Research Article - Discoveries

Molecular evolutionary trends and biosynthesis pathways in the Oribatida revealed by the genome of *Archegozetes longisetosus*

Adrian Brückner[1]*, Austen A. Barnett[2], Igor A. Antoshechkin[1] and Sheila A. Kitchen[1]

[1]Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, United States of America

[2]Department of Biology, DeSales University, 2755 Station Avenue, Center Valley, PA 18034, United States of America

* corresponding author: Adrian Brückner - adrian.brueckner@gmail.com

Abstract

Oribatid mites are a specious order of microarthropods within the subphylum Chelicerata, compromising about 11,000 described species. They are ubiquitously distributed across different microhabitats in all terrestrial ecosystems around the world and were among the first animals colonizing terrestrial habitats as decomposers and scavengers. Despite their species richness and ecological importance genomic resources are lacking for oribatids. Here, we present a 190-Mb genome assembly of the clonal, all-female oribatid mite species *Archegozetes longisetosus* Aoki, a model species used by numerous laboratories for the past 30 years. Comparative genomic and transcriptional analyses revealed patterns of reduced body segmentation and loss of segmental identity gene *abd-A* within Acariformes, and unexpected expression of key eye development genes in these eyeless mites across developmental stages. Consistent with their soil dwelling lifestyle, investigation of the sensory genes revealed a species-specific expansion of gustatory receptors, the largest chemoreceptor family in the genome used in olfaction, and evidence of horizontally transferred enzymes used in cell wall degradation of plant and fungal matter, both components of the *A. longisetosus* diet. Oribatid mites are also noted for their biosynthesis capacities and biochemical diversity. Using biochemical and genomic data, we were able to delineate the backbone biosynthesis of monoterpenes, an important class of compounds found in the major exocrine gland system of Oribatida – the oil glands. Given the mite's strength as an

experimental model, the new high-quality resources provided here will serve as the foundation for molecular research in Oribatida and will enable a broader understanding of chelicerate evolution.

Keywords

soil animal, terpene synthesis, horizontal gene transfer, parthenogenesis, chemoreceptors, Hox genes, model organism, RNAseq, MinION long-read sequencing, Sarcoptiformes

Introduction

In the past couple of years, the number of sequenced animal genomes has increased dramatically, especially for arthropods about 500 genomes sequences are now available (Childers 2020; Thomas et al. 2020). The majority of these genomes, however, belong to the insects (e.g. flies, beetles, wasp, butterflies and bugs (Thomas et al. 2020)) which compromise the most diverse, yet evolutionarily young and more derived taxa of arthropods (Giribet and Edgecombe 2019; Regier et al. 2010). In strong contrast, genome assemblies, many of which are incomplete or not well annotated, exist for the Chelicerata (Childers 2020) – the other major subphylum of arthropods (Giribet and Edgecombe 2019; Regier et al. 2010). Chelicerates include sea spiders, spiders, mites and scorpions among other organisms, as well as several extinct taxa (Ballesteros and Sharma 2019; Dunlop and Selden 1998). Chelicerates originated as marine animals about 500 million years ago (Dunlop and Selden 1998; Dunlop 2010). Molecular analyses suggest that one particular group, the omnivorous and detritivores acariform mites, may have been among the first arthropods that colonized terrestrial habitats and gave rise to ancient, simple terrestrial food webs (Dunlop and Alberti 2008; Schaefer et al. 2010; Walter and Proctor 1999).

So far, the well-annotated genomic data of chelicerates is limited to animal parasites (including human pathogens and ticks), plant parasites, and predatory mites used in pest control (Cornman et al. 2010; Dong et al. 2017; Dong et al. 2018; Grbić et al. 2011; Gulia-Nuss et al. 2016; Hoy et al. 2016; Rider et al. 2015). Other than some lower-quality genome assemblies

(Bast et al. 2016), there are no resources available for free-living soil and litter inhabiting species. Such data are, however, pivotal to understanding the evolution of parasitic lifestyles from a free-living condition and to bridge the gap between early aquatic chelicerates such as horseshoe crabs, and highly derived terrestrial pest species and parasites (Klimov and OConnor 2013; Shingate et al. 2020; Weinstein and Kuris 2016). Because the phylogeny of Chelicerata remains unresolved, additional chelicerate genomes are urgently needed for comparative analyses (Ballesteros and Sharma 2019; Dunlop 2010; Lozano-Fernandez et al. 2019). To help address this deficit, we report here the genome assembly of the soil dwelling oribatid mite *Archegozetes longisetosus* (Aoki, 1965; **Figure 1**) (Aoki 1965) and a comprehensive analysis in the context of developmental genes, feeding biology, horizontal gene transfer and biochemical pathway evolution of chelicerates.

Archegozetes longisetosus (hereafter referred to as *Archegozetes*) is a member of the Oribatida (Acariformes, Sarcoptiformes), an order of chelicerates well-known for their exceptional biosynthesis capacities, biochemical diversity, unusual mode of reproduction, unusually high pulling strength, mechanical resistances and pivotal ecological importance (Brückner et al. 2020; Brückner et al. 2017b; Heethoff and Koerner 2007; Heethoff et al. 2009; Maraun et al. 2007; Maraun and Scheu 2000; Norton and Palmer 1991; Raspotnig 2009; Schmelzle and Blüthgen 2019). *Archegozetes*, like all members of its family Trhypochthoniidae (**Figure 1a**), reproduce via thelytoky (Heethoff et al. 2013). That means the all-female lineages procreate *via* automictic parthenogenesis with an inverted meiosis of the holokinetic chromosomes, resulting in clonal offspring (Bergmann et al. 2018; Heethoff et al. 2006; Palmer and Norton 1992; Wrensch et al. 1994). While studying a parthenogenetic species is useful for the development of genetic tools as stable germ-line modifications can be obtained from the clonal progeny without laboratory crosses, one is confronted with the technical and philosophical problems of species delineation, cryptic diversity and uncertain species distribution (Heethoff et al. 2013; Oxley et al. 2014). Reviewing all available data, Norton (Norton 1994; 2007) and Heethoff et al. (Heethoff et al. 2013) concluded that *Archegozetes* is found widely on continents and islands throughout the tropical and partly subtropical regions of the world and that it is a middle-derived oribatid mite closely related to the suborder Astigmata.

One major feature of most oribatid mites is a pair of opisthonotal oil-glands and *Archegozetes* is no exception (Raspotnig 2009; Sakata and Norton 2001). These are a pair of large exocrine glands, each composed of a single-cell layer invagination of the cuticle, which is the simplest possible paradigm of an animal gland (Brückner and Parker 2020; Heethoff 2012). The biological role of these glands was rather speculative for a long time; ideas ranged from a lubricating and osmo- or thermoregulative function (Riha 1951; Smrž 1992; Zachvatkin 1941) to roles in chemical communication (Heethoff et al. 2011a; Raspotnig 2006; Shimano et al. 2002). So far about 150 different gland components have been identified from oribatid mites, including mono- and sesquiterpenes, aldehydes, esters, aromatics, short-chained hydrocarbons, hydrogen cyanide (HCN) and alkaloids (Brückner et al. 2017b; Brückner et al. 2015; Heethoff et al. 2018; Raspotnig 2009; Saporito et al. 2007). While some chemicals appear to be alarm pheromones (Raspotnig 2006; Shimano et al. 2002), most function as defensive allomones (Heethoff et al. 2011a). Interestingly, alkaloids produced by oribatids mites are the ultimate source of most toxins sequestered by poison-frogs (Saporito et al. 2007; Saporito et al. 2009).

Terrestrial chelicerates predominately ingest fluid food. While phloem-feeding plant pests like spider mites and ecotoparasites likes ticks adapted a sucking feeding mode, scorpions, spiders and others use external, pre-oral digestion before ingestion by morphologically diverse mouthparts (Bensoussan et al. 2016; Cohen 1995; Dunlop and Alberti 2008; Gulia-Nuss et al. 2016). Exceptions from this are Opiliones and sarcoptiform mites, i.e. oribatid and astigmatid mites, all of which ingest solid food (Heethoff and Norton 2009; Norton 2007; Shultz 2007). In general, oribatids feed on a wide range of different resources and show a low degree of dietary specialization (Brückner et al. 2018b). The typical food spectrum of Oribatida, includes leaf-litter, algae, fungi, lichens, nematodes, and small dead arthropods such as collembolans (Heidemann et al. 2011; Riha 1951; Schneider and Maraun 2005; Schneider et al. 2004a; Schneider et al. 2004b). In laboratory feeding trials, oribatid mites tend to prefer dark pigmented fungi, but also fatty acid-rich plant-based food (Brückner et al. 2018b; Schneider and Maraun 2005). Additionally, stable-isotope analyses of ^{15}N and ^{13}C suggested that Oribatida are primary- and secondary decomposers feeding on dead plant material and fungi, respectively (Maraun et al. 2011; Schneider et al. 2004a). The reasons for these preferences are still unknown, but they raise the question of how oribatid mites are able to enzymatically digest the cell walls of plants and

fungi (Brückner et al. 2018a; Brückner et al. 2018b; Schneider et al. 2004b; Smrž and Čatská 2010).

Early studies on *Archegozetes* and other mites found evidence for cellulase, chitinase and trehalase activity which was later attributed to symbiotic gut bacteria (Haq 1993; Luxton 1979; Siepel and de Ruiter-Dijkman 1993; Smrž 2000; Smrž and Čatská 2010; Smrž and Norton 2004; Zinkler 1971). While such bacterial symbionts are a possible explanation, genomic data of other soil organisms and plant-feeding arthropods suggest a high frequency of horizontal transfer of bacterial and fungal genes enabling the digestion of cell walls (Grbić et al. 2011; Mayer et al. 2011; McKenna et al. 2019; Wu et al. 2017; Wybouw et al. 2016; Wybouw et al. 2018). For instance, an in-depth analysis of the spider mite *Tetranychus urticae* revealed a massive incorporation of microbial genes into the mite's genome (Grbić et al. 2011; Wybouw et al. 2018). Horizontal gene transfer appears to be a common mechanism for soil organisms, including mites, to acquire novel metabolic enzymes (Dong et al. 2018; Faddeeva-Vakhrusheva et al. 2016; Grbić et al. 2011; Hoffmann et al. 1998; Mayer et al. 2011; Wu et al. 2017), and hence seems very likely for *Archegozetes* and other oribatid mite species that feed on plant or fungal matter.

Archegozetes has been established as a laboratory model organism for three decades, having been used in studies, ranging from ecology, morphology, development and eco-toxicology to physiology and biochemistry (Barnett and Thomas 2012; 2013a; 2013b; 2018; Brückner et al. 2017a; Brückner et al. 2020; Heethoff et al. 2013). As such, *Archegozetes* is among the few experimentally tractable soil organisms and by far the best-studied oribatid mite species (Barnett and Thomas 2012; Heethoff et al. 2013; Thomas 2002). Since it meets the most desirable requirements for model organisms (Thomas 2002), that is a rapid development under laboratory conditions, a dedicated laboratory strain was named *Archegozetes longisetosus* **ran** in reference to its founder **R**oy **A**. **N**orton (Heethoff et al. 2013, **Figure 1b-c**). Their large number of offspring enables mass cultures of hundreds of thousands of individuals, and their cuticular transparency during juvenile stages, and weak sclerotization as adults are general assets of an amenable model system (Brückner et al. 2018c; Brückner et al. 2016; Heethoff et al. 2013; Heethoff and Raspotnig 2012). In the past 10 years, *Archegozetes* also received attention as a model system for chemical ecology (Brückner and Heethoff 2018; Brückner et al. 2020;

Brückner et al. 2016; Heethoff and Rall 2015; Heethoff and Raspotnig 2012; Raspotnig et al. 2011; Thiel et al. 2018). Some of these studies focusing on the *Archegozetes* gland revealed basic insights into the chemical ecology and biochemical capabilities of arthropods (Brückner et al. 2020; Heethoff and Rall 2015; Thiel et al. 2018). Hence, *Archegozetes* is poised to become a genetically tractable model to study the molecular basis of gland and metabolic biology.

The aim and focus of the current study were three-fold – to provide well-annotated, high-quality genomic and transcriptomic resources for *Archegozetes longisetosus* (**Figure 1**), to reveal possible horizontal gene transfers that could further explain the feeding biology of oribatids, and to present *Archegozetes* as a research model for biochemical pathway evolution. Through a combination of comparative genomic and detailed computational analyses, we were able to generate a comprehensive genome of *Archegozetes* and provide it as an open resource for genomic, developmental and evolutionary research. We further identified candidate horizontal gene transfer events from bacteria and fungi that are mainly related to carbohydrate metabolism and cellulose digestion, features correlated with the mite feeding biology. We also used the genomic data together with stable-isotope labeling experiments and mass spectrometric investigation to delineate the biosynthesis pathway of monoterpenes in oribatid mites.

Results and Discussion

Archegozetes longisetosus genome assembly

Archegozetes longisetosus (**Figure 1**) has a diploid chromosome number (2n) of 18 (Heethoff et al. 2006), most likely comprising 9 autosomal pairs, the typical number of nearly all studied oribatid mite species (Norton et al. 1993). There are no distinct sex chromosomes in *Archegozetes*; this appears to be ancestral in the Acariformes and persisted in the Oribatida (Heethoff et al. 2006; Norton et al. 1993; Wrensch et al. 1994). Even though some XX:XO and XX:XY genetic systems have been described in the closely related Astigmata, the sex determination mechanism in oribatids, including *Archegozetes,* remains unknown (Heethoff et al. 2013; Heethoff et al. 2006; Norton et al. 1993; Oliver Jr 1983; Wrensch et al. 1994). To provide genetic resources, we sequenced and assembled the genome using both Illumina short-read and Nanopore MinION long-read sequencing approaches (**Table 1**; see also "**Materials and**

Methods"). Analyses of the *k-mer* frequency distribution of short reads (**Table 1; Supplementary Figure S1**) resulted in an estimated genome size range of 135-180 Mb, smaller than the final assembled size of 190 Mb (**Table 1**; see also "**Materials and Methods**"). This difference was suggestive of high repetitive content in the genome of *Archegozetes* and indeed, repeat content was predicted to be 32 % of the genome (**see below**) (Alfsnes et al. 2017; Simpson 2014). Compared to genome assemblies of other acariform mites, the assembled genome size of *Archegozetes* is on the large end, but is smaller than that of mesostigmatid mites, ticks and spiders (Bast et al. 2016; Dong et al. 2017; Dong et al. 2018; Grbić et al. 2011; Gulia-Nuss et al. 2016; Hoy et al. 2016; Schwager et al. 2017). In the context of arthropods in general, *Archegozetes*'s genome (**Table 1**) is among the smaller ones and shares this feature with other arthropod model species like the spider mite, *Drosophila*, clonal raider ant and red flour beetle (Consortium 2008; dos Santos et al. 2015; Grbić et al. 2011; Oxley et al. 2014). Even though we surface-washed the mites and only used specimens with empty alimentary tracts for sequencing, we removed 438 contigs with high bacterial or fugal homology making up approximately 8.5 Mb of contamination (**see supplementary Table S1**). The final filtered genome assembly was composed of 1182 contigs with an N_{50} contiguity of 994.5 kb (**Table 1**).

The official gene set and annotation of *Archegozetes*

We generated the official gene set (OGS) for *Archegozetes* by an automated, multi-stage process combining *ab inito* and evidenced-based (RNAseq reads, transcriptomic data and curated protein sequences) gene prediction approaches (see "**Materials and Methods**") yielding 23,825 gene models. In comparison to other mites and ticks as well as insects, this is well within the range of the numbers discovered in other Chelicerata so far (**Figure 2a**). Chelicerates with a large OSG, however, usually possess larger genomes (1-7 Gb), which suggests that *Archegozetes* may have a relatively dense distribution of protein-coding genes in its genome. On the other hand, ticks can have giga-base sized genomes, but only a rather small number of gene models, probably due to high repetitive content (Barrero et al. 2017; Gulia-Nuss et al. 2016; Palmer et al. 1994; Van Zee et al. 2007). Lacking more high-quality genomic resources of mites, it is thus not clear whether the OGS of *Archegozetes* is the rule, or rather the exception within the Oribatida.

To compare if *Archegozetes'* OSG is similar to predicted genes of other oribatid mites as well as Prostigmata and Astigmata, we first clustered genes by ortholog inference (OrthoFinder;

(Emms and Kelly 2015b)), removed species-specific genes and constructed a presence-absence matrix of orthogroups to ordinate the data using non-metric multidimension scaling (NMDS, **Figure 2b**). Ordination revealed that the OGS of *Archegozetes* is well nested with other oribatid mites and clearly separated from their closest relative the astigmatid mites as well as prostigmatid mites (**Figure 2b**). As a first step in annotating the OSG, we ran KOALA (KEGG Orthology And Links Annotation) to functionally characterize the genes (Kanehisa et al. 2016). In total, 10,456 (43.9%) of all genes received annotation and about two thirds of all genes were assigned either as metabolic genes (36%) or genes related to genetic information processing (34%), while the remaining genes fell into different KEGG categories (**Figure 2c**). To further annotate the genome, we followed the general workflow of funannotate with some modifications (Palmer and Stajich 2017, see "Materials and Methods").

Overall, we found 15,236 genes (64%) of the OGS with homology to previously published sequences (**Figure 2d**). For about half of all genes (51%), we were able to assign a full annotation, 4% of all genes only showed homology to bioinformatically predicted proteins of other species, while 9% of all genes only showed homology to hypothetical proteins (**Figure 2d**). As only a few high-quality, annotated mite genomes are available and the two-spotted spider mite is the sole species with any experimentally confirmed gene models, it is not surprising that we were only able to confidently annotate about 55% of all genes of the OGS (**Figure 2d**).

Orthology and comparative genomics of chelicerates

To further access the protein-coding genes of the mite, we compared the OGS to other chelicerates. Both concatenated maximum likelihood and coalescent species-tree phylogenomic approaches based on 1,121 orthologs placed *Archegozetes,* as expected, within the Nothrina (Heethoff et al. 2013; Pachl et al. 2012) with strong support and recovered previously found oribatid clade topologies (**Figure 3a**). Our analysis placed the Astigmata as a sister group of Oribatida and not nested within oribatids as suggested based on life-history, chemical defensive secretions, morphology and several molecular studies (Alberti and Michalik 2004; Dabert et al. 2010; Domes et al. 2007; Klimov et al. 2018; Koller et al. 2012; Li and Xue 2019; Liana and Witaliński 2005; Maraun et al. 2004; Norton 1994; 1998; Pepato and Klimov 2015; Sakata and Norton 2001). The relationship of Oribatida and Astigmata has been challenging to resolve for the past decades and several studies using different set of genes, ultra-conserved elements or

transcriptomic data reconstructed discordant phylogenies, some of which are similar to ours (Dabert et al. 2010; Domes et al. 2007; Klimov et al. 2018; Li and Xue 2019; Lozano-Fernandez et al. 2019; Maraun et al. 2004; Pepato and Klimov 2015; Van Dam et al. 2019). Overall, the Oribatid-Astigmatid relationship remains unresolved and a broader taxon sampling, especially of more basal Astigmata, will be necessary (Domes et al. 2007; Klimov et al. 2018; Lozano-Fernandez et al. 2019; Norton 1994; 1998; Van Dam et al. 2019). We recovered Trombidiformes (Prostigmata and Sphaerolichida) as sister group of the Sarcoptiformes (Oribatida and Astigmata) constituting the Acariformes (**Figure 3a**). Neither the maximum likelihood phylogeny (**Figure 3a**), nor the coalescence-based phylogeny (**Supplementary Figure S2**) reconstructed the Acari (i.e. Acariformes and Parasitiformes) as a monophyletic taxon. Even though there is morphological, ultrastructural and molecular evidence for a biphyletic Acari, aswe recovered here, this relationship and larger-scale chelicerate relationships remain unclear (Alberti 1984; 1991; Dabert 2006; Dunlop and Alberti 2008; Jeyaprakash and Hoy 2009; Li and Xue 2019; Lozano-Fernandez et al. 2019; Van Dam et al. 2019).

To further assess the quality and homology of both the genome assembly (**Table 1**) and the OGS (**Figure 2**), we used the1066 arthropod Benchmarking Universal Single-Copy Ortholog (BUSCO) genes data set (Simão et al. 2015). Nearly all BUSCO genes were present in the *Archegozetes* assembly and OGS (96.2% and 97.3%, respectively; **Figure 3b**). Compared to other genomes sequenced so far, the *Archegozetes* genome has the highest completeness among oribatid mites and the OGS completeness is on par to the well curated genomes of other chelicerate species and *Drosophila melanogaster* (**Figure 3b**). This result is not surprising because the *Archegozetes* genome was assembled from long-read and short-read data, while all other oribatid mite genomes were solely short reads sequenced on older Illumina platforms (Bast et al. 2016). The fraction of duplicated BUSCO genes in *Archegozetes* (4%) was similar to that of the spider mite and deer tick (Grbić et al. 2011; Gulia-Nuss et al. 2016), but very low compared to the house spider (**Figure 3c**), whose genome underwent an ancient whole-genome duplication (Schwager et al. 2017).

Overall, the high quality of both the genome assembly and OGS of *Archegozetes* compared to those of other oribatid mites, strongly indicates the importance of this genomic resource. We next categorized all protein models from the OGS by conversation level based on a

global clustering orthology analysis (OrthoFinder; Emms and Kelly 2015b) of 23 species (**Figure 3c; supplementary Figure S3**) representing Acariformes, Parasitiformes, several other chelicerates and the fly *Drosophila*. As for most other species (Siepel et al. 2005; Thomas et al. 2020), about a third of all orthogroups was highly conserved (**Figure 3c**) across the arthropods, being either in all species (10%; **Figure 3d**) or is most (22%; **Figure 3d**). Only 1% of all *Archegozetes* orthogroups did not show homology and were species specific (**Figure 3c and d**). Only a low proportion (**Figure 3c**) of orthogroups was conserved across the higher taxonomic levels (all <1% in *Archegozetes;* **Figure 3d**), which is in line with previous studies that included prostigmatid and mesostigmatid mites (Dong et al. 2017; Dong et al. 2018; Hoy et al. 2016). Interestingly, there was a large proportion of orthogroups conserved across all Oribatida (43% in *Archegozetes;* **Figure 3d**) and also about 19% of orthogroups in *Archegozetes* were shared only with other Nothrina (**Figure 3d**). A fairly large percentage of these orthogroups may contain potentially novel genes that await experimental verification and functional analyses (Emms and Kelly 2015b; Nagy et al. 2020; Thomas et al. 2020). Especially the lack of homology within the Sarcoptiformes (2-3%; **Figure 3c**) may explain the controversial placement of Astigmata as a sistergroup of Oribatida that we recovered **(Figure 3a**). This grouping is likely caused by a long-branch attraction artifact and the sister relationship was incorrectly inferred (Dabert 2006; Dabert et al. 2010; Domes et al. 2007; Klimov et al. 2018; Pepato and Klimov 2015), because orthogroup clustering could not detect enough homology between oribatids and the Astigmata so far sequenced , which are highly derived. Hence, a broad taxon sampling of basal astigmatid mite genomes seems necessary to resolve Oribatida-Astigmata relationship (Li and Xue 2019; Norton 1994; 1998; Pepato and Klimov 2015; Van Dam et al. 2019).

Repeat content analysis and transposable elements (TEs)

For clonal species like *Archegozetes,* reproducing in the absence of recombination, it has been hypothesized that a reduced efficacy of selection could results in an accumulation of deleterious mutations and repeats in the genome (Arkhipova and Meselson 2000; Barton 2010; Charlesworth 2012; Muller 1964; Nuzhdin and Petrov 2003; Schön et al. 2009). There is, however, no evidence for such an accumulation in oribatids or other arthropods (Bast et al. 2016). Generally, we found that most of the repetitive content in *Archegozetes* could not be classified (57%; **Figure 4a**). The high proportion of unknown repeats likely corresponds to novel

predicted repetitive content, because of limited repeat annotation of mites in common repeat databases such as RepBase (Bast et al. 2016). Regarding the two major classes of repeat content, DNA transposons made up about 32% of total repeats, while only 5% represented retrotransposons (**Figure 4a**). About 6% of total repetitive content comprised simple and low complexity repeats (**Figure 4a**). Overall, the total repetitive content (32%, **Figure 4b**) seems to be within a normal range for chelicerates and arthropods.

