A Metatranscriptomic Approach to the Identification of Microbiota Associated with the Ant Formica exsecta

Johansson, Helena

2013-11-18


http://hdl.handle.net/10138/164273
https://doi.org/10.1371/journal.pone.0079777

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
A Metatranscriptomic Approach to the Identification of Microbiota Associated with the Ant *Formica exsecta*

Helena Johansson, Kishor Dhaygude, Stafva Lindström, Heikki Helanterä, Liselotte Sundström, Kalevi Trontti

Centre of Excellence in Biological Interactions, Department of Biosciences, University of Helsinki, Helsinki, Finland

**Abstract**

Social insects live in cooperative colonies, often in high densities and with closely related individuals, and interact using social contact behaviours. Compared to solitary insects, social insects have evolved multi-level immunity that includes immune responses common to holometabolous insects, and social immunity, which is exclusive to social taxa. This suggests that social insects may be subject to high pathogen pressure, yet relatively little is known about the range of symbiotic and pathogenic microbial communities that associate with social insects. In this study we examined transcriptome data generated from the ant *Formica exsecta* for sequences identifying as microbes (or other organisms potentially of non-ant origin). Sequences showing homology to two viruses and several other potentially or obligate intracellular organisms, such as *Wolbachia*, *Arsenophonus*, *Entomoplasmatales* and *Microsporidia*, were present in the transcriptome data. These homologous sequence matches correspond to genera/species that have previously been associated with a variety of insects, including social insects. There were also sequences with identity to several other microbes such as common moulds and soil bacteria. We conclude that this sequence data provides a starting point for a deeper understanding of the biological interactions between a species of ant and the micro- and macrobiotic communities that it potentially encounters.

**Introduction**

Identifying and classifying pathogens for a species of interest has great significance for fundamental research in ecology and evolution. The recent advent of genomic and transcriptomic methods have opened up the field to rapid, sensitive and comprehensive assessment of microbiota occurring in different natural environments [1,2], including the gut of various species [3,4], diseased individuals [5], and apparently healthy tissues [6,7]. Social insects (ants, bees, wasps and termites) live in cooperative colonies, often in high densities and with closely related individuals. They engage in contact behaviors such as trophallaxis (transfer of food or other fluids through mouth-to-mouth or anus-to-mouth feeding) and allogrooming. Crowded living conditions and contact behaviours allow disease to spread rapidly through colonies, and since nest mates are often related they tend to be genetically susceptible to the same pathogen [8,9]. Compared to solitary insects, social insects are hence believed to be under increased pathogen pressure. In response, social insect have an expanded multilevel immune repertoire. These include highly conserved molecular defense pathways such as Toll, imd, Jak/STAT, JNK and RNAi [10–14], and mechanical and humoral cellular responses such as phagocytosis, nodule formation, encapsulation and antimicrobial peptides [8] that are shared with other holometabolous insects. Furthermore, and in contrast to solitary insects, social insects also have social immunity, i.e. individuals mount immune responses for the benefit of others [9,15–17]. Social immunity in ants and other social insects include the use of specialized glands [8], and a variety of behaviours such as allogrooming to remove parasites from one another [18], task specialization [19,20] and nest maintenance [9]. In spite of their importance for the evolution of immunity, natural pathogen communities have been assessed in detail only for honey bees [5] and one species of ant, the red invasive fire ant *Solenopsis invicta* [21] and references therein.

Another aspect of social contact behaviours such as trophallaxis is that they provide opportunities for frequent transmission of gut bacteria, which is relatively rare in non-social insects [22,23]. Consequently, social insects have some of the most distinctive and consistent gut microbial communities known among the insects. These communities have evolved to provide specialized beneficial functions in nutrition and protection for their host [22,23]. Gut microbiota has been well characterized in the honey bee, and broadly characterized in some 60 species of ant [24–27]. These studies show taxonomic as well as dietary influence on the composition of gut bacterial communities in ants [22,26].

*Formica exsecta* is a common ant species with a distribution range spanning most of Northern Eurasia [28]. It builds nests reaching diameters and depths up to 1–1.5 m, which consist of a soil core overlaid by needles, grass, and small sticks [29,30]. Nests that
reach maturity are relatively long lived (upto 20 years), however many fail to reach maturity, hence the average life span is six to seven years [31,32]. It is an omnivore, feeding mainly on small insects and honey dew collected from aphids [33–35]. Eggs developing into sexuels are laid in May and mature to adults in June/July. Eggs destined to develop into workers are laid intermixed with sexual-destined eggs and directly after them. Workers overwinter as adults and live for about one year, whereas male lifespan is restricted to a few weeks [29]. With regards to potential infectious agents or symbionts, previous studies have shown a high population/nest prevalence of the intracellular parasite Wolbachia pipientis in several populations of F. exsecta, although the rate of individual infection is unknown [36–38]. Several species of mites have been associated with this and other Formica ants [19,39,40]. The bacterial gut communities of this ant species have not been investigated.

This study joins several other recent publications where transcriptomic or genomic approaches have been used to identify sequences with homology to pathogenic and natural communities on and within species. Following the generation and assembly of an F. exsecta transcriptome, we utilized only those contigs that were found to not be of ant origin with the aim of identifying sequences with homology to potential intracellular pathogens and symbionts, including viruses, bacteria, and fungi; and biota closely associated with the ants and their environment. There were sequences with homology to two viruses, to several potential intracellular parasites (including Wolbachia, and candidates such as Arsmnopilus), to potential beneficial symbions (such as Eustomoplasma, Burkholderia and yeasts), to six species of mite and to a range of common environmental fungi and bacteria.

