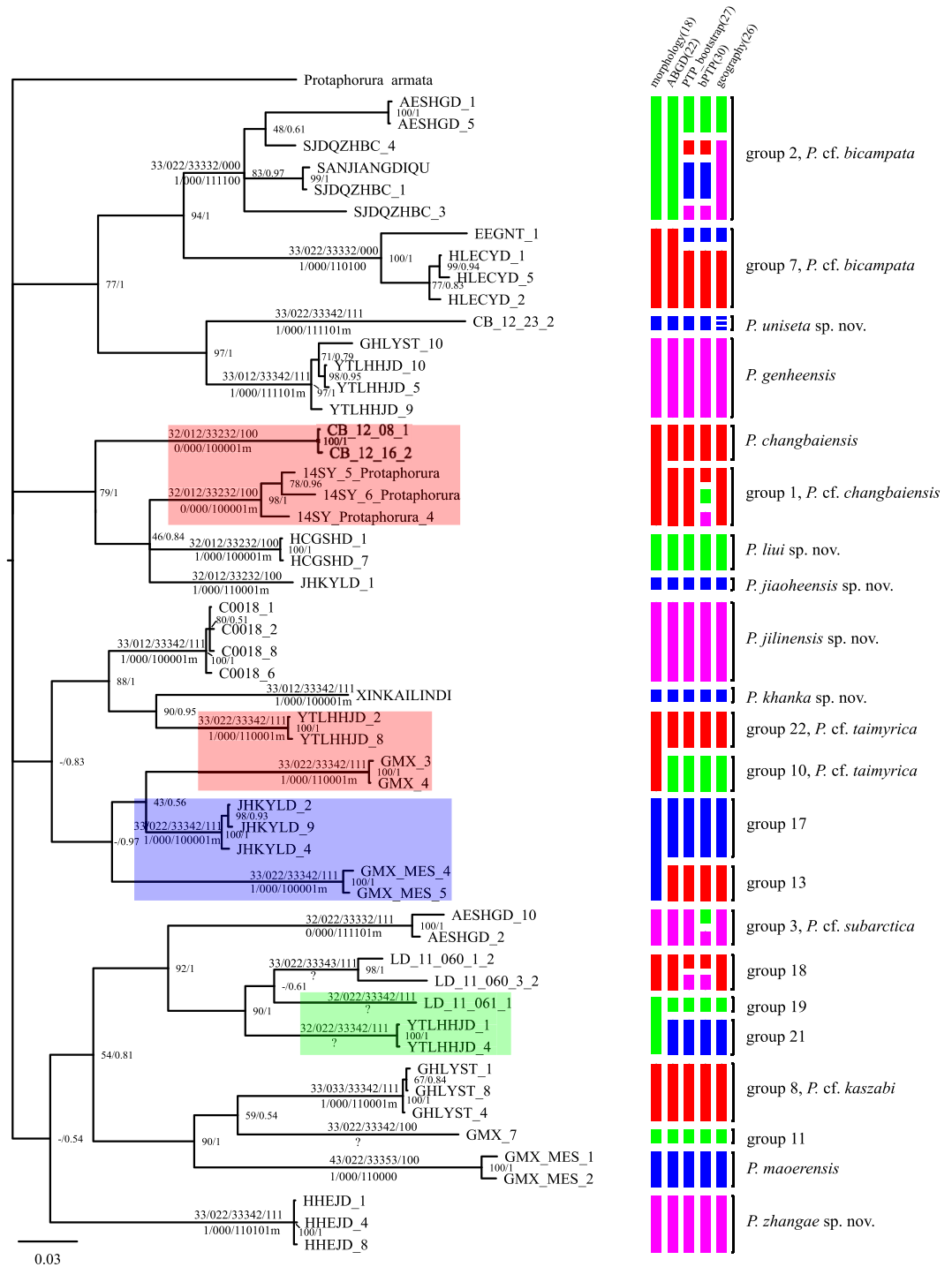
Figure 3. Summary from all species delimitation analyses on a Bayesian consensus tree. Node values represent likelihood bootstrap and posterior probabilities, respectively, with a–indicating nodes not compatible between the analyses. Colored bars represent hypothesized species groupings based on corresponding delimitation analyses. The geographical delimitation indicates the result of threshold of 50 kilometers. Colored clades on phylogeny are taxa morphologically unresolved. Number formulae above and below branches represent dorsal pseudocelli and ventral parapseudocelli, respectively.

