



HAL
open science

Mitochondrial metagenomics reveals the ancient origin and phylodiversity of soil mites and provides a phylogeny of the Acari

Paula Arribas, Carmelo Andujar, Maria Lourdes Moraza, Benjamin Linard, Brent C. Emerson, Alfred P Vogler

► **To cite this version:**

Paula Arribas, Carmelo Andujar, Maria Lourdes Moraza, Benjamin Linard, Brent C. Emerson, et al.. Mitochondrial metagenomics reveals the ancient origin and phylodiversity of soil mites and provides a phylogeny of the Acari. *Molecular Biology and Evolution*, 2020, 37 (3), pp.683-694. 10.1093/molbev/msz255 . hal-02385328

HAL Id: hal-02385328

<https://hal.science/hal-02385328v1>

Submitted on 28 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Mitochondrial metagenomics reveals the ancient origin and
phylodiversity of soil mites and provides a phylogeny of the Acari**

Paula Arribas^{1,2,3} *, Carmelo Andújar^{1,2,3}, María Lourdes Moraza⁴, Benjamin Linard^{2,5,6},
Brent C. Emerson¹ & Alfried P. Vogler^{2,3}

¹*Island Ecology and Evolution Research Group (IPNA-CSIC), Astrofísico Fco. Sánchez 3, 38206 La Laguna, Tenerife, Spain.*

²*Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.*

³*Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK.*

⁴*Departamento de Biología Ambiental, Universidad de Navarra, Facultad de Ciencias, Pamplona 31080, Spain.*

⁵*LIRMM, Université de Montpellier, CNRS, Montpellier, France.*

⁶*ISEM, Université de Montpellier, CNRS, IRD, EPHE, CIRAD, INRAP, Montpellier, France.*

*Corresponding author: pauarribas@ipna.csic.es

Abstract

High-throughput DNA methods hold great promise for phylogenetic analysis of lineages that are difficult to study with conventional molecular and morphological approaches. The mites (Acari), and in particular the highly diverse soil-dwelling lineages, are among the least known branches of the metazoan Tree-of-Life. We extracted numerous minute mites from soils in an area of mixed forest and grassland in southern Iberia. Selected specimens representing the full morphological diversity were shotgun sequenced in bulk, followed by genome assembly of short reads from the mixture, which produced >100 mitochondrial genomes representing diverse acarine lineages. Phylogenetic analyses in combination with taxonomically limited mitogenomes available publicly resulted in plausible trees defining basal relationships of the Acari. Several critical nodes were supported by ancestral-state reconstructions of mitochondrial gene rearrangements. Molecular calibration placed the minimum age for the common ancestor of the superorder Acariformes, which includes most soil-dwelling mites, to the Cambrian-Ordovician (likely within 455–552 Mya), while the origin of the superorder Parasitiformes was placed later in the Carboniferous-Permian. Most family-level taxa within the Acariformes were dated to the Jurassic and Triassic. The ancient origin of Acariformes and the early diversification of major extant lineages linked to the soil are consistent with a pioneering role for mites in building the earliest terrestrial ecosystems.

Keywords: mitochondrial phylogenomics; metagenome skimming; soil biodiversity, arthropod terrestrialisation.

Introduction

Mites (Acari) are present worldwide in virtually any locality capable of supporting life, from Antarctic nunataks to the follicles of human hair (Krantz and Walter 2009). They comprise approximately 55,000 described species, but the actual number is estimated to be 500,000 to 1,000,000, and possibly a lot more, as new mite species are routinely discovered even in previously well-collected places (Walter and Proctor 1999). Mites have a long evolutionary history, dating back to at least 410 million years ago (Mya) based on their presence in early terrestrial fossil layers (Hirst 1923; Dubinin 1962). However, their evolutionary history is only partially understood, and mites remain one of the least studied major branches of the animal Tree-of-Life (Dunlop and Alberti 2008; Krantz and Walter 2009). The uncertainties in particular concern the times of origin of the earliest mite lineages. Recent studies place them either to the pre-Cambrian, implying an early diversification in an aquatic or semiaquatic environment (Schaefer et al. 2010), or to the Devonian well after the formation of terrestrial ecosystems (Dabert et al. 2010; Xue et al. 2017).

The greatest number of mite species is found in soils, where they constitute an essential part of the mesofauna. They predominantly exist in the soil matrix and are implicated in the soil trophic chain as predators, scavengers or primary consumers (Burgess and Raw 1967). Mites are essential for soil viability, aeration and water retention, decomposition of plant matter, microbial biomass control and nutrient cycling (Bardgett et al. 2005; Bardgett and van der Putten 2014). These functional roles, together with their presumed antiquity, point to them as a likely key factor in the evolution of the earliest soils. Studies of the origin and deep-level diversification of the Acari, including the generation of time-calibrated phylogenetic trees, are thus an essential first step toward the reliable characterization of soil mite diversity and community evolution (e.g. Bardgett and van der Putten 2014; Veresoglou et al. 2015).

The study of soil Acari is hampered by extreme body miniaturization, which has likely stifled the interest of taxonomists in this group and delayed the application of new sequencing technologies. As a consequence, DNA-based studies of basal relationships of Acari are scarce and have largely neglected the soil-dwelling groups. Existing data instead are skewed towards lineages closely associated with humans, such as ticks, dust

mites or herbivorous agricultural pests. These groups are also represented by more extensive sequence data, including mitochondrial genomes and transcriptomes (see Sharma et al. 2014; Xue et al. 2016; Xue et al. 2017). However, in comparison with other large groups of arthropods, the study of deep level phylogenetics within Acari is lagging and mostly based on a limited set of mitochondrial and nuclear loci (*cox1*, 18S and 28S rRNA; e.g. Klompen et al. 2007; Dabert et al. 2010; Pepato et al. 2010; Schaefer et al. 2010; Pepato and Klimov 2015) that are only slowly being extended to include a wider range of markers (e.g. Klimov et al. 2018).

Taxonomically, the Acari can readily be separated into two superorders, the Acariformes and Parasitiformes, which broadly correspond to the predominantly soil-dwelling and the parasitic forms, respectively, but most acarologists would agree that both lineages are not closely related and thus Acari are not monophyletic. To date, neither morphological nor molecular studies have convincingly resolved the sister group relationships of either superorder within the wider arachnids (for review see Dunlop and Alberti 2008). Some molecular studies have proposed the Acariformes as sister lineage to Solifugae (Pepato and Klimov 2015), whereas more recent transcriptome studies have found an affiliation with Pseudoscorpiones (Sharma et al. 2014). Within these two superorders, the cladistic hypotheses for Parasitiformes are in general agreement across studies and methodologies, while the systematics of the superorder Acariformes is quite convoluted. For example, the basal dichotomy of the two orders Sarcoptiformes and Trombidiformes has been recently questioned due to the dubious position of Eriophyoidea (gall mites and rust mites) and the uncertain relationships between the non-eriophyoid Trombidiformes and Sarcoptiformes (Klimov et al. 2018).

