Rapid #: -12139849

CROSS REF ID: **730944**

LENDER: **ORE :: Main Library**

BORROWER: **MUM :: Main Library**

TYPE:	Article CC:CCG
JOURNAL TITLE:	Journal of biogeography
USER JOURNAL TITLE:	Journal of biogeography
ARTICLE TITLE:	Strong spatial-genetic congruence between a wood-feeding cockroach and its bacterial endosymbiont, across a topographically complex landscape
ARTICLE AUTHOR:	Garrick, Ryan
VOLUME:	44
ISSUE:	7
MONTH:	
YEAR:	2017
PAGES:	1500-1511
ISSN:	0305-0270
OCLC #:	

Processed by RapidX: 8/4/2017 12:26:52 PM

Journal of Biogeography (J. Biogeogr.) (2017) 44, 1500–1511

ORIGINAL ARTICLE

Strong spatial-genetic congruence between a wood-feeding cockroach and its bacterial endosymbiont, across a topographically complex landscape

Ryan C. Garrick[1]*, Zakee L. Sabree[2†], Benjamin C. Jahnes[2,3] and Jeffrey C. Oliver [4]

[1]Department of Biology, University of Mississippi, Oxford, MS 38677, USA, [2]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH 43210, USA, [3]Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA, [4]University of Arizona Health Sciences Library, University of Arizona, Tucson, AZ 85724, USA

*Correspondence: Ryan C. Garrick, Department of Biology, University of Mississippi, Oxford, MS 38677, USA.
E-mail: rgarrick@olemiss.edu
†Zakee L. Sabree, Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH 43210, USA.
E-mail: sabree.8@osu.edu

ABSTRACT

Aim We analysed data from an insect host (*Cryptocercus punctulatus*) and its maternally-inherited obligate bacterial endosymbiont (*Blattabacterium cuenoti* str. *punctulatus*) to address the following: (1) to what extent do these species exhibit cophylogenetic structure, (2) do the spatial-genetic structures of these species differ, and (3) what is the relative importance of codivergence versus other events in explaining congruence, or instances of incongruence, between their molecular phylogenies?

Location The southern Appalachian Mountains, USA.

Methods We conducted fine-scale population-level sampling and screening of DNA sequence variation in two mitochondrial genes from the host, and four genic or intergenic regions from the endosymbiont. Inferences were made using analyses that have the potential to identify isolated instances of cophylogenetic discord, uncover subtle differences in geographic locations of genetic discontinuities, and disentangle different evolutionary processes that contributed to observed patterns.

Results The host and its endosymbiont showed similar phylogenetic and geographic patterns. Cophylogenetic analyses revealed that while topological discord is rare (and restricted within major clades), some instances are potentially non-negligible. Assessments of spatial-genetic structure showed that most abrupt breaks occur in the same locations, but they differ in strength, again underscoring some subtle discordance. The main process generating observed patterns was inferred to be codivergence due to host-tracking; however, incomplete lineage sorting seems likely to have also played a minor role.

Main conclusions Our overarching finding of strong congruence is reflected by broader-scale cophylogenetic studies of related *Cryptocercus* and *Blattabacterium* taxa. Accordingly, we suggest that members of this symbiosis may provide an excellent opportunity for investigating geographic scaling of processes that affect biogeographic patterns. However, fine-scale sampling coupled with geospatial analyses detected rare and/or minor discordances that appeared to be localized within the most deeply dissected topographic regions of the southern Appalachian Mountains, and these warrant further exploration.

Keywords
A ppalachian Mountains, *Blattabacterium*, codivergence, cophylogeny, *Cryptocercus*, endosymbiosis, genetic structure, host tracking

http://wileyonlinelibrary.com/journal/jbi
doi:10.1111/jbi.12992

Journal of Biogeography

INTRODUCTION

Biogeographic studies focus on identifying geographic patterns in the distributions of lineages, and molecular phylogenies have led to advances in understanding the relative importance of dispersal and vicariance in the formation of ecological guilds or communities (Riddle *et al.*, 2008; Losos & Ricklefs, 2009; Gillespie *et al.*, 2012). Similarly, the use of genetic data to infer evolutionary relationships among lineages of mutualists has provided insights into the antiquity and strength of these co-associations, including the roles of host-switching and codivergence in shaping the present-day patterns of diversity among interacting taxa (Page & Charleston, 1998). Indeed, organismal co-associations that traverse different hierarchical levels of organization (e.g. taxonomic: genus, species, and population; geographic: continent, region, and local site) may be particularly informative for understanding scaling of macro- and micro-evolutionary processes (Araújo & Rozenfeld, 2014 and references therein).