Under the primary hypothesis of 10 morphospecies, 27, 26, 26, 25, 20, 16, 16, 13 putative species were delimited using the distance threshold values 10, 20, 50, 100, 200, 300, 400, 500 kilometers respectively (Supplementary Tables S4 and S5). The estimates for number of species with different geographical threshold values showed an initial plateau around 26 and a second at 20 with a steep decline when distances between populations more than 100 kilometers. Because of the long-term persistence within restricted geographical limits for Collembola[60], the plateau implied a sign of dispersal limits for species. Considering the possibility of no apparent gene flow across

distances of tens of kilometers[60], 26 putative species under a distance threshold of 50 kilometers were accepted for geographical criterion (Fig. 3).

The numbers of putative species estimated by the above methods were much greater than the ten morphospecies. Twenty-two MOTUs (GLC species), corresponding to ≈0.07 divergence threshold, were congruent across ABGD, PTP and geographical delimitations (Fig. 3). Among 22 putative species, the mean intra- and inter-specific distances were 0.93% (0–5.88%) and 15.91% (8.59–20.08%) for COI, respectively. Maximal divergences within species exceeded 5% in lineages group 2 (Gp 2) (Supplementary Tables S1 and S2, Supplementary Fig. S1).

When additional morphological characters were examined in combination with the traditional features, more morphological lineages were recognized, and most of them matched the 22-MOTUs partition. Among them, four twins cannot be distinguished on morphology: Gp 1 and Gp 5, Gp 10 and Gp 22, Gp 13 and Gp 17, Gp 19 and Gp 21 (Fig. 3). Finally, fourteen candidates were congruent across all three species validation criteria, including 2 known (Gp 9 and Gp 12), 6 new to science (Gp 4, Gp 6, Gp 14, Gp 15, Gp 16 and Gp 20), and the remaining which could not be designated a valid species name due to a lack of sufficient diagnostic information (e.g. Gp 11 and Gp 18 could not be further studied due to the lack of additional specimens; Gp 2, Gp 3, Gp 7 and Gp 8 have unclear taxonomical status due to the lack of detailed information available from similar type specimens from other collections) (Supplementary Tables S1–3). The morphological description of the six new species are given in the Supplementary files.

Discussion

In the present study, we combined three types of delimitation data (morphology, molecular markers and geography) to resolve the taxonomical problems in the genus *Protaphorura*. The genus is most abundant and widely distributed group of Onychiuridae in northeast China[33], and this gives us a good chance to collect a great many species and analyze both morphological and genetic information from fresh materials. Molecular markers, as well as geography, proved effective for delimitating. The results from the combination of morphological and molecular analyses, or morphological and geographical analyses were mostly congruent for species delimitation. In particular, morphology with ABGD provided the highest accuracy and could be used as a standard operational workflow in future taxonomy of the Onychiuridae.

Furthermore, the geographical distances may also help to discriminate taxa among closely related species, when the molecular information is unavailable. Geography criterion works well for species delimitation because of the low dispersal capabilities and long-term persistence[60, 61], which result in high endemism and non-overlapping distribution. Geographical delimitation was mostly consistent with molecular approaches except delimitation for the widespread *P. uniseta* sp. nov. in the present study. Many collembolan species, which could be widely distributed in regions of large geographical scale, have been demonstrated to have cryptic diversity[62, 63]. Therefore, geographic criterion should be available for most onychiurids. We don't encourage the single use of geographical distances to delimit species. Endemism extent and range of distributing area have not been estimated for most collembolan groups as well as most insects. As strategies implemented in this study, geographic criteria on the basis of preliminary morphological delimitation using simple morphological characters would be robust and accurate.