Faced with the logistical challenges for a comprehensive phylogenetic analysis of the Acari, next-generation sequencing provides new opportunities for data acquisition from large numbers of very small and morphologically cryptic specimens. In particular, shotgun ‘mitochondrial metagenomics’ (MMG) can be used for the simultaneous sequencing of mitochondrial genomes from numerous specimens, based on metagenome skimming from bulk or pooled samples, which greatly reduces the amount of DNA required for each individual (Crampton-Platt et al. 2015; Linard et al. 2015). Mitochondrial genomes are powerful phylogenetic markers, and have been used successfully for resolving relationships at intra-ordinal and inter-familial levels for

arthropods, particularly under dense taxon sampling and with the use of adequate phylogenetic models (Timmermans et al. 2016; Basso et al. 2017; Wang et al. 2017).

Here we applied the MMG approach to a sample of mite specimens extracted from natural forest and grassland soils in southern Iberia, which generated >100 mitochondrial genomes representing major acarine lineages linked to the soil environment and nearly tripled the number of mite families with publicly available mitochondrial genomes. The increased taxon, lineage and gene sampling allowed for a robust phylogenomic analyses to i) test existing cladistic hypotheses for the relationships among lineages in the Acariformes and Parasitiformes, with special focus on the Trombidiformes – Sarcoptiformes divide; ii) estimate the temporal frame for the origin and early diversification of major mite lineages and evaluate its consistency with fossil evidence and previous molecular clock approaches; and iii) shed light on the origin and phylodiversity of soil mites and their potential role in the earliest terrestrial communities. Additionally, this study serves as paradigmatic example of the 'soup to tree' approach to assemble the Tree-of-Life for the 'hidden biodiversity' of the soil that still remains largely unknown.

Results

The mitochondrial genomes of Acari

Soil samples obtained from superficial and deep soil layers in a landscape of *Quercus* forest and wet grassland at Sierra de Grazalema (Southern Iberian Peninsula) (Table S1) produced large numbers of mites. Specimens were selected for recognizable morphological diversity under a dissection microscope, subjected to non-destructive individual DNA extractions, and morphologically identified to species or genus for adult forms, or to family for immatures (Table S1). The MMG procedure yielded 106 mitogenomes, which each were assigned to a particular morphological specimen by match to a *coxI* (barcode) bait obtained from the same specimens with Sanger sequencing. All baits had an exact sequence match to the corresponding portion of the mitogenome, which together with the strict process of mitogenome assembly by multiple genome assemblers and subsequent steps of gene annotation, alignment and curation, allowed us to validate the dataset and to demonstrate the minimal, if any, impact of mitochondrial nuclear copies or spurious assemblies of the mitogenomes.

A total of 71 mitogenomes were complete (i.e. included 13 protein coding and 2 ribosomal RNA genes) and the remaining 35 contigs ranged from 5000 to 13000 bp in length (see Table S1). The taxonomic profile of the newly assembled mitogenomes covered the main soil lineages across Acari: 22 mitogenomes from the superorder Parasitiformes, all of them from the order Mesostigmata; and 84 from the superorder Acariformes, including 15 from the order Trombidiformes and 69 from the order Sarcoptiformes. The new dataset was taxonomically complementary to 78 mitogenomes of Acari obtainable from GenBank (Mitochondrial Genome database), which mostly represented 'human associated' species: 46 Parasitiformes including 40 Ixodida (ticks); 23 Trombidiformes, of which almost all were agricultural pests (e.g. *Tetranychus*) or mammal parasites (e.g. *Demodex* or *Leptotrombidium*); and nine Sarcoptiformes including eight Astigmatina (dust mites) and a single species of Oribatida (*Steganacarus magnus*). The combined dataset comprised 184 mitogenomes of Acari from 62 families, of which 44 were exclusively present in the newly assembled mitogenomes (Table S1). In addition, 20 outgroups of Arachnida (Amblypygi, Araneae, Opiliones, Pseudoscorpiones, Ricinulei, Scorpiones, Solifugae and Uropygi) and Xiphosura were obtained from GenBank, for a total of 204 mitogenomes and a concatenated matrix of individual gene alignments of 15,047 bp.

The deep level phylogeny of Acari

Under the most appropriate alignment settings and substitution models (Table S2), phylogenetic analyses using different algorithms and data treatments resulted in mostly congruent, well-supported trees (see Figure 1 and Figures S1-S9), whose characteristics were tabulated to indicate monophyly and node support for the main lineages of Acari across the different settings (Table S3). The two superorders Parasitiformes and Acariformes were well supported, but their positions relative to the arachnid outgroups differed across the analyses. All topologies resulted in a diphyletic Acari, placing the Parasitiformes as sister group to a clade comprising the Acariformes and various non-acarine outgroups. The Acariformes were grouped with Pseudoscorpiones in most analyses, but most other basal relationships among the non-acarine groups varied across the analyses and were generally poorly supported. Within the superorder Parasitiformes we obtained highly consistent results, with the three monophyletic orders arranged as (Mesostigmata (Ixodida + Holothyrida)). Within Mesostigmata two major supported

clades include Uropodina together with Cercomegistina and the cohort (infraorder) Gamasina. In the latter, the subcohort (hyporder) Epicriina was sister group to, or included in a clade comprising different lineages of the subcohorts Dermanyssiae and Parasitiae.

Within the superorder Acariformes, the order Trombidiformes in the traditional sense (see below) was always paraphyletic in contrast to a monophyletic Sarcoptiformes. Most analyses recovered a basal split of a well-supported clade composed of the families Eriophyidae and Diptilomiopidae (=Eriophyoidea of the also paraphyletic supercohort Eupodides), which was the sister group to a clade of the remaining Trombidiformes plus Sarcoptiformes. The non-eriophyoid Trombidiformes consisted of three well supported clades, although of uncertain relative position, including i) the supercohort Eupodides (excluding Eriophyoidea) and Labidostommatides, ii) a single representative from the cohort Anystina (supercohort Anystides) and iii) a clade containing the monophyletic supercohort Eleutherengonides as sister group of the monophyletic cohort Parasintengonina (as representative of the supercohort Anystides). ML analyses coding the data as amino acids (*AA* dataset) also recovered the ancestral split of the traditional Trombidiformes, but in this case the families Eriophyidae and Diptilomiopidae (Eriophyoidea) were grouped with the supercohort Eleutherengonides.

Within the Sarcoptiformes (including multiple Oribatida and Astigmatina but lacking samples of the endeostigmatan lineages and Paleosomata that may be part of this group), analyses consistently recovered the monophyletic supercohort Enarthronotides as sister group to all other Sarcoptiformes. The latter was split into four well supported clades, including two clades of the supercohort Mixonomatides and two clades of the supercohort Desmonomatides, including the monophyletic cohort Astigmatina as sister lineage to a well-supported clade comprised of the cohorts Nothrina and Brachypilina. The relationships among these four groups of Sarcoptiformes varied across the analyses, sometimes resulting in a paraphyletic supercohort Mixonomatides. However, the monophyly of the supercohort Desmonomatides was recovered consistently, with Astigmatina as sister clade to the rest of the Desmonomatides. At the family level, a high proportion of the families (>80%) was found as well supported clades in all analyses (for details see Table S3).

