

UC Berkeley

UC Berkeley Previously Published Works

Title

The Genome Sequence of the Leaf-Cutter Ant *Atta cephalotes* Reveals Insights into Its Obligate Symbiotic Lifestyle

Permalink

<https://escholarship.org/uc/item/8xh6k5dd>

Journal

PLOS Genetics, 7(2)

ISSN

1553-7390

Authors

Suen, Garret
Teiling, Clotilde
Li, Lewyn
et al.

Publication Date

2011

DOI

10.1371/journal.pgen.1002007

Peer reviewed

The Genome Sequence of the Leaf-Cutter Ant *Atta cephalotes* Reveals Insights into Its Obligate Symbiotic Lifestyle

Garret Suen^{1,2*}, Clotilde Teiling^{3,9}, Lewyn Li^{4,9}, Carson Holt⁵, Ehab Abouheif⁶, Erich Bornberg-Bauer⁷, Pascal Bouffard^{4a}, Eric J. Caldera^{1,8}, Elizabeth Cash⁹, Amy Cavanaugh^{1,10}, Olgert Denas¹¹, Eran Elhaik¹², Marie-Julie Favé⁶, Jürgen Gadau⁹, Joshua D. Gibson⁹, Dan Graur¹³, Kirk J. Grubbs^{1,14}, Darren E. Hagen¹⁵, Timothy T. Harkins^{3a,b}, Martin Helmkamp⁹, Hao Hu⁵, Brian R. Johnson¹⁶, Jay Kim¹⁷, Sarah E. Marsh^{1,18}, Joseph A. Moeller^{1,2}, Mónica C. Muñoz-Torres¹⁵, Marguerite C. Murphy¹⁹, Meredith C. Naughton²⁰, Surabhi Nigam¹⁹, Rick Overson⁹, Rajendhran Rajakumar⁶, Justin T. Reese¹⁵, Jarrod J. Scott^{1,2,21}, Chris R. Smith²⁰, Shu Tao¹⁵, Neil D. Tsutsui¹⁶, Lumi Viljakainen²², Lothar Wissler⁷, Mark D. Yandell⁵, Fabian Zimmer⁷, James Taylor^{11,23}, Steven C. Slater^{1,2}, Sandra W. Clifton²⁴, Wesley C. Warren²⁴, Christine G. Elsik¹⁵, Christopher D. Smith¹⁷, George M. Weinstock²⁴, Nicole M. Gerardo²³, Cameron R. Currie^{1,2,21*}

1 Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **2** Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **3** Roche Diagnostics, Indianapolis, Indiana, United States of America, **4** 454 Life Sciences, Branford, Connecticut, United States of America, **5** Department of Human Genetics, University of Utah, Salt Lake City, Utah, United States of America, **6** Department of Biology, McGill University, Montreal, Canada, **7** Institute for Evolution and Biodiversity, University of Münster, Münster, Germany, **8** Department of Zoology, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **9** School of Life Sciences, Arizona State University, Tempe, Arizona, United States of America, **10** Department of Biology, University of Wisconsin-Rock County, Janesville, Wisconsin, United States of America, **11** Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia, United States of America, **12** The Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **13** Department of Biology and Biochemistry, University of Houston, Houston, Texas, United States of America, **14** Cellular and Molecular Pathology Graduate Program, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **15** Department of Biology, Georgetown University, Washington, D.C., United States of America, **16** Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, California, United States of America, **17** Department of Biology, San Francisco State University, San Francisco, California, United States of America, **18** Department of Genetics, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **19** Department of Computer Science, San Francisco State University, San Francisco, California, United States of America, **20** Department of Biology, Earlham College, Richmond, Indiana, United States of America, **21** Smithsonian Tropical Research Institute, Balboa, Ancon, Panama, **22** Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America, **23** Department of Biology, Emory University, Atlanta, Georgia, United States of America, **24** The Genome Center, Washington University School of Medicine, St. Louis, Missouri, United States of America

Abstract

Leaf-cutter ants are one of the most important herbivorous insects in the Neotropics, harvesting vast quantities of fresh leaf material. The ants use leaves to cultivate a fungus that serves as the colony's primary food source. This obligate ant-fungus mutualism is one of the few occurrences of farming by non-humans and likely facilitated the formation of their massive colonies. Mature leaf-cutter ant colonies contain millions of workers ranging in size from small garden tenders to large soldiers, resulting in one of the most complex polymorphic caste systems within ants. To begin uncovering the genomic underpinnings of this system, we sequenced the genome of *Atta cephalotes* using 454 pyrosequencing. One prediction from this ant's lifestyle is that it has undergone genetic modifications that reflect its obligate dependence on the fungus for nutrients. Analysis of this genome sequence is consistent with this hypothesis, as we find evidence for reductions in genes related to nutrient acquisition. These include extensive reductions in serine proteases (which are likely unnecessary because proteolysis is not a primary mechanism used to process nutrients obtained from the fungus), a loss of genes involved in arginine biosynthesis (suggesting that this amino acid is obtained from the fungus), and the absence of a hexamerin (which sequesters amino acids during larval development in other insects). Following recent reports of genome sequences from other insects that engage in symbioses with beneficial microbes, the *A. cephalotes* genome provides new insights into the symbiotic lifestyle of this ant and advances our understanding of host-microbe symbioses.

Citation: Suen G, Teiling C, Li L, Holt C, Abouheif E, et al. (2011) The Genome Sequence of the Leaf-Cutter Ant *Atta cephalotes* Reveals Insights into Its Obligate Symbiotic Lifestyle. *PLoS Genet* 7(2): e1002007. doi:10.1371/journal.pgen.1002007

Editor: Gregory Copenhaver, The University of North Carolina at Chapel Hill, United States of America

Received: September 30, 2010; **Accepted:** December 30, 2010; **Published:** February 10, 2011

Copyright: © 2011 Suen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Roche Diagnostics 10 Gigabase Sequencing and Transcriptome Analysis Grant awarded to GS, JT, SCS, SWC, GMW, NMG, and CRC. This work was also funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) supporting GS, CRC, JAM, and SCS; a Volkswagen Foundation grant supporting EB-B and LW; a Deutsche Forschungsgemeinschaft (DFG) grant BO2544-4/1 to EB-B; a National Science Foundation Graduate Research Fellowship supporting EJC; a University of Wisconsin-Madison Colleges Summer Research Grant supporting AC; a US National Library of Medicine Grant LM010009-01 to DG; a Smithsonian Institution Predoctoral Fellowship supporting JJS; a National Institutes of Health grant 5R01HG004694 to MDY supporting the MAKER genome annotation; and a National Institutes of Health NIMH grant 5SC2MH086071 to CDS. This material is also based upon work support by the National Institute of Food and Agriculture, United States Department of Agriculture, under ID number WISO1321, a University of Wisconsin-Madison CALS grant, and the National Science Foundation grants DEB-0747002, MCB-0702025, and MCB-0731822 to CRC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gsuen@wisc.edu (GS); currie@bact.wisc.edu (CRC)

☯ These authors contributed equally to this work.

✉ Current address: Regeneron Pharmaceuticals, Tarrytown, New York, United States of America

✉ Current address: Life Technologies Corporation, Carlsbad, California, United States of America

Introduction

Ants are one of the most successful insects on earth, comprising up to 20% of all terrestrial animal biomass and at least 25% of the entire animal biomass in the New World Tropics [1]. One of the most conspicuous and prolific Neotropical ants are the leaf-cutters (Tribe: Attini), so-called because of their leaf-cutting behavior [2]. Leaf-cutters are unique among ants because they obligately farm a specialized, mutualistic fungus that serves as their primary food source [3]. Using a complex system of trails, foraging ants seek out and cut leaves (Figure 1A) that they use to manure a fungal crop in specialized subterranean fungus gardens (Figure 1B) within their colonies. Fungus farming by ants is exclusive to the New World and is thought to have evolved once 50 million years ago [4], culminating in the leaf-cutter ants. A single mature colony of the genus *Atta* can fill a volume of up to 600 m³ and their fungus gardens can support millions of workers capable of harvesting over 400 kg of leaf material (dry weight) annually [1]. These ants are thus one of the most widespread and important polyphagous insect herbivores in the Neotropics.

