Received: 19 August 2019 | Revised: 4 December 2019 | Accepted: 8 December 2019

DOI: 10.1111/zsc.12408

ORIGINAL ARTICLE

Zoologica Scripta | WILEY

Phylogeny of Neotropical Seirinae (Collembola, Entomobryidae) based on mitochondrial genomes

Nerivânia Nunes Godeiro[1,2,3] | **George Pacheco**[3] | **Shanlin Liu**[3,4] | **Nikolas Gioia Cipola**[5] | **Waldir Miron Berbel-Filho**[6] | **Feng Zhang**[2] | **Marcus Thomas P. Gilbert**[3] | **Bruno Cavalcante Bellini**[1]

[1]Department of Botany and Zoology, Biosciences Center, Federal University of Rio Grande do Norte, Natal, Brazil

[2]Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing, China

[3]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

[4]Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Plant Protection, China Agricultural University, Beijing, China

[5]Laboratório de Sistemática e Ecologia de Invertebrados do Solo, Instituto Nacional de Pesquisas da Amazônia—INPA, CPEN, Manaus, Brazil

[6]Department of Biosciences, College of Science, Swansea University, Swansea, UK

Correspondence
Nerivânia Nunes Godeiro, Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China.
Email: nerivania@gmail.com

Funding information
NNG was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES—Brazil) scholarship (PDSE, Process # BEX 6891/15-2). BCB received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq—Brazil) Universal, Process # 441451/2014-4; PQ2018, Process # 305426/2018-4. GP and WMB-F were funded by a Ciência Sem Fronteiras Full PhD Abroad Scholarship (GP grant: 201761/2014-9; WMB-F grant: 233161/2014-7) granted by the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). NGC is granted by CNPq (PCI-DB, Process # 300925/2019-0).

Abstract

Seirinae is one of the most diverse subfamilies of Collembola. To date no detailed phylogeny of Seirinae has been proposed, which leads to difficulties in the understanding of evolutionary patterns regarding this taxon. The main aim of this study is to clarify the phylogenetic relationships within the Neotropical Seirinae, by generating and analysing the mitochondrial genomes of 26 terminal taxa of Entomobryidae, and one species of Paronellidae. Specifically, we first generated Illumina HiSeq 2000 shotgun sequence data from each species, then reconstructed the mitochondrial genome of each species using two methods: MitoZ and MIRA/MITOBim. Using these data, we were able to generate a well-supported phylogeny that combined all the above species as well as three publicly available mitogenomes from other species. Bayesian and maximum likelihood methods were applied using all 13 protein coding genes. In this way, monophyly for the internal groups of Seirinae was obtained based on molecular evidence for the first time, as was the potential validity of three main internal taxa of the subfamily. We furthermore validated that *Tyrannoseira* is a distinct lineage and propose the elevation of *Lepidocyrtinus* to genus. Lastly, we anticipate that these newly available mitogenomes will serve as a useful dataset for future studies on the evolution of the Collembola and Hexapoda.

KEYWORDS

Entomobryoidea, Lepidocyrtinus, next generation sequencing, *Seira*, *Tyrannoseira*

1 | INTRODUCTION

Phylogenetic analyses that employ molecular techniques are commonly applied today across most of metazoan and plant main clades, but only in the recent years they become more common concerning the Collembola. Indeed, it was only recently that Zhang et al. (2014) reported the first detailed phylogeny for a collembolan family (Entomobryidae). Their results supported the monophyly of the subfamilies Entomobryinae (now also including species with scales), Lepidocyrtinae and Seirinae. The latter two were considered sister groups, although based on weak phylogenetic support. Subsequently, Zhang and Deharveng (2015b) reported a complementary study in which morphological characters were incorporated into the phylogenetic analyses. This study included additional species (a total of 86) in the morphological analysis, and a new character was suggested as diagnostic at the supra-generic level—the S-chaetae tergal pattern. Fusions of the previously recognized tribes were made, and Entomobryidae was subdivided into seven subfamilies (Nothobryinae, Orchesellinae, Heteromurinae, Bessoniellinae, Entomobryinae, Seirinae and Lepidocyrtinae). As for Seirinae, six genera (except *Seira*) were transferred to Entomobryinae. The genus *Tyrannoseira* was not represented in this investigation.

After some months, Zhang, Sun, Yu, and Wang (2015) published a phylogenetic study of the Entomobryoidea, which included nine species of Paronellidae and 35 of Entomobryidae, and concluded that the S-chaetae are more relevant as diagnostic character for the classification of Entomobryoidea than traditional characters, such as furca morphology and presence or absence of body scales. Interestingly, this phylogeny revealed Seirinae to be the sister group of Entomobryinae, contradicting what was previously suggested in 2014 when Seirinae and Lepidocyrtinae were considered to be sister groups.

Despite the importance of these initial studies, the limited range and restricted distribution of the taxa used within the Seirinae limited their ability to clarify its internal phylogenetic relationships. Additionally, the few loci used (at most four) may have not sufficiently overcome commonly found problems on mtDNA-based studies such as introgression and incomplete lineage sorting. Seirinae is one of the most diverse subfamilies of Collembola, comprised of two known genera: *Seira*, with about 200 nominal species; and *Tyrannoseira*, with five described species endemic to Brazil (Bellinger & Janssens, 2019; Cipola, Morais, Godeiro, & Bellini, 2019). Seirinae are well-characterized entomobryids which present: heavily ciliated scales distributed on body and appendages; falcate mucro without basal spine; in almost all species presence of 2/2/3 bothriotricha on abdominal segments II to IV, respectively, and formula of ordinary S-chaetae 11 | 022 + 3 and microsensilla (ms) 10|10100 (Soto-Adames, Barra, Christiansen, & Jordana, 2008; Szeptycki, 1979; Zhang & Deharveng, 2015b).

Analysis of complete mitochondrial genomes holds potential value for solving taxonomic disputes raised by morphological studies. This ability is due to several characteristics the mitochondrial DNA, such as high mutation rate, almost total absence of recombination and exclusive maternal inheritance. Moreover, the fact that any mitochondrial locus is typically presented in many more copies per cell than most nuclear DNA loci can facilitate its amplification and sequencing, which undoubtedly contributed to the widespread use of mitochondrial markers in phylogenetic reconstruction (Gillett et al., 2014).

Nonetheless, currently only 15 mitogenomes have been published for springtails (Carapelli, Comandi, Convey, Nardi, & Frati, 2008; Carapelli, Convey, Nardi, & Frati, 2014; Carapelli, Liò, Nardi, van der Wath, & Frati, 2007; Faddeeva-vakhrusheva et al., 2016, 2017; Nunes et al., 2019; Wu et al., 2017; Zhang et al., 2019) of which only four species belong to Entomobryidae: two of *Orchesella*, one of *Sinella* and one of *Lepidosira*. *Orchesella villosa* was published by Carapelli et al., 2007 together with other 99 mitogenomes of several Pancrustacea groups, and *Orchesella cincta* was accompanied by a detailed description of the first complete genome published for Collembola (Faddeeva-Vakhrusheva et al., 2016). The nuclear genome of *S. curviseta* was published in January 2019 and is the longest of the collembolan genomes assembled until now, with a total length of 381,46 Mb and a mitochondrial genome length of 14,840 bp (Zhang et al., 2019). *Lepidosira neotropicalis* Nunes and Bellini 2019 was sequenced together with the other mitogenomes published in the present study, but the detailed mitogenome was published with the species description (Nunes et al., 2019).

