

Impact of host evolutionary history on endosymbiont genome evolution: a test in *Camponotus* carpenter ants and their *Blochmannia* endosymbionts

Joseph Manthey¹, Jennifer Giron², and Jack Hruska¹

¹Texas Tech University

²Purdue University

December 2, 2020

Abstract

Mutualism discernibly connects the evolution of two or more interacting taxa. Endosymbioses, especially those that are obligate, are an intimate mutualism that link the evolution of host and endosymbiont. In these instances, we may expect codiversification of hosts and endosymbionts as well as host demography discernibly affecting the course of endosymbiont evolution. While many studies have demonstrated cospeciation of hosts and endosymbionts, detailed investigations of the impact of host demography on endosymbiont molecular evolution are generally lacking. Here, we sequenced complete genomes of carpenter ants (Genus *Camponotus*) and their *Blochmannia* endosymbionts to investigate their codiversification and test hypotheses about how host demography impacts molecular evolution in endosymbionts. Using whole genome phylogenomics, we identified strong signatures of codiversification between carpenter ants and their endosymbionts. We found that endosymbiont genes have evolved rapidly, at a pace of ~30x that of their hosts. Using multiple tests for selection in *Blochmannia* genes, we found signatures of positive selection and shifts in selection strength across the phylogeny. We identified a positive relationship between host demography and shifts toward intensified selection in endosymbiont genes, but no relationship between host demography and shifts toward relaxed selection in endosymbiont genes. About 10% of the *Blochmannia* genes exhibited variable presence and absence across endosymbiont genomes. Of those, about half exhibited phylogenetic signal, indicating somewhat random patterns of gene loss in endosymbiont genomes. Lastly, we found no relationship between host demography and endosymbiont gene loss.

INTRODUCTION

Coevolution, the evolutionary change in one population in response to evolutionary change in another population (Janzen 1980), will generally require reciprocal adaptations between partner species (Ehrlich & Raven 1964). Mutualistic evolution may result in codiversification between partners, such as between flowering plants and their pollinators or between endosymbiotic microorganisms and their hosts (Clark *et al.* 2000; Kiestler *et al.* 1984). Endosymbioses are common in insects, may be intra- or extracellular, and the mutualisms or commensalisms range in dependence from facultative to obligate (Kikuchi 2009). Obligate endosymbioses in insects are particularly common and are often implicated in host nutrition, resistance to pathogens, and evolution in general (Anbutsu *et al.* 2017; Brownlie & Johnson 2009; Moreau 2020; Perlmutter & Bordenstein 2020).

Obligate endosymbionts generally have reduced, but not entirely degraded genome sizes (Chong *et al.* 2019; Gil *et al.* 2003; Wernegreen 2002), suggesting they retain some genes for their own function as well as those necessary for their continued retention as symbionts with their respective hosts. As such, we may expect patterns of molecular evolution in endosymbionts to be partially shaped by their hosts' evolutionary and demographic histories. First, selection in endosymbiont genomes will be partially affected by host-level selection, because endosymbiont fitness is partially linked with host fitness (Wernegreen 2002). Second, host

and endosymbiont effective population sizes will be intrinsically linked; as such, the potential strength of selection and efficacy of genetic drift in endosymbiont genomes should both be influenced by host demographic history. Third, if endosymbionts are vertically transmitted, we would expect codiversification of hosts and their endosymbionts. Indeed, there is evidence in several insect clades for codiversification of hosts and endosymbionts (e.g., Chen *et al.* 1999; Clark *et al.* 2001; Clark *et al.* 2000; Distel *et al.* 1994; Lo *et al.* 2003). Despite a relative plethora of evidence for host-endosymbiont codiversification, the effects of insect host demography on endosymbiont molecular evolution is largely unexplored.

The symbiotic relationship between *Camponotus* carpenter ants and their *Blochmannia* bacterial endosymbionts is an ideal system for investigating coevolution of hosts and endosymbionts. *Camponotus* ants are common in woodlands across most of the world and the genus is the most speciose of ant genera (Wilson 1976). Camponotine ants have maintained a relationship with *Blochmannia* for about 40 million years (Wernegreen *et al.* 2009), and this relationship has likely persisted because *Blochmannia* provide essential amino acids to their ant hosts (Feldhaar *et al.* 2007; Russell *et al.* 2017). Consistent with a long-term endosymbiosis, there is some evidence that *Camponotus* and *Blochmannia* have histories of cospeciation (Degnan *et al.* 2004). There have been a few studies to date examining whole *Blochmannia* genomic sequences; in *Blochmannia vafer*, there is some evidence of ongoing purifying selection (Williams & Wernegreen 2012), and in comparative analyses across three *Blochmannia* genomes, gene loss patterns differ across lineages, suggestive of differential, host-specific selective pressures in different host lineages (Williams & Wernegreen 2015).

To investigate coevolutionary patterns between *Camponotus* and *Blochmannia*, we assembled a *de novo* *Camponotus* genome and several *de novo* *Blochmannia* genomes, as well as resequenced genomes for several *Camponotus* individuals to address the following questions and hypotheses.

Do hosts and endosymbionts exhibit strict phylogenomic co-evolutionary histories?

H₀: Largely congruent host-endosymbiont phylogenomic patterns. H_A: Evidence of endosymbiont horizontal transfer (e.g., through host species' hybridization) with conflicting host and endosymbiont phylogenomic histories.

1. How fast do *Blochmannia* genes evolve and is there rate variation across the genome? H₀: Fast rates of *Blochmannia* evolution with similar rates across the genome. H_A: Fast rates of *Blochmannia* evolution with rate variation across the genome indicative of selection.
2. Does host demography shape natural selection strength in endosymbiont genomes? H₀: No relationship between host demography and natural selection in endosymbiont genomes. H_A: Host demography influences patterns of positive selection and / or shifts in selection strength across the endosymbionts' phylogeny.
3. Does host demography influence rates of gene loss in endosymbiont genomes? H₀: No relationship between host demography and endosymbiont gene loss. H_A: Hosts with smaller population sizes will exhibit faster endosymbiont gene loss.

METHODS

Field work

We collected *Camponotus* ant specimens for this study in summer 2018 from Arizona, Colorado, and Utah, USA. We actively searched for *Camponotus* colonies by searching for woody debris during the day and searching for *Camponotus* activity during the night. Specimens for genomics were placed in cryotubes and frozen with liquid nitrogen in the field. Specimens for long-term museum storage (e.g., for photography and morphometrics) were preserved in 100% ethanol. In total, we used individuals from 17 colonies for sequencing in this study (Table S1). Specimens were identified using keys (Mackay 2019) or comparison with available sequences on NCBI's GenBank. All specimen collection numbers in the supplementary table and throughout the manuscript are associated with voucher specimens housed in the Invertebrate Zoology Collection of the Natural Science Research Laboratory, Museum of Texas Tech University.

Camponotus de novo genome assembly and annotation

Genome sample. We used multiple ($N = 4$) individuals from one colony for *de novo* assembly of a high-quality reference genome (Table S1). The only available genome of a *Camponotus* species to date is that of *Camponotus (Myrmothrix) floridanus*. We decided to assemble the genome of a member of the subgenus *Camponotus*, which are more common in western USA. For the reference genome colony, we chose a species that we were currently unable to identify. This sample is either a new species to science or a color polymorphism of other *Camponotus* (perhaps *C. novaeboracensis*?) generally found in northern or eastern North America. We refrain from naming this reference genome colony until we complete a genomic and morphological analysis in the future that includes all *Camponotus* species in the *Camponotus* subgenus described from the USA and Canada. Hereafter in the manuscript, we refer to this species with the code name *Camponotus* sp. (1-JDM).

