**Supplementary File**

**Detailed Materials and Methods**

*Samples*

Nine species selected from all four genera of Speleketorinae and six species selected from two of three genera of Prionoglaridinae were sampled from the family Prionoglarididae. Outgroups represented all three suborders of free-living Psocodea. The tree was rooted between Trogiomorpha and Troctomorpha+Psocomorpha according to the previous study (Yoshizawa et al., 2006). Selected taxa are listed in Table S1, together with GenBank accession numbers. All voucher specimens are deposited in the Hokkaido University Insect Collection, Sapporo, Japan.

*Amplification and sequencing*

Partial sequences of the nuclear 18S rRNA and Histone 3 and mitochondrial 16S rRNA, 12S rRNA, COI and CytB genes were used for the analyses. Methods for DNA extraction, polymerase chain reaction amplification, and sequencing followed Yoshizawa & Johnson (2008), except for CytB, which was amplified and sequenced using the newly designed primer pair COB.F10933trg (5'-TAY GTT YTA CCH TGA GGH CAA ATR TC-3’) and COB.R11367trg (5'-ATA ACH CCW CCT ART TTR TTA GGR AT-3').

Because the only available sample of *Afrotrogla* was old (stored under museum conditions for approximately 15 years) and the available tissue was also small (left over after taxonomic description), only a piece of the 18S rRNA sequence was available for this species.

*Alignment*

Alignment of protein coding genes was straightforward because no gap was identified in the present sequences. Ribosomal RNAs were aligned using MAFFT 6.5 (Katoh and Standley, 2013) with the Q-INS-i option, in which secondary structure information of RNA is considered. Apparent misalignments were corrected manually, and poorly aligned regions were excluded from the analyses.

*Model selection and phylogenetic analyses*

The best substitution models and partition schemes were estimated using PartitionFinder 2.3.3 (Lanfear et al., 2017), with the greedy algorithm. The codon positions for each PCG (3 genes x 3 codons = 9 partitions) and rRNA (3 partitions) were predefined for the PartitionFinder analyses. The best fit partition scheme and models were described in the nexus formatted data matrix available from FigShare at <https://doi.org/10.6084/m9.figshare.6452816>.

We estimated a maximum likelihood tree using IQ-Tree 1.6.3 (Nguyen et al., 2015), with 10,000 replicates of an ultrafast likelihood bootstrap (Hoang et al., 2018) to obtain bootstrap branch support values. A Bayesian analysis was performed using MrBayes (Ronquist and Huelsenbeck, 2003). We performed two runs each with four chains for 1,000,000 generations, and trees were sampled every 1000 generations. The first 10% of sampled trees was excluded as burn-in, and a 50% majority consensus tree was computed to estimate posterior probabilities.

*Morphological observations and coding*

A detached female abdomen was soaked with 10% KOH at room temperature for one night, then soaked with water and 80% ethanol. Observations were made in glycerol. An Olympus SZX16 binocular microscope (Tokyo, Japan) and a Zeiss Axiophot (Oberkochen, Germany) were used for observations. Photographs were taken with an Olympus E-M5 digital camera attached to an Olympus SZX16 binocular microscope (Tokyo, Japan). Partially focused pictures were combined using Helicon Focus (Helicon Soft Ltd., http://www.heliconsoft.com) to obtain images with a high depth of field.

In Figure 1, homologous structures were labelled according to the following colour scheme. Orange: Basal apodeme, where the muscles for moving/protruding the female penis are attached. Homologous throughout Sensitibillini. Light blue: basal ring supporting the apical membrane/sclerite of the female penis. Homologous throughout Sensitibillini. Green: lateral lobes and spines, homologous throughout *Neotrogla*, except for *N. truncata.* Red: dorsal lobes and spines, homologous throughout *Neotrogla*, except for *N. truncata.* Purple: ventral spines, unique to *N. curvata*.

