

# Evolution of Strigamia centipedes (Chilopoda): a first molecular assessment of phylogeny and divergence times

Zoologica Scripta

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

# Evolution of *Strigamia* centipedes (Chilopoda): a first molecular assessment of phylogeny and divergence times

Lucio Bonato, Francesca Bortolin, Leandro Drago, Marco Orlando & László Dányi

Submitted: 4 October 2016
Accepted: 28 December 2016
doi:10.1111/zsc.12234

Bonato, L., Bortolin, F., Drago, L., Orlando, M. and Dányi, L. (2017). Evolution of *Strigamia* centipedes (Chilopoda): a first molecular assessment of phylogeny and divergence times. —*Zoologica Scripta*, *00*, 000–000.

We present a first phylogenetic and temporal framework, with biogeographical insights, for the centipedes of the genus *Strigamia*, which are widespread predators in the forest soils of the Northern Hemisphere and comprise the evo-devo model species *Strigamia maritima*. The phylogeny was estimated by different methods of maximum likelihood and Bayesian inference from sequences of two mitochondrial (*16S*, *COI*) and two nuclear (*18S*, *28S*) genes, obtained from 16 species from all major areas of the global range of the genus and encompassing most of the overall morphological and ecological diversity. Divergence times were estimated after calibration upon the fossil record of centipedes. We found that major lineages of extant species of *Strigamia* separated most probably around 60 million years (Ma) ago. The two most diverse lineages diversified during the last 30 Ma and are today segregated geographically, one in Europe and another in Eastern Asia. This latter region hosts a hitherto underestimated richness and anatomical diversity of species, including three still unknown, yet morphologically well distinct species, which are here described as new: *Strigamia inthanoni* sp. n. from Thailand, *Strigamia korsosi* sp. n. from the Ryukyu Islands and *Strigamia nana* sp. n. from Taiwan. The northern European model species *S. maritima* is more strictly related to the Eastern Asian lineage, from which it most probably separated around 35 Ma ago before the major diversification of the latter.

Corresponding author: *Lucio Bonato*, *Università di Padova*, *Dipartimento di Biologia*, *Via Bassi 58B*, *I-35131 Padova*, *Italy*. *E-mail: lucio.bonato@unipd.it*

*Lucio Bonato*, *Francesca Bortolin*, *Leandro Drago and Marco Orlando*, *Department of Biology*, *University of Padova*, *via Bassi 58B*, *Padova*, *I-35131 Italy*. *E-mails: lucio.bonato@unipd.it*, *francesca.bortolin@unipd.it*, *leandro.drago@unipd.it*, *marco.orlando1991@live.it*

*László Dányi*, *Department of Zoology*, *Hungarian Natural History Museum*, *Baross u. 13*, *Budapest*, *H-1088 Hungary*. *E-mail: laszlodanyi@gmail.com*

## Introduction

Centipedes (Chilopoda) are among the most widespread predators in the soil communities throughout the world, together with some major groups of beetles (especially Carabidae), spiders (Araneae) and a few other taxa. In the forest soils of the entire Northern Hemisphere, the genus *Strigamia* (Chilopoda: Geophilomorpha) is one of the most frequent and diverse of the centipede genera, together with the better known *Cryptops* (Scolopendromorpha) and *Lithobius* (Lithobiomorpha) (Edgecombe & Giribet 2007; Minelli 2011). At least 30 morphologically distinct species are known within *Strigamia* (Bonato *et al.* 2012). Most of these species inhabit North America, Europe and nearby Western Asia, whereas only seven species have been described from Eastern Asia, besides another four nominal taxa of uncertain validity (Murakami 1993; Shinohara 1999). However, other species are expected to exist but still await detection and description, a common situation for the entire centipede biodiversity (Edgecombe 2007).

Most notable among all known species is *Strigamia maritima*. While all other congeners colonize mainly forest soils in inland sites, up to high mountains, *S. maritima* is instead strictly littoral and lives only on seashores of northern Europe (Barber 2009). In relation to its unique habitat, *S. maritima* greatly differs from other *Strigamia* species also in feeding habits and reproductive biology (Lewis 1961). Additionally, because of the unusually high population densities in suitable breeding sites, *S. maritima* has become a model organism to study the developmental processes and the genetic control underlying the segmental patterning of the arthropod body (e.g. Arthur & Chipman 2005; Chipman & Akam 2008; Brena 2014). Consistently, *S. maritima*

has been the first myriapod species for which the genome has been entirely sequenced (Chipman *et al.* 2014; Robertson *et al.* 2015) and the single geophilomorph species for which sexual chromosomes have been demonstrated (Green *et al.* 2016).

Despite the overarching ecological role of *Strigamia* species in most boreal forest ecosystems, and the great interest of the model species *S. maritima* within evolutionary developmental biology, the evolutionary history of the genus *Strigamia* and the specific phylogenetic position of *S. maritima* have never been explored so far. Indeed, phylogenetic inference from morphology is discouraged by the fact that all species of *Strigamia* are very similar and they are usually distinguished only for very few anatomical characters: above all, the number of body segments (Bonato *et al.* 2012), the shape of the so-called forcipules (i.e. the poison fangs; Maruzzo & Bonato 2014) and the arrangement of the so-called coxal organs (i.e. putative osmoregulatory vesicles at the basis of the ultimate legs; Rosenberg 1989). On the other hand, molecular phylogenetic analyses have been hindered so far by the fact that non-European species of *Strigamia* have been rarely collected, especially in the last decades, and suitable samples for DNA extraction are very rare in zoological collections.

We present here a first molecular analysis of the phylogenetic relationships within *Strigamia* based on freshly collected specimens from all major areas of the global range of the genus, and encompassing most of its overall morphological and ecological variation. Our analysis is based on the sequences of four genes, including mitochondrial and nuclear loci, for a total of >4000 base pairs. As a result, we present a first phylogenetic and temporal framework to understand the evolutionary and biogeographical history of these typical predators of the boreal forests, and the differentiation of the ecologically unique model species *S. maritima*.