The repeat content found in other oribatid mites was lower (Bast et al. 2016), but recent studies suggest that sequencing technology, read depth and assembly quality are paramount to the capacity of identifying repeat content and TEs (Bourque et al. 2018; Panfilio et al. 2019). Hence, it is very likely the current genomic data for other Oribatida underestimates the actual total repetitive content. More low-coverage, long-read sequencing could reduce the assembly fragmentation and likely reveal a higher proportion of repeats, closer to the actual repetitiveness of oribatid genomes (Panfilio et al. 2019).

Different classes of transposable elements (TEs) are characterized by the mechanism they use to spread within genomes and are known to influence population dynamics differently (Bourque et al. 2018; Crescente et al. 2018; Finnegan 1989). We therefore analyzed the evolutionary history of TE activity in *Archegozetes* in more detail (**Figure 4c**). The main TE superfamilies were DNA transposons (**Figure 4a and c**), which seems to be a common pattern of oribatid mite genomes. For *Archegozetes,* they appear to have accumulated in the genome for a long time (i.e. they are more divergent from the consensus; (Waterston et al. 2002)) with Tc1/mariner – a superfamily of interspersed repeats DNA transposons (Bourque et al. 2018) – being the most abundant one (**Figure 4c**). Interestingly, we found an increase in TE activity with 0-3% sequence divergence range, indicating a recent burst (**Figure 4c**). This burst contained an enrichment of DNA Mavericks, which are the largest and most complex DNA transposons with homology to viral proteins (Bourque et al. 2018), but also several of retrotransposons. Among these, is the Long Terminal Repeat (LTR) gypsy retroelement (**Figure 4c**), which is closely related to retroviruses (Bourque et al. 2018). Like retroviruses, it encodes genes equivalent to *gag*, *pol* and *env*, but relatively little is known about how it inserts its DNA into the host genome (Dej et al. 1998; Havecker et al. 2004). So far, it is unknown what these TEs do in *Archegozetes*,

but the recent burst in TE abundance might suggest that some changes in the genome might have happened since the became a laboratory model nearly 30 years (Heethoff et al. 2013).

The *Archegozetes* Hox cluster

The Hox genes are a group of highly conserved transcription factor-encoding genes that are used to pattern the antero-posterior axis in bilaterian metazoans (Holland and Hogan 1988; Hrycaj and Wellik 2016). Ancestrally, arthropods likely had ten Hox genes arranged in a cluster (Hughes and Kaufman 2002). During arthropod development, the Hox genes specify the identities of the body segments, and mutations in Hox genes usually result in the transformation of segmental identities (Hughes and Kaufman 2002). The importance of Hox genes in development of metazoans makes knowledge of their duplication and disappearances important for understanding their role in the evolution of body plans (Hughes and Kaufman 2002).

Mites largely lack overt, external signs of segmentation, other than the serially arranged appendages of the prosoma (Dunlop and Lamsdell 2017). Signs of segmentation in the posterior body tagma, the opisthosoma, do exist in adult members of Endeostigmata (van der Hammen 1970). However, these segmental boundaries are largely present only in the dorsal opisthosoma, making it difficult to assess how these correspond to the ventral somites (Dunlop and Lamsdell 2017; van der Hammen 1970). Developmental genetic studies of the spider mite and *Archegozetes* suggest that acariform mites only pattern two segments in the posterior body region, the opisthosoma, during embryogenesis (Barnett and Thomas 2012; 2013b; 2018; Grbić et al. 2011). This stands in stark contrast to other studied chelicerate embryos. For example, during embryogenesis the spider *Parasteatoda tepidariorum* patterns twelve opisthosomal segments (Schwager et al. 2015) and the opilionid *Phalangium opilio* patterns seven (Sharma et al. 2012). Furthermore, a member of Parasitiformes, the tick *Rhipicephalus microplus,* appears to pattern eight opisthosomal segments during embryogenesis (Santos et al. 2013).

Parallel to the observation of segmental reduction in *T. urticae*, genomic evidence suggests that this acariform mite has lost two of its Hox genes, *i.e., Hox3* and *abdominal-A (abd-A*) (Grbić et al. 2011). Interestingly, orthologs of *abd-A* in other studied arthropods pattern the posterior segments as well. A genomic comparison of arthropod Hox clusters has also shown a correlation between independent losses of *abd-A* and a reduction in posterior segmentation (Pace et al. 2016). To investigate whether the loss of segmentation in *Archegozetes* is also due to an

absence in *abd-A*, we annotated its Hox cluster, paying close attention to the region between the Hox genes *Ultrabithorax* (*Ubx*) and *Abdominal-B* (*Abd-B*), which is usually where this gene resides in other arthropods (Hughes and Kaufman 2002). Our results suggest that the *Archegozetes* Hox genes are clustered in a contiguous sequence (tig00005200_pilon, total size ~7.5 Mbp) in the same order as suggested for the ancestral arthropod (Heethoff and Rall 2015). Furthermore, we found no sequences suggestive of an *abd-A* ortholog in *Archegozetes* (**Figure 5a**). These data also support the findings of a previous PCR survey that retrieved no *abd-A* ortholog in *Archegozetes* (Cook et al. 2001). Genomic evidence from the Parasitiformes *Ixodes scapularis* and *Metaseiulus occidentalis* reveal that these taxa maintain orthologs of all ten Hox genes, however in *M. occidentalis* these genes are not clustered as they are in *I. scapularis* (Gulia-Nuss et al. 2016; Hoy et al. 2016).

Taken together, these observations suggest that the last common ancestor of acariform mites likely lost its *abdominal-A* gene as well as experiencing a reduction in opisthosomal segmentation (**Figure 5b**). Alternatively, these shared losses of *abd-A* may be due to convergence due to similar selective pressures favoring a reduction in body size. The dorsal, external segmentation of endeostigmatid mites does not necessarily contradict the hypothesis of a loss of *abd-A* at the base of the acariform mites. As Hox genes are usually deployed after the genetic establishment of segments in arthropods (Hughes and Kaufman 2002), the opisthosomal segments in endeostigmatid mites may still develop in the absence of *abd-A*. However, this hypothesis needs further testing with observations of segmental gene expression in endeostigmatids as well as additional acariform species.

Life-stage specific RNA expression patterns

Developmental and gene expression data from *Archegozetes* embryos (**Figure 5 d** and **e**) have elucidated many of the potential mechanisms driving the morphogenesis of manydevelopmental peculiarities. These peculiarities include the suppression of the fourth pair of walking legs during embryogenesis as well as the reduction of opisthosomal segmentation (Barnett and Thomas 2012; 2013a; 2013b; 2018; Telford and Thomas 1998; Thomas 2002). In typical acariform mites, embryogenesis ends with the first instar, the prelarva, which usually remains within the egg chorion, as in *Archegozeetes*. Hatching releases the second instar, the larva, which is followed by three nymphal instars (proto-, deutero- and tritonymph) and the

384 adult, for a total of six instars. (Heethoff et al. 2007). Thus far, methodological limitations have
385 made it difficult to examine how mite segmentation and limb development progress throughout
386 these instars.

387 To this end, we used RNAseq to calculate the transcripts per million (tpm) values of
388 genes known to be, or suspected to be, involved in limb development and segmentation
389 throughout the six different instars of *Archegozetes*. Prior to comparing these tpm values, gene
390 orthology was confirmed *via* phylogenetic analyses (**supplementary Figures S4-S11**; see **Table**
391 **S2** for phylogenetic statistics and **Table S3** for tpm values). Regarding the total number of genes
392 expressed across the different life stages, we found that earlier instars generally expressed a
393 higher number of genes (**Figure 5c**). While most expressed genes were shared across all instars,
394 more transcripts were shared between the eggs and the larvae and among all five juvenile instars.
395 Additionally, we found that earlier instars expressed a larger number of stage-specific genes as
396 compared to later instars and adults (**Figure 5c**).

397 Gene expression, SEM and time-lapse data have revealed that the development of the
398 fourth pair of walking legs in *Archegozetes* is suppressed until after the larval instar (Barnett and
399 Thomas 2012; 2018; Telford and Thomas 1998). The resulting larva is thus hexapodal (see also
400 embryo in **Figure 5e**), which constitutes a putative synapomorphy of Acari, if they are
401 monophyletic (Dunlop and Alberti 2008). In arthropods, the development of the limbs is
402 generally accomplished via the activity of highly conserved regulatory genes, termed the "limb
403 gap genes." These genes are expressed along their proximo-distal axes to establish the specific
404 identities of the limb podomeres. The limb gap genes include *extradenticle* (*exd*) and
405 *homothorax* (*hth*), which act together to specify the proximal limb podomeres, *dachshund* (*dac*),
406 which specifies the medial podomeres, and *Distal-less* (*Dll*) which specifies the distal-most
407 podomeres. It was previously shown that the deployment of these genes in the anterior
408 appendages of *Archegozetes*, *i.e.*, the chelicerae, pedipalps and first three pairs of walking legs
409 (**Figure 5d** and **e**), is similar to that of other chelicerate taxa (Barnett and Thomas 2013a;
410 Schwager et al. 2015; Sharma et al. 2015). However, in the anlagen of the fourth pair of walking
411 legs, only the proximal-specifying genes, *exd* and *hth*, are expressed (Barnett and Thomas
412 2013a).

Whether the limb gap genes are re-deployed during the transition from the prelarval to larval instars in order to activate the development of the fourth pair of walking legs remains an open question. We therefore compared the average tpm values of verified limb gap genes (i.e., Al-*Dll*, Al-*Hth*, Al-*exd*, and Al-*dac* (Barnett and Thomas 2013a)) in embryos and at each instar stage (**Figure 5f**). We also compared the tpm values of the *Archegozetes* orthologs of *Sp6-9* and *optomotor blind*, genes shown to be involved in limb formation in spiders (Heingård et al. 2019; Königsmann et al. 2017). We hypothesized that limb development genes would show high expression in the larval stage leading to the development of the octopodal protonymph. We did observe an increase in the tpm averages of Al-*hth* as well as Al-*optomotor-blind*, however the aforementioned limb gap gene expression levels were similar between these instars (**Figure 5f**). Taken together, these genes may not be up-regulated for the formation of the fourth pair of walking legs between these two instars.

Chelicerate embryos segment their bodies through a "short/intermediate germ" mechanism, whereby the anterior (prosomal) segments are specified asynchronously (Schwager et al. 2015). This usually occurs well before the sequential addition of posterior segments from a posterior growth zone. Based on neontological and paleontological data, chelicerate arthropods may have ancestrally had an opisthosoma comprised of 12 or more segments (Dunlop and Selden 1998; Dunlop 2010; Dunlop and Lamsdell 2017). Embryonic expression data for the segment polarity genes, those genes that delineate the boundaries of the final body segments, have shown that in most studied chelicerate embryos opisthosomal segments are delineated during embryogenesis (Dunlop and Lamsdell 2017; Schwager et al. 2015). However, as discussed above, expression data in *Archegozetes* embryos suggest that only two opisthosomal segments are patterned during embryogenesis (Barnett and Thomas 2012; 2018); this indicates that mites have significantly reduced their number of opisthosomal segments either by loss or by fusion. Further complicating this is the observation that many mites add segments as they progress through the larval instars, a phenomenon known as anamorphic growth (Dunlop and Lamsdell 2017).

To determine by what genetic process *Archegozetes* may add segments during post-embryonic ontogeny, we assessed the expression of known chelicerate and arthropod segmentation genes in each instar transcriptome (**Figure 5f**) (Schwager et al. 2015). We

observed an up-regulation of the segmentation genes *hedgehog* and *engrailed* in the larvae, as well as the slight up-regulation of *patched* and *pax3/7*. Furthermore, the segmentation gene *wingless* was slightly up-regulated in the protonymph, as well as a slight up-regulation of *hedgehog* in the tritonymph. Lastly, we found that transcripts of the genes *pax3/7* and *runt* were up-regulated in adults. These results suggest that *Archegozetes* does pattern body segments during the progression through the it's instars similar to other Chelicerata.

Another peculiarity of *Archegozetes* is that these mites lack eyes (see more details below). Eye loss has been documented in other arachnid clades, including independently in other members of Acari (Evans 1992; Walter and Proctor 1999), and it has been recently demonstrated that a species of whip spider has reduced its eyes by reducing the expression of retinal determination genes that are shared throughout arthropods (Gainett et al. 2020). We sought to determine if eye loss in *Archegozetes* also is associated with the reduced expression of these genes (see also analysis of photoreceptor genes below). The genes, which have been shown to be expressed in the developing eyes of spiders and whip scorpions, include *Pax-6*, *six1/sine oculis* (*so*), *eyes absent* (*eya*), *Eyegone*, *Six3*/*Optix*, and *atonal* (Gainett et al. 2020; Samadi et al. 2015; Schomburg et al. 2015). We also followed the expression of Al-*orthodenticle*, a gene previously shown to be expressed in the ocular segment of *Archegozetes* (Telford and Thomas 1998). Surprisingly, all of these genes, excluding the *Pax-6* isoform A and *eyegone*, are indeed expressed during embryogenesis (**Figure 5f**). Aside from the larval expression of the *Pax-6* isoform A during the larval stage, these eye-development genes remain quiescent until the adult stage, where all but *Pax-6* isoform A, *six3* and *atonal* are up-regulated (**Figure 5f**). These results are exceedingly surprising, given the conserved role of genes in retinal patterning. They suggest a novel role for these genes, or alternatively, these expression patterns could be the result of early expression of a retinal determination pathway followed by negative regulation by other genes to suppress eye development.

Photoreceptor and chemosensory system of *Archegozetes longisetosus*

Unlike insects and crustaceans, chelicerates do not have compounds eyes – with horseshoe crab being an exception. Generally, mites are eyeless or possess one or two pairs of simple ocelli (Alberti and Coons 1999; Alberti and Moreno-Twose 2012; Exner 1989; Harzsch et al. 2006; Patten 1887). Ocelli are common in Prostigmata and Endeostigmata , among

Acariformes, as well Opilioacarida – the most likely sister group to the Parasitiformes – but are absent in most Oribatida, Astigmata, Mesostigmata and ticks (Norton and Franklin 2018; Norton and Fuangarworn 2015; Walter and Proctor 1998; Walter and Proctor 1999). This suggests that the presence of eyes might be an ancestral condition for both Acariformes and Parasitiformes, while more derived mites rely largely on chemical communication systems (Alberti and Coons 1999).

In oribatid mites, detailed morphological and ultrastructural investigations have suggested that setiform sensilla are the most obvious sensory structures (**Figure 6a**) (Alberti 1998; Alberti and Coons 1999; Walter and Proctor 1999). The trichobothria are very complex, highly modified (e.g. filiform, ciliate, pectinate, variously thickened or clubbed) no-pore setae which are anchored in a cup-like base and likely serve as mechanosensory structures. In contrast, the setal shafts of solenidia and eupathidia (**Figure 6a**) both possess pores (Alberti 1998; Alberti and Coons 1999; Walter and Proctor 1999). Solenidia have transverse rows of small pores visible under a light microscope and likely function in olfaction, while the eupathidia have one or several terminal pores and likely are used as contact/gustatory sensilla (**Figure 6a**) (Alberti 1998; Alberti and Coons 1999). Previous work demonstrated that oribatid mites indeed use olfactory signals in the context of chemical communication and food selection (Brückner et al. 2018a; Brückner et al. 2018b; Heethoff et al. 2011a; Heethoff and Raspotnig 2012; Raspotnig 2006; Shimano et al. 2002).

Interestingly, detailed morphological and ultrastructural studies showed that light-sensitive organs exist in some Palaeosomata and Enarthronota (probably true eyes) as well as in Brachypylina (the secondary lenticulus), representing lower and highly derived oribatid mites, respectively (Alberti and Coons 1999; Alberti and Moreno-Twose 2012; Norton and Franklin 2018; Norton and Fuangarworn 2015). *Archegozetes* and most other oribatids, however, are eyeless, yet there is scattered experimental and some anecdotal evidence that even these mites show some response to light and seem to avoid it ('negative phototropism' or 'negative phototaxis') (Madge 1965; Trägårdh 1933; Walter and Proctor 1999; Woodring 1966). Hence, we mined the genome of *Archegozetes* for potential photoreceptor genes and found two genes of the *all-trans retinal peropsin* class and one gene related to spider mite *rhodopsin-7-like* gene (**Figure 6b**). *Peropsin-like* genes are also present in other eyeless ticks. In jumping spiders it

encodes for nonvisual, photosensitive pigments, while *rhodopsin-7* may be involved in basic insect circadian photoreception (Eriksson et al. 2013; Koyanagi et al. 2008; Nagata et al. 2010; Senthilan et al. 2019; Senthilan and Helfrich-Förster 2016; Shen et al. 2011). Taken together, this might suggest that eyeless species like *Archegozetes* use *peropsin*- and *rhodopsin-7-like* genes for reproductive and diapause behaviors, or to maintain their circadian rhythm, as well as negative phototaxis.

However, the main sensory modality soil mites use is chemical communication via olfaction (Alberti 1998; Alberti and Coons 1999; Brückner et al. 2018a; Brückner et al. 2018b; Raspotnig 2006; Shen et al. 2011; Walter and Proctor 1999). In contrast to insects, but similar to crustaceans and Myriapoda, mites do not have the full repertoire of chemosensory classes, they are missing odorant receptors and odorant-binding proteins (**Table 2**) (Dong et al. 2017; Dong et al. 2018; Hoy et al. 2016; Maraun et al. 2007; Raspotnig 2009; Sánchez-Gracia et al. 2009; Sánchez-Gracia et al. 2011; Vieira and Rozas 2011). Although chemosensory protein (CSP) encoding genes are absent in most mite genomes, we identified one gene encoding for such a protein in *Archegozetes* and one CSP has been previously found in the deer tick (Table 2). Hence, *Archegozetes* should primarily rely on gustatory receptors (GRs) and ionotropic receptors (IRs). Both the number of GRs (68 genes; **Figure 6d**) and IRs (3 genes; **Figure 6c**) was very well within the range of most mites and ticks and there was no evidence for any massive chemoreceptor expansion like in the spider mite (**Table 2**) (Ngoc et al. 2016). This was surprising because *Archegozetes*, like other acariform mites have many multiporous solenidia, present on all legs and the palp, but appear to only have a limited number of chemoreceptors.

Canonical ionotropic glutamate receptors (iGluRs) are glutamate-gated ion channels with no direct role in chemosensation, which come in two major subtypes: either NMDA iGluRs which are sensitive to N-methyl-D-aspartic acid (NMDA) or non-NMDA iGluRs. The latter group – at least in *Drosophila* – seems to have essential functions in synaptic transmission in the nervous system and have been associated with sleep and vision (Benton et al. 2009; Croset et al. 2010; Ngoc et al. 2016; Sánchez-Gracia et al. 2009; Sánchez-Gracia et al. 2011). None of the IRs we found in the *Archegozetes* genome belonged to the NMDA iGluRs and most were classified as non-NMDA iGluRs (**Figure 6c**). Nothing is known about their functions in mites. It is, however, likely that they perform similar tasks in synaptic transmission in the brain and

musculature. In *Drosophila* a specific set of chemosensory IRs, which do not bind glutamate, respond to acids and amines (*IR25a*), but also to temperature (*IR21a*, *IR93a*). For *Archegozetes* we found 3 IRs, like *IR21a* and *IR93a* of *Drosophila*, which fell into the antenna/1st leg IRs category (**Table 2**; **Figure 6c**) (Budelli et al. 2019; Knecht et al. 2016; Rytz et al. 2013). This is consistent with an assumed limited contribution of IRs to the perception of chemical cues. Furthermore, it is so far unclear whether these IRs are expressed in the first pair of legs (**Figure 6a** and **c**) in *Archegozetes*, but similar genes seem to be expressed in the legs of other mite species, which could suggest a similar function as in the fruit fly.

GRs are multifunctional proteins and at least in insects they are responsible for the perception of taste, heat or volatile molecules (Montell 2009). In *Archegozetes* we found 68 GRs, over half of which belonged to a species-specific expansion of the GR gene family (**Figure 6d**). Generally, it is unclear if GRs in *Archegozetes* and other mites have similar functions as in insects, but the GR gene family is heavily expanded in many acariform mites and also is present in ticks (**Table 2**), suggesting an important biological role (Barrero et al. 2017; Dong et al. 2017; Dong et al. 2018; Gulia-Nuss et al. 2016; Hoy et al. 2016; Ngoc et al. 2016). This is supported by experimental evidence which suggested that ticks and other mites, including *Archegozetes,* use chemical cues to find their host, communicate or discriminate food (Barrero et al. 2017; Brückner et al. 2018a; Brückner et al. 2018b; Bunnell et al. 2011; Gulia-Nuss et al. 2016; Kuwahara 2004; Raspotnig 2006; Yunker et al. 1992).

In general, not much is known about the nervous and sensory system of oribatid mites, or about sensory integration or the neuronal bases of their behavior (Alberti 1998; Alberti and Coons 1999; Norton 2007). Modern methods like Synchrotron X-ray microtomography (SRμCT) recently made it possible to investigate the organization and development of the nervous systems of oribatid mites (**Figure 6e;** (Hartmann et al. 2016)) and here we provide the first genomic resource for the investigation of the photo- and chemosensory systems of Oribatida (**Figure 6b-d**).

<u>Horizontal gene transfer event sheds light on oribatid feeding biology</u>

Horizontal gene transfer (HGT) is common among mites and other soil organisms (Dong et al. 2018; Faddeeva-Vakhrusheva et al. 2016; Grbić et al. 2011; Mayer et al. 2011; Wu et al. 2017; Wybouw et al. 2018). In some cases, genes that had been horizontally transferred now

563 have pivotal biological functions. For instance, terpene and carotenoid biosynthesis genes in
564 trombidiid and tetranychid mites, respectively, are found nowhere else in the animal kingdom
565 (Altincicek et al. 2012; Dong et al. 2018). Yet they show high homology with bacterial (terpene
566 synthase) or fungal (carotenoid cyclase/synthase/desaturase) genes, suggesting horizontal gene
567 transfer from microbial donors (Altincicek et al. 2012; Dong et al. 2018). At least the carotenoid
568 biosynthesis genes in spider mites still code for functional enzymes and equip these phytophages
569 with the ability to *de novo* synthesize carotenoids, which can induce diapause in these animals
570 (Altincicek et al. 2012).

571 Soil microarthropods like collembolans show numbers of horizontally transferred genes
572 that are among the highest found in Metazoan genomes, exceeded only by nematodes living in
573 decaying organic matter (Crisp et al. 2015; Faddeeva-Vakhrusheva et al. 2016; Wu et al. 2017).
574 Interestingly, many HGT genes found in collembolans are involved in carbohydrate metabolism
575 and were especially enriched for enzyme families like glycoside hydrolases, carbohydrate
576 esterases or glycosyltransferases (Faddeeva-Vakhrusheva et al. 2016; Wu et al. 2017). All three
577 enzyme families are involved in the degradation of plant and fungal cell walls (Gilbert 2010;
578 Latgé 2007). Hence, it has been hypothesized that cell-wall degrading enzymes acquired by HGT
579 are beneficial for soil organisms as it allowed such animals to access important food source in a
580 habitat that is highly biased towards polysaccharide-rich resources (Faddeeva-Vakhrusheva et al.
581 2016; Faddeeva-Vakhrusheva et al. 2017; Mitreva et al. 2009; Wu et al. 2017).

