Complete genome of a unicellular parasite (*Antonospora locustae*) and transcriptional interactions with its host locust

Longxin Chen1,2,†, Xingke Gao1†, Runting Li2,4, Limeng Zhang2,4, Rui Huang5,6, Lingling Wang7, Yue Song7, Zhenzhen Xing7, Ting Liu2, Xiaoning Nie2, Fangyuan Nie5,6, Shuang Hua7, Zihan Zhang2, Feng Wang3,*, Runlin Z. Ma2,5,6,* and Long Zhang1,*

Abstract

Microsporidia are a large group of unicellular parasites that infect insects and mammals. The simpler life cycle of microsporidia in insects provides a model system for understanding their evolution and molecular interactions with their hosts. However, no complete genome is available for insect-parasitic microsporidian species. The complete genome of *Antonospora locustae*, a microsporidian parasite that obligately infects insects, is reported here. The genome size of *A. locustae* is 3170203 nucleotides, composed of 17 chromosomes onto which a total of 1857 annotated genes have been mapped and detailed. A unique feature of the *A. locustae* genome is the presence of an ultra-low GC region of approximately 25 kb on 16 of the 17 chromosomes, in which the average GC content is only 20%. Transcription profiling indicated that the ultra-low GC region of the parasite could be associated with differential regulation of host defences in the fat body to promote the parasite’s survival and propagation. Phylogenetic gene analysis showed that *A. locustae*, and the microsporidian family in general, is likely at an evolutionarily transitional position between prokaryotes and eukaryotes, and that it evolved independently. Transcriptomic analysis showed that *A. locustae* can systematically inhibit the locust phenoloxidase PPO, TCA and glyoxylate cycles, and PPAR pathways to escape melanization, and can activate host energy transfer pathways to support its reproduction in the fat body, which is an insect energy-producing organ. Our study provides a platform and model for studies of the molecular mechanisms of microsporidium–host interactions in an energy-producing organ and for understanding the evolution of microsporidia.

DATA SUMMARY

The whole-genome sequencing datasets from this study have been submitted to the BioProject database of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) under accession number PRJNA353563. All transcriptome data were uploaded to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) database under the GenBank accession numbers SRR5171247, SRR5171248, SRR5171251-SRR5171254, SRR5171257 and SRR5171258.

INTRODUCTION

Microsporidia are a large group of obligate intracellular parasites of insects and mammals [1–8]. The taxonomic status of this
group of unicellular parasites remains controversial, although recent studies have suggested that microsporidia belong to the fungal kingdom. Regardless of their classification, most studies place these parasites in a unique node position between prokaryotes and unicellular eukaryotes, using them as model organisms for evolutionary studies of interactions with eukaryotic hosts [9–11].

Over 1400 species of microsporidia in 200 genera have been identified since Nägeli first isolated Nosema bombycis from an infected silkworm in 1871 [12, 13]. Both invertebrates and vertebrates fall within the host ranges of microsporidia, including honeybee [14], silkworm [15], fish [7], shrimp [16], swine [17], horse [18], cattle [19] and goat [20]. Approximately 10 species of microsporidia cause human diseases [21]. Symptoms of human infection with microsporidia include keratitis, myositis, encephalitis, cholecystitis, hepatitis, osteomyelitis, pulmonary infection and death [22–26].

Microsporidia that infect insects have relatively simple life cycles, providing ideal model systems for studying evolution and molecular interactions with their hosts. Following the initial genome sequencing work with the parasite Encephalitozoon cuniculi, which infects mammals [2], partial genomic sequences have been produced for N. bombycis [3] and Nosema ceranae [27], whose target tissues of infection are primarily the midgut, silk gland and malphigian tube of host insects. In contrast to N. bombycis and N. ceranae, the major organ targeted for infection by Antonospora locustae is the fat body of locusts. A. locustae, which was formerly known as Nosema locustae and is synonymous with Paranosema locustae, has a fairly narrow host range, only infecting locusts [28]. The fat body in insects is functionally equivalent to the liver in vertebrates, and may provide another special model for interactions between microsporidia and their hosts. These previous works facilitated in-depth investigation of the mechanisms of parasite infection and host–parasite interactions, as well as their control.

Numerous studies have investigated the use of A. locustae as a powerful biological control agent against locusts in agriculture since the 1980s, and it is the only species that has demonstrated strong potential as a bio-insecticide for controlling outbreaks of locusts worldwide [29–31]. A. locustae showed great potential during the locust disaster that broke out at the end of last year, in particular. However, few studies have been carried out on the molecular infection mechanism of A. locustae [32]. Such research would be helpful for improving the application of microsporidia as bio-insecticides. Preliminary sequences of the A. locustae genome and transcriptome have been reported, along with limited information on A. locustae genomics [8].

In this study, we report a full and complete genome sequence of A. locustae based on second- and third-generation genome-sequencing technologies. The genome of the parasite consists of 17 chromosomes on which a total of 1857 coding genes have been annotated and mapped. An ultra-low GC region of approximately 25kb was found in 16 of the 17 chromosomes. Our phylogenetic study based on genomes suggested that microsporidia are a special evolutionary group. A transcriptional study of pathogen-infected and healthy host tissues highlighted the molecular interactions between A. locustae disease and the locust host. Our study provides a platform and model for studies of the molecular mechanisms of microsporidia–host interactions in an energy-producing organ, as well as for understanding the evolution of microsporidia.