## Material and methods

### *Specimen sampling and identification*

Specimens of *Strigamia* were collected in the field, either by the authors or collaborators, in all major subcontinental areas comprising the global range of the genus to obtain suitable samples for DNA extraction from as many species as possible. As the current taxonomic knowledge of *Strigamia* is strongly biased towards the European species, we devoted special efforts to sample the rarely collected and poorly known species inhabiting North America and Eastern Asia, from subarctic to subtropical areas.

Specimens were originally fixed in ethanol 60–96% and later examined independently by two authors (L. Bonato, L. Dányi) in order to identify them at the species level, referring to the most recent taxonomic revision of the genus (Bonato *et al.* 2012). Selected specimens were then transferred to absolute ethanol.

### *DNA extraction and sequencing*

Total DNA was extracted from a few trunk segments of each specimen using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

We amplified and sequenced portions of the mitochondrial genes for cytochrome *c* oxidase subunit I (*COI*) and 16S rRNA (*16S*), and the nuclear genes for 18S and 28S rRNA (*18S* and *28S*). These four genes were chosen because they are among the few genes that have been so far amplified and sequenced successfully in a large sample of Chilopoda, including *Strigamia* species, and have been used successfully in previous phylogenetic analyses (Edgecombe & Giribet 2004; Murienne *et al.* 2010; Bonato *et al.* 2014).

The primers used for amplification are listed in Table S1. For *18S*, three overlapping fragments (I, II, III) were amplified. For *28S*, two overlapping fragments (II-III) and another non-overlapping fragment (I) were amplified. The primers COI StrigFor (ACCTGAGCCGCAA TAGCA), COI StrigRev (TGTTGGTAAAGAATGGG GTC), 28SI StrigFor (CGAAGCCGCCGTCCATTA) and 28SI StrigRev (TACGCCAGATTCCAACTCCA) were designed specifically for *Strigamia* on the fragments amplified with the other primers for *COI* and the fragment I of *28S* (Table S1).

PCRs were performed in 20 $\mu$L reactions containing 4.0 $\mu$L of 5X Flexi Buffer, 0.4 $\mu$L of 10 mM dNTPs, 1.2 $\mu$L of 25 mM $MgCl_2$, 0.5 $\mu$L of 100% DMSO, 0.5–1.0 $\mu$L of each 10 $\mu$M primer, 0.1 $\mu$L of 5U/$\mu$L GoTaq Flexi DNA Polymerase (Promega, Madison, USA), 1 $\mu$L of template DNA and purified water. The reaction was carried out as follows: 1 step at 95 °C for 5 min; 27–35 cycles consisting of 1 min at 94 °C, 1 min at 40–58 °C and 1 min 30 s at 72 °C; a final step at 72 °C for 7 min. Specific annealing conditions are reported in Table S1.

The double-stranded PCR products were verified by 1% agarose gel electrophoresis and purified using the MinElute PCR Purification Kit (Qiagen). The purified PCR products were sequenced directly on both strands with the same primer pairs as used for amplification, by means of an ABI 3730 XL automatic capillary sequencer (Applied Biosystems, Branchburg, USA; service provided by BMR Genomics, Padova, Italy).

The sequences obtained were checked for similarity to other related taxa through BLAST searches. Additionally, to check for the presence of nuclear pseudogenes or misalignments, *COI* sequences were also translated into amino acid

sequences, after choosing the correct open reading frame, using the 'invertebrate mitochondrial' code, with MEGA 6 (Tamura *et al.* 2013).

### *Species data set*

We obtained good-quality DNA sequences for 14 morphologically different species of *Strigamia*, including three species known from North America, five species distributed in Europe and nearby Western Asia, and six species found only in Eastern Asia, from Japan to Indochina (Fig. 1; Table 1). Another two species for which molecular data were already available were included in the analysis, namely *S. maritima* from Europe (sequences from Chipman *et al.* 2014) and *Strigamia inthanoni* sp. n. from Eastern Asia (sequences from Bonato *et al.* 2014, originally referred tentatively to *Strigamia svenhedini*). As a result, the molecular data set included 16 species of *Strigamia*, each represented by a single specimen except for *Strigamia acuminata* (two specimens from different sites) and for *S. maritima* (sequences available from an unknown number of specimens).

### *Sequence alignment and trimming*

Cytochrome *c* oxidase subunit I sequences were aligned with CLUSTALW at default parameters with MEGA 6, whereas the ribosomal gene sequences were aligned with the E-INS-i algorithm of Mafft (Katoh *et al.* 2005) within the web server version of GUIDANCE 2 (Sela *et al.* 2015). Aligned positions with a confidence score <0.5 were trimmed, and random similar sections were excluded with ALISCORE v.2.0 (Misof & Misof 2009) and ALICUT v.2.3 (Kück 2009) at default parameters (window size set to six and gaps/ambiguities treated as fifth character state). For the *28S* gene, the non-overlapping fragment I was aligned and trimmed independently from the overlapping segments II and III, and only later, they were all concatenated with SEQUENCEMATRIX v.1.8 (Vaidya *et al.* 2011). Terminal gaps were eliminated when present in the majority of the sequences. The definitive lengths of the multiple sequence alignments were 578 bp for *COI*, 415 bp for *16S*, 1555 bp for *18S* and 1470 bp for *28S* (with the exceptions of two species, *Strigamia korsosi* sp. n. and *Strigamia platydentata*, for which only a 587-bp portion was sequenced successfully).