Although the function of the female genitalia in *Afrotrogla* has not been confirmed to date, the extremely developed condition shown in Fig. 1f was coded as "presence of female penis", together with *Neotrogla*, for which the intromittent function of the structure was confirmed (Yoshizawa et al., 2014). Although the females of *Sensitibilla* also have a small penis-like organ, it apparently lacks the ability to protrude. Therefore, this condition was coded as "absence of female penis". The presence of lobes and spines labelled by green and red was coded as "presence of anchoring spines (figure 1b–d)", whereas their absence was coded as "absence of anchoring spines (figure 1a)".

*Ancestral state reconstruction*

(1) The presence or absence of the female penis, (2) the reduction or non-reduction of the male paramere, (3) the presence or absence of the female anchoring spines, (4) the presence or absence of the males genital pouch, (5) the presence or absence of the additional ventral spines, and (6) the presence or absence of ventral pouches were mapped on the maximum likelihood tree according to the parsimony and likelihood criteria as implemented in Mesquite 3.50 (Maddison & Maddison, 2018). Characters (1) and (2) (A and a in the figure), (3) and (4) (B and b), and (5) and (6) (C and c) are tightly associated characters during copulations.

As mentioned above, only a piece of the 18S rRNA sequence was available for *Afrotrogla*. This gene is the most slowly evolving one among the gene markers targeted here. Therefore, in the estimated trees, the basal branch leading to *Afrotrogla* was extremely short (Bayes) or almost zero (likelihood) (figure S1–S2). This branch length is likely an inappropriate estimation caused by the absence of rapidly evolving genes from our *Afrotrogla* sample. If we performed the likelihood-based ancestral state estimation using this tree, the terminal character of *Afrotrogla* will be directly assigned to the basal node of *Sensitibila+Afrotrogla*, because there is no chance to evolve new characters along the zero-length branch leading to *Afrotrogla*. To resolve this problem, branch lengths of the IQ-Tree estimated ML tree were re-estimated under the molecular clock model using PAUP\* (Swofford, 2002) (figure 2), and the character states were reconstructed using this tree with the likelihood model as implemented in Mesquite 3.50 (Maddison & Maddison, 2018). Because partitioned model was not allowed in PAUP\*, all data were combined into a single partition and the best fit model (GTR+I+R) was re-estimated using jModeltest 2 (Darriba et al., 2012). The best fit parameters were noted in the data matrix available at <https://doi.org/10.6084/m9.figshare.6452816>.

The ancestral state reconstruction was conducted under the parsimony and likelihood models as implemented in Mesquite 3.50. The Mesquite file for ancestral state estimation (which includes exact statistical values) is also available from <https://doi.org/10.6084/m9.figshare.6452816>. The estimated likelihood values are summarized in table S2.

*Environmental Data*

Most *Neotrogla* species live in extremely dry caves (referring to the absence of liquid water capable of transporting organic matter, both particulate and dissolved). Considering this, a *proxy* of transport of organic matter to caves and/or external productivity (such as precipitation and evapotranspiration) can be informative. Accordingly, datasets from the last 11 years of potential and real evapotranspiration and average precipitation of some municipalities where each *Neotrogla* species occur were obtained from the INMET (Instituto Nacional de Meteorologia from Brazil) (www.inmet.gov.br). Evapotranspiration can be defined as the sum of evaporation and plant transpiration from a land and ocean surface to the atmosphere. Potential evapotranspiration is defined as the ability of the atmosphere to remove water from the surface through evaporation and transpiration assuming no water supply constraints. Real (or actual) evapotranspiration is the quantity of water that is actually removed from a surface due to the processes of evaporation and transpiration. Generally, high gross primary production is strongly related to evapotranspiration, although the exact correspondence between those processes is not always observed since climate variables may drive them in different ways.

The lowest values for both evapotranspiration and average precipitation are observed for the Irecê region, the area where *N. truncata* lives. Therefore, the primary production in this area is the lowest (as indicated by the low evapotranspiration rates). This also shows that less organic matter is produced in the epigean environments and that importation of external fine particulate organic matter (FPOM) to caves by water is lower in that region than the others.