Extraction & Sequencing. We used two sequencing methods for genome assembly: (1) long reads with Pacific Biosciences (PacBio) sequencing, and (2) a Hi-C library sequenced with Illumina technology. For the PacBio sequencing, we used services of RTL Genomics (Lubbock, TX, USA). They performed a high molecular weight DNA extraction using Qiagen’s (Hilden, Germany) MagAttract HMW DNA Kit. A single major worker, with legs and gaster removed, was used for the extraction. The extracted DNA was then used for PacBio SMRTbell library preparation, size selection using a Blue Pippin (Sage Science), and sequencing on four PacBio Sequel SMRTcells 1M v3 with Sequencing 3.0 reagents. We used the services of the Texas A&M University Core Facility to prepare a Hi-C library. They used a single major worker as input for the Arima Genomics Hi-C kit (San Diego, CA, USA). The Hi-C library was then sequenced on a partial lane of an Illumina NovaSeq S1 flow cell at the Texas Tech University Center for Biotechnology and Genomics.

Assembly. We assembled the *Camponotus* sp. (1-JDM) genome in two stages. First, we used Canu v1.9 (Koren *et al.* 2017) *de novo* assemble the PacBio long reads. Second, we used the Hi-C sequence data to scaffold the initial assembly using the 3D-DNA pipeline (Dudchenko *et al.* 2017; Durand *et al.* 2016). All commands for the assembly, annotation, and all further analyses are documented on GitHub (github.com/jdmanthey/camponotus_genomes1). We quality checked the genome assembly for potential contamination with BlobTools v.1.0.1 (DOI:10.5281/zenodo.845347; Laetsch & Blaxter 2017). Briefly, BlobTools attempts to identify contamination through taxonomic annotation and coverage parsing of resequencing data to the reference genome. With the use of BlobTools, we identified ~37kbp of potential contamination attributed to chordates or molluscs that we subsequently removed from our assembly (Fig. S1).

Repetitive Element Annotation. We used a multi-step process to annotate transposable elements (TEs) and repetitive elements in the *Camponotus* sp. (1-JDM) genome: (1) identify *de novo* repeats and over-represented sequences, (2) manually curate repetitive elements, and (3) mask the genome with these elements to create a TE and repetitive element summary file. First, we used RepeatModeler’s (v1.0.11; Smit & Hubley 2008) implementations of RepeatScout, RECON, and Tandem Repeats Finder to identify repeats based on homology, structure, and repetitiveness in the *de novo* assembly (Bao & Eddy 2002; Benson 1999; Price *et al.* 2005). We refined the RepeatModeler output by filtering matches to closely related sequences in the RepBase invertebrate database v24.03 (Jurka *et al.* 2005) and then creating consensus sequences of novel repetitive elements.

First, we removed any RepeatModeler output sequences [?] 98% identical to RepBase sequences. Second, we used BLAST and bedtools (Camacho *et al.* 2009; Quinlan & Hall 2010) to extract genomic regions matching repetitive elements as well as 1000 bp flanking sequences. We used these extracted sequences to develop consensus sequences for novel TEs using the following steps: (1) alignment using MAFFT (Katoh & Standley 2013) implemented in Geneious (BioMatters Ltd), (2) 50% majority consensus sequences in Geneious, and (3) trimming any ambiguous nucleotides on the ends of newly created consensus sequences. For any incomplete consensus sequences where we did not recover TE endpoints, we repeated this prior process up to two times. In addition to identifying *de novo* repeats and manual curation in the *Camponotus* sp. (1-JDM) genome, we also repeated this process for the published *Formica selysi* genome (NCBI: GCA_009859135.1; Brelsford *et al.* 2020). We added this species to increase the diversity of ant TEs in our database, which has been shown to improve annotations by including TEs that may have been missed in other curated species (Boman *et al.*

2019). To help with naming some of our *de novo* TEs, we assessed homology of newly curated sequences to the invertebrate RepBase database using BLAST. Lastly, we used a combination TE library including the RepBase invertebrate database and all newly curated TEs described here for use in RepeatMasker v4.08 (Smit *et al.* 2015). Repeatmasker output included a masked genome and summarized repetitive and TE content in the *Camponotus* sp. (1-JDM) genome.

Gene Annotation. To annotate genes in the *Camponotus* sp. (1-JDM) genome, we used the MAKER v2.31.10 pipeline (Cantarel *et al.* 2008). First, we used MAKER to predict genes using proteins from other ant species: *Camponotus floridanus* (GCF_003227725.1), *Formica exsecta* (GCF_003651465.1), *Lasius niger* (GCA_001045655.1), and *Nylanderia fulva* (GCF_005281655.1) (Bonasio *et al.* 2010; Dhaygude *et al.* 2019). We used these initial MAKER predictions to train SNAP and Augustus (Korf 2004; Stanke & Waack 2003). Lastly, we used the models trained in SNAP and Augustus in a second iteration of MAKER to predict gene models in the *Camponotus* sp. (1-JDM) genome. We used BUSCO v3 (Simão *et al.* 2015) with the hymenoptera single orthologous gene set (set: odb9) to assess genome assembly completeness.

Camponotus molecular clock

We extracted the putative *Camponotus* sp. (1-JDM) coding sequence (CDS) from the assembly using the MAKER output and bedtools. We downloaded the CDS sequences for *Formica exsecta*, *Lasius niger*, and *Nylanderia fulva* (same versions as proteins) for homology-based comparisons. We performed a reciprocal BLAST of all species versus *Camponotus* sp. (1-JDM) using blastn (Camacho *et al.* 2009) to identify putative homologues across datasets.

To align putative homologues between the four ant species, we used T-Coffee (Notredame *et al.* 2000). T-Coffee translates nucleotide sequences, aligns them using several alignment algorithms, takes the averaged best alignment of all alignments, and back translates the protein alignments to provide a nucleotide alignment for each gene. Before the final back-translating, we used trimAl (Capella-Gutiérrez *et al.* 2009) to remove gaps in the protein alignments.

We tested each gene for selection using gene-wide and branch-specific tests for selection in CODEML (Yang 1997). After correcting significance values for multiple testing using the Benjamini and Hochberg (1995) method, we removed any alignments with evidence for selection. We then extracted and concatenated four-fold degenerate sites from the alignments (N = 806,844) using custom R scripts and the R packages Biostrings and seqinr (Charif & Lobry 2007; Pagès *et al.* 2017). With this alignment of four-fold degenerate sites, we identified an appropriate model of sequence evolution using jModelTest (Darriba *et al.* 2012) and used the GTR + I model of sequence evolution in PhyML (Guindon *et al.* 2010) to estimate a phylogenetic tree.

To put the evolution of the CDS four-fold degenerate sites in a timed evolutionary context, we downloaded a recent phylogenomic tree of formicine ants (Blaimer *et al.* 2015) and pruned the tree to the four representative lineages covered by our CDS downloads and novel assembly using the R package ape (Paradis *et al.* 2004) using four species as representatives of those lineages: *Camponotus hyatti*, *Nylanderia dodo*, *Formica neogagates*, and *Lasius niger*. We used the *Camponotus*-specific branch length of the four-fold degenerate sites tree along with divergence time estimates from Blaimer *et al.* (2015) to obtain an estimate of *Camponotus*-specific mutation rates.