**Impact Statement**

Antonospora locustae is the only species that has demonstrated strong potential as a bio-insecticide for controlling outbreaks of locusts worldwide. Following the development of high-throughput sequencing, the complete genome of A. locustae was assembled to the chromosome level based on the method used for combining second- and third-generation high-throughput genome sequencing technologies. Complete genomes of other microsporidia may be obtained in the same way, accelerating research into microsporidia. We have reported the most complete genome sequences of A. locustae to date, along with detailed gene annotation and a relatively clear explanation of the interaction between A. locustae and its locust host. In-depth analysis of the interactions between A. locustae and its host at different stages showed that A. locustae recognizes its host and transitions in the midgut, regulates host energy transfer processes after entering the fat body, inhibits host melanization of A. locustae to evade the host’s immune system, and then completes its generational change in the host’s fat body cells.

**METHODS**

**A. locustae sample preparation and DNA isolation**

The stocks of A. locustae spores used in the experiments were routinely maintained in the Key Laboratory of Biological Control, Ministry of Agriculture, China Agricultural University. To isolate genomic DNA from the parasite, the spores were propagated *in vivo* by infecting its natural host, Locusta migratoria, which were routinely maintained in the same laboratory, following procedures described previously [32].

Briefly, colonies of the host locust were maintained at 28–30°C and 60% relative humidity. For infection, fresh corn leaves coated with *A. locustae* spores were fed to third instar larvae of locusts for 12h. Establishment of *A. locustae* in the fat body of locusts was determined microscopically 15–18 days after infection. The fat body and other control tissues were collected from the infected hosts, and cells were lysed for initial removal of host genomic DNA. After proper filtration, a modified CTAB extraction method was followed to isolate spores of the parasites. The Omniprep DNA extraction kit (G-Biosciences) was then used to extract genomic DNA of the parasite [33]. The isolated DNA was examined for the presence or absence of host DNA contamination, and only host DNA-negative samples were retained for subsequent experiments.
**Genome de novo sequencing through Illumina HiSeq**

DNA sequencing libraries were constructed according to the manufacturer’s instructions (NEBNext Ultra DNA Library Prep Kit for Illumina). For each sample, 2 μg of genomic DNA was randomly fragmented to <500 bp through sonication (Covaris S220). The fragments were treated with End Prep Enzyme Mix for end repair, 5’ phosphorylation, and dA-tailing in one reaction, followed by ligation to adaptors with a ‘T’ base overhang. Size selection of adaptor-ligated DNA was then performed using the AxyPrep Mag PCR Clean-up kit (Axygen), and fragments of ~410 bp (with an approximate insert size of 350 bp) were recovered. Each sample was then amplified via PCR for eight cycles using P5 and P7 primers, with both primers containing sequences that anneal with the flow cell for bridge PCR and the P7 primer containing a six-base index to allow for multiplexing. The PCR products were cleaned up using the AxyPrep Mag PCR Clean-up kit (Axygen), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified with a Qubit2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA).

Libraries with different indices were multiplexed and loaded in an Illumina HiSeq instrument according to the manufacturer’s instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2 × 150 paired-end (PE) configuration; image analysis and base calling were conducted with HiSeq Control Software+OLB+GAPipeline-1.6 (Illumina) on the HiSeq instrument. The average final read depth in the assembly was 2000×. The sequencing results were processed and analysed using GENEWIZ.

**Third-generation de novo sequencing of the A. locustae genome using the PacBio RSII SMRT platform**

To assemble the *A. locustae* genome, 15 μg of high-quality genomic DNA was randomly interrupted using an ultrasound method (Covaris S220) to obtain double-stranded fragments of approximately 5–10 kb, and DNA fragments of more than 2 kb were recovered. The end of the DNA fragment was ligated to the linker of the hairpin structure. Library construction was carried out using the commercial SMRTbell library method. Sequencing of the library was performed using the PacBio RSII SMRT system [34]. Assembly of the PacBio reads was conducted using PBcR WGS-Assembler 8.2 software [35–40]. The average final read depth of the assembly was 200×.

**Assembly and annotation of genomic data**

Based on clean data optimized with the Illumina HiSeq platform, Velvet (version 1.2.10) software was used for k-mer analysis, a de Bruijn diagram was constructed based on the overlapping relationship between k-mers, and the segmented contig sequence was spliced. SSPACE (version 3.0) was used to compare the reads obtained from sequencing of all libraries to the contig sequences obtained using the pairing relationship between paired-end reads and insert size distances, as well as to support primer shifts and design of primers for PCR experiments and further assembly of contig sequences to the scaffold sequence. Finally, using GapFiller (version 1–10) software, all reads from all libraries were aligned to the scaffold sequence. Gaps in the scaffold sequence were complemented by the aligned reads, and thus the scaffold sequence could be extended. Finally, we obtained a scaffold sequence with a lower ratio of unknown bases, N, and increased sequence length.

The PacBio off-machine data obtained using the PacBio RSII SMRT platform were assembled using the assembly software wgs-assembler (version 8.3) to obtain preliminary assembly results. Based on the preliminary assembly results, Illumina HiSeq sequencing data were simultaneously imported with the PacBio assembly data, and this assembly was corrected using the calibration software Quiver (version 1.1.0) to obtain the final assembly result.

