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.brueckner@gmail.com

ABSTRACT

Background

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. Noted for their biosynthesis capacities and biochemical diversity, the majority of oribatid mites possess a pair of exocrine opisthonotal oil-glands used for chemical defense and communication. Genomic resources are lacking for oribatids despite their species richness and ecological importance.

Results

We used a comparative genomic approach to investigate the developmental, sensory and glandular biosynthetic gene repertoire of the clonal, all-female oribatid mite species *Archegozetes*

longisetosus Aoki, a model species used by numerous laboratories for the past 30 years. Here, we present a 190-Mb genome assembly constructed from Nanopore MinION and Illumina sequencing platforms with 23,825 predicted protein-coding genes. 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 the 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 *Archegozetes longisetosus* diet. 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.

Conclusions

With the *Archegozetes longisetosus* genome, we now have the first high-quality, annotated genome of an oribatid mite genome. Given the mite's strength as an experimental model, the new sequence 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, terpenes, 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 (1, 2). The majority of these genomes, however, belong to the insects (e.g. flies, beetles, wasp, butterflies and bugs (2)) which compromise the most diverse, yet evolutionarily young and more derived taxa of arthropods (3, 4). In strong contrast, genome assemblies, many of which are incomplete or not well annotated, exist for the Chelicerata (1) – the other major subphylum of arthropods (3, 4). Chelicerates include sea spiders, spiders, mites and scorpions among other organisms, as well as several extinct taxa (5, 6). Chelicerates originated as marine animals about 500 million years ago (5, 7). 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 (8-10).

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 (11-17). Other than some lower-quality genome assemblies (18), 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 (19-21). Because the phylogeny of Chelicerata remains unresolved, additional chelicerate genomes are urgently needed for comparative analyses (6, 7, 22). To help address this deficit, we report here the genome assembly of the soil dwelling oribatid mite *Archegozetes longisetosus* (Aoki, 1965; **Figure 1**) (23) 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 highpulling strength, mechanical resistances and pivotal ecological importance (24-32). *Archegozetes*, like all members of its family Trhypochthoniidae (Figure 1a), reproduce via thelytoky (33). That means the all-female lineages procreate via automictic parthenogenesis with an inverted meiosis of the holokinetic chromosomes, resulting in clonal offspring (34-37). 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 (33, 38). Reviewing all available data, Norton (39, 40) and Heethoff et al. (33) 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 (30, 41). 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 (42, 43). The biological role of these glands was rather speculative for a long time; idea ranged from a lubricating and osmo- or thermoregulative function (44-46) to roles in chemical communication (47-49). 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 (26, 30, 50-52). While some chemicals appear to be alarm pheromones (47, 49), most function as defensive allomones (48). Interestingly,

alkaloids produced by oribatids mites are the ultimate source of most toxins sequestered by poison-frogs (51, 53).

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 (9, 12, 54, 55). Exceptions from this are Opiliones and sarcoptiform mites, i.e. oribatid and astigmatid mites,all of which ingest solid food (40, 56, 57). In general, oribatids feed on a wide range of different resources and show a low degree of dietary specialization (58). The typical food spectrum of Oribatida, includes leaf-litter, algae, fungi, lichens, nematodes, and small dead arthropods such as collembolans (44, 59-62). In laboratory feeding trials, oribatid mites tend to prefer dark pigmented fungi, but also fatty acid-rich plant-based food (58, 61). 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 (59, 63, 64). 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 (58, 60, 64, 65).

Early studies on *Archegozetes* and other mites found evidence for cellulase, chitinase and trehalase activity which was later attributed to symbiotic gut bacteria (65-71). 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 (11, 72-76). For instance, an in-depth analysis of the spider mite *Tetranychus urticae* revealed a massive incorporation of microbial genes into the mite's genome (11, 75). Horizontal gene transfer appears to be a common mechanism for soil organisms,

including mites, to acquire novel metabolic enzymes (11, 17, 74, 76-78), 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 (27, 33, 79-83). As such, *Archegozetes* is among the few experimentally tractable soil organisms and by far the best-studied oribatid mite species (33, 81, 84). Since *the mite* meets the most desirable requirements for model organisms (84), 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 (33 , **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 (33, 85-87). In the past 10 years, *Archegozetes* also received attention as a model system for chemical ecology (27, 85, 86, 88-91). Some of these studies focusing on the *Archegozetes* gland revealed basic insights into the chemical ecology and biochemical capabilities of arthropods (27, 89, 91). 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

133 events from bacteria and fungi that are mainly related to carbohydrate metabolism and cellulose
134 digestion, features correlated with the mite feeding biology. We also used the genomic data
135 together with stable-isotope labeling experiments and mass spectrometric investigation to
136 delineate the biosynthesis pathway of monoterpenes in oribatid mites.

137

Results and Discussion

Archegozetes longisetosus genome assembly

Archegozetes longisetosus (**Figure 1**) has a diploid chromosome number (2n) of 18 (36), most likely comprising 9 autosomal pairs, the typical number of nearly all studied oribatid mite species (92). There are no distinct sex chromosomes in *Archegozetes*; this appears to be ancestral in the Acariformes and persisted in the Oribatida (35, 36, 92). 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 (33, 35, 36, 92, 93). 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**) (94, 95). 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 (11-13, 15, 17, 18, 96). 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 (11, 38, 97, 98). 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**

160 **supplementary Table S1**). The final filtered genome assembly was composed of 1182 contigs
161 with an N_{50} contiguity of 994.5 kb (**Table 1**).

162 The official gene set and annotation of *Archegozetes*

163 We generated the official gene set (OGS) for *Archegozetes* by an automated, multi-stage
164 process combining *ab inito* and evidenced-based (RNAseq reads, transcriptomic data and curated
165 protein sequences) gene prediction approaches (see "**Materials and Methods**") yielding 23,825
166 gene models. In comparison to other mites and ticks as well as insects, this is well within the range
167 of the numbers discovered in other Chelicerata so far (**Figure 2a**). Chelicerates with a large OSG,
168 however, usually possess larger genomes (1-7 Gb), which suggests that *Archegozetes* may have a
169 relatively dense distribution of protein-coding genes in its genome. On the other hand, ticks can
170 have giga-base sized genomes, but only a rather small number of gene models, probably due to
171 high repetitive content (12, 99-101). Lacking more high-quality genomic resources of mites, it is
172 thus not clear whether the OGS of *Archegozetes* is the rule, or rather the exception within the
173 Oribatida.

174 To compare if *Archegozetes'* OSG is similar to predicted genes of other oribatid mites as
175 well as Prostigmata and Astigmata, we first clustered genes by ortholog inference (OrthoFinder;
176 (102)), removed species-specific genes and constructed a presence-absence matrix of orthogroups
177 to ordinate the data using non-metric multidimension scaling (NMDS, **Figure 2b**). Ordination
178 revealed that the OGS of *Archegozetes* is well nested with other oribatid mites and clearly
179 separated from their closest relative the astigmatid mites as well as prostigmatid mites (**Figure**
180 **2b**). As a first step in annotating the OSG, we ran KOALA (KEGG Orthology And Links
181 Annotation) to functionally characterize the genes (103). In total, 10,456 (43.9%) of all genes
182 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 (104, 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 (33, 105) 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 (39, 41, 106-115). 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 (22, 109, 110, 112-116). Overall, the Oribatid-Astigmatid relationship remains unresolved and a broader taxon sampling, especially of more basal Astigmata, will be necessary (22, 39, 111, 114-116). 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 (9, 22, 113, 116-120).

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 (121). 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 (18). The fraction of duplicated BUSCO genes in *Archegozetes* (4%) was similar to that of the spider mite and deer tick (11, 12), but very low compared to the house spider (**Figure 3c**), whose genome underwent an ancient whole-genome duplication (96).

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; 102) of 23 species (**Figure 3c; supplementary Figure S3**) representing Acariformes, Parasitiformes, several other chelicerates and the fly *Drosophila*.

As for most other species (2, 122), 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 (13, 15, 17). 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 (2, 102, 123). 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 (109, 112, 114, 115, 120), 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 (39, 111-113, 116).

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 (124-129). There is, however, no evidence for such an accumulation in oribatids or other arthropods (18). 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 (18). 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 (18), but recent studies suggest that sequencing technology, read depth and assembly quality are paramount to the capacity of identifying repeat content and TEs (130, 131). 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 (130).

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 (131-133). 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; (134)) with Tc1/mariner – a superfamily of interspersed repeats DNA transposons (131) – 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 (131), but also several of retrotransposons. Among these, is the Long Terminal

Repeat (LTR) gypsy retroelement (**Figure 4c**), which is closely related to retroviruses (131). 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 (135, 136). 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 (33).

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 (137, 138). Ancestrally, arthropods likely had ten Hox genes arranged in a cluster (139). 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 (139). 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 (139).

Mites largely lack overt, external signs of segmentation, other than the serially arranged appendages of the prosoma (140). Signs of segmentation in the posterior body tagma, the opisthosoma, do exist in adult members of Endeostigmata (141). However, these segmental boundaries are largely present only in the dorsal opisthosoma, making it difficult to assess how these correspond to the ventral somites (140, 141). 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 (11, 79-81). This stands in stark contrast to other studied chelicerate embryos. For example, during embryogenesis the spider *Parasteatoda tepidariorum* patterns twelve opisthosomal segments (142) and the opilionid *Phalangium opilio*

297 patterns seven (143). Furthermore, a member of Parasitiformes, the tick *Rhipicephalus microplus,*
298 appears to pattern eight opisthosomal segments during embryogenesis (144).

299 Parallel to the observation of segmental reduction in *T. urticae*, genomic evidence suggests
300 that this acariform mite has lost two of its Hox genes, *i.e., Hox3* and *abdominal-A (abd-A*) (11).
301 Interestingly, orthologs of *abd-A* in other studied arthropods pattern the posterior segments as well.
302 A genomic comparison of arthropod Hox clusters has also shown a correlation between
303 independent losses of *abd-A* and a reduction in posterior segmentation (145). To investigate
304 whether the loss of segmentation in *Archegozetes* is also due to an absence in *abd-A*, we annotated
305 its Hox cluster, paying close attention to the region between the Hox genes *Ultrabithorax* (*Ubx*)
306 and *Abdominal-B* (*Abd-B*), which is usually where this gene resides in other arthropods (139). Our
307 results suggest that the *Archegozetes* Hox genes are clustered in a contiguous sequence
308 (tig00005200_pilon, total size ~7.5 Mbp) in the same order as suggested for the ancestral arthropod
309 (89). Furthermore, we found no sequences suggestive of an *abd-A* ortholog in *Archegozetes*
310 (**Figure 5a**). These data also support the findings of a previous PCR survey that retrieved no *abd-*
311 *A* ortholog in *Archegozetes* (146). Genomic evidence from the Parasitiformes *Ixodes scapularis*
312 and *Metaseiulus occidentalis* reveal that these taxa maintain orthologs of all ten Hox genes,
313 however in *M. occidentalis* these genes are not clustered as they are in *I. scapularis* (12, 13).

314 Taken together, these observations suggest that the last common ancestor of acariform
315 mites likely lost its *abdominal-A* gene as well as experiencing a reduction in opisthosomal
316 segmentation (**Figure 5b**). Alternatively, these shared losses of *abd-A* may be due to convergence
317 due to similar selective pressures favoring a reduction in body size. The dorsal, external
318 segmentation of endeostigmatid mites does not necessarily contradict the hypothesis of a loss of
319 *abd-A* at the base of the acariform mites. As Hox genes are usually deployed after the genetic

establishment of segments in arthropods (139), 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 many developmental peculiarities. These peculiarities include the suppression of the fourth pair of walking legs during embryogenesis as well as the reduction of opisthosomal segmentation (79-82, 84, 147). 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 adult, for a total of six instars. (148). Thus far, methodological limitations have made it difficult to examine how mite segmentation and limb development progress throughout these instars.

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

342 Additionally, we found that earlier instars expressed a larger number of stage-specific genes as
343 compared to later instars and adults (**Figure 5c**).

344 Gene expression, SEM and time-lapse data have revealed that the development of the
345 fourth pair of walking legs in *Archegozetes* is suppressed until after the larval instar (80, 81, 147).
346 The resulting larva is thus hexapodal (see also embryo in **Figure 5e**), which constitutes a putative
347 synapomorphy of Acari, if they are monophyletic (9). In arthropods, the development of the limbs
348 is generally accomplished via the activity of highly conserved regulatory genes, termed the "limb
349 gap genes." These genes are expressed along their proximo-distal axes to establish the specific
350 identities of the limb podomeres. The limb gap genes include *extradenticle* (*exd*) and *homothorax*
351 (*hth*), which act together to specify the proximal limb podomeres, *dachshund* (*dac*), which
352 specifies the medial podomeres, and *Distal-less* (*Dll*) which specifies the distal-most podomeres.
353 It was previously shown that the deployment of these genes in the anterior appendages of
354 *Archegozetes*, *i.e.*, the chelicerae, pedipalps and first three pairs of walking legs (**Figure 5d** and
355 **e**), is similar to that of other chelicerate taxa (82, 142, 149). However, in the anlagen of the fourth
356 pair of walking legs, only the proximal-specifying genes, *exd* and *hth*, are expressed (82).

357 Whether the limb gap genes are re-deployed during the transition from the prelarval to
358 larval instars in order to activate the development of the fourth pair of walking legs remains an
359 open question. We therefore compared the average tpm values of verified limb gap genes (i.e., Al-
360 *Dll*, Al-*Hth*, Al-*exd*, and Al-*dac* (82)) in embryos and at each instar stage (**Figure 5f**). We also
361 compared the tpm values of the *Archegozetes* orthologs of *Sp6-9* and *optomotor blind*, genes
362 shown to be involved in limb formation in spiders (150, 151). We hypothesized that limb
363 development genes would show high expression in the larval stage leading to the development of
364 the octopodal protonymph. We did observe an increase in the tpm averages of Al-*hth* as well as

365 Al-*optomotor-blind*, however the aforementioned limb gap gene expression levels were similar
366 between these instars (**Figure 5f**). Taken together, these genes may not be up-regulated for the
367 formation of the fourth pair of walking legs between these two instars.

368 Chelicerate embryos segment their bodies through a "short/intermediate germ"
369 mechanism, whereby the anterior (prosomal) segments are specified asynchronously (142). This
370 usually occurs well before the sequential addition of posterior segments from a posterior growth
371 zone. Based on neontological and paleontological data, chelicerate arthropods may have
372 ancestrally had an opisthosoma comprised of 12 or more segments (5, 7, 140). Embryonic
373 expression data for the segment polarity genes, those genes that delineate the boundaries of the
374 final body segments, have shown that in most studied chelicerate embryos opisthosomal segments
375 are delineated during embryogenesis (140, 142). However, as discussed above, expression data in
376 *Archegozetes* embryos suggest that only two opisthosomal segments are patterned during
377 embryogenesis (80, 81); this indicates that mites have significantly reduced their number of
378 opisthosomal segments either by loss or by fusion. Further complicating this is the observation
379 that many mites add segments as they progress through the larval instars, a phenomenon known
380 as anamorphic growth (140).

