Genomic and metagenomic analyses of the domestic mite Tyrophagus Putrescentiae identify it as a widespread environmental contaminant and a host of a basal, mite-specific Wolbachia lineage (supergroup Q)

Genomic and metagenomic analyses of a domestic mite *Tyrophagus putrescentiae* identify it as a widespread environmental contaminant and a host of a basal, mite-specific *Wolbachia* lineage

Abstract

Background

Tyrophagus putrescentiae (mold mite) is a global, generalist species that commonly occurs in various human-created habitats, causing allergies and damaging stored food; but it has not been found in healthcare settings so far. The mite's properties as a ubiquitous trophic generalist are attributed to symbiotic bacteria housed in specialized internal mite structures; however, a recent work suggests that horizontal transfer of bacterial/fungal genes may be also involved. Bacterial associations of *T. putrescentiae* include an uncharacterized and genetically divergent Gram-negative bacterium (*Wolbachia*) displaying blocking / microbiome modifying effects. The phylogenomic position and supergroup assignment of this bacterial species are unknown.

Results

An extensive analysis of GenBank data shows that (i) *T. putrescentiae* DNA is a substantial source of contamination in public sequence databases and (ii) the mite occurs in the lab and healthcare settings. Our phylogenomic analysis of *Wolbachia* recovered a basal, mite-specific lineage (supergroup Q) represented by two *Wolbachia* species associated with the mold mite and a gall-inducing plant mite (*Fragariocoptes setiger*). Fluorescence *in situ* hybridization confirmed the presence of *Wolbachia* inside the mold mite. *T. putrescentiae* also forms associations with bacteria implicated in healthcare-acquired infections that can rapidly develop antibiotic resistance (*Pseudomonas aeruginosa*, *Escherichia coli*, and *Stenotrophomonas maltophilia*). Despite the presence of diverse bacterial communities in *T. putrescentiae*, we did not detect any recent horizontal gene transfers in this mite species and/or in astigmatid (domestic) mites in general.

Conclusions

Our data will assist future research in elucidating the nature of the mite bacterial communities and the mite's ability to spread bacteria involved in healthcare-acquired infections The discovery of an early-derivative *Wolbachia* lineage (supergroup Q) in two phylogenetically unrelated and ecologically dissimilar mite species indicates that this endosymbiotic bacterial lineage formed long-term association with mites and provides a unique insight into the early evolution and host associations of this bacterial genus. Further discoveries of the *Wolbachia* diversity in acariform mites are anticipated.

Key words: Domestic mites, mite-bacteria trophic symbiosis, phylogenomic tree of *Wolbachia*, basal mite-specific supergroup of *Wolbachia*, environmental genomics, healthcare-acquired infection.

Introduction

The mold mite, *Tyrophagus putrescentiae*, is a common, microscopic, and globally distributed species of domestic mites [1, 2]. It is a generalist species, living in nearly every terrestrial habitat with relative humidity >65% [3, 4]. It frequently infests stored products, resulting in economic loss, and it also causes anaphylaxis in sensitized individuals consuming mite-contaminated food [5]. After pyroglyphid house dust mites, *T. putrescentiae* is the second most medically important species responsible for indoor allergies in humans [6-8] and domesticated animals [9]. Thanks to its long body setae it can disperse via air currents, or by attaching to human clothing and household items, or it can rapidly move on its own [10, 11]. *T. putrescentiae* is arguably the first known animal space hitchhiker found onboard a human-inhabited spacecraft in low Earth orbit [12]. The ubiquity of *T. putrescentiae* is well documented by a large body of literature [13], however, its reliable identification was only possible after 2007-2009, when accurate methods based on morphology and DNA sequences were developed [1, 13]. Even though the mite has been previously reported to infest fungal, insect, and plant tissue cultures in laboratory and industrial settings [14, 15], its role as a significant factor in DNA sequence contamination is currently underappreciated. Since most bioinformatics tools focus on bacterial and human DNA contamination, DNA of a microscopic eukaryote may evade the NCBI GenBank standard quality check procedure and be inadvertently incorporated into public sequence databases. A whole genome sequence of *T. putrescentiae* would alleviate this situation.

The mold mite is a vector of various bacterial and fungal microorganisms in human-related habitats and agricultural settings [10, 16]. The presence of bacteriocytes, large, compartmentalized bacterial colonies in the mite parenchymal tissues, is one of the most distinctive anatomical features of *T. putrescentiae* [17]. Bacteria may be very important in the mite's nutritional ecology,

cooperating with their acarine hosts to use nutrients from different sources. Associated bacteria may provide chitinolytic enzymes digesting fungal cell walls [18], making this system ecologically important for mobilizing nitrogen from chitin [19]. However, a recent work suggests that bacterial or fungal genes may have been incorporated into the mite genome, and thus extend the mite' nutritional functionality [20]. For example, the horizontally transferred genes encoding UDP glucuronosyltransferases and several fungal cell wall lytic enzymes could enable detoxification and digestive functions of their acarine hosts [20].

Bacterial associations of *T. putrescentiae* include an uncharacterized Gram-negative bacterium (*Wolbachia*) found in several *T. putrescentiae* populations based on previous 16S rRNA and protein sequencing [8, 21]. *Wolbachia* is an intracellular endosymbiont associated with various arthropods and nematodes. This bacterium may form nutritional symbiosis with several hosts [22], and it has been demonstrated to be a useful agent to control human pathogens, such as dengue virus vectored by mosquitoes [23]. A blocking / microbiome modifying effect has been recently shown for the *Wolbachia* from *T. putrescentiae* [8], suggesting that this *Wolbachia* may be potentially useful for disease/pest control applications.

Here we sequenced the whole genome of *T. putrescentiae* from North America and use transcriptomes previously generated by us for several mite populations from Europe and USA. We present a genomic assembly of the mite and metagenomic analyses of its microbiome. Based on these genomic and metagenomic analyses, we answer the following questions: 1) Can the mite be a significant source of DNA contamination in the laboratory, industrial and healthcare settings? 2) Are there recent horizontal gene transfers from bacteria/fungi that can account for the mite' extended nutritional functionality? 3) What are the phylogenomic affinities of the previously uncharacterized mite-specific *Wolbachia*?

Results

Genome of *Tyrophagus putrescentiae*

We conducted a series of independent metagenomic assemblies in metaSpades and Megahit (**Supplementary Tables S1-S4 online**). Based on different metrics, most importantly, transcriptome mapping statistics (**Supplementary Table S1 online**), we selected a metaSpades (k=21,33,55) assembly as our preferred assembly, consisting of 176,943 scaffolds, with the total length of 151,679,586 bp and average coverage of 518.3. Of them, 9,303 scaffolds (79.5 Mb) matched the transcriptomic assembly (**Supplementary Fig. S1 online**). Of the 176,943 initial scaffolds, BLAST searches were able to identify 125,524 scaffolds (136,606,678 bp, average coverage = 440.4) (**Supplementary Fig. S2a online**). This analysis revealed that our metagenomic assembly had a substantial portion of non-mite sequences, mostly bacterial DNA. Most of the non-mite scaffolds had low coverage, i.e. below 100x, with the notable exception of the bacterium *Alcaligenes faecalis*, which had a coverage of slightly above 100x (**Supplementary Fig. S2a online**). Many scaffolds lacking BLAST hits, therefore, still could be confidently classified as belonging to the mite based on their higher coverages, i.e. >600 (**Supplementary Fig. S2a online: grey color**). After filtering scaffolds using a combination of the BLAST classification result and/or coverage information (detailed in the section "Metagenomic decontamination"), our final decontaminated mite assembly had a total of 19,731 scaffolds, with the length of 95,135,691 bp and average coverage of 1024.7 (**Fig. 1a,b, Supplementary Table S1, Fig. S2b online**). Busco statistic using the arachnida_odb10 database (2934 genes) was compatible to that of a recent long + short read assembly of a Hong Kong population: Complete:89.5%[Single-copy:83.4%,Duplicated:6.1%], Fragmented:3.2%, Missing:7.3% vs Complete:89.8%[Single-

copy:85.1%, Duplicated:4.7%], Fragmented:3.1%, Missing:7.1%, (Hong Kong: GCA_021730765.1) (**Supplementary Fig. S3 online**). The mitochondrial genome has the typical gene order of Astigmata [24, 25] (**Fig. 1b**). Top gene ontologies summarized for three sub-ontologies were as follows: membrane and membrane part (cellular component), catalytic activity and binding (molecular function), and metabolic and cellular processes (biological process) (**Supplementary Fig. S4 online**). Our phylogenomic analysis inferred *T. putrescentiae* within the Astigmata, a major general mite lineage (**Fig. 1c**). Astigmata evolved within soil mites (Oribatida) with absolute support (**Fig. 1c**), while other high-level relationships were similar to those inferred previously [20, 26].