The methods used for gene prediction in the genomes as follows. Glimmer gene prediction software was primarily used for the prediction of single-exon genes [41]. De novo gene prediction was carried out using Augustus software [42], combined with existing transcriptome data for mapping of the genome, analysed with StringTie software, and finally, more accurate gene predictions were obtained. Annotation of coding genes was performed by comparison with the Nr database of the National Center for Biotechnology Information (NCBI). Functional annotation of these genes was performed using the GO database [43], pathway annotation was carried out using the KEGG database [44] and systematic classification of proteins encoded by the genes was performed using the Clusters of Orthologous Groups (COG) database [45]. Scanning for transfer RNAs in the genome was mainly performed using rRNAscan-SE software with default parameters [46], and ribosomal RNAs were identified using RNAmmer software [47].

**RNA isolation and sequencing**

Locusts were infected with *A. locustae* as described above. Transcriptome sequencing was divided into four groups, with each group consisting of two independent repeated samples, and a total of eight independent samples. The midgut of healthy locusts (NM), the infected midgut (M), the fat body of healthy locusts (NF) and the fat body of infected locusts (F) were analysed, with numbers representing different individuals. On the 15th day after infection, the fat bodies and midguts of living locusts were collected. All samples were immediately frozen in liquid nitrogen, and the remaining tissues were ground and diluted with deionized water to determine whether the infected group was actually infected with *A. locustae* and the intensity of infection. Only locusts with the same infection intensity were used for RNA extraction and sequencing. In addition, the healthy group of locusts was tested as a control. Locusts in this group were only used for RNA extraction and transcriptome sequencing after passing the infection test.

One microgram of total RNA was used for library construction. For transcriptome sequencing, we used the NEBNext
UltraTM RNA Library Prep kit from Illumina according to the manufacturer's instructions. After mixing the various index-labelled libraries, 2×150bp double-end sequencing (PE) was performed according to the Illumina HiSeq 2500/3000 (Illumina) instrument instruction manual, using the HiSeq Control Software provided by HiSeq and the OL+B+GAPipeline−1.6 (Illumina) program to obtain sequence read data. Four uninfected samples provided approximately 6.0 Gb of data per sequencing library; four infected samples provided approximately 8.0 Gb of data per sample library (including locust and microsporidium).

Data analysis
Clean data for subsequent analysis were obtained by removing the linker and low-quality sequences from the raw data (Pass Filter Data) using the second-generation sequencing data quality statistical software Trimomatic (v0.30) [48, 49]. The filtered clean data were analysed against the locust genome and the A. locustae genome sequenced in this study [50]. The reads obtained from A. locustae RNA sequencing were applied to density statistical analysis of each chromosome using Hisat2 (v2.0.1) software with the default parameters for short-read comparison [51, 52]. BUSCO was applied to evaluate the assembly completeness by identifying a set of highly conserved microsporidia orthologues in the assembly [53]. For the analysis, gene annotation was performed with Augustus, and the analysis for homology and positive matches was performed with HMMPER 3 [54]. The expression levels of each A. locustae chromosome were calculated. Gene expression levels were analysed using the fragment per kilo bases per million reads (FPKM) method with rsem software (V1.2.6).

Differential gene expression analysis was performed using Deseq2 in Bioconductor software (V1.6.3). Screening for differentially expressed genes was based on expression level changes that were greater than twofold with a false discovery rate ≤0.05. Statistical analyses were performed on upregulated and downregulated genes to identify significant differences. All statistical t-tests (and nonparametric tests) followed by two-tailed comparison tests were performed using GraphPad Prism version 6.00 for Windows (GraphPad Software, Inc., La Jolla, CA, USA).

For the genomic data of A. locustae, Saccharomyces cerevisiae, Kazachstania naganishii, Babesia bigemina, Babesia motasi, Encephalitozoon cuniculi and Encephalitozoon hellem, paralogous and syntenic collinear blocks were characterized using the MScanX strategy [55]. Briefly, proteomic sequence data were obtained using the BLASTP algorithm to generate BLAST outputs, which were imported into MScanX software to generate collinearity outputs. Then, a circle plotter program was employed to graph the paralogous and syntenic collinear blocks with the collinearity outputs.

For phylogenetic analysis, the microsporidian protein sequences of frataxin were retrieved from GenBank databases. Orthologous sequences were identified using BLASTP searches at an E-value cutoff of 1E−20, using A. locustae frataxin proteins as queries. Each group of proteins was aligned using the MAFFT program (version 7) with the E-INS-I algorithm [56], and ambiguous regions in the aligned sequences were removed with TriMali [57]. Maximum-likelihood phylogenetic trees were estimated in phyML 3.0 software [58] using the best model of amino acid substitutions estimated with MEGA 6 and the Rerooting (SPR) branch-swapping algorithm.