The evolution of mitochondrial gene rearrangements across Acari

Mitochondrial gene order is known to be highly variable in Acari (Xue et al. 2016) and may be used as a phylogenetic character (e.g. Basso et al. 2017; Wang et al. 2017). Across the complete dataset (204 taxa, including outgroups) we identified 37 different 'mitotypes' defined by different gene orders of protein and rRNA coding genes, in addition to more frequent rearrangements of tRNA genes not specifically studied here. A total of 20 mitotypes were exclusive to the newly assembled dataset. Rearrangements of the ribosomal (*rrnL* and *rrnS*) and adjacent *nad1* and *nad2* genes were the most frequent across Acari (see Table S1 and Fig. S10). *Mitotype 1* represented the putative ancestral arthropod mitochondrial gene order and was also the most widespread in Acari, while 20 of the 37 identified mitotypes were restricted to a single terminal (Fig. 1 and Table S1). Ancestral state reconstructions with BEAST and TreeREx were highly congruent and inferred *mitotype 1* as the ancestral gene order for both Acariformes and Parasitiformes (also shared with the arachnid outgroups). Within Parasitiformes, *mitotype 1* remained almost unchanged within the orders Holothryda and Ixodida, where a single translocation of the *nad1-rrnL-rrnS* gene block occurred within the family Ixodidae. Within the order Mesostigmata, rearrangements were inferred to result from single translocations (e.g. to *mitotype 17* characterising the cohort Uropodina) or required multiple translocation steps (e.g. those within the cohort Gamasina). In Acariformes, three gene orders were derived independently from *mitotype 1* in the early diversification of the group. An inversion of the ribosomal RNA genes to *mitotype 3* was exclusive to the families Eriophyidae and Diptilomiopidae (Eriophyoidea) at their basal split from the rest of Acariformes. Another translocation of the *rrnS* gene (*mitotype 8*) in Trombidiformes gave rise to a complex series of rearrangements across the cohort Parasitengonina and the family Tetranychidae. A further single translocation of the *nad2* gene (*mitotype 20*) supported the Sarcoptiformes, including the supercohorts Enarthonotides and Desmonomatides (excluding Astigmatina). The remaining clades of Sarcoptiformes exhibited various gene orders derived from *mitotype 20*. The Astigmatina was characterised by a complex rearrangement; while two independent translocations derived from *mitotype 20* marked each of the two clades within the supercohort Mixonomatides.

The temporal frame of Acari evolution

The tree was dated using the oldest available fossils for lineages of Acari at various hierarchical levels (reviewed by Dunlop and Selden 2009) (see Table S4 for details). These ages were used as hard minimum bounds for the stem node of the corresponding lineages by applying uniform distributions with uninformative maximum age constraints. This way of calibration avoids an important source of uncertainty that arises from implicitly constraining maximum ages at every calibration point when other parametric functions are used to calibrate (see Methods). We performed the analyses for three nested sets of calibration points using from 14 to 22 fossils, each of which representing the oldest records for the respective taxon (Dunlop and Selden 2009), but whose placement on our trees is less certain for some of them (see Methods and Table S4). The three analyses were conducted with and without constraining the maximum ages of the Acariformes and Parasitiformes crown groups based on the 99% confidence interval age for their first fossil appearance, as estimated by the statistical analysis of the available fossil series (see Methods and Table S4 and S5 for details). The results for the analyses using each of the three sets of minimum hard bounds (14 to 22 fossil minimum ages) were very similar, and comparisons of those with and without maximum age constraints resulted in similar median ages. The range of ages with maximum probabilities also coincided and the 95% highest posterior density intervals (HPD) were fully overlapping but, as expected, the calibration analyses without constrained maximum ages tended to have wider ranges due to longer tails of the marginal probability distributions to the past (Fig. S16). The results of all analyses supported minimum ages for the origin of Acariformes in the Cambrian-Ordovician, at 455 Mya (95% HPD: 417–508 Mya), 458 (419–511) Mya, and 469 (424–512) Mya for the three analyses with maximum constraints, and at 496 (412–739) Mya, at 552 (415–810) Mya and 538 (415–805) Mya for the corresponding analyses without any upper limit to divergence times (Fig. 1 and Figs. S11-S16).

The time line of Acari evolution, estimated under the most conservative set of calibration points and maximum ages constraints (*14min+2max* calibration set, Fig. 1), placed the minimum age for the origin of Acariformes (crown group age) at 455 (417–508) Mya, followed by the early diversification of Eriophyoidea and the Trombidiformes during the Silurian and early Devonian. The minimum age for the origin of the Sarcoptiformes (here **Oribatida and Astigmatina**) was estimated to be in the

early Devonian (389 Mya, HDP: 355–440 Mya), with multiple basal lineages of Oribatida dated back at least to the Devonian and Carboniferous. In Parasitiformes, the age for the crown node was estimated at 310 (268–359) Mya, placing their minimum age to the Carboniferous. The subsequent diversification of Mesostigmata and Ixodida were estimated at 279 (238–329) Mya and 250 (216–289 Mya) Mya respectively. All families represented in the tree (with two or more representatives) dated back at least to the Jurassic and Cretaceous while Oppiidae and Phthiracaridae dated to the Triassic and Trombiculidae to the Permian. The times of origin of the different supercohorts and cohorts varied greatly. The estimated rate of molecular evolution was 0.0146 substitutions per site per million years per lineage (subs/s/Ma/l) (95% HPD: 0.0082–0.022) for the protein coding genes, and an almost 10-fold lower rate for the *rrnL* and *rrnS* genes at 0.00175 and 0.00189 subs/s/Ma/l, respectively. These rates were similar to those estimated for other arthropods in recent studies (e.g. Papadopoulou et al. 2010; Andújar et al. 2012).

Discussion

The deep phylogeny of Acari

The mitogenome data are notable for their power to resolve the deep level systematics of the Acari, despite great divergences of lineages within both superorders. The mitogenomes also produce strong evidence for the monophyly of the two superorders Acariformes and Parasitiformes and support the findings of recent studies that both groups are only distantly related, while the relationships of either group with other Arachnida lineages remain ambiguous. In the ML analysis the Acariformes was a well-supported sister group of Pseudoscorpiones, contrary to some molecular studies placing Acariformes as sister lineage to Solifugae, in the so-called Poecilophysidea (e.g. Pepato and Klimov 2015; Klimov et al. 2018), but in accordance with studies using mitochondrial genomes (Xue et al. 2016; Xue et al. 2017) and more than 3600 orthologs from transcriptome sequencing by Sharma et al. (2014). The latter study attributed the Acariformes + Pseudoscorpiones to spurious long-branch attraction that was not obtained when only considering the 200 slowest-evolving genes, and given the limited taxon sampling, the authors emphasized the need to break long branches by including putatively slowly evolving exemplars. Our analysis showed that the lineages used by

Sharma et al. (2014), specifically Astigmatina and Tetranychidae, also harbor the fastest evolving mitogenomes of all Acariformes (see branch lengths in Figs. S1-S9), but despite the much broader taxon representation that presumably overcomes these long-branch artifacts we still find the Pseudoscorpions + Acariformes clade. However, the PhyloBayes analysis (which uses a model that is less sensitive to rate heterogeneity) did not fully support this node, which leaves open the possibility that biases in evolutionary rates affect the deepest acariform relationships.