Extensive contamination of GenBank databases by *T. putrescentiae* sequences

Our BLAST searches detected extensive contamination of public sequence databases with *T. putrescentiae* sequences (**Table 1**). The GenBank reference genome database contained a '*Rhagoletis zephyria*' genome GCF_001687245 [27] heavily contaminated with *T. putrescentiae* DNA, 4901 scaffolds (**Table 1**). Another important example, is the mite sequence reported as a bacterium "*Shinella* sp." from healthcare settings [28] (**Table 1**). Other contaminated GenBank sequences include many species of insects and fungi, as well as vertebrates, round worms, bacteria, and plants (**Table 1**); these sequences also have matches on the '*Rhagoletis zephyria*' genome, which offers an independent confirmation of contamination. We also run BlobTools analyses based on the GenBank nucleotide database and our clean assembly of *T. putrescentiae* as a query (**Supplementary Fig. S2b online**). The BlobTools analysis largely agrees with our results, and identifying about 96.5 Mb of *T. putrescentiae* DNA in the '*R. zephyria*' genome (**Supplementary Fig. S2a online**). However, except for five GenBank sequences of *Drosophila melanogaster*

classified as contaminants. However, these hits had either low bitscore or identity, or included hits on ultraconservative regions (i.e. nuclear rRNA). We therefore, do not consider these *D. melanogaster* sequences as contaminated (not included in Table 1). A similar situation occurred when a BlobTools analysis was run on a custom BLAST database, including the GenBank nucleotide collection and the '*R. zephyria*' genome. The following taxa were identified by BlobTools as 'contaminants': *Ixodes scapularis*, *Drosophila biramipes*, *Plutella xylostella*, *Cyprinodon variegatus* (**Supplementary Fig. S2b online**), again with low score/low identity, and probably representing false positives.

Mold mite harbors diverse bacterial communities

We profiled our metatranscriptome (Europe) and metagenome (Mexico) NGS datasets using (i) Kraken to classify raw reads and (ii) BLAST to assign a taxonomic classification to scaffolds (assembled reads). For the two classification strategies, we used standard databases plus our clean genomic assembly of *T. putrescentiae* (GCA_012066115) to avoid false positive hits, i.e. incorrect classifications of *T. putrescentiae* sequences as different eukaryotic organisms. Kraken identified the following bacterial species having 2% or more reads classified as Bacteria (**Table 2**): *Alcaligenes faecalis* (58%), *Pseudomonas aeruginosa* (9%), and others (all 3-4%): *Stenotrophomonas maltophilia*, *Stenotrophomonas* sp. PAMC25021, *Advenella kashmirensis*, and *Achromobacter denitrificans*(**Fig. 2a**). Kraken identified the yeast *Candida dubliniensis* as the most abundant fungal species (99% of all fungi, magnitude 245,252). However, the BLAST search did not find this OTU at all; instead, a different species, *Candida parapsilosis*, was detected in trace amounts (1,717 bp assembly, 42 mapped reads) (**Table 2**). We also detected trace amounts of low-scoring reads for Apicomplexa (**Table 3**), which are known internal parasites of mites [29].

The metatranscriptomic dataset contained *Escherichia coli*, *Wolbachia* (unidentified divergent species, see below), and *Bacillus* (including *Bacillus thuringiensis, Bacillus cereus*) with high abundance, 15-36% of all bacterial reads (**Table 2**). Fungi were represented by *Fusarium*; and apicomplexans were represented by an unidentified coccidian taxon (different from the Apicomplexa from the metagenomic dataset), whose nuclear rRNA was similar to that of *Adelina* (92%) (**Table 2**). Here, many species reported by the Kraken software in the transcriptome (European mite populations) could not be confirmed by BLAST searches (**Table 2**). This is an expected outcome, as Kraken's use of short k-mers inherently leads to lower accuracy.

Based on the intersection of the DNA and RNA samples by BLAST, only one species was found to occur in both samples, *Cutibacterium acnes* (**Table 2**). However, this bacterium is a widespread contaminant of DNA extraction kits [30] and we consider its presence as a probable artefact, but see [31].

Fluorescence *in situ* hybridization detected *Wolbachia* (**Fig. 3a**), *Rickettsia* (**Fig. 3b**), in the mite's parenchymal tissue bacteriocytes; Eubacteria were mostly associated with the digestive track and parenchymal tissue (**Fig. 3e, d**); *Wolbachia* was also found in ovaries and eggs (**Fig. 3a, g**).

No recent horizontal gene transfers in the mold mite or Astigmata

We detected eight putative HGT events from the following lineages: Bacteria (5), Fungi (2), and Amoebozoa (1) (**Table 3, Figs 3b-d, Supplementary Figs. S5-11 online, Supplementary Dataset S1 online**). Among them, was the D-Ala-D-Ala dipeptidase gene previously suggested to be laterally transferred to Astigmatid mites. However, all these putative HGTs had very low amino acid similarity (40.3-60.71%) to the corresponding proteins of *T. putrescentiae* (**Table 3**) and also

had significant matches to Oribatida, Endeostigmata, or Trombidiformes, which are major early-derivative acariform mite lineages as compared to Astigmata (**Table 3**). This evidence strongly indicates that no HGT events occurred at the origin and during the evolution of Astigmata, a major lineage that includes the mold mite and other domestic mites.

An early derivative, mite-specific supergroup of *Wolbachia*

Our phylogenomic analysis recovered two species of *Wolbachia* associated with mites (the mold mite *T. putrescentiae* and the gall mite *Fragariocoptes setiger*) forming a monophyletic group in a basal portion of the tree (**Fig. 4**). As genomic-scale data are not available for many *Wolbachia*, we attempted to identify this lineage through the use 16S or multi-locus analyses (many taxa, few loci). These analyses can demonstrate the affinity of an unknown sequence, but generally they cannot resolve phylogenetic relationships among the *Wolbachia* supergroups. In the 5-gene analyses, *Wolbachia* from *T. putrescentiae* (wTp) was grouped with *Wolbachia* from the quill mite, *Torotrogla cardueli* classified previously in supergroup Q [32] (**Supplementary Fig. S12 online**). 16S also placed wTp within a general grouping that includes *Wolbachia* associated with pratylenchid nematodes, *T. cardueli*, and *Bryobia* (a plant-feeding mite) (**Supplementary Fig. S13 online**). We therefore identify genomic sequences of wTp as part of supergroup Q, which was proposed previously based on a 5-locus sequence data and a different set of mite hosts [32].

Supergroup Q was sister to supergroup M (*Wolbachia* from the banana aphid *Pentalonia nigronervosa*). The lineage representing supergroups Q+M was recovered as sistergroup to supergroup L from the lesion pratylenchid nematode *Pratylenchus penetrans* (plus an environmental soil sample). The entire lineage L(Q+M) was recovered as sistergroup to the

199 remaining diversity of *Wolbachia*. In addition, our genomic-scale analysis showed that previously
200 proposed *Wolbachia* supergroup O [32, 33] is nested within supergroup B (**Fig. 3**).
201

202 ## Discussion

203 The common domestic mite, *Tyrophagus putrescentiae* is a ubiquitous generalist species
204 associated with human-created habitats, such as houses, retail stores, storehouses, food-processing
205 facilities, and research laboratories. Here we discuss several questions related to whether it can be
206 a source of significant DNA contamination in public databases, and whether its microbial trophic
207 symbiosis and/or horizontal gene transfers can contribute to its remarkable ability of being a broad
208 dietary generalist. As our GenBank analysis identified it in hospital settings, we briefly discuss
209 whether the mite could be one of the factors influencing the spread of bacteria responsible for
210 hospital acquired infection. Furthermore, we elucidate the phylogenetic relationship of a novel
211 *Wolbachia* bacterium associated with this mite.