The inferences made here based on congruent results across several analyses are considered to be credible for species delimitation[37]. Several studies have combined the evidence from morphology and molecular markers in Collembola[55–58, 63, 64]. In our study, two new species delimitation methods based on molecular data (ABGD and PTP) were performed, and the results were mostly congruent. Coupled with the geographical evidence and using the Generalized Lineage Concept of species delimitation, the species of *Protaphorura* have been well discriminated here. Therefore, integrative taxonomy is essential for the accurate species identification of Onychiuridae, and it should be encouraged in the future taxonomical work.

Burkhardt's work[59] based on the results of molecular and morphological analyses supported Pomorski's conclusion[27], that the position of pseudocelli is more important than their number for taxonomy, especially for the 'deformed' pseudocelli which are reduced or supplementary pseudocelli present on only one side of the body. In a limited number of 'deformed' pseudocelli cases in our study (5 in 108 individuals), the diagnostic value of the pseudocelli formulae was validated in a number of species with the application of integrative taxonomy. Although frequent variations in the number and position of pseudocelli on the dorsal body have been reported in several species[26, 27], we discovered here that the characters were relatively stable intraspecifically in the present 22 MOTUs which were both supported by the molecular and geographical evidence. Overall the results confirmed the importance of pseudocelli formulae for discriminating species in Onychiuridae[17, 18] in the majority of cases.

Parapseudocelli, a long neglected structure in Onychiuridae, was first found by Rusek[28] and never reported in other groups of Collembola. It is usually located on the ventral side of body, subcoxae 1 of legs and anal valves, while on the dorsal side of the body it occurs in some genera of the tribe *Hymenaphorurini* Pomorski, 1996. The parapseudocelli formula was first used as diagnostic character by Weiner[65] and Pomorski[27, 66]. However, parapseudocelli are sometimes difficult to see and stated by several authors as "not seen" or "invisible" rather than "absent" in the descriptions. Therefore, the practicality of using this character in species diagnosis was questionable and it has not been widely used. In the present study, parapseudocelli formulae were shown to be stable and had no variations at the intraspecific level of *Protaphorura*. Variation in the number of the parapseudocelli is not common in other groups of Onychiuridae[12, 29, 67–69] and for this reason they were proposed to be of great taxonomic values in species discrimination[70]. The instances of intraspecific variability reported for this character system before may correspond to inter-specific differences, because the variation reported referred to inter-populational rather than intra-populational comparisons[15, 32]. Compared to ten morphospecies recognized by traditional diagnostic

characters, seven more have been recognized by the addition "parapseudocelli formulae" in our study, thus confirming views of Sun & Li[70] on the diagnostic value of this character system.

While useful, several MOTUs, which were well separated by evidence from DNA sequences and geography, still could not been separated by morphological characters. Thus, other minor characters, which may have differences at the species level, were also explored. The number of chaetae on Th. sternum II, was often recorded in previous studies on Onychiuridae and can be relatively stable intraspecifically[13, 15, 29, 71, 72], although it has not yet been treated as an important diagnostic character. When closely related taxa in this study are compared, the number of chaetae on Th. sternum II enabled G4 to be differentiated from Gp 20 (although this feature was variable in a small percentage of the examined specimens) and here we propose that it could be used as a main diagnostic character. The axial chaetae on abodominal terga IV–VI and the number of chaetae on subcoxa 1 of legs I–III, which are relatively stable at the intraspecific level in many genera of Onychiuridae[12–14, 29, 68, 72], were also checked. However, they had a large ratio of intraspecific variations and could not be used as diagnostic characters in *Protaphorura*.

Despite the findings of this study, inadequacies in the morphological taxonomy of the *Protaphorura* still exist as evidenced by the presence of four pairs of morphologically cryptic species (e.g. Gp 1 and Gp 5, Gp 10 and Gp 22, Gp 13 and Gp 17, Gp 19 and Gp 21). This deficiency is caused by a lack of sufficient morphological characters or the existence of cryptic species, and requires further investigation.