582 To assess the degree of HGT in *Archegozetes* we first used blobtools (v1.0) (Laetsch and
583 Blaxter 2017) to generate a GC proportion vs read coverage plot of our genome assembly, in
584 order to remove scaffolds of bacterial origin (**Figure 7a**; 438 contigs of ~ 8.5 Mb of
585 contamination). Of the remaining scaffolds, candidate HGTs were identified using the Alien
586 Index (Flot et al. 2013; Thorpe et al. 2018), where HGTs are those genes with blast homology
587 (bit score) closer to non-metazoan than metazoan sequences (**supplementary Table S4**). We
588 further filtered these HGT candidates to remove those that overlapped predicted repeats by ≥
589 50%, resulting in 617 genes. As HGT become integrated into the host genome, they begin to
590 mirror features of the host genome, including changes in GC content and introduction of introns
591 (Lawrence 1997). Comparing the GC content of the HGT candidates showed two distinct peaks,
592 one at 54.3% and the other at 35.2%, slightly higher than the remaining *Archegozetes* genes, GC

content of 31.5% (**Figure 7b**). Of the 407 HGT genes that shared similar GC content to the host genome, 73.5% had at least one intron (**Table S4**). In a final step, we used the gene expression data (RNAseq) to filter the list of all putative HGT genes and only retained candidates that were expressed in any life stage of *Archegozetes* (*n*= 298 HGT genes).

The majority of HGT candidates were of bacterial origin (75.2%), followed by genes likely acquired from fungi (13.4%), while transfer from Archaea, plants, virus and other sources was comparatively low (**Figure 7d**). This composition of HGT taxonomic origin is different from genes found in collembolans, which appear to have acquired more genes of fungal and protist origin (Faddeeva-Vakhrusheva et al. 2016; Faddeeva-Vakhrusheva et al. 2017; Wu et al. 2017). Subsequently, we performed an over-representation analysis of GO terms associated with these genes. We found an over-representation of genes with GO terms related to methyl transfer reactions and breaking glycosidic bonds (molecular function; **Figure 7c**) as well as carbohydrate metabolism, among others (biological process; **Figure 7c**) providing a first line of evidence that *Archegozetes* possess HGT related to plant- and fungal cell wall degradation similar to springtails.

As mentioned previously, oribatid mites are among the few Chelicerata that ingest solid food and are primary- and secondary decomposers feeding on dead plant material and fungi (Cohen 1995; Dunlop and Alberti 2008; Heethoff and Norton 2009; Maraun et al. 2011; Norton 2007; Shultz 2007). It was argued for decades that the enzymes necessary to break down these polysaccharide-rich resources originate from the mite's gut microbes (Siepel and de Ruiter-Dijkman 1993; Smrž 1992; Smrž 2000; Smrž and Čatská 2010; Smrž and Norton 2004; Stefaniak 1976; 1981). Microbes might be mixed with the food in the ventriculus and digest it while passing through the alimentary tract as food boli enclosed in a peritrophic membrane (see **Figure 7f** for an example) (Stefaniak 1976; 1981). However, screening the HGT candidate list for potential cell-wall degrading enzymes and mapping their overall and life-stage specific expression in *Archegozetes* using the RNAseq reads, revealed at least seven HGT genes related to polysaccharide breakdown (**Figure 7g**). We found that specifically members of the *glycoside hydrolases family 48* and *cellulose-binding domain* genes showed high expression in most life stages - the egg being an obvious exception (**Figure 7g**). Moreover, the majority of these genes

were flanked by a predicted metazoan gene, suggesting host transcriptional regulation (**Table S4**).

In a last step we blasted the highly expressed HGT candidates (**Figure 7g**) against the non-redundant protein sequence database, aligned the sequences with the highest alignment score. Eventually, we performed a phylogenetic maximum likelihood analysis. For the highest expressed HGT related to cell-wall-degrading enzymes (*glycoside hydrolases family 48 gene II*), we recovered that the *Archegozetes* sequences was well nested within a clade of *GH 48* sequences from herbivores beetles (McKenna et al. 2019), which appear to be related to similar genes from various *Streptomyces* (**Figure 7e**) and we reconstructed similar phylogenies for other highly expressed HGT candidates (**supplementary Figure S12**). All the sequences of beetle *glycoside hydrolases family 48* members (**Figure 7e**) were included in recent studies arguing for a convergent horizontal transfer of bacterial and fungal genes that enabled the digestion of lignocellulose from plant cell walls in herbivores beetles (McKenna et al. 2016; McKenna et al. 2019). They showed that phytophagous beetles likely acquired all genes of the *GH 48 family* from Actinobacteria (including *Streptomyces*) (McKenna et al. 2019) and our phylogenetic analysis (**Figure 7e**) revealed the same pattern as well as a highly similar tree topology (compare to Fig 3B in (McKenna et al. 2019)).

Overall, our findings indicate that genes encoding for enzymes in *Archegozetes* capable of degrading plant and fungal cell walls were likely horizontally transferred from bacteria (likely *Streptomyces*). Bacterial symbionts and commensal living in the mites' gut are still likely to contribute to the breakdown of food (**Figure 7f**). Yet, the high expression of genes encoding cell-wall degrading enzymes (**Figure 7g**) as well as the evolutionary analyses of such genes (**Figure 7e**) suggest that *Archegozetes* – and potentially many other oribatid mites – are able to exploit polysaccharide-rich resources like dead plant material or chitinous fungi without microbial aid. Enzymological and microscopical investigation of *Archegozetes* have suggested that certain digestive enzymes (chitinase and cellulase) are only active when the mites consume a particular type of food (e.g. algae, fungi or filter paper) (Smrž and Norton 2004). These results were interpreted as evidence that these enzymes are directly derived from the consumed food source (Smrž and Norton 2004). By contrast, we argue that this instead confirms our findings of HGT: upon consumption of food containing either chitin or cellulose, gene expression of

652 polysaccharide-degrading enzymes starts, and proteins can readily be detected. Further
653 enzymological studies have placed oribatid mites in feeding guilds based on carbohydrase
654 activity and also found highly similar enzyme activity between samples of mites from different
655 times and locations (Luxton 1972; 1979; 1981; 1982; Siepel and de Ruiter-Dijkman 1993).
656 Future functional studies can disentangle the contribution of the host and microbes to cell wall
657 digestion and novel metabolic roles of the HGTs identified here.

658 Biosynthesis of monoterpenes – a common chemical defense compound class across oribatid
659 mite

660 Oribatid and astigmatid mites are characterized by a highly diverse spectrum of natural
661 compounds that are produced by and stored in so-called oil glands (for an example see **Figure**
662 **8a**) (Heethoff et al. 2016; Raspotnig 2009; Raspotnig et al. 2011). These paired glands are
663 located in the opisthosoma (i.e. the posterior part of chelicerate arthropods, analogous to the
664 abdomen of insects) and are composed of a single-cell layer invagination of the cuticle (**Figure**
665 **8f**). As previously mentioned, mites use chemicals produced by these glands to protect
666 themselves against environmental antagonists (predators or microbes) or use them as
667 pheromones (Brückner et al. 2015; Heethoff et al. 2011a; Heethoff and Rall 2015; Heethoff and
668 Raspotnig 2012; Raspotnig 2006; 2009; Shimano et al. 2002). The monoterpene aldehyde citral –
669 a stereoisomeric mixture of geranial ((*E*)-3,7-dimethylocta-2,6-dienal) and neral ((*Z*)-3,7-
670 dimethylocta-2,6-dienal) – and its derivatives are widely detected compounds in glandular
671 secretions of oribatids and astigmatids (Koller et al. 2012; Kuwahara 2004; Kuwahara et al.
672 2001; Raspotnig et al. 2004; Sakata 1997; Sakata and Norton 2001; Sakata and Norton 2003;
673 Sakata et al. 1995). These monoterpenes have been called "astigmatid compounds" (Sakata and
674 Norton 2001) as they characterize the biochemical evolutionary lineage of major oribatid mite
675 taxa (Mixonomata and Desmonomata) and almost all investigated astigmatid mites (Alberti
676 1984; Kuwahara 2004; Raspotnig 2009; Sakata 1997; Sakata and Norton 2001).

677 The chemical cocktail released by *Archegozetes* consists of a blend of 10 compounds
678 (**Figure 8a**) including two terpenes (approx. 45%)– neral and neryl formate – six hydrocarbons
679 (approx. 15%) and two aromatic compounds (approx. 40%) (Brückner and Heethoff 2017;
680 Sakata and Norton 2003). The hydrocarbons likely serve as solvents, while the terpenes and
681 aromatics are bioactive compounds used in chemical alarm and defense (Heethoff et al. 2011a;

Raspotnig 2006; Sakata and Norton 2003; Shimano et al. 2002). Recently, it was shown that *Archegozetes* synthesizes the two aromatic compounds using a polyketide-like head-to-tail condensation of (poly)-β-carbonyls *via* a horizontally acquired putative polyketide synthetase (Brückner et al. 2020). Studies in Astigmata found that the monoterpenes of these mites appeared to be made *de novo* from (poly)-*β*-carbonyls as well and one study identified a novel geraniol dehydrogenases (GeDH), unrelated to those of bacteria, in *Carpoglyphus lactis* (Morita et al. 2004; Noge et al. 2005; Noge et al. 2008). To learn about the biosynthesis of astigmatid compounds in *Archegozetes* and demonstrate the mite's applicability as research model for biochemical pathway evolution, we used the novel genomic resources presented in this study.

First, we delineated the basic biochemical reactions likely to happen in the *Archegozetes* gland through a stable-isotope labeling experiment. We supplemented the diet of the mite with food containing 25% heavy $^{13}C_6$ D-glucose and 10% antibiotics (a combination of three different antibiotics was fed, because this mixture is able to eliminate nearly all qPCR and FISH detectable bacteria found on the food and in the alimentary tract (Brückner et al. 2020)). To examine the incorporation of heavy $^{13}C_6$ D-glucose and its metabolic products into neral (**Figure 8b**) and neryl formate (**Figure 8c**), we compared selected fragment ions (M^+ and M^+-46, respectively) using single ion mass spectrometry. Both neral and neryl formate showed consistent enrichment in their M^+ to $[M+4]^+$ and $[M-46]^+$ to $[M-46+4]^+$-ion series, indicating that *Archegozetes* used glycolysis breakdown products of $^{13}C_6$ D-glucose for the biosynthesis of their monoterpenes. We then used the OGS mapped to KEGG metabolic pathways (Kanehisa et al. 2007) to reconstruct the backbone synthesis of terpenes in *Archegozetes* (**Figure 8d**). We found mapped mite genes, which suggest that *Archegozetes* synthesizes geranyl pyrophosphate (GPP) – the input substrate for further monoterpene synthesis – via the mevalonate pathway using the Mevalonate-5P to Isopentenyl-PP route (**Figure 8d**). The Mevalonate-5P pathway is used in most higher eukaryotes as compared to the Mevalonate-3P pathway in Archaea and the MEP/DOXP pathway in bacteria, some plants and apicomplexan protists (Breitmaier 2006; Degenhardt et al. 2009; Eisenreich et al. 2004; Miziorko 2011; Oldfield and Lin 2012; Trapp and Croteau 2001). This likely excludes any horizontal gene transfer of mevalonate pathway genes as *Archegozetes* uses enzymes similar to those of other animals.

The biosynthesis of monoterpenes not only depends on very widespread enzymes, but also requires more specific enzymes downstream of GPP (Breitmaier 2006; Degenhardt et al. 2009; Trapp and Croteau 2001). For instance, *Carpoglyphus lactis* expresses a unique geraniol dehydrogenase (GeDH) – catalyzing the oxidation of geraniol to geranial – different from all previously characterized geraniol-related and alcohol dehydrogenases (ADHs) of animals and plants (Noge et al. 2008). We used the functionally validated Carpoglyphus-GeDH (Noge et al. 2008), blasted its sequence against the *Archegozetes* OGS and found a homologous sequence. We used both mite sequences in an alignment with plant, fungal and bacterial GeDHs and animal ADHs and constructed a maximum likelihood phylogeny (**Figure 8e**). Similar to the previous analysis including only Carpoglyphus-GeDH, we found that the *Al-GeDH* represent a new class of geraniol dehydrogenases different from those in plants, fungi or bacteria and not nested within animal ADHs (**Figure 8e**). This is why we hypothesize that *Al-GeDH* is a novel expansion of the geraniol dehydrogenases gene family and has not been acquired by horizontal gene transfer, like other biosynthesis and digestive enzymes in *Archegozetes* (**Figure 7**; (Brückner et al. 2020)).

Based on our mass spectrometry data of stable isotopes and genomic analysis, we propose that the following biochemical pathway leading to monoterpenes is of oribatid mites (**Figure 8f and g**): geraniol is likely to be synthesized from GPP – the universal precursor of all monoterpenes – either enzymatically by a geraniol synthase (GES) or a diphosphate phosphatase (DPP), but possibly also endogenously by dephosphorylation of GPP (Beran et al. 2019; Liu et al. 2015; Oswald et al. 2007; Zhou et al. 2014). For *Archegozetes*, we could not find any *GES* or specific *DPP* in the OGS, thus geraniol might be formed from GPP *via* endogenous dephosphorylation, but further research is required to verify or falsify this hypothesis. Subsequently, geraniol is oxidized to geranial by the pervious described *GeDH* (**Figure 8e**) and readily isomerized to neral. Trace amounts of geranial have been found in *Archegozetes* and it is common among other oribatid and astigmatid mites, supporting this idea (Koller et al. 2012; Kuwahara 2004; Raspotnig et al. 2008; Raspotnig et al. 2004). Also, there is no evidence that geraniol is converted into nerol, or that neral is formed directly via oxidation of nerol (Morita et al. 2004; Noge et al. 2005; Noge et al. 2008). The most parsimonious explanation for neryl formate synthesis would be an esterification of the corresponding terpene alcohol nerol. There is, however, no evidence of nerol in the traces of any oribatid or astigmatid mite species (Kuwahara 2004; Raspotnig 2009; Raspotnig et al. 2011). Aliphatic non-terpene formats in Astigmata are

742 synthesized by dehomologation and generation of a one-carbon–shorter primary alcohol from an
743 aldehyde via hydrolysis of formate in a biological Baeyer–Villiger oxidation catalyzed by a
744 novel, uncharacterized enzyme (Shimizu et al. 2017). A similar reaction to synthesize terpene
745 formates is unlikely, as the terpenoid backbone would be shortened by one-carbon and this does
746 not happen in any possible scenario. The discovery of this Baeyer–Villiger oxidation mechanism,
747 however, highlights the probability that there are many very unusual reactions that remain to be
748 discovered in oribatid mites (Brückner et al. 2020).

749

750 **Conclusion**

751 The integrated genomic and transcriptomic resources presented here for *Archegozetes*
752 *longisetosus* allowed a number of insights into the molecular evolution and basic biology of
753 decomposer soil mites. Our analysis of an oribatid mite genome also provides the foundation for
754 experimental studies building on the long history of *Archegozetes'* as a chelicerate model
755 organism, which now enters the molecular genetics era (Aoki 1965; Heethoff et al. 2013; Norton
756 et al. 1993; Palmer and Norton 1992). This includes the study of biochemical pathways,
757 biochemistry, neuroethological bases of food searching behavior, and environmental impacts on
758 genomes of complex, clonal organisms.

759 Our evolutionary comparisons across the Chelicerata revealed interesting patterns of
760 genome evolution and how horizontal gene transfer might have shaped the feeding mode of soil
761 mites. We also showed how oribatid glandular biology and chemical ecology are reflected in the
762 genome. The community of researchers studying the fundamental biology of oribatid and other
763 free-living, non-parasitic mites is growing. We think that providing these genomic and
764 transcriptomic resources can foster a community effort to eventually transform basic research on
765 these mites into a modern, molecular discipline.

766 Key priorities for a future community research effort include i) sequencing organ-specific
767 transcriptomic data, ii) developing tools for genetic interrogation (RNAi or CRISPR/CAS9), iii)
768 establishing reporter linages with germ-line stable modifications (e.g. GAL4/UAS misexpression
769 systems), iv) constructing an whole-animal single-cell RNAseq expression atlas, and v)
770 gathering more genomic data to improve the genome assembly. Please do not hesitate to contact

771 the corresponding author, if you want to start your own culture of *Archegozetes*. He will be
772 happy to provide you with starter specimens and share rearing protocols with you.

773

774 **Materials and Methods**

775 Mite husbandry

776 The lineage 'ran' (Heethoff et al. 2013) of the pantropical, parthenogenetic oribatid mite
777 *Archegozetes longisetosus* was used in this study. Stock cultures were established in 2015 from
778 an already existing line and fed with wheat grass (*Triticum* sp.) powder from Naturya. Cultures
779 were maintained at 20-24°C and 90% relative humidity. Sterilized water and 3-5 mg wheat grass
780 were provided three times each week.

781 DNA extraction and Illumina sequencing

782 For the short-read library, DNA was extracted from ~200 mites that were taken from the
783 stock culture, starved for 24 h to avoid possible contamination from food in the gut, subsequently
784 washed with 1% SDS for 10 sec. For extraction of living specimens, we used the Quick-DNA
785 Miniprep Plus Kit (Zymo Research) according to the manufacturer's protocol. Amounts and
786 quality of DNA were accessed with Qubit dsDNA HS Kit (ThermoFisher) and with NanoDrop
787 One (ThermoFisher) with target OD 260/280 and OD 260/230 ratios of 1.8 and 2.0-2.2,
788 respectively. Extracted DNA was shipped to Omega Bioservices (Norcross, GA, USA) on dry
789 ice for library preparation and sequencing. DNA library preparation followed the KAPA
790 HyperPrep Kit (Roche) protocol (150 bp insert size) and 200 million reads were sequenced as
791 150bp paired-end on a HighSeq4000 (Illumina) platform.

792 High-molecular weight DNA isolation and Nanopore sequencing

793 Genomic DNA was isolated from ~300-500 mites starved for 24 h using QIAGEN Blood
794 & Cell Culture DNA Mini Kit. Briefly, mites were flash frozen in liquid nitrogen and
795 homogenized with a pestle in 1 ml of buffer G2 supplemented with RNase A and Proteinase K at
796 final concentrations of 200 ng/µl and 1 µg/µl, respectively. Lysates were incubated at 50°C for 2
797 h, cleared by centrifugation at 5 krpm for 5 min at room temperature and applied to Genomic tip
798 G/20 equilibrated with buffer QBT. Columns were washed with 4 ml of buffer QC and genomic

DNA was eluted with 2 ml of buffer QF. DNA was precipitated with isopropanol, washed with 70% EtOH and resuspended in 50 µl of buffer EB. DNA was quantified with Qubit dsDNA HS Kit (ThermoFisher) and the absence of contaminants was confirmed with NanoDrop One (ThermoFisher) with target OD 260/280 and OD 260/230 ratios of 1.8 and 2.0-2.2, respectively. DNA integrity was assessed using Genomic DNA ScreenTape kit for TapeStation (Agilent Technologies).

Libraries for nanopore sequencing were prepared from 1 µg of genomic DNA using 1D Genomic DNA by Ligation Kit (Oxford Nanopore) following manufacturer's instructions. Briefly, unfragmented DNA was repaired and dA tailed with a combination of NEBNext FFPE Repair Mix (New England Biolabs) and NEBNext End repair/dA-tailing Module (New England Biolabs). DNA fragments were purified with Agencourt AMPure XP beads (Beckman Coulter) and Oxford Nanopore sequencing adapters were ligated using NEBNext Quick T4 DNA Ligase (New England Biolabs). Following AMPure XP bead cleanup, ~500 ng of the library was combined with 37.5 µL of SQB sequencing buffer and 25.5 µl of loading beads in the final volume of 75 µl and loaded on a MinION Spot-ON Flow Cell version R9.4 (Oxford Nanopore). Two flow cells were run on MinION device controlled by MinKNOW software version 3.1.13 for 48 hours each with local basecalling turned off generating 9.7 and 5.1 GB of sequence data. Post run basecalling was performed with Guppy Basecalling Software, version 3.4.5 (Oxford Nanopore). After filtering low quality reads (Q<7), the combined output of the two runs was 13.69 GB and 4.7 million reads.

Genome assembly and contamination filtering

Read quality was assessed using FastQC v0.11.8 (Andrews 2010). Illumina adapters, low-quality nucleotide bases (phred score below 15) from the 3' and 5' ends and reads shorter than 50 bp were removed using cutadapt v1.18 (Martin 2011). From the filtered reads, *in silico* genome size estimates were calculated using *k-mer* based tools kmergenie v.1.7048 (Chikhi and Medvedev 2014), GenomeScope v1.0 (Vurture et al. 2017), and findGSE v0.1.0 R package (Sun et al. 2018). The latter two required a *k-mer* histogram computed by jellyfish v2.2.10 (Marçais and Kingsford 2011) with *k-mer* size of 21. The genome was assembled using 4.7 million long reads from two MinION runs (60x coverage) using Canu v1.8 with default settings and setting the expected genome size to 200 Mb (Koren et al. 2017). To improve assembly quality, paired

829 end Illumina reads were mapped to the genome with BWA aligner (Li and Durbin 2009) using
830 BWA-MEM algorithm and polished with Pilon v. 1.23 with '—changes' and '--fix all' options
831 (Walker et al. 2014). Assembled contigs identified as bacterial and fungal contaminants based on
832 divergent GC content from most *Archegozetes* contigs, high coverage and blast homology to the
833 nt database (downloaded February 2019, Evalue $1e^{-25}$) were removed using Blobtools v1.0
834 (Laetsch and Blaxter 2017).

835 Identification, classification and masking of repetitive element

836 Repetitive elements in the genome *Archegozetes* were identified using a species-specific
837 library generated with RepeatModeler v 1.0.11 (Bao et al. 2015; Smit and Hubley 2008) and
838 MITE tracker (Crescente et al. 2018) and annotated by RepeatClassifier, a utility of the
839 RepeatModeler software that uses the RepBase database (version Dfam_Consensus-20181026).
840 Unclassified repeat families from both programs were run through CENSOR v 4.2.29 (Kohany et
841 al. 2006) executable *censor.ncbi* against the invertebrate library v 19.03 to provide further
842 annotation. Predicted repeats were removed if they had significant blast homology (E-value $1e^{-5}$)
843 to genuine proteins in the NCBI nr database and/or a local database of arthropod genomes
844 (*Drosophila melanogaster, Tribolium castaneum, Tetranychus urticae, Leptotrombidium*
845 *deliense, Dinothrombium tinctorium, Sarcoptes scabiei, Euroglyphus maynei, Galendromus*
846 *occidentalis, Dermatophagoides pteronyssinus*). Unclassified repeats with blast homology to
847 known TEs were retained whereas those with no blast homology were removed (Petersen et al.
848 2019). The remaining repeat families were combined with the Arthropoda sequences in RepBase
849 and clustered using vsearch v 2.7.1 (--iddef 1 --id 0.8 --strand both; (Rognes et al. 2016)). The
850 filtered repeat library was used to soft mask the *A. longisetosus* using RepeatMasker v 4.07 (Smit
851 et al. 1996-2010).A summary of the masked repeat content was generated using the
852 "buildSummary.pl" script, the Kimura sequence divergence calculated using the
853 "calcDivergenceFromAlign.pl" script and the repeat landscape visualized using the
854 "createRepeatLandscape.pl" script, all utilities of RepeatMasker.

855 Gene prediction and annotation

856 Both *ab inito* and reference-based tools were used for gene prediction using modified
857 steps of the funannotate pipeline (Palmer and Stajich 2017). The *ab inito* tool GeneMark-ES
858 v4.33 (Ter-Hovhannisyan et al. 2008) was used along with reference based tools BRAKER

v2.1.2 (Bruna et al. 2020) using RNAseq reads discussed below and PASA v 2.3.3 (Haas et al. 2008) using genome-guided transcriptome assembly from Trinity described below. Lastly, *Tetranychus urticae* gene models from the NCBI database (GCF_000239435.1) were aligned to the contigs using GeMoMa (Keilwagen et al. 2019). All gene predictions were combined in EvidenceModeler (Haas et al. 2008) with the following weights: GeMoMa =1, PASA = 10, other BRAKER = 1, and GeneMark = 1. Predicted tRNAs using tRNAscan-SE v 2.0.3 (Chan and Lowe 2019) were combined with the gene predictions in the final gene feature format (GFF) file and filtered for overlap using bedtools (Quinlan and Hall 2010) *intersect* tool (Quinlan and Hall 2010).