212 Bacteria and microscopic fungi are expected to introduce sequence contamination into
213 eukaryotic whole genome sequences because they can be symbionts, and/or originate from the
214 environment, laboratory equipment, DNA extraction kits or reagents [34, 35]. Non-fungal
215 eukaryotic DNA usually is not considered a significant source of contamination, except for human
216 DNA, particularly its repetitive elements [36]. Here we show that the mold mite, *T. putrescentiae*,
217 is an important environmental contaminant that can make its way inside various DNA samples
218 (Table 1), and its contaminating DNA can evade detection by metagenomic methods as commonly
219 used in practice [35, 36]. The most conspicuous example of contamination is the GenBank
220 reference genome of a fly *Rhagoletis zephyria* GCF_001687245 [27] containing 4901 scaffolds
221 and about of 100 MB of *T. putrescentiae* DNA (**Table 1**). For the sequences deposited into

GenBank as insects and fungi, it is very likely that laboratory cultures of these organisms were infested by the mite, and this was unnoticed by the researchers. Suspect sequences of other organisms may result from inadvertent lab contamination. For example, the lab that deposited sequences of *T. putrescentiae* as sequences of a tree (*Intsia palembanica*), also published sequences of *T. putrescentiae*, so spill-over contamination from mite cultures maintained in the same lab is possible. Sequences deposited as a mouse, bird, and worm (**Table 1**) may have resulted from environmental contaminations as the mite is ubiquitous and readily reproduces in many humid environments [37]. This contamination is not surprising because fungal/insect/cell culture infestation by the mite was well known from the literature prior to the advent of molecular techniques [38]. In addition, the case of the phytoseiid *Neoseiulus cucumeris* (**Table 1**), a predatory mite used for biological control, may be attributable to *T. putrescentiae* being used as the food for the predatory mite, which is a common practice in mass production of phytoseiids [39].

Our metagenomic profiling of two independent mite samples indicate that in the Mexican sample, the most abundant bacterial species were *Alcaligenes faecalis* (58% of all Bacteria) and *Pseudomonas aeruginosa* (9%) (**Table 2, Fig. 2a**); in the European sample, *Escherichia coli* (36%), *Bacillus* spp. (18%), and *Wolbachia* (15%) were dominant (**Table 2**). The fact that the two mite microbiomes were completely different across the two samples suggests the importance of local factors (such as food type, habitat, and available bacterial communities) in forming the microbiomes of this globally distributed mite species. The same effect was observed previously in local mite populations that harbored different microbiomes, although only a single taxon, a *Solitalea*-like bacterium, was shared across mite populations [17, 40]. Because our two microbiomes do not display commonalities in taxonomic composition, it is likely that the mite can opportunistically recruit available local bacterial species having a chitinolytic activity and/or other

useful properties rather than form permanent specialized associations with a fixed set of bacterial species. Several bacteria identified by us do display chitinolytic properties, i.e. *Stenotrophomonas maltophilia* (abundance: 3% of all bacterial species) and *Serratia liquefaciens* (2%) from the Mexican sample; and *Bacillus cereus* (5%) from the European sample [19, 41, 42]. These bacteria were isolated from *T. putrescentiae* previously and their chitinolytic properties were demonstrated experimentally [19, 42]. Furthermore, *Alcaligenes faecalis*, the most common bacterium in the Mexican sample, was also found to be an effective chitinase producing bacterium on marine waste [43]. These bacterial chitinolytic features allow metabolizing chitin from fungal cell walls; when their acarine hosts feed on fungus-rich diets, these bacteria probably cooperate with the mites, forming a nutritional symbiosis [19]. In this system, the mite, through its normal feeding activities, shreds the fungal mycelium (thus making it available for the microorganisms) and disperses both bacteria and fungi [18, 19, 44], possibly promoting the spread of genetic variants and increasing recombination rates of these microorganisms on a local scale [45]. Several bacteria associated with *T. putrescentiae* are implicated in hospital acquired infections [46, 47]: *Pseudomonas aeruginosa*, *Escherichia coli*, and *Stenotrophomonas maltophilia* (Table 2). These bacterial taxa are the first, second, and eighth bacterial species most frequently isolated in hospital settings in the US [48]. All these bacteria can rapidly develop resistance to multiple classes of antibiotics [46, 49-51], leading to high morbidity and mortality among hospitalized patients, particularly in intensive care units [46, 47, 49, 52]. The mite itself has been reported in healthcare settings as the bacterium "*Shinella* sp." [28]. Given the known ability of *T. putrescentiae* to vector various microorganisms (see above), it is possible that in hospital settings, the mite could be one of the factors influencing the spread of pathogenic bacteria and promoting exchange of bacterial genetic elements

267 responsible for antibiotic resistance. Further studies on the role of this mite in dispersing antibiotic-
268 resistant bacteria are needed.

269 We did not detect any recent horizontal gene transfer events in the *Tyrophagus*
270 *putrescentiae* genome. All HGT events occurred prior to the origin of Astigmata, in the common
271 ancestor of either acariform or oribatid mites (**Table 3**). This includes D-alanyl-D-alanine
272 dipeptidase, which was previously suggested to occur within Astigmata [20], however, we found
273 this gene in Endeostigmata and Oribatida (**Table 3, id=9**), indicating an earlier origin. Given that
274 all detected HGTs were ancient, we suggest that they do not have immediate relevance to the mite
275 abilities to be a widespread nutritional generalist.

276 A very divergent *Wolbachia* species (wTp) was found in the European sample and in the
277 contaminated '*Rhagoletis zephyria*' genome originated from the USA. On our tree, wTp and
278 *Wolbachia* from the gall mite *Fragariocoptes setiger* formed a basal monophyletic lineage
279 (supergroup Q), indicating that this lineage may be specific to acariform mites (**Fig. 4**). It is
280 possible that wTp may cause cytoplasmic incompatibility in its host, explaining the results of early
281 breeding experiments that demonstrated large-scale reproductive incompatibility between
282 morphologically similar populations of *T. putrescentiae* [53]. However, an alternative explanation
283 of these experiments could be the presence of two sibling mite species, *T. putrescentiae* and *T.*
284 *fanetzhangorum*, which are separated by large genetic distances and probably cannot interbreed
285 [1, 13, 54]. *Wolbachia* has recently gained medical relevance because of their ability to affect
286 transmission of human pathogens such as dengue virus vectored by mosquitoes [23]. This
287 bacterium can manipulate its hosts via pathogen blocking, which limits the ability of many
288 pathogenic viruses, bacteria and nematodes to grow in the host [55, 56]. Several hypotheses have
289 been proposed to explain the mechanism of pathogen blocking . Among them, the lipid

perturbations hypothesis, suggesting that *Wolbachia* may outcompete pathogens for lipids, a critical nutritional resource, seems to be better supported by experimental evidence [57]. A strong pathogen blocking effect has been observed when naturally uninfected mosquitoes were transinfected with *Drosophila*-specific strains of *Wolbachia* [23, 58, 59]. wTp has also been observed to affect the associated microbiomes in its own mite host, *T. putrescentiae* [8]. Here, by analogy with the *Drosophila*-specific *Wolbachia* exhibiting a strong pathogen blocking effect in an unnatural host (mosquitoes), coupled with the ability of wTp to modulate the host's microbiomes, we suggest that further research needs to be done to elucidate whether wTp can be used control pathogens vectored by ticks and other parasitic Acari.

Our mold mite assembly was used to conduct phylogenomic analysis of acariform mites using all available genomes of acariform mites. This analysis nferred Astigmata within soil mites (Oribatida) with absolute support (**Fig. 1c**). This result is consistent with the leading morphological hypothesis suggesting a single origin of opisthosomal glands within the Oribatida+Astigmata lineage [60] and relationships inferred in several multilocus molecular studies based on Sanger sequencing [61, 62], but contrasts with sister group relationships of Astigmata and Oribatida inferred in recent phylogenomic analyses [20, 26, 63, 64].

Conclusion

We assembled a whole genome of the mold mite, *Tyrophagus putrescentiae*, and showed that this mite species is a significant contaminant in laboratory, hospital, and industrial settings. Our genomic assembly was used to detect contamination in GenBank, and should be employed, as a common practice, to prevent further contamination. Microbiome profiles of the samples from Europe and Mexico were completely different, suggesting that mite-bacterial symbiotic

associations are formed via opportunistic recruitment of locally available bacterial species by the acarine host. We also found that the mold mite is a potential carrier of several bacteria associated with hospital-acquired infections, most importantly *Pseudomonas aeruginosa*. Finally, based on whole transcriptome sequence of a novel, mite-specific *Wolbachia* from *T. putrescentiae*, we identified it as part of a basal, mite-specific *Wolbachia* lineage (supergroup Q). These findings provide a unique insight into the early evolution and host associations of this bacterial genus. Based on the previously known blocking / microbiome modifying effect of the *Wolbachia* from *T. putrescentiae*, we believe that this bacterium may be potentially useful for disease/pest control applications. We expect further discoveries of the *Wolbachia* diversity associated with acariform mites.