Mutualisms between insects and microbes are taxonomically widespread, with many well-known cases involving insect hosts whose primary food sources are deficient in essential nutrients. For example, phloem-feeding aphids (Hemiptera) are provisioned with assimilable essential amino acids by their intracellular *Buchnera*. In exchange, the endosymbiont is the primary inhabitant of protective host tissues (i.e. bacteriocytes) and receives host-derived nutrients (Moran *et al.*, 2008). Microbial mutualists also enable insects to specialize on food sources, with wood-feeding insects exhibiting considerable diversity in such relationships. For example, lower termites (Blattodea) maintain obligate associations with hindgut flagellates and bacteria that fix atmospheric nitrogen and aid in the breakdown of cellulose (Brune, 2014), and passalid beetles (Coleoptera) harbour gut-inhabiting yeasts that ferment xylose (Suh *et al.*, 2003). Indeed, many insect-microbe associations are obligate, species-specific, involve vertical transfer of host tissue-restricted symbionts, and have been stable over millions of years (Moran *et al.*, 2008; Duron & Hurst, 2013). These characteristics make insect-microbe species pairs interesting case studies for examining the fidelity of codivergence. Recent advances in cophylogenetic analyses have expanded the scope and complexity of questions that can be addressed (e.g. Conow *et al.*, 2010), and have also enabled assessments of the significance of cophylogenetic structure globally (across large trees, spanning numerous taxonomic levels) and locally (among specific taxon pairs within trees; e.g. Meier-Kolthoff *et al.*, 2007). Despite these advances, however, these methods do not explicitly incorporate geographic information. Accordingly, in this study, we explore ways to integrate geospatial analyses into the assessment of cophylogenetic structure between an insect host and its microbial endosymbiont.

Topographically complex montane regions may be particularly well-suited to assessing spatial variability in the strength of cophylogenetic structure between insect hosts their symbionts. This suitability is due to gene flow limitation among populations being common in montane settings, particularly in low mobility taxa (Garrick, 2011). Thus, marked phylogeographic structuring should facilitate historical reconstruction (Cruzan & Templeton, 2000), including assessments of congruence. The southern Appalachian Mountains represent some of the oldest uplands in North America, and the region's steep environmental gradients have promoted population subdivision within many terrestrial invertebrates (e.g. Thomas & Hedin, 2008; Walker *et al.*, 2009). Among the invertebrate detritivores in the southern Appalachian Mountains, wood-feeding *Cryptocercus* cockroaches play major roles in the breakdown of coarse woody debris (Ulyshen, 2016). Although the taxonomic status of southern Appalachian *Cryptocercus* lineages remains uncertain, the group has several characteristics that make them interesting models for understanding divergence processes. For example, this taxon is comprised of four chromosomal races, they are locally allopatric, and they are genetically differentiated at both mitochondrial and nuclear loci (Luykx, 1983; Kambhampati *et al.*, 1996; Burnside *et al.*, 1999; Nalepa *et al.*, 2002; Everaerts *et al.*, 2008). However, formal assessments of reproductive isolation and ecological divergence are still needed, and so we collectively refer to southern Appalachian lineages as "*Cryptocercus punctulatus*" herein.

Codivergence between *Cryptocercus* wood-roaches and *Blattabacterium cuenoti*, an obligate intracellular bacterial mutualist that resides solely in host abdominal fat bodies and is essential for host growth and reproduction, has previously been explored above and below the species level (Clark *et al.*, 2001; Clark & Kambhampati, 2003; Lo *et al.*, 2003; Maekawa *et al.*, 2005). Given that *B. cuenoti* is vertically transmitted through the matriline in cockroaches (Brooks, 1970), strong congruence between phylogenies derived from host mitochondrial DNA and those from bacterial genes is expected. Indeed, based on direct comparison of molecular phylogenetic trees, Clark *et al.* (2001) detected strong overall similarity, albeit from a sample of only 16 host-symbiont pairs. Interestingly, those authors found that within southern Appalachian lineages of *Cryptocercus*, two localized instances of topological conflict between host and endosymbiont were evident. Furthermore, one of these instances had high bootstrap support, suggesting the conflict was not due to phylogenetic uncertainty. This topological conflict suggests that processes other than codivergence may be operating.