### *Phylogenetic analyses*

Candidate out-group species were selected among the closest relatives of *Strigamia* (according to previously published phylogenetic analyses; Murienne *et al.* 2010; Bonato *et al.* 2014), for which sequences were available for all the four genes. They were compared for the instability of their position in a reference maximum likelihood (ML) tree including only *Strigamia* species, using the evolutionary placement algorithm (Berger *et al.* 2011) in RAXML v.8.2.1 (Stamatakis 2014). The two species with the lowest average entropy were selected as out-groups (*Geophilus flavus* and *Pleurogeophilus mediterraneus*; Table 1).

Maximum likelihood and Bayesian inference (BI) analyses were carried out for the single genes.

Maximum likelihood analyses were performed with PHYML v.3.1 (Criscuolo 2011). For each gene, the best fitting substitution model was selected according to AICc, among 40 general time reversible (GTR) possible schemes, using JMODELTEST v.2.1.7 (Darriba *et al.* 2012). The

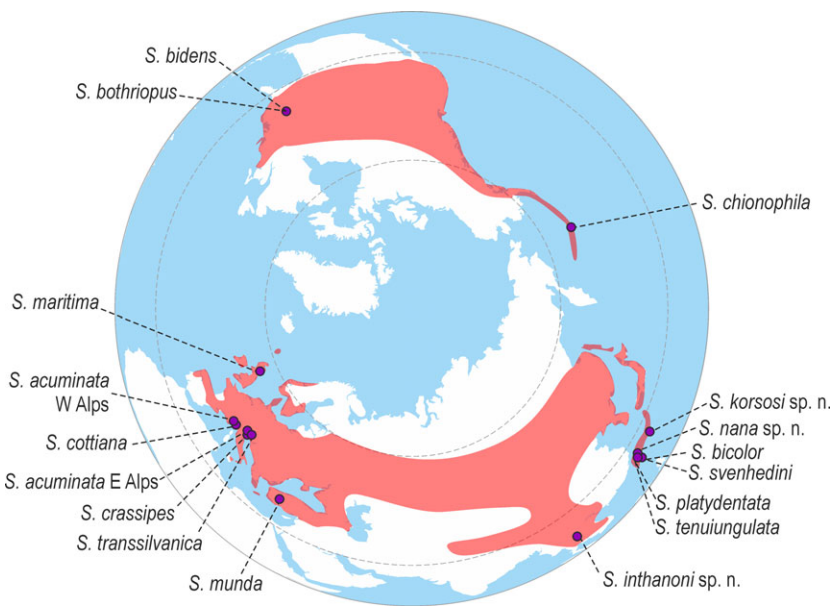


**Fig. 1** Geographical distribution of the genus *Strigamia* (shaded area) and provenance of the specimens of the species included in the analysis (circles). Map: polar projection of the Northern Hemisphere, with the 30°N and 60°N parallels.

**Table 1** Species, specimens and GenBank accession numbers of the gene sequences included in the phylogenetic analyses

| Species | Area | Population | Specimen code | *COI* | *16S* | *18S* | *28S* |
|---|---|---|---|---|---|---|---|
| In-group | | | | | | | |
| *Strigamia acuminata* (Leach, 1815) | W Palearctic | Italy: Western Alps: Valle Varaita | PD1311 | LT630324 | LT702887 | LT702899 | LT702912 |
| | | Italy: Eastern Alps: Val Popena | PD1348 | LT630323 | LT702886 | LT702898 | LT702911 |
| *Strigamia bicolor* Shinohara, 1981 | E Asia | Taiwan: Hualien | HNHM chilo-5252 | LT630325 | LT702888 | LT702900 | LT702913 |
| *Strigamia bidens* Wood, 1862 | N America | USA: Appalachians: Great Smoky Mt | PD1823 | LT630326 | LT702889 | LT702901 | LT702914 |
| *Strigamia bothriopus* Wood, 1862 | N America | USA: Appalachians: Great Smoky Mt | PD1975 | – | – | LT702902 | LT702915 |
| *Strigamia chionophila* Wood, 1862 | N America | USA: Alaska: Kasatochi | UAM 100032858a | LT630327 | LT702890 | LT702903 | LT702916 |
| *Strigamia cottiana* (Verhoeff 1935) | W Palearctic | Italy: Western Alps: Colle del Termine | PD2122 | LT630328 | LT702891 | LT702904 | LT702917 |
| *Strigamia crassipes* (Koch, 1835) | W Palearctic | Italy: Eastern Alps: Susegana | PD692 | KF569303 | LT702892 | LT702905 | KF569282 |
| *Strigamia inthanoni* sp. n. | E Asia | Thailand: Thai highlands: Doi Inthanon | ZMUC1006* | KF569304 | KF569239 | KF569262 | KF569283 |
| *Strigamia korsosi* sp. n. | E Asia | Japan: Okinawa | PD6126 | LT630329 | LT702893 | LT702906 | LT702918 |
| *Strigamia maritima* (Leach, 1817) | W Palearctic | UK: Scotland | Unknown | KP173664 | AY288733 | AF173265 | HM453303 |
| *Strigamia munda* (Chamberlin, 1952) | W Palearctic | Turkey: Anatolia | PD1286 | LT630330 | – | – | – |
| *Strigamia nana* sp. n. | E Asia | Taiwan: Taoyuan: Fuxing | PD6124 | LT630331 | – | – | LT702919 |
| *Strigamia platydentata* Shinohara, 1981 | E Asia | Taiwan: Taichung | HNHM chilo-5253 | LT630332 | LT702894 | LT702907 | LT702920 |
| *Strigamia svenhedini* (Verhoeff, 1933) | E Asia | Taiwan: Nantou | HNHM chilo-5254 | LT630333 | LT702895 | LT702908 | LT702921 |
| *Strigamia tenuiungulata* (Takakuwa, 1938) | E Asia | Taiwan: Taichung | HNHM chilo-5255 | – | LT702896 | LT702909 | LT702922 |
| *Strigamia transsilvanica* (Verhoeff, 1928) | W Palearctic | Italy: Eastern Alps: Sappada | PD908 | LT630334 | LT702897 | LT702910 | LT702923 |
| Out-groups for phylogenetic analyses | | | | | | | |
| *Geophilus flavus* (De Geer, 1778) | – | – | – | JN306685 | KF569229 | KF569252 | KF569273 |
| *Pleurogeophilus mediterraneus* (Meinert, 1870) | – | – | – | KF569299 | KF569232 | KF569255 | KF569277 |
| Out-groups for time estimations | | | | | | | |
| *Cormocephalus hartmeyeri* (Kraepelin, 1908) | – | – | – | KF676531 | KF676491 | KF676445 | AY210812 |
| *Haplophilus subterraneus* (Shaw, 1789) | – | – | – | JN306673 | KF569237 | KF569260 | KF56928 |