Methods

Sample, library preparation, sequencing, and metagenomic assembly

Genomic sequencing was done from a single female reared in a culture maintained at the University of Michigan Museum of Zoology, Ann Arbor, MI, USA, at room temperature, relative humidity 75-100%, using Tetra® TetraMin Large Tropical Flakes as the food source. This culture was started from specimens with the following collecting data: MEXICO: Ciudad de México, Parque Ecológico de Xochimilco, nr. Lago Acitlalin, 19.297115 -99.092799, rotten reed stalk (*Typha*), 03 Jan 2017, P. Klimov (coll.), UMMZ accession BMOC 17-0108-002.

Genomic DNA was extracted from a single female specimen using a QIAamp DNA Micro kit (Qiagen). An Illumina sequencing library was generated from a single mite female using the KAPA HyperPlus Kit. The insert size was 322 bp. Sequencing was done on an Illumina HiSeq-4000 instrument, generating 755,504,138 (377,752,069x2) 150x2 bp paired-end reads. Read

quality was evaluated in FastQC [65]. Quality filtering and adaptor content removal was done in bbtools v. 38.23 (https://sourceforge.net/projects/bbmap/) as detailed in Supplementary Material online: section 1. Four metagenomic assemblies were run (**Table 1**), in Megahit [66] and MetaSPAdes 3.12.0 [67] with three different sets of kmer sizes ("-k"): 21,33,55; 21,33,55,77; 21,33,55,111. For assembly evaluation, the following three statistics were used: (i) basic, reference-free statistics, e.g., N50, L50 in QUAST v.5.0.0 [68] (**Supplementary Table S2 online**); (ii) alignment against the transcriptome using HISAT2 v.2.1.0 [69] and RNAquast v.1.5.1 [70] (**Supplementary Table S1 online**), and (iii) alignment against the "*Rhagoletis zephyria*" genome in QUAST (**Supplementary Table S3 online**). These programs were run using Unix shell scripts, for detail see **Supplementary Material online**: section 2. In addition, our final decontaminated assembly (see below) was evaluated by finding sets of single-copy, orthologous genes specific to Arachnida in BUSCO v.5.3.0 [71]. Our transcriptome assembly (GenBank TSA accession GIFQ00000000.1) was described earlier [8].

Assembly annotation

Gene prediction and annotation was done in the maker genome annotation pipeline v2.31.10 [72] in three steps: (i) we directly used our transcriptome and non-redundant GenBank proteins from Ecdysozoa as the annotation evidence (est=transcriptome.fas; protein=Ecdysozoa_prot.fa; est2genome=1; protein2genome=1 in the configuration file maker_opts.ctl); these imperfect gene models (ii) were then used to train the gene prediction program SNAP [73] bundled with maker (snaphmm=snap1.hmm; est2genome=0; protein2genome=0); a new set of gene annotations generated in this step (iii) was then used to train the gene predictor yet again (snaphmm=snap2.hmm). For conserved protein domains, gene ontologies were determined in

InterProScan v.5.38-76.0 [74]. These ontologies and standardized gene names, were assigned to maker annotations using maker accessory scripts. Top gene ontologies were summarized in WEGO v2.0 [75] using the InterProScan output.

D. farinae miRNAs were annotated by miRDeep2 [76] software and curation of miRNAs based on standard miRNA features. *T. putrescentiae* miRNAs were found using BLAST to identify regions homologous to *D. farinae* miRNAs as well as those deposited for chelicerate arthropods in miRbase. Ago/Piwi homologs were likewise found with BLAST. Each putative Ago/Piwi protein was verified to encode a PAZ and PIWI domain.

Mitochondrial genome annotation was done using multiple lines of evidence: sequence similarity with two related species, *Sancassania berlesei* (KF499016) and *Aleuroglyphus ovatus* (NC_023778.1) [25, 77]; *T. putrescentiae* EST data (GenBank accession: SAMN00174981 ID: 174981); ARWEN [78] with minimal tRNA search methodology as described previously [24]; and the Mitos Web Server for automatic prediction of all mitochondrial genes [79].

Mite phylogenomics

We analyzed 48 genomes of acariform mites (n=34), including our *T. putrescentiae* assembly, parasitiform mites (outgroups, n=13) and *Limulus polyphemus* (distant outgroup, n=1). BUSCO v5.3.0 [71, 80] analyses was run to identify phylogenetically conserved, single-copy genes using the arachnida_odb10 database. Alignments were done in mafft v7.490 [81]: mafft --thread $proc --inputorder --bl 62 $i > $i.mafft. Alignment quality trimming was done in trimAl v.1.4.1 [82]: trimal -in $i -out $i.trimal -automated1 -resoverlap 0.75 -seqoverlap 80. A subset of 415 protein alignments (occupancy≥0.8 and length≥200) were used for a partitioned analyses in IQ-TREE

v.2.2.0 with automatic protein model selection and partition merging [83]: iqtree2 -s $ipf -p $ipf --seqtype AA -T AUTO --merge -rclusterf 10 -m MFP -alrt 1000 -bb 1000 -safe --prefix $ipr.

Metagenomic decontamination

Detailed decontamination procedure is described in Supplementary Material online: section 3. Briefly, to classify the 176,943 initial metagenomic scaffolds, we ran a local BLAST v.2.7.1 [84]. Hits on contaminated GenBank assemblies, Illumina technical sequences (PhiX), and *Homo sapiens* contamination were removed; high-scoring hits on Bacteria and other non-eukaryotic organisms were filtered out; then high-scoring hits on "*Rhagoletis zephyria*" and *Tyrophagus* (bitscore >=300 or identity >=95%) were filtered and their coverages were noted. Entries unclassified by BLAST were classified as mite sequences based on their coverages >600; all sequences shorter than 300 bp were removed. The mitochondrial DNA and ribosomal RNA scaffolds were identified, annotated manually and trimmed. The final assembly, named here P3F6, had a total of 19,731 scaffolds (length 95,135,691 bp). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAAALH000000000 (BioSample SAMN13712654, BioProject PRJNA598686, assembly GCA_012066115.1).

Metagenomic profiling

Detailed methodology is given in Supplementary Material online: section 4. Briefly, raw Illumina reads were processed to remove adapter sequences, low quality data, and artefacts in bbmap 38.51. For these processed reads, we assigned taxonomic classifications in Kraken2 v2.0.8-beta [85] with the confidence parameter of 0.1, followed by abundance estimation in Bracken [86]. Three analyses were run each using a separate Kraken library: (i) Basic with standard Kraken databases:

archaea, bacteria, viral, human, plant, fungi, protozoa; (ii) Custom1 (basic plus the '*Rhagoletis zephyria*' genome); Custom2 (basic plus the *T. putrescentiae* P3.F6 assembly).

To classify scaffolds, we used BLAST searches with the nucleotide blastdb5 database (downloaded May 16 2019). We also constructed several custom BLAST databases using our metagenome, metatranscriptome, *T. putrescentiae* assembly (P3.F6), and the '*Rhagoletis zephyria*' genome. Intersection between the DNA and RNA assemblies was determined using the following criteria to classify sequences as belonging the same species/OTU: bitscore ≥ 500 and identity≥95%. Full taxonomic lineage information was added to blast/diamond outputs using a custom script (Supplementary Material online: section 6). For metagenomic profiling of assemblies, we used BlobTools [87], an analysis which uses three lines of evidence: coverage, GC content, and BLAST or DIAMOND classification. For assigning a unique classification to multiple nt BLAST hits, we used the BlobTools' 'bestsum' algorithm. Because this and other BlobTools classification algorithms may return false positives, we did not BlobTools to automatically remove contaminants.

Horizontal gene transfer (HGT)

Previously described methodology was used [88]. Using the Unix command awk, we parsed the Uniref50 proteins (downloaded Jul 11 2022) into two groups: (1) no Metazoa; and (2) Metazoa minus Arthropoda. TaxIds were extracted from the Uniref50 fasta headers. Blacklisted TaxId (178133=plant+phytophagous eriophyoid mite) and 46 TaxIds not found in GenBank taxonomy were removed. Diamond v0.9.14.115 were used to run the mite coding sequences (see the section Assembly annotation) against two databases build from the two sets of proteins. Using bitscores from the two diamond searches, HGT indices were calculated for each coding sequence as

described previously [88]. Our entire HGT discovery pipeline was documented as a Unix shell script in Supplementary Material online 12: section 7. Because UniRef50 protein clusters (50% sequence similarity) are labelled by common taxonomy of the cluster (rather than by the representative taxon) [89], we considered protein clusters having a high taxonomic rank as conserved. For example, regardless of its HGT score, a cluster labelled as "cellular organisms|Eukaryota|" was deemed as conserved across Eukaryota, rather than being result of HGTs.