The predicted genes were searched against the NCBI nr (February 2019) (Pruitt et al. 2005), SwissProt (February 2019) (Bairoch and Apweiler 2000), a custom-made Chelicerata database including genomes of *Tetranychus urticae, Leptotrombidium deliense, Dinothrombium tinctorium, Sarcoptes scabiei, Euroglyphus maynei, Galendromus occidentalis, Metaseiulus occidentalis, Dermatophagoides pteronyssinus, Trichonephila clavipes, Stegodyphus mimosarum, Centruroides sculpturatus, Ixodes scapularis* and *Parasteatoda tepidariorum* (all downloaded Feb 2019), PFAM (v 32, August 2018) (Bateman et al. 2004), merops (v 12, October 2017) (Rawlings et al. 2010) and CAZY (v 7, August 2018) (Cantarel et al. 2009) databases. The results of the hmm-based (Eddy 2011) PFAM and CAZY searches were filtered using cath-tools v 0.16.2 (https://cath-tools.readthedocs.io/en/; E-value $1e^{-5}$) and the blast-based searches were filtered by the top hit (E-value $1e^{-5}$ threshold). Predicted genes were also assigned to orthologous groups using eggNOG-mapper (Huerta-Cepas et al. 2017). Gene annotation was prioritized by the SwissProt hit if the E-value < $1e^{-10}$ followed by NCBI annotation, the custom Chelicerata database and if no homology was recovered, then the gene was annotated as, "hypothetical protein". Final annotation was added to the GFF file using GAG (Geib et al. 2018).

Analysis of the official gene set (OGS)

To allow the OGS to be used as resources for functional studies, we assigned functional categories based on Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Consortium 2004; Kanehisa and Goto 2000). GO terms for the respective genes models of the OGS were assigned based on the gene id with highest homology from the SwissProt database or NCBI nr database (Bairoch and Apweiler 2000; Pruitt et al. 2005). A

889 custom database of GO terms was created with makeOrgPackage function in the R package
890 AnnotationForge v1.26.0 (Carlson and Pagès 2019). Over-representation analysis of GO terms
891 was tested using the enrichGO function in the R package clusterProfiler v3.12.0 (Yu et al. 2012)
892 with a hypergeometric distribution and a Fisher's Exact test. P-values were adjusted for multiple
893 comparisons using false discovery rate correction (Benjamini and Hochberg 1995). Resulting
894 enriched GO terms were processed with GO slim (Consortium 2019) and the final list of over
895 represented GO terms was used to plot the number of genes in a respective category.

896 KEGG orthology terms were assigned from single-directional best hit BLAST searches
897 of each gene model on the KEGG Automatic Annotation Server (Moriya et al. 2007).
898 Additionally, we ran GhostKOALA (Kanehisa et al. 2016) (GHOSTX searches for KEGG
899 Orthology And Links Annotation) to obtain KEGG orthology terms. Compared to conventional
900 BLAST searches, GhostKOALA is about 100 times more efficient than BLAST to remote
901 homologs by using suffix arrays (Suzuki et al. 2014).

902 Orthology and phylogenomic analyses

903 Orthologs of *A. longisetosus*, other species within Acari, Chelicerata and the fruit fly
904 *Drosophila* were identified using OrthoFinder v 2.3.3 (-M msa –A mafft –T fasttree; (Emms and
905 Kelly 2015a)). Prior to running OrthoFinder, isoform variants were removed from the gene
906 predictions using CD-Hit (Fu et al. 2012). Trees of orthogroups with at least 80% of taxa present
907 (n= 4,553) were constructed using fasttree v 2.1.10 (Price et al. 2010), trimmed with TrimAl v
908 1.4.1 (-keepheader -fasta –gappyout; (Capella-Gutiérrez et al. 2009)) and paralogs pruned using
909 phylotreepruner v 1.0 (min_number_of_taxa =18, bootstrap_cutoff= 0.7, longest sequence for a
910 given orthogroup=u; (Kocot et al. 2013)). Alignments shorter than 100 amino acids were
911 removed, leaving 1,121 orthogroups.

912 For the maximum likelihood analysis, the trimmed and pruned alignments were
913 concatenated into a supermatrix using FasConCat v1.04 (Kück and Meusemann 2010) composed
914 of 377,532 amino acids and the best substitution models determined using PartitionFinder v 2.1.1
915 (Lanfear et al. 2016). The maximum likelihood consensus phylogeny from the supermatrix and
916 partition scheme was constructed using IQ-tree and 1,000 ultrafast bootstrap replicates (Nguyen
917 et al. 2015). For the coalescence species tree reconstruction, gene trees were generated using IQ-
918 tree v 1.6.12 on the trimmed alignments of the 1,121 filtered orthogroups and processed using

ASTRAL v 5.6.3 (Zhang et al. 2018). Branch lengths are presented in coalescent units (differences in the 1,121 gene trees) and the node values reflect the local posterior probabilities.

RNA sequencing and transcriptome assembly

For RNA extraction, about 200 mites of all life stages were taken from stock culture and subsequently washed with 1% SDS for 10 s. RNA was extracted from living specimens using the Quick-RNA MiniPrep Kit (Zymo Research) according to the manufacturer's protocol. Quantity and quality of RNA were accessed using a Qubit fluorometer and NanoDrop One (Thermo Fisher Scientific), respectively.

Extracted RNA was shipped to Omega Bioservices (Norcross, GA, USA) on dry ice for library preparation and sequencing. Whole animal RNA was used for poly-A selection, cDNA synthesis and library preparation following the Illumina TruSeq mRNA Stranded Kit protocol. The library was sequenced with 100 million 150 bp paired-end on a HighSeq4000 platform. For the genome-guided assembly of the transcriptome a bam-file was created from the genome using STAR (Dobin et al. 2013). RNAseq reads were *in silico* normalized and subsequently used together with the bam-file to assemble the transcripts using Trinity v2.8.4 (Grabherr et al. 2011; Haas et al. 2013), yielding an assembly with a total length of 162.8 Mbp, an N50= 2994 bp and a BUSCO score (Simão et al. 2015) of C:96.3% [S:36.5%,D:59.8%], F:1.3%, M:2.4%.

Life-stage specific RNAseq

For life-stage specific RNAseq, we collected 15 specimens per life stage from the stock culture that were split into three replicates of five individuals. Whole animals (for all stages but eggs) were flash frozen in 50 µl TRIzol using a mixture of dry ice and ethanol (100%) and stored at −80°. RNA was extracted using a combination of the TRIzol RNA isolation protocol (Life Technologies) and RNeasy Mini Kit (Qiagen) (Kitchen et al. 2015). The TRIzol protocol was used for initial steps up to and including the chloroform extraction. Following tissue homogenization, an additional centrifugation step was performed at 12,000 × g for 10 min to remove tissue debris. After the chloroform extraction, the aqueous layer was combined with an equal volume of ethanol and the RNeasy Mini Kit was used to perform washes following the manufacturer's protocol. Eggs were crushed using pipette tips and directly stored in a mixture of cell lysis buffer and murine RNase Inhibitor (New England Biolab).

We used the NEBNext® Single Cell/Low Input RNA Library Prep Kit for Illumina® together with NEBNext® Multiplex Oligos for Illumina® (New England Biolab) for library preparation, including reverse transcription of poly(A) RNA, amplification full-length cDNA, fragmentation, ligation and final library amplification according to the manufacturer's protocol. We performed cDNA amplification for 16 (18 for egg samples) PCR cycles and final library amplification 8 PCR cycles. In total, we constructed 18 libraries (three for each life stage). The quality and concentration of the resulting libraries were assessed using the Qubit High Sensitivity dsDNA kit (Thermo Scientific) and Agilent Bioanalyzer High Sensitivity DNA assay. Libraries were sequenced on an Illumina HiSeq2500 platform (single-end with read lengths of 50 bp) with ~18 million reads per library.

Illumina sequencing reads were pseudoaligned to the bulk transcriptome and quantified (100 bootstrap samples) with kallisto 0.46.0 (Bray et al. 2016) using default options for single-end reads. Fragment length sizes were extracted from the Agilent Bioanalyzer runs. For life-stage specific differential expression analysis, kallisto quantified RNAseq data was processes with sleuth 0.30.0 (Pimentel et al. 2017) using Likelihood Ratio tests in R 3.6.1 (R_Core_Team 2019). The average transcripts per million (tpm) values for each target transcript were extracted from the sleuth object (see R script) and used with the Heatmapper tool (Babicki et al. 2016) to produce an unclustered heatmap showing relative expression levels. UpSetR (Conway et al. 2017) was used to compare the number of unique and shared expressed genes across life stages.

Horizontal gene transfer events identification

To detect HGTs, we used the published tool "./Lateral_gene_transfer_predictor.py" (Thorpe et al. 2018) to calculate the Alien Index described by (Gladyshev et al. 2008) and (Flot et al. 2013). All predicted genes were compared to the NCBI nr database as previously described (Thorpe et al. 2018). Results to Arthropoda (tax id 6656) were ignored in the downstream calculations. The HGT candidates were filtered for contamination identified by both Blobtools (Laetsch and Blaxter 2017) and the Alien Index (AI > 30 and >70% percent identity to a non-metazon sequence). The candidates were further filtered for > 50% overlap with predicted repeats using the bedtools intersect tool with the RepeatMasker gff file and expression from any developmental stage. Introns were scored manually from visualization in IGV genome browser

977 (Robinson et al. 2011) and GC content for all predicted genes was calculated using the bedtools
978 nuc tool.

979 Analysis of chemosensory and photoreceptor gene families

980 The search and analysis chemosensory genes largely followed the procedure outlined by
981 Dong et al. (Dong et al. 2018) with slight modifications. First, the *Archegozetes* official gene set
982 (OGS) was searched using BLASTP (E-value, $<1 \times 10^{-3}$) against the following queries for the
983 different chemosensory gene families. The OGS was queried against i) *D. melanogaster*, *D.*
984 *mojavensis*, *Anopheles gambiae*, *Bombyx mori*, *T. castaneum*, *Apis mellifera*, *Pediculus humanus*
985 *humanus*, and *Acyrthosiphon pisum* odorant binding proteins (OBPs) (Vieira and Rozas 2011);
986 ii) *D. melanogaster*, *D. mojavensis*, *A. gambiae*, *B. mori*, *T. castaneum*, *A. mellifera*, *P.*
987 *humanus humanus*, *A. pisum, I. scapularis*, and *Daphnia pulex* small chemosensory proteins
988 (CSP) (Niimura and Nei 2005; Robertson and Wanner 2006; Vieira and Rozas 2011); iii) *D.*
989 *melanogaster* and *A. mellifera* odorant receptors (Niimura and Nei 2005; Robertson and Wanner
990 2006); iv) *D. melanogaster*, *A. mellifera*, *I. scapularis*, *T. urticae*, *T. mercedesae*, and *M.*
991 *occidentalis* gustatory receptors (GRs) (Dong et al. 2017; Gulia-Nuss et al. 2016; Hoy et al.
992 2016; Ngoc et al. 2016; Robertson and Wanner 2006; Robertson et al. 2003); v) a comprehensive
993 list of iGluRs and IRs across vertebrates and invertebrates (Croset et al. 2010), as well as those
994 identified in the *T. mercedesae, D. tinctorium* and *L. deliense* genome projects (Dong et al. 2017;
995 Dong et al. 2018). Second, all candidate *Archegozetes* sequences were reciprocally blasted
996 (BLASTP, E-value $<1 \times 10^{-3}$) against the NCBI database (Pruitt et al. 2005) and all sequences
997 that did not hit one of the respective receptors or transmembrane proteins were removed from the
998 list. Third, for phylogenetic analysis of IRs and GRs from *Archegozetes* were aligned with IRs
999 from *D. melanogaster*, *T. urticae, D. tinctorium* and *L. deliense* and GRs from iv) *D.*
1000 *melanogaster, T. mercedesae*, *I. scapularis*, and *M. occidentalis*, respectively, using MAFFT (v
1001 7.012b) with default settings (Katoh and Standley 2013). Poorly aligned and variable terminal
1002 regions, as well as several internal regions of highly variable sequences were excluded from the
1003 phylogenetic analysis. Fourth, maximum likelihood trees were constructed with the IQ-TREE
1004 pipeline (v 1.6.12) with automated model selection using 1,000 ultrafast bootstrap runs (Nguyen
1005 et al. 2015).

Reference opsin genes and opsin-like sequences were obtained from Dong et al. (Dong et al. 2018) and used to query the *Archegozetes* OGS using BLASTP (E-value, $<1 \times 10^{-5}$). Subsequently, candidates sequenced were reciprocally blasted against NCBI using the same settings and only retained if they hit an opsin or opsin-like gene. The *Archegozetes* candidates were aligned with the query sequence list using MAFFT (v 7.012b) with default settings (Katoh and Standley 2013). This opsin gene alignment phylogenetically analyzed using the IQ-TREE pipeline (v 1.6.12) with automated model selection and 1,000 ultrafast bootstrap runs (Nguyen et al. 2015).

Gene family phylogenies

We used the following workflow to analyses genes related to Figure 5 (hox and developmental genes), Figure 7 (cell wall-degrading enzyme encoding genes) and Figure 8 (alcohol and geraniol dehydrogenases genes). Generally, protein orthologs were retrieved from NCBI (Pruitt et al. 2005), and aligned using MUSCLE (Edgar 2004) or MAFFT (v 7.012b) (Katoh and Standley 2013) and ends were manually inspected and trimmed. The resulting final protein sequence alignments used to construct a maximum likelihood (ML) phylogenetic tree with either i) PhyML with Smart Model Selection (Guindon et al. 2010; Lefort et al. 2017) or ii) the IQ-TREE pipeline with automated model selection (Nguyen et al. 2015). The ML trees were constructed using either 1,000 ultrafast bootstrap runs (IQ-TREE) or approximate-likelihood ratio test (PhyML) was used to assess node support.

Feeding experiments with labelled precursors and chemical analysis (GC/MS)

Stable isotope incorporation experiments were carried out as previously described (Brückner et al. 2020). Briefly, mites were fed with wheat grass containing a 10% (w/w) mixture of three antibiotics (amoxicillin, streptomycin and tetracycline) and additionally, we added 25% (w/w) of the stable isotope-labelled precursors [$^{13}C_6$] D-glucose (Cambridge Isotope Laboratories, Inc.) as well as a control with untreated wheat grass. Cultures were maintained for one generation and glands of adult specimens were extracted one week after eclosion by submersing groups of 15 individuals in 50 µl hexane for 5 min, which is a well-established method to obtain oil gland compounds from mites (Brückner and Heethoff 2016; 2017; Brückner et al. 2017b; Raspotnig et al. 2008).

Crude hexane extracts (2-5 µl) were analysed with a GCMS-QP2020 gas chromatography – mass spectrometry (GCMS) system from Shimadzu equipped with a ZB-5MS capillary column (0.25 mm x 30m, 0.25 µm film thickness) from Phenomenex. Helium was used a carrier gas with a flow rate of 2.14 ml/min, with splitless injection and a temperature ramp was set to increase from 50°C (5 min) to 210°C at a rate of 6°C/min, followed by 35°C/min up to 320°C (for 5 min). Electron ionization mass spectra were recorded at 70 eV and characteristic fragment ions were monitored in single ion mode.

Acknowledgment

We thank Joe Parker for making his laboratory space and resources available to us. Michael Heethoff, Sebastian Schemlzle, Benjamin Weiss and Martin Kaltenpoth graciously allowed us to use some of their unpublished images. Roy A. Norton provided invaluable comments to the manuscript and collected the first specimens of *Archegozetes longisetosus* giving rise to the current laboratory strain. This work was supported by a grant from the Caltech Center for Environmental Microbial Interactions (CEMI) to AB. AB is Simons Fellow of the Life Sciences Research Foundation (LSRF).

Ethics statement

There are no legal restrictions on working with mites.

Authors contributions

AB had the initial idea for the study; AB, AAB and SAK design research; IAA performed long-read sequencing and assembled the genome; AB performed all other experimental work; AAB analyzed hox and life-stage specific expression data; AB analyzed chemical data; SAK and AB performed bioinformatic analyses; AB wrote the first draft of the manuscript with input from AAB and SAK; SAK revised the manuscript. All authors gave final approval for publication.

Data availability

Genomic and transcriptomic data generated for his project can be found on NCBI under the accession numbers PRJNA683935 and PRJNA683999. All other data related to this

manuscript can be found at https://doi.org/10.22002/D1.1876 under a cc-by-nc-4.0 license (Brückner 2021).

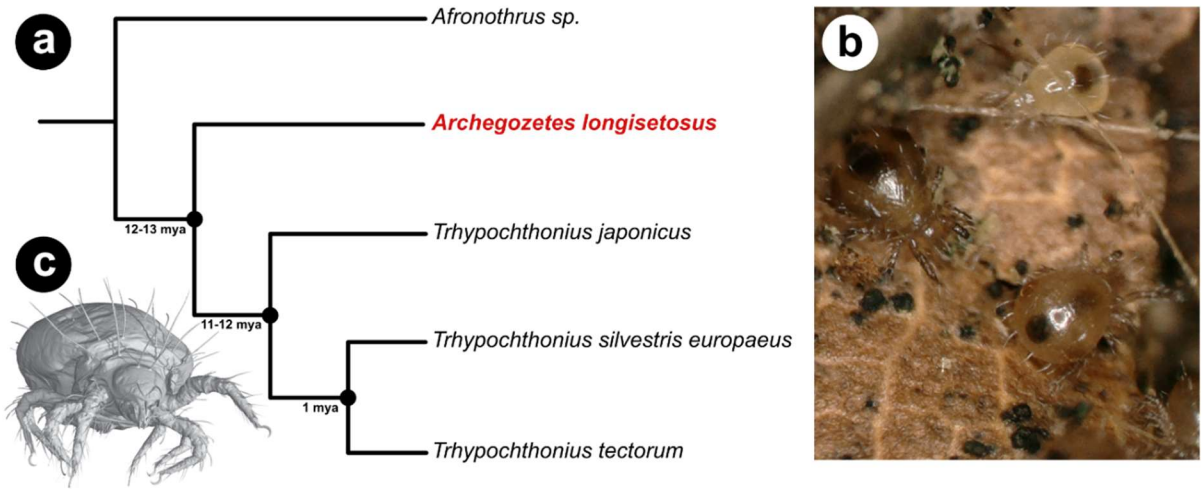

Figure 1 The mite, *Archegozetes longisetosus*, in its phylogenetic and natural environment. **a**: Species tree of selected oribatid mites of the family Trhypochthoniidae based on phylogenetic analyses and divergence time estimates by (Heethoff et al. 2011b). **b**: Two adults and one tritonymph of *Archegozetes* on a piece of leaf litter. The algae growing on the leaf serves as a food source for the mites. **c**: Habitus of an adult mite based on a surface rendering of a µCT-scan reconstruction. Image courtesy of Sebastian Schmelzle.

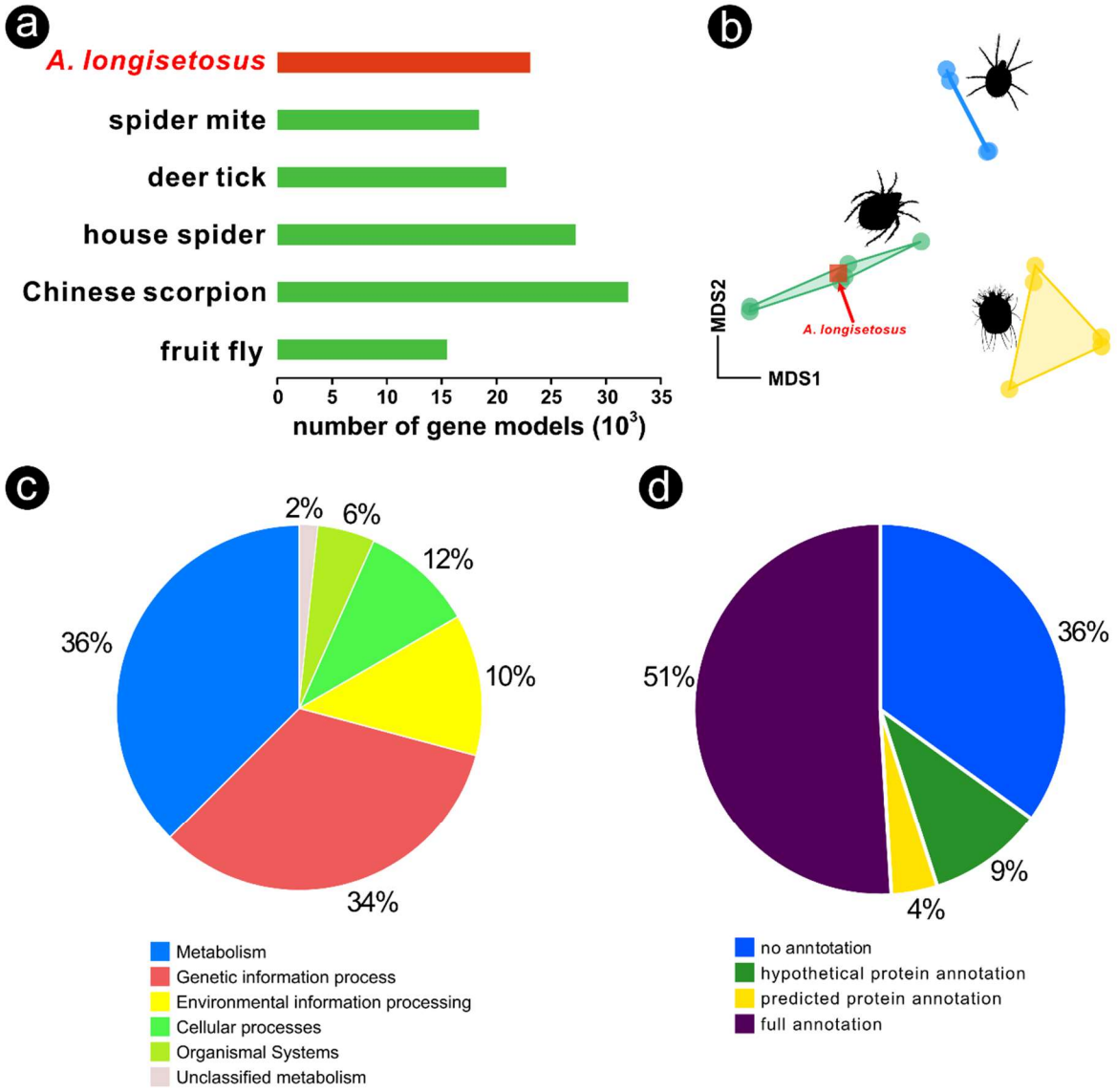

Figure 2 Comparisons and annotations of the official gene set (OGS) of *Archegozetes longisetosus*. **a**: Number of gene models of the mites compared to other mites, chelicerates and the fruit fly (Cao et al. 2013; dos Santos et al. 2015; Grbić et al. 2011; Gulia-Nuss et al. 2016; Schwager et al. 2017). **b**: Non-linear multidimensional scaling plot (NMDS) of clustered orthogroups based on the OGS or predicted proteins of several mite species. *Archegozetes longisetosus* is marked a square, nested within Oribatida. Prostigmata are depicted in blue, Astigmata in yellow and Oribatida in green. **c**: Pie chart showing the percentage composition of genes of the *Archegozetes* annotated to different broad biological categories by GhostKOALA. **d**: Pie chart describing the overall annotation of the OGS of the mite.