Giving the extensive taxonomic uncertainties and frequent cryptic diversity within Collembola groups (Ding, Yu, Guo, Li, & Zhang, 2018; Katz, Giordano, & Soto-adames, 2015; Porco et al., 2012), this study was performed with the intention to the following: (a) reconstruct a phylogeny for Neotropical Seirinae based on mitogenomes; (b) test the hypothesis that Seirinae do not include taxa with bidentate mucro (e.g. *Lepidocyrtoides*) and (c) verify the potential monophyly of Neotropical genera of the subfamily. This study is a first important step in understanding the evolution of Seirinae and may guide future revisions of the subfamily. It also uses a different approach (the use of mitogenomes) to investigate the evolution of part of Entomobryidae, a topic recently well studied based in other techniques.

2 | METHODS

2.1 | Taxon sampling and processing

The mitogenomes of one sample of Paronellidae (*Trogolaphysa* sp.) and 26 samples of Entomobryidae: *Lepidocyrtus* (1 sp.), *Lepidocyrtoides* (2 spp.), *Entomobrya*

(1 sp.), *Seira* (17 spp.) and *Tyrannoseira* (2 spp.) were sequenced in this study. Further details about the species are listed in Table 1, and their respective sampling localities are shown on the map of Figure 1. Specimens were stored postcollection at −20°C in 100% ethanol and prior to genetic analyses were initially analysed to confirm their overall conservation and morphological integrity. Some specimens of each sample were mounted on glass slides following Mari-Mutt (1979) with some modifications for taxonomic identification. All glass slides mounted were identified with the same code of samples in ethanol and are stored at Collembola Collection/UFRN and Soil Invertebrate Systematics and Ecology Laboratory/INPA. The material source is included in Brazilian SisGen online databank (accession number A05460B).

2.2 | DNA extraction

Total genomic DNA was extracted from a single individual from each vial containing several specimens belonging to a given species following the protocol provided by Cell and Tissue DNA Kit (Qiagen) with some modifications: incubation time was increased to 48h with extra addition of 15 µl of proteinase K after the first 24 hr. Before DNA isolation the specimen's cuticle was recovered and maintained in 70% ethanol for further morphological analyses. Purification was automated using KingFisher Duo Prime Purification System (Thermo Fisher Scientific), and the purified genomic DNA was eluted in 30 µl and quantified using Qubit 2.0 Fluorometer (Life Technologies).

2.3 | Libraries construction and sequencing

Genomic DNA was fragmented to an average length of 300-400 bp using Bioruptor (Diagenode). To verify the fragments size and concentration, automatized electrophoresis was made on Tape Station 2200 (Agilent Technologies). Libraries were prepared using NEBNext kit (Biolabs) according to the Illumina multiplex protocol of Meyer and Kircher (2010) using double-indexed library adapters. The quantity (X) of PCR amplification cycles required to generate sufficient library for sequencing was determined by a quantitative real-time PCR performed using Mx3000P (Strategene). Amplification reactions were performed in 50 µl of reaction mixture containing 0.4 µl of 25 mM dNTP, 5 µl of buffer (10 concentrated), 5 µl of 25 mM MgCl2, 1 µl of each primer (10 µM/µl), 1 µl of a forward primer (10 µM/1 µl), 5 µl of total DNA, 1 µl of TaqGold DNA polymerase and 31.6 µl of distilled water to complete to final volume. The cycling profile was 95°C for 10 min, followed by X cycles of 95°C for 30 s, 60°C for 45 s, 72°C for 45 s and 72°C for 5 min during the final cycle. A final purification step was made using QIAquick PCR purification kit (Qiagen), and the quantification and size estimation of libraries were conducted on a Tape Station 2200 (Angilent Technologies). Libraries were pooled and sequenced at the Danish National High-Throughput DNA Sequencing Centre on one lane of Illumina HiSeq 2000, using 80 paired-end cycles with a rapid run. FastQC was used for quality control of raw sequence data. Adapter removal (Schubert, Lindgreen, & Orlando, 2016) was conducted to remove reads containing adapters and low-quality reads from the raw data.

2.4 | Mitogenome assembly

As no closely related Collembola mitochondrial genome was available (of the same genera or even subfamily) to be used as a reference genome, we applied two methodologies to *de novo* reconstruct the mitochondrial genomes of interest based on our shotgun data.

The first methodology reconstructed 19 genomes using MitoZ v. 1.04 (Meng, Li, Yang, & Liu, 2019), and the second reconstructed eight using Mira v. 4.0 (Chevreux, Wetter, & Suhai, 1999) and MITObim 1.8 (Hahn, Bachmann, & Chevreux, 2013). The methodology and the reference genome used for each sample are described in Table 1. Reference genomes used in the second reconstruction resulted from the first methodology. To confirm the identity of the new genomes, BLAST (basic local alignment search tool) searches were conducted on the National Center for Biotechnology Information website (*NCBI*; https://blast.ncbi.nlm.nih.gov/Blast.cgi) and all samples were similar to other Collembola sequences available. Annotations and the visualization of the circular genomes were made with MitoZ v. 1.04. The new 27 mitogenome sequences have been deposited at GenBank (see Table S1), as well as the raw data that are deposited at the SRA (Sequence Read Archive), available at the NCBI portal.

2.5 | Phylogenetic analyses

For phylogenetic reconstruction, the 13 protein coding genes (NAD1, NAD2, NAD3, NAD4, NAD4L, NAD5, NAD6, COX1, COX2, COX3, ATP6, ATP8 and CYTB) of the 27 samples sequenced here, plus *Lepidosira neotropicalis*, *Sinella curviseta* and *Orchesella cincta* (accession numbers: MF716603, MK014212 and KT985987, respectively) were used. All 13 protein sequences were aligned separately with MAFFT 7 website (http://mafft.cbrc.jp/alignment/server/index.html). The best partition model for the ML and BI analyses was found using IQ-Tree Model Finder (Kalyaanamoorthy, Minh, Wong, Haeseler, & Jermiin, 2017), and it is detailed in Table S2. ML analyses

TABLE 1 List of speciess equenced in this work and their respective sampling localities in Brazil