New species are described in Appendix S1. Repositories of specimens: PD, Department of Biology, University of Padova, Italy; HNHM, Hungarian Natural History Museum, Budapest, Hungary; NMNS, National Museum of Natural Science, Taichung, Taiwan; UAM, Museum of the North, University of Alaska, Fairbanks, USA; ZMUC, Zoological Museum, University of Copenhagen, Denmark.

*Sequences obtained by this specimen have been originally referred to *S. svenhedini* (Bonato *et al.* 2014) because of misidentification.

selected models were TrN+G+I for *COI* and *18S*, GTR+G+I for *16S* and *28S*. All parameters, including base frequencies, were estimated by maximizing a likelihood function. Node support was assessed with 1500 bootstrap replicates, and the stabilization of bootstrap frequency values was checked in RAXML v.8.2.1 with the – I option.

Bayesian inference analyses were performed with PHYLOBAYES 4.1b (Lartillot *et al.* 2009). Two identical but independent runs with the same settings were conducted, and the sampled trees were merged subsequently to produce a majority rule extended tree. A model with GTR relative substitution rates and five categories of rate variation among sites, together with a 'cat' base frequencies mixture, was employed. Priors were set at default values. For each run, a Markov chain of 20 000 cycles was performed, one cycle for every 10 was sampled and the first 10% of the retained 2000 cycles was discarded as 'burn-in'. The quality of parameter estimation and the convergence and similarity of both chains were checked following the guidelines provided by Lartillot *et al.* (2009).

For each gene, a tree was generated with MESQUITE 3.0.4 (Maddison & Maddison 2001) from the ML and the BI majority rule consensus trees: nodes were included only if supported by either >50% bootstrap or >0.5 posterior probability, whereas conflicting clades were collapsed. Discordant topologies between the consensus trees of the four genes were tested with the Shimodaira–Hasegawa test implemented in the software CONSEL (Shimodaira & Hasegawa 2002).

A comprehensive analysis was also performed on all the four genes, after concatenating the sequences with SEQUENCEMATRIX v.1.8.

Maximum likelihood analysis of the concatenated sequences was performed using RAXML v.8.2.1, with 200

completely independent inferences, using the GTRGAMMAI model and optimizing base frequencies. One partition per gene resulted in the best fitting partitioning scheme in PARTITIONFINDER v1.1.1 (Lanfear *et al.* 2012), under the AICc.

Bayesian inference analysis of the concatenated sequences was performed in BEAST v.1.8.4 (Drummond *et al.* 2012). A log-normal distributed prior was set for the relative substitution rate, a Yule tree prior was used to model branching times, and a strict clock was assumed with a uniform prior mean rate between 0 and 10 000. Relative substitution rate priors were set to a log-normal distribution. Three identical but independent runs were conducted with the same settings: a Monte Carlo Markov chain of 20 000 000 generations, sampling one tree for every 4000 and discarding the first 10% of the trees as 'burn-in'. The resulting chains were checked for quality and convergence with TRACER 1.6 (Rambaut & Drummond 2013) and merged with LOGCOMBINER v.1.8.2 (part of BEAST package). The same settings were used for a BI using the 'multi-species coalescent' model, which accounts for gene-specific coalescence process within a species, while treating each species lineage as a Wright–Fisher population (Heled & Drummond 2010). The 'maximum clade credibility' species and gene trees were estimated simultaneously (Rannala & Yang 2003; Degnan & Rosenberg 2009). The best model was selected from estimated marginal posterior probability of likelihood values, using AICm (a posterior simulation-based analogue of AIC through Monte Carlo Markov chain, implemented in TRACER 1.6; Baele *et al.* 2012) with 100 bootstrap replicates.

All ML and BI analyses of the concatenated sequences were performed both including all species and excluding three species represented by only one or two of the four genes, namely *Strigamia bothriopus*, *Strigamia munda* and *Strigamia nana* n. sp. 'Rogue' species were searched for with RogueNaRok (Aberer *et al.* 2013) and excluded when detected (only the out-group *G. flavus*).

### *Divergence time estimation*

Divergence times were estimated with a BI analysis in BEAST v.1.8.4 on the concatenated sequences of the four genes, considering the maximum clade credibility tree obtained from precedent analyses and setting the substitution rates of the partitions to continuous time Markov chain reference prior (Ferreira & Suchard 2008). Other priors were set as in the phylogenetic analyses.

Two calibration points were used: 306 million years (Ma) was set as the minimum age of the split between Scolopendromorpha and Geophilomorpha (modelled with a log-normal distribution with average 30 Ma and log standard deviation 0.75), according to the reported age of the fossil *Mazoscolopendra* (Epimorpha crown-group; Mundel 1979; Fernández *et al.* 2014); 170 Ma was set as the average age of the split between Himantarioidea and Geophiloidea (modelled with a normal distribution with standard deviation 25 Ma), according to the average estimate of a previous analysis (Fernández *et al.* 2014) and referring to the most recent phylogenetic hypothesis for Geophilomorpha (Bonato *et al.* 2014). To set these constraints, the out-groups used in the previous analyses were replaced by a species of Scolopendromorpha and a species of Himantarioidea for which molecular sequences were already available (Table 1).