RESULTS
DNA sequencing of the A. locustae genome
Genomic DNA from A. locustae were prepared successfully without contamination from host DNA (Fig. S1, available in the online version of this article). Subsequent DNA sequencing using the Illumina HiSeq II platform generated a total of 87379454 reads with a bidirectional read length of 150 bp. GC content of 42.53% and uniform distribution (Fig. S2). Statistical analysis of the original data (Figs S3 and S4) showed that all reads were of good quality, as indicated by the error rate Q20 (<1%)=95.58%, Q30 (<0.1%)=91.42% and N=6.99/Mb. The data were cleaned and optimized using Trimomatic software, yielding a total of 58316056 reads of 144.70 bp on average, with GC content=41.77%, Q20=99.70%, Q30=98.63% and N=0.77 per million bases. In parallel, PacBio RSII SMRT third-generation high-throughput genome sequencing (GENEWIZ) was used to generate a total of 160777 sequence reads with an average length of 3918.61 bp per read, N50=5237 bp and GC content=41.57%.

The complete genomic sequence of A. locustae, determined to comprise 3170203 nucleotides and encode 1857 predicted genes (Table S1), was successfully assembled de novo using the clean sequence data generated from the PacBio RSII SMRT platform supplemented with sequence data from the Illumina HiSeq II platform. A total of 17 scaffolds, ranging from 88.763 to 388.82 kb, were identified. Each of these scaffolds was assigned to a chromosome of A. locustae, from chromosome 1 to 17 (Table S2). The assembly completeness of the A. locustae genome was evaluated with benchmarked universal single-copy orthologue (BUSCO), for a total of 600 genes, and the HMMPER 3 homology (reference genome: E. cuniculi) search revealed 85% complete single-copy orthologues (C), <1% complete duplicated orthologues (D), <1% fragmented orthologues (F) and 14% missing (M) from the universal orthologue microsporidia database (Fig. S5). Among the genes predicted in the A. locustae genome, 1755 are single-exon genes, accounting for 94.5% of all genes found. By contrast, only 102 genes were found to contain multiple exons, accounting for 5.5% of the entire genome (Table S3). A brief parameter comparison of the A. locustae genome with other available microsporidian genomes is included in Table 1. The assembled sequence was submitted to GenBank. Using a combination of gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) and Eukaryotic Orthologous Groups (KOG) pathways (Tables S4–S6, Figs S6–S8), we successfully constructed a complete chromosome map for the genome of A. locustae (Fig. S9). The gene annotation and the locations of predicted genes are summarized in Fig. 1 (See Table S7 for details).
Table 1. Comparison of genome information between *A. locustae* and other microsporidia

| Genomic features         | *A. locustae* | *N. bombycis* | *N. ceranae* | *E. cuniculi* | *E. intestinalis* | *E. bieneusi* | *Octosporea bayeri* | *Spraguea lophii* | *Bursaphelenchus xylophilus* |
|--------------------------|---------------|---------------|--------------|--------------|-------------------|--------------|---------------------|------------------|-------------------------|
| Chromosome no.           | 17            | 18            | ND           | 11           | 11                | ≥6           | ND                  | 15               | 6                       |
| Genome size (Mbp)        | 3.2           | ND            | 7.7          | 2.9          | 2.3               | 6            | ≤24.2               | 6.2–7.3          | 63–75                   |
| Assembled (Mbp)          | 3.2           | 15.7          | 7.9          | 2.5          | 2.2               | 3.9          | 13.3                | 4.98             | 74.6                    |
| No. of scaffolds/contigs | 17            | 1605          | 5465         | 11           | 137               | 1646         | 41804               | 1392             | 1231                    |
| N50 (bp)                 | 183675        | 57394         | 2902         | ND           | ND                | 2349         | ND                  | 5923             | 1158                    |
| Genome coverage (%)      | 100           | 100           | 90           | 86           | 96                | 64           | 55                  | 70–80            | ND                      |
| G+C content (%)          | 41.6          | 31            | 27           | 48           | 41.4              | 26           | 26                  | 20               | 40.4                    |
| Predicted ORFs           | 1857          | 4458          | 2614         | 1999         | 1833              | 3804         | 2174                | 2543             | 18074                   |

ND, no data.

Fig. 1. A circular representation of the complete genome of *A. locustae*. DNA sequencing revealed that the genome size of *A. locustae* is 3170203 base pairs, with a total of 1857 predicted coding genes distributed on 17 chromosomes. The outermost circle shows chromosome size (kb) and the distribution of KEGG pathways, as indicated with colour-coded bars (see the colour bar for key) for each chromosome. The second circle from the outside shows the variation in GC content for each of the 17 chromosomes, characterized by a sharp decrease in GC content near the centre of each chromosome. The third circle from the outside represents the distribution of coding genes on the positive strand (red) and negative strand (green) of DNA, respectively. The non-coding RNA (ncRNA) detected is shown in the fourth circle. Information about long-fragment repeat sequences is represented in the fifth circle, and genomic long-fragment repeat sequences are indicated on the innermost circle.
Features of the *A. locustae* genome

Compared with the genome of *E. cuniculi* (Fig. S10), each *A. locustae* chromosome contains an ultra-low GC region of approximately 25 kb, except for chromosome 9, which has a contrasting ultra-high GC content (Fig. 2A). No coding genes were identified in any of the ultra-low GC regions. A scatter plot of GC content showed two similar Poisson distributions, with a small number of sequences exhibiting lower GC content (approximately 20%) and the rest exhibiting normal (approximately 45%) GC distribution (Fig. S11a,b), indicating that the genome of *A. locustae* may have two distinct forms of organization, which has not yet been observed in other microsporidian species (Fig. S12).

