Genetic Signature of Reproductive Manipulation in the Phylogeography of the Bat Fly, *Trichobius major*

**JUSTIN B. LACK, RANDILEA D. NICHOLS, GREGORY M. WILSON, AND RONALD A. VAN DEN BUSSCHE**

From the Department of Zoology, 501 Life Sciences West, Oklahoma State University, Stillwater, OK 74078 (Lack, Nichols, and Van Den Bussche); and the Department of Biology, University of Central Oklahoma, Edmond, OK 73034 (Wilson).

Address correspondence to Justin B. Lack, Department of Zoology, 501 Life Sciences West, Oklahoma State University, Stillwater, OK 74078, or e-mail: justin.lack@okstate.edu.

**Abstract**

The bat fly (*Trichobius major*) is a blood-feeding ectoparasite of the cave myotis (*Myotis velifer*). A recent mitochondrial DNA (mtDNA) study examining population structure of *T. major* in the South Central United States detected a single haplotype from all individuals examined (*N* = 48 from 12 different caves), representing one of only a few known examples of such widespread mtDNA uniformity. We examined nuclear genetic diversity using amplified fragment length polymorphism and detected high levels of nuclear genetic diversity in all populations sampled. Amplified fragment length polymorphism analyses indicated significant levels of gene flow among caves 700 km apart, suggesting the absence of mtDNA diversity in *T. major* is the result of a selective sweep, not a demographic event (i.e., a recent bottleneck). One mechanism by which mtDNA sweeps occur in arthropods is through bacterial parasites that manipulate host reproduction and mtDNA inheritance. We used PCR to test for the presence of all known reproductive parasites and detected a widespread infection (91.33% infection rate) of *T. major* with a novel *Arsenophonus* bacterium, as well as the infection of 2 individuals (1.16% infection rate) with a novel strain of *Rickettsia*. We discuss the implications for *T. major* phylogeography and the necessity of a bigenomic approach in arthropod population genetics.

**Key words:** AFLP, *Arsenophonus*, bat fly, *Rickettsia*, reproductive parasites, *Trichobius major*

In studies of population genetics and phylogeography, genetic loci are sampled to examine the evolutionary history and distribution of genetic variation within a species, compare patterns across species, or to identify portions of the genome under natural selection. Most often, any patterns recovered are assumed to be the result of neutral evolution (i.e., drift) or natural selection acting directly on the organism (or more specifically, the genetic locus being examined). Particularly for nonmodel organisms, mitochondrial DNA (mtDNA) has been the locus of choice due to its high substitution rate, 4-fold elevated fixation rate of novel substitutions (due to differences in effective population size), almost complete lack of recombination, and maternal inheritance (Birky 2001; but see Galtier et al. 2009). Owing to the above characteristics as well as the ease of design and near ubiquitous availability of primers, mtDNA analysis has produced a wealth of knowledge for organisms ranging from prokaryotes to mammals; however, over the last decade the utility of mtDNA in assessing evolutionary history and molecular diversity has been questioned. The maternal inheritance of mtDNA means any recovered evolutionary history is that of only females, and for organisms with differing patterns of gene flow between the sexes (i.e., male-biased dispersal and female philopatry in mammals; Greenwood 1980), mtDNA will provide an incomplete history. Also, the permanently linked nature of mtDNA means selective sweeps can result in rapid shifts in standing diversity, eliminating historical signal. The prevalence of these issues is still relatively unknown, but recent evidence suggests that problems may be more widespread than previously thought (Hurst and Jiggins 2005; Galtier et al. 2009).

One mechanism by which mtDNA variation might not accurately reflect the evolutionary history of the organism is through reproductive manipulation by microbial parasites, a phenomenon known to occur in arthropods (Yen and Barr 1971; Johnstone and Hurst 1996; Hurst and Jiggins 2005; Moran et al. 2008). These bacteria are most often vertically transmitted from mother to offspring and through various mechanisms skew reproductive success to increase the probability of a particular bacterial strain or species being propagated (for an extensive review, see Stouthammer...
et al. 1999). Reproductive manipulation has been shown to occur through male-killing, cytoplasmic incompatibility, feminization of males, induced parthenogenesis, and the requirement of infection for oogenesis (Gherna et al. 1991; Southammer et al. 1999; Dedine et al. 2001; Hagimori et al. 2006). These manipulations typically result in linkage disequilibrium between the bacterium and the mtDNA of the infected individuals. This can lead to a drastic reduction in haplotype diversity within or even among entire populations (Turelli et al. 1992; Jiggins 2003). Alternatively, the presence of many strains of symbionts within a population or distinct strains between different populations can lead to uncharacteristically high mtDNA diversity, and bacterial symbionts have even been implicated as potential drivers of speciation (Sharon et al. 2010). These patterns have been observed in a diversity of arthropod taxa (Diptera, Turelli et al. 1992; Isopoda, Marcade et al. 1999; Coleoptera, Schulenburg et al. 2002; Amphipoda, Ironside et al. 2003; Lepidoptera, Jiggins 2003; and Hymenoptera, Reuter and Keller 2003), and surveys suggest the prevalence of these bacteria in some arthropod groups is high (>50% of spiders, Rowley et al. 2004; >30% of insects, Duron et al. 2008).

The streblid bat fly (Triebelobius major) is an obligate blood-feeding ectoparasite of the cave myotis (Myotis velifer), a vespidilionid bat in the southwestern United States and Mexico, although T. major has been found at low frequency on other codistributed bats (Johling 1949; Ross 1961; Overall 1974; Reisen et al. 1976). Triebelobius major exhibits an upward spiraling flight that is utilized for locating hosts in caves but is not conducive to dispersion between caves (Caire and Hornuff 1982). Morphological analysis of T. major from western Oklahoma revealed no significant differences among caves approximately 300 km apart, suggesting significant gene flow between these caves (Caire et al. 1981). Caire et al. (1981) hypothesized that dispersal in T. major occurred via M. velifer, which have been documented making overnight dispersals >22 km and seasonal movements of several hundred kilometers (Hayward 1970), allowing for large amounts of gene flow between relatively distant populations. To test this hypothesis, Wilson et al. (2007) sequenced portions of the cytochrome oxidase I (COI) and NADH dehydrogenase 4 (ND4) genes from T. major collected from caves in Kansas, Oklahoma, and Texas. All individuals examined were identical at both mtDNA loci leading Wilson et al. (2007) to suggest T. major had undergone a recent and significant bottleneck from which mtDNA diversity had not yet recovered. While we do not deny this as a possibility, the fixation of a single haplotype across T. major populations occurring >700 km apart and parasitizing geographically and morphologically distinct host subspecies (M. v. grandis and M. v. incautus; Vaughan 1954; Hayward 1970) suggests that mtDNA similarities across populations may be maintained by disequilibrium between a bacterial symbiont and an infected maternal lineage.