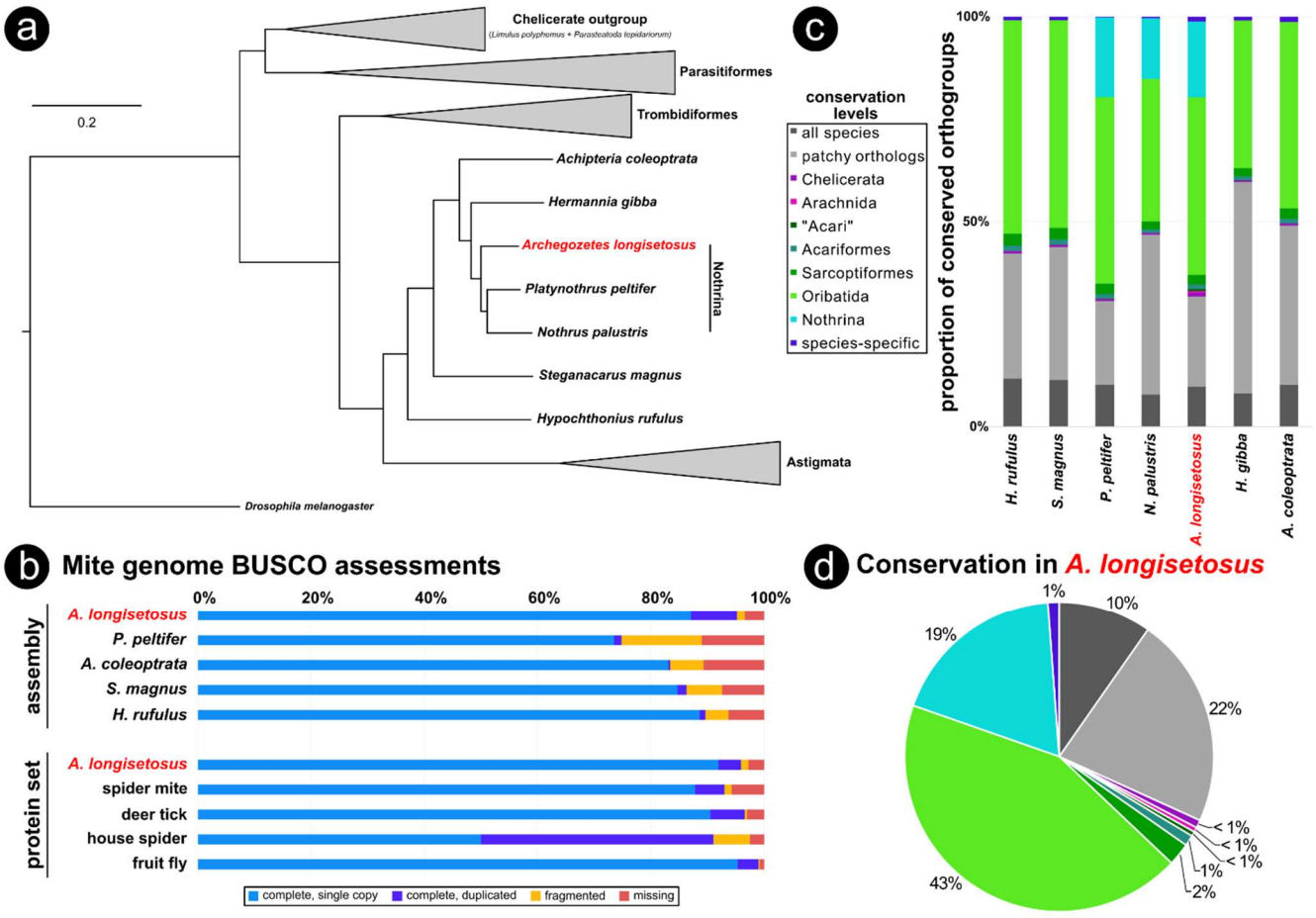

Figure 3 Orthology comparison and phylogenetic placement of *Archegozetes longisetosus* among other chelicerates. **a**: Maximum likelihood phylogeny based on concatenation of 1,121 orthologs showing the mites phylogenetic environment within the Oribatida (all nodes have 100% support; branch length unit is substitutions per site). For the fully expanded tree see supplementary Figure S2. **b**: BUSCO-assessment of the *Archegozetes* genome assembly and protein set for both ortholog presence and copy number compared to other oribatid mites and selected model species, respectively. **c**: Comparisons of protein-coding genes in seven oribatid mite species (for a full comparison of all species see supplementary Figure S3) with *Archegozetes* highlighted in red. The bar charts show the proportion of orthrogroup conservation with each species (see insert legend) based on OrthoFinder clustering. **d**: Detailed pie chart depicting the conservation levels of orthogroup in *Archegozetes*.

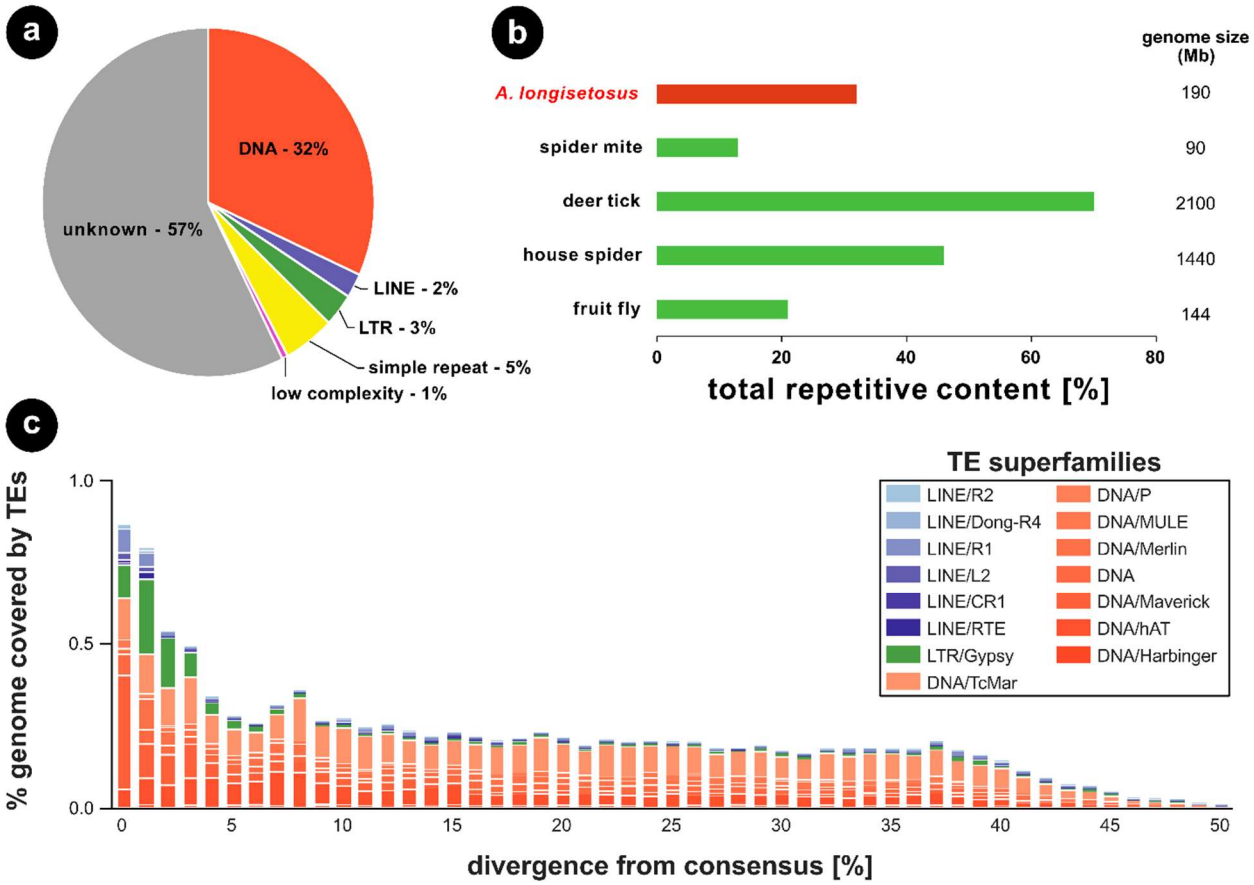

Figure 4 Comparison of repeat content estimations and transposable element (TE) landscape of *Archegozetes longisetosus*. **a:** Repetitive element categories of *Archegozetes* based on the results from RepeatModeler and MITE Tracker. LINE= long interspersed nuclear element, LTR= long terminal repeat. **b:** Comparison of total repetitive content among *Archegozetes*, other model chelicerates and the fly. All values are from the respective genome paper of the species, except for the fly. **c:** Repeat divergence plot showing TE activity through time for the major TE superfamilies of *Archegozetes.* Transposable elements with a low divergence from the consensus were recently active, while TEs diverging from the consensus depicted older activities (x-axis).

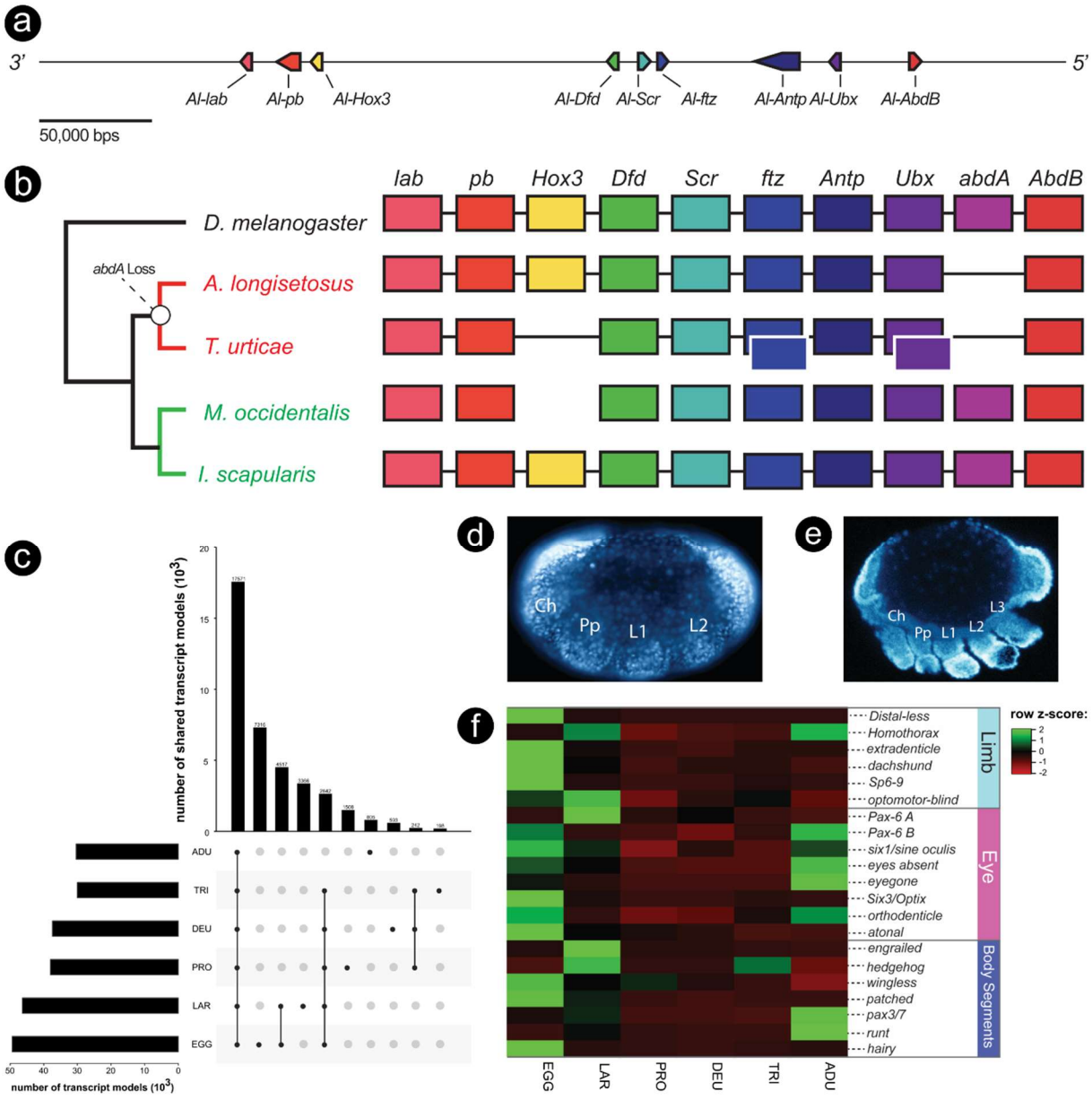

Figure 5 The genomic organization of the Hox genes and life-stage specific expression patters of developmental genes in *Archegozetes longisetosus*. **a:** Schematic of the genomic region enclosing the *Archegozetes* Hox cluster. The genomic organization of the Hox cluster is collinear, as it is in many arthropod taxa, however an abdominal-A ortholog is absent. Arrowed boxes denote the direction of transcription. The scale bar represents 50,000 base pairs. **b:** A comparison of the Hox cluster organization of reported members of Acari with the fruit fly Drosophila melanogaster as the outgroup. The last common ancestor of the parasitiform mites *M. occidentalis* and *I. scapularis* likely had an intact Hox cluster (green branches and labels),

whereas abdominal-A was likely lost in the last common ancestor of acariform mites, as represented by *Archegozetes* and *T. urticae* (red branches and labels). Boxes with white borders represent duplicated Hox genes. Lines through the boxes indicate an intact Hox cluster. See text for further details. **c:** Number of transcripts shared across the different life stages of *Archegozetes*. The barplot panel on the left shows the numbers of transcripts in each stage. Exemplars of (**d**) early and (**e**) mid- germ-band embryos. Ch= chelicera; L1-3= walking legs 1-3; Pp= pedipalp. Embryos are stained with the nuclear dye DAPI and oriented with the anterior to the left of the page. **f:** Non-clustered heatmap showing the relative expression (row z-score based on tpm) patterns of putative limb, eye, and body segmentation genes throughout the embryonic, larval instars, and adult stages of *Archegozetes*. See **supplementary Table S3** for average tpm values. Life stages (for **c** and **f**): EGG= egg; LAR= larva; PRO= protonymph; DEU= deutonymph; TRI= tritonymph; ADU= adult.

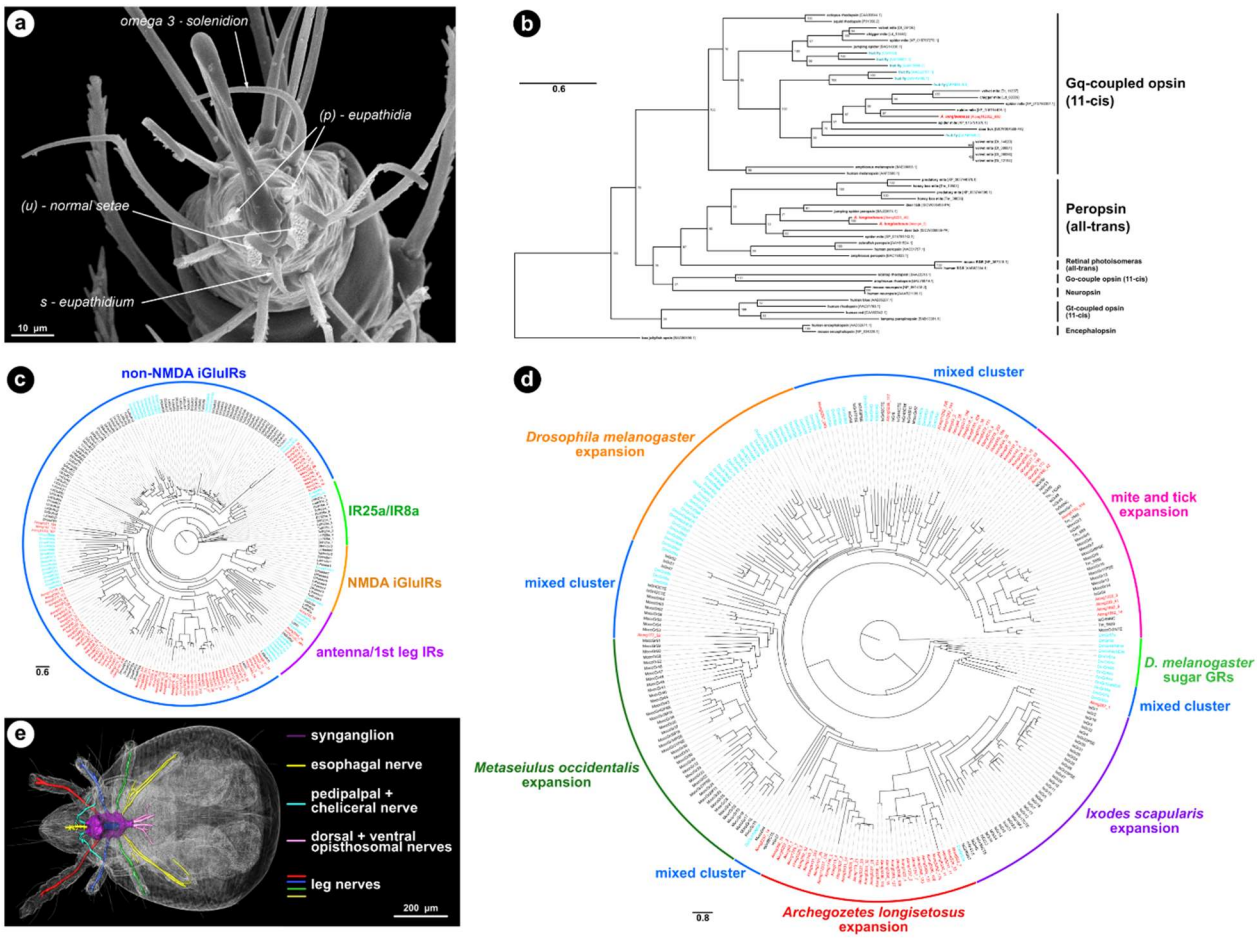

Figure 6 The sensory systems of *Archegozetes longisetosus* and phylogenetic analysis of selected photoreceptor and chemosensory genes. **a**: Scanning electron micrograph (SEM) showing the end of tarsus on *Archegozetes*' first leg. Images shows normal setae, but also modified chemosensory setae, namely eupathidia, both paired (p) and single (s), as well as an omega-3 solenidium. SEM picture courtesy of Michael Heethoff. **b**: Phylogeny and classification of opsin genes across the Metazoa, including those of several Chelicerata. The tree was constructed using a maximum likelihood approach and rooted with a jelly fish opsin. *Archegozetes* sequences are depicted in red, Drosophila in turquoise; branch length unit is substitutions per site. **c**: Maximum likelihood phylogeny of ionotropic receptors and ionotropic glutamate receptors of *Archegozetes* (Along), *Dinothrombium* (Dt), *Leptothrombidium* (Ld), *Tetranychus* (Tu) and *Drosophila* (Dmel). IR25a/IR8a and antenna/1st leg IRs contain genes with known chemosensory function in *Drosophila*. The tree was rooted to the middle point; *Archegozetes* sequences are depicted in red, *Drosophila* in turquoise; branch length unit is

substitutions per site. Bootstrap values can be found in the supplementary Figure S13. **d**: Maximum likelihood phylogenetic tree of gustatory receptors of *Archegozetes* (Along), *Ixodes* (Is), *Tropilaelaps* (Tm), *Metaseiulus* (Mocc) and *Drosophila* (Dmel). The tree was rooted to the middle point; *Archegozetes* sequences are depicted in red, *Drosophila* in turquoise; branch length unit is substitutions per site. Bootstrap values can be found in the supplementary Figure S14. **e**: Combined image of volume rendering (grey) and reconstructed nervous system of *Archegozetes* in dorsal view. Color-code corresponds to different parts of the nervous system, as depicted in the legend. The blue structure in the middle of the synganglion is the part of the esophagus which penetrates the synganglion. Scale bar: 200 µm. Image courtesy of Sebastian Schmelzle based on data in (Hartmann et al. 2016).

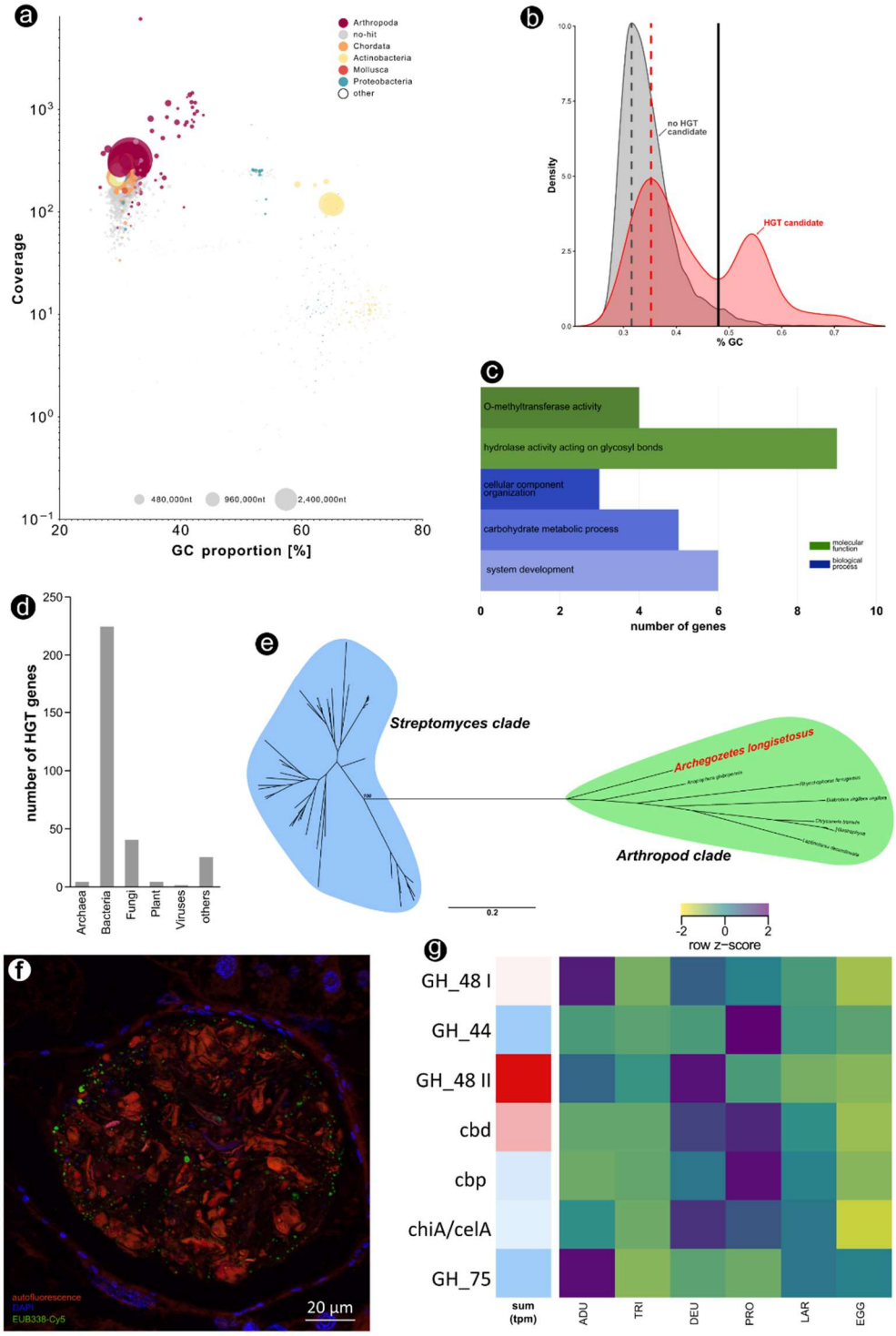

Figure 7 Horizontal gene transfer (HGT) and implications for the feeding biology of *Archegozetes longisetosus*. **a:** Blob-plot of the long-read genome assembly contigs plotting the read coverage against GC proportion [%]. Contigs are colored according to the taxonomic order

of their best Megablast hit to the NCBI nucleotide database. Size of circle corresponds to the nucleotides per contigs. **b:** Comparison of the GC content of HGT and non-HGT genes. HGT genes shifted towards the host genome GC content indicate integration within the host genome while the higher GC content HGT genes might be the product of relatively recent HGT events. **c:** Enrichment of functional categories (GO terms) describing the molecular functions and biological processes related to the HGT candidate genes. **d:** Taxonomic origin of HGT. The category "others" includes mostly protozoan donor genes among other Eukaryotes. **e:** Unrooted maximum-likelihood tree of glycoside hydrolase family 48 members (GH_48) from Streptomyces bacteria and HGT genes from other arthropods as well as *Archegozetes* (GH_48 II). Bootstrap values and the full tree can be found in the supplementary Figure S15. The scale bar denotes substitutions per site. **f:** Fluorescence *in situ* hybridization (FISH) micrograph of a food bolus in the mites' alimentary tract. The food material (wheat grass power) is enclosed in a peritrophic membrane and there is a high bacterial prevalence in the food bolus. Image courtesy of Benjamin Weiss and Martin Kaltenpoth. **g:** RNAseq support of HGT candidates related to cell wall degrading enzymes. The first block (single column) shows the overall RNA expression (tpm) of the HGT in all life stages; red denotes high total expression, while blue depicts low total expression. The second block (six columns) shows the expression (row z-score based on tpm) of the same HGT candidates across the different life stages of *Archegozetes.* Abbreviations: GH_48= glycoside hydrolase family 48, GH_44= glycoside hydrolase family 44, cbd= cellulose-binding domain, cbp= cellulose-binding protein, chiA/celA = chitinase/cellulase, GH_75= glycoside hydrolase family 75.