Family/ Subfamily/ Species	Collection site	Genome assembly method	Reference genome	Average coverage depth (reads)	Ratio of mitogenome-derived reads (%)
Entomobryidae/Entomobryinae					
1. *Entomobrya* sp.	Águas de Lindoia/SP	MitoZ	N/A	281.97	0.31
2. *Lepidocyrtoides* sp.	Abaíra/BA	MitoZ	N/A	118.15	0.13
3. ***Lepidocyrtoides caeruleomaculatus*** Cipola & Bellini, 2017	Rio Preto da Eva/AM	*MIRA/MITOBim*	*Entomobrya* sp.	11.42	0.04
4. *Lepidocyrtus* sp.	Riacho da Cruz/RN	MitoZ	N/A	10.34	0.05
Entomobryidae/ Seirinae					
5. *Seira atrolutea* (Arlé, 1939)	Bariri/SP	MitoZ	N/A	50.14	0.04
6. *Seira brasiliana* (Arlé, 1939)	Medianeira/PR	*MIRA/MITOBim*	*Seira ritae*	27.90	0.09
7. ***Seira coroatensis*** Godeiro & Bellini, 2015	Almino Afonso/RN	*MIRA/MITOBim*	*Seira ritae*	16.27	0.07
8. ***Seira diamantinae***[*] Godeiro & Bellini, 2015	Abaíra/BA	MitoZ	N/A	59.86	0.14
9. *Seira dowlingi* (Wray, 1953)	Vera Cruz/RN	*MIRA/MITOBim*	*Seira ritae*	28.23	0.12
10. ***Seira harena***[*] Godeiro & Bellini, 2014	Areia/PB	*MIRA/MITOBim*	*Seira diamantinae*	13.01	0.03
11. *Seira mendoncae* Bellini, & Zeppelini, 2008	Abaíra/BA	MitoZ	N/A	54.18	0.13
12. ***Seira paraibensis***[*] Bellini & Zeppelini, 2009	Areia/PB	MitoZ	N/A	46.57	0.13
13. *Seira paulae* Cipola, Morais, & Bellini, 2014	Londrina/PR	MitoZ	N/A	37.93	0.12
14. *Seira potiguara* Bellini, Fernandes, & Zeppelini, 2010	Abaíra/BA	*MIRA/MITOBim*	*Seira ritae*	17.96	0.04
15. *Seira ca. prodiga*[*]	Abaíra/BA	MitoZ	N/A	48.97	0.15
16. *Seira ca. prodiga*[*]	Quixadá/CE	MitoZ	N/A	37.70	0.13
17. *Seira ca. prodiga*[*]	Lajes/RN	MitoZ	N/A	44.42	0.20
18. *Seira ritae* Bellini & Zeppelini, 2011	Lajes/RN	MitoZ	N/A	145.09	0.14
19. *Seira* sp. 1	Riacho da Cruz/RN	*MIRA/MITOBim*	*Seira ritae*	21.83	0.03
20. *Seira* sp. 2[*]	Pedro II/PI	MitoZ	N/A	55.55	0.12
21. ***Seira dapeste***[*] Bellini, Santos, & Souza, 2018	Natal/RN	MitoZ	N/A	2.383.893	0.56
22. *Seira* sp. 3	Caracarai/RR	MitoZ	N/A	109.03	0.26

(Continues)

TABLE 1 (Continued)

Family/ Subfamily/ Species	Collection site	Genome assembly method	Reference genome	Average coverage depth (reads)	Ratio of mitogenome-derived reads (%)
23. *Seira tinguira* Cipola, Morais, & Bellini, 2014	Londrina/PR	*MIRA/MITOBim*	*Seira ritae*		0.23
24. *Tyrannoseira bicolorcornuta* (Bellini, Pais, & Zeppelini, 2009)	Apodi/RN	MitoZ	N/A	128.52	0.34
25. *Tyrannoseira bicolorcornuta* (Bellini, Pais, & Zeppelini, 2009)	Sapé/PB	MitoZ	N/A	311.10	0.28
26. *Tyrannoseira raptora* (Zeppelini & Bellini, 2006)	Bananeiras/PB	MitoZ	N/A	247.54	0.37
Paronellidae/Paronellinae					
27. *Trogolaphysa* sp.	Abaíra/BA	MitoZ	N/A	50.14	0.11

Note: In bold are represented species with material collected from type localities, and the species belonging to *Lepidocyrtinus* genus are marked with an asterisk. The method used to reconstruct the genomes, the reference genome and the mitogenome coverage stats are listed as well.

All nominal species of Seirinae with except of *S. brasiliana* are endemic to Brazil.

Abbreviations: AM, Amazonas; BA, Bahia; CE, Ceará; PB, Paraíba; PI, Piauí; PR, Paraná; RN, Rio Grande do Norte; RR, Roraima; SP, São Paulo, all Brazilian federal states. N/A, not applicable.

These species marked with "*" belongs to the new genus proposed in this paper (*Lepidocyrtinus*).

were conducted using IQ-Tree 1.6.10 (Nguyen, Schmidt, Haeseler, & Minh, 2015). Branch support was estimated using Ultrafast option for bootstrap analysis, with 1,000 replicates. The model mitochondrial vertebrate—mtREV (Adachi & Hasegawa, 1996), was used for BI analyses performed on MrBayes v.3.2.6 (Huelsenbeck & Ronquist, 2001) with number of generations 20M; relative burn-in (fractions of samples discarded) of 0.25; print frequency 2,000; sample frequency 2,000 and number of chains 4. Evaluating effective sample size (ESS) values and state convergence were checked in Tracer 1.6.0. (Andrew Rambaut, Drummond, Xie, Baele, & Suchard, 2018), as well the burn-in value. The resulting phylogeny was visualized in FigTree v.1.4.2. (Rambaut, 2016) and edited using Adobe Illustrator CC 2017.

3 | RESULTS AND DISCUSSION

A total of 27 collembolan mitogenomes were recovered, of which 19 were assembled using a de novo method (Meng et al., 2019), and eight were obtained using a reference-based method (Hahn et al., 2013) (Table 1). We retrieved mitogenomes of full length for most of the species (18) and delivered 13 protein coding genes (PCGs) for the remaining species (Figures S1–S27). The length of the mitogenomes ranged from 14,565–15,364 bp, similar to other reported collembolan mitogenomes (Carapelli et al., 2008, 2014, 2007; Faddeeva-vakhrusheva et al., 2016, 2017; Nunes et al., 2019; Wu et al., 2017; Zhang et al., 2019). Mapping statistics including ratio of mitogenome-derived reads and average depth for all species is summarized in Table 1. On average reads derived from mitogenomes accounted for 0.16% in genomic DNA extracts and reached a depth of 79× coverage for all the assemblies (excluding the sample of *Seira dapeste* which had an overload of reads during the sequencing process). The final mitogenomes have been uploaded to Genbank, as well as the raw data are deposited at the SRA (Sequence Read Archive), available at the NCBI portal. Accession numbers are provided in the Table S1.

We calculated the pairwise p-distance (proportion of nucleotide sites at which two compared sequences are different) and G + C content for each PCG showing that ATP8 gene possesses the highest interspecies genetic variances (Figure 2) with the lowest G-C contents (Figure 3). Interestingly, the COI gene—widely adopted as the DNA barcode marker—possesses smallest interspecies genetic variances (Figure 2) but the highest G + C contents (Figure 3).

Bayesian (BI) and maximum likelihood (ML) analyses using the reconstructed mitochondrial genomes and three previously published sequences (*Lepidosira neotropicalis*, *Sinella curviseta* and *Orchesella cincta*) resulted in a