Analysis of repetitive sequences in the *A. locustae* genome resulted in the identification of a total of 298 simple tandem repeats. The number of low complexity repeats and microRNA was 86 and 45, respectively. Only one long terminal repeat was found in the whole genome, and no microsatellite DNA sequences were found (Table S8).

Transcriptional analysis of non-coding RNAs in the ultra-low GC regions showed a noticeable difference between the fat body and midgut (Fig. 2a). In the fat body, transcript levels in the ultra-low GC regions were relatively low (Fig. 2b), but the levels of the same DNA regions were significantly higher than those in the midgut for the majority of chromosomes (Fig. 2c). This finding indicates that non-coding RNA in the ultra-low GC content region is relatively highly expressed in the fat body. While the density of microsporidia in the midgut was very low after infection, that in the fat body was extremely high (Fig. 3d). As no coding genes were identified in the ultra-low GC regions, we speculated that these regions have an evolutionary effect that is currently unknown. Given that *A. locustae* carries out schizogamy in the host fat body, including cell division, we hypothesize that the ultra-low GC content regions may be associated with cell division and schizogamy.

Evolutionary analysis

Genomic and protein sequences of *A. locustae* were compared with those of several other single-cell organisms to assess genetic synteny and collinearity using the MCScanX method [55]. Organisms within the same taxonomic class generally showed markedly higher degrees of homology, while those in different classes showed little or almost no homology (Fig. 4a, b), and those in the microsporidian family showed greater homology (Fig. 4c). Within the microsporidia family, although genes from different host species showed high variability and some microsporidia exhibit gene loss and inversion, most genes in the microsporidian genomes still have good collinearity (Fig. 4d). Systematic analysis of *frataxin*, a key single-exon gene in the *A. locustae* mitosome, also found in other organisms such as fungi, prokaryotes, protozoa, etc. [59], provided an insight into the evolutionary status of the parasite, which is located at the base of the fungal evolutionary tree (Fig. 4e). In addition, the observation that single-exon genes occupy a predominant portion of the *A. locustae* genome is similar to observations in prokaryotic organisms (Fig. S13a, Table S3). This evidence suggested that
**A. locustae**, and the microsporidian family in general, are more primitive eukaryotes.

**Transcriptomic profiling of A. locustae in host tissues**

We calculated the differential locust genes in healthy and diseased locusts, and also the biological process of differential genes in *A. locustae* and locusts. The results showed that the number of fat body and midgut transcripts varied greatly between diseased and healthy locusts; in general, after *A. locustae* infected the locust, there were significantly more differential transcripts for fat body mobilization than for the midgut. At the same time, the expression level of transcripts in the diseased locust showed a downward trend (Fig. S13b-d). As for the biological process of differential genes, we can see that it mainly participates in the metabolic process (Fig. S13e).

The transcriptome profiles of *A. locustae* differed significantly in the fat body and midgut during the middle and late stages after infection (Fig. 3a, b), suggesting that *A. locustae* is present in the locust and exercises different functions affecting the host in these two kinds of tissue. In addition, we found from principal component analysis (PCA) that the main components of the transcripts in the fat bodies of diseased and healthy locusts differed, while those in diseased and healthy locust midguts did not differ significantly (Fig. 3c). Thus, the transcriptome in the host’s fat body responded strongly, further illustrating the main site affected by *A. locustae* is the fat body. As a control, the differential gene expression of *A. locustae* in the midgut and fat body of healthy locusts (NF-VS-NM) was 0 (Fig. S13a).

Based on GO terms and KEGG and KOG pathway analysis (Table S9), the upregulated microsporidian genes in the midgut include LRR receptor-like protein kinase, adenylyl cyclase and a large variety of leucine-rich repeat units and MEIS1 transcription factors, which are mainly involved in activating parasite–host membrane surface signalling pathways, thereby promoting microsporidian invasion of the host through the midgut. However, we did not detect differences in the expression of microsporidian polar tube proteins in different locust tissues, suggesting that a high level of polar tube expression should occur in the intestinal tract outside the midgut. In addition, during the late infection stage, the load of microsporidia in the fat body was much larger than that in the midgut (Fig. 3d), indicating that the fat body is the site where *A. locustae* eventually multiplies. These findings provide useful information on the molecular relationships driving spore reproduction.
Analysis of transcripts from the midgut of diseased locusts showed that a huge number of host genes were activated in response to *A. locustae* infection as compared to the uninfected healthy midgut (Table S10). The microsporidian spore load in the midgut was controlled at a low level in the host midgut (Fig. 3d), and this control was correlated with abundant expression of antimicrobial peptides and other defence genes, such as peroxiredoxin and amine oxidase. Although it appeared that *A. locustae* could inhibit the melanization
pathway, microsporidia can only survive temporarily in the midgut before being carried to fat body cells through vesicle transport (Fig. 5, Tables S9 and S10).

After the pathogen entered the host fat body, the spore load in the fat body increased greatly compared to that of the midgut (Fig. 3d), and this change was correlated with obvious inhibition of melanization compared to that in the healthy fat body. In particular, several critical phenol oxidases and peroxisome proliferator-activated receptors in the locust were inhibited, reducing melanization in the locust and also enabling immune escape by the parasite, aiding its survival and proliferation (Fig. 5, Tables S9 and S11).