381 To determine by what genetic process *Archegozetes* may add segments during post-
382 embryonic ontogeny, we assessed the expression of known chelicerate and arthropod segmentation
383 genes in each instar transcriptome (**Figure 5f**) (142). We observed an up-regulation of the
384 segmentation genes *hedgehog* and *engrailed* in the larvae, as well as the slight up-regulation of
385 *patched* and *pax3/7*. Furthermore, the segmentation gene *wingless* was slightly up-regulated in the
386 protonymph, as well as a slight up-regulation of *hedgehog* in the tritonymph. Lastly, we found that
387 transcripts of the genes *pax3/7* and *runt* were up-regulated in adults. These results suggest that

388 *Archegozetes* does pattern body segments during the progression through the it's instars similar to
389 other Chelicerata.

390 Another peculiarity of *Archegozetes* is that these mites lack eyes (see more details below).
391 Eye loss has been documented in other arachnid clades, including independently in other members
392 of Acari (10, 152), and it has been recently demonstrated that a species of whip spider has reduced
393 its eyes by reducing the expression of retinal determination genes that are shared throughout
394 arthropods (153). We sought to determine if eye loss in *Archegozetes* also is associated with the
395 reduced expression of these genes (see also analysis of photoreceptor genes below). The genes,
396 which have been shown to be expressed in the developing eyes of spiders and whip scorpions,
397 include *Pax-6*, *six1/sine oculis* (*so*), *eyes absent* (*eya*), *Eyegone*, *Six3*/*Optix*, and *atonal* (153-155).
398 We also followed the expression of Al-*orthodenticle*, a gene previously shown to be expressed in
399 the ocular segment of *Archegozetes* (147). Surprisingly, all of these genes, excluding the *Pax-6*
400 isoform A and *eyegone*, are indeed expressed during embryogenesis (**Figure 5f**). Aside from the
401 larval expression of the *Pax-6* isoform A during the larval stage, these eye-development genes
402 remain quiescent until the adult stage, where all but *Pax-6* isoform A, *six3* and *atonal* are up-
403 regulated (**Figure 5f**). These results are exceedingly surprising, given the conserved role of genes
404 in retinal patterning. They suggest a novel role for these genes, or alternatively, these expression
405 patterns could be the result of early expression of a retinal determination pathway followed by
406 negative regulation by other genes to suppress eye development.

407 <u>Photoreceptor and chemosensory system of *Archegozetes longisetosus*</u>

408 Unlike insects and crustaceans, chelicerates do not have compounds eyes – with horseshoe
409 crab being an exception. Generally, mites are eyeless or possess one or two pairs of simple ocelli
410 (156-160). Ocelli are common in Prostigmata and Endeostigmata, among Acariformes, as well

411 Opilioacarida – the most likely sister group to the Parasitiformes – but are absent in most Oribatida,
412 Astigmata, Mesostigmata and ticks (10, 161-163). This suggests that the presence of eyes might
413 be an ancestral condition for both Acariformes and Parasitiformes, while more derived mites rely
414 largely on chemical communication systems (156).

415 In oribatid mites, detailed morphological and ultrastructural investigations have suggested
416 that setiform sensilla are the most obvious sensory structures (**Figure 6a**) (10, 156, 164). The
417 trichobothria are very complex, highly modified (e.g. filiform, ciliate, pectinate, variously
418 thickened or clubbed) no-pore setae which are anchored in a cup-like base and likely serve as
419 mechanosensory structures. In contrast, the setal shafts of solenidia and eupathidia (**Figure 6a**)
420 both possess pores (10, 156, 164). Solenidia have transverse rows of small pores visible under a
421 light microscope and likely function in olfaction, while the eupathidia have one or several terminal
422 pores and likely are used as contact/gustatory sensilla (**Figure 6a**) (156, 164). Previous work
423 demonstrated that oribatid mites indeed use olfactory signals in the context of chemical
424 communication and food selection (47-49, 58, 64, 86).

425 Interestingly, detailed morphological and ultrastructural studies showed that light-sensitive
426 organs exist in some Palaeosomata and Enarthronota (probably true eyes) as well as in
427 Brachypylina (the secondary lenticulus), representing lower and highly derived oribatid mites,
428 respectively (156, 160-162). *Archegozetes* and most other oribatids, however, are eyeless, yet there
429 is scattered experimental and some anecdotal evidence that even these mites show some response
430 to light and seem to avoid it ('negative phototropism' or 'negative phototaxis') (10, 165-167).
431 Hence, we mined the genome of *Archegozetes* for potential photoreceptor genes and found two
432 genes of the *all-trans retinal perosin* class and one gene related to spider mite *rhodopsin-7-like*
433 gene (**Figure 6b**). *Perosin-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 (168-173). Taken together, this might suggest that eyeless species like *Archegozetes* use *perosin*- 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 (10, 49, 58, 64, 156, 164, 173). 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**) (13, 15, 17, 30, 31, 174-176). 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**) (177). 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 (175-179). 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**) (180-182). 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 (183). 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 (12, 13, 15, 17, 101, 177). 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 (12, 49, 58, 64, 101, 184-186).

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 (40, 156, 164). 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;** (187)) and here we provide the first genomic resource for the investigation of the photo- and chemosensory systems of Oribatida (**Figure 6b-d**).

Horizontal gene transfer event sheds light on oribatid feeding biology

Horizontal gene transfer (HGT) is common among mites and other soil organisms (11, 17, 74-77). In some cases, genes that had been horizontally transferred now have pivotal biological functions. For instance, terpene and carotenoid biosynthesis genes in trombidiid and tetranychid mites, respectively, are found nowhere else in the animal kingdom (17, 188). Yet they show high homology with bacterial (terpene synthase) or fungal (carotenoid cyclase/synthase/desaturase) genes, suggesting horizontal gene transfer from microbial donors (17, 188). At least the carotenoid biosynthesis genes in spider mites still code for functional enzymes and equip these phytophages with the ability to *de novo* synthesize carotenoids, which can induce diapause in these animals (188).

Soil microarthropods like collembolans show numbers of horizontally transferred genes that are among the highest found in Metazoan genomes, exceeded only by nematodes living in decaying organic matter (76, 77, 189). Interestingly, many HGT genes found in collembolans are involved in carbohydrate metabolism and were especially enriched for enzyme families like glycoside hydrolases, carbohydrate esterases or glycosyltransferases (76, 77). All three enzyme families are involved in the degradation of plant and fungal cell walls (190, 191). Hence, it has been hypothesized that cell-wall degrading enzymes acquired by HGT are beneficial for soil organisms as it allowed such animals to access important food source in a habitat that is highly biased towards polysaccharide-rich resources (76, 77, 192, 193).

To assess the degree of HGT in *Archegozetes* we first used blobtools (v1.0) (194) to generate a GC proportion vs read coverage plot of our genome assembly, in order to remove scaffolds of bacterial origin (**Figure 7a**; 438 contigs of ~ 8.5 Mb of contamination). Of the remaining scaffolds, candidate HGTs were identified using the Alien Index (195, 196), where HGTs are those genes with blast homology (bit score) closer to non-metazoan than metazoan

503 sequences (**supplementary Table S4**). We further filtered these HGT candidates to remove those
504 that overlapped predicted repeats by ≥ 50%, resulting in 617 genes. As HGT become integrated
505 into the host genome, they begin to mirror features of the host genome, including changes in GC
506 content and introduction of introns (197). Comparing the GC content of the HGT candidates
507 showed two distinct peaks, one at 54.3% and the other at 35.2%, slightly higher than the remaining
508 *Archegozetes* genes, GC content of 31.5% (**Figure 7b**). Of the 407 HGT genes that shared similar
509 GC content to the host genome, 73.5% had at least one intron (**Table S4**). In a final step, we used
510 the gene expression data (RNAseq) to filter the list of all putative HGT genes and only retained
511 candidates that were expressed in any life stage of *Archegozetes* (*n*= 298 HGT genes).

512 The majority of HGT candidates were of bacterial origin (75.2%), followed by genes likely
513 acquired from fungi (13.4%), while transfer from Archaea, plants, virus and other sources was
514 comparatively low (**Figure 7d**). This composition of HGT taxonomic origin is different from
515 genes found in collembolans, which appear to have acquired more genes of fungal and protist
516 origin (76, 77, 192). Subsequently, we performed an over-representation analysis of GO terms
517 associated with these genes. We found an over-representation of genes with GO terms related to
518 methyl transfer reactions and breaking glycosidic bonds (molecular function; **Figure 7c**) as well
519 as carbohydrate metabolism, among others (biological process; **Figure 7c**) providing a first line of
520 evidence that *Archegozetes* possess HGT related to plant- and fungal cell wall degradation similar
521 to springtails.

522 As mentioned previously, oribatid mites are among the few Chelicerata that ingest solid
523 food and are primary- and secondary decomposers feeding on dead plant material and fungi (9, 40,
524 55-57, 63). It was argued for decades that the enzymes necessary to break down these
525 polysaccharide-rich resources originate from the mite's gut microbes (45, 65-67, 71, 198, 199).

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) (198, 199). 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 (72), 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 (72, 200). They showed that phytophagous beetles likely acquired all genes of the *GH 48 family* from Actinobacteria (including *Streptomyces*) (72) and our phylogenetic analysis (**Figure 7e**) revealed the same pattern as well as a highly similar tree topology (compare to Fig 3B in (72)).

549 Overall, our findings indicate that genes encoding for enzymes in *Archegozetes* capable of
550 degrading plant and fungal cell walls were likely horizontally transferred from bacteria (likely
551 *Streptomyces*). Bacterial symbionts and commensal living in the mites' gut are still likely to
552 contribute to the breakdown of food (**Figure 7f**). Yet, the high expression of genes encoding cell-
553 wall degrading enzymes (**Figure 7g**) as well as the evolutionary analyses of such genes (**Figure**
554 **7e**) suggest that *Archegozetes* – and potentially many other oribatid mites – are able to exploit
555 polysaccharide-rich resources like dead plant material or chitinous fungi without microbial aid.
556 Enzymological and microscopical investigation of *Archegozetes* have suggested that certain
557 digestive enzymes (chitinase and cellulase) are only active when the mites consume a particular
558 type of food (e.g. algae, fungi or filter paper) (71). These results were interpreted as evidence that
559 these enzymes are directly derived from the consumed food source (71). By contrast, we argue
560 that this instead confirms our findings of HGT: upon consumption of food containing either chitin
561 or cellulose, gene expression of polysaccharide-degrading enzymes starts, and proteins can readily
562 be detected. Further enzymological studies have placed oribatid mites in feeding guilds based on
563 carbohydrase activity and also found highly similar enzyme activity between samples of mites
564 from different times and locations (67, 68, 201-203). Future functional studies can disentangle the
565 contribution of the host and microbes to cell wall digestion and novel metabolic roles of the HGTs
566 identified here.

567 <u>Biosynthesis of monoterpenes – a common chemical defense compound class across oribatid mite</u>

568 Oribatid and astigmatid mites are characterized by a highly diverse spectrum of natural
569 compounds that are produced by and stored in so-called oil glands (for an example see **Figure 8a**)
570 (30, 88, 204). These paired glands are located in the opisthosoma (i.e. the posterior part of
571 chelicerate arthropods, analogous to the abdomen of insects) and are composed of a single-cell

572 layer invagination of the cuticle (**Figure 8f**). As previously mentioned, mites use chemicals
573 produced by these glands to protect themselves against environmental antagonists (predators or
574 microbes) or use them as pheromones (30, 47-49, 52, 86, 89). The monoterpene aldehyde citral –
575 a stereoisomeric mixture of geranial ((*E*)-3,7-dimethylocta-2,6-dienal) and neral ((*Z*)-3,7-
576 dimethylocta-2,6-dienal) – and its derivatives are widely detected compounds in glandular
577 secretions of oribatids and astigmatids (41, 107, 184, 205-209). These monoterpenes have been
578 called "astigmatid compounds" (41) as they characterize the biochemical evolutionary lineage of
579 major oribatid mite taxa (Mixonomata and Desmonomata) and almost all investigated astigmatid
580 mites (30, 41, 118, 184, 208).

581 The chemical cocktail released by *Archegozetes* consists of a blend of 10 compounds
582 (**Figure 8a**) including two terpenes (approx. 45%) – neral and neryl formate – six hydrocarbons
583 (approx. 15%) and two aromatic compounds (approx. 40%) (206, 210). The hydrocarbons likely
584 serve as solvents, while the terpenes and aromatics are bioactive compounds used in chemical
585 alarm and defense (47-49, 206). Recently, it was shown that *Archegozetes* synthesizes the two
586 aromatic compounds using a polyketide-like head-to-tail condensation of (poly)-β-carbonyls *via* a
587 horizontally acquired putative polyketide synthetase (27). Studies in Astigmata found that the
588 monoterpenes of these mites appeared to be made *de novo* from (poly)-β-carbonyls as well and
589 one study identified a novel geraniol dehydrogenases (GeDH), unrelated to those of bacteria, in
590 *Carpoglyphus lactis* (211-213). To learn about the biosynthesis of astigmatid compounds in
591 *Archegozetes* and demonstrate the mite's applicability as research model for biochemical pathway
592 evolution, we used the novel genomic resources presented in this study.

593 First, we delineated the basic biochemical reactions likely to happen in the *Archegozetes*
594 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 (27)). 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 (214) 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 prostists (215-220). 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 (218-220). 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 (212). We used the functionally validated Carpoglyphus-GeDH (212), 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

618 ADHs and constructed a maximum likelihood phylogeny (**Figure 8e**). Similar to the previous
619 analysis including only Carpoglyphus-GeDH, we found that the *Al-GeDH* represent a new class
620 of geraniol dehydrogenases different from those in plants, fungi or bacteria and not nested within
621 animal ADHs (**Figure 8e**). This is why we hypothesize that *Al-GeDH* is a novel expansion of the
622 geraniol dehydrogenases gene family and has not been acquired by horizontal gene transfer, like
623 other biosynthesis and digestive enzymes in *Archegozetes* (**Figure 7**; (27)).

624 Based on our mass spectrometry data of stable isotopes and genomic analysis, we propose
625 that the following biochemical pathway leading to monoterpenes is of oribatid mites (**Figure 8f**
626 **and g**): geraniol is likely to be synthesized from GPP – the universal precursor of all monoterpenes
627 – either enzymatically by a geraniol synthase (GES) or a diphosphate phosphatase (DPP), but
628 possibly also endogenously by dephosphorylation of GPP (221-224). For *Archegozetes*, we could
629 not find any *GES* or specific *DPP* in the OGS, thus geraniol might be formed from GPP *via*
630 endogenous dephosphorylation, but further research is required to verify or falsify this hypothesis.
631 Subsequently, geraniol is oxidized to geranial by the pervious described *GeDH* (**Figure 8e**) and
632 readily isomerized to neral. Trace amounts of geranial have been found in *Archegozetes* and it is
633 common among other oribatid and astigmatid mites, supporting this idea (107, 184, 209, 225).
634 Also, there is no evidence that geraniol is cconverted into nerol, or that neral is formed directly via
635 oxidation of nerol (211-213). The most parsimonious explanation for neryl formate synthesis
636 would be an esterification of the corresponding terpene alcohol nerol. There is, however, no
637 evidence of nerol in the traces of any oribatid or astigmatid mite species (30, 88, 184). Aliphatic
638 non-terpene formats in Astigmata are synthesized by dehomologation and generation of a one-
639 carbon–shorter primary alcohol from an aldehyde via hydrolysis of formate in a biological Baeyer–
640 Villiger oxidation catalyzed by a novel, uncharacterized enzyme (226). A similar reaction to

641 synthesize terpene formates is unlikely, as the terpenoid backbone would be shortened by one-
642 carbon and this does not happen in any possible scenario. The discovery of this Baeyer–Villiger
643 oxidation mechanism, however, highlights the probability that there are many very unusual
644 reactions that remain to be discovered in oribatid mites (27).