Bacterial reproductive parasites have the ability to manipulate the distribution and frequency of haplotypes in space and can be the primary driver of shifts in mtDNA diversity over time. When inferring phylogeographical relationships or historical demography of an organism, these are the primary parameters of interest. Therefore, it can be difficult to distinguish between recent demographic events such as a bottleneck or founder effect and sweeps that occur due to selection on an infecting reproductive parasite (Hurst and Jiggins 2005). To distinguish between these two processes (bottleneck vs. genetic sweep), it is necessary to examine nuclear data as demographic events will affect both the nuclear and mitochondrial genomes, while maternally inherited reproductive parasites will only affect mitochondrial markers, a pattern confirmed in multiple taxa (Ballard and Kreitman 1994; Ballard 2000; Jiggins 2003). In addition, for species where mtDNA diversity has been manipulated by bacterial reproductive parasites, mitochondrial markers will not provide accurate depictions of dispersal and gene flow. Therefore, we examined nuclear genetic diversity using amplified fragment length polymorphism (AFLP) from the previously examined populations of Wilson et al. (2007) to address the following questions: 1) Is the extreme reduction in T. major mtDNA diversity the result of recent demographic events or disequilibrium between a bacterial reproductive parasite and the infected mtDNA lineage? 2) To what extent is gene flow occurring between populations of T. major? Also, we used PCR detection to address a third question: 3) Do any known bacterial reproductive parasites occur in T. major?

**Materials and Methods**

**Sample Collection and DNA Isolation**

We collected 173 T. major from 5 cave systems in the South Central United States: 4 caves in southern Kansas, 2 caves in the Alabaster cave system in northwestern Oklahoma, 3 caves in the Selman cave system also in northwestern Oklahoma, 2 unnamed caves in a single system in central Oklahoma, and a single cave in South Central Texas (Figure 1). We refer to the 2 central Oklahoma caves as “Washita caves 1 and 2” and members of the “Washita” cave system to keep with the terminology previously used (Wilson et al. 2007), although these caves are found in Blaine County, OK. Kansas and Oklahoma specimens were collected in December 2002, December 2003, March 2004, and January 2005 by removing individual bat flies from the ears of hibernating M. velifer. Texas specimens were collected in July 2004 from the surface of rocks at the entrance of a cave as well as by netting bats at the cave entrance and removing bat flies from their body. Following collection, each individual was placed in a 1.5-µl tube containing 500 µl of lysis buffer (Longmire et al. 1997). The entire fly was macerated and whole-genomic DNA was extracted using the protocol of Longmire et al. (1997).

**AFLP Data Collection and Analysis**

AFLP profiles were generated following a modified version of the protocol described by Vos et al. (1995).
Approximately 200–1000 ng whole-genomic DNA was digested in 20 µl volume using 2 units of EcoRI (New England Biolabs, Beverly, MA) and 2 units of MseI (New England Biolabs) at 37 °C for 3 h. Following digestion and deactivation of enzymes, specific adapters (Table 1) were ligated by adding 3 units of T4 DNA ligase (Invitrogen) and 75 pmoles of each adapter to the digest and incubating at 16 °C for 10 h. Following ligation, samples were diluted 1:10, and preselective PCR was conducted with primers (Table 1) annealing to adapters, but with the addition of one base at the 3′ end. Preselective PCR conditions were 72 °C for 60 s, 20 cycles of 94 °C for 50 s, 56 °C for 60 s, and 72 °C for 120 s. The resulting product was again diluted 1:10 and used as a template for the selective PCR under the following conditions: 13 cycles of 94 °C for 50 s, 65 °C for 60 s, 72 °C for 120 s with the annealing temperature lowered 0.7 °C each cycle, and followed by 23 cycles of 94 °C for 50 s, 56 °C for 60 s, and 72 °C for 120 s. Eight selective primer pairs (Table 1) were used to produce the AFLP profiles analyzed.

Selective PCR products were electrophoresed in an ABI 3130 genetic analyzer (Applied Biosystems, Inc., Foster City, CA). Band scoring was performed manually by R.D.N for all individuals included and then was repeated by J.B.L. for 24 arbitrarily selected runs from each 96-well plate. In addition, AFLP profiles were generated a second time for 24 arbitrarily selected individuals to ensure profiles and scoring were reproducible. Bands ranging from 50 to 500 base pairs were scored, and any locus for which there was conflict between the 2 independent scorers was removed from all analyses so that only distinct and unambiguous bands were included in the final data set. All bands occurring in a frequency <5% were excluded from the final data set. Mean heterozygosity was estimated using a Bayesian approach implemented in Hickory v1.1 that relaxes

**Table 1** Adapters and selective PCR primer combination used to generate the *Trichobius major* AFLP profiles

<table>
<thead>
<tr>
<th>Primer/adapter combinations (5’ to 3’)</th>
<th>Primer/adapter combinations (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI adapter A</td>
<td>EcoRI adapter B</td>
</tr>
<tr>
<td>CTCGTA GACGTGTA AATTACA</td>
<td>AATTTGAGCGCTGCTAC</td>
</tr>
<tr>
<td>MseI adapter A</td>
<td>MseI adapter B</td>
</tr>
<tr>
<td>GACGAT GAGGTGCTGAG</td>
<td>TACTCAGGACTCAT</td>
</tr>
<tr>
<td>Selective EcoRI primer</td>
<td>Selective MseI primer</td>
</tr>
<tr>
<td>ACTGGTAGACAATTAACA</td>
<td>GATGAGTCCTGAGATACCA</td>
</tr>
<tr>
<td>ACTGGTAGAACATTAACA</td>
<td>GATGAGTCCTGAGATACCA</td>
</tr>
<tr>
<td>ACTGGTAGACAAATTAG</td>
<td>GATGAGTCCTGAGATACCA</td>
</tr>
<tr>
<td>ACTGGTAGAACTTATG</td>
<td>GATGAGTCCTGAGATACCA</td>
</tr>
<tr>
<td>ACTGGTAGACAATATG</td>
<td>GATGAGTCCTGAGATACCA</td>
</tr>
<tr>
<td>ACTGGTAGAATATG</td>
<td>GATGAGTCCTGAGATACCAT</td>
</tr>
</tbody>
</table>

Although only selective primer sequences are shown, preselective primers were identical to the selective primers, but with the 2 most terminal 3′ bases removed.