The importance of leaf-cutter ants in Neotropical rainforest ecology lies in their ability to substantially alter arboreal foliage through their extensive leaf-cutting activities. Estimates suggest that leaf-cutter ants remove 12–17% of the total leaf production in tropical rainforests [1]. As a group, they harvest more plant biomass than any other Neotropical herbivore including mammals and other insects. As a result, leaf-cutter ants are a major human agricultural pest, responsible for billions of dollars in economic loss each year [5]. These ants do, however, have a positive impact on rainforest ecosystems, as they contribute to rapid soil turnover through their nest excavation activities [6], stimulate plant growth by cutting vegetation [7], and help to recycle organic carbon [1].

In addition to their importance in Neotropical ecosystems, leaf-cutter ants also serve as a model for understanding the ecology and evolution of host-microbe symbioses [8]. In return for receiving a continuous supply of leaf-material, protection from competitors, and dispersal, the fungus these ants grow provide nutrients in the form of specialized hyphal swellings called gongylidia. Gongylidia, which contain a mixture of carbohydrates, amino acids, proteins, lipids, and vitamins [9], is the sole food source for developing larvae. The fungus garden is also known to harbor other microbial symbionts including nitrogen-fixing bacteria that provide both fungus and ants with nitrogen [10], and a diverse community of fungus garden bacteria that appear to help the fungus degrade plant biomass [11]. The complexity of the leaf-cutter ant symbiosis is further highlighted by the presence of a specialized microfungus pathogen that exploits the ant-fungus mutualism [12,13]. As a result, the leaf-cutter ant symbiosis comprises at least three established mutualists and one specialized pathogen. With the

reported presence of additional microbial symbionts from *Acromyrmex* leaf-cutter ants [14–19], and the isolation of numerous microbes from other fungus-growing ants [20–22], this ant-microbe symbiosis is perhaps one of the most complex examples of symbiosis currently described.

Leaf-cutter ants in the genus *Atta* are also known for their morphologically diverse caste system (Figure 1C), which reflects their complex division of labor [23,24]. For example, the overall body size of *Atta cephalotes* workers varies tremendously (i.e., head widths (HW) ranging from 0.6 mm to 4.5 mm [23]), and these differences correspond to the tasks performed by workers. The smallest workers (HW 0.8–1.6 mm) engage in gardening and brood care as their small mandibles allow them to manage the delicate fungal hyphae and manipulate developing larvae. Some of these workers are also responsible for processing plant material collected by foragers by clipping large pieces of leaf material into smaller fragments to manure the fungus. Larger workers (HW >1.6 mm) are responsible for foraging, as they have mandibles powerful enough to cut through leaves and other vegetation [24]. The largest workers form a true soldier caste, which are involved primarily in nest excavation and colony defense [23,24].

To gain a better understanding of the biology of leaf-cutter ants, we sequenced the genome of *Atta cephalotes* using 454 pyrosequencing technology [25] and generated a high-quality *de novo* assembly and annotation. Analysis of this genome sequence reveals a loss of genes associated with nutrient acquisition and amino acid biosynthesis. These genes appear to be no longer required because the fungus may provide these nutrients. With the recent reports of genomes from other social hymenopterans [26,27] and insects that engage in microbial mutualisms [28,29], the *A. cephalotes* genome contributes to our understanding of social insect biology and provides insights into the interactions of host-microbe symbioses.

Results/Discussion

Sequencing, Assembly, and Annotation of the *Atta cephalotes* Genome

Three males from a mature *Atta cephalotes* colony in Gamboa, Panama were collected and sequenced using 454-based pyrosequencing [25] with both fragment and paired-end sequencing approaches. A total of 12 whole-genome shotgun fragment runs were performed using the 454 FLX Titanium platform in addition to two sequencing runs of an 8 kbp insert paired-end library, and one run of a 20 kbp insert paired-end library. Assembly of these data resulted in a genome sequence of 290 Mbp, similar to the 300 Mbp genome size previously estimated for *A. cephalotes* [30]. The genome is spread across 42,754 contigs with an average length of 6,788 bp and an N50 of 14,240 bp (Table 1). Paired-end

Author Summary

Leaf-cutter ant workers forage for and cut leaves that they use to support the growth of a specialized fungus, which serves as the colony's primary food source. The ability of these ants to grow their own food likely facilitated their emergence as one of the most dominant herbivores in New World tropical ecosystems, where leaf-cutter ants harvest more plant biomass than any other herbivore species. These ants have also evolved one of the most complex forms of division of labor, with colonies composed of different-sized workers specialized for different tasks. To gain insight into the biology of these ants, we sequenced the first genome of a leaf-cutter ant, *Atta cephalotes*. Our analysis of this genome reveals characteristics reflecting the obligate nutritional dependency of these ants on their fungus. These findings represent the first genetic evidence of a reduced capacity for nutrient acquisition in leaf-cutter ants, which is likely compensated for by their fungal symbiont. These findings parallel other nutritional host-microbe symbioses, suggesting convergent genomic modifications in these types of associations.

sequencing (8 kbp and 20 kbp inserts) generated 2,835 scaffolds covering 317 Mbp with an N50 scaffold size of 5,154,504 bp. The disparity between contig and scaffold size may be accounted for by the number of repeats present in this genome (see below) leading

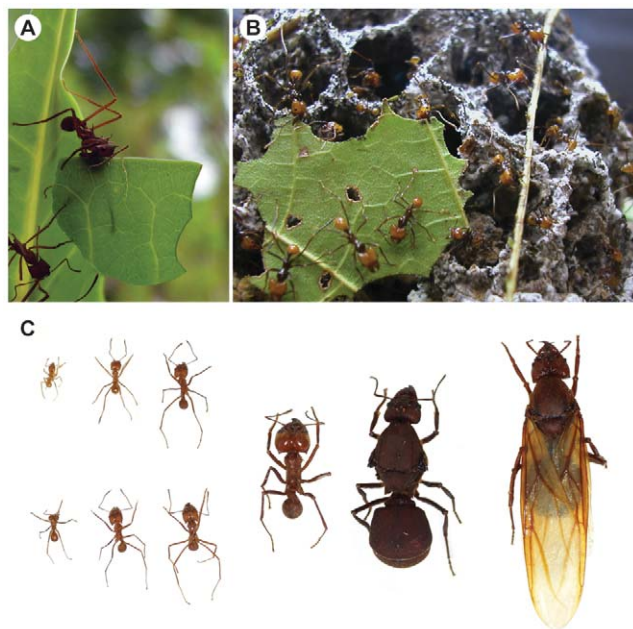


Figure 1. The leaf-cutter ant *Atta cephalotes*. Leaf-cutter ants harvest fresh leaf material which they cut from Neotropical rainforests (a) and use them to grow a fungus that serves as the colony's primary food source (b). These ants display a morphologically diverse caste system that reflects a complex division of labor (c) correlated to specific tasks within the colony. These include small workers that undertake garden management and brood care, medium workers that forage leaves, large workers that can serve as soldiers, and winged sexuals that lose their wings after mating. [Photo Credits: foraging workers, Jarrod J. Scott/University of Wisconsin-Madison; fungus garden, Austin D. Lynch/University of Wisconsin-Madison; caste morphology, used under the GNU Free Documentation License version 1.3]. doi:10.1371/journal.pgen.1002007.g001

Table 1. General assembly statistics for the genome of the leaf-cutter ant *Atta cephalotes*.

Statistic	<i>Atta cephalotes</i>
Number of contigs	42,754
Average length of contigs	6,788 bp
Total length of contigs	290,223,730 bp
Contig size N50	14,240 bp
Number of scaffolds	2,835
Average length of scaffolds	112,054 bp
Total length of all scaffolds	317,672,992 bp
Scaffold size N50	5,154,504 bp
Total coverage	18-20X
Predicted genes (OGS1.2)	18,093

doi:10.1371/journal.pgen.1002007.t001

to an inflated assembly size due to chimeric contigs. Based on the total amount of base pairs generated and its predicted genome size, we estimate that the coverage of the *A. cephalotes* genome is 18-20X.