Material and Methods

Taxon sampling. Samples were collected from areas of high biodiversity in northeast China, including the Changbai Mountain Range, the Greater and Lesser Khingan Range and the Sanjiang Plain (Fig. 2). Specimens were collected by Berlese extraction, and stored in 99% ethanol at −20 °C. One hundred and forty-four individuals from 17 populations were selected for molecular analysis. In this study, a population was defined as individuals collected from sites within a 5 kilometer radius of each other. The samples were stored pending analysis in the Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun (NEIGA), China and the Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing (NJAU), China.

DNA extraction and sequencing. DNA was extracted using an Ezup Column Animal Genomic DNA Purification Kit (Sangon Biotech, Shanghai, China) following the manufacturer's standard protocols. Specimens were kept in 75% ethanol for further morphological examination after the 1h-lysis buffer was transferred to the pipette containing silica column. Extractions were performed non-destructively (small tissue biopsy) allowing further morphological examination of the specimens. Primers were LCO1490/HCO2198 commonly used for metazoa[73]. Amplification volume and procedure were listed in Zhang *et al.*[63]. All PCR products were checked on a 1% agarose gel. Successful products were purified and sequenced in both directions by Majorbio (Shanghai, China) on the ABI 3730XL DNA Analyser (Applied Biosystems). Sequences were assembled in Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, USA), and deposited in GenBank (Supplementary Table S3). Sequences were blasted in GenBank and checked for possible errors, then were preliminarily aligned using MAFFT v7.149 by the Q-INS-I strategy[74]. Alignments were checked and corrected manually, with a final 658 bp for COI.

Phylogenetic analyses. Neighbour-joining (NJ) trees and genetic distances were calculated in MEGA 5.0[75], with the Kimura-2 parameter model (K2P, Kimura 1980) and pairwise deletion for gaps. Node supports of NJ tree were evaluated through 1000 bootstrap replications.

Phylogenetic trees were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) on online CIPRES services[76]. *Protaphorura armata* (Tullberg, 1869) from Europe was selected as the outgroup. To avoid the incorrect likelihood calculation, identical sequences were removed from the analyses. Best-fitting substitution models were assessed under the AIC and BIC criteria in jModelTest 2.1.7[77], and the GTR + I + Γ model selected for subsequently analyses. ML trees were reconstructed in RAxML v8.2.4[78] with the GTRGAMMAI model and 1000 bootstrap replicates. BI-analyses were conducted in MrBayes 3.2.6[79]. To avoid the problem of branch-length overestimation, the compound Dirichlet priors "brlenspr = unconstrained: gammadir (1, 1, 1, 1)" for branches lengths were incorporated[80]. The number of generations for the total analysis was set to 50 million, with the chain sampled every 5000 generations. The burn-in value was 25% and other parameters were set as default options. Evaluating effective sample size (ESS) values and state convergence were checked in Tracer 1.5[81].

Species delimitation. Species delimitation was performed using three data, morphology, mitochondrial marker COI, and geography. Using multiple approaches provided an acceptable error rate for species delimitation of 0.027[35].

Morphology. The specimens for morphological examination were cleared in lactic acid, mounted in Marc André II solution and then studied using a Nikon Eclipse 80i microscope. Primary discrimination for morphospecies relied on the most traditionally important diagnostic characters among *Protaphorura* species including the formula and the positions of the pseudocelli on the dorsal body and subcoxa 1 of the legs, the presence or absence of additional microchaetae on abdominal terga I–III and V, and the relative position of the prespinal microchaetae on abdominal tergum VI. Other minor additional characters were carefully examined for assessing their potential values including the number and arrangement of parapseudocelli, the number of ventral mesothoracic chaetae, the axial chaetae on abdominal terga IV–VI, and the number of chaetae on subcoxa 1 of legs I–III. Specimens which displayed bilateral asymmetry in key diagnostic features were considered deformed and not included in our analysis.