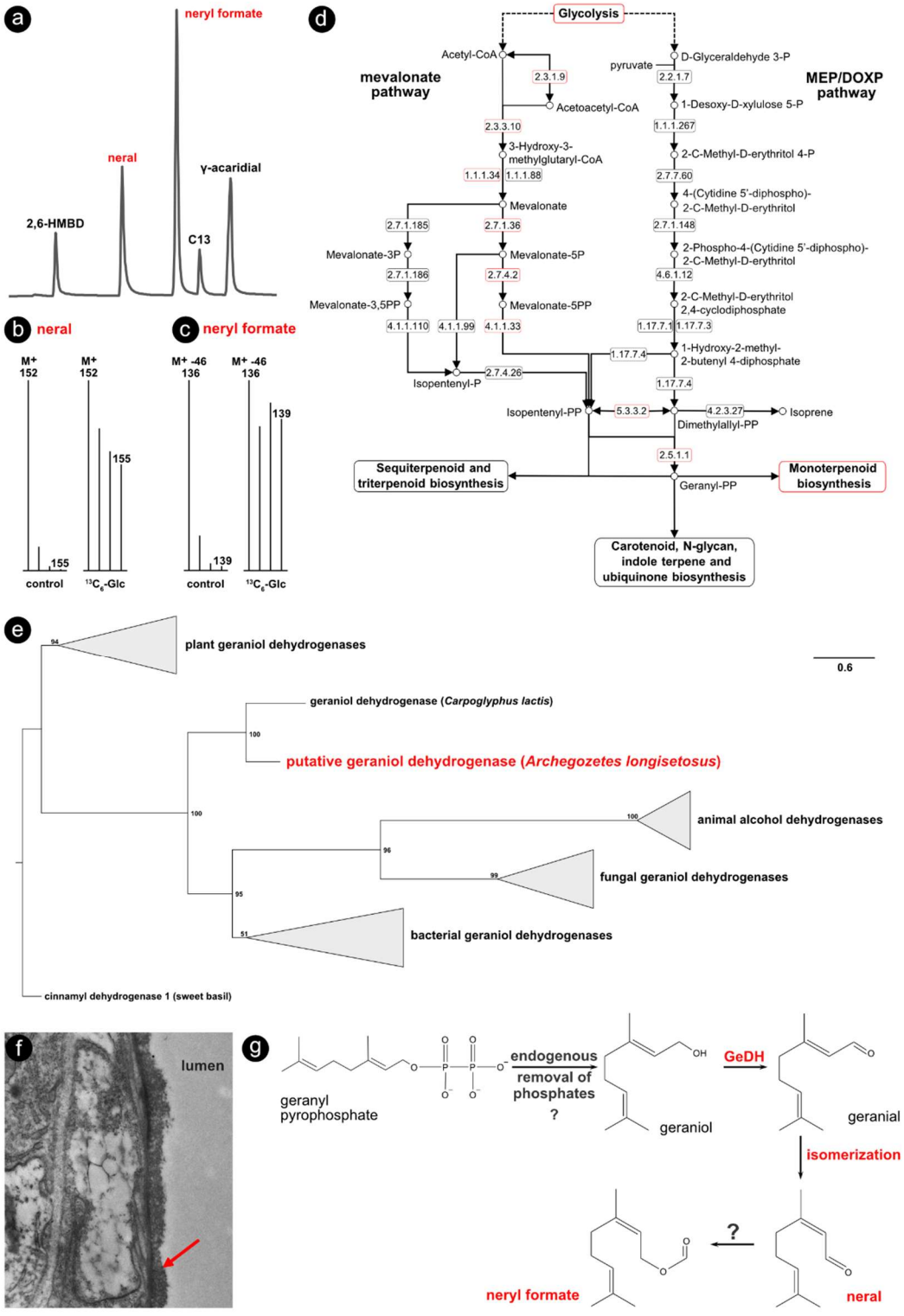

Figure 8 Reconstruction of the biosynthetic pathway leading to monoterpenes in *Archegozetes longisetosus*. **a**: Representative gas chromatogram of the mite' gland content; in order of retention time: 2-hydroxy-6-methyl-benzaldehyde (2,6-HMBD), neral ((*Z*)-3,7-dimethylocta-2,6-

dienal) neryl formate ((*Z*)-3,7-dimethyl-2,6-octadienyl formate), tridecane, 3-hydroxybenzene-1,2-dicarbaldehyde (γ-acaridial). Further alkanes/alkenes (pentadec-7-ene, pentadecane, heptadeca-6,9-diene, heptadec-8-ene, heptadecane) are not shown. Monoterpenes are marked in red. **b and c**: Representative mass spectra of neral (**b**) and neryl formate (**c**) extracted from defensive glands of mites fed with unlabeled wheatgrass powder (control), or wheatgrass infused with $^{13}C_6$-labelled glucose recorded in single-ion mode. The mass spectra for neral (**b**) shows the M+-ion series, while the spectra for neryl formate (**c**) show the diagnostic ion series at $[M-46]^+$. Mites fed with the $^{13}C_6$ glucose infused wheatgrass showed enriched ions. **d:** KEGG reference pathway map for terpenoid backbone biosynthesis. Mapping genes from the *Archegozetes* genome encoding for pathway enzymes (labeled in red) revealed that the mite can produce geranyl pyrophosphate (GPP) *via* the mevalonate pathway from precursors provided by glycolysis. Enzymes names correspond to EC numbers: 2.3.1.9= acetyl-CoA C-acetyltransferase; 2.3.3.10= hydroxymethylglutaryl-CoA synthase; 1.1.1.34= hydroxymethylglutaryl-CoA reductase; 2.7.1.36= mevalonate kinase; 2.7.4.2= phosphomevalonate kinase; 4.1.1.33= diphosphomevalonate decarboxylase; 5.3.3.2= isopentenyl-diphosphate delta-isomerase; 2.5.1.1= farnesyl diphosphate synthase. **e:** Maximum-likelihood tree based on an alignment of plant, fungal and bacterial geraniol dehydrogenases, animal alcohol dehydrogenase and two mite (*Carpoglyphus lactis* and *Archegozetes*) geraniol dehydrogenases (GeDH). Bootstrap values (based on 1000 replicates) are indicated along branches and the scale bar denotes substitutions per site. The tree was rooted by the outgroup cinnamyl dehydrogenase from sweet basil. **f:** Ultrastructure of the gland-tissue of *Archegozetes*, as observed by transmission electron microscopy (TEM). Red error shows the border between the gland cell and the glandular lumen. TEM picture courtesy of Michael Heethoff. **g**: Proposed biochemical pathway scenario leading to neral and neryl formate in *Archegozetes* starting with GGP from the terpenoid backbone biosynthesis.

1207 **Table 1** *Archegozetes longisetosus* genome metrics

1208

Feature	**Value**
Estimated genome size	135-180 Mb
Assembly size	190 Mb
Coverage based on assembly (short/long)	200x (short), ~60x (long)
# contigs	1182
N50 (contigs)	994.5 kb
Median contig length	50.3 kb
GC content	30.9%
# gene models	23,825

1209 **Table 2** Comparison of chemosensory receptor repertoires between *Archegozetes longisetosus*
1210 and other arthropods. GR= gustatory receptor, OR= odorant receptor, IR= ionotropic receptor,
1211 OBP= odorant binding protein, CSP= chemosensory protein.

	Chemosensory receptors				
	GR	**OR**	**IR**	**OBP**	**CSP**
A. longesitosus	68	0	3	0	1
spider mite	689	0	4	0	0
deer tick	60	0	22	0	1
house spider	634	0	108	4	0
fruit fly	73	62	66	51	4

1212
1213

Supplementary Material

Supplementary Figures

Figure S1 Results of *in silico* genome size estimations based on jellyfish *k-mer* counting using **a:** GenomeScope v1.0 and **b** and **c**: the findGSE v0.1.0 R package (Sun et al. 2018).

Figure S2 Phylogenetic placement of *Archegozetes longisetosus* among other chelicerates. **a**: Maximum likelihood phylogeny based on concatenation of 1,121 orthologs Branch lengths unit is substitutions per site and the node values reflect bootstrap supports. **b**: Coalescence species tree reconstruction of the 1,121 filtered orthogroups. Branch lengths are presented in coalescent units (differences in the 1,121 gene trees) and the node values reflect the local posterior probabilities.

Figure S3 Comparisons of protein-coding genes of 23 arthropod species, including *Archegozetes*. The bar charts show the proportion of orthrogroup conservation with each species (see insert legend) based on OrthoFinder clustering.

Figure S4 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Paired protein orthologs. **(a)** Maximum likelihood tree showing the relationship of the Eyegone, Pax-3/7, and Pax-6 clades as collapsed subtrees. **(b)** The un-collapsed clade in **A** showing the phylogenetic relationships of selected Eyegone proteins and the putative *A. longisetosus* Eyegone ortholog. **(c)** The un-collapsed clade in **A** showing the phylogenetic relationships of selected Pax-3/7 proteins and the putative *A. longisetosus* Pax-3/7 orthologs. **(d)** The un-collapsed clade in **A** showing the phylogenetic relationships of selected Pax-6 proteins and the putative *A. longisetosus* Pax-6 ortholog. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S5 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Eyes absent (Eya) protein ortholog and selected metazoan Eya proteins. The *A. longisetosus* ortholog is in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S6 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Hairy protein ortholog and selected metazoan Hairy proteins. Hairy/E(spl) proteins were used as an outgroup. The *A. longisetosus* ortholog is in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S7 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Omb, T-box H15, and TBX1 protein orthologs and selected metazoan T-box proteins. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S8 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Runt protein ortholog and selected arthropod Runt proteins. The *A. longisetosus* ortholog is in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers. See

Figure S9 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Six family protein orthologs and selected metazoan Six family proteins. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S10 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Sp-family protein orthologs and selected metazoan Sp-family proteins. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S11 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Wnt-family protein orthologs and selected metazoan Wnt proteins. The tree is organized as a cladogram for easier viewing. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

Figure S12 Unrooted maximum-likelihood phylogenetic trees of cell-wall degrading enzymes based on the alignment of amino acid sequences. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports. *Archegozetes* sequences are highlighted in red.

Figure S13 Maximum likelihood phylogeny of ionotropic receptors and ionotropic glutamate receptors of *Archegozetes* (Along), *Dinothrombium* (Dt), *Leptothrombidium* (Ld), *Tetranychus* (Tu) and *Drosophila* (Dmel). The tree was rooted to the middle point. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports.

Figure S14 Maximum likelihood phylogenetic tree of gustatory receptors of *Archegozetes* (Along), *Ixodes* (Is), *Tropilaelaps* (Tm), *Metaseiulus* (Mocc) and *Drosophila* (Dmel). The tree was rooted to the middle point. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports.

Figure S15 Unrooted maximum-likelihood tree of glycoside hydrolase family 48 members (GH_48) from Streptomyces bacteria and HGT genes from other arthropods as well as *Archegozetes* (GH_48 II). Branch lengths unit is substitutions per site and the node values reflect bootstrap supports.

Supplementary Table

Table S1 Contamination contigs identified from Blobtools.

Table S2 Phylogenetic statistics of the PhyML constructed trees as well as the matrices selected by the SMS model selection tool.

Table S3 The average transcript per million (tpm) values for the transcripts highlighted in the heatmap (**Figure 5f**) for each instar stage.

Table S4 Candidate HGTs identified from the *Archegozetes* genome. The genes were filtered first if they were predicted to be contamination from Blobtools and the Alien Index report, second if they overlapped predicted repeats by ≥ 50%, and third if they were not expressed in any developmental stage. Annotation is provided from similarity searches against the NCBI nr database, other oribatid mite and eggNOG database. The taxonomy of the sequences upstream and downstream of each candidate HGT was determine using the eggNOG predicted taxonomic group.

References

Alberti G. 1984. The contribution of comparative spermatology to problems of acarine systematics. *Acarology VI*.479-490.

Alberti G. 1991. Spermatology in the Acari: systematic and functional implications. In: Schuster R, Murphy PW, editors. The Acari - Reproduction, Development and Life-History Strategies. London: Chapman & Hall. p. 77-105.

Alberti G. 1998. Fine structure of receptor organs in oribatid mites (Acari). In: Ebermann E, editor. Arthropod biology: Contributions to morphology, ecology and systematics. Wien: Austrian Academy of Sciences Press p. 27-77.

Alberti G, Coons LB. 1999. Acari-Mites. New York: Wiley.

Alberti G, Michalik P. 2004. Feinstrukturelle Aspekte der Fortpflanzungssysteme von Spinnentieren (Arachnida). *Denisia*. 12(14):1-62.

Alberti G, Moreno-Twose AI. 2012. Fine structure of the primary eyes in *Heterochthonius gibbus* (Oribatida, Heterochthoniidae) with some general remarks on photosensitive structures in oribatid and other actinotrichid mites. *Soil Org*. 84(2):391-408.

Alfsnes K, Leinaas HP, Hessen DO. 2017. Genome size in arthropods; different roles of phylogeny, habitat and life history in insects and crustaceans. *Ecol Evol*. 7(15):5939-5947.

Altincicek B, Kovacs JL, Gerardo NM. 2012. Horizontally transferred fungal carotenoid genes in the two-spotted spider mite Tetranychus urticae. *Biol Lett*. 8(2):253-257.

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data. *Ref Source*.https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Aoki J. 1965. Oribatiden (Acarina) Thailands. I. *Nat Life Southeast Asia*. 4:129-193.

Arkhipova I, Meselson M. 2000. Transposable elements in sexual and ancient asexual taxa. *PNAS*. 97(26):14473-14477.

Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS. 2016. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res*. 44(W1):W147-W153.

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 28(1):45-48.

1336 Ballesteros JA, Sharma PP. 2019. A critical appraisal of the placement of *Xiphosura*
1337 (Chelicerata) with account of known sources of phylogenetic error. *Syst Biol*. 68(6):896-
1338 917.
1339 Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in
1340 eukaryotic genomes. *Mobile DNA*. 6(1):11.
1341 Barnett AA, Thomas RH. 2012. The delineation of the fourth walking leg segment is temporally
1342 linked to posterior segmentation in the mite *Archegozetes longisetosus* (Acari: Oribatida,
1343 Trhypochthoniidae). *Evol Dev*. 14(4):383-392.
1344 Barnett AA, Thomas RH. 2013a. The expression of limb gap genes in the mite *Archegozetes*
1345 *longisetosus* reveals differential patterning mechanisms in chelicerates. *Evol Dev*.
1346 15(4):280-292.
1347 Barnett AA, Thomas RH. 2013b. Posterior Hox gene reduction in an arthropod: Ultrabithorax
1348 and Abdominal-B are expressed in a single segment in the mite *Archegozetes*
1349 *longisetosus*. *EvoDevo*. 4(1):23.
1350 Barnett AA, Thomas RH. 2018. Early segmentation in the mite *Archegozetes longisetosus*
1351 reveals conserved and derived aspects of chelicerate development. *Dev Genes Evol*.
1352 228(5):213-217.
1353 Barrero RA, Guerrero FD, Black M, McCooke J, Chapman B, Schilkey F, de Leon AAP, Miller
1354 RJ, Bruns S, Dobry J. 2017. Gene-enriched draft genome of the cattle tick *Rhipicephalus*
1355 *microplus*: assembly by the hybrid Pacific Biosciences/Illumina approach enabled
1356 analysis of the highly repetitive genome. *Int J Parasitol*. 47(9):569-583.
1357 Barton NH. 2010. Mutation and the evolution of recombination. *Philos Trans R Soc Lond B Biol*
1358 *Sci*. 365(1544):1281-1294.
1359 Bast J, Schaefer I, Schwander T, Maraun M, Scheu S, Kraaijeveld K. 2016. No accumulation of
1360 transposable elements in asexual arthropods. *Mol Biol Evol*. 33(3):697-706.
1361 Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M,
1362 Moxon S, Sonnhammer EL. 2004. The Pfam protein families database. *Nucleic Acids*
1363 *Res*. 32(suppl_1):D138-D141.
1364 Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and
1365 Powerful Approach to Multiple Testing. *J R Stat Soc Series B*. 57(1):289-300.

1366 Bensoussan N, Santamaria ME, Zhurov V, Diaz I, Grbić M, Grbić V. 2016. Plant-herbivore
1367 interaction: dissection of the cellular pattern of *Tetranychus urticae* feeding on the host
1368 plant. *Front Plant Sci*. 7:1105.
1369 Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant ionotropic glutamate
1370 receptors as chemosensory receptors in *Drosophila*. *Cell*. 136(1):149-162.
1371 Beran F, Köllner TG, Gershenzon J, Tholl D. 2019. Chemical convergence between plants and
1372 insects: biosynthetic origins and functions of common secondary metabolites. *New*
1373 *Phytol*.
1374 Bergmann P, Laumann M, Norton RA, Heethoff M. 2018. Cytological evidence for automictic
1375 thelytoky in parthenogenetic oribatid mites (Acari, Oribatida): Synaptonemal complexes
1376 confirm meiosis in *Archegozetes longisetosus*. *Acarologia*. 58(2):342-356.
1377 Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák
1378 Z, Levin HL, Macfarlan TS. 2018. Ten things you should know about transposable
1379 elements. *Genome Biol*. 19(1):1-12.
1380 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq
1381 quantification. *Nat Biotech*. 34(5):525-527.
1382 Breitmaier E. 2006. Terpenes: flavors, fragrances, pharmaca, pheromones. Weinheim: John
1383 Wiley & Sons.
1384 Brückner A. 2021. Data related to "The *Archegozetes longisetosus* genome project" 1.0 ed.
1385 Pasadena: CaltechDATA https://doi.org/10.22002/D1.1876
1386 Brückner A, Heethoff M. 2016. Scent of a mite: origin and chemical characterization of the
1387 lemon-like flavor of mite-ripened cheeses. *Exp Appl Acarol*. 69(3):249-261.
1388 Brückner A, Heethoff M. 2017. The ontogeny of oil gland chemistry in the oribatid mite
1389 *Archegozetes longisetosus* Aoki (Oribatida, Trhypochthoniidae). *Int J Acarol*. 43(5):337-
1390 342.
1391 Brückner A, Heethoff M. 2018. Nutritional effects on chemical defense alter predator–prey
1392 dynamics. *Chemoecology*. 28(1):1-9.
1393 Brückner A, Hilpert A, Heethoff M. 2017a. Biomarker function and nutritional stoichiometry of
1394 neutral lipid fatty acids and amino acids in oribatid mites. *Soil Biol Biochem*. 115:35-43.

1395 Brückner A, Kaltenpoth M, Heethoff M. 2020. De novo biosynthesis of simple aromatic
1396 compounds by an arthropod (*Archegozetes longisetosus*). *Proc R Soc Lond Biol*.
1397 287(1934):20201429.
1398 Brückner A, Parker J. 2020. Molecular evolution of gland cell types and chemical interactions in
1399 animals. *J Exp Biol*. 223(Suppl 1).
1400 Brückner A, Raspotnig G, Wehner K, Meusinger R, Norton RA, Heethoff M. 2017b. Storage and
1401 release of hydrogen cyanide in a chelicerate (*Oribatula tibialis*). *PNAS*. 114(13):3469-
1402 3472.
1403 Brückner A, Schuster R, Smit T, Heethoff M. 2018a. Imprinted or innated food preferences in
1404 the model mite *Archegozetes longisetosus* (Actinotrichida, Oribatida,
1405 Trhypochthoniidae). *Soil Org*. 90(1):23-26.
1406 Brückner A, Schuster R, Smit T, Pollierer MM, Schäffler I, Heethoff M. 2018b. Track the snack
1407 – Olfactory cues shape foraging behaviour of decomposing soil mites (Oribatida).
1408 *Pedobiologia*. 66(74-80).
1409 Brückner A, Schuster R, Wehner K, Heethoff M. 2018c. Effects of nutritional quality on the
1410 reproductive biology of *Archegozetes longisetosus* (Actinotrichida, Oribatida,
1411 Trhypochthoniidae) *Soil Org*. 90(1):1-12.
1412 Brückner A, Stabentheiner E, Leis HJ, Raspotnig G. 2015. Chemical basis of unwettability in
1413 Liacaridae (Acari, Oribatida): specific variations of a cuticular acid/ester-based system.
1414 *Exp Appl Acarol*. 66(3):313-335.
1415 Brückner A, Wehner K, Neis M, Heethoff M. 2016. Attack and defense in a gamasid-oribatid
1416 mite predator-prey experiment - sclerotization outperforms chemical repellency.
1417 *Acarologia*. 56(4):451-461.
1418 Bruna T, Hoff K, Stanke M, Lomsadze A, Borodovsky M. 2020. BRAKER2: Automatic
1419 Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a
1420 Protein Database. *bioRxiv*.
1421 Budelli G, Ni L, Berciu C, van Giesen L, Knecht ZA, Chang EC, Kaminski B, Silbering AF,
1422 Samuel A, Klein M. 2019. Ionotropic receptors specify the morphogenesis of phasic
1423 sensors controlling rapid thermal preference in *Drosophila*. *Neuron*. 101(4):738-747.
1424 e733.

1425 Bunnell T, Hanisch K, Hardege JD, Breithaupt T. 2011. The fecal odor of sick hedgehogs
1426 (*Erinaceus europaeus*) mediates olfactory attraction of the tick *Ixodes hexagonus*. *J*
1427 *Chem Ecol*. 37(4):340.
1428 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The
1429 Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics.
1430 *Nucleic Acids Res*. 37(suppl_1):D233-D238.
1431 Cao Z, Yu Y, Wu Y, Hao P, Di Z, He Y, Chen Z, Yang W, Shen Z, He X. 2013. The genome of
1432 *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nature*
1433 *communications*. 4(1):1-10.
1434 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated
1435 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15):1972-
1436 1973.
1437 Carlson M, Pagès H. 2019. AnnotationForge: Tools for building SQLite-Based Annotation Data
1438 Packages. *R package version*. 1(0).
1439 Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences.
1440 Gene Prediction. Springer. p. 1-14.
1441 Charlesworth B. 2012. The effects of deleterious mutations on evolution at linked sites. *Genetics*.
1442 190(1):5-22.
1443 Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome
1444 assembly. *Bioinformatics*. 30(1):31-37.
1445 Sequenced Arthropod Genomes. 2020. Manhattan, KS: i5k initiative; [accessed 10/27/2020].
1446 Cohen AC. 1995. Extra-oral digestion in predaceous terrestrial Arthropoda. *Annual Rev Entomo*.
1447 40(1):85-103.
1448 Consortium GO. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic*
1449 *Acids Res*. 32(suppl_1):D258-D261.
1450 Consortium GO. 2019. The gene ontology resource: 20 years and still GOing strong. *Nucleic*
1451 *Acids Res*. 47(D1):D330-D338.
1452 Consortium TGS. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*.
1453 452(7190):949.
1454 Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of
1455 intersecting sets and their properties. *Bioinformatics*. 33(18):2938-2940.