*Supplementary References*

Darriba D, Taboada GL, Doll R, Posada D (2012) jModeltest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:772.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34:772–773.

Madditon WP, Maddison DR (2018) Mesquite: a modular system for evolutionary analysis. Version 3.50 http://mesquiteproject.org

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35:518–522.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-Tree: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Swofford DL (2002) PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Sinauer Associates, Sunderland, MA.

Yoshizawa K, Lienhard C, Johnson KP (2006) Molecular systematics of the suborder Trogiomorpha (Insecta: Psocodea: "Psocoptera"). *Zoological Journal of the Linnean Society* 146:287–299.

Yoshizawa K, Ferreira RL, Kamimura Y, Lienhard C (2014) Female penis, male vagina, and their correlated evolution in a cave insect. *Current Biology* 24:1006–1010.

*Supplementary Figures and Tables*

**Figure S1**. The maximum likelihood tree estimated by IQ-Tree. Branch lengths are proportional to ML-estimated branch lengths. The numbers correspond the node ID in table S2.

(separate file)

**Figure S2**. The Bayesian tree estimated by MrBayes. Branch lengths are proportional to the branch length estimated by Bayesian MCMC.

(separate file)

**Table S1**. List of taxa examined in this study with GenBank accession numbers. “–“ indicates missing data.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Suborder** | **Family** | **Tribe** | **Species** | **Locality** | **ID** | **18S** | **H3** | **16S** | **12S** | **COI** | **CytB** |
| **Trogiomorpha** | **Prionoglarididae** | **Prionoglaridini** | ***Prionoglaris* sp*.*** | Greece | KY249 | AY630456 | DQ104773 | DQ104745 | LC425155 | – | LC425195 |
|  |  |  | ***Prionoglaris* sp*.* TK** | Turkey | KY468 | LC425184 | LC425231 | LC425176 | LC425156 | LC425217 | LC425196 |
|  |  |  | ***Prionoglaris stygia*** | Luxenbourg | KY478 | LC425185 | LC425232 | MG255141 | MG255141 | MG255141 | MG255141 |
|  |  | **Siamoglarinae** | ***Siamoglaris theresiae*** | Thailand | KY466 (paratype) | LC425186 | LC425233 | LC425177 | LC425157 | LC425218 | – |
|  |  |  | ***Siamoglaris zebrina*** | Thailand | KY255 (paratype) | DQ104798 | LC425234 | DQ104746 | LC425158 | LC425219 | LC425197 |
|  |  | **Speleketorini** | ***Speleketor irwini*** | California, USA | KY308 | DQ104799 | DQ104774 | DQ104747 | MG255142 | MG255142 | MG255142 |
|  |  | **Sensitibilini** | ***Afrotrogla oryx*** | South Africa | KY406 (paratype) | LC425187 | – | – | – | – | – |
|  |  |  | ***Sensitibilla strinatii*** | Central Namibia | KY251 | DQ104800 | DQ104775 | LC425178 | LC425159 | – | – |
|  |  |  | ***Sensitibilla etosha*** | Northern Namibia | KY410 (paratype) | LC425188 | LC425235 | LC425179 | LC425160 | LC425220 | LC425198 |
|  |  |  | ***Neotrogla truncata*** | Bahia, Brazil | KY507.508 | LC425189 | LC425236 | LC425180 | LC425161 | LC425221 | LC425199 |