**Methods**

**Sampling and Sample Preparation**

Formica exsecta queens, workers, males and cocoons developing to these casts were collected from six localities (three island localities: Rovholmarna, Joskar, Furuskar, and three mainland populations: Harparoskog, Prastkulla, Ingå) spanning an area of approximately 50 km² (Table S1). These islands are relatively pristine and uninhabited, and form part of a nature conservancy area in which agricultural activities, including apiculture, are strictly forbidden. The mainland localities were sampled in recently (<10–30 yrs.) clear-cut economical forest or meadowland close to small villages. No specific permission was required for sampling these ants. This species is not endangered or protected by Finnish or international laws. Finnish land legislation permits everyone the right to access all land without permission of the owner (Everyman’s right).

Established colony queens and overwintered workers (hereafter old queens and workers) were sampled straight from the field without permission of the owner (Everyman’s right). Finnish land legislation permits everyone the right to access land, uninhabited, and form part of a nature conservancy area in which agricultural activities, including apiculture, are strictly forbidden. The mainland localities were sampled in recently (<10–30 yrs.) clear-cut economical forest or meadowland close to small villages. No specific permission was required for sampling these ants. This species is not endangered or protected by Finnish or international laws. Finnish land legislation permits everyone the right to access all land without permission of the owner (Everyman’s right).

Established colony queens and overwintered workers (hereafter old queens and workers) were sampled straight from the field colonies in late April and early May 2011. They were immediately transported to the laboratory and frozen alive at −80°C. Cocoons were maintained in the laboratory until sex and caste could be reliably assessed from external morphology and size, and then divided into three developmental classes: young (white cuticle and eyes), intermediate (white cuticle but pigmented eyes), and old (pigmented cuticle and eyes). A proportion of the cocoons eclosed to new queens, males and workers. Samples were frozen as above at the appropriate developmental stage. For the purposes of this study we used the three casts: queens, workers and males. For the workers and queens we used three age classes: old, new and cocoon. New males and male cocoons were analyzed together due to the low sample numbers. Note that the cocoons were not analyzed by developmental stage. Table 1 lists detailed numbers of colonies, individuals, casts, ages and developmental classes.

**Table 1. Number of individuals pooled in each of the total of 28 libraries sequenced by FIMM or BGI.**

<table>
<thead>
<tr>
<th>Caste/sex</th>
<th>Stage</th>
<th>FIMM</th>
<th>BGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queen</td>
<td>Old overwintered</td>
<td>6 (6)</td>
<td>4 (4)</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>5 (10)</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td>Old cocoon</td>
<td>8 (4)</td>
<td>6 (3)</td>
</tr>
<tr>
<td></td>
<td>Intermediate cocoon</td>
<td>8 (4)</td>
<td>6 (3)</td>
</tr>
<tr>
<td></td>
<td>Young cocoon</td>
<td>8 (4)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>Worker</td>
<td>Old overwintered</td>
<td>18 (6)</td>
<td>30 (8)</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>10 (5)</td>
<td>15 (6)</td>
</tr>
<tr>
<td></td>
<td>Old cocoon</td>
<td>8 (4)</td>
<td>6 (3)</td>
</tr>
<tr>
<td></td>
<td>Intermediate cocoon</td>
<td>8 (4)</td>
<td>6 (3)</td>
</tr>
<tr>
<td></td>
<td>Young cocoon</td>
<td>8 (4)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>Male</td>
<td>Adult</td>
<td>3 (3)</td>
<td>3 (3)</td>
</tr>
<tr>
<td></td>
<td>Old cocoon</td>
<td>3 (3)</td>
<td>3 (3)</td>
</tr>
<tr>
<td></td>
<td>Intermediate cocoon</td>
<td>3 (3)</td>
<td>3 (3)</td>
</tr>
<tr>
<td></td>
<td>Young cocoon</td>
<td>3 (3)</td>
<td>3 (3)</td>
</tr>
</tbody>
</table>

Castes and developmental classes of F. exsecta are given together with number of source colonies in parentheses. doi:10.1371/journal.pone.0079777.t001

The nests were not directly assessed for disease symptoms, however all nests were sampled by experienced researchers who noted no obvious signs of disease with regards to nest appearance, individual ant appearance and general worker behaviour. It was noted that many individuals collected in the field or housed in the laboratory carried mites that likely originated from the field colonies.

Prior to RNA extraction, samples were rinsed in 96% ethanol and any unspecific material, including mites, was removed under a microscope. Total RNA for RNA-seq library construction was extracted from whole queens, workers, males, and cocoons using TrisSure reagent (Bioline), following the manufacturer’s protocol. Each sample was extracted individually, high quality RNA was determined by the presence of an 18S rRNA peak using the Agilent 2100 bioanalyzer, and RNA was pooled after extraction into the respective libraries (Table 1) based on caste and stage so that each RNA sample had equal representation in the pool. The exception to this was the library of old workers sequenced in BGI, which was normalized according to populations (not individuals) due to difficulties in obtaining high quality RNA. The final number of ants that were used was 209 individuals from 39 colonies.