Conclusion

The integrated genomic and transcriptomic resources presented here for *Archegozetes longisetosus* allowed a number of insights into the molecular evolution and basic biology of decomposer soil mites. Our analysis of an oribatid mite genome also provides the foundation for experimental studies building on the long history of *Archegozetes'* as a chelicerate model organism, which now enters the molecular genetics era (23, 33, 34, 92). This includes the study of biochemical pathways, biochemistry, neuroethological bases of food searching behavior, and environmental impacts on genomes of complex, clonal organisms.

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

Key priorities for a future community research effort include i) sequencing organ-specific transcriptomic data, ii) developing tools for genetic interrogation (RNAi or CRISPR/CAS9), iii) establishing reporter linages with germ-line stable modifications (e.g. GAL4/UAS misexpression systems), iv) constructing an whole-animal single-cell RNAseq expression atlas, and v) gathering more genomic data to improve the genome assembly. Please do not hesitate to contact the corresponding author, if you want to start your own culture of *Archegozetes*. He will be happy to provide you with starter specimens and share rearing protocols with you.

667 **Materials and Methods**

668 Mite husbandry

669 The lineage 'ran' (33) of the pantropical, parthenogenetic oribatid mite *Archegozetes*
670 *longisetosus* was used in this study. Stock cultures were established in 2015 from an already
671 existing line and fed with wheat grass (*Triticum* sp.) powder from Naturya. Cultures were
672 maintained at 20-24°C and 90% relative humidity. Sterilized water and 3-5 mg wheat grass were
673 provided three times each week.

674 DNA extraction and Illumina sequencing

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

685 High-molecular weight DNA isolation and Nanopore sequencing

686 Genomic DNA was isolated from ~300-500 mites starved for 24 h using QIAGEN Blood
687 & Cell Culture DNA Mini Kit. Briefly, mites were flash frozen in liquid nitrogen and homogenized
688 with a pestle in 1 ml of buffer G2 supplemented with RNase A and Proteinase K at final

689 concentrations of 200 ng/µl and 1 µg/µl, respectively. Lysates were incubated at 50°C for 2 h,
690 cleared by centrifugation at 5 krpm for 5 min at room temperature and applied to Genomic tip
691 G/20 equilibrated with buffer QBT. Columns were washed with 4 ml of buffer QC and genomic
692 DNA was eluted with 2 ml of buffer QF. DNA was precipitated with isopropanol, washed with
693 70% EtOH and resuspended in 50 µl of buffer EB. DNA was quantified with Qubit dsDNA HS
694 Kit (ThermoFisher) and the absence of contaminants was confirmed with NanoDrop One
695 (ThermoFisher) with target OD 260/280 and OD 260/230 ratios of 1.8 and 2.0-2.2, respectively.
696 DNA integrity was assessed using Genomic DNA ScreenTape kit for TapeStation (Agilent
697 Technologies).

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

711 Genome assembly and contamination filtering

712 Read quality was assessed using FastQC v0.11.8 (227). Illumina adapters, low-quality
713 nucleotide bases (phred score below 15) from the 3' and 5' ends and reads shorter than 50 bp were
714 removed using cutadapt v1.18 (228). From the filtered reads, *in silico* genome size estimates were
715 calculated using *k-mer* based tools kmergenie v.1.7048 (229), GenomeScope v1.0 (230), and
716 findGSE v0.1.0 R package (231). The latter two required a *k-mer* histogram computed by jellyfish
717 v2.2.10 (232) with *k-mer* size of 21. The genome was assembled using 4.7 million long reads from
718 two MinION runs (60x coverage) using Canu v1.8 with default settings and setting the expected
719 genome size to 200 Mb (233). To improve assembly quality, paired end Illumina reads were
720 mapped to the genome with BWA aligner (234) using BWA-MEM algorithm and polished with
721 Pilon v. 1.23 with '—changes' and '--fix all' options (235). Assembled contigs identified as
722 bacterial and fungal contaminants based on divergent GC content from most *Archegozetes* contigs,
723 high coverage and blast homology to the nt database (downloaded February 2019, Evalue $1e^{-25}$)
724 were removed using Blobtools v1.0 (194).

725 Identification, classification and masking of repetitive element

726 Repetitive elements in the genome *Archegozetes* were identified using a species-specific
727 library generated with RepeatModeler v 1.0.11 (236, 237) and MITE tracker (132) and annotated
728 by RepeatClassifier, a utility of the RepeatModeler software that uses the RepBase database
729 (version Dfam_Consensus-20181026). Unclassified repeat families from both programs were run
730 through CENSOR v 4.2.29 (238) executable *censor.ncbi* against the invertebrate library v 19.03
731 to provide further annotation. Predicted repeats were removed if they had significant blast
732 homology (E-value $1e^{-5}$) to genuine proteins in the NCBI nr database and/or a local database of
733 arthropod genomes (*Drosophila melanogaster, Tribolium castaneum, Tetranychus urticae,*
734 *Leptotrombidium deliense, Dinothrombium tinctorium, Sarcoptes scabiei, Euroglyphus maynei,*

735 *Galendromus occidentalis, Dermatophagoides pteronyssinus*). Unclassified repeats with blast
736 homology to known TEs were retained whereas those with no blast homology were removed (239).
737 The remaining repeat families were combined with the Arthropoda sequences in RepBase and
738 clustered using vsearch v 2.7.1 (--iddef 1 --id 0.8 --strand both; (240)). The filtered repeat library
739 was used to soft mask the *A. longisetosus* using RepeatMasker v 4.07 (241).A summary of the
740 masked repeat content was generated using the "buildSummary.pl" script, the Kimura sequence
741 divergence calculated using the "calcDivergenceFromAlign.pl" script and the repeat landscape
742 visualized using the "createRepeatLandscape.pl" script, all utilities of RepeatMasker.

743 Gene prediction and annotation

744 Both *ab inito* and reference-based tools were used for gene prediction using modified steps
745 of the funannotate pipeline (104). The *ab inito* tool GeneMark-ES v4.33 (242) was used along with
746 reference based tools BRAKER v2.1.2 (243) using RNAseq reads discussed below and PASA v
747 2.3.3 (244) using genome-guided transcriptome assembly from Trinity described below. Lastly,
748 *Tetranychus urticae* gene models from the NCBI database (GCF_000239435.1) were aligned to
749 the contigs using GeMoMa (245). All gene predictions were combined in EvidenceModeler (244)
750 with the following weights: GeMoMa =1, PASA = 10, other BRAKER = 1, and GeneMark = 1.
751 Predicted tRNAs using tRNAscan-SE v 2.0.3 (246) were combined with the gene predictions in
752 the final gene feature format (GFF) file and filtered for overlap using bedtools (247) *intersect* tool
753 (247).

754 The predicted genes were searched against the NCBI nr (February 2019) (248), SwissProt
755 (February 2019) (249), a custom-made Chelicerata database including genomes of *Tetranychus*
756 *urticae, Leptotrombidium deliense, Dinothrombium tinctorium, Sarcoptes scabiei, Euroglyphus*
757 *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) (250), merops (v 12, October 2017) (251) and CAZY (v 7, August 2018) (252) databases. The results of the hmm-based (253) 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 (254). 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 (255).

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) (256, 257). 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 (248, 249). A custom database of GO terms was created with makeOrgPackage function in the R package AnnotationForge v1.26.0 (258). Over-representation analysis of GO terms was tested using the enrichGO function in the R package clusterProfiler v3.12.0 (259) with a hypergeometric distribution and a Fisher's Exact test. P-values were adjusted for multiple comparisons using false discovery rate correction (260). Resulting enriched GO terms were processed with GO slim (261) and the final list of over represented GO terms was used to plot the number of genes in a respective category.

780 KEGG orthology terms were assigned from single-directional best hit BLAST searches of
781 each gene model on the KEGG Automatic Annotation Server (262). Additionally, we ran
782 GhostKOALA (103) (GHOSTX searches for KEGG Orthology And Links Annotation) to obtain
783 KEGG orthology terms. Compared to conventional BLAST searches, GhostKOALA is about 100
784 times more efficient than BLAST to remote homologs by using suffix arrays (263).

785 Orthology and Phylogenomic analyses

786 Orthologs of *A. longisetosus*, other species within Acari, Chelicerata and the fruit fly
787 *Drosophila* were identified using OrthoFinder v 2.3.3 (-M msa –A mafft –T fasttree; (102)). Prior
788 to running OrthoFinder, isoform variants were removed from the gene predictions using CD-Hit
789 (264). Trees of orthogroups with at least 80% of taxa present (n= 4,553) were constructed using
790 fasttree v 2.1.10 (265), trimmed with TrimAl v 1.4.1 (-keepheader -fasta –gappyout; (266)) and
791 paralogs pruned using phylotreepruner v 1.0 (min_number_of_taxa =18, bootstrap_cutoff= 0.7,
792 longest sequence for a given orthogroup=u; (267)). Alignments shorter than 100 amino acids were
793 removed, leaving 1,121 orthogroups.

794 For the maximum likelihood analysis, the trimmed and pruned alignments were
795 concatenated into a supermatrix using FasConCat v1.04 (268) composed of 377,532 amino acids
796 and the best substitution models determined using PartitionFinder v 2.1.1 (269). The maximum
797 likelihood consensus phylogeny from the supermatrix and partition scheme was constructed using
798 IQ-tree and 1,000 ultrafast bootstrap replicates (270). For the coalescence species tree
799 reconstruction, gene trees were generated using IQ-tree v 1.6.12 on the trimmed alignments of the
800 1,121 filtered orthogroups and processed using ASTRAL v 5.6.3 (271). Branch lengths are
801 presented in coalescent units (differences in the 1,121 gene trees) and the node values reflect the
802 local posterior probabilities.

803 RNA sequencing and transcriptome assembly

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

809 Extracted RNA was shipped to Omega Bioservices (Norcross, GA, USA) on dry ice for
810 library preparation and sequencing. Whole animal RNA was used for poly-A selection, cDNA
811 synthesis and library preparation following the Illumina TruSeq mRNA Stranded Kit protocol. The
812 library was sequenced with 100 million 150 bp paired-end on a HighSeq4000 platform. For the
813 genome-guided assembly of the transcriptome a bam-file was created from the genome using
814 STAR (272). RNAseq reads were *in silico* normalized and subsequently used together with the
815 bam-file to assemble the transcripts using Trinity v2.8.4 (273, 274), yielding an assembly with a
816 total length of 162.8 Mbp, an N50= 2994 bp and a BUSCO score (121) of C:96.3%
817 [S:36.5%,D:59.8%], F:1.3%, M:2.4%.

818 Life-stage specific RNAseq

819 For life-stage specific RNAseq, we collected 15 specimens per life stage from the stock
820 culture that were split into three replicates of five individuals. Whole animals (for all stages but
821 eggs) were flash frozen in 50 µl TRIzol using a mixture of dry ice and ethanol (100%) and stored
822 at −80°. RNA was extracted using a combination of the TRIzol RNA isolation protocol (Life
823 Technologies) and RNeasy Mini Kit (Qiagen) (275). The TRIzol protocol was used for initial steps
824 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 (276) 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 (277) using Likelihood Ratio tests in R 3.6.1 (278). 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 (279) to produce an unclustered heatmap showing relative expression levels.

847 UpSetR (280) was used to compare the number of unique and shared expressed genes across life
848 stages.

849 Horizontal gene transfer events identification

850 To detect HGTs, we used the published tool "./Lateral_gene_transfer_predictor.py" (196)
851 to calculate the Alien Index described by (281) and (195). All predicted genes were compared to
852 the NCBI nr database as previously described (196). Results to Arthropoda (tax id 6656) were
853 ignored in the downstream calculations. The HGT candidates were filtered for contamination
854 identified by both Blobtools (194) and the Alien Index (AI > 30 and >70% percent identity to a
855 non-metazon sequence). The candidates were further filtered for > 50% overlap with predicted
856 repeats using the bedtools intersect tool with the RepeatMasker gff file and expression from any
857 developmental stage. Introns were scored manually from visualization in IGV genome browser
858 (282) and GC content for all predicted genes was calculated using the bedtools nuc tool.

859 Analysis of chemosensory and photoreceptor gene families

860 The search and analysis chemosensory genes largely followed the procedure outlined by
861 Dong et al. (17) with slight modifications. First, the *Archegozetes* official gene set (OGS) was
862 searched using BLASTP (E-value, $<1 \times 10^{-3}$) against the following queries for the different
863 chemosensory gene families. The OGS was queried against i) *D. melanogaster*, *D. mojavensis*,
864 *Anopheles gambiae*, *Bombyx mori*, *T. castaneum*, *Apis mellifera*, *Pediculus humanus humanus*,
865 and *Acyrthosiphon pisum* odorant binding proteins (OBPs) (174); ii) *D. melanogaster*, *D.*
866 *mojavensis*, *A. gambiae*, *B. mori*, *T. castaneum*, *A. mellifera*, *P. humanus humanus*, *A. pisum, I.*
867 *scapularis*, and *Daphnia pulex* small chemosensory proteins (CSP) (174, 283, 284); iii) *D.*
868 *melanogaster* and *A. mellifera* odorant receptors (283, 284); iv) *D. melanogaster*, *A. mellifera*, *I.*

869 *scapularis*, *T. urticae*, *T. mercedesae*, and *M. occidentalis* gustatory receptors (GRs) (12, 13, 15,
870 177, 284, 285); v) a comprehensive list of iGluRs and IRs across vertebrates and invertebrates
871 (178), as well as those identified in the *T. mercedesae, D. tinctorium* and *L. deliense* genome
872 projects (15, 17). Second, all candidate *Archegozetes* sequences were reciprocally blasted
873 (BLASTP, E-value $<1 \times 10^{-3}$) against the NCBI database (248) and all sequences that did not hit
874 one of the respective receptors or transmembrane proteins were removed from the list. Third, for
875 phylogenetic analysis of IRs and GRs from *Archegozetes* were aligned with IRs from *D.*
876 *melanogaster*, *T. urticae, D. tinctorium* and *L. deliense* and GRs from iv) *D. melanogaster, T.*
877 *mercedesae*, *I. scapularis*, and *M. occidentalis*, respectively, using MAFFT (v 7.012b) with default
878 settings (286). Poorly aligned and variable terminal regions, as well as several internal regions of
879 highly variable sequences were excluded from the phylogenetic analysis. Fourth, maximum
880 likelihood trees were constructed with the IQ-TREE pipeline (v 1.6.12) with automated model
881 selection using 1,000 ultrafast bootstrap runs (270).