To determine the completeness of the *A. cephalotes* genome sequence, we performed three analyses. First, we compared the *A. cephalotes* genome annotation against a set of core eukaryotic genes using CEGMA [31], and found that 234 out of 248 core proteins (94%) were present and complete, while 243 (98%) were present and partially represented. Second, we analyzed the cytoplasmic ribosomal proteins (CRPs) in the *A. cephalotes* genome and identified a total of 89 genes (Text S1). These encode the full complement of 79 CRPs known to exist in animals, nine of which are represented by gene duplicates (RpL11, RpL14, RpS2, RpS3, RpS7, RpS13, RpS19, RpS28) or triplicates (RpL22). The presence of a complete set of these numerous genes, which are widely distributed throughout the genome, confirmed the high-quality of the *A. cephalotes* genome sequence (Text S2). Finally, we found that the genome of *A. cephalotes* contains 66 of the 67 known oxidative phosphorylation (OXPHOS) nuclear genes in insects (Text S3). The only OXPHOS gene missing, *cox7a*, we found to also be missing in the two ants *Camponotus floridanus* and *Harpegnathos saltator* and the honey bee *Apis mellifera*. The presence of this gene in the jewel wasp *Nasonia vitripennis* (along with other holometabolous insects), suggests an aculeate Hymenoptera-specific loss, rather than a lack of genome coverage for *A. cephalotes*.

We also generated an annotation for the *A. cephalotes* genome using a combined approach of electronically-generated annotations followed by manual review and curation of a subset of gene models. Expressed Sequence Tags (ESTs) generated from a pool of workers consisting of different ages and castes from a laboratory-maintained colony of *A. cephalotes* was used in conjunction with the MAKER [32] automated annotation pipeline to generate an initial genome annotation. This electronically-generated annotation set (OGS1.1) contained a total of 18,153 gene models encoding 18,177 transcripts (See Materials and Methods), 7,002 of which had EST splice site confirmation and 7,224 had at least partial EST overlap. The MAKER-produced gene annotations were used for further downstream review and manual curation of over 500 genes across 16 gene categories (Table S1). Significant findings from this annotation are highlighted below, with additional details of our full analysis described in Text S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20).

In addition to the *A. cephalotes* genome sequence, we also recovered an 18–20X coverage complete and circular mitochondrial genome, which showed strong whole sequence identity to the mitochondrial genome sequence reported for the solitary wasp *Diadegma semiclausum* [33]. A synteny analysis of the predicted genes on the *A. cephalotes* mitochondrial genome showed near-identical gene order with that of *A. mellifera* [34] (Text S4).

Repetitive DNA

The *A. cephalotes* assembly contains 80 Mbp of repetitive elements, which accounts for 25% of the predicted assembly (Table S2). The large majority of these are interspersed repeats, which account for 70 Mbp (21%). Many of these repeats are transposable elements (TEs), with DNA TEs the most abundant and accounting for 14.3 Mbp (4.5%). A large number of retroviral element fragments were also identified, with Gypsy/DIRS1 and L2/CR1/Rex as the most abundant. However, the majority of interspersed elements (51.8 Mbp) were similar to *de novo* predictions that we could not be classified to a specific family (Table S2). Improvements to the assembly, integration of repeat annotation evidence, and manual curation will be necessary to determine if these elements represent new TE families or complex nests of interspersed repeats.

Given the obligate association between *A. cephalotes* and its fungal cultivar, we investigated the possibility that the *A. cephalotes* genome might contain transposable elements commonly found in fungi. This was done by re-analyzing the genome using a TE library optimized for the detection of Fungi and Viridiplantae. We did not find evidence for any high-scoring or full-length retroviral or DNA TEs from either of these taxa present in the *A. cephalotes* genome.

Our estimate that 25% of *A. cephalotes* assembly contains repetitive elements may be ambiguous because our assembly spans 317 Mbp and the estimated genome size for *A. cephalotes* is 300 Mbp [30]. These predictions are, however, more similar to other ant species [27] and *N. vitripennis* [35] than to *A. mellifera* [28], which lacks the majority of retroviral elements and other transposable elements (TE) found in *A. cephalotes*.

Global Compositional Analysis

Eukaryotic genomes can be understood from the perspective of their nucleotide topography, particularly with respect to their GC content. Previous work has shown that animal genomes are not uniform, but are composed of compositional domains including homogeneous and nonhomogeneous stretches of DNA with varying GC composition [36]. A global composition analysis was performed for *A. cephalotes* and the compositional distribution was compared to those of other insect genomes, as described in Text S5. This analysis revealed that *A. cephalotes* has a compositional distribution similar to other animal genomes, with an abundance of short domain sequences and few long domain sequences. *A. cephalotes* also has the largest number of long GC-rich domain sequences when compared to other insect genomes, with over six times the number of long GC-rich domain sequences than the *N. vitripennis* genome. When genes are mapped to compositional domains in the *A. cephalotes* genome, we find that they are uniformly distributed across the entire genome, in contrast to *N. vitripennis* and *A. mellifera*, which have genes occurring in more GC-poor regions of their genomes.

DNA Methylation

The methylation of genes has been reported for other hymenopterans including *A. mellifera* [37] and *N. vitripennis* [35]. In insects, it is thought that this process contributes to gene

silencing [37], but recent reports suggest a positive correlation between DNA methylation and gene expression [38,39]. DNA methylation is thought to involve three genes: *dnmt1*, *dnmt2*, and *dnmt3* [40], although the precise role of *dnmt2* remains unresolved. We found all three genes as single copies in *A. cephalotes*, which is similar to the other ants [27] but in contrast to *A. mellifera* and *N. vitripennis* where *dnmt1* has expanded to two and three copies, respectively [35] (Text S6). *Dnmt3* is known to be involved in caste development in *A. mellifera* [41], and the presence of this gene in *A. cephalotes* may therefore indicate a similar role.

RNAi

RNA interference is a mechanism through which the expression of RNA transcripts is modulated [42]. We annotated a total of 29 different RNAi-related genes in *A. cephalotes*, including most of the genes involved in the microRNA pathway, the small interfering RNA pathway, and the piwi-interacting RNA pathway (Text S7). All detected RNAi genes were found as single copies except for two copies of the gene *loquacious*. One of these contains three double-stranded RNA binding domains characteristic of *loquacious* in *D. melanogaster* [43], whereas the other contains only two. It is not known what role this second *loquacious*-like gene plays in *A. cephalotes* and future work is needed to deduce its role.

The Insulin Signaling Pathway

The insulin signaling pathway is a highly-conserved system in insects that plays a key role in many processes including metabolism, reproduction, growth, and aging [44]. An analysis of the insulin signaling system in *A. cephalotes* reveals that it has all of the core genes known to participate in this pathway (Text S8). One of the hallmarks of *A. cephalotes* biology is its complex size-based caste system and, although virtually nothing is known about the genetic basis of caste development in this ant, it is currently thought that it is intrinsically linked to brood care and the amount of nutrients fed to developing larvae [1]. Given the importance of the insulin signaling system in nutrition, it is likely that this pathway is involved in caste differentiation in *A. cephalotes*, as has been shown for *A. mellifera* [45].

Yellow and Major Royal Jelly Proteins

The yellow/major royal jelly proteins are encoded by an important class of genes and in *A. mellifera* they are thought to be integral to many major aspects of eusocial behavior [46]. For example, members of these genes are implicated in both caste development and sex determination. An analysis of this gene family in *A. cephalotes* revealed a total of 21 genes, 13 of which belong to the yellow genes and 8 of which encode major royal jelly proteins (MRJP) (Text S9). In general, the yellow genes display one-to-one orthology with yellow genes in other insects like *Drosophila melanogaster* and *N. vitripennis*. With eight members in the MRJP subfamily, which is restricted to Hymenoptera, the number of MRJP genes in *A. cephalotes* is similar to the number reported for other Hymenoptera [35,46]. However, five of the eight genes in *A. cephalotes* are putative pseudogenes. This may indicate that a high copy number of MRJPs may be an ancestral feature and that *Atta* is in the process of losing these genes. The loss of MRJPs may be a common theme among ants, as the recently reported genome sequences for *C. floridanus* and *H. saltator* revealed only one and two MRJP genes, respectively [27].

Wing Polyphenism

Wing polyphenism is a universal feature of ants that has contributed to their evolutionary success [1]. The gene network

that underlies wing polyphenism in ants responds to environmental cues such that this network is normally expressed in winged queens and males, but is interrupted at specific points in wingless workers [47]. We therefore predict that the differential expression of this network between queens and workers may be regulated by epigenetic mechanisms as has been demonstrated in honey bees [41]. In *A. mellifera*, developmental and caste specific genes have a distinct DNA methylation signature (high-CpG dinucleotide content) relative to other genes in the genome [48]. Because *A. cephalotes* has more worker castes than other ant species [23] (Figure 1C), we predict that the DNA methylation signature of genes underlying wing polyphenism will also be distinct relative to other genes in its genome. To test this prediction, we analyzed the sequence composition of wing development genes in *A. cephalotes*, and found that they exhibit a higher CpG dinucleotide content than the rest of the genes in the genome (Text S10). Previous experiments have shown that genes with a high-CpG dinucleotide content can be differentially methylated in specific tissues or different developmental stages [49]. Therefore, DNA methylation may facilitate the caste-specific expression of genes that underlie wing polyphenism in *A. cephalotes*. This may be a general feature of genes that underlie polyphenism.