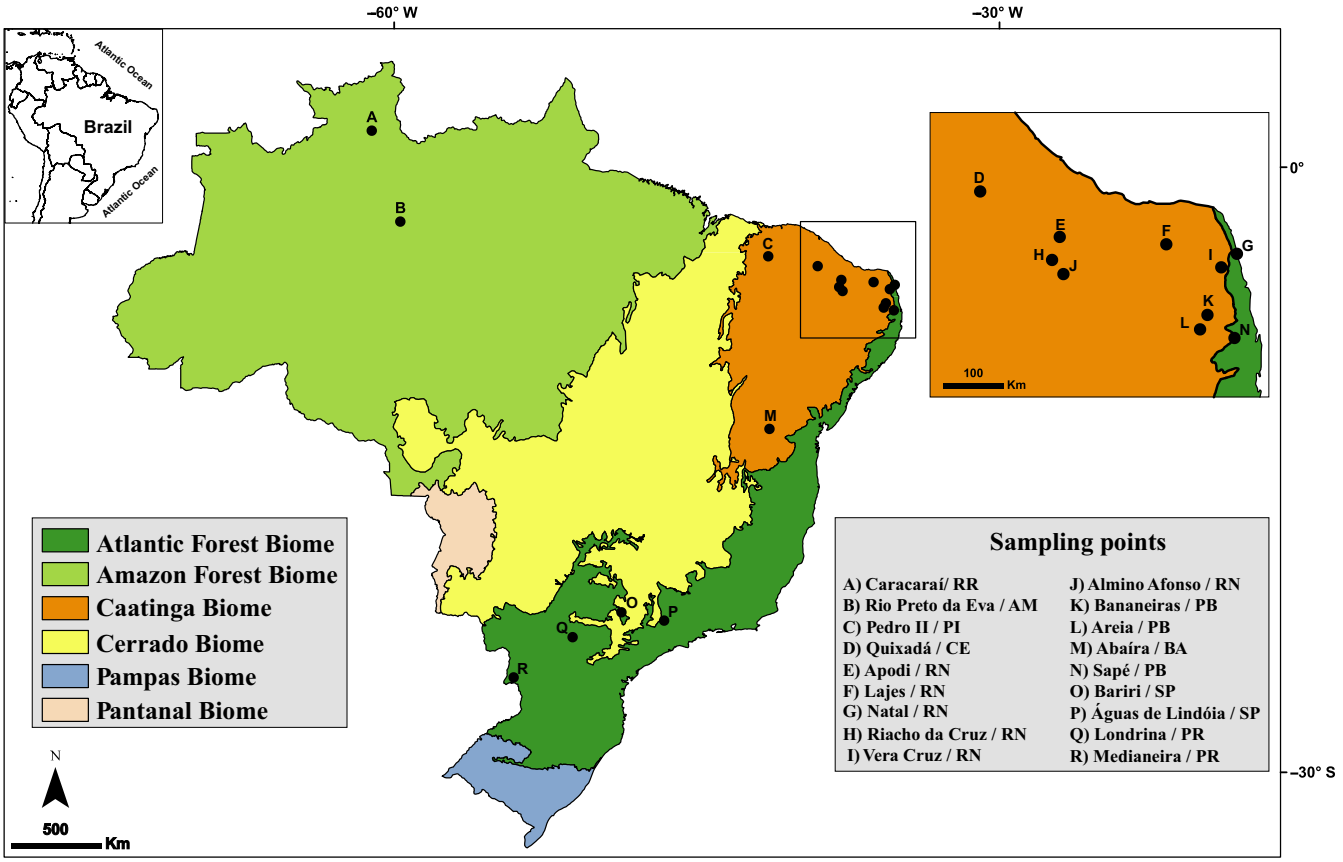

FIGURE 1 Map of Brazilian biomes showing the source areas of samples, highlighting nearby locations in northeastern region. Species from Atlantic and Amazon Forest biomes were collected from forested areas or their surroundings. Species collected from Caatinga Biome were collected from forested areas, exposed soil or over rocks. AM, Amazonas; BA, Bahia; CE, Ceará; PB, Paraíba; PI, Piauí; PR, Paraná; RN, Rio Grande do Norte; RR, Roraima and SP, São Paulo, all Brazilian states [Colour figure can be viewed at wileyonlinelibrary.com]

consensus tree with high support values for the majority of the nodes (Figure 4). The single Paronellidae species (*Trogolaphysa* sp.) was recovered as a sister group of Entomobryinae with medium support (Bootstrap = 50% and PP = 0.95). Seirinae was presented as a monophyletic taxon (Bootstrap = 100% and PP = 1), including only species with scales and falcate mucro. *Lepidosira* and *Lepidocyrtoides* were found to group within Entomobryinae (Bootstrap = 98% and PP = 1, in Figure 4), and Seirinae was recovered as the sister group of Lepidocyrtinae (*Lepidocyrtus* sp.), rather than Entomobryinae (Bootstrap = 92% and PP = 0.98).

Within the Neotropical Seirinae, three clades were recovered with high support in all nodes. The first major event of internal cladogenesis in the taxon occurred with the appearance of two groups of species: those mostly with modified chaetae on dorso-distal manubrium and proximal dens, with extra anterior macrochaetae on mesothorax and large lateral teeth on ungues (indicated as *Lepidocyrtinus* in Figure 4) and those lacking such structures. The second major event of cladogenesis occurred with the emergence of species with clear male dimorphic legs, belonging to *Tyrannoseira* genus, compared with the other members of *Seira* s. str.

The reconstructed phylogeny (Figure 4) corroborates the results of Zhang et al. (2014) in that species (eg. *Lepidosira*, *Lepidocyrtoides*) with strongly ciliated scales, S-chaetotaxy formula 22|122 + 3 dense coverage of dorsal macrochaetae and bidentate mucro of Seirini sensu Soto-Adames et al., 2008 belong to Entomobryinae sensu Zhang & Deharveng, 2015b. This observation reinforces the hypothesis that densely ciliated scales with different apex morphologies emerged independently among different Entomobryidae taxa (Zhang et al., 2014, 2015; Zhang & Deharveng, 2015a) or are derived from a basal group (probably close related to Heteromurini) and were secondarily lost in different branches of Entomobryinae and Paronellinae. The condition of a high number of dorsal macrochaetae in both Seirinae and Entomobryinae is somehow plesiomorphic, and for this reason could not support the inclusion of *Lepidosira* and *Lepidocyrtoides* aside with the Seirinae with falcate mucro. This can be explained by the fact that basal Entomobryidae such as Capbryinae and Nothobryini sensu Soto-Adames et al., 2008 (Nothobryinae sensu Zhang & Deharveng, 2015b) also shows dense coverage of multiciliated macrochaetae, a feature that has been further reduced in some taxa such as Heteromurinae and