The resulting chains were checked for quality and convergence with TRACER 1.6. Another analysis with the same settings, but sampling from prior only, was performed as suggested by Heled & Drummond (2011), to check for consistency between the prior distributions and the analysis-driven marginal prior densities for each calibrated node and to control if posterior time estimates are conditioned by prior densities.

### *Description of new species*

Specimens that were morphologically different from all known species of *Strigamia* in major recognized species-diagnostic characters (see Bonato *et al.* 2012), and that resulted different from all other sampled species also in the molecular sequences, were described as new species (Appendix S1). Morphological descriptions were based on microscopic examination after specimen preparation and dissection (Pereira 2000) and following the standard terminology for Chilopoda (Bonato *et al.* 2010). Measures were taken by means of a micrometre applied to the ocular lens. Line-drawings were produced by a single author (L. Dányi) by means of a camera lucida applied to a Leica DM 1000 microscope.

## Results

### *Phylogeny*

The gene sequences obtained from the species of *Strigamia* (13–15 species, depending on the gene) and from the two most stable out-group species (Table 1) were all homogeneous in base frequencies for all four genes ($\chi^2$ test and parametric bootstrapping: $P > 0.05$). Rogue species were detected only when analysing the *16S* (*S. inthanoni* sp. n. and the out-groups).

For each gene, ML and BI analyses produced largely consistent topologies, but the consensus tree between the two methods showed generally low statistical support (Fig. S1). It was especially low for the two shorter mitochondrial genes (*COI* and *16S*), whereas the most resolved consensus tree was obtained from *28S*. The trees obtained from different genes were largely compatible with each other, and the few topological incongruences (especially in

the position of the species pair *Strigamia cottiana* + *Strigamia transsilvanica*) were not statistically significant (Shimodaira–Hasegawa tests: $P > 0.05$).

Maximum likelihood and BI analyses of the concatenated genes produced better supported and largely consistent trees when including only species represented by at least three of the four genes (Fig. 2A vs. Fig. S2). However, while the ML analysis suggests that the species pair *S. cottiana* + *S. transsilvanica* is closer to *Strigamia crassipes* than to *S. acuminata*, the BI analysis favours the alternative hypothesis, with a posterior probability 0.62 for the clade [(*S. cottiana*, *S. transsilvanica*) *S. acuminata*]. For most other species, the favoured position is the same between ML and BI analyses (ML bootstrap values >95% and BI posterior probabilities ~1), but the positions of *Strigamia bidens* and *S. maritima* are strongly supported only by BI, whereas that of *S. korsosi* sp. n. is supported only moderately by the two analyses. An alternative BI analysis performed on the four genes under the 'multispecies coalescent' model favoured a similar topology (Fig. 2B), however with generally lower posterior probabilities and differing in the precise position of *S. bidens* and *Strigamia chionophila*. When the three species represented by only one or two genes were included in the analysis (Fig. S3), their position was inconsistent between ML and BI, and the statistical support of most clades was lowered.

Taking into account all analyses, most of the species of *Strigamia* can confidently be assigned to one or the other of two major clades. A first clade comprises *S. acuminata*, *S. cottiana*, *S. crassipes* and *S. transsilvanica*, that is all species sampled from those inhabiting Europe to the exclusion of the strictly littoral, northern European *S. maritima* (Fig. 1). Within this clade, *S. cottiana* and *S. transsilvanica* are resolved as sister species, whereas it is uncertain whether they are closer to *S. acuminata* or to *S. crassipes*. The other major clade comprises *Strigamia bicolor*, *S. platydentata*, *S. svenhedini*, *Strigamia tenuiungulata*, three new species (here described as *S. inthanoni* sp. n., *S. korsosi* sp. n., *S. nana* sp. n.; Appendix S1) and *S. maritima*, that is all species sampled from those inhabiting Eastern Asia, together with the strictly littoral, northern European *S. maritima* (Fig. 1). Within this clade, the first four species are more strictly related, with a well-resolved topology, whereas *S. maritima* most probably emerged basal to the other species. Outside the two well-supported clades, *S. munda* remains unresolved, whereas *S. bothriopus*, *S. chionophila* and *S. bidens*, which are all from North America (Fig. 1), are variously resolved at the basis of one or the other of the two major clades, depending on genes and analyses, but they do not group together.