Within the Acariformes, the mitogenomes support the Sarcoptiformes but reject Trombidiformes in the traditional sense due to the absence of the Eriophyoidea. Recent molecular studies have already moved away from the Trombidiformes-Sarcoptiformes dichotomy of earlier cladistic analyses and are producing evidence for multiple basal splits (Klimov et al. 2018). In our analyses, the deepest branch was occupied by Eriophyoidea, whose phylogenetic position is still unclear but recently hypothesised to be near the family Nematalycidae, a member of the ancient Endeostigmata (Bolton et al. 2017; Klimov et al. 2018). The latter is composed of several families of vermiform mites inhabiting the deep-soil, and maybe an early branch of, or even the sister group to all other Acariformes. The critical Endeostigmata unfortunately were not encountered in our sample although several families are expected in this habitat. If the placement of Eriophyoidea within Nematalycidae is correct, these mites are the only representative of Endeostigmata available for this study, although they constitute a very derived lineage not associated with the soil habitat. Yet, this placement is consistent with the initial separation of the Eriophyoidea from all other Acariformes in our trees, which is additionally characterized by an exclusive mitochondrial rearrangement (mitotype 3), derived from the ancestral order of Acariformes (mitotype 1) by a single inversion of the ribosomal genes, and which is now a possible clade marker for the wider Endeostigmata.

After excluding Eriophyoidea, the remaining lineages of Trombidiformes were strongly supported as sister group to Sarcoptiformes. Previous studies differed regarding this conclusion dependent on the chosen genes and partition strategies (Xue et al. 2016; Bolton et al. 2017; Xue et al. 2017; Klimov et al. 2018). Within the (non-eriophyoid) Trombidiformes, two basal branches contributed by newly sampled soil mitogenomes corresponding to the suborders Eupodides, Labiostommatides and Anystides, while Eleutherongonides was represented by existing sequences mostly of spider mites

(Tetranychidae). The results on the internal classification of the trombidiform supercohorts are only partially congruent with the two main hypotheses from morphology presented in Norton et al. (1993) and Lindquist (1996) and further support the need for a deep review of relationships across the entire eriophyoid-trombidiform complex, with focus on the potentially non-monophyletic supercohorts Eupodides and Anystides.

On the contrary, for the Sarcoptiformes the mitochondrial genomes largely confirm existing cladistic hypotheses, including i) the monophyly of the order Sarcoptiformes, ii) the sister relationship of the supercohort Enarthronotides and Mixonomatides + Desmonomatides, and iii) the monophyly of the supercohort Desmonomatides (Pepato and Klimov 2015; Klimov et al. 2018). However, our results raise the potential paraphyly of the supercohort Mixonomatides, which will require further review. The position of Astigmatina (a hyperdiverse group including many medically and economically important species) within Desmonomatides is consistent with recent molecular and some morphology-based studies (OConnor 1984; Norton 1998). This finding is evidence that mitochondrial genome data are in this case less susceptible to long-branch attraction than nuclear ribosomal genes (Dabert et al. 2010).

Finally, the second major branch of the Acari, the superorder Parasitiformes, constitutes a deep lineage within the class Arachnida that goes back even further than the Acariformes, but the crown group is much younger. Our study recovered a great diversity of hitherto unsampled lineages in the order Mesostigmata, which unlike many other Parasitiformes are free-living and include predatory soil mites. The tree is highly consistent with a previous molecular study based on rRNA markers (Klompen et al. 2007), with two well supported clades, corresponding respectively to the orders Mesostigmata and Ixodida plus Holothyrida. Within the Mesostigmata, relationships among the present cohorts are in agreement with the rRNA study and further effort should be focused on adding missing lineages such as Sejida, the presumed most primitive suborder of Mesostigmata.

The ancient diversity of Acari in soils and their role in early terrestrial ecosystems

Our broad focus on hitherto neglected soil-dwelling lineages of mites and largely complete mitogenomes for the estimation of divergence times places the minimum age

of Acariformes to the Cambrian-Ordovician, with median estimated ages of the Acariformes crown node for the different molecular calibration analyses ranging from 455 to 552 Mya. The corresponding marginal probability distributions (95% HPD intervals) extend these ranges to 417 to 512 Mya for calibrations schemes using maximum age constraints and 413 to 810 Mya without such constraints (Fig. S16). The favored Cambrian-Ordovician origin of the Acariformes is included in the age ranges obtained with all calibration schemes, while the dates extending far into the Precambrian are only obtained under the broad intervals obtained without maximum age constraints. The current study therefore suggests an earlier origin of Acariformes than previous molecular calibrations from rRNA genes and mitogenomes dated to the early Devonian (e.g. Dabert et al. 2010; Xue et al. 2017), but is in good agreement with recent molecular estimations for the origin of some of the main terrestrial arthropod lineages (Rota-Stabelli et al. 2013; Lozano-Fernandez et al. 2016). The estimated origin of Sarcoptiformes in the Silurian-early Devonian (Oribatida and Astigmatina in our study) matches the classic view based on the fossil record of Oribatida (Norton et al. 1988; Subías and Arillo 2002). However, these estimates are less ancient than those obtained by Schaefer et al. (2010) of a Precambrian origin, a hypothesis which is still plausible but unlikely with our results, as again the upper bound ages of the estimated 95% confidence intervals are overlapping with the Precambrian for our most relaxed analyses, but not overlapping with the more restricted ones (with maximum age constraints). While the precise age for the origin of the Acariformes is still an open topic, the range of minimum ages here estimated at the Cambrian-Ordovician supports an aquatic or semiaquatic origin of the group, as similarly has been proposed for other terrestrial arthropod lineages (Lozano-Fernandez et al. 2016).

The current age estimates may still be compromised to some extent by taxon sampling. For example, the earliest branch of the Acariformes in our tree is represented by the Eriophyoidea, but earlier diverging lineages could be present with more complete sampling of the presumably non-monophyletic Endeostigmata. However, these lineages are unlikely to change the estimated age of Acariformes dramatically because the branch length between the crown node of Acariformes and the stem node defining the split from the Pseudoscorpiones (but also Ricinulei, Opiliones) corresponds to only approximately 25 Mya. Thus, any more basal branches that might be added by the

inclusion of further Endeostigmata can only be placed within this limited age range. In a similar way, the choice of calibration constraints may cause uncertainties. For example, the time calibration of the Sarcoptiformes was constrained using the oldest fossil for the order dated to 392 Mya (Dunlop and Selden 2009), whereas an alternative calibration point could be the next inner Enarthronotides stem node. The fossil in question (*Protochthonius gilboa*, *Devonacarus sellnicki*) represents extinct lineages of monobasic genera and families (only established to accommodate these fossils), both considered early derivative members of Enarthronotides but with uncertain relationships (Norton et al. 1988). Therefore, we avoided the use of this fossil to constrain the Enarthronotides stem node, but favored a position deeper in the tree. Another reason supporting this deeper constraint is the uncertain monophyly of Enarthronotides, particularly regarding the early derivative lineages (see Schaefer et al. 2010; Klimov et al. 2018), and which in our trees is only represented by members of the family Lomanniidae. Overall, we consider that the use of the 392 Mya calibration for constraining the stem node of Sarcoptiformes (i.e. the node subtending the stem of Enarthronotides) is a more conservative and robust option to the goal of estimating minimum ages of origin for Sarcoptiformes, even if it may lead to a slight underestimate of the time of origin of the order.