**DISCUSSION**

The genome of *A. locustae* was sequenced by the Marine Biological Laboratory (USA) in 2002 with first-generation sequencing technology, obtaining approximately 648 contigs with a total size of approximately 2.1 Mb. However, few reports on the molecular biology of *A. locustae* were based on this sequence, and incompleteness of the dataset may have been an important limiting factor. In this study, an elaborate genome map of *A. locustae* was obtained through a combination of second- and third-generation sequencing technologies. A total of 17 chromosomes and 1857 genes were obtained, with a total size of approximately 3.2 Mb. The features of this microsporidium have been characterized. The 25 kb structure in the ultra-low GC region of the *A. locustae* genome was unique in that it is present in 16 of the 17 chromosomes sequenced, and this phenomenon has not been previously reported in other microsporidian species. We suspect that the sequence of the ultra-low GC region could represent the microsporidian centromere region, which is associated with cell division and schizogamy. Some centromere-related genes have been identified in the *A. locustae* genome, such as gene18 encoding the centromere-associated protein NUF2 and gene538 encoding the centromere-associated protein HEC1. These two proteins are involved in cell cycle control, cell division and chromosome partitioning. In addition, we found putative centromere/microtubule-binding protein 5 encoded by gene1155, which is homologous to the *family Encephalitozoon*, and centromere protein F encoded by gene1349, which is homologous to that of *Propithecus coquereli*. However, there are currently no reports on the centromere region of microsporidia related to proliferation, division, or the regulation of gene expression. Comparison with the midgut to determine whether the non-coding RNA encoded by the ultra-low GC content region of microsporidia in the fat body transcriptome is related to energy metabolism or immune evasion in the host–parasite interaction remains to be conducted.

The taxonomic status of microsporidia is a controversial topic. The level of conserved genes is close to that of fungi, representing either a basal branch or sister group [60]. Genomic and protein sequences of *A. locustae* were compared with those of several other unicellular organisms to identify genetic synteny and collinearity using the MCScanX method, and the results showed that *A. locustae* has high homology within the microsporidia group, while those in different classes showed little or essentially no homology. Within the microsporidia, despite large variation in genes among different host species and the presence of gene loss and inversion in some microsporidia, most genes in the microsporidian genome exhibit good collinearity. Additionally, our evolutionary analyses on the important mitosome gene *frataxin* showed that the microsporidia evolved side by side with fungi and prokaryotes, each as an independent group. Single-exon genes occupy a predominant portion of the *A. locustae* genome, similar to observations in prokaryotic organisms. These findings
provide some evidence that *A. locustae*, and the microsporidian family in general, are more primitive eukaryotes.

Through a combination of genome and transcriptome sequencing, a striking picture of the intensive interactions between parasite and host has been revealed. *A. locustae* proliferates in the fat body, and turns glucose into pyruvate through a series of reactions in the mitosome residual of the mitochondria (Fig. 5). Due to the lack of related enzymes in the mitosome, pyruvate may undergo decarboxylation through the action of pyruvate dehydrogenase E1 component (PDH-E1) [2]. By comparing the transcriptome before and after infection of a host locust, we found that *A. locustae* increased the expression of trehalose-6-phosphate synthase in the sugar metabolism pathway after infection, indicating accelerated glucose metabolism, which may lay the foundation for evasion of host immunity and rapid reproduction. In addition, expression of RAB5 in *A. locustae* increased after infection (Fig. S14), indicating an increase in vesicle transport of parasite spores, similar to that of macromolecules [2]. The elevated levels of otsA and RAB5, involved in energy metabolism and material transportation, serve as a sign of intensive interactions between *A. locustae* and the locust.

This study found that *A. locustae* could systematically inhibit essential pathways in the locust, including phenoloxidase PPO, the TCA cycle, the glyoxylate cycle and PPAR pathways in the host fat body, making it difficult for the host to melanize the pathogen (Fig. 5). In addition, *A. locustae* activated the locust energy transfer pathway, which transports ATP produced by the locust to the microsporidium, and ADP produced by the microsporidium was transported back to locust cells for reuse in the synthesis of ATP, meaning that the microsporidia could use energy substances in the locust fat body for reproduction. *A. locustae* uses surface protein recognition to identify the membrane proteins of the midgut, constructs a polar tube to pierce the midgut and host-to-parasite horizontal gene transfer. *mBio* 2015;6.