Molecular approaches. Both distance-based (ABGD) and evolutionary model-based (PTP) methods were employed for the single mitochondrial marker COI without a priori species hypothesis. Prior intraspecific divergence varied from 0.001 (Pmin, a single nucleotide difference) to 0.14 (Pmax, threshold value applied in Collembola of Churchill[64]). Relative gap width was set to 1, with 20 recursive steps, 40 bids for graphic histogram of distances, K2P model for distance calculation and other parameters as default. An unrooted ML tree was generated in RAxML v8.2.4 with the GTRGAMMA model and 1000 bootstrap replicates. Identical sequences were removed and Bayesian PTP (bPTP) delimitation was performed on single unrooted tree in python script bPTP.py v0.51[50]. A total of 5×10^5 generations were run with first 20% as burn-in. Maximum likelihood delimitation was also calculated in PTP.py 2.2 given 1000 trees derived from RAxML bootstrap analysis.

Geographical criterion. Because of the very low dispersal capabilities of springtails (except for some species which are in maritime islands and nunataks), vicariance and long-term persistence can occur within restricted geographical limits, even across distances of tens of kilometers[60, 61, 82]. Putative geographical species were delimited on the basis of primary hypotheses of morphospecies. For one morphospecies occurring in several geographical sites, populations, separated by distances greater than the hypothetical values (10–500 kilometers), were recognized as distinct putative species.

Species validation criteria. We used Generalized Lineage Concept (GLC) to determine the species boundary from our data. In this study, the operational criteria for final species recognition are conservative with trusted delimitations congruent across all analyses[37]: (i) monophyletic lineage (GLC) confirmed by phylogenetic inferences (it is also the a priori species hypothesis hidden in PTP); (ii) congruent across geographical and molecular delimitations; (iii) reliable morphological diagnosis with characters stably differing from other lineages. Lineages which met all of the above criteria were given a formal biological scientific name. Different lineages consistent across molecular and geographical analyses but of morphological stasis were considered cryptic species.