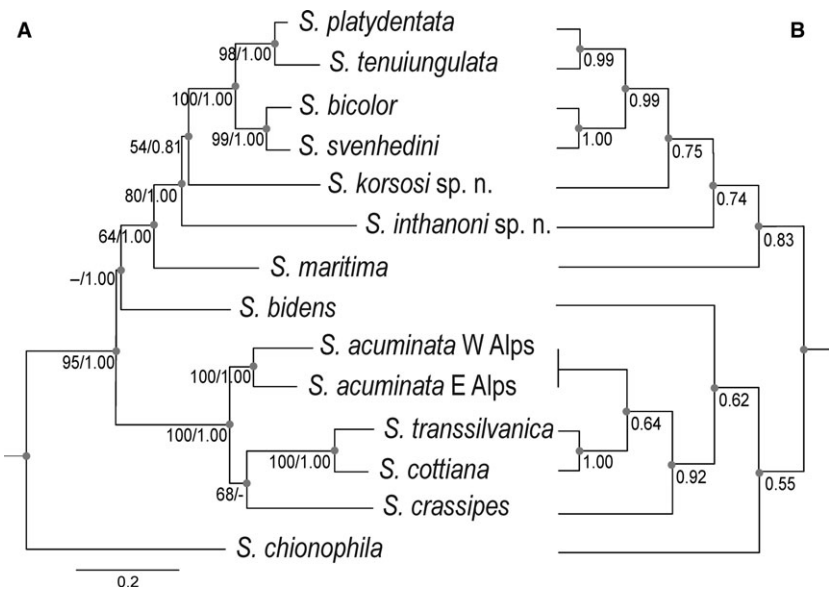


**Fig. 2** Phylogeny of *Strigamia* species.—A. Maximum likelihood (ML) tree of the concatenated sequences of the four genes. Only species represented by at least three of the four genes have been included (see Fig. S2 for the analysis including all species). Branches are proportional to the expected probability of substitutions per site. Bootstrap values of the ML analysis and posterior probabilities of the corresponding Bayesian inference (BI) analysis are indicated at each node, when higher than 50%/0.5.—B. BI tree of the four genes analysed under the 'multispecies coalescent' model (see Material and methods). Only species represented by at least three of the four genes have been included (see Fig. S3 for the analysis including all species). Posterior probabilities are indicated at each node, when higher than 0.5. Out-groups are omitted.

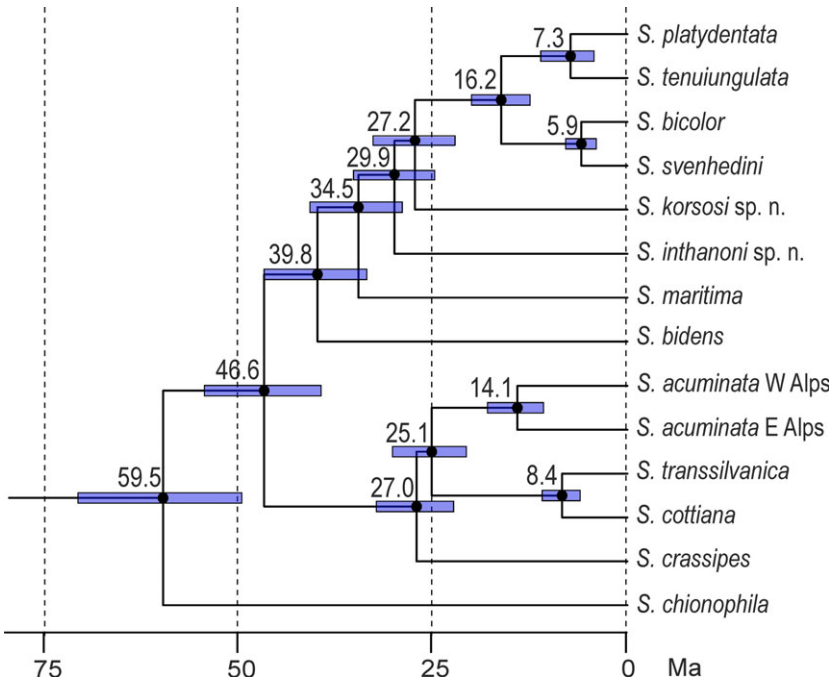


**Fig. 3** Divergence times estimated with a Bayesian inference analysis of the concatenated sequences of the four genes. Median values are indicated for each node, with 95% confidence intervals represented by bars. Out-groups are omitted.

***Divergence times***

Through the fossil-calibrated BI analysis of the sequence differentiation of the four genes, the last common ancestor of all the sampled species of *Strigamia* was estimated to have existed around 60 Ma ago (95% confidence interval: 49–70 Ma; Fig. 3). The two major clades of species (mainly European vs. mainly Eastern Asian) were estimated to have split around 47 Ma ago (39–54 Ma). Within the mainly Eastern Asian clade, *S. maritima* was estimated to have diverged from the closest relatives since around 35 Ma ago (29–41 Ma). The last common ancestor of the mainly European clade was estimated at around 27 Ma (22–32 Ma). For all the pairs of sister species in our sample, the divergence ages were comprised in the range 4–11 Ma, but an even more ancient divergence was estimated between the sequences of the two putative conspecific specimens of *S. acuminata*.

After the BI analysis of the time divergence, we have verified that the posteriorly sampled marginal prior densities are consistent with the prior distributions and that the posterior time estimates of the non-calibrated nodes are not conditioned by prior densities (Fig. S4).

## Discussion

By sampling species of *Strigamia* from throughout the global range of the genus, and by exploiting molecular sequences from both mitochondrial and nuclear genes, we could obtain a first preliminary hypothesis on the evolutionary history of *Strigamia*. Actually, we could analyse only about half of the total number of species recognized to date, and other species are probably still unknown. Nevertheless, considering the diversity of morphology and the geographical occurrence of these species upon their calibrated molecular phylogeny allows some insights into the history of diversification of this widespread genus throughout the Northern Hemisphere.

1. The extant global diversity of *Strigamia* most probably differentiated during the entire Cenozoic. In particular, we detected multiple major lineages that separated since the early Cenozoic, but we also estimated a late Cenozoic branching history for the two most diverse of these lineages.
2. The different assemblages of *Strigamia* species currently inhabiting different subcontinents mainly originated from different deeply separated lineages. In particular, the species assemblage inhabiting a large part of the Western Palearctic region (Europe, from Great Britain and Scandinavia southwards to Sicily, and nearby Asia; Fig. 1) and the comparably diverse species assemblage inhabiting the Oriental region (from south-eastern Siberia and Sakhalin Island, southwards to the inner mountains of Indochina; Fig. 1) originated mainly through the independent diversification of two major lineages separated from each other since the early

Cenozoic. Instead, the different species assemblage inhabiting North America (northwards to Alaska and the Aleutian islands, southwards to California and other southern States) had a more composite origin, from multiple deeply diverging lineages.