882 Reference opsin genes and opsin-like sequences were obtained from Dong et al. (17) and
883 used to query the *Archegozetes* OGS using BLASTP (E-value, $<1 \times 10^{-5}$). Subsequently,
884 candidates sequenced were reciprocally blasted against NCBI using the same settings and only
885 retained if they hit an opsin or opsin-like gene. The *Archegozetes* candidates were aligned with the
886 query sequence list using MAFFT (v 7.012b) with default settings (286). This opsin gene
887 alignment phylogenetically analyzed using the IQ-TREE pipeline (v 1.6.12) with automated model
888 selection and 1,000 ultrafast bootstrap runs (270).

889 Gene family phylogenies

890 We used the following workflow to analyses genes related to Figure 5 (hox and
891 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 (248), and aligned using MUSCLE (287) or MAFFT (v 7.012b) (286) 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 (288, 289) or ii) the IQ-TREE pipeline with automated model selection (270). 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 (27). 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 (26, 210, 225, 290).

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. AB is Simons Fellow of the Life Sciences Research Foundation (LSRF). This project was supported by a grant from the Caltech Center for Environmental Microbial Interactions (CEMI) to AB. 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.

Ethics statement

There are no legal restrictions on working with mites.

Authors contributions

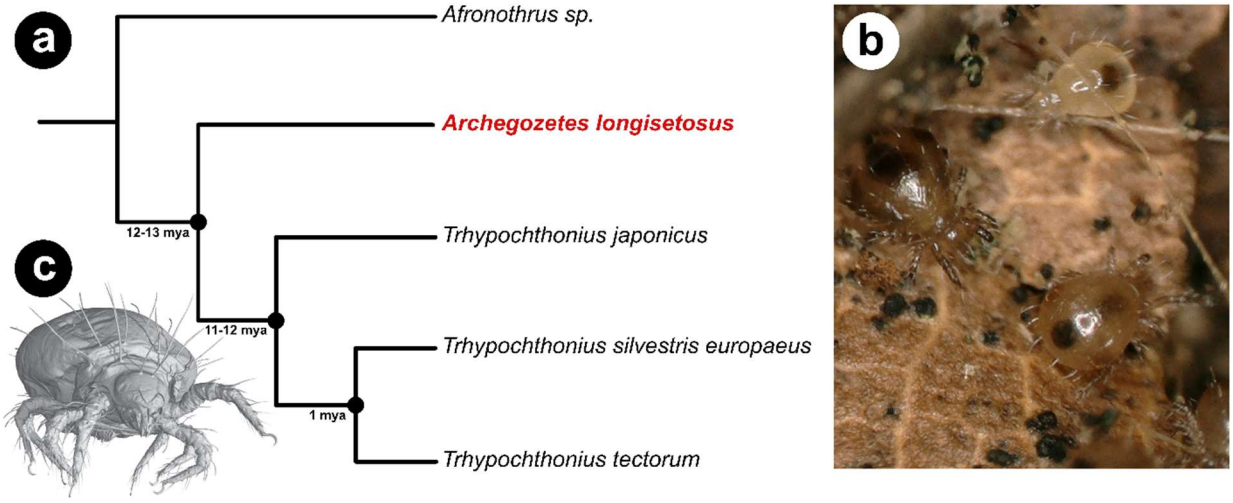
All authors gave final approval for publication.

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.



936

937 **Figure 1** The mite, *Archegozetes longisetosus*, in its phylogenetic and natural environment. **a**:
938 Species tree of selected oribatid mites of the family Trhypochthoniidae based on phylogenetic
939 analyses and divergence time estimates by (291). **b**: Two adults and one tritonymph of
940 *Archegozetes* on a piece of leaf litter. The algae growing on the leaf serves as a food source for the
941 mites. **c**: Habitus of an adult mite based on a surface rendering of a µCT-scan reconstruction. Image
942 courtesy of Sebastian Schmelzle.

943

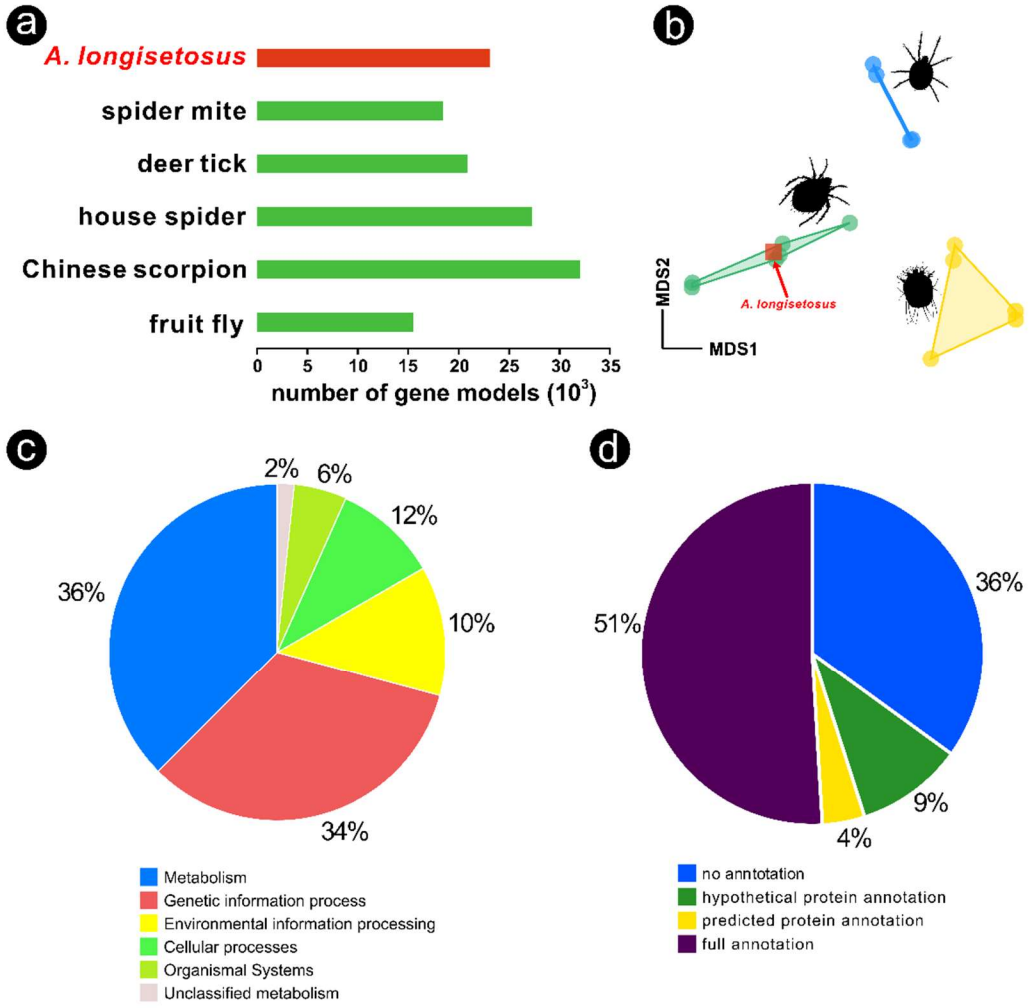
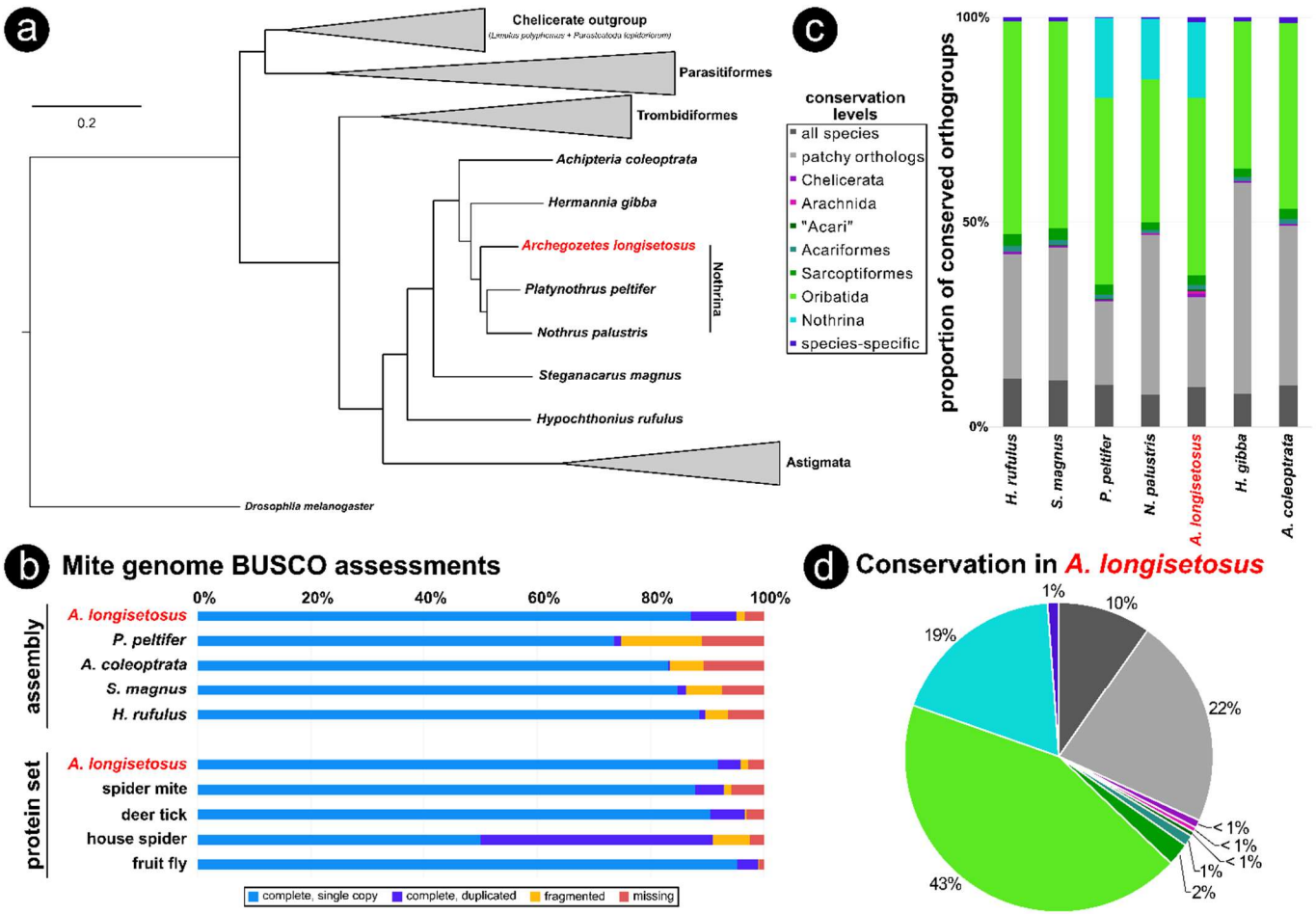
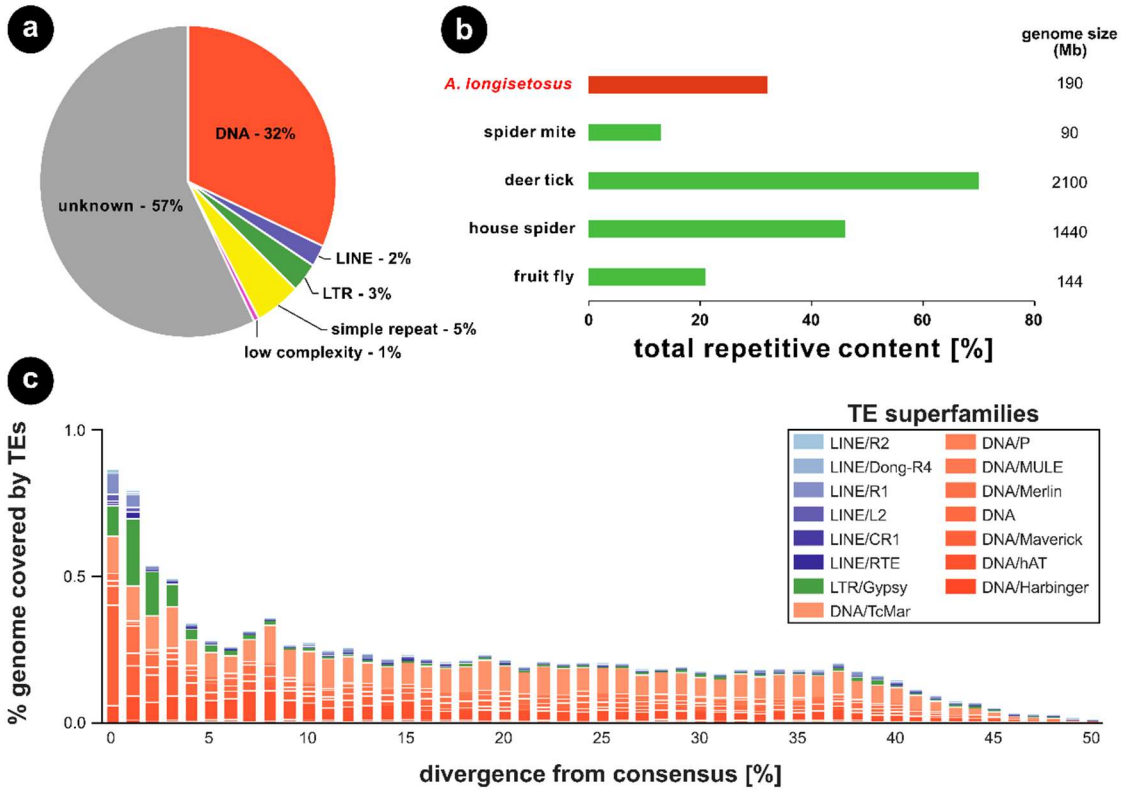


Figure 2 Comparisons and annotations of the official gene set (OGS) of *Archegozetes longisetosus*. **a**: Number of gene models of the mites compared to other mites, chelicerates and the fruit fly (11, 12, 96, 97, 292). **b**: Non-linear multidimensional scaling plot (NMDS) of clustered orthogroups 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.