References

1. Bellinger, P. F., Christiansen, K. A. & Janssens, F. *Checklist of the Collembola of the World*. Available from: http://www.collembola.org/ (accessed 8 January 2016) (1996–2016).
2. Hopkin, S. P. *Biology of the Springtails (Insecta: Collembola)*. New York: Oxford University Press (1997).
3. Filser, J. The role of Collembola in carbon and nitrogen cycling in soil. *Pedobiologia* **46**, 234–245 (2002).
4. Johnson, D. *et al*. Soil invertebrates disrupt carbon flow through fungal networks. *Science* **309**(5737), 1047–1047 (2005).
5. Chamberlain, P. M., McNamara, N. P., Chaplow, J., Stott, A. W. & Black, H. I. J. Translocation of surface litter carbon into soil by Collembola. *Soil Biol. Biochem.* **38**, 2655–2664 (2006).
6. Rusek, J. Biodiversity of Collembola and their functional role in the ecosystem. *Biodivers. Conserv.* **7**, 1207–1219 (1998).
7. Chen, J. X., Ma, Z. C., Yan, H. J. & Zhang, F. Roles of springtails in soil ecosystem. *Biodivers. Sci.* **15**(2), 154–161 (2007).
8. Deharveng, L. Recent advances in Collembola systematics. *Pedobiologia* **48**, 415–433 (2004).
9. Baker, A. N. & Dunning, R. A. Association of populations of Onychiurid Collembola with damage to sugar-beet seedlings. *Plant Pathol.* **24**(3), 150–154 (1975).
10. Hurej, M., Debek, J. & Pomorski, R. J. Investigations on damage to sugar-beet seedlings by the springtail *Onychiurus armatus* (Collembola, Onychiuridae) in lower Silesia (Poland). *Acta Entomol. Bohem.* **89**, 403–407 (1992).
11. Joseph, S. V., Bettiga, C., Ramirez, C. & Soto-Adames, F. N. Evidence of *Protaphorura fimata* (Collembola: Poduromorpha: Onychiuridae) feeding on germinating lettuce in the Salinas Valley of California. *J. Econ. Entomol.* **108**(1), 228–236 (2015).
12. Pomorski, R. J., Furgoł, M. & Christiansen, K. Review of North American species of the genus *Onychiurus* (Collembola: Onychiuridae), with a description of four new species from caves. *Ann. Entomol. Soc. Am.* **102**(6), 1037–1049 (2009).
13. Sun, X., Chen, J. X. & Deharveng, L. Six new species of *Thalassaphorura* (Collembola, Onychiuridae) from southern China, with a key to world species of the genus. *Zootaxa* **2627**, 20–38 (2010).
14. Weiner, W. M. & Kaprus', I. J. Revision of Palearctic species of the genus *Dimorphaphorura* (Collembola: Onychiuridae: Onychiurinae: Oligaphorurini) with description of new species. *J. Insect Sci.* **14**, 1–30 (2014).
15. Babenko, A. B. & Fjellberg, A. Subdivision of the tribe Oligaphorurini in the light of new and lesser known species from North-East Russia (Collembola, Onychiuridae, Onychiurinae). *ZooKeys* **488**, 47–75 (2015).
16. Gisin, H. Notes sur les Collemboles, avec démembrement des espèces des Onychiurus armatus, ambulans, et fimetarius auctorum. *Mitteilungen der Schweizerischen Entomologischen Gesellschaft* **25**, 1–22 (1952).
17. Gisin, H. Nouvelles contributions au démembrement des espèces d'*Onychiurus* (Collembola). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft* **29**, 229–352 (1956).
18. Christiansen, K. A. The Collembola of Lebanon and Western Syria. Part I. General consideration and the family Onychiuridae. *Psyche* **63**, 119–133 (1956).
19. Selga, D. Tres especies nuevas de Colémbolos del puerto de Navacerrada (Guadarrama). *Publicaciones del Instituto de Biologia Aplicada Barcelona* **33**, 33–41 (1962).
20. Weiner, W. M. Onychiuridae of Poland. New species of *Protaphorura* Absolon, 1901 from the Tatra Mts. *Acta Zool. Crac.* **33**(18), 453–457 (1990).
21. Stach, J. The Apterygotan fauna of Poland in relation to the world-fauna of this group of insects. Family: Onychiuridae. Kraków: Panstwowe Wydawnictwo Naukowe (1954).
22. Salmon, J. T. Concerning the Collembola Onychiuridae. *Trans. Roy. Ent. Soc. Lond.* **111**, 119–156 (1959).
23. Choudhuri, D. K. Taxonomic observations on the British species of the *armatus* and *fimetarius* groups of the genus *Onychiurus* (Collembola: Onychiuridae). *Proc. Nat. Acad. Sci. India B* **34**, 260–282 (1964).
24. Bödvarsson, H. Studien über die Variation einiger systematischen Charaktere bei *Onychiurus armatus* (Tullberg, 1989) (Collembola). *Opuscula Entomol.* **24**, 225–245 (1959).
25. Bödvarsson, H. Studies of *Onychiurus armatus* (Tullberg) and *Folsomia quadrioculata* (Tullberg) (Collembola) with reference to morphology and taxonomy. *Opuscula Entomol.* **36**, 1–182 (1970).
26. Pomorski, R. J. Morphological-systematic studies on the variability of pseudocellae and some morphological characters in '*armatus*-group' (Collembola, Onychiuridae). Part I. *Onychiurus* (*Protaphorura*) *fimatus* Gisin 1952. *Polskie Pismo Entomol.* **56**(3), 531–556 (1986).
27. Pomorski, R. J. Morphological-systematic studies on the variability of pseudocelli and some morphoiogical characters in *Onychiurus* of the "*armatus*-group" (Collembola, Onychiuridae). Part II. On synonyms within the "*armatus*-group", with special reference to diagnostic characters. *Ann. Zool.* **43**(26), 535–576 (1990).
28. Rusek, J. A new location and type of pseudocelli in *Onychiurus* spp. (Collembola, Onychiuridae). *Annls. Soc. R. Zool. Belg.* **14**, 3–7 (1984).
29. Pomorski, R. J. *Onychiurinae of Poland (Collembola: Onychiuridae)*. Wrocklaw: Polish Taxonomical Society (1998).