Wolbachia endosymbiont: Metatranscriptomic assembly and phylogenetics

We sequenced a metatranscriptome of *T. putrescentiae* from Europe (GIFQ00000000) and assembled it in CLC Genomics Workbench v11 (Qiagen). Collection detail and bioinformatics methodology for this sample were described previously [8]. *Wolbachia* contigs were identified using BLAST and DIAMOND v0.9.24.125 [90]. Because the rRNA fraction was depleted in the transcriptome, the 16S rRNA gene was recovered separately by mapping adaptor-free, artefact-free, quality trimmed and filtered reads onto the known 16S *Wolbachia* sequence (GCA_000829315.1) following assembly in rnaSPAdes v3.13.0. The final *Wolbachia* transcriptomic assembly had 280 contigs with a total length of 925,767 bp (average coverage 742.8), approaching the typical *Wolbachia* genome size range, 1.3-1.6 Mb [91]. Phylogenetic inferences were done for four datasets: genomic, five standard phylogenetic loci [32], five MLST loci [92], and 16S rRNA (see supplementary table S4 online: standard phylogenetic loci accession ids; **Supplementary Datasets S2-4 online**: nexus alignments). For the former analysis, we used 169 GenBank genomes downloaded from GenBank plus our assembly: 31 outgroups (*Ehrlichia*, *Anaplasma*) and 139 *Wolbachia* ingroups. BUSCO5 analyses was run to identify phylogenetically

conserved, single-copy proteins at the level of Rickettsiales. A Maximum Likelihood tree was inferred in IQ-TREE v.2.2.0 using a partitioned analysis and automatic model selection [83]. Our entire workflow is described in detail in **Supplementary Material online: section 5**.

Fluorescence in situ hybridization (FISH)

FISH was performed using universal and specific bacterial probes. *T. putrescentiae* adults were first fixed in 4% formaldehyde. For the hybridizations we followed the FISH protocol described in Perotti et al. [93], then mounted as whole-specimens. Samples were incubated at 45°C in darkness for up to 20 h, washed for 1 h in hybridization buffer followed by PBTA (phosphate buffer with Triton X-100 plus sodium azide) at room temperature. Then, mites were mounted in PBS/ glycerol and photographed under the confocal microscope. A number of bacterial probes were used in different observations: EUB-338 [94] and EUB-338 II and III [95], *Rickettsia* [96] and *Wolbachia* [97] (equimolar mixed in the hybridization buffer (following remarks of ProBase)). No probe and competition suppression controls were performed. A Confocal Zeiss LSM510 microscope with Coherent Multiphoton laser was used. For these experiments, we used cultures originating from the Crop Research Institute, Prague (Czech Republic) [8] and the University of Reading lab, UK (maintained since 2008, stock colony received from the Food and Environment Research Agency, UK).

Data Availability Statement

The data underlying this article are available in GenBank: assembled genome of *Tyrophagus putrescentiae* (GCA_012066115.1, WGS project JAAALH000000000), assembled metagenome of *T. putrescentiae* (GCA_013316015.1, JAALOO000000000) and the source short Illumina reads

(SRA accession: PRJNA598686; SRA run: SRR11069688); assembled metatranscriptome of *T. putrescentiae* (GIFQ00000000) and the source short Illumina reads (SRR7903714-SRR7903734); assembled transcriptome of *Wolbachia* endosymbiont of *T. putrescentiae* (GIJY01000000).

Supplementary Material

Supplementary data are available at BMC Genomics online.

Acknowledgements

Datasets were processed on the Tyumen University high-performance computing computer cluster (Tyumen State University, Russia), Sagarana cluster CEPAD-ICB-UFMG (Federal University of Minas Gerais, Brazil), and the Bell community cluster at Purdue University (West Lafayette, USA). José Miguel Ortega (Federal University of Minas Gerais, Brazil) made the Sagarana cluster available to us and gave related technical advice. This work was supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of the Federal Scientific and Technical Program for the Development of Genetic Technologies for 2019-2027 (agreement №075-15-2021-1345, unique identifier RF----193021X0012) to A. V. Lisitsa.

Author contributions

All authors contributed to data analysis and all approved the final version of the manuscript. P.B.K designed the study, collected the Mexican sample, and performed all analyses. J.H. provided the metatranscriptomic dataset and assembly. M.A.P and H.R.B. provided the FISH data. P.B.K, J.H., T.E., Y.C., M.A.P., H.R.B., A.F., and Q.H. interpreted the data and wrote the manuscript.

Literature Cited

1. Klimov PB, OConnor BM. Conservation of the name *Tyrophagus putrescentiae*, a medically and economically important mite species (Acari: Acaridae). *International Journal of Acarology*. 2009, 35(2):95-114.
2. Hughes AM. The mites of stored food and houses, 2 edn. London: Her Majesty's Stationery Office; 1976.
3. Canfield MS, Wrenn WJ. Tyrophagus putrescentiae mites grown in dog food cultures and the effect mould growth has on mite survival and reproduction. Vet Dermatol. 2010;21(1):58-63.
4. Erban T, Rybanska D, Hubert J. Population Growth of the Generalist Mite Tyrophagus putrescentiae (Acari: Acaridida) Following Adaptation to High- or Low-Fat and High- or Low-Protein Diets and the Effect of Dietary Switch. Environ Entomol. 2015;44(6):1599-1604.
5. Sánchez-Borges M, Fernandez-Caldas E. Hidden allergens and oral mite anaphylaxis: the pancake syndrome revisited. Curr Opin Allergy Clin Immunol. 2015;15(4):337-343.
6. Arlian LG, Geis DP, Vyszenski-Moher DL, Bernstein IL, Gallagher JS. Antigenic and allergenic properties of the storage mite Tyrophagus putrescentiae. J Allergy Clin Immunol. 1984;74(2):166-171.
7. Cui Y, Yu L, Teng F, Zhang C, Wang N, Yang L, Zhou Y.ranscriptomic/proteomic identification of allergens in the mite Tyrophagus putrescentiae. Allergy. 2016;71(11):1635-1639.
8. Hubert J, Nesvorna M, Klimov P, Dowd SE, Sopko B, Erban T. Differential allergen expression in three *Tyrophagus putrescentiae* strains inhabited by distinct microbiome. *Allergy*. 2019; 74(12):2502-2508.
9. Mueller RS, Janda J, Jensen-Jarolim E, Rhyner C, Marti E. Allergens in veterinary medicine. Allergy. 2016;71(1):27-35.
10. Hubert J, Nesvorna M, Sopko B, Smrz J, Klimov P, Erban T. Two Populations of Mites (Tyrophagus putrescentiae) Differ in Response to Feeding on Feces-Containing Diets. Front Microbiol. 2018;9:2590. Published 2018 Oct 30.
11. Freitag JA, Kells SA. Efficacy and application considerations of selected residual acaricides against the mold mite Tyrophagus putrescentiae (Acari: Acaridae) in simulated retail habitats. J Econ Entomol. 2013;106(4):1920-1926.
12. Ott CM, Bruce RJ, Pierson DL. Microbial characterization of free floating condensate aboard the Mir space station. Microb Ecol. 2004;47(2):133-136.
13. Fan Q-H, Zhang Z-Q. *Tyrophagus* (Acari: Astigmata: Acaridae). Fauna of New Zealand = Ko te Aitanga Pepeke o Aotearoa (no. 56). Lincoln, New Zeland: Manaaki Whenua Press; 2007.
14. Duek L, Kaufman G, Palevsky E, Berdicevsky I. Mites in fungal cultures. Mycoses. 2001;44(9-10):390-394.
15. Walter DE, Hudgens RA, Freckman DW. Consumption of nematodes by fungivorous mites, *Tyrophagus* spp. (Acarina: Astigmata: Acaridae). Oecologia. 1986;70(3):357-361.
16. Hubert J, Stejskal V, Kubátová A, Munzbergová Z, Vánová M, Zd'árková E. Mites as selective fungal carriers in stored grain habitats. Exp Appl Acarol. 2003;29(1-2):69-87.