Most lineages sampled for mitogenomes to date are of medical or agricultural importance, with very limited sampling from the soil (Xue et al. 2017). These lineages are generally younger in origin, particularly within the widely sampled Parasitiformes, whose crown group was placed in the late Carboniferous just over 300 Mya. The much increased inclusion of soil fauna phylodiversity (labelled in red letters in Fig. 1) adds critical early branches in both the Trombidiformes and Sarcoptiformes, and reveals approximately a dozen lineages of Acariformes to have originated prior to the end of the Devonian and 19 lineages dated to before 250 Mya, i.e. originating in the Paleozoic. Equally, the phylogenetic diversity within individual families in almost all cases dates back to the Jurassic or even the Triassic. Our time-calibrated tree places the diversification of major extant soil lineages to the earliest existence of terrestrial ecosystems, and so also supports the role of soil mites in early soil communities. Our conclusions do not strictly require that association with the soil habitat was the ancestral condition, but a simplified ancestral-state reconstruction of the habitat type performed

across the tree on major taxonomic groups of Acariformes and Parasitiformes, also supports the antiquity of the association of mite lineages with soil-related habitats (see Table S6 and Fig. S17 for details). The preponderance of soil-inhabiting taxa representing early-branching lineages gives weight to the hypothesis that mites, possibly together with springtails, were among the few macroscopic animal groups inhabiting the soil matrix during the Ordovician-Silurian, and underwent the process of land colonization in synchrony with early terrestrial plants (Lenton et al. 2016). Other mesofauna of modern soil ecosystems such as the nematodes colonized the land much later (Rota-Stabelli et al. 2013), while some early land lineages such as myriapods and other arachnids were mostly living on the soil surface. Thus, the early diversification of major soil mite lineages with various functional roles, including particulate-feeding scavengers, fungivores and microbiovores, and eventually predators, is in agreement with a pioneering role of mites in the evolution of early terrestrial ecosystems.

Mitochondrial metagenomics and further advance in the Tree-of-Life for the mites

The 106 newly assembled mitochondrial genomes markedly extend the available phylogenetic coverage and now include representatives of seven of the nine supercohorts of Acariformes and three of the four orders within Parasitiformes. It is common practice in phylogenetic analysis of basal relationships to target specific exemplars of the focal group, which typically involves broad geographic and ecological sampling, or the use of museum collections, to achieve a comprehensive representation of major evolutionary lineages. This taxonomy-based approach may be fraught logistically and may lead to taxonomic bias, in particular in small-bodied and poorly characterized groups, while ignoring potentially valuable non-target taxa encountered during field collecting. Here instead we maximized the taxon sampling by gathering the full diversity of lineages from a small study area and applying a flotation method (Arribas et al. 2016) for extraction of a maximum number of species from the soil matrix. This ‘site-based’ approach makes full use of all specimens obtained, in this case from about 100 kg of soil in a semi-natural landscape (12 km maximum distance) of forest and grassland, which also produced many thousand further specimens that are being sequenced with PCR-based bulk approaches (community metabarcoding). If applied more widely, the site-based taxon selection in phylogenetic studies may overcome the current undersampling of soil-dwelling lineages of Acariformes and the

bias towards free-living and parasitic forms in public sequence databases. If applied to further sites and biogeographic regions, MMG can be a synergistic and cost-effective way for assembling the Tree-of-Life in hyperdiverse but elusive taxa (Crampton-Platt et al. 2016; Giribet 2016).

Site-based sampling will inevitably miss major lineages distributed elsewhere or not sampled by chance. An obvious group missing from our analysis are the members of Endeostigmata, which are common in soils everywhere but extremely small in body size, and their absence here is most likely explained by failure at the sequencing stage. Very small individuals frequently do not produce sufficient DNA for the construction of shotgun sequencing libraries, and this problem is alleviated here by pooling of DNA extracts from multiple specimens. The approach is most efficient if all species in the mixture are present in roughly equimolar proportions, which is possible if DNA extractions are performed on individuals, as was the case here. However, very small specimens (producing <1 ng of total genomic DNA) may still be underrepresented in the metagenomic libraries.

Bulk sequencing of multiple soil mite lineages from landscape scale sampling improved the topology and support of the tree of Acari and greatly extends the available taxon and gene sampling. The study revealed the great diversity of ancient lineages at the local scale and illustrates the potential of metagenomics to overcome limitations for phylogenetic and evolutionary analysis of small-bodied metazoans. The generally well-supported phylogenomic tree demonstrates that mitochondrial genomes are suitable for inferring the evolutionary history of this ancient group. Further phylogenetic efforts can now build on the current MMG dataset and methodology by including key lineages not represented, in particular: the presumed early-branching Endeostigmata; the Heterostigmata, the largest radiation of Trombidiformes with various lineages common in the soil; representatives of groups whose monophyly is questionable, such as Eupodina; or undersampled groups, e.g. various deep lineages of Sarcoptiformes. This first application of MMG to analyze site-based soil phylodiversity provides a framework that if applied to additional sites and biogeographic regions, could be a cost-effective way towards a more complete Tree of the mites or other hyperdiverse but elusive taxa. This source of phylogenetic information also is crucial to be incorporated in community level studies (e.g. the phylogenetic placement of metabarcodes) if we are

to understand patterns and processes driving soil biodiversity from a phylogenetic and functional perspective.

Materials & Methods

Specimen sampling, DNA extraction and sequencing

Specimens (see Supplementary Material, Table S1 for details on collecting locations) for each sample were obtained from the superficial (1 m² to 5 cm depth) and deep soil (40 cm diameter core to 35 cm depth) layers. Specimens were extracted from the soil sample following the flotation-Berlese-flotation protocol (Arribas et al. 2016) that allows for the extraction of 'clean' minute arthropods from large volumes of soil. Selected specimens were subjected to non-destructive individual DNA extractions and morphologically identified at least to family level (Table S1). DNA extractions were performed with the BioSprint 96 DNA Kit and a BioSprint 96 Workstation (Qiagen). To generate 'bait' sequences for associating specimens with the assembled mitochondrial genomes, the 5' end of the *cox1* gene (barcode fragment) was PCR amplified (see Arribas et al. 2016) and Table S7 for primer sequences and PCR conditions) and Sanger sequenced with ABI technology.

The mitochondrial metagenomics approach (MMG) of Crampton-Platt et al. (2015) was followed for the assembly of mitochondrial genomes from shotgun sequences of equimolar DNA mixtures of specimens. Genome assembly from shotgun reads favors the contig formation from the high-copy mitochondrial fraction and produces complete or partial mitochondrial genome assemblies for individual species in the sample. The approach contains different steps for the assembly, identification, verification and curation of the mitogenomes (see below) that allow to validate the reliability and homology of the data (Gillett et al. 2014; Andújar et al. 2015; Gómez-Rodríguez et al. 2015). For metagenomic sequencing, the concentration of each specimen DNA extract was measured using the Qubit dsDNA HS Assay Kit (Invitrogen). Aliquots of ≈ 1.5 ng of DNA (for many of these tiny specimens this was almost all extracted DNA) were pooled to prepare a single TruSeq Nano library for shotgun sequencing on a MiSeq sequencer (2 x 300 bp paired-end reads) using 120% of a flow cell (Illumina).