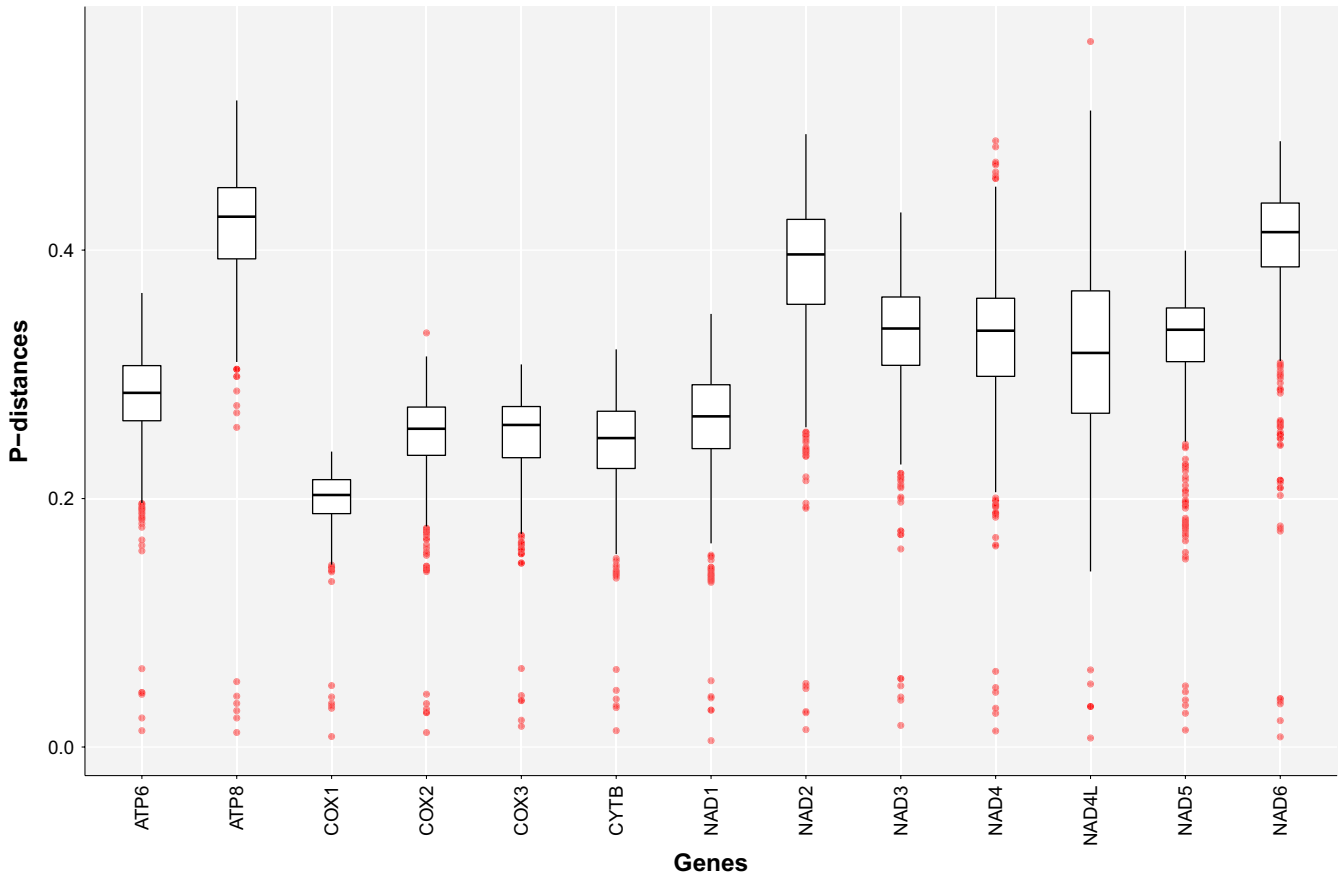

FIGURE 2 Boxplot showing the P-distances within each of the 13 genes analysed across all the 27 samples studied. Outlier values are depicted as red points [Colour figure can be viewed at wileyonlinelibrary.com]

Lepidocyrtinae (Baquero-Martín, Jordana, & Christiansen, 2004; Cipola, Oliveira, Morais, & Bellini, 2016; Jordana, 2012; Zhang & Deharveng, 2015a).

The monophyly of Seirinae that is evident from our results is reaffirmed by a clear diagnosis among the Entomobryoidea of the following combination of characters: antennae with four segments; fourth antennomere bearing an apical bulb; postantennal organ (PAO) absent; 8 + 8 eyes present; strongly ciliate scales with rounded or lanceolate apexes covering dorsal body and appendices (including ventral part of furcula); dorsal macrochaetotaxy moderately plurichaetotic, and most macrochaetae present a truncated apex; chaetae m4 and p4 on second abdominal segment absent in adult specimens (possibly modified into scales during development); bothriotricha formula from second to fourth abdominal segments 2/3/3 (except in *S. rowani* Yosii, 1959, with 2/3/2); S-chaetae formula from Th. II to fifth abdominal segment 11 | 022–3 and the microsensilla 10 | 10100; fourth abdominal segment at least 2.5 times longer than third in the body midline; dens dorsally crenulated without spines and falcate mucro separated from dens, without mucronal chaeta (Christiansen & Bellinger, 2000; Soto-Adames et al., 2008; Szeptycki, 1979; Zhang & Deharveng, 2015b). Although none of these characters alone is synapomorphic of Seirinae, their combination delimits the taxon. On the other hand, recent morphological and phylogenetic reviews indicate that diagnoses in Entomobryinae and Lepidocyrtinae are challenging. Both taxa are apparently paraphyletic as they may include different lineages of Paronellidae. In this sense, the Paronellinae possibly emerged from one (or more) of the Entomobryinae lineages, and the Cyphoderinae are derived from some form of Lepidocyrtinae, possibly *Pseudosinella* (Szeptycki, 1979; Zhang et al., 2015), or at least both Cyphoderinae and Lepidocyrtinae are sister groups. Our data also corroborate the paraphyly of Entomobryinae, since the sole species of Paronellinae used in our analysis, *Trogolaphysa* sp., is apparently related to Entomobryinae taxa.

Our analysis shows Lepidocyrtinae as the sister group of Seirinae, with high support (Bootstrap = 92% and PP = 0.98). This condition is in disagreement with the most recent Entomobryoidea phylogeny studies (Zhang et al., 2014; Zhang, Ma, & Greenslade, 2017). Our results cannot be taken as conclusive since we used a single species of Lepidocyrtinae in our analysis (*Lepidocyrtus* sp.), and we could not include representatives of all subfamilies of Entomobryidae. Morphologically, this relationship may be related to the loss of S-chaetae on the second and third thorax and first abdominal segments (Zhang & Deharveng,