1456 Cook CE, Smith ML, Telford MJ, Bastianello A, Akam M. 2001. Hox genes and the phylogeny
1457 of the arthropods. *Curr Biol*. 11(10):759-763.
1458 Cornman RS, Schatz MC, Johnston JS, Chen Y-P, Pettis J, Hunt G, Bourgeois L, Elsik C,
1459 Anderson D, Grozinger CM. 2010. Genomic survey of the ectoparasitic mite Varroa
1460 destructor, a major pest of the honey bee *Apis mellifera*. *BMC Genomics*. 11(1):602.
1461 Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach
1462 to identify miniature inverted-repeat transposable elements in large genomes. *BMC*
1463 *Bioinf*. 19(1):348.
1464 Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. 2015. Expression of multiple
1465 horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes.
1466 *Genome Biol*. 16(1):1-13.
1467 Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R.
1468 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the
1469 evolution of insect taste and olfaction. *PLoS genetics*. 6(8):e1001064.
1470 Dabert M. 2006. DNA markers in the phylogenetics of the Acari. *Biological Lett*. 43(2):97-107.
1471 Dabert M, Witalinski W, Kazmierski A, Olszanowski Z, Dabert J. 2010. Molecular phylogeny of
1472 acariform mites (Acari, Arachnida): Strong conflict between phylogenetic signal and
1473 long-branch attraction artifacts. *Mol Phylogenet Evol*. 56:222-241.
1474 Degenhardt J, Köllner TG, Gershenzon J. 2009. Monoterpene and sesquiterpene synthases and
1475 the origin of terpene skeletal diversity in plants. *Phytochem*. 70(15-16):1621-1637.
1476 Dej KJ, Gerasimova T, Corces VG, Boeke JD. 1998. A hotspot for the *Drosophila* gypsy
1477 retroelement in the ovo locus. *Nucleic Acids Res*. 26(17):4019-4024.
1478 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
1479 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29(1):15-21.
1480 Domes K, Althammer M, Norton RA, Scheu S, Maraun M. 2007. The phylogenetic relationship
1481 between Astigmata and Oribatida (Acari) as indicated by molecular markers. *Exp Appl*
1482 *Acarol*. 42(3):159-171.
1483 Dong X, Armstrong SD, Xia D, Makepeace BL, Darby AC, Kadowaki T. 2017. Draft genome of
1484 the honey bee ectoparasitic mite, *Tropilaelaps mercedesae*, is shaped by the parasitic life
1485 history. *GigaScience*. 6(3):gix008.

1486 Dong X, Chaisiri K, Xia D, Armstrong SD, Fang Y, Donnelly MJ, Kadowaki T, McGarry JW,
1487 Darby AC, Makepeace BL. 2018. Genomes of trombidid mites reveal novel predicted
1488 allergens and laterally transferred genes associated with secondary metabolism.
1489 *GigaScience*. 7(12):giy127.
1490 dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB,
1491 Gelbart WM, Consortium F. 2015. FlyBase: introduction of the *Drosophila melanogaster*
1492 Release 6 reference genome assembly and large-scale migration of genome annotations.
1493 *Nucleic Acids Res*. 43(D1):D690-D697.
1494 Dunlop J, Alberti G. 2008. The affinities of mites and ticks: a review. *J Zool Syst Evol Res*.
1495 46(1):1-18.
1496 Dunlop J, Selden P. 1998. The early history and phylogeny of the chelicerates. In: Fortey RA,
1497 Thomas RH, editors. Arthropod Relationships The Systematics Association Special
1498 Volume Series. Dordrecht: Springer. p. 221-235.
1499 Dunlop JA. 2010. Geological history and phylogeny of Chelicerata. *Arthropod Struct Dev*. 39(2-
1500 3):124-142.
1501 Dunlop JA, Lamsdell JC. 2017. Segmentation and tagmosis in Chelicerata. *Arthropod Struct*
1502 *Dev*. 46(3):395-418.
1503 Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comp Biol*. 7(10):e1002195.
1504 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
1505 throughput. *Nucleic Acids Res*. 32(5):1792-1797.
1506 Eisenreich W, Bacher A, Arigoni D, Rohdich F. 2004. Biosynthesis of isoprenoids via the non-
1507 mevalonate pathway. *Cellular and molecular life sciences : CMLS*. 61(12):1401-1426.
1508 Emms DM, Kelly S. 2015a. OrthoFinder: solving fundamental biases in whole genome
1509 comparisons dramatically improves orthogroup inference accuracy. *Genome biology*.
1510 16(1):157.
1511 Emms DM, Kelly S. 2015b. OrthoFinder: solving fundamental biases in whole genome
1512 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*.
1513 16(1):157.
1514 Eriksson BJ, Fredman D, Steiner G, Schmid A. 2013. Characterisation and localisation of the
1515 opsin protein repertoire in the brain and retinas of a spider and an onychophoran. *BMC*
1516 *Evol Biol*. 13(1):186.

1517 Evans GO. 1992. Principles of Acarology. Wallingford: CAB International.
1518 Exner S. 1989. The physiology of the compound eyes of insects and crustaceans. Berlin,
1519 Heidelberg: Springer-Verlag GmbH & Co. KG.
1520 Faddeeva-Vakhrusheva A, Derks MF, Anvar SY, Agamennone V, Suring W, Smit S, van
1521 Straalen NM, Roelofs D. 2016. Gene family evolution reflects adaptation to soil
1522 environmental stressors in the genome of the collembolan *Orchesella cincta*. *Genome*
1523 *Biol Evol*. 8(7):2106-2117.
1524 Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MF, Anvar SY, Agamennone V, Suring W,
1525 Kampfraath AA, Ellers J, Le Ngoc G, van Gestel CA. 2017. Coping with living in the
1526 soil: the genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics*.
1527 18(1):493.
1528 Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet*.
1529 5:103-107.
1530 Flot J-F, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EG, Hejnol A, Henrissat B, Koszul R,
1531 Aury J-M. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta
1532 vaga. *Nature*. 500(7463):453-457.
1533 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation
1534 sequencing data. *Bioinformatics*. 28(23):3150-3152.
1535 Gainett G, Ballesteros JA, Kanzler CR, Zehms JT, Zern JM, Aharon S, Gavish-Regev E, Sharma
1536 PP. 2020. Systemic paralogy and function of retinal determination network homologs in
1537 arachnids. *BMC Genomics*. 21:811.
1538 Geib SM, Hall B, Derego T, Bremer FT, Cannoles K, Sim SB. 2018. Genome Annotation
1539 Generator: a simple tool for generating and correcting WGS annotation tables for NCBI
1540 submission. *GigaScience*. 7(4):giy018.
1541 Gilbert HJ. 2010. The biochemistry and structural biology of plant cell wall deconstruction.
1542 *Plant Physiol* 153(2):444-455.
1543 Giribet G, Edgecombe GD. 2019. The phylogeny and evolutionary history of arthropods. *Curr*
1544 *Biol*. 29(12):R592-R602.
1545 Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid
1546 rotifers. *Science*. 320(5880):1210-1213.

1547 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
1548 Raychowdhury R, Zeng Q. 2011. Trinity: reconstructing a full-length transcriptome
1549 without a genome from RNA-Seq data. *Nat Biotech*. 29(7):644.
1550 Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W,
1551 Ngoc PCT, Ortego F. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest
1552 adaptations. *Nature*. 479(7374):487-492.
1553 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New
1554 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
1555 performance of PhyML 3.0. *Syst Biol*. 59(3):307-321.
1556 Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, Sattelle DB, De
1557 La Fuente J, Ribeiro JM, Megy K. 2016. Genomic insights into the *Ixodes scapularis* tick
1558 vector of Lyme disease. *Nature communications*. 7(1):1-13.
1559 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
1560 Li B, Lieber M. 2013. De novo transcript sequence reconstruction from RNA-seq using
1561 the Trinity platform for reference generation and analysis. *Nat Protoc*. 8(8):1494-1512.
1562 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR.
1563 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the
1564 Program to Assemble Spliced Alignments. *Genome Biol*. 9(1):R7.
1565 Haq MA. 1993. Symbiotic association of mites and microbes in cellulose degradation. *Soil Org*
1566 *Sustain*. 4:81-85.
1567 Hartmann K, Laumann M, Bergmann P, Heethoff M, Schmelzle S. 2016. Development of the
1568 synganglion and morphology of the adult nervous system in the mite *Archegozetes*
1569 *longisetosus* Aoki (Chelicerata, Actinotrichida, Oribatida). *J Morphol*. 277(4):537-548.
1570 Harzsch S, Vilpoux K, Blackburn DC, Platchetzki D, Brown NL, Melzer R, Kempler KE,
1571 Battelle BA. 2006. Evolution of arthropod visual systems: development of the eyes and
1572 central visual pathways in the horseshoe crab *Limulus polyphemus* Linnaeus, 1758
1573 (Chelicerata, Xiphosura). *Dev Dyn*. 235(10):2641-2655.
1574 Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. *Genome Biol*.
1575 5(6):1-6.
1576 Heethoff M. 2012. Regeneration of complex oil-gland secretions and its importance for chemical
1577 defense in an oribatid mite. *J Chem Ecol*. 38(9):1116-1123.

1578 Heethoff M, Bergmann P, Laumann M, Norton RA. 2013. The 20th anniversary of a model mite:
1579 A review of current knowledge about *Archegozetes longisetosus* (Acari, Oribatida).
1580 *Acarologia*. 53(4):353-368.
1581 Heethoff M, Bergmann P, Norton RA. 2006. Karyology and sex determination of oribatid mites.
1582 *Acarologia*. 46(1-2):127-131.
1583 Heethoff M, Brückner A, Schmelzle S, Schubert M, Bräuer M, Meusinger R, Dötterl S, Norton
1584 RA, Raspotnig G. 2018. Life as a fortress–structure, function, and adaptive values of
1585 morphological and chemical defense in the oribatid mite *Euphthiracarus reticulatus*
1586 (Actinotrichida). *BMC Zoology*. 3(1):7.
1587 Heethoff M, Koerner L. 2007. Small but powerful: the oribatid mite *Archegozetes longisetosus*
1588 Aoki (Acari, Oribatida) produces disproportionately high forces. *J Exp Biol*.
1589 210(17):3036-3042.
1590 Heethoff M, Koerner L, Norton RA, Raspotnig G. 2011a. Tasty but protected-first evidence of
1591 chemical defense in oribatid mites. *J Chem Ecol*. 37:1037-1043.
1592 Heethoff M, Laumann M, Bergmann P. 2007. Adding to the reproductive biology of the
1593 parthenogenetic oribatid mite, *Archegozetes longisetosus* (Acari, Oribatida,
1594 Trhypochthoniidae). *Turk J Zool*. 31:151-159.
1595 Heethoff M, Laumann M, Weigmann G, Raspotnig G. 2011b. Integrative taxonomy: Combining
1596 morphological, molecular and chemical data for species delineation in the
1597 parthenogenetic *Trhypochthonius tectorum* complex (Acari, Oribatida,
1598 Trhypochthoniidae). *Front Zool*. 8:2.
1599 Heethoff M, Norton RA. 2009. A new use for synchrotron X-ray microtomography: three-
1600 dimensional biomechanical modeling of chelicerate mouthparts and calculation of
1601 theoretical bite forces. *Inver Biol*. 128(4):332-339.
1602 Heethoff M, Norton RA, Raspotnig G. 2016. Once Again: Oribatid Mites and Skin Alkaloids in
1603 Poison Frogs. *J Chem Ecol*. 42(8):841-844.
1604 Heethoff M, Norton RA, Scheu S, Maraun M. 2009. Parthenogenesis in Oribatid Mites (Acari,
1605 Oribatida): Evolution Without Sex. In: Schön I, Martens K, van Dijk P, editors. Lost Sex:
1606 The Evolutionary Biology of Parthenogenesis. Dordrecht: Springer. p. 241-257.
1607 Heethoff M, Rall BC. 2015. Reducible defence: chemical protection alters the dynamics of
1608 predator–prey interactions. *Chemoecology*. 25(2):53-61.

1609 Heethoff M, Raspotnig G. 2012. Expanding the 'enemy-free space' for oribatid mites: evidence
1610 for chemical defense of juvenile *Archegozetes longisetosus* against the rove beetle *Stenus*
1611 *juno*. *Exp Appl Acarol*. 56:93-97.
1612 Heidemann K, Scheu S, Ruess L, Maraun M. 2011. Molecular detection of nematode predation
1613 and scavenging in oribatid mites: Laboratory and field experiments. *Soil Biol Biochem*.
1614 43:229-236.
1615 Heingård M, Turetzek N, Prpic N-M, Janssen R. 2019. FoxB, a new and highly conserved key
1616 factor in arthropod dorsal–ventral (DV) limb patterning. *EvoDevo*. 10(1):1-16.
1617 Hoffmann A, Thimm T, Dröge M, Moore ER, Munch JC, Tebbe CC. 1998. Intergeneric transfer
1618 of conjugative and mobilizable plasmids harbored by *Escherichia coli* in the gut of the
1619 soil microarthropod *Folsomia candida* (Collembola). *Appl Environ Microbiol*.
1620 64(7):2652-2659.
1621 Holland P, Hogan B. 1988. Expression of homeo box genes during mouse development: a
1622 review. *Gene Devol*. 2(7):773-782.
1623 Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, Pomerantz AF, Simao FA,
1624 Thomas J, Jiggins FM. 2016. Genome sequencing of the phytoseiid predatory mite
1625 *Metaseiulus occidentalis* reveals completely atomized Hox genes and superdynamic
1626 intron evolution. *Genome Biol Evol*. 8(6):1762-1775.
1627 Hrycaj SM, Wellik DM. 2016. Hox genes and evolution. *F1000Research*. 5.
1628 Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017.
1629 Fast genome-wide functional annotation through orthology assignment by eggNOG-
1630 mapper. *Mol Biol Evol*. 34(8):2115-2122.
1631 Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. *Evol*
1632 *Dev*. 4(6):459-499.
1633 Jeyaprakash A, Hoy MA. 2009. First divergence time estimate of spiders, scorpions, mites and
1634 ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp Appl Acarol*.
1635 47(1):1-18.
1636 Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S,
1637 Okuda S, Tokimatsu T. 2007. KEGG for linking genomes to life and the environment.
1638 *Nucleic Acids Res*. 36(suppl_1):D480-D484.

1639 Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids*
1640 *Res*. 28(1):27-30.
1641 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for
1642 functional characterization of genome and metagenome sequences. *J Mol Biol*.
1643 428(4):726-731.
1644 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
1645 improvements in performance and usability. *Mol Biol Evol*. 30(4):772-780.
1646 Keilwagen J, Hartung F, Grau J. 2019. GeMoMa: Homology-based gene prediction utilizing
1647 intron position conservation and RNA-seq data. Gene Prediction. Amsterdam: Springer.
1648 p. 161-177.
1649 Kitchen SA, Crowder CM, Poole AZ, Weis VM, Meyer E. 2015. De novo assembly and
1650 characterization of four anthozoan (Cnidaria) transcriptomes. *G3*. 5(11):2441-2452.
1651 Klimov PB, OConnor B. 2013. Is permanent parasitism reversible?—Critical evidence from
1652 early evolution of house dust mites. *Syst Biol*. 62(3):411-423.
1653 Klimov PB, OConnor BM, Chetverikov PE, Bolton SJ, Pepato AR, Mortazavi AL, Tolstikov
1654 AV, Bauchan GR, Ochoa R. 2018. Comprehensive phylogeny of acariform mites
1655 (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a
1656 long branch. *Mol Phylogenet Evol*. 119:105-117.
1657 Knecht ZA, Silbering AF, Ni L, Klein M, Budelli G, Bell R, Abuin L, Ferrer AJ, Samuel AD,
1658 Benton R. 2016. Distinct combinations of variant ionotropic glutamate receptors mediate
1659 thermosensation and hygrosensation in *Drosophila*. *eLife*. 5:e17879.
1660 Kocot KM, Citarella MR, Moroz LL, Halanych KM. 2013. PhyloTreePruner: a phylogenetic
1661 tree-based approach for selection of orthologous sequences for phylogenomics. *Evol*
1662 *Bioinform Online*. 9:429-435.
1663 Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of
1664 repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*.
1665 7(1):1-7.
1666 Koller LM, Wirth S, Raspotnig G. 2012. Geranial-rich oil gland secretions: a common
1667 phenomenon in the Histiostomatidae (Acari, Astigmata)? *Int J Acarol*. 38(5):420-426.

1668 Königsmann T, Turetzek N, Pechmann M, Prpic N-M. 2017. Expression and function of the zinc
1669 finger transcription factor Sp6–9 in the spider *Parasteatoda tepidariorum*. *Dev Genes*
1670 *Evol*. 227(6):389-400.
1671 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable
1672 and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
1673 *Genome Res*. 27(5):722-736.
1674 Koyanagi M, Nagata T, Katoh K, Yamashita S, Tokunaga F. 2008. Molecular evolution of
1675 arthropod color vision deduced from multiple opsin genes of jumping spiders. *J Mol*
1676 *Evol*. 66(2):130-137.
1677 Kück P, Meusemann K. 2010. FASconCAT: Convenient handling of data matrices. *Mol*
1678 *Phylogenet Evol*. 56(3):1115-1118.
1679 Kuwahara Y. 2004. Chemical ecology of astigmatid mites. In: Cardé RT, Millar JG, editors.
1680 Advances in Insect Chemical Ecology. Cambridge: Cambridge University Press. p. 76-
1681 109.
1682 Kuwahara Y, Ibi T, Nakatani Y, Ryouno A, Mori N, Sakata T, Okabe K, Tagami K, Kurosa K.
1683 2001. Chemical ecology of astigmatid mites LIX. Neral, the alarm pheromone of
1684 Schwiebea elongata (Banks)(Acari: Acaridae). *J Acarol Soc Japan*. 10:19-25.
1685 Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies.
1686 *F1000Research*. 6(1287):1287.
1687 Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: new
1688 methods for selecting partitioned models of evolution for molecular and morphological
1689 phylogenetic analyses. *Mol Biol Evol*. 34(3):772-773.
1690 Latgé JP. 2007. The cell wall: a carbohydrate armour for the fungal cell. *Mol Microbiol*.
1691 66(2):279-290.
1692 Lawrence JG. 1997. Selfish operons and speciation by gene transfer. *Trends Microbiol*. 5(9):355-
1693 359.
1694 Lefort V, Longueville J-E, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol*
1695 *Evol*. 34(9):2422-2424.
1696 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
1697 *Bioinformatics*. 25(14):1754-1760.

Li W-N, Xue X-F. 2019. Mitochondrial genome reorganization provides insights into the relationship between oribatid mites and astigmatid mites (Acari: Sarcoptiformes: Oribatida). *Zoolo J Lin Soc*. 187(3):585-598.

Liana M, Witaliński W. 2005. Sperm structure and phylogeny of Astigmata. *J Morphol*. 265(3):318-324.

Liu W, Zhang R, Tian N, Xu X, Cao Y, Xian M, Liu H. 2015. Utilization of alkaline phosphatase PhoA in the bioproduction of geraniol by metabolically engineered *Escherichia coli*. *Bioengineered*. 6(5):288-293.

Lozano-Fernandez J, Tanner AR, Giacomelli M, Carton R, Vinther J, Edgecombe GD, Pisani D. 2019. Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nature Comm*. 10(1):1-8.

Luxton M. 1972. Studies on oribatid mites of a Danish beech wood soil .1. Nutritional Biology. *Pedobiologia*. 12(6):434-463.

Luxton M. 1979. Food and energy processing by oribatid mites. *Rev Ecol Biol Sol*. 16:103-111.

Luxton M. 1981. Studies on the oribatid mites of a Danish beech wood soil .7. Energy Budgets. *Pedobiologia*. 22(2):77-111.

Luxton M. 1982. The bology of mites from beech woodland soil. *Pedobiologia*. 23(1):1-8.

Madge D. 1965. Further studies on the behaviour of *Belba geniculosa* Oudms. in relation to various environmental stimuli. *Acarologia*. 7(4):744-757.

Maraun M, Erdmann G, Fischer BM, Pollierer MM, Norton RA, Schneider K, Scheu S. 2011. Stable isotopes revisited: Their use and limits for oribatid mite trophic ecology. *Soil Biol Biochem*. 43(5):877-882.

Maraun M, Heethoff M, Schneider K, Scheu S, Weigmann G, Cianciolo J, Thomas RH, Norton RA. 2004. Molecular phylogeny of oribatid mites (Oribatida, Acari): evidence for multiple radiations of parthenogenetic lineages. *Exp Appl Acarol*. 33(3):183-201.

Maraun M, Schatz H, Scheu S. 2007. Awesome or ordinary? Global diversity patterns of oribatid mites. *Ecography*. 30(2):209-216.

Maraun M, Scheu S. 2000. The structure of oribatid mite communities (Acari, Oribatida): Patterns, mechanisms and implications for future research. *Ecography*. 23(3):374-383.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 27(6):764-770.

1729 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1730 *EMBnet J*. 17(1):10-12.
1731 Mayer WE, Schuster LN, Bartelmes G, Dieterich C, Sommer RJ. 2011. Horizontal gene transfer
1732 of microbial cellulases into nematode genomes is associated with functional assimilation
1733 and gene turnover. *BMC Evol Biol*. 11(1):1-10.
1734 McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse
1735 RM, Ahn S-J, Arsala D. 2016. Genome of the Asian longhorned beetle (*Anoplophora*
1736 *glabripennis*), a globally significant invasive species, reveals key functional and
1737 evolutionary innovations at the beetle–plant interface. *Genome Biol*. 17(1):1-18.
1738 McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE,
1739 Friedrich F, Letsch H. 2019. The evolution and genomic basis of beetle diversity.
1740 *Proceedings of the National Academy of Sciences*. 116(49):24729-24737.
1741 Mitreva M, Smant G, Helder J. 2009. Role of horizontal gene transfer in the evolution of plant
1742 parasitism among nematodes. Horizontal Gene Transfer. Amsterdam: Springer. p. 517-
1743 535.
1744 Miziorko HM. 2011. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch*
1745 *Biochem Biophys*. 505(2):131-143.
1746 Montell C. 2009. A taste of the *Drosophila* gustatory receptors. *Curr Opin Neurobiol*. 19(4):345-
1747 353.
1748 Morita A, Mori N, Nishida R, Hirai N, Kuwahara Y. 2004. Neral biosynthesis via the
1749 mevalonate pathway, evidenced by D-glucose-1-13C feeding in *Carpoglyphus lactis* and
1750 13C incorporation into other opisthonotal gland exudates. *J Pest Sci*. 29:27-32.
1751 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome
1752 annotation and pathway reconstruction server. *Nucleic Acids Res*. 35(suppl_2):W182-
1753 W185.
1754 Muller HJ. 1964. The relation of recombination to mutaional advance. *Mutation Res* 106:2-9.
1755 Nagata T, Koyanagi M, Tsukamoto H, Terakita A. 2010. Identification and characterization of a
1756 protostome homologue of peropsin from a jumping spider. *J Comp Physiol A*. 196(1):51.
1757 Nagy LG, Merényi Z, Hegedüs B, Bálint B. 2020. Novel phylogenetic methods are needed for
1758 understanding gene function in the era of mega-scale genome sequencing. *Nucleic Acids*
1759 *Res*. 48(5):2209-2219.