Assembly, annotation and alignment of Acari mitochondrial genomes

Raw reads were processed following a bioinformatic pipeline for assembly and further validation of obtained mitogenomes (Andújar et al. 2015; Crampton-Platt et al. 2015). Raw read quality was assessed with FastQC v0.10.1, low quality reads were discarded with Prinseq v0.19.2 (Schmieder and Edwards 2011) and Illumina adapter remnants were trimmed with Trimmomatic v0.30 (Bolger et al. 2014). Remaining reads were filtered for putative Arthropoda mitochondrial reads against a reference database of all complete Arthropoda mitochondrial genomes from NCBI (November 2016) using BLAST (Altschul et al. 1990) ($e\text{-value} = 10^{-5}$). Putative mitochondrial reads were assembled using three independent assemblers: Celera Assembler v7.0 (Myers 2000), IDBA-UD v1.1.1 (Peng et al. 2012) and Newbler (J.R. Miller et al. 2010). Independently assembled contigs were further validated by comparing among assemblers and only identical or highly similar contigs produced by at least two of the three different assemblers were merged with the 'De Novo Assembly' function in Geneious v7.1.9 (minimum overlap = 1000 bp; maximum mismatches per read = 1%; maximum gaps per read = 1%). The resulting mitochondrial genomes were identified by BLAST alignment to the Sanger 'bait' *cox1* sequences and annotated using MITOS (Bernt et al. 2013) followed by manual refinement. Protein-coding and ribosomal gene sequences were extracted gene by gene from these mitochondrial genomes and those downloaded from the NCBI (November 2016) (see Table S1 for accession numbers). Each gene extraction was independently aligned with MAFFT v6.240 (Katoh and Toh 2008), and the reading frame was verified across all protein coding gene alignments. The concatenation of the alignments for the 15 mitochondrial genes constituted the matrix subsequently used for phylogenetic analyses.

Mitochondrial phylogenomics of Acari

Three different datasets were generated from the 15-gene matrix: (i) the *DNAall* dataset including the DNA sequences for all ribosomal and protein coding genes, (ii) the *DNArecoded* dataset including the DNA sequences for complete ribosomal genes and modified protein coding genes: third codon positions excluded and first position recoded as R (for purines A or G) and Y (for pyrimidines T or C), and (iii) the *AA* dataset including amino acid sequences for the protein coding genes, with ribosomal genes excluded. These three datasets were used for Bayesian inference with PhyloBayes (Lartillot and Philippe 2004) and Maximum likelihood (ML) in RAxML v8.2

(Stamatakis 2014) and IQ-TREE (Nguyen et al. 2015). For PhyloBayes analyses two independent chains were run for a minimum of 25,000 generations under a GTR-CAT model for each dataset, and a consensus tree was obtained combining trees from both chains after discarding the first 10,000 generations as a burn-in fraction and retaining 1 tree at every 10 generations (*bpcomp* options *-x 10000 10 -c 0*).

RAxML analyses for each dataset were computed on the CIPRES portal (M.A. Miller et al. 2010) using 100 alternative runs to select the best tree and generating 1000 bootstrap replicates for node support. Analyses were performed using the best partition scheme and corresponding best models of substitution for each dataset (considering only those supported by RAxML for nucleotide (i.e. GTR) and amino acid sequences (GTR with different transition matrixes)) as determined with Partition Finder (Lanfear et al. 2017), using a partition scheme by gene and by codon position as input for the DNA datasets and a partition by gene for the AA dataset (see Table S2 for details). IQ-TREE analyses were performed on the three datasets using the best partition scheme and models of substitution selected by Partition Finder, as before, but using all available models (Table S2). Node support was calculated with 250 standard bootstrap replicates.

Dating Acari origin and diversification

Dating was performed using BEAST v.1.8.4 (Drummond et al. 2012). Analyses were done using the *DNAall* dataset and applying the best-fitting substitution model and partition scheme as estimated in Partition Finder (Lanfear et al. 2017) for all available models (Table S2) and using the dataset partitioned by gene and codon as input. For the molecular clock settings, the dataset was partitioned in three partitions (protein coding genes, *rrnL* and *rrnS*) applying an uncorrelated lognormal (ULN) clock to each partition. Due to the high complexity and time required for calibration analyses on our dataset, together with the fact that preliminary calibration analyses showed long branch attraction problems (that were not confounding the tree searches performed before with other methods), we opted for fixing the topology in the calibration analyses to the optimal topology obtained for the *AA* dataset with PhyloBayes, a software particularly robust under high rate heterogeneity scenarios. Note that this is a well-established and common practice in phylogenetic inference using complex datasets (e.g. Brown and Yang 2010; Pons et al. 2010) for which the co-estimation approach usually results in

poor tree topologies. A birth-death speciation prior was applied and analyses were run twice in parallel with 100 million generations sampling one tree every 10,000 generations. Consensus trees were estimated with TREEANNOTATOR (Drummond et al. 2012) combining both runs and discarding the 50% initial trees as burn-in after checking the ESS of the tree likelihood and ensuring that values had reached a plateau in TRACER v. 1.6 (<http://beast.bio.ed.ac.uk/Tracer>). Posterior probabilities were considered as a measure of node support. Calibration analyses were computed on CIPRES (M.A. Miller et al. 2010).

The oldest fossil age in each taxon was used as hard minimum bound (using uniform distributions with uninformative maximum ages) for the age of the stem node of the corresponding lineage (calibration point in the phylogeny). Note that the use of uniform distributions with uninformative maximum ages avoids giving an implicit maximum age constraint not supported by fossil data at every calibration point, which happens when log-normal or other functions are used to constrain node ages (Ho and Phillips 2009; Donoghue and Yang 2016). Among all the available oldest fossils from the comprehensive compilation of Dunlop and Selden (2009) we performed the analyses for three nested sets of calibration points including an increased number of fossils but also resulting in increased uncertainty of the phylogenetic placement of these fossils in our trees (see Table S4). A first analysis was performed using the most conservative set of calibration points for our trees (*14min* set) including constraints for Acariformes, Sarcoptiformes and those families (with more than one representative) found to be monophyletic and well-supported, i.e. Damaeidae, Erythraeidae, Galumnidae, Gymnodamaeidae, Hermanniellidae, Ixodidae, Oppiidae, Parasitidae, Phenopelopidae, Phthiracaridae, Scheloribatidae and Tetranychidae (see Table S4 for details). A second analysis used the 14 previous plus six additional calibration points (*20min* set) on families with a single representative in our trees (i.e. Cepheidae, Camisiidae, Bdellidae, Anystidae, Ceratozetidae and Scutoverticidae). Finally, a third analysis was performed including all the previous calibration points and two additional constraints (*22min* set) to integrate some of the oldest known fossils from families that are not represented in our trees (and that previously were not used to constrain the minimum age at the stem node of the complete Sarcoptiformes (i.e. 392 Mya)). These are Hypochthoniidae (326 Mya), Cosmochthoniidae (326 Mya) and Protoplophoridae (326 Mya) constraining the

minimum age for stem node of Enarthronotides, and Hydrozetidae (197 Mya), Trhypochthoniidae (145 Mya) and Cymbaereidae (145 Mya) for which the calibration point in our phylogeny coincided with the node of the split between Nothrina and Brachypylina (stem node of both Nothrina and Brachypylina respectively) that was constrained to the age of the oldest available fossil for Hydrozetidae (197 Mya).