|  |  |  | ***Neotrogla brasiliensis*** | Minas Gerais, Brazil | KY408 (paratype) | LC425190 | LC425237 | – | LC425162 | LC425222 | LC425200 |
|  |  |  | ***Neotrogla aurora*** | Tocantins, Brazil | KY467 (paratype) | LC425191 | – | LC425181 | LC425163 | LC425223 | LC425201 |
|  |  |  | ***Neotrogla aurora*** | Tocantins, Brazil | KY503 | LC425192 | – | LC425182 | LC425164 | LC425224 | LC425202 |
|  |  |  | ***Neotrogla* sp.5** | Minas Gerais, Brazil | KY501/ Genbank | LC425193 | – | MG255139 | MG255139 | MG255139 | MG255139 |
|  |  |  | ***Neotrogla curvata*** | Bahia, Brazil | KY476 (paratype) | LC425194 | – | LC425183 | LC425165 | LC425225 | LC425203 |
| **Outgroups** |  |  |  |  |  |  |  |  |  |  |  |
| **Trogiomorpha** | **Psyllipsocidae** |  | ***Dorypteryx domestica*** | Switzerland | GenBank | AY630454 | DQ104777 | DQ104749 | MG255136 | MG255136 | MG255136 |
|  | **Psoquillidae** |  | ***Rhyopsocus* sp*.*** | USA, California | KY297 | DQ104801 | DQ104778 | DQ104750 | LC425167 | LC425227 | LC425205 |
|  | **Trogiidae** |  | ***Cerobasis alpha*** | USA, Arizona | KY321 | DQ104803,4 | DQ104787 | DQ104760 | LC425166 | LC425226 | LC425204 |
|  | **Lepidopsocidae** |  | ***Soa* sp*.*** | PNG | KY323 | DQ104802 | DQ104780 | DQ104752 | LC425168 | – | – |
|  |  |  | ***Echmepteryx madagascariensis*** | Japan | KY61 | AY630447 | DQ104781 | DQ104753 | AY139915 | LC425228 | LC425206 |
| **Troctomorpha** | **Troctopsocidae** |  | ***Selenopsocus* sp*.*** | Malaysia | KY198 | AY630457 | AB919025 | AB918939 | LC425169 | LC425229 | LC425207 |
|  |  |  | ***Thaipsocus* sp*.*** | Malaysia | KY258 | AB919005 | AB919026 | AB918940 | LC425170 | AB918977 | LC425208 |
|  | **Amphientomidae** |  | ***Stimulopalpus japonicus*** | Japan/USA | GenBank | AY900127 | GU569345 | MG255143 | MG255143 | MG255143 | MG255143 |
|  |  |  | ***Cymatopsocus* sp*.*** | Malaysia | KY220 | AY630460 | AB919021 | AB918935 | LC425171 | AB918976 | LC425209 |
|  |  |  | **Genus sp*.*** | Malaysia | KY197,256 | AY630458 | AB919022 | AB918936 | – | LC425230 | LC425210 |
| **Psocomorpha** | **Archipsocidae** |  | ***Archipsocus nomas*** | Florida, USA | GenBank | AY900133 | AB919034 | MG255135 | MG255135 | MG255135 | MG255135 |
|  | **Dasydemellidae** |  | ***Matsumuraiella radiopicta*** | Japan | KY236 | AY630493 | DQ104797 | DQ104770 | AB856912 | GU569275 | LC425211 |
|  | **Mesopsocidae** |  | ***Mesopsocus hongkongensis*** | Japan | KY224 | AY630516 | DQ104794 | DQ104767 | LC425172 | GU569268 | LC425212 |
|  | **Philotarsidae** |  | ***Aaroniella* sp*.*** | Japan | KY216 | AY630533 | AB919037 | AB918951 | LC425173 | AB918985 | LC425213 |
|  | **Epipsocidae** |  | ***Epipsocus* sp*.*** | Malaysia | KY205 | AY630539 | GU569314 | GU569189 | LC425174 | GU569249 | LC425214 |
|  | **Hemipsocidae** |  | ***Hemipsocus* sp*.*2** | Malaysia | KY228 | AY630544 | DQ104792 | DQ104765 | LC425175 | EF662063 | LC425215 |
|  | **Psocidae** |  | ***Ptycta johnsoni*** | Japan | KY235 | AY630553 | EF662175 | AY139954 | AY139907 | EF662093 | LC425216 |