**Figure 1.** Geographic location of *Trichobius major* populations sampled and the distribution of their host, *M. velifer*. The distribution of *M. velifer* subspecies is shown in gray. The caves sampled were Double Entrance (1), Sink Valley (2), Parker Bat Cave (3), and Swartz Cave (4) in the Kansas Cave System; Alabaster Caverns (5) and Bear Cave (6) in the Alabaster Cave System; Skylight (7), Skunk eater (8), and Green’s Cathedral (9) in the Selman Cave System; Washita Cave 1 (10) and Washita Cave 2 (11) in the Washita Cave System; and the Eckert James River Bat Cave (12) in South Central Texas.
assumption of Hardy–Weinberg (H-W) equilibrium (Holsinger et al. 2002).

We examined the partitioning of genetic variation among populations and cave systems using two approaches. We conducted a standard hierarchical analysis of molecular variance (AMOVA) partitioned by population and cave system in GenAlEx v6.4 (Peakall and Smouse 2006). Genetic distance (D_{ST}/C0) was calculated following the method of Evanno et al. (2005). Analyses were conducted in GenAlEx v6.4. This analysis was only conducted in GenAlEx v6.4. The Bayesian model-based clustering method implemented in the program STRUCTURE (Pritchard et al. 2000) was used to infer population structure and assign individuals to populations. STRUCTURE gives the probability of assignment of each individual to postulated clusters, allowing for the identification of individuals with ancestry attributable to multiple populations. Following a burn-in of 10^5 generations, data were collected for 2 × 10^5 generations under both no-admixture and admixture models, with no a priori assumption of population origin, with the “correlated allele frequencies” option (as recommended for dominant data; Pritchard and Wen 2004) and with the input data formatted following Falush et al. (2007). Analyses were run at K ranging from 1 to 12 with 5 replicates at each value of K. The optimal value of K was calculated following the method of Evanno et al. (2005). The 5 replicates for each K were then combined using the program CLUMPP (Jakobsson and Rosenberg 2007), which identifies common modes among replicate runs. Results from STRUCTURE were used to construct graphs using DISTRACT (Rosenberg 2004).

Bacterial Symbiont Data Collection and Analysis

Although Wolbachia is the most thoroughly studied bacterial reproductive parasite, several other bacterial lineages are known to manipulate reproduction in arthropods. Cardinium is a bacterium with reproductive manipulations that rival Wolbachia in diversity (Weeks et al. 2001; Provencher et al. 2005; Gotou et al. 2007; Kenyon and Hunter 2007), and reproductive manipulation has also been documented in Arsenophonus (Gherna et al. 1991), Rickettsia (Lawson et al. 2001; Hagimori et al. 2006), non-Cardinium members of the Flavobacterium clade (Harst et al. 1999), and Spiroplasma (Tinsley and Majorus 2006). We screened all T. major for the presence of bacterial reproductive parasites using primers designed to amplify either a portion of the 16S rDNA (Wolbachia, Arsenophonus, Cardinium, Flavobacterium sp., Spiroplasma) locus or a portion of the 17-kDa gene (Rickettsia) from all known bacterial reproductive parasites (see Table 2 for a list of all primers).

Polymerase chain reactions were carried out in 30-µl reactions containing 200–500 ng of DNA, 0.14 mM of each deoxynucleoside triphosphate, 6 µl of 10× buffer, 3.5 mM of MgCl2, 0.8 mg/ml of bovine serum albumin, 0.15 µM of each primer, 1 unit Taq polymerase, and double distilled water to volume. Thermal profiles for PCR amplification were as follows: denaturation for 2 min at 95 °C, 45 cycles of 95 °C for 30 s, 50–55 °C for 30 s (see Duron et al. 2008 for specific annealing temperatures for each primer pair), and 72 °C for 1 min, followed by a final extension of 72 °C for 5 min. All PCRs included both positive and negative controls. For any symbiont that returned at least 1 positive amplification at 45 cycles, we reran all negative T. major again with 60 cycles of amplification. All positive PCR products were purified using the Wizard SV Gel PCR Prep DNA Purification System (Promega, Madison, WI). To confirm all positives, both strands of the purified PCR product was sequenced using Big Dye 1.1 chain terminators and an ABI 3130 Genetic Analyzer (Applied Biosystems). Contigs were assembled using Geneious v5.1.7 (Drummond et al. 2010). All sequences generated were subjected to a basic alignment search tool (BLAST) search against the National Center for Biotechnology Information database. Once tentative sequence identity was obtained, we used CLUSTALX
Table 2  Loci and primers used to detect bacterial reproductive parasites in *Trichobius major*