The mRNA was selected from the total RNA (50–80 ug) pools using two rounds of Poly-A-Purist MAG kit (Ambion) and/or custom assays of library providers (see below). The first set of libraries was constructed by the Finnish Institute for Molecular Medicine (FIMM; University of Helsinki, Finland). In short, after the second round of poly-A selection and DNase treatment cDNA was fragmented and sizes 200–500 bases were selected to construct Pair-End sequencing Illumina libraries. The second library set was constructed in the same manner by the Beijing Genomic Institute (BGI, China) except selecting only approximately 200 base fragments. Sequencing of the first set of libraries was conducted by FIMM (PE-99) and the second by BGI (PE-91), in both cases using two lanes of Illumina HighSeq 2000. The raw reads of the meta-transcriptome will be available from 2014-08-29 on GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under bio-

---

**Microbiota Associated with the Ant Formica exsecta**
Basic Bioinformatic Analyses and NCBI BLAST Searches

An initial quality check of the raw fastq data was done by assessing the phred score (or Q-value, ranging from 0–40, where 40 denotes the highest quality) and trimming the low quality bases using Fastx toolkit v. 0.0.13 [41]. Transcriptome assembly from the trimmed reads was performed with the software Trinity (Release 2012-05-18, [42]). An initial transcriptome (IT) assembly was performed, followed by realignment of the original transcripts to the IT. This was a quality control step for all transcripts which ensured that at least five (forward and reverse) paired-end reads were used and removed transcripts which had very low reads per kilobase per million reads (RPKM, see below). Only transcripts with a confirmed minimum of five paired-end reads were used for BLAST searches and mapping against protein sequences from Harpegnathos saltator, Camponotus floridanus [43], Atta cephalotes [44], Linepithema humile [45], Pogonomyrmex barbatus [46], S. invicta [47] and Acromyrmex echinatior [48]. Full details of the final ant transcriptome assembly will be published elsewhere (Trontti et al. unpublished data).

Table 2. Sequences matching microbes, and a mite, associated with F. exsecta, identified from GenBank BLAST searches.

<table>
<thead>
<tr>
<th>Function</th>
<th>Kingdom</th>
<th>Phyla</th>
<th>Genus</th>
<th>Expression value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular</td>
<td>Virus</td>
<td>Discistrovirus</td>
<td>Discistrovirus</td>
<td>4945.261</td>
<td>Similar to IAPV, ABPV and KBV in honey bees (Apis mellifera).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iflaviruses</td>
<td>Iflaviruses</td>
<td>2984.429</td>
<td>Similar to DWV in honey bees.</td>
</tr>
<tr>
<td>Bacteria</td>
<td>Proteobacteria</td>
<td>Wolbachia, 978</td>
<td>321.331</td>
<td>Obligate endoparasite, widespread in arthropods, including F. exsecta.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Wolbachia, 416</td>
<td>3510.295</td>
<td>A second strain of Wolbachia, known from previous studies on Formica ants.</td>
<td></td>
</tr>
<tr>
<td>Fungi*</td>
<td>Microsporidia</td>
<td>Encephalitozoon</td>
<td>17.902</td>
<td>Unicellular parasites, related microsporidians (Nosema spp.) cause infection in bees.</td>
<td></td>
</tr>
<tr>
<td>Fungi</td>
<td>Basidimycota</td>
<td>Cryptococcus*</td>
<td>9.771</td>
<td>Some members of this genus are parasitic – their effect on insects (if any) is unknown</td>
<td></td>
</tr>
<tr>
<td>(Likely) Extracellular</td>
<td>Bacteria</td>
<td>Proteobacteria</td>
<td>Acidiphilium</td>
<td>8147.115</td>
<td>Acidophile, widespread in soils.</td>
</tr>
<tr>
<td>Fungi</td>
<td>Actinobacteria</td>
<td>Rhodococcus</td>
<td>7949.379</td>
<td>Generalist soil genus.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ascomycota</td>
<td>Aspergillus*</td>
<td>41.899</td>
<td>Pervasive in the environment, including indoor air, some opportunistic pathogens</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neurospora</td>
<td>Penicillium</td>
<td>110.729</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gibberella</td>
<td>Neurospora</td>
<td>2888.620</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Podospora</td>
<td>5.751</td>
<td>Plant pathogen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basidiomycota</td>
<td>Eremothecium</td>
<td>26.331</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Animalia</td>
<td>Basidiomycota</td>
<td>Ustilago</td>
<td>5.080</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Fungi</td>
<td>Ascomycota (yeast)</td>
<td>Candida</td>
<td>3.144</td>
<td>Often found as symbionts in insects, including ants. Common and diverse in soils.</td>
</tr>
</tbody>
</table>

Total expression value represents the combined expression value for all ages and castes, from both FIMM and BGI data. Intracellular microbes are known parasites or have at least one stage of their lifecycle inside the host. Extracellular microbes do not usually have lifecycles within hosts, but can in some cases be opportunistic pathogens. Unknown is used for yeast, these are pervasive in environments that contain moisture and carbohydrate sources, and may form symbiotic relationships with ants in their guts, or they can grow in the nest material, or be carried in via aphid dew that ants feed on. IAPV = Israel Acute Paralysis Virus, Acute Bee Paralysis Virus, KBV = Kashmir Bee Virus, DWV = deformed wing virus.