3. A hitherto underestimated diversity of *Strigamia* species is actually present in Eastern Asia. It is at least comparable, possibly surpassing, that known in North America and Europe, in terms of number of species, divergence age, geographical extent and morphological disparity. A long-standing view was that the bulk of *Strigamia* richness and diversity was in North America and Europe, whereas only a much lower number of quite similar species was present in Eastern Asia. However, such a biased view was conditioned by the unbalanced taxonomic surveys carried out so far between different continents. Compelling evidence comes from the fact that three of the seven species sampled from Eastern Asia for our analysis turned out to be morphologically distinct but still undescribed, and that they contribute to significantly widen the overall anatomical diversity of *Strigamia* species occurring in that region (Appendix S1). With regard to body size, *S. korsosi* sp. n. and *S. nana* sp. n. do not surpass 15 mm in length and specimens of the latter may have as low as 31 pairs of legs, whereas the previously known smallest Asian species (*Strigamia sibirica*) reaches 19 mm at least and has at least 33 pairs of legs. With regard to the number and arrangement of the pores of the coxal organs, S. *nana* sp. n. develops no more than 5 pores per side when fully grown, whereas all previously known Eastern Asian species develop at least 8–9 pores per side; moreover, *S. korsosi* sp. n. features as the single species in that area to develop all coxal pores clustered mesally, while at least a single pore invariantly develops separated from the remaining pores in *S. svenhedini*.
4. *Strigamia maritima*, which is currently limited to the seashores of northern Europe, from Great Britain eastwards to Scandinavia, is possibly representative of a quite ancient clade that is more strictly related to the Eastern Asian major lineage than to the European one, and separated most probably before the great diversification of the two lineages. Some morphological characters had already prompted past taxonomists to recognize *S. maritima* as quite distinct from most other European species. Indeed, Verhoeff (1935) speculated explicitly on a recent evolutionary separation between *S. maritima* and *Strigamia japonica*, which was originally described by him as a subspecies of the former, during the recent climatic fluctuations. Our analysis offers the first insights into the phylogenetic position, tempo and biogeographical context of *S. maritima*, which singles out among all other species of *Strigamia* for its unique littoral ecology and is of primary interest as model species within evolutionary developmental biology.

## Acknowledgements

We thank many colleagues who provided precious specimens for the analyses, including S. Andreev, P. Beron, H. Enghoff, S. Hasegawa, S. Karasawa and Z. Korsós. The research was supported by the Italian MIUR. L. Dányi was supported by a Hungarian Scientific Research Grant (OTKA No. 100369) and a grant of the National Science Council, Taiwan (NSC 98-2911-I-029-003-2), and F. Bortolin by a grant of the University of Padova (CPDA115439/11).

## References

Aberer, A. J., Krompass, D. & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology*, *62*, 162–166.

Arthur, W. & Chipman, A. D. (2005). The centipede *Strigamia maritima*: what it can tell us about the development and evolution of segmentation. *BioEssays*, *27*, 653–660.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A. & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, *29*, 2157–2167.

Barber, A. D. (2009). Littoral myriapods: a review. *Soil Organisms*, *81*, 735–760.

Berger, S. A., Krompass, D. & Stamatakis, A. (2011). Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology*, *60*, 291–302.

Bonato, L., Edgecombe, G. D., Lewis, J. G. E., Minelli, A., Pereira, L. A., Shelley, R. M. & Zapparoli, M. (2010). A common terminology for the external anatomy of centipedes (Chilopoda). *ZooKeys*, *69*, 17–51.

Bonato, L., Dányi, L., Socci, A. A. & Minelli, A. (2012). Species diversity of *Strigamia* Gray, 1843 (Chilopoda: Linotaeniidae): a preliminary synthesis. *Zootaxa*, *3593*, 1–39.

Bonato, L., Drago, L. & Murienne, J. (2014). Phylogeny of Geophilomorpha (Chilopoda) inferred from new morphological and molecular evidence. *Cladistics*, *30*, 485–507.

Brena, C. (2014). The embryoid development of *Strigamia maritima* and its bearing on post-embryonic segmentation of geophilomorph centipedes. *Frontiers in Zoology*, *11*, 58.

Chipman, A. D. & Akam, M. (2008). The segmentation cascade in the centipede *Strigamia maritima*: involvement of the Notch pathway and pair-rule gene homologues. *Developmental Biology*, *319*, 160–169.

Chipman, A. D., Ferrier, D. E., Brena, C., Qu, J., Hughes, D. S., Schröder, R., Torres-Oliva, M., Znassi, N., Jiang, H., Almeida, F. C., Alonso, C. R., Apostolou, Z., Aqrawi, P., Arthur, W., Barna, J. C., Blankenburg, K. P., Brites, D., Capella-Gutiérrez, S., Coyle, M., Dearden, P. K., Du Pasquier, L., Duncan, E. J., Ebert, D., Eibner, C., Erikson, G., Evans, P. D., Extavour, C. G., Francisco, L., Gabaldón, T., Gillis, W. J., Goodwin-Horn, E. A., Green, J. E., Griffiths-Jones, S., Grimmelikhuijzen, C. J.,

Gubbala, S., Guigó, R., Han, Y., Hauser, F., Havlak, P., Hayden, L., Helbing, S., Holder, M., Hui, J. H., Hunn, J. P., Hunnekuhl, V. S., Jackson, L., Javaid, M., Jhangiani, S. N., Jiggins, F. M., Jones, T. E., Kaiser, T. S., Kalra, D., Kenny, N. J., Korchina, V., Kovar, C. L., Kraus, F. B., Lapraz, F., Lee, S. L., Lv, J., Mandapat, C., Manning, G., Mariotti, M., Mata, R., Mathew, T., Neumann, T., Newsham, I., Ngo, D. N., Ninova, M., Okwuonu, G., Ongeri, F., Palmer, W. J., Patil, S., Patraquim, P., Pham, C., Pu, L. L., Putman, N. H., Rabouille, C., Ramos, O. M., Rhodes, A. C., Robertson, H. E., Robertson, H. M., Ronshaugen, M., Rozas, J., Saada, N., Sánchez-Gracia, A., Scherer, S. E., Schurko, A. M., Siggens, K. W., Simmons, D., Stief, A., Stolle, E., Telford, M. J., Tessmar-Raible, K., Thornton, R., van der Zee, M., von Haeseler, A., Williams, J. M., Willis, J. H., Wu, Y., Zou, X., Lawson, D., Muzny, D. M., Worley, K. C., Gibbs, R. A., Akam, M. & Richards, S. (2014). The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLOS Biology*, *12*, e1002005.