Desaturases

An important aspect of the eusocial lifestyle is communication between colony members, specifically in differentiating between individuals that belong to the same colony and those that do not. Nestmate recognition in many ants is mediated by cuticular hydrocarbons (CHCs) [50], and nearly 1,000 of these compounds have been described. In ants, CHC biosynthesis involves $\Delta 9/\Delta 11$ desaturases, which are known to produce alkene components of CHC profiles [51]. We analyzed the $\Delta 9$ desaturases in the genome of *A. cephalotes* and detected nine genes localized to a 200 kbp stretch on a single scaffold in addition to four other $\Delta 9$ desaturase genes on other scaffolds (Text S11). In contrast, the seven genes found in *D. melanogaster* are more widely distributed along one chromosome. The number of $\Delta 9$ desaturase genes in *A. cephalotes* is similar to the 9 and 16 found in *A. mellifera* and *N. vitripennis*, respectively. A phylogenetic analysis of these genes supports their division into five clades, with eight $\Delta 9$ desaturase genes falling in a single clade suggesting an expansion of these genes possibly related to an increased demand for chemical signal variability during ant evolution (Text S11). Interestingly, the phylogeny also supports an expansion in this type of $\Delta 9$ desaturase genes within *N. vitripennis* but not in *A. mellifera*.

Immune Response

All insects have innate immune defenses to deal with potential pathogens [52] and *A. cephalotes* is no exception with a total of 84 annotated genes found to be involved in this response (Text S12). These include the intact immune signaling pathways Toll, Imd, Jak/Stat, and JNK. When compared to solitary insects like *D. melanogaster* and *N. vitripennis*, *A. cephalotes* has fewer immune response genes and better resembles what is known for the eusocial *A. mellifera* [53]. The presence of other defenses in *A. cephalotes*, such as antibiotics produced by metapleural glands [54–56], may account for the paucity of immune genes. Furthermore, social behavioral defenses may also participate in the immune response, as has been suggested for *A. mellifera* [53].

Orthology Analysis

A set of shared orthologs was determined among *A. cephalotes*, *A. mellifera*, *N. vitripennis*, and *D. melanogaster* (Figure 2). A total of 5,577 orthologs were found conserved across all four insect genomes,

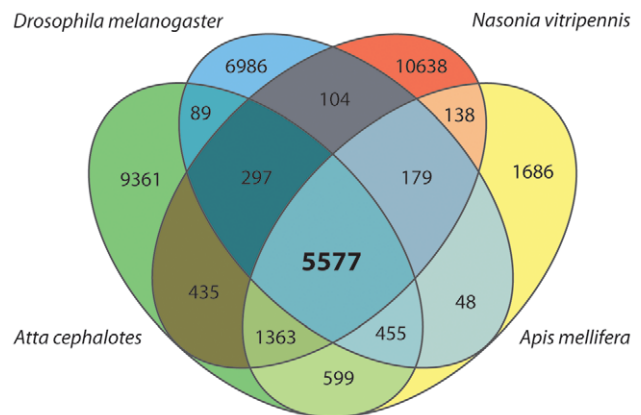


Figure 2. Orthology analysis of the *Atta cephalotes* predicted peptide sequences (green) against the proteomes of the fly *Drosophila melanogaster* (blue), the wasp *Nasonia vitripennis* (red), and the honey bee *Apis mellifera* (yellow).
doi:10.1371/journal.pgen.1002007.g002

with an additional 1,363 orthologs conserved across the three hymenopteran genomes. A further, 599 orthologs were conserved between *A. cephalotes* and *A. mellifera*, perhaps indicating genes that are specific to a eusocial lifestyle. We also found 9,361 proteins that are unique to *A. cephalotes*, representing over half of its predicted proteome. These proteins likely include those specific to ants or to *A. cephalotes*.

We then analyzed the proteins that were found to be specific to *A. cephalotes* and determined those Gene Ontology (GO) [57] terms that are enriched in these proteins, relative to the rest of the genome (Table S3). We found many GO terms that reflect the biology of *A. cephalotes* and ants in general. For example, we find proteins with GO terms that reflect the importance of communication. These include proteins associated with olfactory receptor activity, odorant binding function, sensory perception, neurological development, localization at the synapse, and functions involved in ligand-gated and other membrane channels.

Gene Comparisons within Hymenopteran Genomes

To focus on Hymenoptera evolution, we compared the *A. cephalotes* genome to 4 other hymenopterans including the ants *C. floridanus* and *H. saltator*, the honey bee *A. mellifera*, and the solitary parasitic jewel wasp *N. vitripennis*. We used the eukaryotic clusters of orthologous groups (KOG) ontology [58] to annotate the predicted proteins from all of these genomes and performed an enrichment analysis by comparing the KOGs of the social insects *A. cephalotes*, *C. floridanus*, *H. saltator*, and *A. mellifera* against the KOGs of the non-social *N. vitripennis* as shown in Table S4.

A detailed analysis of KOGs within each over- and under-represented category is highly suggestive of *A. cephalotes* biology (Table S5). One of the most over-represented KOGs in *A. cephalotes* includes the 69 copies of the RhoA GTPase effector *diaphanous* (KOG1924). In contrast, all of the other hymenopteran genomes have substantially less copies of this gene. RhoA GTPase *diaphanous* is known to be involved in actin cytoskeleton organization and is essential for all actin-mediated events [59]. The large number of these genes in *A. cephalotes* may relate to the extensive cytoskeletal changes that occur during caste differentiation. One of these genes (*ACEP_00016791*) was found to exhibit high single nucleotide polymorphism (SNPs) (Text S13). Given that genes involved in caste development in other social insects like *A. mellifera* also have high SNPs [60,61], this may indicate that this gene is important for

caste determination in *A. cephalotes*. *A. cephalotes* is also significantly over-represented in the dosage compensation complex subunit (KOG0921), the homeobox transcription factor SIP1 (KOG3623), the muscarine acetylcholine receptor (KOG4220), the cadherin EGF LAG seven-pass GTP-type receptor (KOG4289), and the calcium-activated potassium channel *slowpoke* (KOG1420), relative to *N. vitripennis*. Many of these genes have been implicated in *D. melanogaster* larval development, specifically during nervous system formation [62,63]. As a result, an over-representation of these genes in *A. cephalotes* relative to *N. vitripennis* may indicate their association with a eusocial lifestyle, and in particular, caste and subcaste differentiation.

Genes that were found to be under-represented in *A. cephalotes* relative to *N. vitripennis* include core histone genes, nucleosome-binding factor genes, serine protease trypsins, and cytochrome P450s (Table S5). These findings were confirmed by a domain-based comparison between *A. cephalotes* and all other sequenced insects (Text S14). One of the most under-represented KOGs is trypsin, a serine protease used in the degradation of proteins into their amino acid constituents. Trypsins in *N. vitripennis* are known to be part of the venom cocktail injected into its host, which helps necrotization and initiates the process of amino acid acquisition for developing larvae [35,64]. In contrast to the protein-rich diet of *N. vitripennis*, *A. cephalotes* feed on gongylidia produced by their fungus, which represents a switch to a carbohydrate-rich (60% of mixture) diet [65]. These differences in diet may explain the under-representation of trypsin in *A. cephalotes*, as trypsin is likely not the primary mechanism used to digest nutrients obtained from the fungal cultivar. Our analysis also revealed a reduction of trypsin genes in the other social insects relative to *N. vitripennis*, and this may also reflect their diets. For example, honey dew is a major component of the diet of *C. floridanus* and contains primarily sugars [1], while the honey/pollen diet of *A. mellifera* is composed primarily of carbohydrates, lipids, carbohydrates, vitamins, and some proteins [66]. Because this under-representation of trypsin is consistent across social insects when compared to other sequenced insects (Table S5, Text S14), this reduction may reflect the specific dietary features of these insects, or could indicate a loss of these genes across eusocial insects.