17. Erban T, Klimov PB, Smrz J, Phillips TW, Nesvorna M, Kopecky J, Hubert J. Populations of Stored Product Mite Tyrophagus putrescentiae Differ in Their Bacterial Communities. Front Microbiol. 2016;7:1046.
18. Smrž J, Soukalová H, Čatská V, Hubert J. Feeding Patterns of Tyrophagus putrescentiae (Sarcoptiformes: Acaridae) Indicate That Mycophagy Is Not a Single and Homogeneous Category of Nutritional Biology. J Insect Sci. 2016;16(1):94.
19. Smrz J, Catska V. Mycophagous mites and their internal associated bacteria cooperate to digest chitin in soil. Symbiosis. 2010, 52(1):33-40.
20. Xiong Q, Wan ATY, Liu XY, Fung CSH, Xiao XJ, Malainual N, Hou JP, Wang LY, Wang MQ, Yang KY *et al*.Comparative Genomics Reveals Insights into the Divergent Evolution of Astigmatic Mites and Household Pest Adaptations. Mol Biol Evol. 2022;39(5):msac097.
21. Erban T, Klimov PB, Harant K, Talacko P, Nesvorna M, Hubert J. Label-free proteomic analysis reveals differentially expressed Wolbachia proteins in Tyrophagus putrescentiae: Mite allergens and markers reflecting population-related proteome differences. J Proteomics. 2021;249:104356.
22. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. Evolutionary origin of insect-Wolbachia nutritional mutualism. Proc Natl Acad Sci U S A. 2014;111(28):10257-10262.
23. Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, Leong YS, Dong Y, Axford J, Kriesner P, Lloyd AL, Ritchie SA, O'Neill SL, Hoffmann AA. The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. Nature. 2011;476(7361):450-453.
24. Klimov PB, Oconnor BM. Improved tRNA prediction in the American house dust mite reveals widespread occurrence of extremely short minimal tRNAs in acariform mites. BMC Genomics. 2009;10:598.
25. Sun ET, Li CP, Nie LW, Jiang YX. The complete mitochondrial genome of the brown leg mite, Aleuroglyphus ovatus (Acari: Sarcoptiformes): evaluation of largest non-coding region and unique tRNAs. Exp Appl Acarol. 2014;64(2):141-157.
26. Klimov PB, Chetverikov PE, Dodueva IE, Vishnyakov AE, Bolton SJ, Paponova SS, Lutova LA, Tolstikov AV. Symbiotic bacteria of the gall-inducing mite *Fragariocoptes setiger* (Eriophyoidea) and phylogenomic resolution of the eriophyoid position among Acari. *Sci Rep.* 2022;12(1):3811.
27. Dowle EJ, Powell THQ, Doellman MM, Meyers PJ, Calvert MB, Walden KKO, Robertson HM, Berlocher SH, Feder JL, Hahn DA, Ragland GJ. Genome-wide variation and transcriptional changes in diverse developmental processes underlie the rapid evolution of seasonal adaptation. Proc Natl Acad Sci U S A. 2020;117(38):23960-23969.
28. Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, Banfield JF.Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat Commun. 2017;8(1):1814.
29. Poinar G Jr, Poinar R. Parasites and pathogens of mites. Annu Rev Entomol. 1998;43:449-469.
30. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8:24.
31. Campisano A, Ometto L, Compant S, Pancher M, Antonielli L, Yousaf S, Varotto C, Anfora G, Pertot I, Sessitsch A, Rota-Stabelli O. Interkingdom transfer of the acne-causing

agent, Propionibacterium acnes, from human to grapevine. Mol Biol Evol. 2014;31(5):1059-1065.
32. Glowska E, Dragun-Damian A, Dabert M, Gerth M. New Wolbachia supergroups detected in quill mites (Acari: Syringophilidae). Infect Genet Evol. 2015;30:140-146.
33. Bing XL, Xia WQ, Gui JD, Yan GH, Wang XW, Liu SS. Diversity and evolution of the Wolbachia endosymbionts of Bemisia (Hemiptera: Aleyrodidae) whiteflies. Ecol Evol. 2014;4(13):2714-2737.
34. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW.Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
35. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE. Conducting a microbiome study. Cell. 2014;158(2):250-262.
36. Breitwieser FP, Pertea M, Zimin AV, Salzberg SL. Human contamination in bacterial genomes has created thousands of spurious proteins. Genome Res. 2019;29(6):954-960.
37. Barker PS. The effects of high humidity and different temperatures on the biology of *Tyrophagus putrescentiae* (Schrank) (Acarina: Tyrogliphidae). Can J Zool. 1967; 45:91-96.
38. Amon RB, Hanfin JM, Campbell MC. Contaminated fungal cultures owing to Tyrophagus putrescentiae. Arch Dermatol. 1977;113(11):1614-1615.
39. Pekas A, Palevsky E, Sumner JC, Perotti MA, Nesvorna M, Hubert J. Comparison of bacterial microbiota of the predatory mite *Neoseiulus cucumeris* (Acari: Phytoseiidae) and its factitious prey *Tyrophagus putrescentiae* (Acari: Acaridae). Sci Rep. 2017; 7(1):2.
40. Lee J, Kim JY, Yi MH, Hwang Y, Lee IY, Nam SH, Yong D, Yong TS.Comparative microbiome analysis of *Dermatophagoides farinae*, *Dermatophagoides pteronyssinus*, and *Tyrophagus putrescentiae*. J Allergy Clin Immunol. 2019;143(4):1620-1623.
41. Swiontek Brzezinska M, Jankiewicz U, Burkowska A, Walczak M. Chitinolytic microorganisms and their possible application in environmental protection. Curr Microbiol. 2014;68(1):71-81.
42. Erban T, Rybanska D, Harant K, Hortova B, Hubert J. Feces Derived Allergens of Tyrophagus putrescentiae Reared on Dried Dog Food and Evidence of the Strong Nutritional Interaction between the Mite and Bacillus cereus Producing Protease Bacillolysins and Exo-chitinases. Front Physiol. 2016;7:53.
43. Annamalai N, Veeramuthu Rajeswari M, Vijayalakshmi S, Balasubramanian T. Purification and characterization of chitinase from Alcaligenes faecalis AU02 by utilizing marine wastes and its antioxidant activity. Ann Microbiol. 2011;61(4):801-807.
44. Van Asselt L. Interactions between domestic mites and fungi. Indoor and Built Environment. 1999;8(4):216-220.
45. Brasier CM. Mites and reproduction in *Ceratocystis ulmi* and other fungi. Transactions of the British Mycological Society. 1978;70(Feb):81-89.
46. Peleg AY, Hooper DC. Hospital-acquired infections due to gram-negative bacteria. N Engl J Med. 2010;362(19):1804-1813.
47. Brooke JS. *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin Microbiol Rev.* 2012; 25(1):2-41.
48. Lockhart SR, Abramson MA, Beekmann SE, Gallagher G, Riedel S, Diekema DJ, Quinn JP, Doern GV. Antimicrobial resistance among gram-negative bacilli causing infections in