References

1. Williams BAP. Unique physiology of host-parasite interactions in microsporidia infections. Cell Microbiol 2009;11:1551–1560.
2. Katinka MD, Duprat S, Cornillet E, Méténier G, Thomarat F et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 2001;414:450–453.
3. Pan G, Xu J, Li T, Xia Q, Liu S-L et al. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. BMC Genomics 2013;14:186.
4. Corradi N, Haag KL, Pombert J-F, Ebert D, Keeling PJ. Draft genome sequence of the Daphnia pathogen *Octosporea bayeri*: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. Genome Biol 2009;10:R106.
5. Pombert JF, Haag KL, Beidas S, Ebert D, Keeling PJ. The Ordspora colligata genome: evolution of extreme reduction in microsporidia and host-to-parasite horizontal gene transfer. *mBio* 2015;6.
6. Chen Yping, Pettis JS, Zhao Y, Liu X, Tallon LJ et al. Genome sequencing and comparative genomics of honey bee microsporidia. Nseeoma Apis reveal novel insights into host-parasite interactions. BMC Genomics 2013;14:451.
7. Campbell SE, Williams TA, Yousef A, Soanes DM, Paszkiewicz KH et al. The genome of *Spraguea lophii* and the basis of host-microsporian interactions. *PloS Genet* 2013;9:e1003676.
8. Corradi N, Akiyoshi DE, Morrison HG, Feng X, Weiss LM et al. Patterns of genome evolution among the microsporidian parasites *Encephalitozoon cuniculi*, *Antanospora locustae* and *Enterocytozoon bieneusi*. *PloS One* 2007;2:e1277.
9. Germot A, Philippe H, Le Guyader H. Evidence for loss of mitochondria in microsporidia from a mitochondrial-type Hsp70 in *Nosema locustae*. *Mol Biochem Parasitol* 1997;87:159–168.
10. Fast NM, Logsdon JM, Doolittle WF. Phylogenetic analysis of the TATA box binding protein (TBP) gene from *Nosema locustae*: evidence for a microsporidium-fungi relationship and spliceosomal intron loss. *Mol Biol Evol* 1999;16:1415–1419.
11. Keeling PJ, Luker MA, Palmer JD. Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol Biol Evol* 2000;17:23–31.
12. Jones MD, Forn I, Gadelia C, Egan MJ, Bass D et al. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 2011;474:200–203.
13. Szumowski SC, Troemel ER. Microsporidia-host interactions. *Current opinion in microbiology*, Research Support, N.I.H., Extra- mural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. *Review* 2015;26:10–16.
14. Cameron SA, Lim HC, Lozier JD, Duennes MA, Thorp R. Test of the invasive pathogen hypothesis of bumble bee decline in North America. *Proc Natl Acad Sci U S A* 2016;113:4386–4391.
15. Xu X, Shen Z, Zhu F, Tao H, Tang X et al. Phylogenetic characterization of a microsporidium (*Endoreticulatus sp. Zenjiang*) isolated from the silkworm, *Bombyx mori*. Parasitol Res 2012;110:815–819.

Acknowledgements

The authors thank Jay Zhang for assistance in optimization of DNA sequencing, and also thank GENEWIZ information engineers Zhwei Ni and Yu Jiang for assistance with bioinformatics software and analysis.

Author contributions

Conceptualization: L.C., R.Z.M., L.Z. Data curation: L.C., X.G. Formal analysis: L.C., X.G., R.L. Funding acquisition: F.W., R.Z.M., L.Z. Investigation: L.C., X.G. Methodology: L.C., X.G., R.L., L.Z., R.H., L.W., Y.S., Z.X., T.L., X.N., F.N., S.H., Z.Z. Project administration: R.Z.M., L.Z. Resources: L.Z. Supervision: F.W., R.Z.M., L.Z. Validation: X.G., F.W., R.Z.M., L.Z. Visualization: L.C., X.G., R.L., R.H. Writing – original draft: L.C., X.G., R.L., R.Z.M., L.Z. Writing – review and editing: L.C., X.G., R.L., F.W., R.Z.M., L.Z.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

This work was supported in part by grants from the China Ministry of Agriculture (2014-Z18 to L. Z.), National Natural Science Foundation of China (81871130 to R. Z. M.), and Zhengzhou Key Laboratory of Molecular Biology (N2013GC1502). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
16. Han JE, Tang KFJ, Pantoja CR, Lightner DV, Redman RM et al. Detection of a new microsporidium Perezia sp. in shrimps Penaeus monodon and P. indicus by histopathology, in situ hybridization and PCR. Dis Aquat Organ 2016;120:165–171.

17. Jeong D-K, Won G-Y, Park B-K, Hur J, You J-Y et al. Occurrence and genotypic characteristics of Encephalitozoon cuniculi in pigs with diarrhea. Parasitol Res 2007;102:123–128.

18. Goodwin D, Gennari SM, Howe DK, Dubey JP, Zajac AM et al. Prevalence of antibodies to Encephalitozoon cuniculi in horses from Brazil. Vet Parasitol 2006;142:380–382.

19. Santin M, Fayer R. A longitudinal study of Encephalitozoon cuniculi in dairy cattle. Parasitol Res 2009;105:141–144.

20. Cisläková L, Literák I, Bálent P, Hipiková V, Levkutová M et al. Prevalence of antibodies to Encephalitozoon cuniculi (microsporidia) in angora goats—a potential risk of infection for breeders. Ann Agric Environ Med 2001;8:289–291.

21. Stentiford GD, Becnel J et al. Microsporidia − emergent pathogens in the global food chain. Trends Parasitol 2016;32:336–348.

22. Patel AK, Patel KK, Chickabasaviah YT, Shah SD, Patel DJ et al. Microsporidial polymyositis in human immunodeficiency virus-infected patients, a rare life-threatening opportunistic infection: clinical suspicion, diagnosis, and management in resource-limited settings. Muscle Nerve 2015;51:775–780.