Resequencing genomes of multiple Camponotus species

Lab work. We aimed to resequence genomes for 17 *Camponotus* individuals from 7 species (Table S1) at moderate sequencing coverage (e.g., ~10–30x). Each individual is a single representative of many ants collected per each colony. For each individual, we performed two DNA extractions: (1) gaster, and (2) head + mesosoma. We did this because the gaster has a plethora of *Blochmannia* DNA relative to ant DNA. For each extraction, we froze the sample with liquid nitrogen and subsequently pulverized the sample with a sterile mortar and pestle. We then used the pulverized material as the input for DNA extraction with QIAGEN (Hilden, Germany) DNeasy blood and tissue kits. We quantified DNA concentrations from the extractions with Invitrogen (Carlsbad, California) Qubit fluorescent quantitation, and pooled the head +

mesosoma and gaster extracts at a 0.8:0.2 ratio, respectively, so as to have good representation of both ant and endosymbiont DNA for sequencing. Genomic DNA extractions were sent to the Texas Tech University Center for Biotechnology and Genomics for standard Illumina shotgun sequencing library creation and subsequent sequencing on a partial lane of an S4 flow cell on the Illumina NovaSeq6000.

Filtering and Genotyping. First, we downloaded Illumina sequencing reads from the NCBI SRA of two published datasets to use as outgroups: *Camponotus floridanus* (SRX022802) and *Cataglyphis niger* (SRX5650044). With our newly generated data and the downloaded data,

we trimmed adapters and quality filtered the raw sequencing data using the `bbduk.sh` script of the `bbmap` package (Bushnell 2014). We then aligned the filtered data to the *de novo* *Camponotus* sp. (1-JDM) reference genome with BWA (Li & Durbin 2009) using the `BWA-MEM` command. We used `samtools` v1.4.1 (Li, et al. 2009) to convert the BWA output SAM file to BAM format, and lastly cleaned, sorted, added read groups to, and removed duplicates from each BAM file using the Genome Analysis Toolkit (GATK) v4.1.0.0 (McKenna *et al.* 2010). We used GATK’s functions `HaplotypeCaller` and `GenotypeGVCFs` to genotype all individuals for both variant and invariant sites on all scaffolds at least two Mbp in length. We measured the distribution of sequencing coverage using the `samtools` ‘depth’ command (Fig. S2). We used `VCFtools` v0.1.14 (Danecek, et al. 2011) to initially filter all variant and invariant site calls using the following restrictions: (1) genotyped in [?] 70% of individuals, (2) minimum site quality of 20, (3) minimum genotype quality of 20, (4) minimum depth of coverage of 5, and (5) maximum mean depth of coverage of 70.

Blochmannia genome assemblies and annotation

Assembly. With the raw sequencing data, we used the MinYS pipeline (Guyomar *et al.* 2020) to assemble *Blochmannia* genomes for each sample. MinYS used samples mixed with host and bacterial DNA in a pipeline that allows targeted assembly of bacterial genomes. First, it maps metagenomic reads to a reference genome using BWA. Here, we used a *Blochmannia pennsylvanicus* genome (NC_007292.1) as our target. Next, the pipeline assembles these recruited reads using the program `Minia` (github.com/GATB/minia), followed by gapfilling the contigs using the program `MindTheGap` (Rizket *et al.* 2014). Finally, the pipeline simplifies the graphical fragment assembly (GFA) output of `MindTheGap`. The resulting GFA output was then visualized in `Bandage` (see Fig. S3; Wick *et al.* 2015), regions with multiple paths merged by coverage, and output in FASTA format. This process assembled a circular genome for each of the samples in our study. We also downloaded the sequence and annotation of the *Blochmannia* endosymbiont of *Camponotus floridanus* for use as an outgroup (NC_005061.1).

Annotation. We used the NCBI prokaryotic genome annotation pipeline (Tatusova *et al.* 2016) to annotate genes in each of the *Blochmannia* genomes.

Camponotus phylogenomics and population genomics

Phylogenomics. We estimated “gene trees” for non-overlapping 50 kbp sliding windows using `RAxML` v8.2.12 (Stamatakis 2014) with the `GTRGAMMA` model of sequence evolution. We summarized the phylogenies ($n = 4784$). From these gene trees, we estimated a species tree using two methods: (1) maximum clade credibility tree of all input trees using `DendroPy` (Sukumaran & Holder 2010), and (2) the coalescent-based species tree approach `ASTRAL III` (Zhang *et al.* 2018).

Genetic diversity and demography. We estimated genetic diversity for each individual in two ways. First, we estimated observed heterozygosity simply as the proportion of bi-allelic to invariant sites for each individual for all genotyped sites. Because sequencing depth has the potential to impact estimates of genetic diversity, we also used the program `ROHan` (Renaud *et al.* 2019). `ROHan` uses a Bayesian framework to estimate rates of heterozygosity while accounting for sequencing depth and per-base quality scores. We found the two estimates to be highly correlated in ingroup samples ($R^2 = 0.7113$, $p << 0.001$), so we consider only raw estimates of heterozygosity hereafter.

To estimate demography for each individual, we used the program `MSMC2` v1.1.0 (Schiffels & Durbin 2014). For use in `MSMC`, we masked genomic regions not genotyped, as these would otherwise be mistaken for

runs of homozygosity. It is relevant to note that MSMC estimates are accurate in panmictic populations, but population structure or changes in connectivity between populations through time may mimic changes in population sizes (Chikhi *et al.* 2018; Mazet *et al.* 2016). Because of this, some caution should be used when interpreting raw demographic history results. We largely used the demographic histories to estimate harmonic mean population sizes over the past half million years, which is highly correlated with observed heterozygosity ($R^2 = 0.7497$, $p \ll 0.001$). When running MSMC, we allowed up to 20 iterations and up to 23 inferred distinct time segments. We performed bootstrap replicates for each individual to see how signal could vary using different genomic regions. For this, we bootstrapped 1 Mbp segments of the genomes, with a total of ten bootstrap replicates. We decided to use MSMC for each individual rather than aggregating samples per species because of several reasons: (1) uncertainty of population structure amongst sampling locations, (2) uneven sampling sizes per species, and (3) low certainty with phasing necessary to run the program with multiple individuals, again because of the small sample sizes per species.

MSMC output is presented in relative to a species' generation time and mutation rate. Because there are not good estimates of generation times in *Camponotus* ants, we used a conservative proxy generation time used in other studies: double the age of sexual maturity (Nadachowska-Brzyska *et al.* 2015). In *Camponotus*, previous studies have suggested that the earliest age of queens producing winged reproductives is a minimum of two years following colony formation, with the first winged individuals overwintering until the third year (Fowler 1986; Pricer 1908). As such, we used three years as the age of reproductive maturity and six years as the generation time for demographic analyses. This value is generally consistent with generation time estimates of seven to eight years in red harvester ant (*Pogonomyrmex barbatus*) colonies kept in captivity (Ingram *et al.* 2013).

Blochmannia phylogenomics and population genomics

Phylogenomics. We extracted CDS regions from each *Blochmannia* genome using bedtools. We then used BLAST to match each gene to genes in the outgroup *Blochmannia floridanus* to identify putatively homologous genes from each *Blochmannia* genome. For further analysis we kept 507 genes present in all samples. We aligned all sequences for each gene using T-Coffee and trimmed any portions of the alignments not present in all samples using trimAl. Next, we used RAxML v8.2.12 (Stamatakis 2014) with the GTRGAMMA model to estimate a phylogeny for each of the *Blochmannia* genes. From these gene trees, we estimated a species tree using two methods: (1) maximum clade credibility tree of all input trees using DendroPy (Sukumaran & Holder 2010), and (2) the coalescent-based species tree approach ASTRAL III (Zhang *et al.* 2018).

Tests for selection. We used the HyPhy software package (Pond & Muse 2005) to test for selection in the *Blochmannia* genes in a phylogenetic framework. We tested for positive selection using aBSREL (Smith *et al.* 2015) and we tested for shifts in selection strength across the phylogeny using RELAX (Wertheim *et al.* 2015). We ran aBSREL in exploratory mode where all branches are tested for positive selection followed by a Holm-Bonferroni correction (Holm 1979) for multiple testing. RELAX requires a set of test branches and reference branches to identify shifts in selection strength. As such, we ran RELAX six times, once for each *Camponotus* species in the study with more than a single individual (i.e., excluding *C. ocreatus* and the outgroup).