*current classification.
**Some Kluyveromyces are telemorphs of Candida.
*Fungi that gave hits on both common and species specific genes in GenBank.

doi:10.1371/journal.pone.0079777.t002

Microbiota Associated with the Ant Formica exsecta

project ID PRJNA213662, sample accession numbers: SAMN02297446, SAMN02297447, SAMN02297448, SAMN02297449, SAMN02297450, SAMN02297451 and SAMN02297452.
BLAST searched against the virus, fungal, protozoan and bacterial genome databases at NCBI (BLAST version 2.2.26+). As cut off criteria in the BLAST searches we used a minimum alignment length of 100nt, E-value 0.001 a word size of 11, and a minimum of 70% sequence identity. Identified microbiota were then BLAST searched back against the *F. exsecta* transcriptome. By doing this we ensured that more than one gene in the identified microbiota was active so that sequence matches did not just represent single highly conserved genes/domains.

Expression values for the GenBank matches were calculated by aligning and counting all reads of each library, utilizing both FIMM and BGI data. From these, RPKM values (expression value in RNASeq) were calculated as follows: RPK = number of mapped reads/length of transcript in kb (transcript length/1000). This method quantifies gene expression from RNA sequencing data by normalizing for total read length and the number of sequencing reads. We give a total expression value, calculated over all ages and castes in Tables 2–4, and a breakdown of expression values by caste, age class and sequencing provider in Table S2. Note that the number of individuals from each age class and caste varied, from 3 to 30. The samples were pooled by sex and age-class (see above) prior to transcriptome expression in order to obtain sufficient amounts of RNA, hence the libraries were not analyzed by locality. Aware that only a fraction of microbiological diversity is described and/or present on GenBank, we chose to conservatively identify microbiota to genus level.

Ribosomal RNA

To cast the net wider we also isolated ribosomal GenBank matches from the transcriptome, namely the 16S, 18S and 28S ribosomal subunit genes for analysis. Sequences matching 18S rRNA eukaryotes were identified using NCBI BLAST searches. The Ribosomal Database Project (RDP) Classifier tool (http://rdp.cme.msu.edu/, [49]) was used to classify sequences corresponding to the 16S rRNA gene in bacteria and the 28S gene (Large Subunit (LSU)) in fungi. The RDP Classifier tool implements a naïve Bayesian Classifier to taxonomic classification by choosing the most likely taxon that a sequence corresponds to based on the training set. The sequences matching eukaryotes of non-fungal origin identified from the 18S rRNA gene. The Ribosomal Database Project (RDP) Classifier tool (http://rdp.cme.msu.edu/) was used to classify sequences matching 18S rRNA eukaryotes.

Table 3. Sequences matching eukaryotes of non-fungal origin identified from the 18S rRNA gene.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Superorder</th>
<th>Order</th>
<th>“cohort”</th>
<th>genus</th>
<th>Expression value</th>
<th>Genbank match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthropoda</td>
<td>Acariformes</td>
<td>Sarcoptiformes</td>
<td>Astigmata</td>
<td>Acarus</td>
<td>3.427</td>
<td>EU152495.1.1543</td>
</tr>
<tr>
<td></td>
<td>Astigmata</td>
<td></td>
<td></td>
<td>Ewingia</td>
<td>7.687</td>
<td>EU152502.1.1702</td>
</tr>
<tr>
<td></td>
<td>Astigmata</td>
<td></td>
<td></td>
<td>Histiostoma</td>
<td>73.472</td>
<td>GQ864328.1.1656</td>
</tr>
<tr>
<td></td>
<td>Astigmata</td>
<td></td>
<td></td>
<td>Tyroborus</td>
<td>13.140</td>
<td>EF203768.1.1516</td>
</tr>
<tr>
<td>Parasitiformes</td>
<td>Mesostigmata</td>
<td>Dermanysina</td>
<td>Holostapis</td>
<td>23.156</td>
<td>FJ911851.1.1592</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Garoaolaps</td>
<td>3.265</td>
<td>FJ911852.1.1593</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vegaiidae</td>
<td>7.186</td>
<td>GQ864298.1.1627</td>
</tr>
</tbody>
</table>

GenBank match denotes the GenBank accession number of the closest GenBank BLAST match. Total expression value represents the combined expression value for all ages and castes, from both FIMM and BGI data.
seven ant genomes (H. saltator, C. floridanus, A. cephalotes, L. humile, P. barbatus, S. invicta and A. echinatior), annotated using protein sequences. Of those contigs, 5746 returned sequence matches for organisms entered on the individual GenBank databases.