In addition to trypsin, cytochrome P450s were also found to be under-represented in both *A. cephalotes* and *A. mellifera*, relative to *N. vitripennis*, with reductions in both CYP3- and CYP4-type P450s (Table S5). P450s in insects are important enzymes known to be involved in a wide range of metabolic activities, including xenobiotic degradation, and pheromone metabolism [67]. We identified a total of 52 and 62 P450s in *A. cephalotes* and *A. mellifera*, respectively, which is similar to the low numbers reported for another insect, the body louse *Pediculus humanus* [29]. These values represent some of the smallest amounts of P450s reported for any insect genome, and may represent the minimal number of P450s required by insects to survive. Comparison of the *A. cephalotes* P450s against those of *A. mellifera* and *P. humanus* reveals that while there are some shared P450s, many are specific to each insect (Text S15).

In *A. mellifera*, the paucity of P450s is thought to be associated with the evolutionary underpinnings of its eusocial lifestyle [68], although an enrichment of P450s in the ants *C. floridanus* and *H. saltator* [27] would seem to contradict this prediction. It is therefore unclear why *A. cephalotes* has a small number of P450s relative to other ants, and future work will be necessary to provide insight into this apparent discrepancy. A SNP analysis of the P450 genes in *A. cephalotes* did reveal that one of these, *ACEP_00016463*, has 20 SNPs/kbp (Text S13). Since P450s are known to undergo accelerated duplication and divergence [67], the high number of

SNPs in this particular P450 may reflect positive selection for new functions.

Comparative Metabolic Reconstruction Analysis

Given the tight obligate association that *A. cephalotes* has with its fungal mutualist, one might predict that it acquires amino acids from its fungus in a manner similar to that of the pea aphid *Acyrtosiphon pisum*, which obtains amino acids from its bacterial symbionts [28]. To test this, we performed a metabolic reconstruction analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [69]. *A. cephalotes* contains a nearly identical set of amino acid biosynthesis genes as *A. mellifera*, *C. floridanus*, *H. saltator*, and *N. vitripennis*, all of which are incapable of synthesizing histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine *de novo*. The only exception is arginine, and only *A. cephalotes* was found to lack the genes necessary for its biosynthesis (Figure 3). Arginine, which is produced through the conversion of citrulline and aspartate [70,71], is predicted to be synthesized at levels too low to support growth in insects [72].

In *A. cephalotes* the 2 genes that catalyze the synthesis of arginine, argininosuccinate synthase (EC 6.3.4.5) and argininosuccinate lyase (EC 4.3.2.1), were not found (Figure 3). The loss of these two genes suggests a dependence on externally-acquired arginine, which we hypothesize, is provided by their fungus. In the carpenter ant *C. floridanus*, arginine is thought to be synthesized from citrulline provided by its endosymbiont *Blochmannia floridanus* [73], and this dependency is predicted to play an essential role in maintaining the carpenter ant-bacteria mutualism. An extreme case has been reported for the pea aphid, which has lost its urea pathway and depends entirely on its endosymbiont, *Buchnera aphidicola*, for arginine [28]. The loss of arginine biosynthesis in *Atta* may similarly be important for maintaining the leaf-cutter ant-fungus mutualism. In line with this prediction, the fungus the ants cultivate contains all of the amino acids that *A. cephalotes* can not synthesize, including arginine [65].

Comparison of Hexamerins

In addition to arginine biosynthesis, *A. cephalotes* may have also lost the need to rely on hexamerins as a source of amino acids during development. In many insects, hexamerin proteins are synthesized by developing larvae and used as amino acid sources during development into the adult stage [74]. Four hexamerins are commonly found across insects, including *hex 70a*, *hex 70b*, *hex 70c*, and *hex 110*. Comparison among the hymenopteran genomes reveals the presence of all hexamerins in varying copy number across all genomes except for *A. cephalotes*, which is missing *hex 70c* (Figure 4) (Text S16). In *A. mellifera*, hexamerins are expressed at different times, with *hex 70a* and *hex 110* expressed during the larval, pupal and adult stage of workers, and *hex 70b* and *hex 70c* only expressed during the larval stage [74]. The specific expression of *hex 70b* and *hex 70c* in larvae may reflect the increased need for these nutrients during early development. Given that *A. cephalotes* larvae feed primarily on gongylidia, it is possible that amino acids supplemented by the fungus over the millions of years of this mutualism has relaxed selection for maintaining larval-stage hexamerins, and thus *hex 70c* may have been lost. Future expression analyses of these genes at different life stages, in different castes, and under different nutritional conditions will likely confirm and elucidate their role.

Conclusion

Here we have presented the first genome sequence for a fungus-growing ant and show that its genomic features potentially reflect

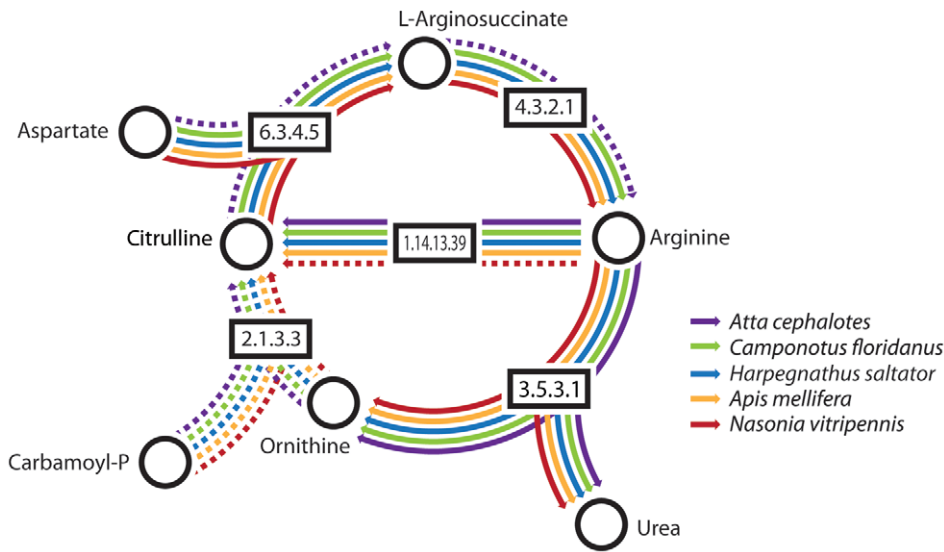


Figure 3. Predicted arginine biosynthesis pathway map in *Atta cephalotes*, *Camponotus floridanus*, *Harpegnathus saltator*, *Apis mellifera*, and *Nasonia vitripennis*. This pathway in *A. cephalotes* was found to be missing the two enzymes agininosuccinate synthase (EC 6.3.4.5) and argininosuccinate lyase (EC 4.3.2.1), which catalyzes the conversion of aspartate and citrulline into arginine. Other enzymes in this pathway include ornithine cabamoyltransferase (EC 2.1.3.3), arginase (EC 3.5.3.1) and nitric oxide synthase (EC 1.14.13.39). Dotted arrows indicate genes encoding proteins which were not found.
doi:10.1371/journal.pgen.1002007.g003

its obligate symbiotic lifestyle and developmental complexity. An initial analysis of its genome reveals many characteristics that are similar to both solitary and eusocial insect genomes. One hypothesis, based on the obligate mutualism of *Atta cephalotes* and its fungus, is that its genome exhibits reductions related to this relationship. We have provided some evidence that *A. cephalotes* has gene reductions related to nutrient acquisition, and these losses may be compensated by the provision of these nutrients from the fungus. For example, the extensive reduction in serine proteases may reflect the lack of proteins in its diet since the fungus primarily provides nutrients in the form of carbohydrates and free amino acids. Furthermore, the loss of the arginine biosynthesis pathway in *A. cephalotes* may indicate the obligate reliance that it has on the fungus, as arginine is part of the nutrients that it provides to the ant. This type of relationship

appears to be conserved in other insect-microbe mutualisms, specifically in the pea aphid [28] and the carpenter ant [73]. Finally, *A. cephalotes* appears to have lost a hexamerin protein that is conserved across all other insect genome sequences reported to date. Loss of this protein, which is associated with amino acid sequestration during larval development, may be tolerated because larvae have a ready source of amino acids from the fungus. These genomic features may serve as essential factors that have stabilized the mutualism over its coevolutionary history. The sequencing and analysis of this genome will be a valuable addition to the growing number of insect genomes, and in particular will provide insight into both host-microbe symbiosis and eusociality in hymenoptera.