30. Shaw, P., Faria, C. & Emerson, B. Updating taxonomic biogeography in the light of new methods – examples from Collembola. *Soil Organisms* **85**(3), 161–170 (2013).
31. Sun, X., Zhang, B. & Wu, D. H. Two new species and one new country record of *Protaphorura* Absolon, 1901 (Collembola: Onychiuridae) from northeast China. *Zootaxa* **3673**(2), 207–220 (2013).
32. Babenko, A. B. & Kaprus', I. J. Species of the genus *Protaphorura* (Collembola: Onychiuridae) described on material of Yu. I. Chernov from Western Taimyr. *Entomol. Rev.* **94**(4), 581–601 (2014).
33. Sun, X., Chang, L. & Wu, D. H. New species and new records of *Protaphorura* species from northeast China (Collembola: Onychiuridae). *Zootaxa* **3920**(2), 381–392 (2015).
34. de Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886 (2007).
35. Schlick-Steiner, B. C. *et al.* Integrative taxonomy: a multisource approach to exploring biodiversity. *Annu. Rev. Entomol.* **55**, 421–438 (2010).
36. de Queiroz, K. The general lineage concept of species, species criteria, and the process of speciation. In: Howard SJ, Berlocher SH, eds. *Endless Forms: Species and Speciation.* New York: Oxford University Press, 57–78 (1998).
37. Carstens, B. C., Pelletier, T. A., Reid, N. M. & Satler, J. D. How to fail at species delimitation. *Mol. Ecol.* **22**, 4369–4383 (2013).
38. Dayrat, B. Towards integrative taxonomy. *Biol. J. Linn. Soc.* **85**, 407–415 (2005).
39. DeSalle, R., Egan, M. G. & Siddall, M. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. B* **360**, 1905–1916 (2005).
40. Rubinoff, D., Cameron, S. & Will, K. A genomic perspective on the shortcomings of mitochondrial DNA for 'barcoding' identification. *J. Heredity* **97**, 581–594 (2006).
41. Will, K. W., Mishler, B. D. & Wheeler, Q. D. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* **54**, 844–851 (2005).
42. Will, K. W. & Rubinoff, D. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* **20**, 47–55 (2004).
43. Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. & Hebert, P. D. N. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* **103**, 968–971 (2007).
44. Hebert, P. D. N. & Gregory, T. R. The Promise of DNA Barcoding for Taxonomy. *Syst. Biol.* **54**(5), 852–859 (2005).
45. Hebert, P. D. N., Ratnasingham, S. & DeWaard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. Ser. B* **270**, 96–99 (2003).
46. Rougerie, R. *et al.* DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuária Brasileira* **44**(8), 789–802 (2009).
47. Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. A plea for DNA taxonomy. *Trends Ecol. Evol.* **18**, 70–74 (2003).
48. Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol. Evol.* **24**(2), 110–117 (2009).
49. Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864–1877 (2012).
50. Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876 (2013).
51. Pons, J. *et al.* Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Syst. Biol.* **55**(4), 595–609 (2006).
52. Yang, Z. & Rannala, B. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. USA* **107**, 9264–9269 (2010).
53. O'Meara, B. C. New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* **59**, 59–73 (2010).
54. Frati, F., Dell'Ampio, E., Casasanta, S., Carapelli, A. & Fanciulli, P. P. Large amounts of genetic divergence among Italian species of the genus *Orchesella* (Insecta, Collembola) and the relationships of two new species. *Mol. Phylogenet. Evol.* **17**, 456–461 (2000).
55. Porco, D., Bedos, A. & Deharveng, L. Description and DNA barcoding assessment of the new species *Deutonura gibbosa* (Collembola: Neanuridae: Neanurinae), a common springtail of Alps and Jura. *Zootaxa* **2639**, 59–68 (2010).
56. Soto-Adames, F. N. Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii's 'color pattern species'. *Mol. Phylogenet. Evol.* **25**, 27–42 (2002).
57. Schneider, C. & D'Haese, C. A. Morphological and molecular insights on *Megalothorax*: the largest Neelipleona genus revisited (Collembola). *Invertebr. Syst.* **27**(3), 317–364 (2013).