### GenBank BLAST Searches

Sequences with identity to *Wolbachia* were expressed in many contigs and with high expression values (Table 2, Table S2). Combined expression values for the two strains were lowest in the old queens and workers, and highest in cocoons. Ant-derived transcripts showed homology with over 300 environmental prokaryotes, seven of which were yeasts, two were plant pathogens, one a facultative endoparasite (*Cryptococcus*), and one an intracellular microsporidian from the genus *Encephalitozoon* grouped with all other fungi (Table 2). This collapsed list included since many of these fungi may share conserved genes we collapsed this list by removing sequences with homology only to genes shared with all other fungi (Table S2). Annotated virus draft genome assemblies are available under accession numbers KF500001 (Dicistrovirus) and KF500002 (Iflavirus) on GenBank. We identified the contig ‘isotig09016’ in the published genome (Figure 1). The ant-derived Dicistrovirus had 75% nucleotide sequence similarity to the Kashmir Bee Virus, and the ant-derived Iflavirus had 70% nucleotide sequence similarity to the Deformed Wing Virus. The sizes of both virus genomes from the initial assembly (Dicistrovirus: 9554 bp, Iflavirus: 9160 bp) were consistent with genome sizes found in the other viruses in Figure 1 (range: ca 8000–10500 bp). Sequences matching the Dicistrovirus were expressed in all ages and castes, with the highest expression value found in new workers, and the lowest expression in old queens. Sequences matching the Iflavirus were expressed in all ages and castes, with the highest expression value in new queens and the lowest in new workers (Table S2). Annotated virus draft genome assemblies are available under accession numbers KF500001 (Dicistrovirus) and KF500002 (Iflavirus) on GenBank. We identified the contig ‘isotig09016’ in the published *F. exsecta* transcriptome generated by Badouin et al., [58] which showed 99% identity with sequences yielding identity to the Iflavirus in our transcriptome.

Sequences with homology to seventeen species of fungi were present in the transcriptome data, seven of which were yeasts, two were common moulds (*Aspergillus, Penicillium*), six were plant pathogens, one a facultative endoparasite (*Cryptococcus*) and one an intracellular microsporidian from the genus *Entecephalitozoon*. Since many of these fungi may share conserved genes we collapsed this list by removing sequences with homology only to genes shared with all other fungi (Table 2). This collapsed list included sequences with high similarity to only four species: *Aspergillus oryzae*, *P. barbatus*, *S. invicta* and *A. echinatior*, annotated using protein sequences. Of those contigs, 5746 returned sequence matches for organisms entered on the individual GenBank databases.

### GenBank BLAST Searches

Sequences with identity to *Wolbachia* were expressed in many contigs and with high expression values (Table 2, Table S2). Combined expression values for the two strains were lowest in the old queens and workers, and highest in cocoons. Ant-derived transcripts showed homology with over 300 environmental prokaryotes, seven of which were yeasts, two were plant pathogens, one a facultative endoparasite (*Cryptococcus*), and one an intracellular microsporidian from the genus *Entecephalitozoon*. Since many of these fungi may share conserved genes we collapsed this list by removing sequences with homology only to genes shared with all other fungi (Table 2). This collapsed list included sequences with high similarity to only four species: *Aspergillus oryzae*,

---

**Table 4.** 16s rRNA sequence analysis.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Subphylum</th>
<th>Genus</th>
<th>Expression value</th>
<th>GenBank match</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteria</td>
<td>Actinomycetales</td>
<td>Micrococcineae</td>
<td>38.48</td>
<td>JQ559818</td>
<td>This phylum and genus are very common soil bacteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>37.70</td>
<td>AY189308</td>
<td>Soil</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>37.88</td>
<td>JQ559818</td>
<td>Soil</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Sphingobactericeae</td>
<td>Pedobacter</td>
<td>5.63</td>
<td>CP001681</td>
<td>Soil</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>Streptococcaceae</td>
<td>Streptococcus</td>
<td>4.38</td>
<td>A8682354</td>
<td>Soil</td>
</tr>
<tr>
<td>Lactobacillaceae</td>
<td>Lactobacillus</td>
<td>77.63</td>
<td>FN667112</td>
<td>Common in soils. Symbiotic with e.g. honey bees</td>
<td></td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Acetobacteraceae</td>
<td>Saccharibacter</td>
<td>67.51</td>
<td>JN846886</td>
<td>Camponotus fragilis gut bacteria.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99.24</td>
<td>JN846886</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>96.64</td>
<td>JN846887</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anaplasmataceae</td>
<td>Anaplasma</td>
<td>3793.28</td>
<td>CP001391</td>
<td>Wolbachia was closest GenBank match</td>
</tr>
<tr>
<td>Burkholderiaceae</td>
<td>Burkholderia</td>
<td>8.07</td>
<td>JF763883</td>
<td>Common soil bacteria. Often symbiotic with arthropods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moraxellaceae</td>
<td>Enhydrobacter</td>
<td>46.32</td>
<td>EU706098</td>
<td>Soil/symbionts</td>
</tr>
<tr>
<td></td>
<td>Acinetobacter</td>
<td>37.37</td>
<td>JQ435689</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudomonadaceae</td>
<td>Azomonas</td>
<td>78.38</td>
<td>JQ659356</td>
<td>Common in soils, also symbiotic or pathogenic in several taxa</td>
</tr>
<tr>
<td></td>
<td>Azomonas</td>
<td>56.54</td>
<td>EU072020</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudomonas</td>
<td>110.73</td>
<td>HE566388</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenericutes</td>
<td>Entomoplasmataceae</td>
<td>Entomoplasma</td>
<td>50.24</td>
<td>U67946</td>
<td>Obligate intracellular microbes. Often in association with insects</td>
</tr>
<tr>
<td>Mesoplasma</td>
<td>43.20</td>
<td>HM996851</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A. niger, Kluyveromyces and Cryptococcus, which were expressed in all castes and ages.