Gene loss. We tested for gene loss in all *Blochmannia* genomes assembled for this study. First, we performed an all-to-all protein BLAST (blastp) of all amino acid sequences from coding genes annotated from all samples. From the BLAST analysis results, we tabulated a gene presence / absence matrix for each *Blochmannia* genome (N = 607 unique coding genes identified from all samples).

Camponotus and *Blochmannia* coevolution

Evolutionary rates. We explored evolutionary rates of evolution in *Blochmannia* genomes in multiple ways. First, we examined variation in *Blochmannia* gene phylogenies relative to the host species tree. To do this we calculated the Kuhner-Felsenstein (1994) distance between the host species tree and *Blochmannia* gene trees, implemented in the R package ape (Paradis *et al.* 2004). Second, we explored rates of evolutionary change in a phylogenetic context by measuring relative branch lengths in the host species tree versus the

Blochmannia gene trees. We calculated these rates of evolutionary change in three groups: (1) subgenus *Camponotus*, (2) subgenus *Tanaemyrmex* excluding *C. ocreatus* because it is on a long branch by itself, and (3) the combined group of the subgenera *Camponotus* and *Tanaemyrmex*. Third, we measured nucleotide percent identity for all gene alignments, excluding indels, in the same three groups as aforementioned.

Demographic influences on selection and gene loss. For each of the six ingroup species (i.e., excluding *C. ocreatus*), we explored the relationship between host population sizes and (1) changes in selection strength in *Blochmannia* genes and (2) patterns of gene loss in *Blochmannia* genomes. First, we performed linear regression on the relationship between host harmonic mean population size and aBSREL results (per species), specifically both (1) intensified and (2) relaxed selection regimes in *Blochmannia* genes. Second, we used phylogenetic independent contrasts (Felsenstein 1985) to explore the relationship between host harmonic mean population size and endosymbiont *Blochmannia* gene count.

RESULTS

Camponotus reference genome characteristics and molecular clock

Our *de novo* *Camponotus* sp. (1-JDM) reference genome was highly contiguous (contig L50 ~ 500 kbp) with a small number of scaffolds composing the majority of the assembly (scaffold L90 ~ 3.58 Mbp, N90 = 29; see Table S2). Overall, we had 31 scaffolds greater than 2 Mbp. Although there are no karyotypes for North American *Camponotus* (*Camponotus*), there are estimates for *Camponotus ligniperda* (haploid N = 14), *C. japonicus* (N = 13 or 14), and *C. obscuripes* (N = 14) from the eastern Palearctic (Hauschteck 1983; Imai & Yosida 1964; Imai 1969). As such, it appears we have generated a genome with the contiguity of about two scaffolds per chromosome. While we did have additional signal in the Hi-C contacts to potentially further scaffold the genome (Fig. S4), we chose to be somewhat conservative and only link genomic regions where we were confident of the signal (Fig. S4).

The difficulty in fully scaffolding the genome may relate to the repetitive nature of the genome; the genome averaged 24.7% repetitive element content with many large portions of the genome exhibiting greater than 70% repetitive content (Fig. 1). Overall, about 13% of genomic windows contained more than 60% repetitive content. A large proportion of the repetitive content was DNA transposons (Fig. S5), both previously described and those manually curated for this study. The repetitive landscape is consistent with other ant species exhibiting “islands” of extreme repetitive content in a background of lower genomic repetitive content (Schradler *et al.* 2014).

Coding gene content is heterogeneous across the genome and is negatively correlated with both repetitive element content and GC% in 100 kbp sliding windows (Fig 1, Fig. S6). BUSCO results suggest our genome is nearly complete and representative of other hymenopterans, containing 98% complete genes and 1.2% fragmented genes of the 4415 hymenopteran near-universal single copy orthologs (Table S3). Using the four-fold degenerate sites from the genome’s CDS regions, we estimated a substitution rate of 1.983877×10^{-9} substitutions / site / year.

Phylogenomics

We estimated an ant host species tree using 4770 “gene trees” estimated in 50 kbp sliding windows across the genome. Both methods we used to create the species tree—maximum clade credibility and ASTRAL—identified an identical topology (Fig. 2). Here, each species was monophyletic. *Camponotus* (*Camponotus*), including four species, was monophyletic; in contrast, most *Camponotus* (*Tanaemyrmex*) formed a clade, but *C. ocreatus* was recovered as more closely related to *C. floridanus* (subgenus *Myrmotherix*) than other species in the *Tanaemyrmex* subgenus. Between 40% and 99% of gene trees supported the relationships identified in the species tree, while each node had 100% support in ASTRAL analyses (Fig. 2).

The *Blochmannia* species tree estimated from 507 gene trees identified a strongly supported phylogeny with a nearly identical topology to the host species tree (Fig. 2). The only differences were some relationships between individuals within species (colored orange in Fig. 2). In contrast to varying proportions of gene trees matching species tree relationships in the ant hosts, a majority of *Blochmannia* gene trees matched

relationships of the species tree (67% to 99% support for each node). We measured the KF94 distance between each *Blochmannia* gene tree and the host species tree to identify if any regions of the *Blochmannia* genome were relatively divergent from the overall phylogenomic signal. In general, the phylogenetic concordance, as measured by the KF94 metric, were consistent across the entire *Blochmannia* genome (Fig. 3A).

Blochmannia genome sizes, gene composition, rates of molecular evolution

The *Blochmannia* genomes assembled here largely varied per subgenus. In the subgenus *Camponotus*, the genomes varied in size from ~783 to 792 kbp (Fig. 3). This is consistent with *Blochmannia* genomes from this subgenus already on GenBank [*B. pennsylvanicus* (NC_007292.1) = 791 kbp; *B. chromaiodes* (NC_020075.1) = 791 kbp]. In contrast, individuals in the subgenus *Tanaemyrmex* had highly variable *Blochmannia* genome sizes. Five of the genomes ranged in size from 775 kbp to 781 kbp, while *B. ocreatus* was ~746 kbp and a sequence from GenBank for *B. vafer* (NC_014909.2) was ~722 kbp (Fig. 3D).

The *Blochmannia* genomes contained between 576 and 601 genes, and in total across all genomes, 607 unique coding genes were annotated. In total, 65 genes exhibited variable presence / absence among the samples sequenced here (Fig. 4). Of those 65 genes, 37 exhibited phylogenetic signal of gene loss (Fig. 4, Fig. S7).

The *Blochmannia* genes evolved at ~20–30x the rate of the ant hosts, with slight variation across the *Blochmannia* genome (Fig. 3B). Additionally, it appeared that *Blochmannia* genes in ant hosts of the subgenus *Camponotus* had a slightly slower rate of evolution than those of the ant host subgenus *Tanaemyrmex* (Fig. 3B). If we use the *Camponotus* rate of molecular evolution to put *Blochmannia* rates in a timed absolute context, the mean *Blochmannia* gene evolution rate is about 5.474×10^{-8} substitutions / site / year (range = 1.454×10^{-8} to 1.256×10^{-7} substitutions / site / year).

Blochmannia gene identity within host subgenera was generally consistent across the endosymbiont genome, suggestive of similar evolutionary forces acting across most of the genome at the taxonomic scale of host clades (Fig. 3C). We also tried to identify relative rates of evolution and percent sequence identity in intergenic regions. To do this, we performed whole genome alignments using progressiveMauve (Darling *et al.* 2010). However, endosymbiont intergenic sequences were so divergent between host subgenera, and in some cases between host species, that we were unable to recover any high-quality alignments in these regions (e.g., large non-overlapping sections in these regions of the alignments). Needless to say, the rates of evolution in these intergenic regions are likely much higher than the genic rates in Fig. 3B.