1760 Ngoc PCT, Greenhalgh R, Dermauw W, Rombauts S, Bajda S, Zhurov V, Grbić M, Van de Peer
1761 Y, Van Leeuwen T, Rouze P. 2016. Complex evolutionary dynamics of massively
1762 expanded chemosensory receptor families in an extreme generalist chelicerate herbivore.
1763 *Genome Biol Evol*. 8(11):3323-3339.
1764 Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
1765 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*.
1766 32(1):268-274.
1767 Niimura Y, Nei M. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and
1768 tetrapods. *PNAS*. 102(17):6039-6044.
1769 Noge K, Kato M, Iguchi T, Mori N, Nishida R, Kuwahara Y. 2005. Biosynthesis of neral in
1770 *Carpoglyphus lactis* (Acari: Carpoglyphidae) and detection of its key enzyme, geraniol
1771 dehydrogenase, by electrophoresis. *J Acarol Soc Japan*. 14:75-81.
1772 Noge K, Kato M, Mori N, Kataoka M, Tanaka C, Yamasue Y, Nishida R, Kuwahara Y. 2008.
1773 Geraniol dehydrogenase, the key enzyme in biosynthesis of the alarm pheromone, from
1774 the astigmatid mite *Carpoglyphus lactis* (Acari: Carpoglyphidae). *The FEBS journal*.
1775 275(11):2807-2817.
1776 Norton RA. 1994. Evolutionary aspects of oribatid mite life histories and consequences for the
1777 origin of the Astigmata. In: Houck MA, editor. Mites: Ecological and Evolutionary
1778 Analyses of Life-History Patterns. Chapman & Hall ed. New York: Chapman & Hall. p.
1779 99-135.
1780 Norton RA. 1998. Morphological evidence for the evolutionary origin of Astigmata (Acari :
1781 Acariformes). *Exp Appl Acarol*. 22(10):559-594.
1782 Morales-Malacara JB, Behan-Pelletier V, Ueckermann E, Perez TM, Estrada-Venegas EG, Badii
1783 M, editors. Holistic acarology and ultimate causes: examples from the oribatid mites.
1784 Acarology XI: Proceedings of the International Congress; 2007; Mexico. Sociedad
1785 Latinoamericana de Acarologia.
1786 Norton RA, Franklin E. 2018. *Paraquanothrus* n. gen. from freshwater rock pools in the USA,
1787 with new diagnoses of Aquanothrus, Aquanothrinae, and Ameronothridae (Acari,
1788 Oribatida). *Acarologia*. 58(3):557-627.

1789 Norton RA, Fuangarworn M. 2015. Nanohystricidae n. fam., an unusual, plesiomorphic
1790 enarthronote mite family endemic to New Zealand (Acari, Oribatida). *Zootaxa*.
1791 4027(2):151-204.
1792 Norton RA, Kethley JB, Johnston DE, O'Connor BM. 1993. Phylogenetic perspectives on
1793 genetic systems and reproductive modes of mites. In: Wrensch D, Ebbert M, editors.
1794 Evolution and Diversity of Sex Ratio in Insects and Mites. London: Chapman & Hall. p.
1795 8-99.
1796 Norton RA, Palmer SC. 1991. The distribution, mechanisms and evolutionary significance of
1797 parthenogenesis in oribatid mites. In: Schuster R, Murphy PW, editors. The Acari -
1798 Reproduction, Development and Life-History Strategies. London: Chapman & Hall. p.
1799 107-136.
1800 Nuzhdin SV, Petrov DA. 2003. Transposable elements in clonal lineages: lethal hangover from
1801 sex. *Biol J Lin Soc*. 79(1):33-41.
1802 Oldfield E, Lin FY. 2012. Terpene biosynthesis: modularity rules. *Angew Chem Int Ed Engl*.
1803 51(5):1124-1137.
1804 Oliver Jr JH. 1983. Chromosomes, genetic variance and reproductive strategies among mites and
1805 ticks. *Bull Entomol Soc Am*. 29(2):8-17.
1806 Oswald M, Fischer M, Dirninger N, Karst F. 2007. Monoterpenoid biosynthesis in
1807 *Saccharomyces cerevisiae*. *FEMS Yeast Res* 7(3):413-421.
1808 Oxley PR, Ji L, Fetter-Pruneda I, McKenzie SK, Li C, Hu H, Zhang G, Kronauer DJ. 2014. The
1809 genome of the clonal raider ant *Cerapachys biroi*. *Curr Biol*. 24(4):451-458.
1810 Pace RM, Grbić M, Nagy LM. 2016. Composition and genomic organization of arthropod Hox
1811 clusters. *EvoDevo*. 7(1):11.
1812 Pachl P, Domes K, Schulz G, Norton RA, Scheu S, Schaefer I, Maraun M. 2012. Convergent
1813 evolution of defense mechanisms in oribatid mites (Acari, Oribatida) shows no "ghosts of
1814 predation past". *Mol Phylogenet Evol*. 65(2):412-420.
1815 Palmer J, Stajich J. 2017. Funannotate: eukaryotic genome annotation pipeline.
1816 https://funannotate.readthedocs.io/en/latest/.
1817 Palmer M, Bantle J, Guo X, Fargoxy1 WS. 1994. Genome size and organization in the ixodid
1818 tick *Amblyomma americanum* (L.). *Insect Mol Biol*. 3(1):57-62.

1819 Palmer SC, Norton RA. 1992. Genetic diversity in thelytokous oribatid mites (Acari;
1820 Acariformes: Desmonomata). *Biochem Syst Ecol*. 20(3):219-231.
1821 Panfilio KA, Jentzsch IMV, Benoit JB, Erezyilmaz D, Suzuki Y, Colella S, Robertson HM,
1822 Poelchau MF, Waterhouse RM, Ioannidis P. 2019. Molecular evolutionary trends and
1823 feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome.
1824 *Genome Biol*. 20(1):64.
1825 Patten W. 1887. Eyes of molluscs and arthropods. *J Morphol*. 1(1):67-92.
1826 Pepato A, Klimov P. 2015. Origin and higher-level diversification of acariform mites–evidence
1827 from nuclear ribosomal genes, extensive taxon sampling, and secondary structure
1828 alignment. *BMC Evol Biol*. 15(1):178.
1829 Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof
1830 B. 2019. Diversity and evolution of the transposable element repertoire in arthropods
1831 with particular reference to insects. *BMC Evol Biol*. 19(1):11.
1832 Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq
1833 incorporating quantification uncertainty. *Nature Meth* 14(7):687.
1834 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately Maximum-Likelihood trees
1835 for large alignments. *PLoS One*. 5(3):e9490.
1836 Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-
1837 redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*.
1838 33(suppl_1):D501-D504.
1839 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
1840 features. *Bioinformatics*. 26(6):841-842.
1841 R_Core_Team. 2019. R: A language and environment for statistical computing. R Foundation
1842 for Statistical Computing, Vienna, Austria. 2019.
1843 Raspotnig G. 2006. Chemical alarm and defence in the oribatid mite *Collohmannia gigantea*
1844 (Acari: Oribatida). *Exp Appl Acarol*. 39(3):177-194.
1845 Sabelis MW, Bruin J, editors. Oil gland secretions in Oribatida (Acari). Trends in Acarology;
1846 2009; Dordrecht. Springer.
1847 Raspotnig G, Kaiser R, Stabentheiner E, Leis HJ. 2008. Chrysomelidial in the Opisthonotal
1848 Glands of the Oribatid Mite, *Oribotritia berlesei*. *J Chem Ecol*. 34(8):1081-1088.

1849 Raspotnig G, Norton RA, Heethoff M. 2011. Oribatid mites and skin alkaloids in poison frogs.
1850 *Biol Lett*. 7:555-556.
1851 Raspotnig G, Schuster R, Krisper G. 2004. Citral in oil gland secretions of Oribatida (Acari): a
1852 key component for phylogenetic analyses. *Abh Ber Naturkundemus Görlitz*. 76:43-50.
1853 Rawlings ND, Barrett AJ, Bateman A. 2010. MEROPS: the peptidase database. *Nucleic Acids*
1854 *Res*. 38(suppl_1):D227-D233.
1855 Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW.
1856 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-
1857 coding sequences. *Nature*. 463(7284):1079-1083.
1858 Rider SD, Morgan MS, Arlian LG. 2015. Draft genome of the scabies mite. *Parasite Vectors*.
1859 8(1):1-14.
1860 Riha G. 1951. Zur Ökologie der Oribatiden in Kalksteinböden. *Zoolo Jahrb*. 80:407-450.
1861 Robertson HM, Wanner KW. 2006. The chemoreceptor superfamily in the honey bee, *Apis*
1862 *mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res*.
1863 16(11):1395-1403.
1864 Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor
1865 gene superfamily in *Drosophila melanogaster*. *PNAS*. 100(suppl 2):14537-14542.
1866 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.
1867 2011. Integrative genomics viewer. *Nat Biotechnol*. 29(1):24-26.
1868 Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source
1869 tool for metagenomics. *PeerJ*. 4:e2584.
1870 Rytz R, Croset V, Benton R. 2013. Ionotropic receptors (IRs): chemosensory ionotropic
1871 glutamate receptors in *Drosophila* and beyond. *Insect Biochem Mol Biol*. 43(9):888-897.
1872 Sakata T. 1997. Natural Chemistry of Mite Secretions. [Kyoto, Japan]: Kyoto University.
1873 Sakata T, Norton RA. 2001. Opisthonotal gland chemistry of early-derivative oribatid mites
1874 (Acari) and its relevance to systematic relationships of Astigmata. *Int J Acarol*.
1875 27(4):281-292.
1876 Sakata T, Norton RA. 2003. Opisthonotal gland chemistry of a middle-derivative oribatid mite,
1877 *Archegozetes longisetosus* (Acari : Trhypochthoniidae). *Int J Acarol*. 29(4):345-350.
1878 Sakata T, Tagami K, Kuwahara Y. 1995. Chemical ecology of oribatid mites. I. Oil gland
1879 components of *Hydronothrus crispus* Aoki. *J Acarol Soc Japan*. 4(2):69-75.

1880 Samadi L, Schmid A, Eriksson BJ. 2015. Differential expression of retinal determination genes
1881 in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *EvoDevo*.
1882 6(1):16.
1883 Sánchez-Gracia A, Vieira F, Rozas J. 2009. Molecular evolution of the major chemosensory
1884 gene families in insects. *Heredity*. 103(3):208-216.
1885 Sánchez-Gracia A, Vieira FG, Almeida FC, Rozas J. 2011. Comparative genomics of the major
1886 chemosensory gene families in Arthropods.
1887 *eLS*.https://doi.org/10.1002/9780470015902.a9780470022848.
1888 Santos VT, Ribeiro L, Fraga A, de Barros CM, Campos E, Moraes J, Fontenele MR, Araujo HM,
1889 Feitosa NM, Logullo C. 2013. The embryogenesis of the tick *Rhipicephalus (Boophilus)*
1890 *microplus*: the establishment of a new chelicerate model system. *Genesis*. 51(12):803-
1891 818.
1892 Saporito RA, Donnelly MA, Norton RA, Garraffo HM, Spande TF, Daly JW. 2007. Oribatid
1893 mites as a major dietary source for alkaloids in poison frogs. *PNAS*. 104(21):8885-8890.
1894 Saporito RA, Spande TF, Garraffo HM, Donnelly MA. 2009. Arthropod alkaloids in poison
1895 frogs: A review of the dietary hypothesis. *Heterocycles*. 79:277-297.
1896 Schaefer I, Norton RA, Scheu S, Maraun M. 2010. Arthropod colonization of land - Linking
1897 molecules and fossils in oribatid mites (Acari, Oribatida). *Mol Phylogenet Evol*. 57:113-
1898 121.
1899 Schmelzle S, Blüthgen N. 2019. Under pressure: force resistance measurements in box mites
1900 (Actinotrichida, Oribatida). *Front Zool*. 16(1):24.
1901 Schneider K, Maraun M. 2005. Feeding preferences among dark pigmented fungal taxa
1902 ("Dematiacea") indicate limited trophic niche differentiation of oribatid mites (Oribatida,
1903 Acari). *Pedobiologia*. 49(1):61-67.
1904 Schneider K, Migge S, Norton RA, Scheu S, Langel R, Reineking A, Maraun M. 2004a. Trophic
1905 niche differentiation in soil microarthropods (Oribatida, Acari): evidence from stable
1906 isotope ratios (N-15/N-14). *Soil Biol Biochem*. 36(11):1769-1774.
1907 Schneider K, Renker C, Scheu S, Maraun M. 2004b. Feeding biology of oribatid mites: a
1908 minireview. *Phytophaga*. 14:247-256.

1909 Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic N-M, Posnien N. 2015.
1910 Molecular characterization and embryonic origin of the eyes in the common house spider
1911 *Parasteatoda tepidariorum*. *EvoDevo*. 6(1):1-14.
1912 Schön I, Martens K, van Dijk P. 2009. Lost Sex - The Evolutionary Biology of Parthenogenesis.
1913 Dordrecht: Springer.
1914 Schwager EE, Schönauer A, Leite DJ, Sharma PP, McGregor AP. 2015. Chelicerata.
1915 Amsterdam: Springer.
1916 Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y,
1917 Esposito L, Bechsgaard J, Bilde T. 2017. The house spider genome reveals an ancient
1918 whole-genome duplication during arachnid evolution. *BMC Biology*. 15(1):1-27.
1919 Senthilan PR, Grebler R, Reinhard N, Rieger D, Helfrich-Förster C. 2019. Role of rhodopsins as
1920 circadian photoreceptors in the *Drosophila melanogaster*. *Biology*. 8(1):6.
1921 Senthilan PR, Helfrich-Förster C. 2016. Rhodopsin 7–the unusual rhodopsin in *Drosophila*.
1922 *PeerJ*. 4:e2427.
1923 Sharma PP, Schwager EE, Extavour CG, Giribet G. 2012. Hox gene expression in the
1924 harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate
1925 opisthosoma. *Evol Dev*. 14(5):450-463.
1926 Sharma PP, Tarazona OA, Lopez DH, Schwager EE, Cohn MJ, Wheeler WC, Extavour CG.
1927 2015. A conserved genetic mechanism specifies deutocerebral appendage identity in
1928 insects and arachnids. *Proc R Soc Lond Biol*. 282(1808):20150698.
1929 Shen WL, Kwon Y, Adegbola AA, Luo J, Chess A, Montell C. 2011. Function of rhodopsin in
1930 temperature discrimination in *Drosophila*. *Science*. 331(6022):1333-1336.
1931 Shimano S, Sakata T, Mizutani Y, Kuwahara Y, Aoki J-i. 2002. Geranial: the alarm pheromone
1932 in the nymphal stage of the oribatid mite, *Nothrus palustris*. *J Chem Ecol*. 28(9):1831-
1933 1837.
1934 Shimizu N, Sakata D, Schmelz EA, Mori N, Kuwahara Y. 2017. Biosynthetic pathway of
1935 aliphatic formates via a Baeyer-Villiger oxidation in mechanism present in astigmatid
1936 mites. *PNAS*. 114(10):2616-2621.
1937 Shingate P, Ravi V, Prasad A, Tay B-H, Garg KM, Chattopadhyay B, Yap L-M, Rheindt FE,
1938 Venkatesh B. 2020. Chromosome-level assembly of the horseshoe crab genome provides
1939 insights into its genome evolution. *Nature Comm*. 11(1):1-13.

1940 Shultz JW. 2007. A phylogenetic analysis of the arachnid orders based on morphological
1941 characters. *Zoolo J Lin Soc*. 150(2):221-265.
1942 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J,
1943 Hillier LW, Richards S. 2005. Evolutionarily conserved elements in vertebrate, insect,
1944 worm, and yeast genomes. *Genome Res* 15(8):1034-1050.
1945 Siepel H, de Ruiter-Dijkman EM. 1993. Feeding guilds of oribatid mites based on their
1946 carbohydrase activities. *Soil Biol Biochem*. 25(11):1491-1497.
1947 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:
1948 assessing genome assembly and annotation completeness with single-copy orthologs.
1949 *Bioinformatics*. 31(19):3210-3212.
1950 Simpson JT. 2014. Exploring genome characteristics and sequence quality without a reference.
1951 *Bioinformatics*. 30(9):1228-1235.
1952 RepeatMasker Open-3.0. 1996-2010. [accessed]. http://www.repeatmasker.org.
1953 Smit AF, Hubley R. 2008. RepeatModeler Open-1.0.
1954 Smrž J. 1992. Some adaptive features in the microanatomy of moss-dwelling oribatid mites
1955 (Acari: Oribatida) with respect to their ontogenetical development. *Pedobiologia*.
1956 36(5):306-320.
1957 Smrž J. 2000. A modified test for chitinase and cellulase activity in soil mites. *Pedobiologia*.
1958 44(2):186-189.
1959 Smrž J, Čatská V. 2010. Mycophagous mites and their internal associated bacteria cooperate to
1960 digest chitin in soil. *Symbiosis*. 52(1):33-40.
1961 Smrž J, Norton RA. 2004. Food selection and internal processing in *Archegozetes longisetosus*
1962 (Acari : Oribatida). *Pedobiologia*. 48(2):111-120.
1963 Stefaniak O. 1976. The microflora of the alimentary canal of *Achipteria coleoptrata* (Acarina,
1964 Oribatei). *Pedobiologia*. 16(185-194).
1965 Stefaniak O. 1981. The effect of fungal diet on the development of *Oppia nitens* (Acari, Oribatei)
1966 and on the microflora of its alimentary tract. *Pedobiologia*. 21:202-210.
1967 Sun H, Ding J, Piednoël M, Schneeberger K. 2018. findGSE: estimating genome size variation
1968 within human and Arabidopsis using k-mer frequencies. *Bioinformatics*. 34(4):550-557.

1969 Suzuki S, Kakuta M, Ishida T, Akiyama Y. 2014. GHOSTX: an improved sequence homology
1970 search algorithm using a query suffix array and a database suffix array. *PloS One*.
1971 9(8):e103833.
1972 Telford MJ, Thomas RH. 1998. Expression of homeobox genes shows chelicerate arthropods
1973 retain their deutocerebral segment. *PNAS*. 95(18):10671-10675.
1974 Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in
1975 novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome*
1976 *Res*. 18(12):1979-1990.
1977 Thiel T, Brechtel A, Brückner A, Heethoff M, Drossel B. 2018. The effect of reservoir-based
1978 chemical defense on predator-prey dynamics. *Theo Ecol*. 12(3):365-378.
1979 Thomas GW, Dohmen E, Hughes DS, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub
1980 NA, Batterham P, Bellair M. 2020. Gene content evolution in the arthropods. *Genome*
1981 *Biol* 21(1):1-14.
1982 Bernini F, Nannelli R, Nuzzaci G, de Lillo E, editors. Mites as models in development and
1983 genetics. Acarid Phylogeny and Evolution: Adaptation in Mites and Ticks: Proceedings
1984 of the IV Symposium of the European Association of Acarologists; 2002; Dordrecht.
1985 Kluwer Academic Publishers.
1986 Thorpe P, Escudero-Martinez CM, Cock PJ, Eves-van den Akker S, Bos JI. 2018. Shared
1987 transcriptional control and disparate gain and loss of aphid parasitism genes. *Genome*
1988 *Biol Evol*. 10(10):2716-2733.
1989 Trägårdh I. 1933. Methods of automatic collecting for studying the fauna of the soil. *Bull*
1990 *Entomol Res*. 24(2):203-214.
1991 Trapp SC, Croteau RB. 2001. Genomic organization of plant terpene synthases and molecular
1992 evolutionary implications. *Genetics*. 158(2):811-832.
1993 Van Dam MH, Trautwein M, Spicer GS, Esposito L. 2019. Advancing mite phylogenomics:
1994 Designing ultraconserved elements for Acari phylogeny. *Mol Ecol Res*. 19(2):465-475.
1995 van der Hammen L. 1970. La segmentation primitive des Acariens. *Acarologia*. 12(1):3-10.
1996 Van Zee JP, Geraci N, Guerrero F, Wikel S, Stuart J, Nene V, Hill C. 2007. Tick genomics: the
1997 Ixodes genome project and beyond. *Int J Parasitol*. 37(12):1297-1305.

1998 Vieira FG, Rozas J. 2011. Comparative genomics of the odorant-binding and chemosensory
1999 protein gene families across the Arthropoda: origin and evolutionary history of the
2000 chemosensory system. *Genome Biol Evol*. 3:476-490.
2001 Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC.
2002 2017. GenomeScope: fast reference-free genome profiling from short reads.
2003 *Bioinformatics*. 33(14):2202-2204.
2004 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
2005 Wortman J, Young SK. 2014. Pilon: an integrated tool for comprehensive microbial
2006 variant detection and genome assembly improvement. *PloS One*. 9(11):e112963.
2007 Walter DE, Proctor HC. 1998. Feeding behaviour and phylogeny: observations on early
2008 derivative Acari. *Exp Appl Acarol*. 22(1):39-50.
2009 Walter DE, Proctor HC. 1999. Mites: ecology, evolution, and behaviour. Amsterdam: Springer
2010 Netherlands.
2011 Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R,
2012 Ainscough R, Alexandersson M, An P. 2002. Initial sequencing and comparative analysis
2013 of the mouse genome. *Nature*. 420(6915):520-562.
2014 Weinstein SB, Kuris AM. 2016. Independent origins of parasitism in Animalia. *Biol Lett*.
2015 12(7):20160324.
2016 Woodring J. 1966. Color phototactic responses of an eyeless oribatid mite. *Acarologia*. 8(2):382-
2017 388.
2018 Wrensch DL, Kethley JB, Norton RA. 1994. Cytogenetics of holokinetic chromosomes and
2019 inverted meiosis: keys to the evolutionary success of mites, with generalizations on
2020 eukaryotes. In: Houck MA, editor. Mites: Ecological and Evolutionary Analyses of Life-
2021 History Patterns. New York: Chapman & Hall. p. 282-343.
2022 Wu C, Jordan MD, Newcomb RD, Gemmell NJ, Bank S, Meusemann K, Dearden PK, Duncan
2023 EJ, Grosser S, Rutherford K. 2017. Analysis of the genome of the New Zealand giant
2024 collembolan (*Holacanthella duospinosa*) sheds light on hexapod evolution. *BMC*
2025 *Genomics*. 18(1):795.
2026 Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal gene transfer contributes
2027 to the evolution of arthropod herbivory. *Genome Biol Evol*. 8(6):1785-1801.

Wybouw N, Van Leeuwen T, Dermauw W. 2018. A massive incorporation of microbial genes into the genome of *Tetranychus urticae*, a polyphagous arthropod herbivore. *Insect Mol Biol*. 27(3):333-351.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16(5):284-287.

Yunker C, Peter T, Norval R, Sonenshine D, Burridge M, Butler J. 1992. Olfactory responses of adult *Amblyomma hebraeum* and *A. variegatum* (Acari: Ixodiae) to attractant chemicals in laboratory tests. *Exp Appl Acarol*. 13(4):295-301.

Zachvatkin AA. 1941. Tyroglyphoidae (Acari). Moscow: Zoological Institute of the Acaemy of Science of the U.S.S.R.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 19(6):153.

Zhou J, Wang C, Yoon S-H, Jang H-J, Choi E-S, Kim S-W. 2014. Engineering *Escherichia coli* for selective geraniol production with minimized endogenous dehydrogenation. *J Biotech*. 169:42-50.

Zinkler D. 1971. Vergleichende Untersuchungen zum Wirkungsspektrum der Carbohydrasen laubstreubewohnender Oribatiden. *Zool Ges Verh*.149-153.