In spite of careful removal of mites from the ants, there were sequences with identity to *V. destructor* genes present in the transcriptome data (Table 2, Table S2). Notably, sequences matching *V. destructor* were not present among the 18S rRNA sequence data generated from our transcriptome data (see below), nor from Badouin et al.’s [58] transcriptome. *Varroa destructor* is the sole mite species that has had its genome sequenced, so similar genes and sequences from other species of mites will most likely also show homology with *V. destructor* sequences submitted to GenBank. The expression value was highest in new queens, and lowest in old workers, and the expression values were overall similar to expression patterns for total mite 18S rRNA sequences.

**Ribosomal RNA BLAST Searches and RDP Classifier Results**

The 18S rRNA sequences yielded matches to genes of seven main mite species (Table 3). Both super-orders, Acariformes and Parasitiformes, were represented, sorting into suborders Sarcoptiformes; Astigmata (four members) for the former and Mesostigmata; Dermanyssina and Mesostigmata: Veigaiidae for the latter (three members). Sequences yielding identity to *HISTOSTOMA* showed the highest expression level overall (Table 3) and were expressed in all castes and age classes with the exception of old queens (Table S2), with the highest expression value in new queens. Sequences yielding identity to *VEGANUS* had relatively low expression values, however they were expressed in all ages and castes (Table 3, Table S2). Total expression values for sequences matching mite 18S rRNA was highest in new workers and queens, and lowest in old queens and worker cocoons. No sequences with similarity to 18S rRNA mite genes were detected in the *F. exsecta* transcriptome data (Table 2, Table S2).

Figure 1. Phylogeny of the available genomes of the insect-specific *Discistroviridae* and *Iridoviridae* from GenBank. Social insect viruses are highlighted in blue and ant-derived assembled virus genomes from this study are highlighted in red.

doi:10.1371/journal.pone.0079777.g001

**Microbiota Associated with the Ant *Formica exsecta***

Among the 21 16S rRNA sequences retained from the RDP classifier there were sequences with homology to bacteria whose life cycles are either completely within a host cell or require some stage of their life cycle within a host cell (hereafter intracellular). These were: Proteobacteria (Wolbachia), Enterobacteriaceae (Arsenophonus), and Tenericutes (Entomoplasmata). Sequences with identity to Actobacteraceae matched most closely to gut bacteria isolated from the ant *Camponotus fragilis*, and may also be intracellular. Total expression values when combined by genus showed that sequences with identity to Wolbachia had the highest value, followed by sequences yielding identity to Actobacteraceae and sequences with homology to Entomoplasmata. Sequences matching Arsenophonus did not have high expression values and were less encountered than the above genera (Table 3). Sequences matching Wolbachia and Actobacteraceae were expressed in all castes and ages (Wolbachia expression patterns as above), whereas sequences matching Arsenophonus were not expressed in new queens and new workers (Table S2). The remaining sequences matched bacteria from subphyla commonly found in soils (subphyla Actinomycetales, Sphingobacteriaceae, Streptococcales, Lactobacillales, Burkholderiaceae, Moraxellaceae and Pseudomonadales). Several members of these subphyla are capable of a beneficial or pathogenic symbiotic lifestyle with insect hosts, in particular Lactobacillales, Burkholderiaceae and Pseudomonadales. Sequences matching these latter three genera were expressed in all castes and ages (Table S2).
Fifty-four LSU sequences were retrieved from the transcriptome data. Of these, eight failed to be classified as fungi using the RDP classifier, another 35 were classified as fungi, but no further taxonomic classification was possible, leaving only eight sequences that could be classified with some confidence (>80%) to species level. Of these, four sequences were classified as Basidiomycota, species *Hohenbuehelia* (wood decaying fungi), and the remaining sequences were classified as Ascomycota, two classifying as *Plectosphaerella* (plant pathogen) and two as *Parmotrema* (lichen) (Table S3).

**Discussion**

Sequences with identity to a wide range of microbes, and some species of mite, were retrieved from within and on the ant *Formica exsecta*, by analysing data of non-ant origin from transcriptome assemblies. These included sequences yielding identity to two viruses, two *Wolbachia* strains, several potential intracellular symbionts (*Arsenophonus*, *Lactobacillus* and *Microsporidium*) and potentially opportunistic pathogens (*Aspergillus* spp. and *Pseudomonas* spp.). Since transcriptome data is generated from RNA, the sequence matches most likely represent some of the most active microbes associated with the ant *F. exsecta*. 

---

**Figure 2. Phylogenetic tree of bacterial 16S rRNA sequences, with closest type strains obtained from the Ribosomal Database project in black font.** Ant-derived contig IDs are given in red font. Red clade = *Firmicutes*, blue clade = *Bacteroidetes*, black clade = *Proteobacteria*, yellow clade = *Tenericutes*, green clade = *Actinobacteria*. doi:10.1371/journal.pone.0079777.g002
The successful and expected retrieval of sequences matching Wolbachia from many contigs and among the 16S rRNA sequence data attests to the feasibility of this approach. Sequences matching Wolbachia showed high levels of identity with two strains that commonly infect Eurasian and Finnish populations of several closely related Formica wood ants, including Formica exsecta [36–39,61]. Wolbachia is a very common and widespread intracellular symbionts of insects [62], including ants [36]. It usually shows no noticeable adverse effect but occasionally affects its host negatively [63]. It is readily horizontally and vertically transmitted within and between Formica species, and has not been associated with any detrimental effects in F. exsecta [38,61,64]. Expression of sequences matching Wolbachia appeared to decrease with age, in line with previous results on F. exsecta by Keller et al. [61]. The mechanism and significance of this pattern remains unknown.