The *C. punctulatus*/*B. cuenoti* str. *punctulatus* ("*B. cuenoti*" herein) species pair is well-suited to investigating whether codivergence alone is sufficient for explaining the distribution of genetic variation in a maternally-inherited bacterial endosymbiont, or alternatively, whether duplication or lineage sorting/loss may be more prevalent than is currently recognized. As the southern Appalachian Mountains have a long history of climate-driven vicariance (Thomas & Hedin, 2008; Walker *et al.*, 2009), thereby providing repeated opportunities for lineage splitting in the host (Nalepa *et al.*, 2002), the *C. punctulatus*/*B. cuenoti* species pair is also suitable for assessing geographic variability in the strength of congruence in genetic structure.

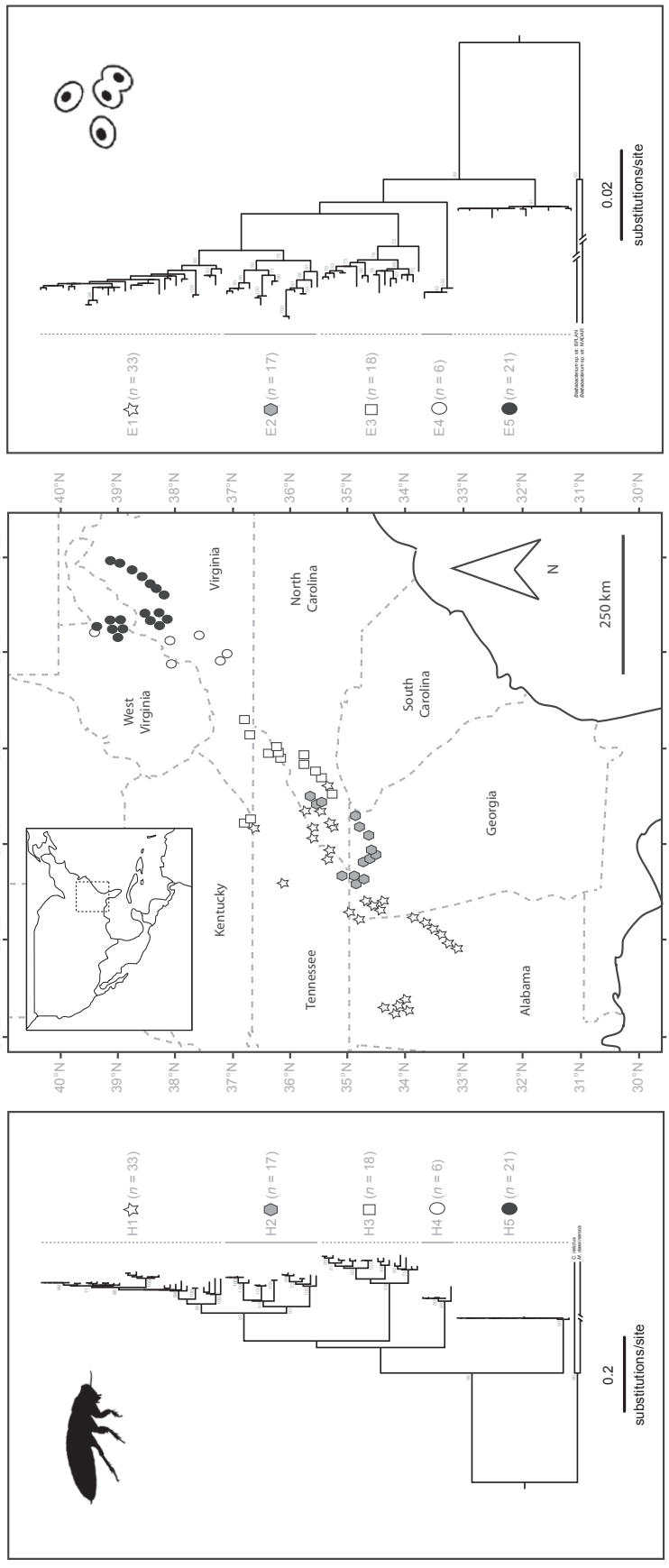

Figure 1 Maximum likelihood phylogenetic trees showing relationships among five major clades identified for *Cryptocercus punctulatus* (left panel) and *Blattabacterium cuenoti* str. *punctulatus* (right panel), and their geographic distributions (middle panel, with inset showing location of study area within North America). Numbers on nodes are bootstrap values (only those ≥ 70% are shown). Each clade is labelled 1–5 with the prefix "H" (host) or "E" (endosymbiont), and sample sizes (*n*) are given in parentheses. Symbols for each clade are used to denote their spatial distribution on the map, where US states are labelled and demarcated by dashed grey lines.

In the present study, the level of congruence between *C. punctulatus* and *B. cuenoti* was investigated using dense geographic sampling, newly-developed genetic markers, and a complementary set of cophylogenetic and geospatial analyses. We address the following questions: (1) to what extent do *C. punctulatus* and *B. cuenoti* exhibit cophylogenetic structure, (2) do their spatial-genetic structures differ, and (3) what is the relative importance of codivergence versus other events in explaining congruence (or incongruence) between the two species' molecular phylogenies? The first two questions can be broadly categorized as relating to pattern, whereas the latter specifically addresses underlying processes.

MATERIALS AND METHODS

Study species and population sampling

The taxonomic status of southern Appalachian *Cryptocercus* is contentious. Briefly, Kambhampati *et al.* (1996) first reported the existence of four chromosomal races (XO ♂ karyotypes: 2n = 37, 39, 43 or 45). Burnside *et al.* (1999) subdivided the taxon into four species, but the basis for this has been repeatedly questioned (Nalepa *et al.*, 2002; Lo *et al.*, 2006; Everaerts *et al.*, 2008). Herein, we simply refer to them as the *C. punctulatus* species complex.