<table>
<thead>
<tr>
<th>Organism</th>
<th>Locus</th>
<th>Primer name</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rickettsia</td>
<td>17 kDa</td>
<td>R1</td>
<td>GCCCTTGCAACTTCTAGTT</td>
<td>Werren et al. (1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R2</td>
<td>CATTTGTCTGCAATGTTGCC</td>
<td></td>
</tr>
<tr>
<td>Wolbachia</td>
<td>16S rDNA</td>
<td>16Swolb76-99f</td>
<td>TTTGAGGGCTGCTATGGTTAACT</td>
<td>O’Neil et al. (1992)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16Swolb1012-994r</td>
<td>GAATGGATATGATTTTGAGTA</td>
<td></td>
</tr>
<tr>
<td>Arsenophonus</td>
<td>16S rDNA</td>
<td>ArsF2</td>
<td>CCCCAGCTTAACCTFAGGA</td>
<td>Duron et al. (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ArsR2</td>
<td>GTAAGGCTCTCTGGAAGGCC</td>
<td></td>
</tr>
<tr>
<td>Cardinium</td>
<td>16S rDNA</td>
<td>CLO-r1</td>
<td>GGAACCTACCTGGCTAGAATGTATT</td>
<td>Gotoh et al. (2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLO-r1</td>
<td>GCCACTGCTCTCAAGCTTACCAAC</td>
<td></td>
</tr>
<tr>
<td>Flavobacterium sp.</td>
<td>16S rDNA</td>
<td>FlavF</td>
<td>CGAATTAGTGTGCCCCACTCCG</td>
<td>Duron et al. (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FlavR</td>
<td>CTAACGTRTTCTGAGCCTATCG</td>
<td></td>
</tr>
<tr>
<td>Spiroplasma icestis</td>
<td>16S rDNA</td>
<td>SpixoF</td>
<td>TTAGGGGCTCAACCCCTAACC</td>
<td>Duron et al. (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SpixoR</td>
<td>CTGGCCATTGGAACCTCCTC</td>
<td></td>
</tr>
<tr>
<td>Spiroplasma poulsonii</td>
<td>16S rDNA</td>
<td>SpoulF</td>
<td>GCTTAACCTCAATGTGCC</td>
<td>Duron et al. (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SpoulR</td>
<td>CCTGTCTCAATGTGTTACCCT</td>
<td></td>
</tr>
</tbody>
</table>

(Thompson et al. 1997) to generate multiple sequence alignments for each bacterial genus using all sequences obtained from *T. major* and all homologous sequences available on GenBank (see Figures 6 and 7 for a complete list of sequences). For bacterial clades with reliable outgroups available, we used MrBayes (Huelsenbeck and Ronquist 2001) to perform a Bayesian analysis using the general time reversible + I + G model of nucleotide substitution; values for model parameters were not defined a priori but were treated as unknown variables with uniform priors. The analysis was run for 5,000,000 generations, trees were sampled every 100 generations, and “temperature” was set at 0.02. Resulting burn-in values (the point at which the model parameters and tree scores reached stationarity) were determined empirically by evaluating likelihood scores. All runs were checked for sufficient mixing, stable convergence on a unimodal posterior, and effective sample sizes (Drummond et al. 2002) > 100 for all parameters using TRACER version 1.5 (Drummond and Rambaut 2003). For bacterial clades with no known outgroup available (i.e., the 17-kDa locus for *Rickettsia* is known only from that genus), we used PAUP* (Swofford 2003) to perform maximum parsimony (MP) analyses using equal weighting of substitutions, the heuristic search option, tree bisection–reconnection branch swapping, and 25 random additions of input taxa. Reliability of clades was evaluated by performing 1000 bootstrap iterations, and MP phylogenies were left unrooted.

**Results**

**Trichobius major** Nuclear Diversity

Nuclear AFLP analysis of *T. major* revealed clear differences from the mtDNA analysis of Wilson et al. (2007). While the mtDNA analysis recovered no diversity among populations >700 km apart, the 8 selective primer combinations generated a total of 224 bands that were unambiguously scored, of which 166 (74.11%) were polymorphic, suggesting considerable nuclear diversity. Estimates of heterozygosity and polymorphism as well as the number of bands with a frequency ≥5% was comparable across individual caves (Table 3), suggesting that populations in each cave (and cave system) are large, stable, and likely have not undergone a recent significant bottleneck.

**Trichobius major** Population Structure

The distribution of bands among populations indicated little population structure, with few bands restricted to individual caves or cave systems. The majority of caves had no private bands, while Parker Bat Cave, Kansas, Green’s Cathedral and Skunkater caves in the Selman cave system, and Washita cave 1 each possessed a single private band. At the level of the cave system, private bands were again very few, and no system possessing >2 private bands (Table 3). The standard AMOVA and the Bayesian approach produced different estimates of population subdivision (*Φ*<sub>b</sub> = 0.028 and *θ*<sup>(II)</sup> = 0.465, respectively; Table 4); however, the proportion of variance partitioned among caves and among cave systems was low for both methods (*Φ*<sub>PR</sub> = 0.018, *Φ*<sub>WT</sub> = 0.012, respectively; among caves *θ*<sup>(II)</sup> = 0.015, among-cave system *θ*<sup>(II)</sup> = 0.010). The large difference between *Φ*<sub>PR</sub> and *θ*<sup>(II)</sup> suggests deviation from H-W equilibrium and/or significant correlation of allele frequencies among populations and regions (Song et al. 2006). This was supported in the model comparisons (Table 5), where the full model was significantly favored (DIC = 4916.11) over the *f* = 0 model (DIC = 5003.39), indicating deviation from H-W equilibrium. The full model was also favored over the *θ*<sup>(I)</sup> = 0 model (DIC = 5208.14), indicating significant population structure.

The NJ network showed no geographic clustering pattern among individuals (Figure 2), and no relationships were supported with bootstrap support (≥70%; data not shown). The PCoA of individuals indicated some weak clustering (Figure 3A), with most individuals from northwestern Oklahoma caves and Kansas caves grouping together and with most individuals from central Oklahoma and Texas appearing distinct from all others. However, these groups overlap significantly, suggesting weak isolation with intergradation. The PCoA among caves (Figure 3B)
also suggested significant gene flow among all populations, with no clear relationship among geography and genetic differentiation. When data were averaged across cave systems (Figure 3C), geographical groupings become evident, with northwestern Oklahoma and southwestern Kansas cave systems grouping closely and Washita and Texas cave systems distinct from all others. For the Bayesian clustering analysis conducted in STRUCTURE, the optimal value of K consisted of 4 clusters, but no cluster appeared to be geographically distinct suggesting weak population structure (Figure 4). Admixture and no-admixture models returned essentially identical results, so only results of the admixture analysis are shown. The majority of individuals from northwestern Oklahoma caves and Kansas caves were assigned to a single cluster (blue in Figure 4), the majority of Washita individuals were assigned to a second cluster (green in Figure 4), and individuals assigned to the 2 remaining clusters (yellow and red in Figure 4) are concentrated in the Texas cave, but also occur in all other cave systems. The Mantel test for isolation-by-distance (IBD) returned a significantly positive correlation between population differentiation and geographic distance (r = 0.534, P = 0.031; Figure 5), indicating gene flow among cave systems fits a model of IBD.