In addition to sequences yielding identity to Wolbachia, there were sequences with homology to other intracellular microbes. These were to the fungi Microsporidia, and to bacteria belonging to the Enterobacteriaceae (Arsenophonus), the Entomoplasmatales, and the Acrorhabdaceae (Saccharibacter). Microsporidia are unicellular, fungal intracellular parasites. Present as spores in the soil, they are ingested and replicate in a wide range of hosts, including insects [65]. Well-studied microsporidia in social insects include Kneallhazia solenopsae [66–68] and Varroa mites [69] that both infect the fire ant Solenopsis invicta. Kneallhazia solenopsae cause substantial brood reductions, queen debilitation and premature death of infected queens [70,71], and V. mites cause significant reduction in nest growth and higher sensitivity to starvation [72,73]. Both of these microsporidia are present in larvae and adults. Bacteria belonging to the Enterobacteriaceae (Arsenophonus), and the Entomoplasmatales (particularly Spiroplasmas) are known reproductive parasites among various insect species, including social Hymenoptera [27,62]. Screening across taxa and species has revealed that these bacteria are usually found in lower prevalence than Wolbachia [62], which expression values suggested was also the case for our data. Arsenophonus is best known from its association with the parasitic wasp Nasonia vitripennis, where it causes elevated death rates among male pupae [74], however recent research suggest that it is one of the richest and most widespread genera of insect associates, with effects ranging from male-killing to symbiosis [75]. Entomoplasmatales are also primarily insect associated bacteria, found in a diversity of roles, again ranging from male-killing to beneficial symbiosis [76]. Entomoplasmatales have been found in ant guts, forming a host-specific clade among the army ants [76], and Spiroplasmas were recently suggested to be prime candidates as a beneficial symbionts in two Solenopsis species [27]. Several 16S rRNA sequences had their nearest GenBank match to bacteria found in the gut of another ant, Camponotus frigidus (GenBank IDs: JN846886 and JN846887). These were classified as Acrorhabdaceae by the RDP classifier. Acrorhabdaceae have been found in honey bee guts, and one such bacteria, Saccharibacter, has also been isolated from pollen, and may hence represent bacteria taken up through foraging [77]. The expression patterns in our data for sequences matching potential intracellular microbes varied and included: present in all ages and castes (e.g. Acrorhabdaceae), absent only in the old age classes (Entomoplasmatales), present in only in new queens and workers (Microsporidia), and no discernable pattern (Arsenophonus). Both beneficial and pathogenic intracellular microbes may predominantly occur in certain age classes [22], hence it is impossible to derive any definitive indication of role from this data.

Two apparent entire RNA virus genomes were assembled from the Formica exsecta transcriptome. Both were positive-sense single-stranded RNA viruses, one belonging to the class Iridoviridae and the other to the Dicistroviridae. These viruses grouped phylogenetically with the KBV, ABV and IAPV viruses that plague apiculture [78,79], and with the Solenopsis invicta virus-1 that infects S. invicta [80]. These viruses are rapidly evolving due to their lack of repair mechanisms and often the same virus can infect several different insect hosts [56,57]. The virus genomes assembled from the F. exsecta transcriptome appear rather dissimilar to honey bee viruses, with no more than 75% nucleotide sequence identity to the phylogenetically closest honey bee viruses. Hence, these viruses possibly warrant classification as new viruses, however such verification and analysis (e.g. electron microscopy; negative strand qPCR and epidemiology) is beyond the scope of this study. None of the ants sampled here showed any obvious symptoms similar to those reported for bees with high levels of infection of Dicistroviruses (trembling paralysis) or Iridoviruses (discolouring), and the pathology and extent of infection in natural ant populations are unknown. Several viruses infecting ants have been described: Solenopsis invicta virus-1, 2 and 3 characterised for Solenopsis invicta [80–82], and a Rhabdovirus, a picorna-like virus and a virus of uncertain classification recently identified from transcriptome data in the Caribbean Crazy ant, Nylanderia fabens [83]. Viruses from several different classes hence appear common among ants.

One route for RNA viruses to enter insects is through ingestion, e.g two ant species, Formica rufa and Camponotus vagus were found to have ingested honey bee viruses by feeding on dead Varroa in honey bee hives [84], and Valles et al [85], reported sequences matching a cricket virus gene in ant-derived expressed sequence tag data, thought to originate from laboratory antfed. In our transcriptome there were sequence matches to V. destructor, a mite that attacks the honey bee (Apis mellifera) [5,86]. However, there were no sequences matching Varroa 18S rRNA, a gene which is used for mite species identification [87]. Instead there were 18S rRNA sequence matches to six species of mite from genera with known ant associations [43,93,94]. Moreover, expression patterns of sequences matching total mite 18S rRNA and expression values for those sequences matching F. destructor were very similar. Hence, sequences matching Varroa were likely an artefact as genes from other species of mites are likely to match V. destructor, currently the only mite genome available in GenBank. No sequences matching Varroa 18S rRNA were present in the second transcriptome generated from F. exsecta ants living in habitat where apiculture is forbidden, yet sequences matching the ILLavirus was present. Sequences matching the Dicistrovirus were not present in this second transcriptome, however this is not surprising since the lowest expression values for sequences matching this virus in our data was in old workers (the age and caste sampled for the Badouin et al. transcriptome).