The use of uniform distributions with uninformative maximum age constraints avoids including an important source of uncertainty that could strongly impact the results (Ho and Phillips 2009; Donoghue and Yang 2016). This way of calibration is however very sensitive to take arbitrarily high ages while the estimated substitution rate is driven toward very low values, due to the absence of any upper limit to divergence times (Hug and Roger 2007; Ho and Phillips 2009). For this reason, we also conducted the time calibration analysis adding two maximum age constraints applied to the base for the Acariformes and Parasitiformes lineages. While to establish maximum age constraints on clade ages inevitably involves some arbitrariness, the statistical analysis of the fossil series can provide very useful information to define hypothetical maximum ages to be further used in calibration analyses (Holland 2016). Therefore, we performed calibration analyses for each set of minimum hard bounds also constraining the maximum ages of Acariformes and Parasitiformes crown nodes (*2max* set, Table S4) defined by the 99% confidence interval age for the first fossil appearance, as estimated using the oldest gap approach (Solow 2003) on the available fossil series at Fossilworks database (Behrensmeyer and Turner 2013) (July 2017) and Dunlop and Selden (2009) (see Table S5 for used fossils) and the R function CladeAge (Claramunt and Cracraft 2015).

Ancestral reconstruction of mitochondrial gene rearrangements

To infer the evolutionary pathways of mitochondrial gene rearrangements, each of the 204 Acari and outgroups mitogenomes were coded as different *mitotypes* according to the gene order (excluding tRNA genes). The much higher observed number of translocations of tRNA genes was ignored as they may mask the phylogenetic signal of rare changes in protein coding and rRNA genes (Xue et al. 2016). Using this information associated to each terminal we reconstructed the ancestral *mitotypes* in BEAST 1.8.4, which simultaneously estimated branch lengths and assigned trait states

to all its nodes (Drummond et al. 2012). Ancestral *mitotype* reconstruction was performed using an asymmetric trait evolution model (unordered states) with other parameters and inputs as in the calibration analyses above.

We also used the program TreeREx 1.85 (Bernt et al. 2008) for parsimony inference of gene rearrangement events (including reversals, transpositions, reverse transpositions, and tandem duplication with random loss of genes within the mitochondrial genome) on an existing tree. A matrix was generated containing the specific gene order for each mitogenome (tRNAs were again omitted). In this case only the 167 taxa with complete mitogenomes were used because the software requires that analyzed mitogenomes include an identical set of genes. Using the phylogenetic tree obtained in PhyloBayes for the *AA* dataset, the most parsimonious transformational pathway was computed using common intervals mapped along the branches with TreeREx, which also inferred the putative gene order at the internal nodes and its reliability. The TreeREx analysis was performed with default settings, as suggested at the website (<http://pacosy.informatik.uni-leipzig.de/185-0-TreeREx.html>).

Data Availability

Datasets, matrices and phylogenetic trees available from Dryad Digital Repository: [<https://doi.org/10.5061/dryad.ghx3ffbjb>]. GenBank Accessions numbers are given in Table S1.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgements

This research was funded by the NHM Biodiversity Initiative and the CGL2015-66192-R and CGL2015-74178-JIN projects. PA was supported by the Royal Society UK (Newton International Program) and the Spanish Ministry of Economy and Competitiveness (Juan de la Cierva Formación); BL by the NHM Biodiversity Initiative and French Labex Agro-CeMEB-NUMEV. Thanks are due to the regional government of Andalucía and the staff of the Sierra de Grazalema Natural Park; to P. Foster (NHM) and J. Arribas for their technical assistance and to four anonymous reviewers and the editor for their comments on the manuscript.

Competing interests

APV is cofounder of a UK company, NatureMetrics, which provides metabarcoding and metagenomics services to the public and private sector.

References

- Altschul S, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Andújar C, Arribas P, Ruzicka F, Platt AC, Timmermans MJTNMJTNJTJN, Vogler APAAP, Fernández CA, Arribas P, Ruzicka F, Crampton-Platt A, et al. 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Mol. Ecol.* 24:3603–3617.
- Andújar C, Serrano J, Gómez-Zurita J. 2012. Winding up the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. *BMC Evol. Biol.* 12:40.
- Arribas P, Andújar C, Hopkins K, Shepherd M, Vogler AP, Andújar C, Hopkins K, Shepherd M, Vogler APAP. 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods Ecol. Evol.* 7:1071–1081.
- Bardgett RD, van der Putten WH. 2014. Belowground biodiversity and ecosystem functioning. *Nature* 515:505–511.
- Bardgett RD, Usher MB, Hopkins DWW. 2005. Biological diversity and function in soils. Cambridge: Cambridge University Press
- Basso A, Babbucci M, Pauletto M, Riginella E, Patarnello T, Negrisolo E. 2017. The highly rearranged mitochondrial genomes of the crabs *Maja crispata* and *Maja squinado* (Majidae) and gene order evolution in Brachyura. *Sci. Rep.* 7:4096.
- Behrensmeyer AK, Turner A. 2013. Taxonomic occurrences of Donacinae recorded in the Paleobiology Database. Fossilworks. Available from: <http://fossilworks.org>
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69:313–319.
- Bernt M, Merkle D, Middendorf M. 2008. An algorithm for inferring mitogenome rearrangements in a phylogenetic tree. In: RECOMB-CG '08 Proceedings of the international workshop on Comparative Genomics. p. 143–157.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bolton SJ, Chetverikov PE, Klompen H. 2017. Morphological support for a clade comprising two vermiform mite lineages: Eriophyoidea (Acariformes) and

- Nematalycidae (Acariformes). *Syst. Appl. Acarol.* 22:1096–1131.
- Brown RP, Yang Z. 2010. Bayesian dating of shallow phylogenies with a relaxed clock. *Syst. Biol.* 59:119–131.
- Burges A, Raw F. 1967. *Soil Biology*. London: Academic Press
- Claramunt S, Cracraft J. 2015. A new time tree reveals Earth history 's imprint on the evolution of modern birds. *Sci. Adv.* 1:e1501005.
- Crampton-Platt A, Timmermans MJTN, Gimmel ML, Kutty SN, Cockerill TD, Vun Khen C, Vogler AP. 2015. Soup to Tree: The Phylogeny of Beetles Inferred by Mitochondrial Metagenomics of a Bornean Rainforest Sample. *Mol. Biol. Evol.* 32:2302–2316.
- Crampton-Platt A, Yu DW, Zhou X, Vogler AP. 2016. Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience* 5:15.
- Dabert M, Witalinski W, Kazmierski A, Olszanowski Z, Dabert J. 2010. Molecular phylogeny of acariform mites (Acari, Arachnida): strong conflict between phylogenetic signal and long-branch attraction artifacts. *Mol. Phylogenet. Evol.* 56:222–241.
- Donoghue PCJ, Yang Z. 2016. The evolution of methods for establishing evolutionary timescales. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20160020.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.
- Dubinin VB. 1962. Class Acaromorpha: mites or gnathosomic chelicerate arthropods. In: Rodendorf B, editor. *Fundamentals of palaeontology*. Moscow: Academy of Sciences of the USSR. p. 447–473.
- Dunlop JA, Alberti G. 2008. The affinities of mites and ticks: A review. *J. Zool. Syst. Evol. Res.* 46:1–18.
- Dunlop JA, Selden PA. 2009. Calibrating the chelicerate clock: A paleontological reply to Jeyaprakash and Hoy. *Exp. Appl. Acarol.* 48:183–197.
- Gillett CPDT, Crampton-Platt A, Timmermans MJTN, Jordal BH, Emerson BC, Vogler AP. 2014. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol. Biol. Evol.* 31:2223–2237.
- Giribet G. 2016. Genomics and the animal tree of life: conflicts and future prospects. *Zool. Scr.* 45:14–21.