Bacterial Reproductive Parasites

PCR surveys for the presence of previously known bacterial reproductive parasites returned positive amplifications for Arsenophonus- and Rickettsia-specific primers only. For Arsenophonus, the infection rate was high, with 67.05% (116) of all individuals positive at 45 amplification cycles and 91.33% (158) positive at 60 amplification cycles. All aligned 16S rDNA sequences were identical in T. major, indicating infection with a single strain of Arsenophonus (GenBank accession JN561339). The top results of the BLAST search confirmed that the amplified fragment was 16S rDNA from the Arsenophonus group of bacterial endosymbionts. The alignment of all available Arsenophonus 16S rDNA sequences resulted in 625 aligned positions, and the Bayesian phylogenetic analysis (Figure 6) placed the T. major endosymbiont sequence basal in a clade containing Arsenophonus isolated from neotropical Trichobius and the hippoboscid Lipeocera cervi (Bayesian Posterior Probability

Table 3  Mean heterozygosity ($H_e$) and descriptive statistics of Trichobius major AFLP

<table>
<thead>
<tr>
<th>Cave system/population</th>
<th>n</th>
<th>Mean heterozygosity ($H_e$)</th>
<th>95% CI ($H_e$)</th>
<th>No. private alleles</th>
<th>Polymorphic (%)</th>
<th>LCB LCB LCB ≤ ≤ ≤ 25% 50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kansas</td>
<td>58</td>
<td>0.205</td>
<td>0.191–0.217</td>
<td>1</td>
<td>67.4</td>
<td>194</td>
</tr>
<tr>
<td>Double Entrance</td>
<td>22</td>
<td>0.203</td>
<td>0.188–0.215</td>
<td>0</td>
<td>49.1</td>
<td>181</td>
</tr>
<tr>
<td>Parker</td>
<td>8</td>
<td>0.205</td>
<td>0.189–0.218</td>
<td>1</td>
<td>34.8</td>
<td>177</td>
</tr>
<tr>
<td>Sink Valley</td>
<td>19</td>
<td>0.210</td>
<td>0.195–0.222</td>
<td>0</td>
<td>56.7</td>
<td>206</td>
</tr>
<tr>
<td>Swartz</td>
<td>9</td>
<td>0.205</td>
<td>0.189–0.217</td>
<td>0</td>
<td>32.6</td>
<td>179</td>
</tr>
<tr>
<td>Alabaster</td>
<td>40</td>
<td>0.202</td>
<td>0.188–0.214</td>
<td>0</td>
<td>54.0</td>
<td>189</td>
</tr>
<tr>
<td>Alabaster Cavern</td>
<td>20</td>
<td>0.204</td>
<td>0.190–0.217</td>
<td>0</td>
<td>46.9</td>
<td>195</td>
</tr>
<tr>
<td>Bear Cave</td>
<td>20</td>
<td>0.201</td>
<td>0.186–0.214</td>
<td>0</td>
<td>43.8</td>
<td>191</td>
</tr>
<tr>
<td>Selman</td>
<td>39</td>
<td>0.205</td>
<td>0.192–0.218</td>
<td>2</td>
<td>59.4</td>
<td>196</td>
</tr>
<tr>
<td>Green's Cathedral</td>
<td>28</td>
<td>0.202</td>
<td>0.187–0.215</td>
<td>1</td>
<td>52.7</td>
<td>183</td>
</tr>
<tr>
<td>Skunkfeather</td>
<td>4</td>
<td>0.208</td>
<td>0.193–0.221</td>
<td>1</td>
<td>33.0</td>
<td>176</td>
</tr>
<tr>
<td>Skylight</td>
<td>7</td>
<td>0.209</td>
<td>0.195–0.221</td>
<td>0</td>
<td>39.7</td>
<td>188</td>
</tr>
<tr>
<td>Washita</td>
<td>21</td>
<td>0.207</td>
<td>0.193–0.219</td>
<td>1</td>
<td>52.2</td>
<td>180</td>
</tr>
<tr>
<td>Cave 1</td>
<td>7</td>
<td>0.205</td>
<td>0.193–0.220</td>
<td>1</td>
<td>43.3</td>
<td>184</td>
</tr>
<tr>
<td>Cave 2</td>
<td>14</td>
<td>0.208</td>
<td>0.195–0.221</td>
<td>0</td>
<td>54.6</td>
<td>193</td>
</tr>
<tr>
<td>Austin, TX</td>
<td>15</td>
<td>0.209</td>
<td>0.195–0.221</td>
<td>1</td>
<td>46.4</td>
<td>195</td>
</tr>
<tr>
<td>Total</td>
<td>173</td>
<td>0.209</td>
<td>0.195–0.219</td>
<td>—</td>
<td>74.1</td>
<td>224</td>
</tr>
</tbody>
</table>

CI, confidence interval; LCB, locally common bands. Private alleles are those found only in a single population or cave system. LCB refer to those at a frequency ≥5%, and LCB ≤25% and ≤50% refers to locally common bands found in ≤25% and ≤50% of the populations, respectively. Cave systems from which multiple caves were sampled are italicized.

Table 4  Standard and Bayesian AMOVA results illustrating the partitioning of variance within and among populations of Trichobius major and among cave systems