953 **Figure 3** Orthology comparison and phylogenetic placement of *Archegozetes longisetosus* among other chelicerates. **a**: Maximum
954 likelihood phylogeny based on concatenation of 1,121 orthologs showing the mites phylogenetic environment within the Oribatida (all
955 nodes have 100% support; branch length unit is substitutions per site). For the fully expanded tree see supplementary Figure S2. **b**:
956 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***.**



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

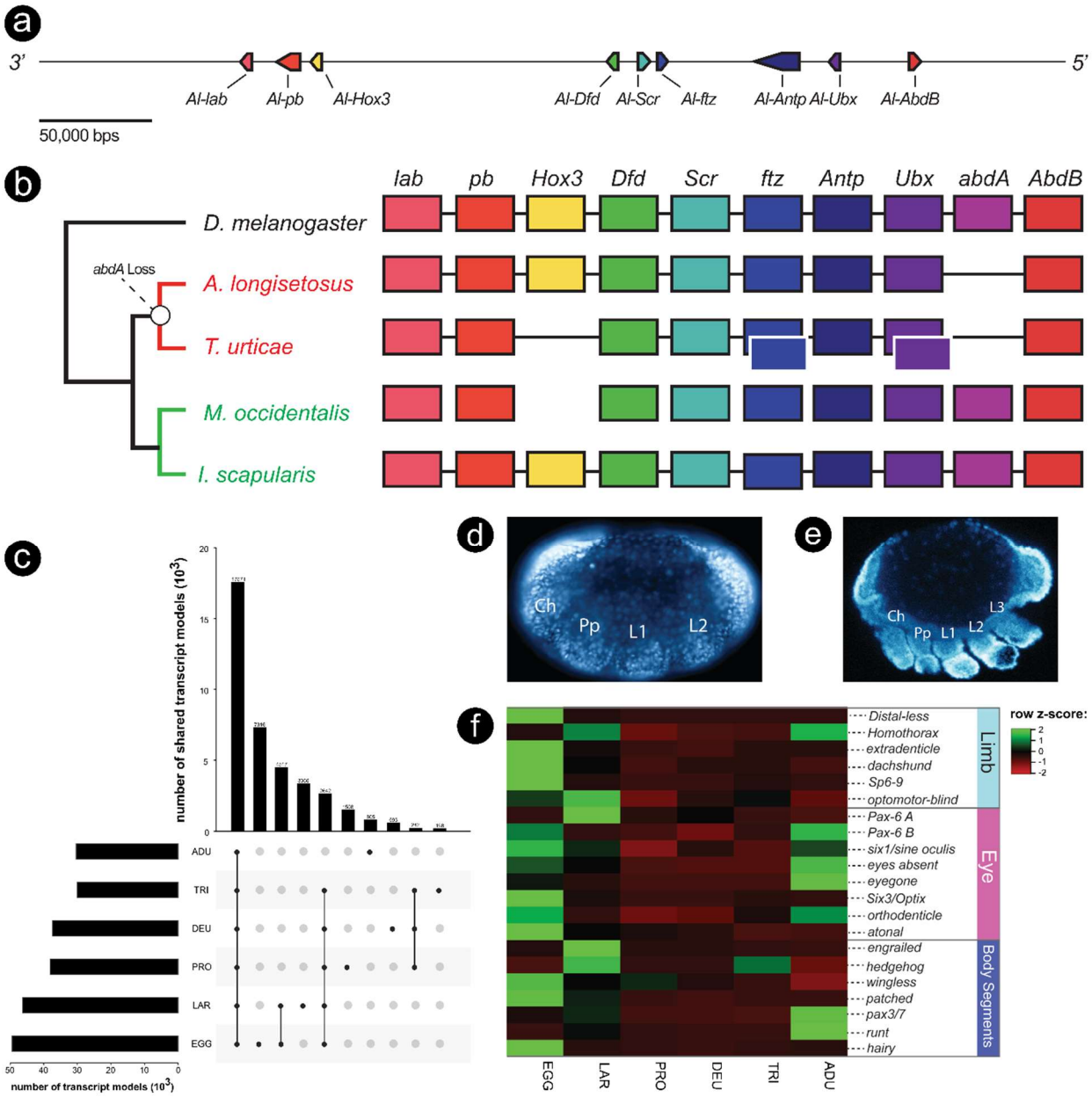


Figure 5 The genomic organization of the Hox genes and life-stage specific expression patters of developmental genes in *Archegozetes longisetosus*. **a:** Schematic of the genomic region enclosing the *Archegozetes* Hox cluster. The genomic organization of the Hox cluster is collinear, as it is in many 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.

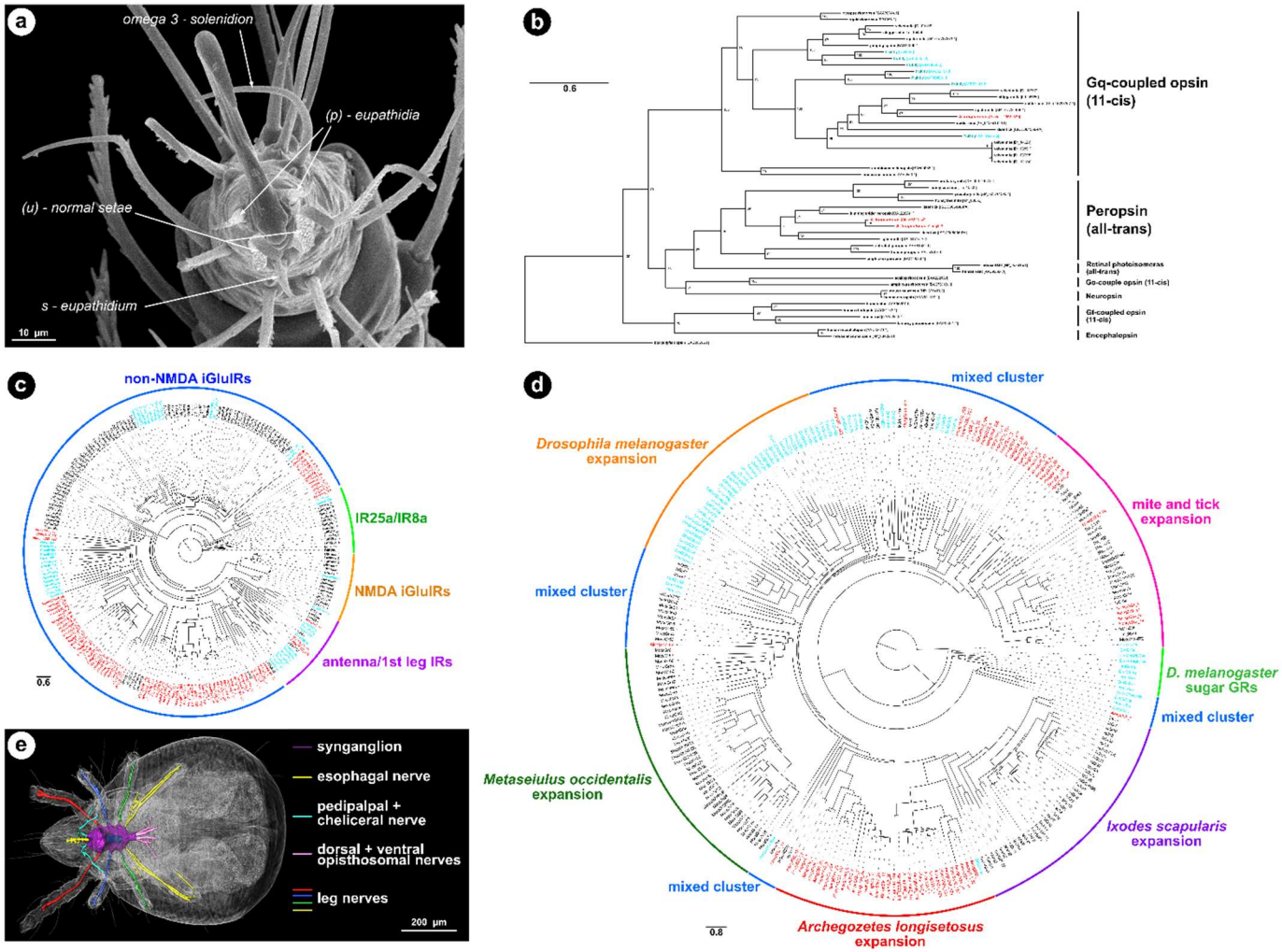
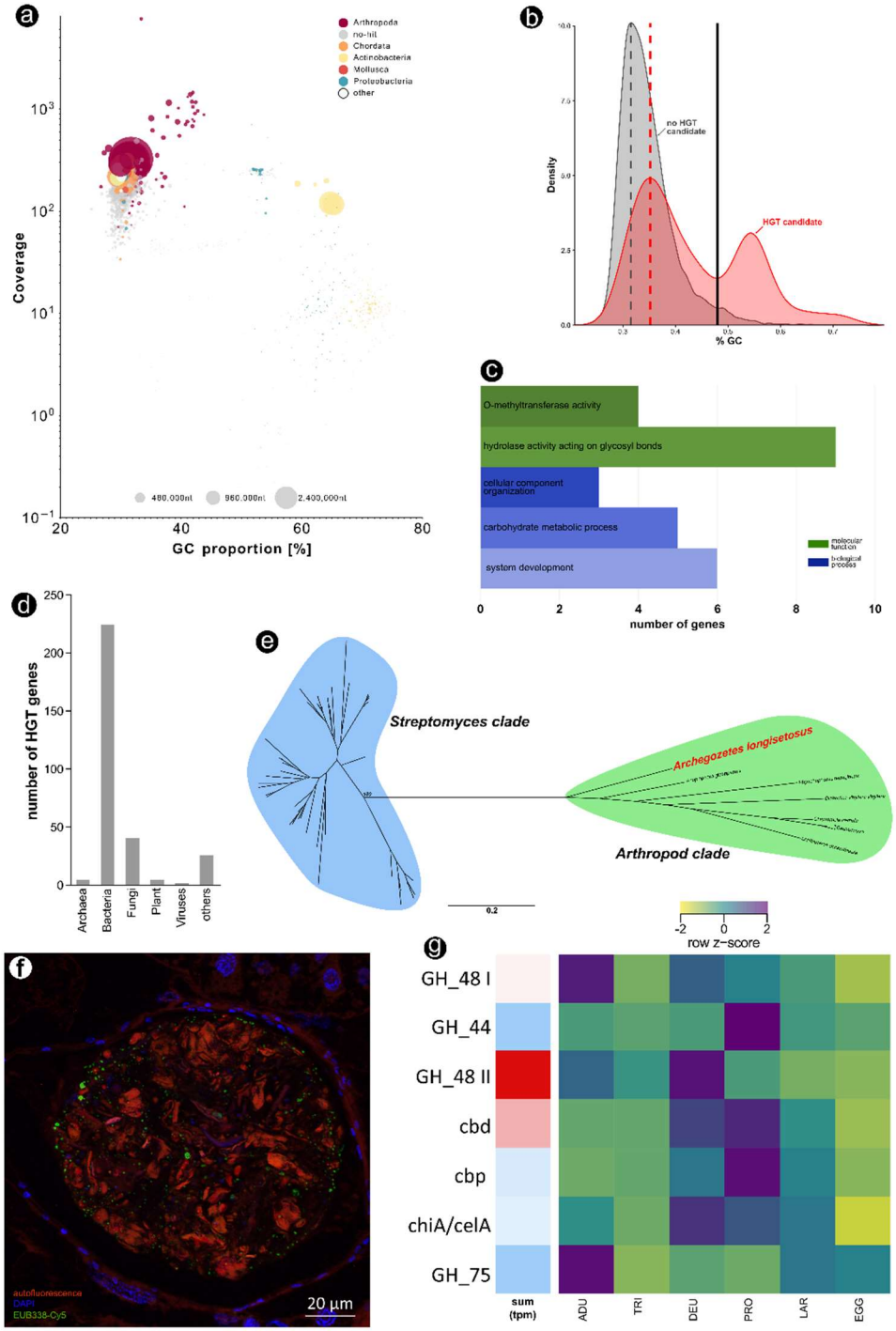


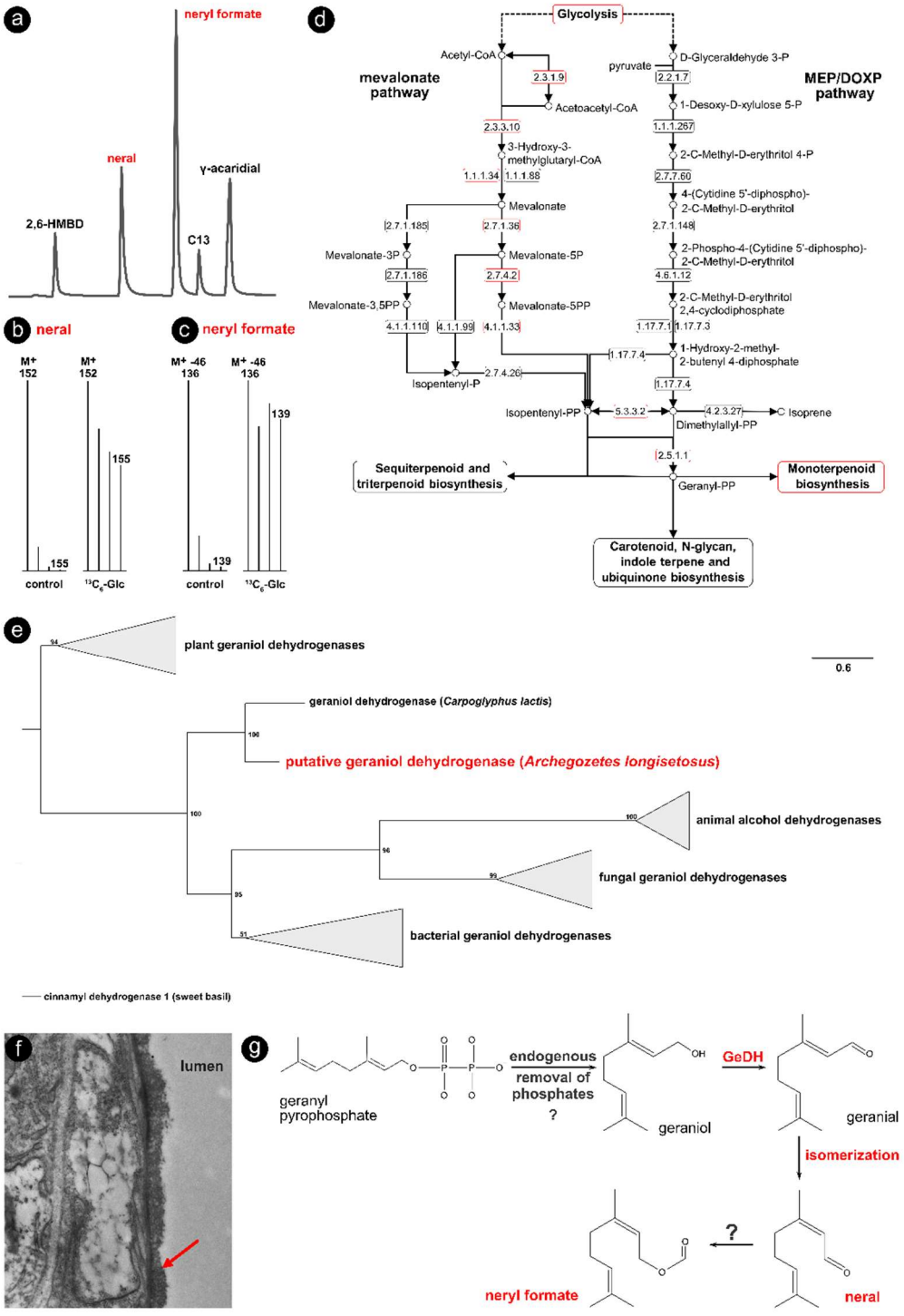
Figure 6 The sensory systems of *Archegozetes longisetosus* and phylogenetic analysis of selected photoreceptor and chemosensory genes. **a**: Scanning electron micrograph (SEM) showing the end of tarsus on *Archegozetes*' first leg. Images shows normal setae, but also modified chemosensory setae, namely 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 (187).



1003 **Figure 7** Horizontal gene transfer (HGT) and implications for the feeding biology of *Archegozetes*
1004 *longisetosus*. **a:** Blob-plot of the long-read genome assembly contigs plotting the read coverage
1005 against GC proportion [%]. Contigs are colored according to the taxonomic order of their best
1006 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.