Impacts of host demography on endosymbiont evolution

Contemporary estimates of host effective population sizes ranged from ~20,000 to 300,000 and, with a couple exceptions, were largely consistent within species (Fig. S8). Variation in demographic histories within species may be indicative of variation in among population gene flow (i.e., a lack of panmixia), and therefore the overall demographic trends for each individual should be interpreted with this in mind. Overall, however, harmonic mean population size through the last half million years was highly correlated with observed heterozygosity for each individual ($R^2 = 0.7497$, $p \ll 0.001$), and suggests that the MSMC population size estimates reflect population history, even if not simply population size trends (e.g., variance in estimates due to differential population structure). As such, we looked for correlations between endosymbiont traits and host population sizes using these harmonic mean population size estimates.

In positive selection tests, 19 *Blochmannia* genes showed evidence for selection (Table S4). These signatures of positive selection appeared somewhat randomly in the phylogeny (Fig. S9; Table S5). We also tested for shifts in selection strength (i.e., intensified or relaxed) among the host lineages for all *Blochmannia* genes. Here, as we were trying to both identify relative numbers of shifts in selection strength as well as loci with extreme shifts in selection strength, we report results for selection tests with and without multiple testing correction. We found a positive relationship between host population size estimates and number of endosymbiont genes with shifts toward intensified selection pressures (Fig. 5A, Fig. S10). In contrast, we found no relationship between number of genes with relaxed selection strength and host demography (Fig. 5B, Fig. S10). As previously mentioned, some gene loss in endosymbiont genomes exhibited phylogenetic

signal (Fig. 4, Fig. S7). Despite this, we found no evidence for a relationship between host population size and gene loss in endosymbiont genomes using phylogenetic independent contrasts (Fig. S11).

DISCUSSION

We sequenced several carpenter ant host and endosymbiont *Blochmannia* genomes to address questions about coevolutionary dynamics. We added a whole genome resource for a *Camponotus* (*Camponotus*) species and more than doubled the number of publicly available *Blochmannia* full-genome sequences. With these resources, we investigated questions related to (1) codiversification of hosts and endosymbionts, (2) variation in lineage sorting and molecular evolution across the endosymbiont genome, and (3) effects of host demography on endosymbiont patterns of gene loss and natural selection.

Codiversification of carpenter ant hosts and Blochmannia endosymbionts

Using whole-genome sequencing of both carpenter ant hosts and their bacterial endosymbionts, we identified generally strict codiversification (Fig. 2). There was some phylogenetic incongruence between host and endosymbiont trees amongst individuals within species, but all species-level relationships were completely congruent. These patterns are consistent with expectations of co-speciation between hosts and vertically-transmitted endosymbionts; similar evidence of codiversification between hosts and endosymbionts has been found in weevils (Toju *et al.* 2013), flies (Chen *et al.* 1999; Hosokawa *et al.* 2012), cockroaches (Clark *et al.* 2001; Loet *et al.* 2003), bivalves (Distel *et al.* 1994), aphids (Clark *et al.* 2000), psyllids (Thao *et al.* 2000), and previous studies in carpenter ants (Degnan *et al.* 2004). Generally, previous studies investigating codiversification have inferred phylogenies using one or a few molecular markers; in contrast, by sequencing full genomes for both hosts and endosymbionts, we were able to obtain strongly supported species trees as well as estimate variation in lineage sorting across *Blochmannia* genes (Fig. 3A) with phylogenetic statistics. Here, we expected one of two patterns: (1) consistent signal across the genome with relatively similar overall rates of evolution across *Blochmannia* genes, or (2) a highly variable landscape of phylogenetic congruence and incongruence caused by variable selective pressures across the genome. Indeed, we found a generally consistent pattern of lineage sorting across the genome as evidenced by a stable estimate of the KF94 statistic across the *Blochmannia* genome.

Rates of molecular evolution in Blochmannia endosymbionts

We found that *Blochmannia* genes evolved at a rate $\sim 30x$ faster than the host genome (Fig. 3). In addition, intergenic regions were so divergent across lineages that we were not able to align them properly. This endosymbiont-host relative evolution rate is similar to the level reported in *Buchnera* bacterial endosymbionts of aphids ($\sim 36x$) by Moran and colleagues (1995). On an absolute scale, the rates of evolution identified here in *Blochmannia* average about an order of magnitude faster than those reported in *Buchnera* (Clark *et al.* 1999). Relatively fast evolution rates are expected in endosymbionts because of their life histories; insect endosymbionts' asexuality and propensity to undergo regular bottlenecks because of their mode of inheritance lead to small effective population sizes and relatively fast evolution (Mira & Moran 2002; Wernegreen 2002). As such, endosymbionts also have faster relative evolutionary rates compared to their free-living relatives, including increased rates of evolution at nonsynonymous coding sites (Moran 1996). However, endosymbiont molecular evolution rate varies somewhat across the genome and may have host-lineage specific rates of molecular evolution (Kuo & Ochman 2009). These previous results are corroborated by molecular evolution rates identified here in *Blochmannia* genomes (Fig. 3). Overall, these results corroborate previous evidence that endosymbionts have faster rates of evolution relative to both their hosts and to their free-living bacterial relatives.

Does host demography shape endosymbiont evolution?

Because population genomic processes are influenced by effective population size, and endosymbiont effective population size is intrinsically linked with host effective population sizes (Mira & Moran 2002; Wernegreen 2002), we may have a null expectation that host demographic patterns partially influence endosymbiont molecular evolution. Here, we investigated whether host demography influenced two factors of endosymbiont

genome evolution: (1) patterns of natural selection, and (2) patterns of gene loss.

We found no relationship between host demography and both signatures of positive selection and relaxation of selection strength in *Blochmannia* genes (Figs. 5, S9, S10). In contrast, we found a positive relationship between host population sizes and shifts toward intensified selection pressures in *Blochmannia* genes (Figs. 5, S10; Table S6). In endosymbionts in general, we may expect relaxed selection relative to patterns in free-living bacteria (Wernegreen 2002). Indeed, selection is often identified in insect endosymbionts, but generally only in a small fraction of genes (Alleman *et al.* 2018; Chong *et al.* 2019; Williams & Wernegreen 2012). Based on our results (Figs. 5, S10), it appears that shifts in selection pressures may at least in part be influenced by host demographic processes.

We also tested for an effect of host demography on patterns of endosymbiont gene loss and found no significant relationship between the two (Fig. S11). We initially anticipated that genetic drift would occur faster, along with decreased selection pressures, in endosymbionts with small host population sizes, thereby leading to faster rates of gene loss. This was not the case in the entire dataset, but the species with the smallest estimated population sizes and estimates of genetic drift—*C. laevigatus*—did have the most endosymbiont gene loss (Fig. 4). We found that about half of the gene loss was phylogenetically informative, while the rest was not (Figs. 4, S7; Table S4). This suggests relatively random patterns of gene loss in the phylogeny; a majority of gene loss lacking phylogenetic signal were singleton gene losses (Fig. 4). This is consistent with previous research in *Blochmannia* endosymbionts identifying lineage-specific gene loss largely due to relaxed selection constraints and genetic drift (Williams & Wernegreen 2015).

CONCLUSIONS

We used whole-genome sequencing of both carpenter ant hosts and their endosymbionts to investigate their coevolution. We identified strict codiversification of *Camponotus* hosts and their *Blochmannia* endosymbionts. *Blochmannia* genes are evolving about 30x faster than host genomes, with relatively consistent evolutionary rates across the *Blochmannia* genome. In *Blochmannia* genes, we found some evidence for positive selection and shifts in selection strength across the host phylogeny. We found that some, but not all, patterns of natural selection were in part shaped by host demographic history. Lastly, we found somewhat random endosymbiont gene loss that was not related to host demographic histories.