<table>
<thead>
<tr>
<th>Partition</th>
<th>df</th>
<th>Percentage</th>
<th>Standard AMOVA</th>
<th>Bayesian AMOVA (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among cave systems</td>
<td>4</td>
<td>100</td>
<td>ΦRT = 0.012 (P = 0.003)</td>
<td>Among-cave system θ(II) = 0.010 (0.006–0.014)</td>
</tr>
<tr>
<td>Among populations</td>
<td>6</td>
<td>100</td>
<td>ΦRT = 0.018 (P = 0.001)</td>
<td>Among-population θ(I) = 0.015 (0.010–0.021)</td>
</tr>
<tr>
<td>Within populations</td>
<td>162</td>
<td>97</td>
<td>ΦRT = 0.030 (P &lt; 0.001)</td>
<td>θ(I) = 0.465 (0.415–0.521)</td>
</tr>
</tbody>
</table>

df, degrees of freedom; CI, confidence interval. Bayesian parameter estimates are those from an analysis using the full model (see Table 5 for model comparison). For the standard AMOVA, ΦRT is the dominant data analogue of Wright’s FST, ΦPC corresponds to the proportion of variance partitioned among regions, and ΦPR corresponds to the proportion of variance partitioned among populations within regions. For the Bayesian parameters, θ(I) corresponds directly to Wright’s FST and θ(II) refers to the proportion of variance partitioned among populations.
of the 17-kDa locus from search confirmed that the amplified fragments were a portion differed at 3 of 435 aligned positions (0.7%), and the BLAST accessions JN561340 and JN561341). These 2 sequences Green's Cathedral in the Selman cave system (GenBank the Alabaster cave system and a single individual from (1.16% infection rate); a single individual from Bear Cave in other
Arsenophonus isolated from Trichobius. Also, Arsenophonus isolated from the hippoboscid Lipoptena were nested within the Trichobius clade with high support (Bayesian PP = 0.95). The Bayesian phylogeny was rooted with Xenorhabdus and Photorhabdus as outgroups as these have been shown to be closely related to Arsenophonus (Trowbridge et al. 2005; Novakova et al. 2009).

We detected positive PCRs for Rickettsia in 2 T. major (1.16% infection rate); a single individual from Bear Cave in the Alabaster cave system and a single individual from Green's Cathedral in the Selman cave system (GenBank accessions JN561340 and JN561341). These 2 sequences differed at 3 of 435 aligned positions (0.7%), and the BLAST search confirmed that the amplified fragments were a portion of the 17-kDa locus from Rickettsia. Because the 17-kDa locus is found only in Rickettsia, no outgroup for the genus is available, and the MP phylogeny (Figure 7) is unrooted. The 2 Rickettsia sequences isolated from T. major were sister taxa in the MP analysis (MP bootstrap [MPBS] = 100%), but their position within Rickettsia could not be resolved with the 17-kDa marker.

Discussion
Under neutral evolution, differences in genetic diversity between mitochondrial and nuclear genomes are expected due to the differences in effective population size and mutation rates (Kimura 1983). These expectations are well understood, and models exist to differentiate between divergence under neutral and nonneutral evolution between 2 loci with different biological properties (e.g., differences in effective population size; Hudson et al. 1987). The dominant nature of AFLP loci prohibits such comparisons between the nuclear data generated here and the mtDNA data of Wilson et al. (2007), but the clear differences between the 2 data sets suggest different demographic histories between the mitochondrial and nuclear genomes. We are aware of only 2 previous cases where a complete lack of mtDNA diversity was detected at such a large geographic scale—Araea encedana butterflies from Uganda (Jiggins 2003) and Drosophila simulans in California (Turelli et al. 1992)—although other instances undoubtedly exist. In both of these species, bacterial reproductive parasites were identified as the causative agent behind the loss of mtDNA diversity through linkage disequilibrium between the infected mtDNA lineage and the bacterial parasite, and we suggest a similar mechanism for T. major.

**Trichobius major** Dispersal and Gene Flow
The clumsy flight exhibited by T. major, where individuals fly in a spiraling motion toward the cave ceiling, is likely adapted for host searching and movement within a cave but is not conducive to independent dispersal among caves (Caire and Hornuff 1982). While no study has directly examined the ability of T. major to self-disperse, Overal (1980) performed an extensive study on the biology of T. major in southern Kansas and never observed T. major outside the caves that contain their hosts. In addition, T. major requires regular blood meals, and when removed from bats do not survive more than a day (W. Caire, personal communication), which is consistent for other streblid bat flies (Dick and Patterson 2006). With their weak flight and need for frequent blood meals, it is highly unlikely T. major could self-disperse among colonies of roosting M. velifer to the extent that the AFLP data presented here suggest.

For obligate ectoparasites, dispersal via the host and phylogeographical concordance between the host and parasite is common (Nieberding and Olivieri 2006). For T. major, we found little population structure and evidence for high amounts of gene flow among all sampled populations. The Bayesian AMOVA indicated significant correlation of allele frequencies among populations and very little variation partitioned among individual caves or cave systems (Table 4). When the data were analyzed without

---

**Table 5** Parameters for DIC model comparison for the Bayesian AMOVA conducted in Hickory v1.1

<table>
<thead>
<tr>
<th>Model</th>
<th>Dbar</th>
<th>Dhat</th>
<th>pD</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>4556.49</td>
<td>4196.88</td>
<td>359.62</td>
<td>4916.11</td>
</tr>
<tr>
<td>$f = 0$</td>
<td>4551.70</td>
<td>4100.01</td>
<td>451.69</td>
<td>5003.39</td>
</tr>
<tr>
<td>$\theta^{(5)} = 0$</td>
<td>5059.50</td>
<td>4910.86</td>
<td>148.64</td>
<td>5208.14</td>
</tr>
</tbody>
</table>

**Figure 2.** Neighbor-joining network of individual *Trichobius major* based on 166 polymorphic AFLP loci. Branch color corresponds to the cave system from which the individuals were collected.
combining individuals a priori—as in the NJ network (Figure 2), the STRUCTURE analysis (Figure 4), and the PCoA of individuals (Figure 3A)—we see a high dispersion of the data and little structure. Similarly, the PCoA of individual caves (Figure 3B) suggests little structure, with the Skylight and Skunkfeather caves from the Selman system in northwestern Oklahoma and the Sink Valley Cave from southern Kansas clustering more closely with the Washita and Texas caves than with geographically more proximate caves in northwestern Oklahoma or Kansas. When data are grouped according to cave system (Figure 3C), there is geographic clustering, and the Mantel test (Figure 5) indicates gene flow among cave systems fits an IBD model. These patterns suggest high levels of gene flow among even the most geographically distant sampled populations and that *T. major* is likely dispersing via its bat host.