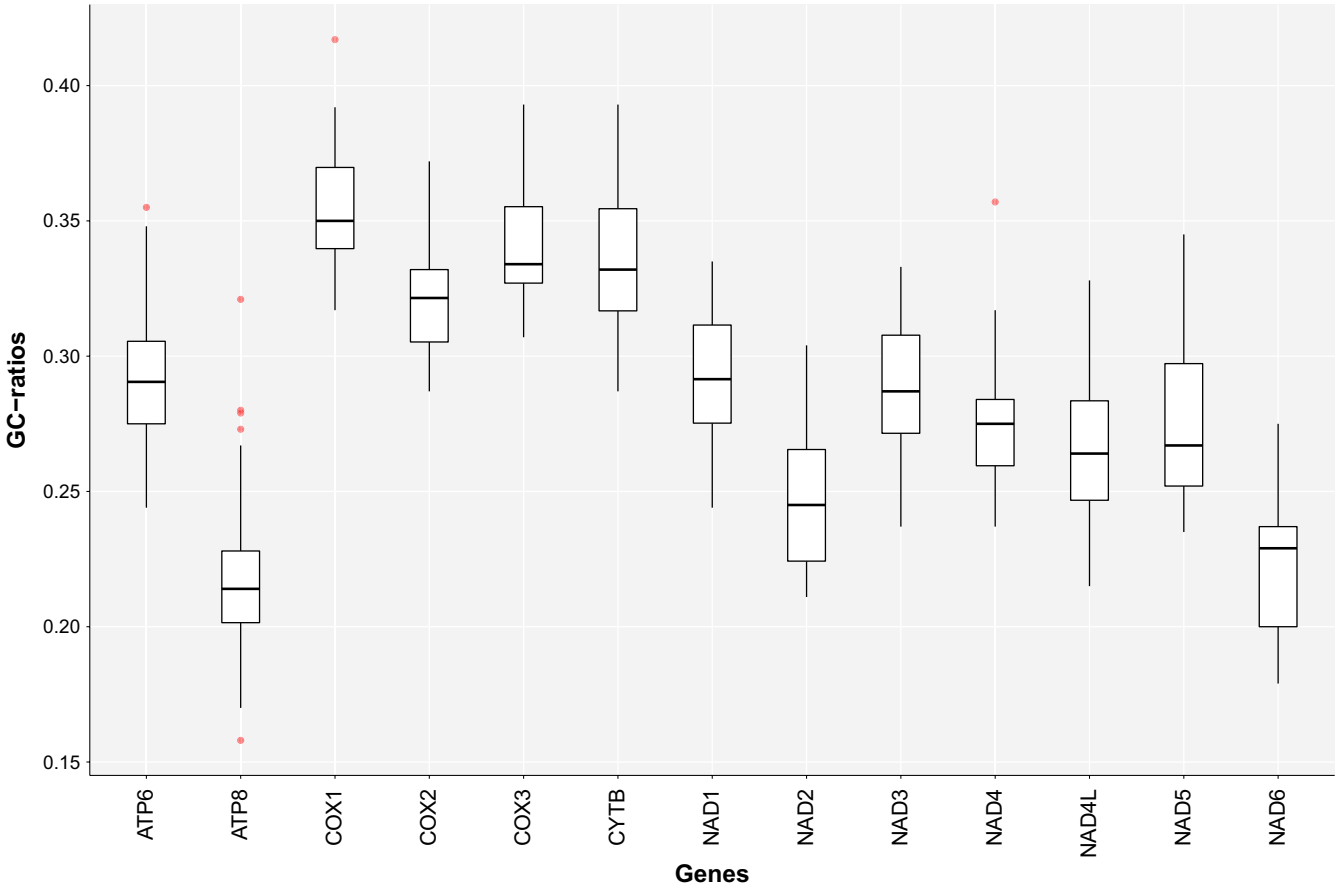

FIGURE 3 Boxplot showing the GC-ratio within each of the 13 genes analysed across all the 27 samples studied. Outlier values are depicted as red points [Colour figure can be viewed at wileyonlinelibrary.com]

2015b). Also, the lack of nuclear genes in our analysis may have contributed to the uncertainty about the relations of deeper branches of Entomobryoidea, as subfamilies affinities. In this sense, our aim in this study was to verify the internal relationships of Seirinae only. We believe the addition of a greater number of species of other Entomobryidae subfamilies, more outgroups and the use of nuclear genes in the analysis could clarify the subfamilies affinities in the family.

The subgenus *Lepidocyrtinus* sensu Yosii, 1959 was validated here with a maximum value of posterior probability (1) and medium bootstrap value (72) (Figure 4), and its main characteristics are the presence of macrochaetae on manubrial plates and on dorso-proximal region of dens of most species, and enlarged lateral teeth on ungues. Also, most Neotropical species present enlarged Th. II with extra multiples of anterior chaetae above m1-2 complex and antennae longer than trunk. *Lepidocyrtinus* was created by Börner (1903) as a subgenus of *Lepidocyrtus*, but in subsequent morphological revisions, the group was considered a variation of *Seira* by its falcate mucro and presence of 3 + 3 bothriotricha on fourth abdominal segment (Szeptycki, 1979; Yosii, 1959). Thus, other members of *Lepidocyrtinus* were later integrated into *Seira* such as its Neotropical species (Christiansen & Bellinger, 2000; Godeiro & Bellini, 2014, 2015). In fact, the Neotropical taxa of *Lepidocyrtinus* were first designated in *Ctenocyrtinus*, a genus that was later synonymized with *Seira* by (Christiansen & Bellinger, 2000). The data presented in Figure 4 supports *Lepidocyrtinus* as a monophyletic taxon. In this context, we therefore raise *Lepidocyrtinus* to generic status here.

Tyrannoseira, the most recently described genus of Seirinae, was recovered as a monophyletic taxon and the sister group of *Seira* s. str. with high support. Such data support the view of *Tyrannoseira* as a full genus within the Seirinae as recently proposed (Bellini & Zeppelini, 2011; Cipola et al., 2019). The main feature that differentiates *Tyrannoseira* from Neotropical *Seira* s. str. is the male sexual dimorphism on the first pair of legs. Males have the first pair of legs modified into prehensile structures: femora are variably enlarged and possess a series of short, multiciliated spines in inner projection; tibiotarsus has an inner row of long ciliated spines (Bellini & Zeppelini, 2011; Christiansen & Bellinger, 2000; Mari-Mutt, 1986). Although some species of the *domestica* group of Old World *Seira* (e.g. *S. domestica*, *S. mantis* and *S. uwei*) also present sexual dimorphism related to the first pair of legs (Cipola et al., 2018), such species are widely different

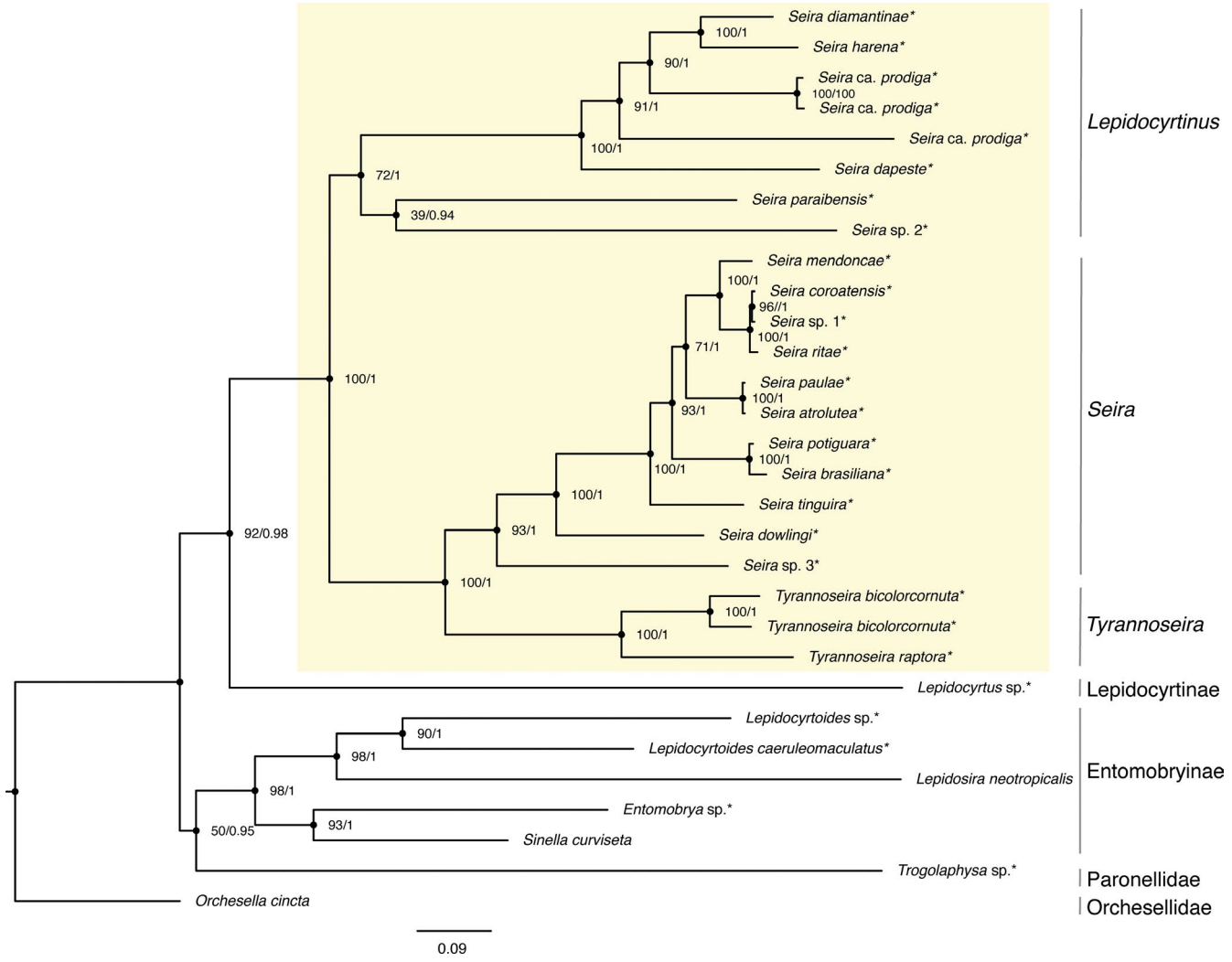

FIGURE 4 Phylogenetic relationships among 22 taxa of Seirinae (highlighted in yellow) and eight outgroup taxa. The tree was inferred from the dataset of the 13 mitochondrial protein coding genes. The ML inference made using IQ-TREE under the partition schemes and best-fit models selected by IQ-TREE Model Finder. First node numbers show bootstrap support values, and second node numbers show posterior probabilities values from Bayesian analyses performed on MrBayes [Colour figure can be viewed at wileyonlinelibrary.com]