AUTHOR CONTRIBUTIONS

All authors contributed to (1) development of ideas in the project, (2) fieldwork for collecting samples, and (3) revisions of initial writing drafts. JPH performed lab work. JDM analyzed the data and wrote the first draft of the manuscript. JCG and JDM took macro photos of ant specimens.

DATA AVAILABILITY

All raw sequencing data is currently being submitted to NCBI's SRA. All *Blochmannia* assemblies in FASTA format are being submitted to NCBI's GenBank, while all GFA format assemblies will be uploaded to Dryad. The de novo *Camponotus* genome assembly and annotations will be uploaded to Dryad until a further taxonomic investigation allows us to either place the species or describe it as a new taxon, to be subsequently followed by uploading the genome to NCBI's Genome archive. Output files, including those from tests for selection and gene loss in *Blochmannia*, will be uploaded to Dryad. All code used for analyses in this project are available on GitHub: <https://github.com/jdmanthey/camponotus-genomes1>.

ACKNOWLEDGEMENTS

This work was supported by Texas Tech University startup funds to JDM. JDM thanks John Jisha for getting him interested in *Camponotus* more than a decade ago with a month-long field trip.

FIGURES & TABLES

Figure 1. *Camponotus* sp. (1-JDM) *de novo* genome assembly characteristics, including coding content (CDS), repetitive and transposable elements, and GC content. Points indicate summary statistics in 100

kbp non-overlapping sliding windows, while solid lines indicate ten window (i.e., 1 Mbp) mean estimates.

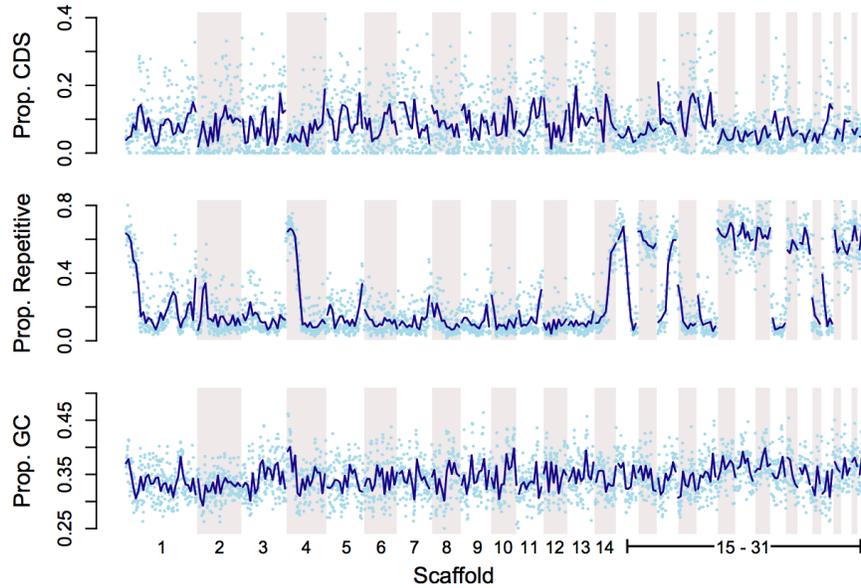


Figure 2. Phylogenetic congruence of host ants and their endosymbionts. Branch labels indicate proportion of trees supporting this phylogenetic hypothesis. The ASTRAL species tree topologies were identical to these phylogenies and exhibited 100% quartet support for every relationship. Host trees were rooted with the *Cataglyphis niger* sample. The *Blochmannia* tree was midpoint rooted. Orange branches in the *Blochmannia* tree indicate branches that vary between the host and endosymbiont phylogenies.

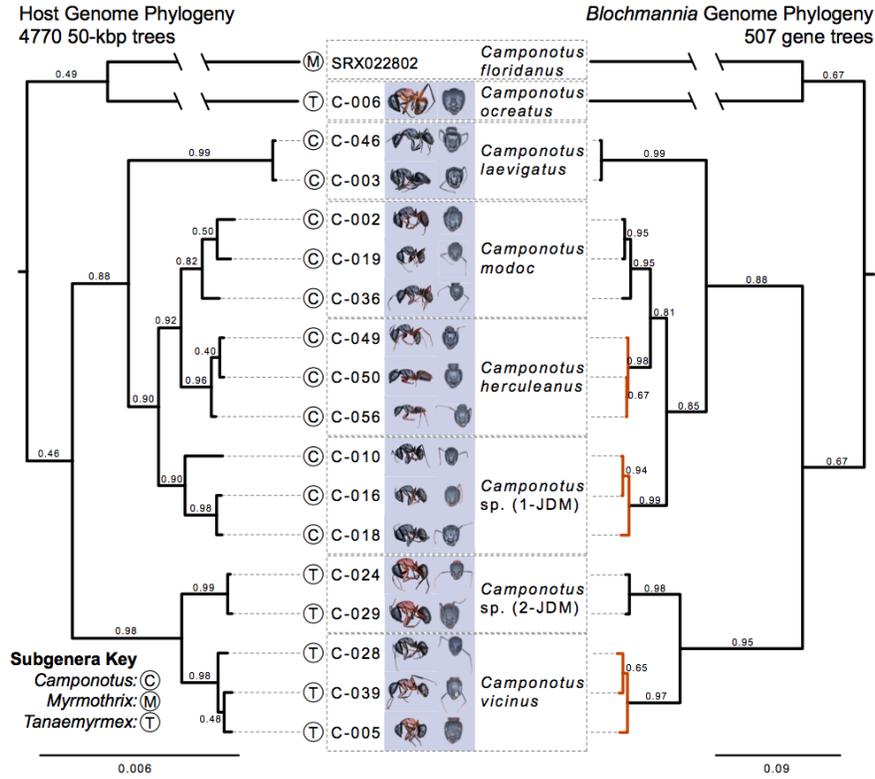


Figure 3. Variation in molecular evolution across the *Blochmanniagenomes*. (A) The KF94 distance between the host species tree and *Blochmannia* genes. The KF94 distance measures differences between phylogenetic topologies including branch lengths, where a value of zero is an identical tree. (B) Relative rates of molecular evolution in *Blochmannia* genes relative to the host species tree. (C) Gene identity percentage in *Blochmannia* genes for alignments with indels removed. For all statistics, lines indicate mean values across windows of ten genes. (D) *Blochmannia* genome assembly sizes for all new assemblies in this study as well as those published on GenBank for these *Camponotus* subgenera. GenBank sequences: *Blochmannia pennsylvanicus* (NC_007292.1), *B. vafer* (NC_014909.2), *B. chromaiodes* (NC_020075.1).