Our sampling regime for *T. major* includes 2 host subspecies and >700 km between the most distant caves. The individuals collected from the Eckert James Bat Cave in Austin, TX, occur on *M. v. incautus* while all other sampled *T. major* populations were taken from hibernating colonies of *M. v. grandis*. These subspecies are differentiated by morphology, with the primary distinction being pelage color and size, and have a distinct distribution gap between them (Hayward 1970; Fitch et al. 1981; see Figure 1 for *M. velifer* subspecies distributions). The STRUCTURE analysis (Figure 4) and PCoA of individuals (Figure 3A) did not reveal any clear distinction of Texas *T. major* from other populations, and the PCoA of cave systems (Figure 3C) indicated the Texas population is no more differentiated from the northwestern Oklahoma and Kansas caves than the Washita cave system is. The overall lack of structure among all sampled populations of *T. major* is congruent with the phylogeographical patterns detected for its host, *M. velifer* (Parlos 2008). Utilizing both mtDNA and nuclear markers, Parlos (2008) detected high levels of gene flow for *M. velifer* among all the recognized subspecies, suggesting that significant population mixing is occurring even among populations from California to Oklahoma. This is likely due to seasonal movements between maternity roosts, summer roosts, and hibernacula (Hayward 1970; Fitch et al. 1981). Our AFLP examination of *T. major* is highly congruent with these results, providing strong support for host-mediated dispersal in *T. major*.

**Figure 3.** PCoA of genetic variation at 166 polymorphic AFLP loci from *Trichobius major*. (A) Genetic structuring among the 173 individuals from 12 cave localities. The variance explained by coordinate 1 (Coord. 1) and coordinate 2 (Coord. 2) was 24.40% and 21.12%, respectively. (B) Genetic structuring among the 12 caves sampled. The variance explained by Coord. 1 and Coord. 2 was 62.27% and 13.39%, respectively. (C) Genetic structuring among the 5 cave systems sampled. The variance explained by Coord. 1 and Coord. 2 was 68.81% and 19.76%, respectively.
Multiple lines of evidence from the nuclear data presented here provide support for the role of a selective sweep in reducing mtDNA diversity in *T. major*. Demographic events such as bottlenecks and founder effects can produce the mtDNA patterns observed by Wilson et al. (2007). Infection by a bacterial reproductive parasite that skews reproductive success toward infected female lineages will also produce this reduction in mtDNA diversity through an indirect selective sweep on the infected maternal lineage (Maynard-Smith and Haigh 1974), and differentiating between demographic processes and a selective sweep is essentially impossible with mtDNA alone (Tajima 1989). Adding nuclear data allows for differentiation between these 2 processes because demographic events will impact both the mitochondrial and nuclear genomes proportionally, while an mtDNA selective sweep will reduce mtDNA diversity much more drastically than nuclear diversity. This is consistent with the pattern observed for *T. major*. Estimates of nuclear AFLP heterozygosity and polymorphism from all sampled populations were consistent across populations (\(H_e = 0.201–0.210\), polymorphism = \(32.6–56.7%\); Table 3) and moderate to high relative to other studies utilizing AFLP analysis to examine insect nuclear DNA diversity (Grapputo et al. 2005; Sei and Porter 2007; Simonato et al. 2007). In addition, the *T. major* samples analyzed here are the same individuals (although we have expanded the sample sizes for each population) used in the mtDNA study of Wilson et al. (2007), ruling out any temporal or sampling effects as an explanation for the difference in genetic diversity between the mitochondrial and nuclear genomes of *T. major*.

The significant difference between the full model and the \(f = 0\) model indicates *T. major* populations are not in H-W equilibrium (Table 5). Several factors can cause populations to fall out of H-W equilibrium, such as inbreeding and selection. For *T. major*, inbreeding is unlikely, as populations appear to be large (Overal 1980; Caire and Hornuff 1986) and possess moderate levels of nuclear diversity (Table 3). However, a recent mtDNA selective sweep would result in a reduction in the effective population size of *T. major*. The most common mechanism of reproductive manipulation is cytoplasmic incompatibility (Hurst and Jiggins 2005). Under this type of manipulation, only copulation between infected individuals will result in offspring. Therefore, successful reproduction will deviate significantly from random, and infected populations are likely to fall out of H-W equilibrium. While we do not know the mechanism of reproductive manipulation in *T. major*, the cytoplasmic incompatibility phenotype results in a nearly 100% infection rate (Stouthamer et al. 1999) as was observed for *Arsenophonus* in *T. major*. Also, because cytoplasmic incompatibility results in the selective killing of uninfected individuals, it can spread rapidly across large geographic distances, even in relatively structured populations (Hurst and Jiggins 2005). The geographically widespread...
Arsenophonus infection in *T. major* (infection with the same strain across two host subspecies and >700 km) suggests cytoplasmic incompatibility as the most likely mechanism. However, it is purely speculative as to whether *Arsenophonus* is the bacterium responsible for the *T. major* mtDNA sweep. Due to the often transient nature of reproductive endosymbiont infection (Hurst and Jiggins 2005), it is quite possible that the bacterium responsible for the *T. major* mtDNA selective sweep is no longer present, and detailed analyses of offspring sex ratios and endosymbiont are needed.

**Bacterial Reproductive Parasites of *T. major***

*Arsenophonus* is a clade of approximately 100 known symbiotic bacteria broadly associated with insects (Novakova et al. 2009). The only species known to act as a reproductive parasite is *Arsenophonus nasoniae*, a male-killing symbiont of the wasp *Nasonia vitripennis*. Essentially nothing is known of the other *Arsenophonus*, and the majority are thought to be primary and secondary endosymbionts with no negative effects (Novakova et al. 2009). Phylogenetic analysis of the 16S rDNA fragment from the *Arsenophonus* strain isolated from *T. major* and all available *Arsenophonus* sequences yielded little in terms of phylogenetic resolution, with the exception of the *Trichobius* clade (Figure 6).

The clade containing the sequence isolated from *T. major*, all other *Trichobius*, and the deer fly *L. cervi*—also a member of Hippoboscoidea—was strongly supported (PP 1.0). Furthermore, the position of *L. cervi* basal to all other *Trichobius*-derived *Arsenophonus*, with the exception of *T. major*, was also supported.