**Table S2**. The likelihood of the ancestral character states estimated by the likelihood criterion using Mesquite. The numbers correspond the node IDs of figure S1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species and node ID | presence of female penis | absence of male paramere | presence of female spines (dorsal+lateral) | presence of male pouches (dorsal+lateral) | presence of female spines (ventral) | presence of male pouches (ventral) |
| Neot. brasiliensis | 1 | 1 | 1 | 1 | 0 | 0 |
| Neot. curvata | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 0.999 | 0.999 | 0.994 | 0.994 | 0.004 | 0.004 |
| Neot. truncata | 1 | 1 | 0 | 0 | 0 | 0 |
| 2 | 0.999 | 0.999 | 0.925 | 0.925 | 0 | 0 |
| Neot. aurora | 1 | 1 | 1 | 1 | 0 | 0 |
| Neot. aurora | 1 | 1 | 1 | 1 | 0 | 0 |
| 3 | 0.999 | 0.999 | 0.999 | 0.999 | 0 | 0 |
| Neot. sp. | 1 | 1 | 1 | 1 | 0 | 0 |
| 4 | 0.999 | 0.999 | 0.997 | 0.997 | 0 | 0 |
| 5 | 0.999 | 0.998 | 0.925 | 0.925 | 0 | 0 |
| Sen. strinatii | 0 | 0 | 0 | 0 | 0 | 0 |
| Sen. etosha | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0.005 | 0.014 | 0 | 0 | 0 | 0 |
| Afrotrogla | 1 | 1 | 0 | 0 | 0 | 0 |
| 7 | 0.112 | 0.147 | 0 | 0 | 0 | 0 |
| 8 | 0.113 | 0.154 | 0.012 | 0.012 | 0 | 0 |
| Speleketor | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0.023 | 0.054 | 0.002 | 0.002 | 0 | 0 |
| Soa | 0 | 0 | 0 | 0 | 0 | 0 |
| Echmepteryx | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cerobasis | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rhyopsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 |
| Dorypteryx | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| Priono. sp.488 | 0 | 0 | 0 | 0 | 0 | 0 |
| Priono. stygia | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| Priono. sp.249 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 |
| Siamo. theresiae | 0 | 0 | 0 | 0 | 0 | 0 |
| Siamo. zebrina | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| Aaroniella | 0 | 0 | 0 | 0 | 0 | 0 |
| Epipsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0.001 | 0 | 0 | 0 | 0 |
| Hemipsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| Ptycta | 0 | 1 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0.039 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0.001 | 0 | 0 | 0 | 0 |
| Matsumuraiella | 0 | 0 | 0 | 0 | 0 | 0 |
| Mesopsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0.001 | 0 | 0 | 0 | 0 |
| 24 | 0 | 0.001 | 0 | 0 | 0 | 0 |
| Archipsocus | 0 | 1 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0.001 | 0 | 0 | 0 | 0 |
| Stimulopalpus | 0 | 0 | 0 | 0 | 0 | 0 |
| Genus | 0 | 0 | 0 | 0 | 0 | 0 |
| 26 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cymatopsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| 27 | 0 | 0 | 0 | 0 | 0 | 0 |
| Selenopsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| Thaipsocus | 0 | 0 | 0 | 0 | 0 | 0 |
| 28 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | 0 | 0.001 | 0 | 0 | 0 | 0 |
| 30 | 0 | 0.002 | 0 | 0 | 0 | 0 |

**Table S3**. Environmental data from the municipalities where *Neotrogla* species occur. The data represent an average (per month) of each parameter from the last 11 years (2006-2017) and are all presented in mm/month.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Species | Locality | Biome | Potential Evapotranspiration | Real Evapotranspiration | Average precipitation |
| *N. truncata* | Irecê (BA) | Semi-arid | 108.48 | 42.01 | 16 |
| *N. aurora* | Peixe (TO) | Savannah | 123.49 | 75.48 | 25 |
| *N. curvata* | Correntina (BA) | Semi-arid | 115.03 | 69.51 | 25.16 |
| *N. brasiliensis*  *Neotrogla* sp. | Januária (MG) | Savannah | 123.54 | 78.29 | 34.35 |