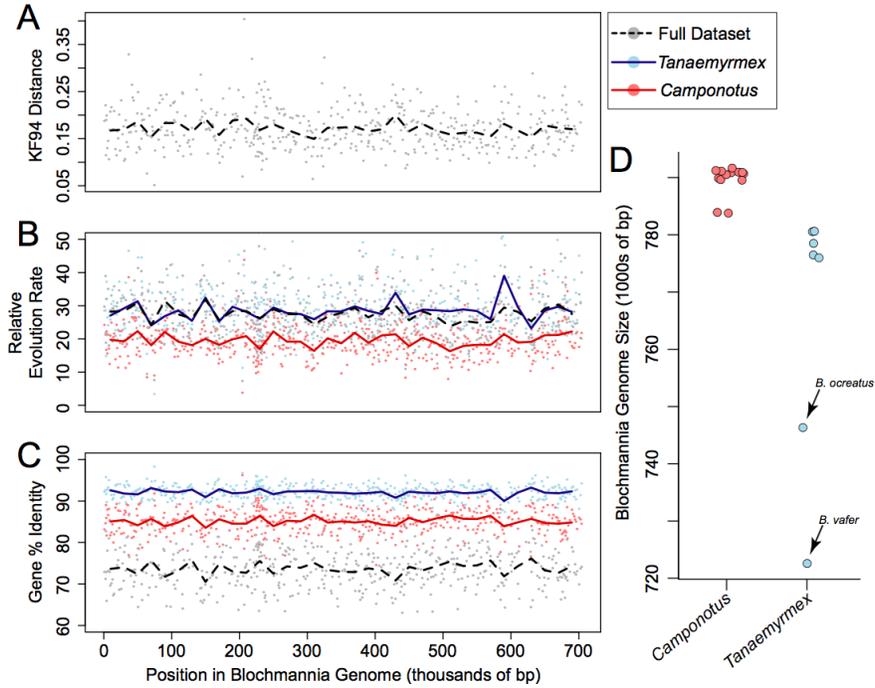


Figure 4. *Blochmannia* gene presence /absence phylogenetic heatmap. Of 607 total genes annotated, 65 varied in presence / absence among samples (Table S4). 37 of 65 genes lost in at least one individual exhibit phylogenetic signal (Fig. S7). Dark green indicates presence, while light yellow indicates absence.

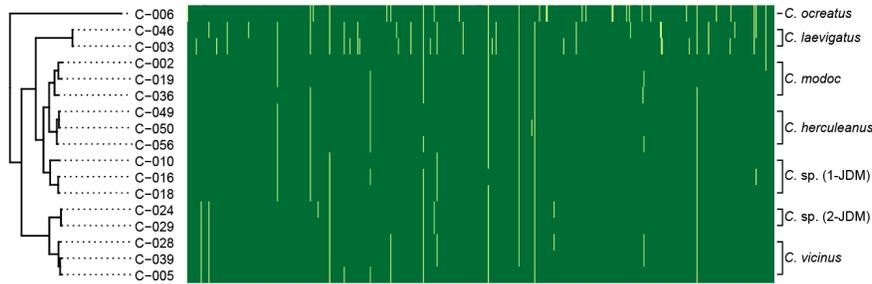
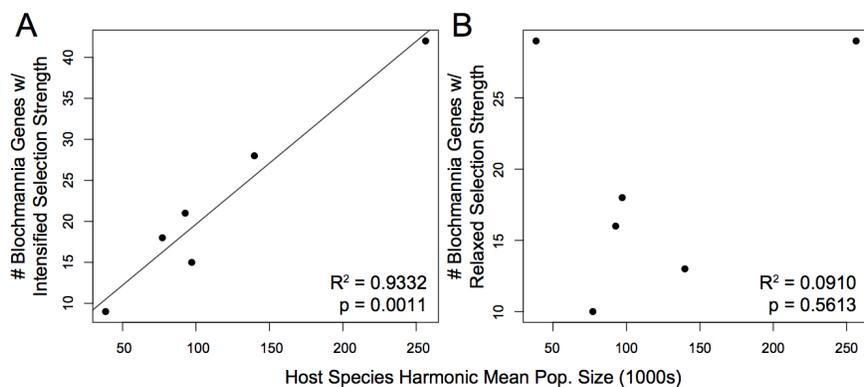


Figure 5. Relationships of shifts in selection strength and host demography. Shifts in selection strength were obtained using the program aBSREL. Results presented here are for non-multiple testing corrected results. Results with multiple testing correction are consistent with these results and presented in Figure S10.



LITERATURE CITED

Alleman A, Hertweck KL, Kambhampati S (2018) Random Genetic Drift and Selective Pressures Shaping the Blattabacterium Genome. *Scientific Reports* **8** , 1-12.

Anbutsu H, Moriyama M, Nikoh N, *et al.* (2017) Small genome symbiont underlies cuticle hardness in beetles. *Proceedings of the National Academy of Sciences, USA* **114** , E8382-E8391.

Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research* **12** , 1269-1276.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* , 289-300.

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27** , 573-580.

Blaimer BB, Brady SG, Schultz TR, *et al.* (2015) Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. *BMC Evolutionary Biology* **15** , 271.

Boman J, Frankl-Vilches C, da Silva dos Santos M, *et al.* (2019) The genome of Blue-capped Cordon-Bleu uncovers hidden diversity of LTR retrotransposons in zebra finch. *Genes* **10** , 301.

Bonasio R, Zhang G, Ye C, *et al.* (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329** , 1068-1071.

Brelsford A, Purcell J, Avril A, *et al.* (2020) An ancient and eroded social supergene is widespread across *Formica* ants. *Current Biology* **30** , 304-311. e304.

Brownlie JC, Johnson KN (2009) Symbiont-mediated protection in insect hosts. *Trends in Microbiology* **17** , 348-354.

Bushnell B (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US).

Camacho C, Coulouris G, Avagyan V, *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10** , 1.

Cantarel BL, Korf I, Robb SM, *et al.* (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18** , 188-196.

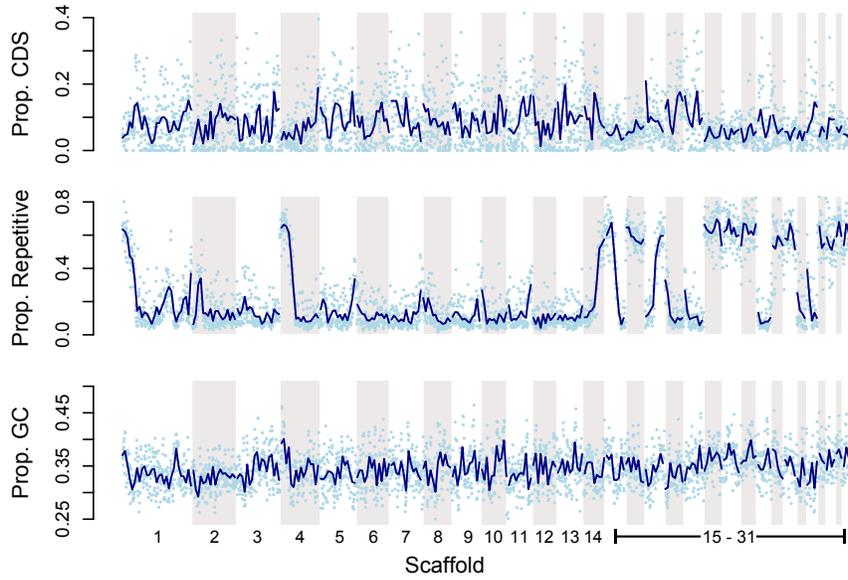
Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25** , 1972-1973.

- Charif D, Lobry JR (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: *Structural Approaches to Sequence Evolution* , pp. 207-232. Springer.
- Chen X, Li S, Aksoy S (1999) Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *Journal of Molecular Evolution* **48** , 49-58.
- Chikhi L, Rodríguez W, Grusea S, *et al.* (2018) The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity***120** , 13-24.
- Chong RA, Park H, Moran NA (2019) Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Molecular Biology and Evolution***36** , 1481-1489.
- Clark JW, Hossain S, Burnside CA, Kambhampati S (2001) Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268** , 393-398.
- Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Molecular Biology and Evolution* **16** , 1586-1598.
- Clark MA, Moran NA, Baumann P, Wernegreen JJ (2000) Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* **54** , 517-525.
- Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5** , e11147.
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods***9** , 772.
- Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ (2004) Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus Blochmannia*. *Systematic Biology* **53** , 95-110.
- Dhaygude K, Nair A, Johansson H, Wurm Y, Sundström L (2019) The first draft genomes of the ant *Formica exsecta*, and its *Wolbachia* endosymbiont reveal extensive gene transfer from endosymbiont to host. *BMC Genomics* **20** , 1-16.
- Distel D, Felbeck H, Cavanaugh C (1994) Evidence for phylogenetic congruence among sulfur-oxidizing chemoautotrophic bacterial endosymbionts and their bivalve hosts. *Journal of Molecular Evolution* **38** , 533-542.
- Dudchenko O, Batra SS, Omer AD, *et al.* (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356** , 92-95.
- Durand NC, Robinson JT, Shamim MS, *et al.* (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3** , 99-101.
- Ehrlich PR, Raven PH (1964) Butterflies and plants: a study in coevolution. *Evolution* , 586-608.
- Feldhaar H, Straka J, Krischke M, *et al.* (2007) Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biology* **5** , 48.
- Felsenstein J (1985) Phylogenies and the comparative method. *The American Naturalist* **125** , 1-15.
- Fowler HG (1986) Polymorphism and colony ontogeny in North American carpenter ants (Hymenoptera: Formicidae: *Camponotus pennsylvanicus* and *Camponotus ferrugineus*). *Zoologische Jahrbücher. Abteilung für allgemeine Zoologie und Physiologie der Tiere* **90** , 297-316.