58. Katz, A. D., Giordano, R. & Soto-Adames, F. N. Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). *Zool. J. Linn. Soc.* **173**, 818–840 (2015).
59. Burkhardt, U. Species identification of Collembola by means of PCR-based marker systems. Unpulished doctoral Thesis, University of Bremen (2005).
60. Cicconardi, F., Nardi, F., Emerson, B. C., Frati, F. & Fanciulli, P. P. Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Mol. Ecol.* **19**, 386–400 (2010).
61. Garrick, R. C., Sands, C. J., Rowell, D. M., Hillis, D. M. & Sunnucks, P. Catchments catch all: long-term population history of a giant springtail from the southeast Australian highlands - a multigene approach. *Mol. Ecol.* **16**, 1865–1882 (2007).
62. Porco, D. *et al.* Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. *Invertebr. Syst.* **26**, 470–477 (2012).
63. Zhang, F. *et al.* Cryptic diversity, diversification and vicariance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. *Zool. Scr.* **43**, 393–404 (2014).
64. Porco, D., Skarżyński, D., Decaëns, T., Hebert, P. D. N. & Deharveng, L. Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area. *Mol. Ecol. Resour.* **14**, 249–261 (2014).
65. Weiner, W. M. Onychiurinae (Onychiuridae, Collembola) of North Korea: species of the *Paronychiurus flavescens* (Kinoshita, 1916) group. *Acta Zool. Cracoviensia* **32**(5), 85–92 (1989).
66. Pomorski, R. J. Two new species of *Protaphorura* Absolon, 1901 from north Karelia (Russia), with notes on the position of altered pseudocelli (psx) in the *armatus*-group (Collembola, Onychiuridae). *Genus* **4**(2), 121–128 (1993).
67. Babenko, A. B. Two new species of the genus *Uralaphorura* Martynova, 1978 (Collembola: Onychiuridae) from Siberia. *Zootaxa* **2108**, 37–44 (2009).
68. Sun, X. & Wu, D. H. Review of Chinese Oligaphorurini (Collembola, Onychiuridae) with descriptions of two new Palaearctic species. *ZooKeys* **192**, 15–26 (2012).
69. Sun, X., Deharveng, L. & Wu, D. H. Broadening the definition of the genus *Thalassaphorura* Bagnall, 1949 (Collembola, Onychiuridae) with a new aberrant species from China. *ZooKeys* **364**, 1–9 (2013).
70. Sun, X. & Li, Y. New Chinese species of the genus *Thalassaphorura* Bagnall, 1949 (Collembola: Onychiuridae). *Zootaxa* **3931**(2), 261–271 (2015).
71. Fjellberg, A. The Collembola of Fennoscandia and Denmark. Part I: Poduromorpha. *Fauna Entomol. Scandinavica* **35**, 1–183 (1998).
72. Babenko, A. B., Chimitova, A. B. & Stebaeva, S. K. New Palaearctic species of the tribe Thalassaphorurini Pomorski, 1998 (Collembola, Onychiuridae). *ZooKeys* **126**, 1–38 (2011).
73. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
74. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**(4), 772–780 (2013).

75. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
76. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*. New Orleans, 1–8 (2010).
77. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
78. Stamatakis, A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **30**(9), 1312–1313 (2014).
79. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
80. Zhang, C., Rannala, B. & Yang, Z. Robustness of compound Dirichlet priors for Bayesian inference of branch lengths. *Syst. Biol.* **61**, 779–784 (2012).
81. Rambaut, A. & Drummond, A. J. Tracer v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer (accessed 7 May 2014) (2007).
82. Cicconardi, F., Fanciulli, P. P. & Emerson, B. C. Collembola, the biological species concept and the underestimation of global species richness. *Mol. Ecol.* **22**, 5382–5396 (2013).

Acknowledgements

Author Contributions

Sun, X., Zhang, F., Li, Y. and Wu, D.H. conceived the experiments, Sun, X. and Ding, Y.H. conducted the experiments, Sun, X. and Zhang, F. analyzed the results and wrote the paper, Davies, T.W. provided critical revision of the article. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-08381-4

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.