1027 **Figure 8** Reconstruction of the biosynthetic pathway leading to monoterpenes in *Archegozetes*
1028 *longisetosus*. **a**: Representative gas chromatogram of the mite' gland content; in order of retention
1029 time: 2-hydroxy-6-methyl-benzaldehyde (2,6-HMBD), neral ((*Z*)-3,7-dimethylocta-2,6-dienal)
1030 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.

1054 **Table 1** *Archegozetes longisetosus* genome metrics

1055

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

1056 **Table 2** Comparison of chemosensory receptor repertoires between *Archegozetes longisetosus* and
1057 other arthropods. GR= gustatory receptor, OR= odorant receptor, IR= ionotropic receptor, OBP=
1058 odorant binding protein, CSP= chemosensory protein.

	Chemosensory receptors				
	GR	**OR**	**IR**	**OBP**	**CSP**
A. longisetosus	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

1059

1060

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 (231).

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

1. Childers A (2020) Sequenced Arthropod Genomes. in http://i5k.github.io/arthropod_genomes_at_ncbi (i5k initiative).
2. Thomas GW, *et al.* (2020) Gene content evolution in the arthropods. *Genome biology* 21(1):1-14.
3. Giribet G & Edgecombe GD (2019) The phylogeny and evolutionary history of arthropods. *Current Biology* 29(12):R592-R602.
4. Regier JC, *et al.* (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463(7284):1079-1083.
5. Dunlop J & Selden P (1998) The early history and phylogeny of the chelicerates. *Arthropod relationships*, (Springer), pp 221-235.
6. Ballesteros JA & Sharma PP (2019) A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Systematic Biology* 68(6):896-917.
7. Dunlop JA (2010) Geological history and phylogeny of Chelicerata. *Arthropod structure & development* 39(2-3):124-142.
8. Schaefer I, Norton RA, Scheu S, & Maraun M (2010) Arthropod colonization of land - Linking molecules and fossils in oribatid mites (Acari, Oribatida). *Molecular Phylogenetics and Evolution* 57:113-121.
9. Dunlop J & Alberti G (2008) The affinities of mites and ticks: a review. *Journal of Zoological Systematics and Evolutionary Research* 46(1):1-18.
10. Walter DE & Proctor HC (1999) Mites: ecology, evolution, and behaviour.

11. Grbić M, *et al.* (2011) The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature* 479(7374):487-492.

12. Gulia-Nuss M, *et al.* (2016) Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nature communications* 7(1):1-13.

13. Hoy MA, *et al.* (2016) Genome sequencing of the phytoseiid predatory mite Metaseiulus occidentalis reveals completely atomized Hox genes and superdynamic intron evolution. *Genome biology and evolution* 8(6):1762-1775.

14. Rider SD, Morgan MS, & Arlian LG (2015) Draft genome of the scabies mite. *Parasites & vectors* 8(1):1-14.

15. Dong X, *et al.* (2017) Draft genome of the honey bee ectoparasitic mite, Tropilaelaps mercedesae, is shaped by the parasitic life history. *Gigascience* 6(3):gix008.

16. Cornman RS, *et al.* (2010) Genomic survey of the ectoparasitic mite Varroa destructor, a major pest of the honey bee Apis mellifera. *BMC genomics* 11(1):602.

17. Dong X, *et al.* (2018) Genomes of trombidid mites reveal novel predicted allergens and laterally transferred genes associated with secondary metabolism. *GigaScience* 7(12):giy127.

18. Bast J, *et al.* (2016) No accumulation of transposable elements in asexual arthropods. *Molecular biology and evolution* 33(3):697-706.

19. Klimov PB & OConnor B (2013) Is permanent parasitism reversible?—Critical evidence from early evolution of house dust mites. *Systematic biology* 62(3):411-423.

20. Weinstein SB & Kuris AM (2016) Independent origins of parasitism in Animalia. *Biology Letters* 12(7):20160324.

21. Shingate P, *et al.* (2020) Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nature communications* 11(1):1-13.

22. Lozano-Fernandez J, *et al.* (2019) Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nature communications* 10(1):1-8.

23. Aoki J (1965) Oribatiden (Acarina) Thailands. I. *Nature and Life in Southeast Asia* 4:129-193.

24. Schmelzle S & Blüthgen N (2019) Under pressure: force resistance measurements in box mites (Actinotrichida, Oribatida). *Frontiers in zoology* 16(1):24.

25. Heethoff M & Koerner L (2007) Small but powerful: the oribatid mite Archegozetes longisetosus Aoki (Acari, Oribatida) produces disproportionately high forces. *Journal of Experimental Biology* 210(17):3036-3042.

26. Brückner A, *et al.* (2017) Storage and release of hydrogen cyanide in a chelicerate (Oribatula tibialis). *Proc Natl Acad Sci U S A* 114(13):3469-3472.

27. Brückner A, Kaltenpoth M, & Heethoff M (2020) De novo biosynthesis of simple aromatic compounds by an arthropod (Archegozetes longisetosus). *Proceedings of the Royal Society B* 287(1934):20201429.

28. Norton RA & Palmer SC (1991) The distribution, mechanisms and evolutionary significance of parthenogenesis in oribatid mites. *The Acari - Reproduction, Development and Life-History Strategies*, eds Schuster R & Murphy PW (Chapman & Hall, London), pp 107-136.

29. Heethoff M, Norton RA, Scheu S, & Maraun M (2009) Parthenogenesis in Oribatid Mites (Acari, Oribatida): Evolution Without Sex. *Lost Sex: The Evolutionary Biology of*

Parthenogenesis, eds Schön I, Martens K, & van Dijk P (Springer, Dordrecht), pp 241-257.

30. Raspotnig G (2009) Oil gland secretions in Oribatida (Acari). *Trends in Acarology*, eds Sabelis MW & Bruin J (Springer), pp 235-239.

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

32. 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.

33. Heethoff M, Bergmann P, Laumann M, & Norton RA (2013) The 20th anniversary of a model mite: A review of current knowledge about Archegozetes longisetosus (Acari, Oribatida). *Acarologia* 53(4):353-368.

34. Palmer SC & Norton RA (1992) Genetic Diversity in Thelytokous Oribatid Mites (Acari; Acariformes: Desmonomata). *Biochemical Systematics and Ecology* 20(3):219-231.

35. Wrensch DL, Kethley JB, & Norton RA (1994) Cytogenetics of holokinetic chromosomes and inverted meiosis: keys to the evolutionary success of mites, with generalizations on eukaryotes. *Mites: Ecological and Evolutionary Analyses of Life-History Patterns*, ed Houck MA (Chapman & Hall, New York), pp 282-343.

36. Heethoff M, Bergmann P, & Norton RA (2006) Karyology and sex determination of oribatid mites. *Acarologia* 46(1-2):127-131.

37. Bergmann P, Laumann M, Norton RA, & Heethoff M (2018) Cytological evidence for automictic thelytoky in parthenogenetic oribatid mites (Acari, Oribatida): Synaptonemal complexes confirm meiosis in Archegozetes longisetosus. *Acarologia* in press.

38. Oxley PR, *et al.* (2014) The genome of the clonal raider ant Cerapachys biroi. *Current Biology* 24(4):451-458.

39. Norton RA (1994) Evolutionary aspects of oribatid mite life histories and consequences for the origin of the Astigmata. *Mites: Ecological and Evolutionary Analyses of Life-History Patterns*, ed Houck MA (Chapman & Hall, New York), Chapman & Hall Ed, pp 99-135.

40. Norton RA (2007) Holistic acarology and ultimate causes: examples from the oribatid mites. *Acarology XI: Proceedings of the International Congress.*, eds Morales-Malacara JB, Behan-Pelletier V, Ueckermann E, Perez TM, Estrada-Venegas EG, & Badii M (Sociedad Latinoamericana de Acarologia), pp 3-20.

41. Sakata T & Norton RA (2001) Opisthonotal gland chemistry of early-derivative oribatid mites (Acari) and its relevance to systematic relationships of Astigmata. *International Journal of Acarology* 27(4):281-292.

42. Heethoff M (2012) Regeneration of complex oil-gland secretions and its importance for chemical defense in an oribatid mite. *J Chem Ecol* 38(9):1116-1123.

43. Brückner A & Parker J (2020) Molecular evolution of gland cell types and chemical interactions in animals. *Journal of Experimental Biology* 223(Suppl 1).

44. Riha G (1951) Zur Ökologie der Oribatiden in Kalksteinböden. *Zoologische Jahrbücher* 80:407-450.

45. Smrž J (1992) Some adaptive features in the microanatomy of moss-dwelling oribatid mites (Acari: Oribatida) with respect to their ontogenetical development. *Pedobiologia* 36(5):306-320.

46. Zachvatkin AA (1941) Fauna of the U.S.S.R., Vol. VI, No.1 Tyroglyphoidae (Acari). Zoological Institute of the Acaemy of Science of the U.S.S.R., Moscow.

47. Shimano S, Sakata T, Mizutani Y, Kuwahara Y, & Aoki J-i (2002) Geranial: the alarm pheromone in the nymphal stage of the oribatid mite, Nothrus palustris. *Journal of Chemical Ecology* 28(9):1831-1837.

48. Heethoff M, Koerner L, Norton RA, & Raspotnig G (2011) Tasty but protected-first evidence of chemical defense in oribatid mites. *Journal of Chemical Ecology* 37:1037-1043.

49. Raspotnig G (2006) Chemical alarm and defence in the oribatid mite Collohmannia gigantea (Acari: Oribatida). *Experimental and Applied Acarology* 39(3):177-194.

50. Heethoff M*, et al.* (2018) Life as a fortress–structure, function, and adaptive values of morphological and chemical defense in the oribatid mite Euphthiracarus reticulatus (Actinotrichida). *BMC Zoology* 3(1):7.

51. Saporito RA*, et al.* (2007) Oribatid mites as a major dietary source for alkaloids in poison frogs. *Proceedings of the National Academy of Sciences of the USA* 104(21):8885-8890.

52. Brückner A, Stabentheiner E, Leis HJ, & Raspotnig G (2015) Chemical basis of unwettability in Liacaridae (Acari, Oribatida): specific variations of a cuticular acid/ester-based system. *Exp Appl Acarol* 66(3):313-335.

53. Saporito RA, Spande TF, Garraffo HM, & Donnelly MA (2009) Arthropod alkaloids in poison frogs: A review of the dietary hypothesis. *Heterocycles* 79:277-297.

54. Bensoussan N*, et al.* (2016) Plant-herbivore interaction: dissection of the cellular pattern of Tetranychus urticae feeding on the host plant. *Front Plant Sci* 7:1105.

55. Cohen AC (1995) Extra-oral digestion in predaceous terrestrial Arthropoda. *Annual review of entomology* 40(1):85-103.

56. Shultz JW (2007) A phylogenetic analysis of the arachnid orders based on morphological characters. *Zoological Journal of the Linnean Society* 150(2):221-265.

57. Heethoff M & Norton RA (2009) A new use for synchrotron X-ray microtomography: three-dimensional biomechanical modeling of chelicerate mouthparts and calculation of theoretical bite forces. *Invertebrate Biology* 128(4):332-339.

58. Brückner A*, et al.* (2017) Track the snack – Olfactory cues shape foraging behaviour of decomposing soil mites (Oribatida). *Pedobiologia* in press - online early.

59. Schneider K*, et al.* (2004) Trophic niche differentiation in soil microarthropods (Oribatida, Acari): evidence from stable isotope ratios (N-15/N-14). *Soil Biology and Biochemistry* 36(11):1769-1774.

60. Schneider K, Renker C, Scheu S, & Maraun M (2004) Feeding biology of oribatid mites: a minireview. *Phytophaga* 14:247-256.

61. Schneider K & Maraun M (2005) Feeding preferences among dark pigmented fungal taxa ("Dematiacea") indicate limited trophic niche differentiation of oribatid mites (Oribatida, Acari). *Pedobiologia* 49(1):61-67.

62. Heidemann K, Scheu S, Ruess L, & Maraun M (2011) Molecular detection of nematode predation and scavenging in oribatid mites: Laboratory and field experiments. *Soil Biology and Biochemistry* 43:229-236.

63. Maraun M*, et al.* (2011) Stable isotopes revisited: Their use and limits for oribatid mite trophic ecology. *Soil Biology and Biochemistry* 43(5):877-882.

64. Brückner A, Schuster R, Smit T, & Heethoff M (2018) Imprinted or innated food preferences in the model mite Archegozetes longisetosus (Actinotrichida, Oribatida, Trhypochthoniidae). *Soil Organisms* submitted.

65. Smrž J & Čatská V (2010) Mycophagous mites and their internal associated bacteria cooperate to digest chitin in soil. *Symbiosis* 52(1):33-40.

66. Smrz J (2000) A modified test for chitinase and cellulase activity in soil mites. *Pedobiologia* 44(2):186-189.

67. Siepel H & de Ruiter-Dijkman EM (1993) Feeding guilds of oribatid mites based on their carbohydrase activities. *Soil Biology and Biochemistry* 25(11):1491-1497.

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

69. Haq MA (1993) Symbiotic association of mites and microbes in cellulose degradation. *Soil Organisms and Sustainability* 4:81-85.

70. Zinkler D (1971) Vergleichende Untersuchungen zum Wirkungsspektrum der Carbohydrasen laubstreubewohnender Oribatiden. *Zool Ges Verh*.

71. Smrž J & Norton RA (2004) Food selection and internal processing in Archegozetes longisetosus (Acari : Oribatida). *Pedobiologia* 48(2):111-120.

72. McKenna DD*, et al.* (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences* 116(49):24729-24737.

73. Wybouw N, Pauchet Y, Heckel DG, & Van Leeuwen T (2016) Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome biology and evolution* 8(6):1785-1801.

74. Mayer WE, Schuster LN, Bartelmes G, Dieterich C, & Sommer RJ (2011) Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC evolutionary biology* 11(1):1-10.

75. 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 molecular biology* 27(3):333-351.

76. Wu C*, et al.* (2017) Analysis of the genome of the New Zealand giant collembolan (Holacanthella duospinosa) sheds light on hexapod evolution. *BMC genomics* 18(1):795.

77. Faddeeva-Vakhrusheva A*, et al.* (2016) Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan Orchesella cincta. *Genome biology and evolution* 8(7):2106-2117.

78. Hoffmann A*, et al.* (1998) Intergeneric transfer of conjugative and mobilizable plasmids harbored by Escherichia coli in the gut of the soil microarthropod Folsomia candida (Collembola). *Applied and Environmental Microbiology* 64(7):2652-2659.

79. Barnett AA & Thomas RH (2013) Posterior Hox gene reduction in an arthropod: Ultrabithorax and Abdominal-B are expressed in a single segment in the mite Archegozetes longisetosus. *EvoDevo* 4(1):23.

80. Barnett AA & Thomas RH (2018) Early segmentation in the mite Archegozetes longisetosus reveals conserved and derived aspects of chelicerate development. *Development genes and evolution* 228(5):213-217.