intensive care unit patients in the united states between 1993 and 2004. J Clin Microbiol .2007, 45(10):3352-3359.
49. Lambert ML, Suetens C, Savey A, Palomar M, Hiesmayr M, Morales I, Agodi A, Frank U, Mertens K, Schumacher M, Wolkewitz M. Clinical outcomes of health-care-associated infections and antimicrobial resistance in patients admitted to European intensive-care units: a cohort study. Lancet Infect Dis. 2011;11(1):30-38.
50. Nathwani D, Raman G, Sulham K, Gavaghan M, Menon V. Clinical and economic consequences of hospital-acquired resistant and multidrug-resistant Pseudomonas aeruginosa infections: a systematic review and meta-analysis. Antimicrob Resist Infect Control. 2014;3(1):32.
51. Lister PD, Wolter DJ, Hanson ND. Antibacterial-resistant Pseudomonas aeruginosa: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. Clin Microbiol Rev. 2009;22(4):582-610.
52. Giske CG, Monnet DL, Cars O, Carmeli Y; ReAct-Action on Antibiotic Resistance. Clinical and economic impact of common multidrug-resistant gram-negative bacilli. Antimicrob Agents Chemother. 2008;52(3):813-821.
53. Griffiths DA. The morpho-species and its relationship to the biological species in the genus *Tyrophagus* (Acaridae, Acarina). In: *Recent Advances in Acarology Proceeding of the V International Congress of Acarology*. Edited by Rodrigues JG, vol. 1. New York: Academic Press. 1979: 199-212.
54. Murillo P, Klimov PB, Hubert J, OConnor BM. Investigating species boundaries using DNA and morphology in the mite *Tyrophagus curvipenis* (Acari: Acaridae), an emerging invasive pest, with a molecular phylogeny of the genus *Tyrophagus*. Exp Appl Acarol. 2018;75(2):167-189.
55. Hedges LM, Brownlie JC, O'Neill SL, Johnson KN. *Wolbachia* and virus protection in insects. *Science*. 2008;322(5902):702-702.
56. Kambris Z, Cook PE, Phuc HK, Sinkins SP. Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. Science. 2009;326(5949):134-136.
57. Geoghegan V, Stainton K, Rainey SM, Ant TH, Dowle AA, Larson T, Hester S, Charles PD, Thomas B, Sinkins SP. Perturbed cholesterol and vesicular trafficking associated with dengue blocking in *Wolbachia*-infected *Aedes aegypti* cells. Nature Communications. 2017, 8.
58. McMeniman CJ, Lane RV, Cass BN, Fong AWC, Sidhu M, Wang YF, O'Neill SL. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. Science. 2009;323(5910):141-144.
59. Diaz-Nieto LM, Gil MF, Lazarte JN, Perotti MA, Beron CM. *Culex quinquefasciatus* carrying *Wolbachia* is less susceptible to entomopathogenic bacteria. Sci Rep. 2021;11(1):1094
60. Pepato AR, Klimov PB. Origin and higher-level diversification of acariform mites - evidence from nuclear ribosomal genes, extensive taxon sampling, and secondary structure alignment. BMC Evol Biol. 2015;15:178.
61. Klimov PB, OConnor BM, Chetverikov PE, Bolton SJ, Pepato AR, Mortazavi AL, Tolstikov AV, Bauchan GR, Ochoa R.Comprehensive phylogeny of acariform mites (Acariformes) provides insights on the origin of the four-legged mites (Eriophyoidea), a long branch. Mol Phylogenet Evol. 2018;119:105-117.

62. Pepato AR, Klimov PB. Origin and higher-level diversification of acariform mites - evidence from nuclear ribosomal genes, extensive taxon sampling, and secondary structure alignment. BMC Evol Biol. 2015;15:178.
63. Van Dam MH, Trautwein M, Spicer GS, Esposito L. Advancing mite phylogenomics: Designing ultraconserved elements for Acari phylogeny. Mol Ecol Resour. 2019;19(2):465-475.
64. Liu Q, Deng Y, Song A, Xiang Y, Chen D, Wei L. Comparative analysis of mite genomes reveals positive selection for diet adaptation. Commun Biol. 2021;4(1):668.
65. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. 2018.
66. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-1676.
67. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27(5):824-834.
68. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-1075.
69. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907-915.
70. Bushmanova E, Antipov D, Lapidus A, Suvorov V, Prjibelski AD. rnaQUAST: a quality assessment tool for de novo transcriptome assemblies. Bioinformatics. 2016;32(14):2210-2212.
71. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210-3212.
72. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18(1):188-196.
73. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.
74. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236-1240.
75. Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, Zhang X, Bolund L, Chen Q, Wang J, Yang H, Fang L, Shi C. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Res. 2018;46(W1):W71-W75.
76. An J, Lai J, Lehman ML, Nelson CC. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. Nucleic Acids Res. 2013;41(2):727-737.
77. Sun ET, Li CP, Li S, Gu SL, Nie LW.Complete mitochondrial genome of *Caloglyphus berlesei* (Acaridae: Astigmata): The first representative of the genus *Caloglyphus*. Journal of Stored Products Research. 2014; 59:282-284.
78. Laslett D, Canbäck B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics. 2008;24(2):172-175.
79. ernt M, Donath A, Jühling F, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69(2):313-319.

80. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol. 2019;1962:227-245.
81. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-780.
82. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-1973.
83. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-274.
84. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-410.
85. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.
86. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. Peerj Computer Science. 2017.
87. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research. 2017;6(1287):1-16.
88. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol. 2015;16(1):50.
89. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH; UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31(6):926-932.
90. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60.
91. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA. . Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2004;2(3):E69.
92. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MC, Tettelin H, Werren JH. Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Appl Environ Microbiol. 2006;72(11):7098-7110.
93. Perotti MA, Allen JM, Reed DL, Braig HR. Host-symbiont interactions of the primary endosymbiont of human head and body lice. FASEB J. 2007;21(4):1058-1066.
94. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol. 1990;56(6):1919-1925.
95. Daims H, Brühl A, Amann R, Schleifer KH, Wagner M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. Syst Appl Microbiol. 1999;22(3):434-444.
96. Perotti MA, Clarke HK, Turner BD, Braig HR. Rickettsia as obligate and mycetomic bacteria. FASEB J. 2006;20(13):2372-2374.

766 97. Balmand S, Lohs C, Aksoy S, Heddi A. Tissue distribution and transmission routes for the
767 tsetse fly endosymbionts. J Invertebr Pathol. 2013;112 Suppl(0):S116-S122.
768

769

Table 1. **Select GenBank sequences attributable to *Tyrophagus putrescentiae* contamination based on NCBI BLAST search**. Bitscore shows quality and significance of alignment between query and subject sequences. Low bitscore matches represent short, full-length GenBank sequences. **TP**=match with our decontaminated T3F6 assembly of *T. putrescentiae* (scaffold ids are given). **RZ**=match with GenBank *"Rhagoletis zephyria"* genome (contaminated with *T. putrescentiae* DNA); this information offers independent validation of our data. *=values averaged; **=the GenBank "*Rhagoletis*" genome does not have the mite mitochondrial DNA; this sequence matches GenBank *T. putrescentiae* mitochondrial genome (from China); ***=the lab that produced these sequences is known to maintain *T. putrescentiae* cultures; ****=hospital surfaces and sink.

GenBank title	GenBank id	Identity (%)	bitscore	TP scaffold id	RZ	Country
Insects						
Rhagoletis zephyria genome	GCF_001687245.1	94.0*	909*	4901 scaffolds total	y	USA
Ostrinia nubilalis	AF398406.1	99.1	623	947	y	USA
Simulium damnosum sp.complex	KY631747.1	99.0	695	3270	y	Nigeria
Predatory mites						
Neoseiulus cucumeris	AY099366	98	678	692	**	USA
Cheyletus malaccensis	KP938898.1	98.7	689	692	**	China
Vertebrates						
Mus musculus	AK041295.1, AK041150.1	97.5-99.4	1118-1308	2104,6946,10139	y	Japan
Ardea herodias	AF447969.1	99.02	366	4160	y	USA
Round worms						
Ancylostoma caninum	DQ841142.1, DQ841148.1	98.5-99.1	580-963	1820,7390	y	USA
Fungi						
Colletotrichum gloeosporioides	JQ862580.1	99.0	1317	905	y	Colombia
Parastagonospora forlicesenica	KY769662.1	100	1905	3270	y	Italy
Fusarium equiseti	MG751111.1,MG751119.1,MG751114.1	99.1-100	327-545	43,57,3166	y(first)	USA
Fusarium graminearum	MG751125.1,MG751131.1	97.2-97.8	388-424	273,6583	y	USA
Plants						
Intsia palembanica	FJ448223.2,FJ448461.2,FJ448035.2	97	407-749	650, 3618, 4637	y(first two)***	Singapore
Bacteria						
Shinella sp.	QFOR01000162	100	2353	6121	y	USA****

Table 2. **Metagenomic profiles of two geographically isolated samples of *Tyrophagus putrescentiae* based on next-generation sequencing datasets.** We analyzed our genomic dataset (from Mexico) and transcriptomic dataset (from Europe) using both unassembled reads (Kraken/Bracken analyses) and metagenomic scaffolds (megablast/blobtools analyses). We only present high-scoring Operational Taxonomic Units (OTUs). The bacterium *Cutibacterium acnes* was found in both datasets; however, it is considered a contaminant. a=not found by BLAST, values given for *Candida parapsilosis*; b=for *Babesia bigemina*; c=for *Babesia ovata*; d= *Bacillus thuringiensis* and *Bacillus cereus*; e=for *Fusarium proliferatum*; f=high abundance in the transcriptome because of the presence of highly expressed mitochondrial genes; g=for *Klossiella equi* (mitochondrion) and *Adelina bambarooniae* (18S rRNA), respectively; * = found in both samples; []=spurious Kraken result, not confirmed by BLAST.