Materials and Methods

Sample Collection, DNA Extraction, and Sequencing

Three males from a single mature *Atta cephalotes* colony were collected in June 2009 in Gamboa, Panama (latitude 9° 7' 0" N, longitude 79° 42' 0" W) and designated males A, B, and C. Genomic DNA from these males was extracted using a modified version of a Genomic-tip extraction protocol for mosquitoes and other insects (QIAGEN, Valencia, CA). Sequencing was performed using the 454 FLX Titanium pyrosequencing platform [25] at the 454 Life Sciences Sequencing Center (Branford, CT) as follows. A whole-genome shotgun fragment library was constructed for male A and sequenced using a single run, generating 539,113,701 bp of sequence. For male B, a whole-genome shotgun fragment library was also constructed and sequenced using 11 runs, generating a total of 4,209,396,304 bp of sequence. An 8 kbp insert paired-end library was also generated for male B and sequenced using two runs, generating a total of 818,851,400 bp of sequence. A 20 kbp paired-end library was generated for male C, and sequenced using a single run, generating 349,435,001 bp. In total, 5,916,796,406 bp of sequence were generated for all three ants.

	hex 70a	hex 70b	hex 70c	hex 110
<i>Atta cephalotes</i>	hex 70a	hex 70b		hex 110
<i>Camponotus floridanus</i>	Cflo_07104	Cflo_07103 Cflo_12692	Cflo_12693	Cflo_01253
<i>Harpegnathus saltator</i>	Hsal_02045 Hsal_02046	Hsal_02044 Hsal_02043	Hsal_02042	Hsal_0629
<i>Apis mellifera</i>	hex 70a	hex 70b	hex 70c	hex 110
<i>Nasonia vitripennis</i>	hex 81	hex 75 hex 75	NV12470 hex 83 hex 94	hex 192 hex 109

Figure 4. Distribution of hexamerin genes in the genomes of *Atta cephalotes*, *Camponotus floridanus*, *Harpegnathus saltator*, *Apis mellifera*, and *Nasonia vitripennis*. Four hexamerins with varying copy number are found within these genomes except for *A. cephalotes* which is missing hex 70c. Many of these genes are found to be syntenic along chromosomes/scaffolds, as shown (not drawn to scale).
doi:10.1371/journal.pgen.1002007.g004

Genome Assembly

All generated sequences were assembled using the 454 GS *de novo* assembler software (March 06 2010 R&D Release). The *Atta cephalotes* whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the project number 48117 and accession ADTU00000000. The version described in this paper is the first version, ADTU01000000.

Transcript Sequencing and Assembly

Workers from a healthy *Atta cephalotes* colony (JS090510-01) collected from Gamboa, Panama and maintained in the laboratory of Cameron Currie at the University of Wisconsin-Madison were used to generate transcript sequences. A pool of 169 workers across different age and size classes was selected and total RNA was extracted using a modified version of a phenol-chloroform protocol previously described [75]. This sample was normalized and a fragment library was generated before subsequent sequencing using a single run of a 454 FLX Titanium pyrosequencer [25] at the Genome Center at Washington University (St. Louis, MO), generating a total of 462,755,799 bp of sequence. Transcript sequences were assembled using the Celera assembler (wgs-assembler 6.0 beta) [76] with standard assembly parameters.

Genome Annotation

Annotations for the *Atta cephalotes* genome was generated using the automated genome annotation pipeline MAKER [32]. The MAKER annotation pipeline consists of 4 general steps. First, RepeatMasker (<http://www.repeatmasker.org>) and RepeatRunner [77] were used to identify and mask repetitive elements in the genome. Second, gene prediction programs including Augustus [78], Snap [79], and GeneMark [80] were employed to generate *ab-initio* (non-evidence informed) gene predictions. Next a set of expressed sequence tags (ESTs) and proteins from related organisms were aligned against the genome using BLASTN and BLASTX [81], and these alignments were further refined with respect to splice sites using the computer program Exonerate [82]. Finally, the EST and protein homology alignments and the *ab-initio* gene predictions were integrated and filtered by MAKER to produce a set of evidence informed gene annotations. This gene set was then further refined to remove all putative repeat elements and to include gene models initially rejected by MAKER but found to contain known protein domains using the program InterProScan [83]. The resulting gene set (OGS 1.1) then became the substrate for further analysis and manual curation. Over 500 genes in OGS 1.1 were manually curated (Table S1), producing OGS 1.2, which is publicly available at the Hymenopteran Genome Database (http://HymenopteraGenome.org/atta/genome_consortium).

The general manual curation process used for generating OGS 1.2 was based on a standardized protocol and conducted as follows. For each gene family, query sequences were obtained first from FlyBase [84] and supplemented with known gene models from the other sequenced hymenopteran genomes, *Apis mellifera* [26] and *Nasonia vitripennis* [35]. BLAST was used to align these gene models against putative sequences in the *A. cephalotes* genome predicted by MAKER. The sequence analysis program Apollo [85] was then used by all annotators to contribute their annotations to a centralized Chado [86] database. In general, putative gene models in *A. cephalotes* were confirmed by investigating the placement of introns and exons, the completeness of sequences, evaluating sequencing errors, and syntenic information. A final homology search was also performed with the putative

A. cephalotes gene model by comparing it against the non-redundant protein database in NCBI to confirm its match against known insect models.

Repeat Analysis

We used PILER-DF [87], RepeatModeler (Smit A, Hubley R, Green P. RepeatModeler Open-1.0. 2008-2010 <http://www.repeatmasker.org>), RECON [88], and RepeatScout [89] to generate *de novo* transposable element (TE) predictions. We found 1,381 *de novo* repeat predictions including 264 from RepeatScout, 26 from PILER-DF, and 1091 from RECON. We simplified the complexity of our *de novo* TE predictions by removing elements that were over 80% similar over 80% of their length [90] and also screened out elements with more than 50% sequence identity to Uniprot [91] genes. This resulted in a final *A. cephalotes*-specific repeat library containing 1,252 elements (1048 RECON, 195 RepeatScout, 9 PILER-DF), which were then classified using RepeatMasker (Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996-2010 <<http://www.repeatmasker.org>>) and custom scripts that identify TIR and LTR sequences. This curated library was converted to EMBL format, appended to a RepBase [92] library and used to mask the *A. cephalotes* genome assembly.

Orthology Analysis

An orthology analysis was performed between the proteins from *Atta cephalotes* (OGS1.2), *Apis mellifera* (preOGS2) [26], *Nasonia vitripennis* (OGSI 1.2) [35], and *Drosophila melanogaster* (Release 5.29) [93]. Using these protein sets, we reduced each dataset to contain only the single longest isoform using custom Perl scripts. An all-by-all BLAST was performed using the computer program OrthoMCL [94] and the best reciprocal orthologs, inparalogs, and co-orthologs were determined. We used the MCL v09-308 Markov clustering algorithm [95] to define final ortholog, inparalog, and co-ortholog groups between the datasets. For all OrthoMCL analyses, the suggested parameters were used.

We then annotated those proteins in *A. cephalotes* that did not have any orthologs to the 3 other insects and performed a gene ontology enrichment analysis. This was done by annotating all *A. cephalotes* proteins using Interproscan [83] to generate Gene Ontology (GO) [57] terms. This resulted in 6,971 (41%) proteins receiving at least one GO annotation. GO-TermFinder [96] was then used to determine those proteins that were enriched for specific GO terms in the *A. cephalotes*-specific proteins, relative to the entire *A. cephalotes* OGS1.2 dataset.

KOG Enrichment Analysis

We performed a eukaryotic orthologous groups (KOG) [58] enrichment analysis for the genomes of *Atta cephalotes*, *Camponotus floridanus* [27], *Harpegnathos saltator* [27], *Apis mellifera* [26], and *Nasonia vitripennis* [35]. The KOG database was obtained from NCBI and RPSBLAST [97] (e-value: 1e-05) was used to compare the predicted proteins from *A. cephalotes* (OGS1.2), *C. floridanus* (OGS3.3), *H. saltator* (OGS3.3), *A. mellifera* (preOGS2), and *Nasonia vitripennis* (OGS r.1). Each KOG hit was tabulated according to its gene category, and Fisher's exact test was then applied to determine which categories were over- or under-represented. This was done for *A. cephalotes*, *C. floridanus*, *H. saltator*, and *A. mellifera* against *N. vitripennis*, respectively, as shown in Table S4. We then determine for each over- and under-represented KOG category in *A. cephalotes* relative to *N. vitripennis*, the specific KOGs within each category that were significantly enriched or under-enriched. This was done by comparing the

total number of *A. cephalotes* KOGs within each of these categories against those in *N. vitripennis* using Fisher's exact test, as shown in Table S5.