- Gil R, Silva FJ, Zientz E, *et al.* (2003) The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences, USA* **100** , 9388-9393.
- Guindon S, Dufayard J-F, Lefort V, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59** , 307-321.
- Guyomar C, Delage W, Legeai F, *et al.* (2020) MinYS: Mine Your Symbiont by targeted genome assembly in symbiotic communities. *NAR Genomics and Bioinformatics* **2** , lqaa047.
- Hauschteck J (1983) Ant Chromosomes II. Karyotypes of Western Palearctic species. . *Insectes Sociaux* **30** , 149-164.
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* , 65-70.
- Hosokawa T, Nikoh N, Koga R, *et al.* (2012) Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *The ISME journal* **6** , 577-587.
- Imai H, Yosida T (1964) Chromosome observations in Japanese ants. *Annual report of the National Institute of Genetics (Japan)* **15** , 64-66.
- Imai HT (1969) Karyological studies of Japanese ants. I. Chromosome evolution and species differentiation in ants. . *Science Report Tokyo Kyoiku Daigaku Section B* **14** , 27-46.
- Ingram KK, Pilko A, Heer J, Gordon DM (2013) Colony life history and lifetime reproductive success of red harvester ant colonies. *Journal of Animal Ecology* **82** , 540-550.
- Janzen DH (1980) When is it coevolution? *Evolution* **34** , 611-612.
- Jurka J, Kapitonov VV, Pavlicek A, *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110** , 462-467.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30** , 772-780.
- Kiester AR, Lande R, Schemske DW (1984) Models of coevolution and speciation in plants and their pollinators. *The American Naturalist* **124** , 220-243.
- Kikuchi Y (2009) Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes and Environments* , 0908180109-0908180109.
- Koren S, Walenz BP, Berlin K, *et al.* (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* , 215087.215116.
- Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5** , 59.
- Kuhner MK, Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* **11** , 459-468.
- Kuo C-H, Ochman H (2009) Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biology Direct* **4** , 35.
- Laetsch DR, Blaxter ML (2017) BlobTools: Interrogation of genome assemblies. *F1000Research* **6** , 1287.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25** , 1754-1760.
- Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T (2003) Evidence for cocoladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Molecular Biology and Evolution* **20** , 907-913.

- Mackay W (2019) *New World Carpenter Ants of the Hyperdiverse Genus Camponotus. Volume 1: Introduction, Keys to the Subgenera and Species Complexes and the Subgenus Camponotus* Lambert Academic Publishing.
- Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity* **116** , 362-371.
- McKenna A, Hanna M, Banks E, *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20** , 1297-1303.
- Mira A, Moran NA (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microbial Ecology* **44** , 137-143.
- Moran NA (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences, USA* **93** , 2873-2878.
- Moran NA, von Dohlen CD, Baumann P (1995) Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *Journal of Molecular Evolution* **41** , 727-731.
- Moreau CS (2020) Symbioses among ants and microbes. *Current Opinion in Insect Science* **39** , 1-5.
- Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H (2015) Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Current Biology* **25** , 1375-1380.
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302** , 205-217.
- Pagès H, Aboyoun P, Gentleman R, DebRoy S (2017) Biostrings: Efficient manipulation of biological strings. *R Package Version 2.0* .
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20** , 289-290.
- Perlmutter JJ, Bordenstein SR (2020) Microorganisms in the reproductive tissues of arthropods. *Nature Reviews Microbiology* , 1-15.
- Pond SLK, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. In: *Statistical Methods in Molecular Evolution* , pp. 125-181. Springer.
- Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* **21** , i351-i358.
- Pricer JL (1908) The life history of the carpenter ant. *The Biological Bulletin* **14** , 177-218.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26** , 841-842.
- Renaud G, Hanghøj K, Korneliussen TS, Willerslev E, Orlando L (2019) Joint Estimates of Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples. *Genetics* , genetics. 302057.302019.
- Rizk G, Gouin A, Chikhi R, Lemaitre C (2014) MindTheGap: integrated detection and assembly of short and long insertions. *Bioinformatics* **30** , 3451-3457.
- Russell JA, Sanders JG, Moreau CS (2017) Hotspots for symbiosis: function, evolution, and specificity of ant-microbe associations from trunk to tips of the ant phylogeny (Hymenoptera: Formicidae). *Myrmecological News* **24** , 43-69.
- Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* **46** , 919.

- Schrader L, Kim JW, Ence D, *et al.* (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications* **5** , 5495.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31** , 3210-3212.
- Smit A, Hubley R, Green P (2015) RepeatMasker Open-4.0. 2013–2015. *Institute for Systems Biology*. <http://repeatmasker.org> .
- Smit AF, Hubley R (2008) RepeatModeler Open-1.0. Available from <http://www.repeatmasker.org> .
- Smith MD, Wertheim JO, Weaver S, *et al.* (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution* **32** , 1342-1353.
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30** , 1312-1313.
- Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** , ii215-ii225.
- Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26** , 1569-1571.
- Tatusova T, DiCuccio M, Badretdin A, *et al.* (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* **44** , 6614-6624.
- Thao ML, Moran NA, Abbot P, *et al.* (2000) Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Applied and Environmental Microbiology* **66** , 2898-2905.
- Toju H, Tanabe AS, Notsu Y, Sota T, Fukatsu T (2013) Diversification of endosymbiosis: replacements, co-speciation and promiscuity of bacteriocyte symbionts in weevils. *The ISME journal* **7** , 1378-1390.
- Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. *Nature reviews genetics* **3** , 850-861.
- Wernegreen JJ, Kauppinen SN, Brady SG, Ward PS (2009) One nutritional symbiosis begat another: Phylogenetic evidence that the ant tribe Camponotini acquired Blochmanniaby tending sap-feeding insects. *BMC Evolutionary Biology* **9** , 292.
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution* **32** , 820-832.
- Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31** , 3350-3352.
- Williams LE, Wernegreen JJ (2012) Purifying selection, sequence composition, and context-specific indel mutations shape intraspecific variation in a bacterial endosymbiont. *Genome Biology and Evolution* **4** , 44-51.
- Williams LE, Wernegreen JJ (2015) Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among Blochmannia spanning the origin of the ant tribe Camponotini. *PeerJ* **3** , e881.
- Wilson EO (1976) Which are the most prevalent ant genera. *Studia Entomologica* **19** , 187-200.
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13** , 555-556.
- Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** , 153.



Hosted file

Fig_2_camponotus_tree_reduced.pdf available at <https://authorea.com/users/380686/articles/496575-impact-of-host-evolutionary-history-on-endosymbiont-genome-evolution-a-test-in-camponotus-carpenter-ants-and-their-blochmannia-endosymbionts>