Criscuolo, A. (2011). MorePhyML: improving the phylogenetic tree space exploration with PhyML 3. *Molecular Phylogenetics and Evolution*, *61*, 944–948.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012). JModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*, 772.

Degnan, J. H. & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, *24*, 332–340.

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973.

Edgecombe, G. D. (2007). Centipede systematics: progress and problems. *Zootaxa*, *1668*, 327–341.

Edgecombe, G. D. & Giribet, G. (2004). Adding mitochondrial sequence data (16S rRNA and cytochrome c oxidase subunit I) to the phylogeny of centipedes (Myriapoda: Chilopoda): an analysis of morphology and four molecular loci. *Journal of Zoological Systematics and Evolutionary Research*, *42*, 89–134.

Edgecombe, G. D. & Giribet, G. (2007). Evolutionary biology of centipedes (Myriapoda: Chilopoda). *Annual Review of Entomology*, *52*, 151–170.

Fernández, R., Laumer, C. E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P. P., Pérez-Porro, A. R., Edgecombe, G. D. & Giribet, G. (2014). Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. *Molecular Biology and Evolution*, *31*, 1500–1513.

Ferreira, M. A. R. & Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics*, *36*, 355–368.

Green, J. E., Dalíková, M., Sahara, K., Marec, F. & Akam, M. (2016). XX/XY system of sex determination in the geophilomorph centipede *Strigamia maritima*. *PLoS One*, *11*, e0150292.

Heled, J. & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580.

Heled, J. & Drummond, A. J. (2011). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, *61*, 138–149.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, *33*, 511–518.

Kück, P. (2009). ALICUT: A perlscript which cuts ALISCORE identified RSS. Ver. 2. Bonn: Department of Bioinformatics, Zoologisches Forschungsmuseum A. Koenig.

Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701.

Lartillot, N., Lepage, T. & Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, *25*, 2286–2288.

Lewis, J. G. E. (1961). The life history and ecology of the littoral centipede *Strigamia* (=*Scolioplanes*) *maritima* (Leach). *Proceedings of the Zoological Society of London*, *137*, 221–248.

Maddison, W. P. & Maddison, D. R. (2001). Mesquite: a modular system for evolutionary analysis. Ver. 2.75. Available via http://mesquiteproject.org

Maruzzo, D. & Bonato, L. (2014). Morphology and diversity of the forcipules in *Strigamia* centipedes (Chilopoda, Geophilomorpha). *Arthropod Structure & Development*, *43*, 17–25.

Minelli, A. (Ed.) (2011). *Treatise on Zoology – The Myriapoda*. Volume 1. Leiden: Brill.

Misof, B. & Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology*, *58*, 21–34.

Mundel, P. (1979). The centipedes (Chilopoda) of the Mazon Creek. In M. H. Nitecki (Ed.) *Mazon Creek Fossils* (pp. 361–378). New York: Academic Press.

Murakami, Y. (1993). Chilopoda. In *List of Animals and Plants in Japan. Invertebrates 1* (pp. 103–105). Tokyo: Environment Agency Japan, Nature and Environment Research Center.

Murienne, J., Edgecombe, G. D. & Giribet, G. (2010). Including secondary structure, fossils and molecular dating in the centipede tree of life. *Molecular Phylogenetics and Evolution*, *57*, 301–313.

Pereira, L. A. (2000). The preparation of centipedes for microscopical examination with particular reference to the Geophilomorpha. *Bulletin of the British Myriapod Group*, *16*, 22–25.

Rambaut, A. & Drummond, A. (2013). Tracer. Ver. 1.6. Available via http://tree.bio.ed.ac.uk/software/tracer

Rannala, B. & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*, 1645–1656.

Robertson, H. E., Lapraz, F., Rhodes, A. C. & Telford, M. J. (2015). The complete mitochondrial genome of the geophilomorph centipede *Strigamia maritima*. *PLoS One*, *10*, e121369.

Rosenberg, J. (1989). Bestimmungsschlüssel für metteleuropäische Erdläufer (Geophilomorpha) anhand der Coxalporen. *Acta Biologica Benrodis*, *1*, 133–141.

Sela, I., Ashkenazy, H., Katoh, K. & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, *43*, W7–W14.

Shimodaira, H. & Hasegawa, M. (2002). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, *17*, 1246–1247.

Shinohara, K. (1999). Chilopoda. In J. Aoki (Ed.) *Pictorial Keys to Soil Animals of Japan* (pp. 693–709). Tokyo: Tokay University Press.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729.

Vaidya, G., Lohman, D. J. & Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*, 171–180.

Verhoeff, K. W. (1935). Über *Scolioplanes* (Chilopoda). *Zoologischer Anzeiger*, *111*, 10–23.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Description of the new species of *Strigamia*.

**Table S1.** Primers and temperature conditions used for gene amplification.

**Table S2.** Range of variation of some anatomical characters among species of *Strigamia* inhabiting the three major parts of the global distribution of the genus.

**Fig. S1.** Consensus trees between ML and BI analyses of single genes.

**Fig. S2.** ML tree of the concatenated sequences of the four genes, including all species. Bootstrap values and posterior probabilities of the corresponding BI analysis are indicated at each node, when higher than 50%/0.5.

**Fig. S3.** BI tree of the four genes analysed under the 'multispecies coalescent' model, including all species.

**Fig. S4.** Divergence times estimated with a BI analysis of the concatenated sequences of the four genes, by sampling from the prior only.