Nevertheless, mites (Acari) are ubiquitous and diverse in Formica wood ant nests [39]. They associate with ants mainly for dispersal (phoresy), or for feeding on bacteria and detritus in the nest or on the ants [40]. Since it is the task of adult workers to forage we would expect the the highest expression values for sequences matching both mites and viruses in this caste and age class if either were indeed ingested. If the viruses were transmitted by mites, we would expect some similarity in expression patterns of the sequences with homology to mite genes and those with homology to viruses. To the contrary, old workers had relatively low expression values for 18S rRNA sequences matching any mites, high expression levels for sequences matching the ILLavirus and low expression values for sequences matching the Dicistrovirus. The highest expression value for 18S rRNA sequences matching mites was found in new queens, which had low expression values for sequences matching the ILLavirus, and relatively high expression...
values for sequences matching the Dicistrovirus. In short, there appears to be no association in this data between sequences matching mites and expression of sequences matching either of the viruses.

Some of the microbes for which there were sequence matches quite possible have roles both as beneficial ant symbionts and in the soil or nest environments. Yeasts, for example, are common in the environment [88] and are unlikely insect pathogens [89]. Many insects are dependent on yeasts as sources of sterols and vitamins. A study in S. invicta found several yeasts present in the nest and guts of both larvae and workers. These provided nutrients during overwintering, thereby increasing nest survival and productivity [90]. Among bacteria, Burkholderiales and Pseudomonadales are commonly found in ant guts and are thought to be beneficial symbionts [26,27], however they are also widespread in soils and may simply have been ingested by F. exsecta. Lactobacillus are commonly found in acidic environments such as composts, yet some species are also well characterised beneficial gut symbionts in e.g. honey bees and bumble bees [77] and have also been recovered from ant guts, although any functional role in ants is unknown [26]. Sequences matching Burkholderiales, Pseudomonadales, Lactobacillus and yeasts were expressed in all age classes, hence intra- or extracellular origin cannot be determined in this study.

Common soil bacteria or fungi may also act as opportunistic pathogens. There are several known generalist pathogens in the genus Aspergillus, but none of the species are exclusively entomopathogenic. Aspergillus ochraceus is an example of a common environmental Aspergillus with entomopathogenic properties. This fungi infects the ant Alta hispifera resulting in about 50% mortality [91]. Another potentially pathogenic fungi for which there were sequence matches in our data is Cryptococcus, which is widely distributed in soils and best known for causing infection in immunocompromised humans [92]. However, strains of Cryptococcus have been isolated from dead insects or insect chaff [92], although their role in insects is unknown.

The remaining sequences matched mainly fungal plant pathogens and common soil bacteria. These may be derived from ant feed or nest material. There were very few sequences matching saprotrophic wood or plant decaying fungi considering that F. exsecta mounds are made of plant material (including needles and other material from coniferous trees). The analysis of fungal LSU genes showed the difficulty of molecular identification of fungal species [93] compared to the bacterial classification scheme based on 16S rRNA. Unexpectedly, the entomopathogenic fungi Metarhizium anisopliae and Beauveria bassiana, common in soil samples in Finland [94], and known to infect ants [95], did not match sequences in the transcriptome data. Ants almost certainly carry fungal spores into the nest, however if these fail to germinate in the dry and warm environment of the nests, or on the ants themselves, they will not be detectable in transcriptomic data.

Two factors are likely to have affected our results. Firstly, a poly-A treatment of the RNA was performed to reduce the non-eukaryotic content in the samples, and secondly, we applied strict criteria for the bioinformatic analysis. As a result, we fully expect that there are even more microbes and other taxa associated with these ants. Certain trends were apparent and we are confident that the biota presented here is truly associated with, and of importance, in F. exsecta. Similar to S. invicta (in its native range) [21], F. exsecta is infected by Wolbachia and may be infected by several viruses and possibly also microsporidia. Transcriptome data derived from F. exsecta showed sequence matches to Burkholderiales, Lactobacillus, Acetobacteraeaceae and Pseudomonadales which are bacterial genera that often form part of the highly distinct bacterial gut communities in social insects [26,77]. There were also sequence matches to common reproductive parasites, of social as well as solitary insects, such as Enteromoptilinales and Aedosporales. Our study provides a starting point for further research to establish which of these potential associates are neutral elements, random benefactors, pathogens, or have decisive positive effects on the fitness of F. exsecta ants and their nests.

Supporting Information

Table S1 Geographic localities, numbers of F. exsecta nests sampled and number of individuals from those nests.

Table S2 Expression values by sex and caste for all sequence matches in Tables 2–4, from both sequencing providers (Beijing Institute of Genomics, China, (BGI) and Finnish Institute of Molecular Medicine, Finland (FIMM)).

Table S3 RDP classifier classification of the fungal large subunit (LSU) gene sequences.

Acknowledgments

We thank Jenni Paviala and Claire Morandin for assistance with sampling and RNA extractions, and Owen Rowe for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: HJ KD. Analyzed the data: HJ KD. Contributed reagents/materials/analysis tools: KT HH LS. Wrote the paper: HJ KD SL HH LS KT.

References


Coptoformica

Seifert B (2000) A taxonomic revision of the ant subgenus Formica (Coptoformica) exsecta

Pisarski B (1982) Characteristique de Formica

Pamilo P (1991) Life span of queens in the ant Formica exsecta