KEGG Reconstruction Analysis

The predicted peptides for *Atta cephalotes* were used to reconstruct putative metabolic pathways using the Kyoto Encyclopedia of Genes and Genomes [69]. This was performed using the KEGG Automated Annotation Server (KAAS), which annotates proteins according to the KEGG database and reconstructs full pathways displaying them as maps. Similar maps were also constructed using KAAS for the predicted peptide sequences of *Camponotus floridanus* (OGS3.3) and *Harpegnathos saltator* (OGS3.3). These maps were compared against the maps currently available in KEGG for *Apis mellifera*, *Drosophila melanogaster*, and *Nasonia vitripennis*. For proteins in *A. cephalotes* that were not found in our KEGG reconstruction analysis, relative to other insects (e.g. argininosuccinate synthase (EC 6.3.4.5) and argininosuccinate lyase (EC 4.3.2.1)), we investigated those reads that were not incorporated into the *A. cephalotes* assembly to confirm that these did not contain potential gene fragments corresponding to these genes.

Supporting Information

Table S1 Total number of genes annotated in the *A. cephalotes* genome according to gene family. (DOC)

Table S2 Repetitive elements identified in the *A. cephalotes* genome. (DOC)

Table S3 Gene Ontology enrichment of proteins specific to *A. cephalotes* relative to *A. mellifera*, *D. melanogaster*, and *N. vitripennis*. (DOC)

Table S4 Enrichment comparison of proteins in categories of KOGs for *Atta cephalotes*, *Camponotus floridanus*, *Harpegnathos saltator*, and *Apis mellifera* relative to *Nasonia vitripennis*. (DOC)

Table S5 Over- and under-represented KOGs in *Atta cephalotes* (Acep), *Camponotus floridanus* (Cflo), *Harpegnathos saltator* (Hsal), and *Apis mellifera* (Amel) compared to *Nasonia vitripennis* (Nvit) according to category. (DOC)

Text S1 Cytoplasmic Ribosomal Proteins. (DOC)

Text S2 Evaluating Sequence Quality and Coverage using Cytoplasmic Ribosomal Proteins. (DOC)

Text S3 Oxidative Phosphorylation Proteins. (DOC)

Text S4 The *Atta cephalotes* Mitochondrial Genome. (DOC)

Text S5 Global Composition Analysis. (DOC)

Text S6 DNA Methylation Tool Kit. (DOC)

Text S7 RNAi Genes. (DOC)

Text S8 Insulin Signalling Pathway Genes. (DOC)

Text S9 Yellow/Major Royal Jelly Protein Family. (DOC)

Text S10 Wing Polyphenism Gene Networks. (DOC)

Text S11 The $\Delta 9$ Desaturase Genes. (DOC)

Text S12 Immune Genes. (DOC)

Text S13 SNP Analysis. (DOC)

Text S14 Protein Domain Contraction and Expansion. (DOC)

Text S15 Cytochrome p450s. (DOC)

Text S16 Hexamerins. (DOC)

Text S17 Alien Gene Fragments. (DOC)

Text S18 Chitinase Genes. (DOC)

Text S19 Hox Genes. (DOC)

Text S20 Micro RNAs. (DOC)

Acknowledgments

We would like to thank the sequencing and production team at the 454 Sequencing Center for their help and expertise during this project; P. Nangle and S. M. Adams for facilitating conference call logistics throughout this project; P. Minx, S. Hou, L. Ye, and R. K. Wilson for transcript assembly and genome sequence deposition assistance; the Smithsonian Tropical Research Institute in Panama for logistical support during sample collection, especially M. Paz, O. Arosemena, Y. Clemons, L. Seid, and R. Urriola for housing access, permit acquisition, and laboratory assistance; G. Starrett for technical assistance with figure generation; and all members of the Currie Lab for their critical reading of this manuscript, their encouragement, and their support.

Author Contributions

Conceived and designed the experiments: GS LL CT TTHJT SCS SWC WCW CGE GMW NMG CRC. Performed the experiments: CT LL PB EJC SEM WCW. Analyzed the data: GS LL CH EA EBB EJC EC AC EE HH MJF JG JDG DG KJG DEH MH CH HH BRJ JK SEM JAM MCM MCMT MCN SN RO RR JTR CRS ST NDT LW MDY LV FZ CGE CDS NMG CRC. Contributed reagents/materials/analysis tools: CT LL CH PB HH JJS MDY TTH SWC WCW CGE CDS GMW. Wrote the paper: GS. Project Management: GS CT TTHJT SCS SWC CGE GMW NMG CRC.