81. Barnett AA & Thomas RH (2012) The delineation of the fourth walking leg segment is temporally linked to posterior segmentation in the mite Archegozetes longisetosus (Acari: Oribatida, Trhypochthoniidae). *Evol Dev* 14(4):383-392.

82. Barnett AA & Thomas RH (2013) The expression of limb gap genes in the mite Archegozetes longisetosus reveals differential patterning mechanisms in chelicerates. *Evol Dev* 15(4):280-292.

83. Brückner A, Hilpert A, & Heethoff M (2017) Biomarker function and nutritional stoichiometry of neutral lipid fatty acids and amino acids in oribatid mites. *Soil Biology & Biochemistry* 115:35-43.

84. Thomas RH (2002) Mites as models in development and genetics. *Acarid Phylogeny and Evolution: Adaptation in Mites and Ticks: Proceedings of the IV Symposium of the European Association of Acarologists*, eds Bernini F, Nannelli R, Nuzzaci G, & de Lillo E (Kluwer Academic Publishers), pp 21-26.

85. Brückner A, Wehner K, Neis M, & Heethoff M (2016) Attack and defense in a gamasid-oribatid mite predator-prey experiment - sclerotization outperforms chemical repellency. *Acarologia* 56(4):451-461.

86. Heethoff M & Raspotnig G (2012) Expanding the 'enemy-free space' for oribatid mites: evidence for chemical defense of juvenile Archegozetes longisetosus against the rove beetle Stenus juno. *Experimental and Applied Acarology* 56:93-97.

87. Brückner A, Schuster R, Wehner K, & Heethoff M (2018) Adding to the reproductive biology of Archegozetes longisetosus (Actinotrichida, Oribatida, Trhypochthoniidae) again: the role of nutritional quality. *Soil Organisms* (submitted).

88. Raspotnig G, Norton RA, & Heethoff M (2011) Oribatid mites and skin alkaloids in poison frogs. *Biology Letters* 7:555-556.

89. Heethoff M & Rall BC (2015) Reducible defence: chemical protection alters the dynamics of predator–prey interactions. *Chemoecology* 25(2):53-61.

1374 90. Brückner A & Heethoff M (2018) Nutritional effects on chemical defense alter predator–
1375 prey dynamics. *Chemoecology* 28(1):1-9.
1376 91. Thiel T, Brechtel A, Brückner A, Heethoff M, & Drossel B (2018) The effect of
1377 reservoir-based chemical defense on predator-prey dynamics. *Theoretical Ecology*:1-14.
1378 92. Norton RA, Kethley JB, Johnston DE, & O'Connor BM (1993) Phylogenetic
1379 perspectives on genetic systems and reproductive modes of mites. *Evolution and*
1380 *Diversity of Sex Ratio in Insects and Mites*, eds Wrensch D & Ebbert M (Chapman &
1381 Hall, London), pp 8-99.
1382 93. Oliver Jr JH (1983) Chromosomes, genetic variance and reproductive strategies among
1383 mites and ticks. *Bulletin of the ESA* 29(2):8-17.
1384 94. Simpson JT (2014) Exploring genome characteristics and sequence quality without a
1385 reference. *Bioinformatics* 30(9):1228-1235.
1386 95. Alfsnes K, Leinaas HP, & Hessen DO (2017) Genome size in arthropods; different roles
1387 of phylogeny, habitat and life history in insects and crustaceans. *Ecology and Evolution*
1388 7(15):5939-5947.
1389 96. Schwager EE*, et al.* (2017) The house spider genome reveals an ancient whole-genome
1390 duplication during arachnid evolution. *BMC biology* 15(1):1-27.
1391 97. dos Santos G*, et al.* (2015) FlyBase: introduction of the Drosophila melanogaster Release
1392 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic*
1393 *acids research* 43(D1):D690-D697.
1394 98. Consortium TGS (2008) The genome of the model beetle and pest Tribolium castaneum.
1395 *Nature* 452(7190):949.

99. Palmer M, Bantle J, Guo X, & Fargoxy1 WS (1994) Genome size and organization in the ixodid tick Amblyomma americanum (L.). *Insect molecular biology* 3(1):57-62.

100. Van Zee JP*, et al.* (2007) Tick genomics: the Ixodes genome project and beyond. *International journal for parasitology* 37(12):1297-1305.

101. Barrero RA*, et al.* (2017) Gene-enriched draft genome of the cattle tick Rhipicephalus microplus: assembly by the hybrid Pacific Biosciences/Illumina approach enabled analysis of the highly repetitive genome. *International journal for parasitology* 47(9):569-583.

102. Emms DM & Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* 16(1):157.

103. Kanehisa M, Sato Y, & Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology* 428(4):726-731.

104. Palmer J & Stajich J (2017) Funannotate: eukaryotic genome annotation pipeline.

105. Pachl P*, et al.* (2012) Convergent evolution of defense mechanisms in oribatid mites (Acari, Oribatida) shows no “ghosts of predation past”. *Molecular Phylogenetics and Evolution* 65(2):412-420.

106. Alberti G & Michalik P (2004) *Feinstrukturelle Aspekte der Fortpflanzungssysteme von Spinnentieren (Arachnida)* (Denisia, Wien).

107. Koller LM, Wirth S, & Raspotnig G (2012) Geranial-rich oil gland secretions: a common phenomenon in the Histiostomatidae (Acari, Astigmata)? *International Journal of Acarology* 38(5):420-426.

108. Liana M & Witaliński W (2005) Sperm structure and phylogeny of Astigmata. *Journal of morphology* 265(3):318-324.

109. Dabert M, Witalinski W, Kazmierski A, Olszanowski Z, & Dabert J (2010) Molecular phylogeny of acariform mites (Acari, Arachnida): Strong conflict between phylogenetic signal and long-branch attraction artifacts. *Molecular Phylogenetics and Evolution* 56:222-241.

110. Maraun M*, et al.* (2004) Molecular phylogeny of oribatid mites (Oribatida, Acari): evidence for multiple radiations of parthenogenetic lineages. *Experimental and Applied Acarology* 33(3):183-201.

111. Norton RA (1998) Morphological evidence for the evolutionary origin of Astigmata (Acari : Acariformes). *Experimental and Applied Acarology* 22(10):559-594.

112. Pepato A & Klimov P (2015) Origin and higher-level diversification of acariform mites–evidence from nuclear ribosomal genes, extensive taxon sampling, and secondary structure alignment. *BMC evolutionary biology* 15(1):178.

113. Li W-N & Xue X-F (2019) Mitochondrial genome reorganization provides insights into the relationship between oribatid mites and astigmatid mites (Acari: Sarcoptiformes: Oribatida). *Zoological Journal of the Linnean Society* 187(3):585-598.

114. Domes K, Althammer M, Norton RA, Scheu S, & Maraun M (2007) The phylogenetic relationship between Astigmata and Oribatida (Acari) as indicated by molecular markers. *Experimental and Applied Acarology* 42(3):159-171.

115. Klimov PB*, et al.* (2018) Comprehensive phylogeny of acariform mites (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a long branch. *Molecular Phylogenetics and Evolution* 119:105-117.

116. Van Dam MH, Trautwein M, Spicer GS, & Esposito L (2019) Advancing mite phylogenomics: Designing ultraconserved elements for Acari phylogeny. *Molecular ecology resources* 19(2):465-475.

117. Jeyaprakash A & Hoy MA (2009) First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Experimental and Applied Acarology* 47(1):1-18.

118. Alberti G (1984) The contribution of comparative spermatology to problems of acarine systematics. *Acarology VI*, pp 479-490.

119. Alberti G (1991) Spermatology in the Acari: systematic and functional implications. *The Acari - Reproduction, Development and Life-History Strategies*, eds Schuster R & Murphy PW (Chapman & Hall, London), pp 77-105.

120. Dabert M (2006) DNA markers in the phylogenetics of the Acari. *Biological Letters* 43(2):97-107.

121. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210-3212.

122. Siepel A*, et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* 15(8):1034-1050.

123. Nagy LG, Merényi Z, Hegedüs B, & Bálint B (2020) Novel phylogenetic methods are needed for understanding gene function in the era of mega-scale genome sequencing. *Nucleic Acids Research* 48(5):2209-2219.

124. Charlesworth B (2012) The effects of deleterious mutations on evolution at linked sites. *Genetics* 190(1):5-22.

125. Barton NH (2010) Mutation and the evolution of recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1544):1281-1294.

126. Schön I, Martens K, & van Dijk P (2009) *Lost Sex - The Evolutionary Biology of Parthenogenesis* (Springer, Dordrecht).

127. Muller HJ (1964) The relation of recombination to mutaional advance. *Mutation Research* 106:2-9.

128. Arkhipova I & Meselson M (2000) Transposable elements in sexual and ancient asexual taxa. *Proceedings of the National Academy of Sciences* 97(26):14473-14477.

129. Nuzhdin SV & Petrov DA (2003) Transposable elements in clonal lineages: lethal hangover from sex. *Biological Journal of the Linnean Society* 79(1):33-41.

130. Panfilio KA*, et al.* (2019) Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome biology* 20(1):64.

131. Bourque G*, et al.* (2018) Ten things you should know about transposable elements. *Genome biology* 19(1):1-12.

132. Crescente JM, Zavallo D, Helguera M, & Vanzetti LS (2018) MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC bioinformatics* 19(1):348.

133. Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends in genetics* 5:103-107.

134. Waterston RH*, et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520-562.

135. Dej KJ, Gerasimova T, Corces VG, & Boeke JD (1998) A hotspot for the Drosophila gypsy retroelement in the ovo locus. *Nucleic acids research* 26(17):4019-4024.

136. Havecker ER, Gao X, & Voytas DF (2004) The diversity of LTR retrotransposons. *Genome biology* 5(6):1-6.

137. Hrycaj SM & Wellik DM (2016) Hox genes and evolution. *F1000Research* 5.

138. Holland P & Hogan B (1988) Expression of homeo box genes during mouse development: a review. *Genes & development* 2(7):773-782.

139. Hughes CL & Kaufman TC (2002) Hox genes and the evolution of the arthropod body plan 1. *Evolution & development* 4(6):459-499.

140. Dunlop JA & Lamsdell JC (2017) Segmentation and tagmosis in Chelicerata. *Arthropod structure & development* 46(3):395-418.

141. van der Hammen L (1970) La segmentation primitive des Acariens. *Acarologia* 12(1):3-10.

142. Schwager EE, Schönauer A, Leite DJ, Sharma PP, & McGregor AP (2015) Chelicerata. *Evolutionary Developmental Biology of Invertebrates 3*, (Springer), pp 99-139.

143. Sharma PP, Schwager EE, Extavour CG, & Giribet G (2012) Hox gene expression in the harvestman Phalangium opilio reveals divergent patterning of the chelicerate opisthosoma. *Evolution & development* 14(5):450-463.

144. Santos VT*, et al.* (2013) The embryogenesis of the tick Rhipicephalus (Boophilus) microplus: the establishment of a new chelicerate model system. *Genesis* 51(12):803-818.

145. Pace RM, Grbić M, & Nagy LM (2016) Composition and genomic organization of arthropod Hox clusters. *Evodevo* 7(1):11.

146. Cook CE, Smith ML, Telford MJ, Bastianello A, & Akam M (2001) Hox genes and the phylogeny of the arthropods. *Current Biology* 11(10):759-763.

147. Telford MJ & Thomas RH (1998) Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proceedings of the National Academy of Sciences of the USA* 95(18):10671-10675.

148. Heethoff M, Laumann M, & Bergmann P (2007) Adding to the reproductive biology of the parthenogenetic oribatid mite, Archegozetes longisetosus (Acari, Oribatida, Trhypochthoniidae). *Turkish Journal of Zoology* 31:151-159.

149. Sharma PP*, et al.* (2015) A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proceedings of the Royal Society B: Biological Sciences* 282(1808):20150698.

150. Königsmann T, Turetzek N, Pechmann M, & Prpic N-M (2017) Expression and function of the zinc finger transcription factor Sp6–9 in the spider Parasteatoda tepidariorum. *Development genes and evolution* 227(6):389-400.

151. Heingård M, Turetzek N, Prpic N-M, & Janssen R (2019) FoxB, a new and highly conserved key factor in arthropod dorsal–ventral (DV) limb patterning. *EvoDevo* 10(1):1-16.

152. Evans GO (1992) *Principles of Acarology* (CAB International, Wallingford) p 563.

153. Gainett G*, et al.* (2020) Systemic paralogy and function of retinal determination network homologs in arachnids. *BMC Genomics* 21:811.

154. Samadi L, Schmid A, & Eriksson BJ (2015) Differential expression of retinal determination genes in the principal and secondary eyes of Cupiennius salei Keyserling (1877). *EvoDevo* 6(1):16.

155. Schomburg C*, et al.* (2015) Molecular characterization and embryonic origin of the eyes in the common house spider Parasteatoda tepidariorum. *EvoDevo* 6(1):1-14.

156. Alberti G & Coons LB (1999) *Acari-Mites* (Wiley, New York) p 1265.

157. Patten W (1887) Eyes of molluscs and arthropods. *Journal of Morphology* 1(1):67-92.

158. Exner S (1989) *The physiology of the compound eyes of insects and crustaceans* (Springer-Verlag GmbH & Co. KG).

159. Harzsch S*, et al.* (2006) Evolution of arthropod visual systems: development of the eyes and central visual pathways in the horseshoe crab Limulus polyphemus Linnaeus, 1758 (Chelicerata, Xiphosura). *Developmental Dynamics* 235(10):2641-2655.

160. 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 Organisms* 84(2):391-408.

161. Norton RA & Fuangarworn M (2015) Nanohystricidae n. fam., an unusual, plesiomorphic enarthronote mite family endemic to New Zealand (Acari, Oribatida). *Zootaxa* 4027(2):151-204.

162. Norton RA & Franklin E (2018) Paraquanothrus n. gen. from freshwater rock pools in the USA, with new diagnoses of Aquanothrus, Aquanothrinae, and Ameronothridae (Acari, Oribatida). *Acarologia* 58(3):557-627.

163. Walter DE & Proctor HC (1998) Feeding behaviour and phylogeny: observations on early derivative Acari. *Experimental & applied acarology* 22(1):39-50.

164. Alberti G (1998) Fine structure of receptor organs in oribatid mites (Acari). *Arthropod biology: Contributions to morphology, ecology and systematics. Biosystematics and Ecology Series* 14:27-77.

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

166. Woodring J (1966) Color phototactic responses of an eyeless oribatid mite. *Acarologia* 8(2):382-388.

167. Trägårdh I (1933) Methods of automatic collecting for studying the fauna of the soil. *Bulletin of Entomological Research* 24(2):203-214.

168. Nagata T, Koyanagi M, Tsukamoto H, & Terakita A (2010) Identification and characterization of a protostome homologue of peropsin from a jumping spider. *Journal of Comparative Physiology A* 196(1):51.

169. Koyanagi M, Nagata T, Katoh K, Yamashita S, & Tokunaga F (2008) Molecular evolution of arthropod color vision deduced from multiple opsin genes of jumping spiders. *Journal of Molecular Evolution* 66(2):130-137.

170. Eriksson BJ, Fredman D, Steiner G, & Schmid A (2013) Characterisation and localisation of the opsin protein repertoire in the brain and retinas of a spider and an onychophoran. *BMC evolutionary biology* 13(1):186.