- Gómez-Rodríguez C, Crampton-Platt A, Timmermans MJTN, Baselga A, Vogler AP. 2015. Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods Ecol. Evol.* 6:883–894.
- Hirst S. 1923. On some arachnid remains from the Old Red Sandstone (Rhynie Chert Bed, Aberdeenshire). *Ann. Mag. Nat. Hist.* 12:455–474.
- Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58:367–380.
- Holland SM. 2016. The non-uniformity of fossil preservation. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150130.
- Hug LA, Roger AJ. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. *Mol. Biol. Evol.* 24:1889–1897.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9:286–298.
- Klimov PB, OConnor BM, Chetverikov PE, Bolton SJ, Pepato AR, Mortazavi AL, Tolstikov A V., Bauchan GR, Ochoa R. 2018. Comprehensive phylogeny of acariform mites (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a long branch. *Mol. Phylogenet. Evol.* 119:105–117.
- Klompen H, Lekveishvili M, Black IV WC, Black WC. 2007. Phylogeny of parasitiform mites (Acari) based on rRNA. *Mol. Phylogenet. Evol.* 43:936–951.
- Krantz GW, Walter DE. 2009. *A Manual of Acarology*. Texas Tech Univ, Press
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34:772–773.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21:1095–1109.
- Lenton TM, Dahl TW, Daines SJ, Mills BJW, Ozaki K, Saltzman MR, Porada P. 2016. Earliest land plants created modern levels of atmospheric oxygen. *Proc. Natl. Acad. Sci.* 113:9704–9709.
- Linard B, Crampton-Platt A, Timmermans MJTN, Vogler AP. 2015. Metagenome skimming of insect specimen pools: potential for comparative genomics. *Genome Biol. Evol.* 7:1474–1489.
- Lindquist EE. 1996. Phylogenetic relationships. In: Lindquist EE, Sabelis MW, Bruin J,

- editors. Eriophyoid Mites. Their Biology, Natural Enemies and Control. World Crop Pests. Amsterdam: Elsevier. p. 301–327.
- Lozano-Fernandez J, Carton R, Tanner AR, Puttick MN, Blaxter M, Vinther J, Olesen J, Giribet G, Edgecombe GD, Pisani D. 2016. A molecular palaeobiological exploration of arthropod terrestrialization. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150133.
- Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95:315–327.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*. New Orleans. p. 1–8.
- Myers EW. 2000. A whole-genome assembly of *Drosophila*. *Science* (80-.). 287:2196–2204.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.
- Norton RA. 1998. Morphological evidence for the evolutionary origin of Astigmata (Acari: Acariformes). *J. Exp. Acarol.* 22:559–594.
- Norton RA, Bonamo PM, Grierson JD, Shear WA. 1988. Oribatid mite fossils from a terrestrial Devonian deposit near Gilboa. New York. *J. Paleontol.* 62:259–269.
- Norton RA, Kethley JB, Johnston DE, OConnor BM. 1993. Phylogenetic perspectives on genetic systems and reproductive modes of mites. In: Wrensch DL, Ebbert MA, editors. *Evolution and diversity of sex ratio in insects and mites*. New York: Chapman and Hall. p. 8–99.
- OConnor BM. 1984. Phylogenetic relationships among higher taxa in the Acariformes, with particular reference to the Astigmata. In: Griffiths DA, Bowman CE, editors. *Acarology VI*. Chichester: Ellis-Horwood Ltd. p. 19–27.
- Papadopoulou A, Anastasiou I, Vogler AP. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* 27:1659–1672.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428.

- Pepato AR, Klimov PB. 2015. Origin and higher-level diversification of acariform mites - evidence from nuclear ribosomal genes, extensive taxon sampling, and secondary structure alignment. *BMC Evol. Biol.* 15:178.
- Pepato AR, da Rocha CEF, Dunlop JA. 2010. Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evol. Biol.* 10:235.
- Pons J, Ribera I, Bertranpetit J, Balke M. 2010. Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Mol. Phylogenet. Evol.* 56:796–807.
- Rota-Stabelli O, Daley AC, Pisani D. 2013. Molecular timetrees reveal a cambrian colonization of land and a new scenario for ecdysozoan evolution. *Curr. Biol.* 23:392–398.
- Schaefer I, Norton RA, Scheu S, Maraun M. 2010. Arthropod colonization of land - Linking molecules and fossils in oribatid mites (Acari, Oribatida). *Mol. Phylogenet. Evol.* 57:113–121.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864.
- Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G. 2014. Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* 31:2963–2984.
- Solow AR. 2003. Estimation of stratigraphic ranges when fossil finds are not randomly distributed. *Paleobiology* 29:181–185.
- Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Subías LS, Arillo A. 2002. Oribatid mite fossils from the Upper Devonian of South Mountain, New York and the Lower Carboniferous of County Antrim Northern Ireland (Acariformes, Oribatida). *Est. Mus. Cienc. Nat. Alava* 17:93–106.
- Timmermans MJTN, Barton C, Haran J, Ahrens D, Culverwell CL, Ollikainen A, Dodsworth S, Foster PG, Bocak L, Vogler AP. 2016. Family-level sampling of mitochondrial genomes in Coleoptera: Compositional heterogeneity and phylogenetics. *Genome Biol. Evol.* 8:161–175.
- Veresoglou SD, Halley JM, Rillig MC. 2015. Extinction risk of soil biota. *Nat. Commun.* 6:8862.

- Walter DE, Proctor HC. 1999. Feeding behaviour and phylogeny: observations on early derivative Acari. *Exp. Appl. Acarol.* 22:39–50.
- Wang Y, Liu X, Garzón-Orduña IJ, Winterton SL, Yan Y, Aspöck U, Aspöck H, Yang D. 2017. Mitochondrial phylogenomics illuminates the evolutionary history of Neuropterida. *Cladistics* 33:617–636.
- Xue XF, Dong Y, Deng W, Hong XY, Shao R. 2017. The phylogenetic position of eriophyoid mites (superfamily Eriophyoidea) in Acariformes inferred from the sequences of mitochondrial genomes and nuclear small subunit (18S) rRNA gene. *Mol. Phylogenet. Evol.* 109:271–282.
- Xue XF, Guo JF, Dong Y, Hong XY, Shao R. 2016. Mitochondrial genome evolution and tRNA truncation in Acariformes mites: new evidence from eriophyoid mites. *Sci. Rep.* 6:18920.

Figure captions

Figure 1. Time-calibrated tree and evolution of mitochondrial gene rearrangements across the phylogeny of Acari using the *14min+2max* set of calibration points. Colored circles: posterior probabilities of the PhyloBayes analysis using the *AA* dataset (black for >0.95). Grey bars: 95% confidence intervals for node ages (Mya). Colored triangles: ancestral *mitotypes* by BEAST (upper) and TreeREx (lower). Circles above branches: rearrangement events inferred by TreeREx (T: transposition, I: inversion, iT: inverse transposition and star: multiple steps rearrangement events). Boxes/vertical bars: solid line for monophyletic supercohort/cohort/subcohort and discontinuous for non-monophyletic ones. *Eriophyoidea was recovered as sister group to the Trombidiformes plus Sarcoptiformes lineage thus breaking the monophyly of Trombidiformes in the traditional sense. Vertical discontinuous lines: oldest terrestrial fossils for some of the earliest land colonizers. Color of terminals: red for newly generated soil Acari mitogenomes, black and blue for Acari and outgroups from GenBank respectively. See Table S1 for details on the terminals.