In a previous phylogenetic analysis of *Arsenophonus*, sequences isolated from *Trichobius* (not including an isolate from *T. major*) formed a monophyletic clade with relationships congruent with the host phylogeny (Novakova et al. 2009). This is in contrast to the phylogenetic pattern recovered for *Wolbachia* (Shoemaker et al. 2002), where there is a general lack of congruence between the *Wolbachia* and host phylogenies. The lack of host–parasite phylogenetic congruence for *Wolbachia* has been explained by the nature of *Wolbachia* association with its hosts, where infections are short lived, new infections are common, and infections likely do not span the time required for speciation to occur (Hurst and Jiggins 2005). The *Arsenophonus* strain isolated from *T. major* does not fall within the *Trichobius* clade, but is basal to a clade containing *Arsenophonus* isolated from both *Trichobius* and *Lipoptena*, and therefore is not congruent with the corresponding insect phylogeny. In a similar analysis of *Arsenophonus* symbionts isolated from white flies (Aleyrodidae), Thao and Baumann (2004) documented a lack.

---

**Figure 6.** Bayesian 50% majority rule consensus phylogram for a fragment of the 16S rDNA locus from *Arsenophonus* isolated from *Trichobius major* as well as all available sequences deposited in GenBank. The phylogeny was rooted with homologous sequences obtained for *Xenorhabdus* and *Photorhabdus*. Numbers at nodes are the Bayesian posterior probabilities. Tip labels refer to the *Arsenophonus* species, the host organism from which the bacterium was isolated in parentheses, and the accession number for the sequence.
of congruence between the Arsenophonus and host phylogenies, similarly suggesting frequent horizontal transfer and a more dynamic relationship between Arsenophonus and its host species than previously thought. For Trichobius, we acknowledge that the sparse sampling prohibits any concrete conclusions concerning Arsenophonus/Trichobius co-cladogenesis, but we can say that the Arsenophonus phylogeny (Figure 6) does not contradict the prediction that bacterial reproductive parasites are less likely to co-speciate with their hosts than bacterial mutualists.

In addition to Arsenophonus, we detected Rickettsia in 2 T. major from northwestern Oklahoma. Rickettsia are notorious as insect-borne pathogens of vertebrates (Raoult and Roux 1997) but have also been identified as reproductive parasites in at least 4 hosts (Perlman et al. 2006; Duron et al. 2008). Rickettsia has traditionally been divided into 2 groups—the typhus group and the spotted fever group (Weiss and Moulder 1984)—but with the discovery of new Rickettsia and the addition of phylogenetic studies, 6 Rickettsia species groups are now recognized (Roux and Raoult 2000). The phylogenetic analysis of the 17-kDa fragment from all available Rickettsia and the 2 sequences obtained from T. major provided little resolution, but any clades where statistical resolution was obtained (≥70% MPBS; Figure 7) were congruent with the currently recognized species groups (Roux and Raoult 2000; Perlman et al. 2006; Vitorino et al. 2007). The Rickettsia isolates we obtained from T. major were sister taxa with high support (MPBS = 100%), but the position of these isolates within Rickettsia was unresolved. In spite of this, the Rickettsia isolated from T. major likely represent a previously unknown strain and should be examined for potential disease implications. In terms of reproductive manipulation of T. major, the low infection rate (1.16%) and detection only in northwestern Oklahoma suggests the Rickettsia isolated from T. major are likely not responsible for the observed mtDNA selective sweep. However, the transient nature of reproductive parasite infection means a low contemporary infection rate, or even the absence of infection, does not rule out prior infection.

Conclusions

The significance of bacterial symbionts in arthropod evolution is only beginning to be understood. While the majority of cases point to reproductive manipulation altering standing mtDNA diversity within a species, bacterial symbionts have more recently been implicated in homogenizing mtDNA among species through the introgressive transfer of symbionts (Shaw 2002; Jiggins 2003) and driving divergence (and potentially speciation) through the incompatibility of parallel infections (James and Ballard 2000; Miller et al. 2010; Sharon et al. 2010). We present here evidence for a widespread mtDNA selective sweep in T. major, and evidence for infection with members of at least two known clades of bacterial reproductive parasites. Further work is necessary to identify the bacterial (or other)
agent responsible for the purge in mtDNA diversity, but unequivocal identification may not be possible. As mentioned previously, infection with reproductive parasites is typically transient, and it is possible that standing mtDNA diversity may be the result of a past infection by a bacterium no longer present in the host (Shoemaker et al. 2002; Hurst and Jiggins 2005). On these grounds, we acknowledge that the detection of Arsenophonus and Rickettsia in *T. major* is anecdotal evidence. Nonetheless, the presence of Arsenophonus across many insect groups (Duron et al. 2008; Novakova et al. 2009) and its known role in reproductive manipulation suggests Arsenophonus may be a clade of reproductive parasites responsible for widespread manipulation and warrants further investigation.

Population analyses of nuclear AFLP data from *T. major* illustrate the point articulated in the literature that, while the results of mtDNA-only studies should be approached with caution (Ballard and Whitlock 2004; Galtier et al. 2009), this is especially true for studies of insect phylogeography (and possibly all arthropods), where there is a relatively high incidence of bacterial reproductive parasites that manipulate host genealogies. The AFLP analysis of *T. major* indicates substantial nuclear diversity and high levels of gene flow among populations >700 km apart. In addition, our results are congruent with host *M. velifer* phyleogeographical patterns, suggesting dispersal in *T. major*, a relatively poor flier that is never seen outside of caves without its bat host, is occurring via dispersal and seasonal movements of its host.

**Funding**

National Science Foundation (REU supplement to NSF Award DEB-0610844 to R.A.V.D.B.); University of Central Oklahoma, Dr Joe C. Jackson College of Graduate Studies and Research Faculty Grant (to G.M.W).

**Acknowledgments**

We thank William Caie, Kendra Dougherty, Ty Reidenbaugh, Kevin Pagerter, and Monte and Kathleen Thies, who helped to collect bat flies in Oklahoma and Texas. We also thank Mrs Betty Selman and the Selman Living Laboratory, and the University of Central Oklahoma for allowing access to the Selman Cave system in Woodward County, OK. Doug Fransen allowed access to the caves in Blaine County, OK. The Nature Conservancy allowed access to the Eckert James River Bat Cave Preserve in Mason County, TX. In addition, we are grateful to Stan Roth who assisted in locating and collecting samples from 4 caves in Comanche County, KS.

**References**


Received March 16, 2011; Revised June 01, 2011; Accepted July 22, 2011

Corresponding Editor: Adalgisa Caccone