from *Tyrannoseira* in dorsal chaetotaxy, especially on Th. III and Abd. I, which have several extra macrochaetae in *domestic*-group species, while in *Tyrannoseira,* this pattern is reduced (Cipola et al., 2019). Also, the femora are not clearly dilated in these species, as well as in any other Entomobryoidea genus. Adult specimens of *Tyrannoseira* always have the first abdominal segment devoid of macrochaetae, a less common feature in *Seira* s. str. The absence of macrochaetae in this segment is either homoplasious or the primitive condition of *Seira* s. str.+*Tyrannoseira* clade. In the phylogeny presented in Figure 4, it is observed that *Seira* sp. 3, which does not have macrochaetae in the Abd. I, is the closest species to *Tyrannoseira* clade.

Following our revision, a future step to better understand the evolution within Seirinae will be the inclusion of species from other regions. Concerning *Seira*, for instance, the *domestica* group and other Palaeartic taxa present a more complex macrochaetotaxy (as described in Cipola et al., 2018), which suggests they could belong to a more primitive branch of the genus. Inclusions like this will help us to understand which species are more related to the primitive stock of each genera and the subfamily itself and whether Seirinae can be subdivided into other well-supported genera or subgenera.

ACKNOWLEDGEMENTS

The authors thank the staff of the Danish National High-Throughput Sequencing Centre for technical assistance.

AUTHOR CONTRIBUTIONS

N.N.G., B.C.B., G.P. and M.T.P.G. designed the research; N.N.G. performed the laboratory work assisted by G.P.; N.N.G., G.P., S.L. and F. Z. analysed and interpreted the molecular data; N.N.G. and B.C.B wrote the draft paper with

input from G.P., M.T.P.G., S.L., W.M.B-F, F. Z. and N.C.; all authors revised and approved the final manuscript.

ORCID

Nerivânia Nunes Godeiro https://orcid.org/0000-0002-1669-6124
Nikolas Gioia Cipola https://orcid.org/0000-0002-7594-9574
Waldir Miron Berbel-Filho https://orcid.org/0000-0001-6991-4685
Feng Zhang https://orcid.org/0000-0002-1371-266X
Bruno Cavalcante Bellini https://orcid.org/0000-0001-7881-9436

REFERENCES

Adachi, J., & Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution*, *42*(4), 459–468. https://doi.org/10.1007/BF02498640

Arlé, R. (1939). Collembola, anexo N. 2, ao relatório da excursão científica do Instituto Oswaldo Cruz realizada na zona da E. F. N. O. B., em outubro de 1938. *Boletim Biológico*, *4*(2), 295–300.

Baquero-Martín, E., Jordana, R., & Christiansen, K. (2004). Redescription of *Nothobrya schubarti* Arlé, 1961 (Collembola, Entomobryomorpha). *Entomological News*, *115*, 31–34.

Bellinger, P. F., Christiansen, K. A., & Janssens, F. (2019). *Checklist of the Collembola of the world*. Retrieved from http://www.collembola.org.

Bellini, B. C., Santos, N. M. C., Souza, P. G. C., & Weiner, W. M. (2018). Two new species of Brazilian springtails (Hexapoda: Collembola) with comments on Neotropical *Brachystomella* Ågren and Seira (Lepidocyrtinus) Börner. *Insect Systematics & Evolution*, 1–30. https://doi.org/10.1163/1876312X-00002192

Bellini, B. C., Pais, A. P., & Zeppelini, D. (2009). A new species of *Seira* Lubbock (Collembola: Entomobryidae) from Brazil with sexually dimorphic legs. *Zootaxa*, *2080*, 38–46.

Bellini, B. C., & Zeppelini, D. (2008). A new species of *Seira* (Collembola: Entomobryidae) from northeastern Brazil. *Zoologia*, *25*, 724–727.

Bellini, B. C., & Zeppelini, D. (2009). A new species of *Seira* Lubbock (Collembola, Entomobryidae), with a key to the species of Paraíba, Brazil. *Revista Brasileira de Entomologia*, *53*, 266–271.

Bellini, B. C., Fernandes, L. H., & Zeppelini, D. (2010). Two new species of *Seira* (Collembola, Entomobryidae) from Brazilian Coast. *Zootaxa*, *2448*, 53–60.

Bellini, B. C., & Zeppelini, D. (2011). New genus and species of Seirini (Collembola, Entomobryidae) from Caatinga Biome, Northeastern Brazil. *Zoosystema*, *33*, 545–555. https://doi.org/10.5252/z2011n4a6

Carapelli, A., Comandi, S., Convey, P., Nardi, F., & Frati, F. (2008). The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). *BMC Genomics*, *9*, 315. https://doi.org/10.1186/1471-2164-9-315

Carapelli, A., Convey, P., Nardi, F., & Frati, F. (2014). The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda: Collembola), and an update on the phylogeny of the collembolan lineages based on mitogenomic data. *Entomologia*, *2*(190), 46–55.

Carapelli, A., Liò, P., Nardi, F., Wath, E., & Frati, F. (2007). Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evolutionary Biology*, *7*(Suppl 2), S8. https://doi.org/10.1186/1471-2148-7-S2-S8

Chevreux, B., Wetter, T., & Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. *German Conference on Bioinformatics*, *99*, 45–56. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.7465&rep=rep1&type=pdf

Christiansen, K. A., & Bellinger, P. (2000). A Survey of the Genus *Seira* (Hexapoda: Collembola: Entomobryidae). *Sciences-New York*, *36*(1), 39–75.

Cipola, N. G., Arbea, J., Baquero, E., Jordana, R., Morais, J. W., & Bellini, B. C. (2018). The survey of *Seira* Lubbock, 1870 (Collembola, Entomobryidae, Seirinae) from Iberian Peninsula and Canary Islands, including three new species. *Zootaxa*, *4458*(1), 1–66. https://doi.org/10.11646/zootaxa.4458.1.1

Cipola, N. G., Morais, J. W., & Bellini, B. C. (2014). Two new species of *Seira* Lubbock (Collembola, Entomobryidae, Seirini) from South Brazil. *Zootaxa*, *3793*, 147–164. http://dx.doi.org/10.11646/zootaxa.3793.1.7

Cipola, N. G., Morais, J. W., & Bellini, B. C. (2017). The discovery of *Lepidocyrtoides* Schött, 1917 (Collembola, Entomobryidae, Entomobryinae) from the New World, including three new species from Brazil and one from Australia. *Zootaxa*, *4324*(2), 201–248. https://doi.org/10.11646/zootaxa.4324.2.2

Cipola, N. G., Morais, J. W., Godeiro, N. N., & Bellini, B. C. (2019). Taxonomic revision of Brazilian genus *Tyrannoseira* Bellini & Zeppelini, 2011 (Collembola, Entomobryidae, Seirinae). *Zootaxa*, *4586*(2), 201–248.