23. Garg P. Microsporidia infection of the cornea—a unique and challenging disease. Cornea, Review 2013;32:533–38.

24. Heyworth MF. Changing prevalence of human microsporidiosis. Transactions of the Royal Society of tropical medicine and hygiene 2012;106:202–204.

25. Didier ES, Weiss LM. Microsporidiosis: not just in AIDS patients. Curr Opin Infect Dis 2011;24:490–495.

26. Didier ES, Stovall ME, Green LC, Brindley PJ, Sestak K et al. Epidemiology of microsporidiosis: sources and modes of transmission. Vet Parasitol 2006;126:145–166.

27. Corman RS, Chen YP, Schatz MC, Street C, Zhao Y et al. Genomic analyses of the microsporidian Nosema ceranae, an emergent honey bee pathogen. PLoS Pathog 2009;5:e1000466.

28. Canning EU. The life cycle of Nosema locustae Canning in Locusta migratoria migratoriorides Reiche and Fairmaire, and its infectivity to other hosts. J Insect Pathol 1962;4:237–247.

29. Henry JE. Experimental application of Nosema locustae for control of grasshoppers. J Invertebr Pathol 1971;18:389–394.

30. Brooks WM. CRC Handbook of Natural Pesticides. Vol V. Microbial Insecticides. Part A. Entomogenous Protozoa and Fungi. Ignoffo CM. Entomogenous protozoa: CRC Press; 1988. pp. 1–49.

31. Zhang L, Lecq M, Latchininsky A, Hunter D. Locust and grasshopper management. Annu Rev Entomol 2019;64:15–34.

32. Chen L, Li R, You Y, Zhang K, Zhang L. A Novel Spore Wall Protein from Antonospora locustae (Microsporidia: Nosematidae) Contributes to Sporulation. J Eukaryot Microbiol 2017;64:779–791.

33. Solter LF, Becnel JJ, Varva J. Manual of Techniques in Invertebrate Pathology. Academic Press; 2012.

34. McCarthy A. Third generation DNA sequencing: Pacific biosciences’ single molecule real time technology. Chem Biol 2010;17:675–676.

35. Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM et al. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 2015;33:623–630.

36. Goldberg SMD, Johnson J, Busam D, Feldblyum T, Ferriera S et al. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc Natl Acad Sci U S A 2006;103:11240–11245.

37. Istrail S, Sutton GG, Florea L, Halpern AL, Mоборрmy CM et al. Whole-Genome shotgun assembly and comparison of human genome assemblies. Proc Natl Acad Sci U S A 2004;101:1916–1921.

38. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL et al. The diploid genome sequence of an individual human. PLoS Biol 2007;5:e254.

39. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP et al. A whole-genome assembly of Drosophila. Science 2000;287:2196–2204.

40. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ et al. The sequence of the human genome. Science 2001;291:1304–1315.

41. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 2007;23:673–679.

42. Keller O, Kolimar M, Stanke M, Waack S. A novel hybrid genome prediction method employing protein multiple sequence alignments. Bioinformatics 2011;27:757–763.

43. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M et al. The gene ontology (go) database and informatics resource. Nucleic Acids Res 2004;32:D258–261.

44. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.

45. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 2003;4:41.

46. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–960.

47. Lagesen K, Hallin P, Redland EA, Staerfeldt HH, Rognes T et al. RNAServer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–3108.

48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–2120.

49. Li X, Nair A, Wang S, Wang L. Quality control of RNA-Seq experiments. Methods Mol Biol 2015;1269:137–146.

50. Wang X, Fang X, Yang P, Jiang X, Jiang F et al. The locust genome provides insight into Swarm formation and long-distance flight. Nat Commun 2014;5:2957.

51. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–628.

52. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–360.

53. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015;31:3210–3212.

54. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol 2011;7:e1002195.

55. Wang Y, Tang H, Debarry JD, Tan X, Li J et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 2012;40:464.

56. Katah K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–780.

57. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009;25:1972–1973.

58. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52:696–704.

59. Bridwell-Rabb J, Iannuzzi C, Pastore A, Barondeau DP. Effector role reversal during evolution: the case of frataxin in Fe-S cluster biosynthesis. Biochemistry 2012;51:2506–2514.

60. Han B, Weiss LM. Microsporidia: obligate intracellular pathogens within the fungal Kingdom. Microbiol spectr 2017;5.

61. Cerenius L, Kawabata S, Lee BL, Nonaka M, Soderhall K. Proteolytic cascades and their involvement in invertebrate immunity. Trends Biochem Sci 2010;35:575–583.

62. Guo Y, Zhang J, Yu R, Zhu KY, Guo Y et al. Identification of two new cytchrome P450 genes and RNA interference to evaluate their roles in detoxification of commonly used insecticides in Locusta migratoria. Chemosphere 2012;87:709–717.
63. Park DW, Kim JS, Chin BR, Baek SH. Resveratrol inhibits inflammation induced by heat-killed *Listeria monocytogenes*. *J Med Food* 2012;15:788–794.

64. Ragab A, Buechling T, Gesellchen V, Spirohn K, Boettcher AL et al. Drosophila Ras/MAPK signalling regulates innate immune responses in immune and intestinal stem cells. *Embo J* 2011;30:1123–1136.

---

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.