References

- Hölldobler B, Wilson EO (1990) The ants. CambridgeMass: Harvard University Press. 732 p.
- Belt T (1874) The naturalist in Nicaragua. London: E. Bumpus. 306 p.
- Weber NA (1966) Fungus-growing ants. Science 153: 587–604.
- Schultz TR, Brady SG (2008) Major evolutionary transitions in ant agriculture. Proc Natl Acad Sci U S A 105: 5435.
- Fowler G, Forri L (1986) Economics of grass-cutting ants. In: Lofgren C, Vander Meer R, eds. Fire ants and leaf cutting ants: biology and management. BoulderCO: Westview Press. pp 123–145.
- Hölldobler B, Wilson EO (2008) The superorganism: the beauty, elegance, and strangeness of insect societies. New York: W.W. Norton. 544 p.
- Wirth R, Herz H, Ryel RJ, Beyschlag W, Hölldobler B (2003) Herbivory of leaf-cutting ants. A case study on *Atta colombica* in the tropical rain forest of Panama. Berlin, Heidelberg: Springer. 230 p.
- Currie CR (2001) A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. Annu Rev Microbiol 55: 357–380.
- Mueller UG, Schultz TR, Cameron RC, Adams RMM, Malloch D (2001) The origin of the attine ant-fungus mutualism. Q Rev Biol 76: 169–197.
- Pinto-Tomas AA, Anderson MA, Suen G, Stevenson DM, Chu FST, et al. (2009) Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science 326: 1120–1123.
- Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, et al. (2010) An insect herbivore microbiome with high plant biomass-degrading capacity. PLoS Genet 6: e1001129. doi:10.1371/journal.pgen.1001129.
- Currie CR, Wong B, Stuart AE, Schultz TR, Rehner SA, et al. (2003) Ancient tripartite coevolution in the attine ant-microbe symbiosis. Science 299: 386–388.
- Currie CR (2001) Prevalence and impact of a virulent parasite on a tripartite mutualism. Oecologia 128: 99–106.
- Currie CR, Scott JA, Summerbell RC, Malloch D (1999) Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. Nature 398: 701–704.
- Currie CR, Bot ANM, Boomsma JJ (2003) Experimental evidence of a tripartite mutualism: bacteria protect ant fungus gardens from specialized parasites. Oikos 101: 91–102.
- Oh D-C, Poulsen M, Currie CR, Clardy J (2009) Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. Nat Chem Biol 5: 391–393.
- Little AE, Currie CR (2007) Symbiotic complexity: discovery of a fifth symbiont in the attine ant-microbe symbiosis. Biol Lett 3: 501–504.
- Little AE, Currie CR (2008) Black yeast symbionts compromise the efficiency of antibiotic defenses in fungus-growing ants. Ecology 89: 1216–1222.
- Frost CL, Fernandez-Marin H, Smith JE, Hughes WOH (2010) Multiple gains and losses of *Wolbachia* symbionts across a tribe of fungus-growing ants. Mol Ecol 19: 4077–4085.
- Pagnocca FC, Legaspe MFC, Rodrigues A, Ruivo CCC, Nagamoto NS, et al. (2010) Yeasts isolated from a fungus-growing ant nest, including the description of *Trichosporon chirellii* sp. nov., an anamorphic basidiomycetous yeast. Int J Syst Evol Microbiol 60: 1454–1459.
- Haeder S, Wirth R, Herz H, Spitter D (2009) Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Excoelopsis*. Proc Natl Acad Sci U S A 106: 4742–4746.
- Bacci M, Jr., Ribeiro SB, Casarotto MEF, Pagnocca FC (1995) Biopolymer-degrading bacteria from nests of the leaf-cutting ant *Atta sexdens rubropilosa*. Braz J Med Biol Res 28: 79–82.
- Wilson EO (1983) Caste and division of labor in leaf-cutter ants (Hymenoptera: Formicidae: Atta): IV. colony ontogeny of *A. cephalotes*. Behav Ecol Sociobiol 14: 55–60.
- Wilson EO (1983) Caste and division of labor in leaf-cutter ants (Hymenoptera: Formicidae: Atta): III. ergonomic resiliency in foraging by *A. cephalotes*. Behav Ecol Sociobiol 14: 47–54.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.
- Honey Bee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443: 931–949.
- Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, et al. (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329: 1068–1071.
- The International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. PLoS Biol 8: e1000313. doi:10.1371/journal.pbio.1000313.
- Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, et al. (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A 107: 12168–12173.
- Tsutsui N, Suarez A, Spagna J, Johnston JS (2008) The evolution of genome size in ants. BMC Evol Biol 8: 64.
- Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061–1067.
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18: 188–196.
- Wei S, Shi M, He J, Sharkey M, Chen X (2009) The complete mitochondrial genome of *Diadegma semiclausum* (Hymenoptera: Ichneumonidae) indicates extensive independent evolutionary events. Genome 52: 308–319.
- Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics 133: 97–117.
- Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, et al. (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327: 343–348.
- Karlin S, Mrazek J (1997) Compositional differences within and between eukaryotic genomes. Proc Natl Acad Sci U S A 94: 10227–10232.
- Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, et al. (2006) Functional CpG methylation system in a social insect. Science 314: 645–647.
- Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A 107: 8689–8694.
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916–919.
- Field LM, Lyko F, Mandrioli M, Pranter G (2004) DNA methylation in insects. Insect Mol Biol 13: 109–115.
- Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. Science 319: 1827–1830.
- Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. Nature 431: 343–349.
- Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, et al. (2005) Normal microRNA maturation and germ-line stem cell maintenance requires loquacious, a double-stranded RNA-binding domain protein. PLoS Biol 3: e236. doi:10.1371/journal.pbio.0030236.
- Wu Q, Brown MR (2006) Signaling and function of insulin-like peptides in insects. Annu Rev Entomol 51: 1–24.
- de Azevedo SV, Hartfelder K (2008) The insulin signaling pathway in honey bee (*Apis mellifera*) caste development - differential expression of insulin-like peptides and insulin receptors in queen and worker larvae. J Insect Physiol 54: 1064–1071.
- Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R (2006) Evolution of the yellow/major royal jelly protein family and the emergence of social behavior in honey bees. Genome Res 16: 1385–1394.
- Abouheif E, Wray GA (2002) Evolution of the gene network underlying wing polyphenism in ants. Science 297: 249–252.
- Elango N, Hunt BG, Goodisman MAD, Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc Natl Acad Sci U S A 106: 11206–11211.
- Illingworth R, Kerr A, DeSousa D, Jurgensen H, Ellis P, et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol 6: e22. doi:10.1371/journal.pbio.0060022.
- Wagner D, Tissot M, Cuevas W, Gordon DM (2000) Harvester ants utilize cuticular hydrocarbons in nestmate recognition. J Chem Ecol 26: 2245–2257.
- Martin S, Drijfhout F (2009) A review of ant cuticular hydrocarbons. J Chem Ecol 35: 1151–1161.
- Vilmos P, Kurucz É (1998) Insect immunity: evolutionary roots of the mammalian innate immune system. Immunol Lett 62: 59–66.
- Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, et al. (2006) Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Mol Biol 15: 645–656.
- Do Nascimento RR, Schoeters E, Morgan ED, Billen J, Stradling DJ (1996) Chemistry of metapleural gland secretions of three attine ants, *Atta sexdens rubropilosa*, *Atta cephalotes*, and *Acromyrmex octospinosus* (Hymenoptera: Formicidae). J Chem Ecol 22: 987–1000.
- Poulsen M, Hughes WOH, Boomsma JJ (2006) Differential resistance and the importance of antibiotic production in *Acromyrmex echinatior* leaf-cutting ant castes towards the entomopathogenic fungus *Aspergillus nomius*. Insect Soc 53: 349–355.
- Fernandez-Marin H, Zimmerman JK, Rehner SA, Weislo WT (2006) Active use of the metapleural glands by ants in controlling fungal infection. Proc R Soc London Ser B Biol Sci 273: 1689–1695.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25: 25.
- Tatusov R, Fedorova N, Jackson J, Jacobs A, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.
- Wasserman S (1998) FH proteins as cytoskeletal organizers. Trends in Cell Biol 8: 111–115.
- Smith CR, Toth AL, Suarez AV, Robinson GE (2008) Genetic and genomic analyses of the division of labour in insect societies. Nat Rev Genet 9: 735–748.
- Beye M, Hasselmann M, Fondrk MK, Page Jr. RE, Omholt SW (2003) The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114: 419–429.
- Blake AD, Anthony NM, Chen HH, Harrison JB, Nathanson NM, et al. (1993) *Drosophila* nervous system muscarinic acetylcholine receptor: transient functional expression and localization by immunocytochemistry. Mol Pharmacol 44: 716–724.
- Becker MN, Brenner R, Atkinson NS (1995) Tissue-specific expression of a *Drosophila* calcium-activated potassium channel. J Neurosci 15: 6250–6259.

64. De Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, et al. (2010) Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. *Insect Mol Biol* 19: 11–26.
65. Martín MM, Carman RM, Macconnell JG (1969) Nutrients derived from the fungus cultured by the fungus-growing ant *Atta colombica* tansipes. *Ann Ent Soc Am* 62: 11–13.
66. Haydak MH (1970) Honey bee nutrition. *Ann Rev Entomol* 15: 143–156.
67. Feyereisen R (1999) Insect P450 enzymes. *Annu Rev Entomol* 44: 507–533.
68. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, et al. (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol* 15: 615–636.
69. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2009) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38: D355–D360.
70. Hinton T, Noyes DT, Ellis J (1951) Amino acids and growth factors in a chemically defined medium for *Drosophila*. *Physiol Zool* 24: 335–353.
71. Reddy SRR, Campbell JW (1969) Arginine metabolism in insects: properties of insect fat body arginase. *Comp Biochem Physiol* 28: 515–534.
72. Raghupathi Rami Reddy S, Campbell J (1977) Enzymic basis for the nutritional requirement of arginine in insects. *Cell Mol Life Sci* 33: 160–161.
73. Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, et al. (2003) The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc Natl Acad Sci U S A* 100: 9388–9393.
74. Martins J, Nunes F, Cristino A, Simoes Z, Bitondi M (2010) The four hexamerin genes in the honey bee: structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. *BMC Mol Biol* 11: 23.
75. Chen J, Weimer PJ (2001) Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. *Microbiology* 147: 21–30.
76. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24: 2818–2824.
77. Smith CD, Edgar RC, Yandell MD, Smith DR, Celniker SE, et al. (2007) Improved repeat identification and masking in Dipterans. *Gene* 389: 1–9.
78. Stanke M, Tzvetkova A, Morgenstern B (2006) AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* 7: S11.
79. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5: 59.
80. Lukashin AV, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 26: 1107–1115.
81. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
82. Slater G, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.
83. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* 37: D211–D215.
84. Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, et al. (2009) FlyBase: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res* 37: D555–D559.
85. Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, et al. (2002) Apollo: a sequence annotation editor. *Genome Biol* 3: RESEARCH0082.
86. Mungall CJ, Emmert DB, The FlyBase Consortium (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics* 23: i337–i346.
87. Edgar RC, Myers EW (2005) PILER: identification and classification of genomic repeats. *Bioinformatics* 21: i152–i158.
88. Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* 12: 1269–1276.
89. Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21: i351–i358.
90. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8: 973.
91. The UniProt Consortium (2010) The universal protein resource (UniProt) in 2010. *Nucleic Acids Res* 38: D142–D148.
92. Jurka J, Kapitonov V, Pavlicek A, Klonowski P, Kohany O, et al. (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110: 462–467.
93. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185–2195.
94. Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13: 2178–2189.
95. van Dongen S (2000) Graph clustering by flow simulation: Phd Thesis, Universiteit Utrecht.
96. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, et al. (2004) GO::TermFinder - open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. *Bioinformatics* 20: 3710–3715.
97. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, et al. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 35: D237–D240.