Cipola, N. G., Oliveira, F. G. L., Morais, J. W., & Bellini, B. C. (2016). The *Heteromurini* Absolon & Ksenemann (Collembola, Entomobryidae): A review of the genera status and diagnoses, keys for species of *Alloscopus* Börner and *Heteromurtrella* Mari Mutt and description of a new species. *Zootaxa*, *4084*(2), 151–186. https://doi.org/10.11646/zootaxa.4084.2.1

Ding, Y. H., Yu, D. Y., Guo, W. B., Li, J. N., & Zhang, F. (2018). Molecular phylogeny of *Entomobrya* (Collembola: Entomobryidae) from China: Color pattern groups and multiple origins. *Insect Science*, *26*(3):587–597. https://doi.org/10.1111/1744-7917.12559

Faddeeva-vakhrusheva, A., Derks, M. F. L., Anvar, S. Y., Agamennone, V., Suring, W., Smit, S., … Roelofs, D. (2016). Gene Family Evolution Reflects Adaptation to Soil Environmental Stressors in the Genome of the Collembolan *Orchesella cincta*. *Genome Biology and Evolution*, *8*(7), 2106–2117. https://doi.org/10.1093/gbe/evw134

Faddeeva-Vakhrusheva, A., Kraaijeveld, K., Derks, M. F. L., Anvar, S. Y., Agamennone, V., Suring, W., … Roelofs, D. (2017). Coping with living in the soil: The genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics*, *18*(493), 1–14. https://doi.org/10.1186/s12864-017-3852-x

Gillett, C. P. D. T., Crampton-Platt, A., Timmermans, M. J. T. N., Jordal, B. H., Emerson, B. C., & Vogler, A. P. (2014). Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, *31*(8), 2223–2237. https://doi.org/10.1093/molbev/msu154

Godeiro, N. N., & Bellini, B. C. (2014). Three new species of *Seira* Lubbock (Collembola, Entomobryidae) from Caatinga Domain, northeastern Brazil. *Zootaxa*, *3764*(2), 131–151. https://doi.org/10.11646/zootaxa.3764.2.2

Godeiro, N. N., & Bellini, B. C. (2015). Two new species and two detailed chaetotaxy descriptions of *Seira* (Collembola: Entomobryidae) from Brazil. *Zootaxa*, *3972*(2), 208–230. https://doi.org/10.11646/zootaxa.3972.2.4

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Jordana, R. (2012). *Synopses on Palaeartic Collembola, Vol. 7/1: Capbryinae & Entomobryini*. Museum of Natural History Görlitz, Görlitz.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katz, A. D., Giordano, R., & Soto-adames, F. N. (2015). Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). *Zoological Journal of the Linnean Society*, *173*(4), 818–840. https://doi.org/10.1111/zoj.12220

Mari-Mutt, J. A. (1986). Puerto Rican species of Lepidocyrtus and *Pseudosinella* (Collembola: Entomobryidae). *Caribbean Journal of Science*, *22*(1–2), 1–48.

Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), e63. https://doi.org/10.1093/nar/gkz173

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448–pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Nguyen, L.-T., Schmidt, H. A., Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nunes, R. C., Godeiro, N. N., Pacheco, G., Liu, S., Gilbert, M. T. P., Alvarez-Valin, F., … Bellini, B. C. (2019). The discovery of Neotropical *Lepidosira* (Collembola, Entomobryidae) and its systematic position. *Zoologica Scripta*, *48*, 783–800. https://doi.org/10.1111/zsc.12377

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarżyński, D., Stevens, M. I., … Deharveng, L. (2012). Challenging species delimitation in Collembola: Cryptic diversity among common springtails unveiled by DNA barcoding. *Invertebrate Systematics*, *26*, 470–477. https://doi.org/10.1071/IS12026

Rambaut, A. (2016). *FigTree v. 1.4.3*. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*(1), 1–7. https://doi.org/10.1186/s13104-016-1900-2

Soto-Adames, F. N., Barra, J.-A., Christiansen, K., & Jordana, R. (2008). Suprageneric Classification of Collembola Entomobryomorpha. *Annals of the Entomological Society of America*, *101*(3), 501–513. https://doi.org/10.1603/0013-8746(2008)101[501:SCOCE]2.0.CO;2

Szeptycki, A. (1979). *Morpho-Systematic Studies on Collembola. IV. Chaetotaxy of the Entomobryidae and Its Phylogenetical Significance*.

Wu, C., Jordan, M. D., Newcomb, R. D., Gemmell, N. J., Bank, S., Meusemann, K., … Buckley, T. R. (2017). Analysis of the genome of the New Zealand giant collembolan (*Holacanthella duospinosa*) sheds light on hexapod evolution. *BMC Genomics*, *18*(1), 1–19. https://doi.org/10.1186/s12864-017-4197-1

Wray, D. L. (1953). Some new species of springtail insects (Collembola). *Nature Notes, Occasional Papers*, *1*, 1–7.

Yosii, R. (1959). Collembolan fauna of the Cape Province, with special reference to the genus *Seira* Lubbock. *Special publications from the Seto Marine Biological Laboratory*, *1*, 1–24. https://doi.org/10.5134/176436

Zeppelini, D., & Bellini, B. C. (2006). Two *Seira* Lubbock 1869 (Collembola, Arthropleona, Entomobryidae) new to science, with remarkable secondary sexual characters. *Zootaxa*, *1185*, 21–35.

Zhang, F., Chen, Z., Dong, R.-R., Deharveng, L., Stevens, M. I., Huang, Y.-H., & Zhu, C.-D. (2014). Molecular phylogeny reveals independent origins of body scales in Entomobryidae (Hexapoda: Collembola). *Molecular Phylogenetics and Evolution*, *70*, 231–239. https://doi.org/10.1016/j.ympev.2013.09.024

Zhang, F., & Deharveng, L. (2015a). First instar tibiotarsal chaetotaxy supports the Entomobryidae and Symphypleona (Collembola) forming a cluster in a phylogenetic tree. *Zootaxa*, *3955*, 487–504. https://doi.org/10.11646/zootaxa.3955.4.2

Zhang, F., & Deharveng, L. (2015b). Systematic revision of Entomobryidae (Collembola) by integrating molecular and new morphological evidence. *Zoologica Scripta*, *44*(3), 298–311. https://doi.org/10.1111/zsc.12100

Zhang, F., Ding, Y., Zhou, Q.-S., Wu, J., Luo, A., & Zhu, C.-D. (2019). A High-quality Draft Genome Assembly of *Sinella curviseta*: A soil model Organism (Collembola). *Genome Biology and Evolution*, *11*(2), 521–530. https://doi.org/10.1093/gbe/evz013

Zhang, F., Ma, Y., & Greenslade, P. (2017). New Australian Paronellidae (Collembola) reveal anomalies in existing tribal diagnoses. *Invertebrate Systematics*, *31*(4), 375–393. https://doi.org/10.1071/IS16073

Zhang, F., Sun, D.-D., Yu, D.-Y., & Wang, B.-X. (2015). Molecular phylogeny supports S-chaetae as a key character better than jumping organs and body scales in classification of Entomobryoidea (Collembola). *Nature Publishing Group*, *5*(1), 1–12. https://doi.org/10.1038/srep12471

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Godeiro NN, Pacheco G, Liu S, et al. Phylogeny of Neotropical Seirinae (Collembola, Entomobryidae) based on mitochondrial genomes. *Zool Scr*. 2020;49:329–339. https://doi.org/10.1111/zsc.12408