171. Senthilan PR, Grebler R, Reinhard N, Rieger D, & Helfrich-Förster C (2019) Role of rhodopsins as circadian photoreceptors in the Drosophila melanogaster. *Biology* 8(1):6.

172. Senthilan PR & Helfrich-Förster C (2016) Rhodopsin 7–the unusual rhodopsin in Drosophila. *PeerJ* 4:e2427.

173. Shen WL, *et al.* (2011) Function of rhodopsin in temperature discrimination in Drosophila. *Science* 331(6022):1333-1336.

174. Vieira FG & Rozas J (2011) Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biology and Evolution* 3:476-490.

175. Sánchez-Gracia A, Vieira FG, Almeida FC, & Rozas J (2011) Comparative genomics of the major chemosensory gene families in Arthropods. *eLS*.

176. Sánchez-Gracia A, Vieira F, & Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103(3):208-216.

177. Ngoc PCT, *et al.* (2016) Complex evolutionary dynamics of massively expanded chemosensory receptor families in an extreme generalist chelicerate herbivore. *Genome biology and evolution* 8(11):3323-3339.

178. Croset V, *et al.* (2010) Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS genetics* 6(8):e1001064.

179. Benton R, Vannice KS, Gomez-Diaz C, & Vosshall LB (2009) Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell* 136(1):149-162.

180. Knecht ZA, *et al.* (2016) Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in Drosophila. *Elife* 5:e17879.

181. Budelli G, *et al.* (2019) Ionotropic receptors specify the morphogenesis of phasic sensors controlling rapid thermal preference in Drosophila. *Neuron* 101(4):738-747. e733.

182. Rytz R, Croset V, & Benton R (2013) Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in Drosophila and beyond. *Insect biochemistry and molecular biology* 43(9):888-897.

183. Montell C (2009) A taste of the Drosophila gustatory receptors. *Current opinion in neurobiology* 19(4):345-353.

184. Kuwahara Y (2004) Chemical ecology of astigmatid mites. *Advances in Insect Chemical Ecology*, eds Cardé RT & Millar JG (Cambridge University Press, Cambridge), pp 76-109.

185. Bunnell T, Hanisch K, Hardege JD, & Breithaupt T (2011) The fecal odor of sick hedgehogs (Erinaceus europaeus) mediates olfactory attraction of the tick Ixodes hexagonus. *Journal of chemical ecology* 37(4):340.

186. Yunker C*, et al.* (1992) Olfactory responses of adultAmblyomma hebraeum andA. variegatum (Acari: Ixodiae) to attractant chemicals in laboratory tests. *Experimental & applied acarology* 13(4):295-301.

187. Hartmann K, Laumann M, Bergmann P, Heethoff M, & Schmelzle S (2016) Development of the synganglion and morphology of the adult nervous system in the mite Archegozetes longisetosus Aoki (Chelicerata, Actinotrichida, Oribatida). *J Morphol* 277(4):537-548.

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

189. Crisp A, Boschetti C, Perry M, Tunnacliffe A, & Micklem G (2015) Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome biology* 16(1):1-13.

190. Gilbert HJ (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant physiology* 153(2):444-455.

191. Latgé JP (2007) The cell wall: a carbohydrate armour for the fungal cell. *Molecular microbiology* 66(2):279-290.

192. Faddeeva-Vakhrusheva A*, et al.* (2017) Coping with living in the soil: the genome of the parthenogenetic springtail Folsomia candida. *BMC genomics* 18(1):493.

193. Mitreva M, Smant G, & Helder J (2009) Role of horizontal gene transfer in the evolution of plant parasitism among nematodes. *Horizontal Gene Transfer*, (Springer), pp 517-535.

194. Laetsch DR & Blaxter ML (2017) BlobTools: Interrogation of genome assemblies. *F1000Research* 6(1287):1287.

195. Flot J-F*, et al.* (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. *Nature* 500(7463):453-457.

196. Thorpe P, Escudero-Martinez CM, Cock PJ, Eves-van den Akker S, & Bos JI (2018) Shared transcriptional control and disparate gain and loss of aphid parasitism genes. *Genome Biol. Evol.* 10(10):2716-2733.

197. Lawrence JG (1997) Selfish operons and speciation by gene transfer. *Trends in microbiology* 5(9):355-359.

198. Stefaniak O (1981) The effect of fungal diet on the development of Oppia nitens (Acari, Oribatei) and on the microflora of its alimentary tract.

199. Stefaniak O (1976) The microflora of the alimentary canal of Achipteria coleoptrata (Acarina, Oribatei).

200. McKenna DD*, et al.* (2016) Genome of the Asian longhorned beetle (Anoplophora glabripennis), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome biology* 17(1):1-18.

201. Luxton M (1972) Studies on Oribatid Mites of a Danish Beech Wood Soil .1. Nutritional Biology. *Pedobiologia* 12(6):434-463.

202. Luxton M (1981) Studies on the Oribatid Mites of a Danish Beech Wood Soil .7. Energy Budgets. *Pedobiologia* 22(2):77-111.

203. Luxton M (1982) The Biology of Mites from Beech Woodland Soil. *Pedobiologia* 23(1):1-8.

204. Heethoff M, Norton RA, & Raspotnig G (2016) Once Again: Oribatid Mites and Skin Alkaloids in Poison Frogs. *J Chem Ecol* 42(8):841-844.

205. Kuwahara Y*, et al.* (2001) Chemical ecology of astigmatid mites LIX. Neral, the alarm pheromone of Schwiebea elongata (Banks)(Acari: Acaridae). *Journal of the Acarological Society of Japan* 10:19-25.

206. Sakata T & Norton RA (2003) Opisthonotal gland chemistry of a middle-derivative oribatid mite, Archegozetes longisetosus (Acari : Trhypochthoniidae). *International Journal of Acarology* 29(4):345-350.

207. Sakata T, Tagami K, & Kuwahara Y (1995) Chemical ecology of oribatid mites. I. Oil gland components of Hydronothrus crispus Aoki. *Journal of the Acarological Society of Japan* 4(2):69-75.

208. Sakata T (1997) Natural Chemistry of Mite Secretions. PhD-thesis (Kyoto University, Kyoto, Japan).

209. Raspotnig G, Schuster R, & Krisper G (2004) Citral in oil gland secretions of Oribatida (Acari): a key component for phylogenetic analyses. *Abh Ber Naturkundemus Görlitz* 76:43-50.

210. Brückner A & Heethoff M (2017) The ontogeny of oil gland chemistry in the oribatid mite Archegozetes longisetosus Aoki (Oribatida, Trhypochthoniidae). *International Journal of Acarology* 43(5):337-342.

211. Noge K, *et al.* (2005) Biosynthesis of neral in Carpoglyphus lactis (Acari: Carpoglyphidae) and detection of its key enzyme, geraniol dehydrogenase, by electrophoresis. *Journal of the Acarological Society of Japan* 14:75-81.

212. Noge K, *et al.* (2008) Geraniol dehydrogenase, the key enzyme in biosynthesis of the alarm pheromone, from the astigmatid mite Carpoglyphus lactis (Acari: Carpoglyphidae). *The FEBS journal* 275(11):2807-2817.

213. Morita A, Mori N, Nishida R, Hirai N, & Kuwahara Y (2004) Neral (the Alarm Pheromone) Biosynthesis via the Mevalonate Pathway, Evidenced by D-Glucose-1-13C Feeding in Carpoglyphus lactis and 13C Incorporation into Other Opisthonotal Gland Exudates. *Journal of Pesticide Science* 29:27-32.

214. Kanehisa M, *et al.* (2007) KEGG for linking genomes to life and the environment. *Nucleic acids research* 36(suppl_1):D480-D484.

215. Eisenreich W, Bacher A, Arigoni D, & Rohdich F (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cellular and Molecular Life Sciences CMLS* 61(12):1401-1426.

216. Miziorko HM (2011) Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch Biochem Biophys* 505(2):131-143.

217. Oldfield E & Lin FY (2012) Terpene biosynthesis: modularity rules. *Angewandte Chemie International Edition* 51(5):1124-1137.

218. Breitmaier E (2006) *Terpenes: flavors, fragrances, pharmaca, pheromones* (John Wiley & Sons).

219. Degenhardt J, Köllner TG, & Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70(15-16):1621-1637.

220. Trapp SC & Croteau RB (2001) Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158(2):811-832.

221. Liu W*, et al.* (2015) Utilization of alkaline phosphatase PhoA in the bioproduction of geraniol by metabolically engineered Escherichia coli. *Bioengineered* 6(5):288-293.

222. Oswald M, Fischer M, Dirninger N, & Karst F (2007) Monoterpenoid biosynthesis in Saccharomyces cerevisiae. *FEMS yeast research* 7(3):413-421.

223. Zhou J*, et al.* (2014) Engineering Escherichia coli for selective geraniol production with minimized endogenous dehydrogenation. *J Biotechnol* 169:42-50.

224. Beran F, Köllner TG, Gershenzon J, & Tholl D (2019) Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. *New Phytologist*.

225. Raspotnig G, Kaiser R, Stabentheiner E, & Leis HJ (2008) Chrysomelidial in the Opisthonotal Glands of the Oribatid Mite, Oribotritia berlesei. *Journal of Chemical Ecology* 34(8):1081-1088.

226. Shimizu N, Sakata D, Schmelz EA, Mori N, & Kuwahara Y (2017) Biosynthetic pathway of aliphatic formates via a Baeyer-Villiger oxidation in mechanism present in astigmatid mites. *P Natl Acad Sci USA* 114(10):2616-2621.

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

228. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17(1):10-12.

229. Chikhi R & Medvedev P (2014) Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30(1):31-37.

230. Vurture GW*, et al.* (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14):2202-2204.

231. Sun H, Ding J, Piednoël M, & Schneeberger K (2018) findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics* 34(4):550-557.

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

233. Koren S*, et al.* (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* 27(5):722-736.

234. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* 25(14):1754-1760.

235. Walker BJ*, et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* 9(11):e112963.

236. Smit AF & Hubley R (2008) RepeatModeler Open-1.0. *Available fom http://www. repeatmasker. org*.

237. Bao W, Kojima KK, & Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna* 6(1):11.

238. Kohany O, Gentles AJ, Hankus L, & Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7(1):1-7.

239. Petersen M*, et al.* (2019) Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Evol. Biol.* 19(1):11.

240. Rognes T, Flouri T, Nichols B, Quince C, & Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.

241. Smit A, Hubley R, & Green P (1996-2010) RepeatMasker Open-3.0.

242. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, & Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18(12):1979-1990.

243. Bruna T, Hoff K, Stanke M, Lomsadze A, & Borodovsky M (2020) BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *bioRxiv*.

244. Haas BJ*, et al.* (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9(1):R7.

245. Keilwagen J, Hartung F, & Grau J (2019) GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Gene Prediction*, (Springer), pp 161-177.

246. Chan PP & Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. *Gene Prediction*, (Springer), pp 1-14.

247. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.

248. Pruitt KD, Tatusova T, & Maglott DR (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research* 33(suppl_1):D501-D504.

249. Bairoch A & Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic acids research* 28(1):45-48.

250. Bateman A*, et al.* (2004) The Pfam protein families database. *Nucleic acids research* 32(suppl_1):D138-D141.

251. Rawlings ND, Barrett AJ, & Bateman A (2010) MEROPS: the peptidase database. *Nucleic acids research* 38(suppl_1):D227-D233.

252. Cantarel BL*, et al.* (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research* 37(suppl_1):D233-D238.

253. Eddy SR (2011) Accelerated profile HMM searches. *PLoS computational biology* 7(10):e1002195.

254. Huerta-Cepas J*, et al.* (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34(8):2115-2122.

255. Geib SM*, et al.* (2018) Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. *GigaScience* 7(4):giy018.

256. Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28(1):27-30.

257. Consortium GO (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic acids research* 32(suppl_1):D258-D261.

258. Carlson M & Pagès H (2019) AnnotationForge: Tools for building SQLite-Based Annotation Data Packages. *R package version* 1(0).

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

260. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57(1):289-300.

261. Consortium GO (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic acids research* 47(D1):D330-D338.

262. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, & Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research* 35(suppl_2):W182-W185.

263. Suzuki S, Kakuta M, Ishida T, & Akiyama Y (2014) GHOSTX: an improved sequence homology search algorithm using a query suffix array and a database suffix array. *PloS one* 9(8):e103833.

264. Fu L, Niu B, Zhu Z, Wu S, & Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150-3152.

265. Price MN, Dehal PS, & Arkin AP (2010) FastTree 2 – approximately Maximum-Likelihood trees for large alignments. *PLoS One* 5(3):e9490.

266. Capella-Gutiérrez S, Silla-Martínez JM, & Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972-1973.

267. Kocot KM, Citarella MR, Moroz LL, & Halanych KM (2013) PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinform. Online* 9:429-435.

268. Kück P & Meusemann K (2010) FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* 56(3):1115-1118.

269. Lanfear R, Frandsen PB, Wright AM, Senfeld T, & Calcott B (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34(3):772-773.

270. Nguyen L-T, Schmidt HA, Von Haeseler A, & Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 32(1):268-274.

271. 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.

272. Dobin A*, et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15-21.

273. Grabherr MG*, et al.* (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology* 29(7):644.

274. Haas BJ*, et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8(8):1494-1512.

275. Kitchen SA, Crowder CM, Poole AZ, Weis VM, & Meyer E (2015) De novo assembly and characterization of four anthozoan (phylum Cnidaria) transcriptomes. *G3: Genes, genomes, genetics* 5(11):2441-2452.

276. Bray NL, Pimentel H, Melsted P, & Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34(5):525-527.

277. Pimentel H, Bray NL, Puente S, Melsted P, & Pachter L (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* 14(7):687.

278. R_Core_Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019.

279. Babicki S*, et al.* (2016) Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* 44(W1):W147-W153.

280. Conway JR, Lex A, & Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18):2938-2940.

281. Gladyshev EA, Meselson M, & Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320(5880):1210-1213.

282. Robinson JT*, et al.* (2011) Integrative genomics viewer. *Nat. Biotechnol.* 29(1):24-26.

283. Niimura Y & Nei M (2005) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences* 102(17):6039-6044.

284. Robertson HM & Wanner KW (2006) The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family. *Genome research* 16(11):1395-1403.

285. Robertson HM, Warr CG, & Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* 100(suppl 2):14537-14542.

286. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4):772-780.

287. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-1797.

288. Guindon S*, et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59(3):307-321.

289. Lefort V, Longueville J-E, & Gascuel O (2017) SMS: smart model selection in PhyML. *Molecular biology and evolution* 34(9):2422-2424.

290. Brückner A & Heethoff M (2016) Scent of a mite: origin and chemical characterization of the lemon-like flavor of mite-ripened cheeses. *Exp Appl Acarol* 69(3):249-261.

291. Heethoff M, Laumann M, Weigmann G, & Raspotnig G (2011) Integrative taxonomy: Combining morphological, molecular and chemical data for species delineation in the parthenogenetic Trhypochthonius tectorum complex (Acari, Oribatida, Trhypochthoniidae). *Frontiers in Zoology* 8:2.

292. Cao Z*, et al.* (2013) The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. *Nature communications* 4(1):1-10.