Species	Kraken % of bacterial/fungal/eukaryotic reads	Kraken Magnitude	Blob Reads mapped on assembly	Blob assembly size (bp)	BLAST best hit (bistcore)	Identity for BLAST best hit
		Metagenome (Mexico)				
	Bacteria					
Alcaligenes faecalis	58	1,154,214	2,450,807	4,220,682	370,900	98.881
Pseudomonas aeruginosa	9	186,466	2,129,316	4,743,839	10,405	100.000
Stenotrophomonas sp. PAMC25021	4	76,490	1,191	9,969	5,317	99.286
Achromobacter denitrificans	4	74,386	94,154	1,554,668	2,248	99.043
Advenella kashmirensis	3	67,222	331,132	3,470,439	6,248	98.477
Stenotrophomonas maltophilia	3	63,477	307,833	3,505,590	5,345	99.456
**Cutibacterium acnes*	2	36,049	54,577	1,060,224	2,830	99.677
	Fungi (Dikarya)					
[*Candida dubliniensis*]	96	221,846	42[a]	1,717[a]	713[a]	100[a]
	Apicomplexa					
Low-scoring OTU(s)	0.0002	592	10[c]	101[c]	556[b]	98.418[b]
		Metatranscriptome (Europe)				
	Viruses					
Wolbachia phage	not recovered	not recovered	507	2,573	1908	81.732
	Bacteria					
Escherichia coli	36	130,213	14,451,554	119,701	9,356	99.98

Bacillus spp	18	63,648	83,438	122,804	12,412 [d]	99.911 [d]
Wolbachia	15	54,356	1,592,451	622,805	10,035	80.264
[*Salmonella enterica*]	8	30,667	not recovered	not recovered	2,564	99.434
[*Yersinia pestis*]	6	22,138	not recovered	not recovered	not recovered	not recovered
**Cutibacterium acnes*	0.7	2,675	13,221	1,361	2,071	94.635
	Fungi (Dikarya)					
[*Zymoseptoria tritici*]	40	11,594	not recovered	not recovered	187	78.84
Fusarium	8	2,423	18,016,619 [f]	58,270 [f]	11,753 [e]	99.597 [e]
[*Botrytis cinerea*]	8	2,224	not recovered	not recovered	not recovered	not recovered
[*Colletotrichum higginsianum*]	7	1,980	n/a	n/a	87.9	85.714
[*Sporisorium graminicola*]	7	1,941	not recovered	not recovered	not recovered	not recovered
	Apicomplexa					
Low-scoring OTU(s)	0.05	104,121	9,053,673	434,697	2857, 1890 [g]	82.403, 92.528 [g]

Table 3. **Horizontal gene transfers detected in acariform mites detected in the *Tyrophagus putrescentiae* genome (GCA_012066115)**. HGT analyses were run using two protein subsets extracted from the Uniref50 database (all_no_Metazoa vs Metazoa_no_Arthropoda). For the source, best tblastx matches for the NCBI nr database (no Metazoa) are given; for the target, best tblastx matches for the NCBI wgs database (Endeostigmata+Trombidiformes+Oribatida) are given. Per. Ident= percent identity; Total score =total bitscore. More detail is given in supplementary dataset S1 online.

id	Locus	Query (TP)	Source	Total Score	Per. Ident	Accession	Target	Total Score	Per. Ident	Accession
1	Nitroreductase	GMOD_00004316-RA	Bacteria: Verrucomicrobia	180	45.05	MBS0604115.1	Acariformes/ Parasitiformes	589	50.92	JAEMBT020000006.1
2	NADPH dehydrogenase/NADH:flavin oxidoreductase	GMOD_00003087-RA	Bacteria: Cyanobacteria	373	53.37	WP_250121213.1	Acariformes/Parasitiformes	1945	65.65	JAEMBT020000008.1
3	Glycoside hydrolase family 28/ polygalacturonase	GMOD_00002110-RA	Bacteria: Bacteroidetes	275	43.96	WP_130856424.1	Acariformes/ Parasitiformes	2232	54.81	JAEMBT020000008.1
4	Glucan endo-1,3-beta-glucosidase A1-like/ Glycoside hydrolase family 16	GMOD_00001330-RA	Eukaryota: Amoebozoa	259	51.38	PRP81173.1	Oribatida	2032	57.71	JAEMBT020000001.1
5a	Discoidin domain-containing protein / mycodextranase	GMOD_00003189-RA	Bacteria: Actinobacteria	631	58.61	WP_067367716.1	Oribatida	2877	69.95	CAJPVJ010009338.1
5b	copy 2	GMOD_00003959-RA	Bacteria: Actinobacteria	504	60.71	MBD0736812.1	Oribatida	2445	65.84	JAEMBT020000001.1
7a	Heparinase II/III-like protein	GMOD_00002564-RA	Bacteria: Bacteroidetes	291	42.33	WP_168862112.1	Acariformes	2179	66	JAEMBT020000009.1
7b	copy 2	GMOD_00003781-RA	Bacteria: Bacteroidetes	169	44.20	WP_113615130.1	Acariformes	1405	65.82	JAEMBT020000009.1
8a	NAD-dependent 5,10-methylenetetrahydrafolate dehydrogenase	GMOD_00004475-RA	Eukaryota: Glomeromycetes	277	54.47	CAG8533486.1	Acariformes	942	65.25	CAEY01000550.1
8b	copy 2	GMOD_00003902-RA	Eukaryota: Glomeromycetes	278	54.47	CAG8533486.1	Acariformes	950	65.25	CAEY01000550.1
9	D-Ala-D-Ala dipeptidase	JAAALH010000366.1: 11632-12279	Bacteria: Alphaproteobacteria	134	40.30	WP_231555892.1	Acariformes/ Parasitiformes	323	59.09	LBFO01075982.1

Figure legends

Figure 1. **Decontaminated genomic assembly of *Tyrophagus putrescentiae* and phylogenomic tree of acariform mites.** Basic assembly statistics (**a**). Mitochondrial genome (for each strand, arrows show direction of transcription; inner circle shows GC content) (**b**). Phylogenomic maximum likelihood analysis of acariform mites (**c**). Genomes of acariform mites and outgroups (parasitiform mites, *Limulus*) were downloaded from GenBank, single-copy orthologs were extracted using the BUSCO arachnida_odb10 database, and then aligned and trimmed; 415 protein alignments (occupancy≥0.8 and length≥200) were used for a partitioned analyses in IQ-TREE with automatic partition merging. For each branch, SH-aLRT and ultrafast bootstrap support values are given (in that order) unless both measures are equal 100%.

Figure 2. Bacterial metagenomic profile of *Tyrophagus putrescentiae* from Mexico and ancient horizontal transfer events. Bacterial metagenomic profile of *Tyrophagus putrescentiae* from Mexico; abundance was estimated by Kraken2/Bracken analyses based on Illumina short reads (**a**). Ancient horizontal transfer events from bacteria (**b**, **c**) or fungi to mites (**d**): Nitroreductase (**b**), Heparinase II/III-like protein (**c**), NAD-dependent 5,10-methylenetetrahydrafolate dehydrogenase (**d**); for detail, see Table 3 and supplementary dataset S1 online.

Figure 3. **Bacterial endosymbionts of *Tyrophagus putrescentiae* visualized by different FISH probes**: *Wolbachia*-specific (**a, g**) (red channel), *Rickettsia*-specific (**b)** (yellow channel), *Wolbachia*+ *Rickettsia* (**c**), Eubacterial (**e,f**) (green channel), and control (**d**). *Wolbachia* and *Ricketsia* are localized (**a, b**, arrowheads), while *Wolbachia* is also found in the area of the ovaries

(**a**, arrowhead), female from the UK culture; much of eubacterial signal is associated with the digestive track (**e**) and parenchymal tissue bacteriocytes (**e,f**), two food boluses show autofluorescence (**e),** female from the Czech culture; *Wolbachia* (arrowheads) in egg, from UK culture (egg membranes showing autofluorescence and not signal in the red and green channels).

Figure 4. Maximum likelihood phylogenomic tree of endosymbiotic bacterial genus *Wolbachia*. This inference is based on 276 orthologous loci, 112,298 amino acid alignment positions, 139 *Wolbachia* ingroups, 31 outgroups. *Wolbachia* supergroups and hosts are shown. Previously recognized supergroup O was recovered as part of supergroup B.