Blattabacterium cuenoti is obligately intracellular, resides solely in host fat body tissues, and is strictly maternally transmitted to offspring during embryonic development (Brooks, 1970). This bacterium has a highly reduced genome size (632 Kb), having lost many genes involved in the synthesis of essential amino acids (Neef *et al.*, 2011). The symbiosis between *B. cuenoti* and its host is estimated to have established > 90 Ma (Patiño-Navarrete *et al.*, 2013).

From 2012–2014, *C. punctulatus* were collected from 95 rotting logs spanning the southern Appalachian Mountains and surrounding areas (Fig. 1; also see Appendix Table S1.1 in Supporting Information). This sampling encompassed seven eastern US states, and covered a broad elevational range (179–1714 m) and many different forest types. One adult cockroach per log was used for subsequent analyses.

DNA isolation and genetic markers

Host DNA was obtained from a rear leg and extracted using a DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's recommendations. Sections of the mitochondrial *cytochrome c oxidase subunit I* (*COI*) and *subunit II* (*COII*) genes were amplified via polymerase chain reaction (PCR). Bacterial endosymbiont DNA was obtained by dissecting the host's abdominal segments and then extracting from the fatty tissue. A portion of the 16S rRNA gene was targeted, and three additional loci (*CP*50, *CP*63 and *CP*78), each based on a different intergenic region (IGR), were developed using genomic resources available in GenBank (see Appendix S2 for details). All loci were amplified using primers reported in Table S2.2 in Appendix S2.

Sequencing, alignment and genetic marker polymorphism

PCR products were purified using ExoSAP-IT® (Affymetrix) and sequenced on an ABI 3730 Genetic Analyser at Yale University. Chromatograms were edited and aligned in MEGA 6.06 (Tamura *et al.*, 2013). Alignment of host *COI* and *COII* and of bacterial 16S rRNA and *CP*78 was unambiguous. Two bacterial IGR loci (*CP*50 and *CP*63) contained several repetitive regions with insertion-deletion (indel) mutations, and manual adjustments were necessary. Following alignment of sequences from each of locus, nucleotide composition and polymorphism were characterized using standard summary statistics (see Appendix S3 for details). Individual-based GenBank accession numbers for DNA sequences from each host and bacterial isolate are given in Table S4.4 in Appendix S4.

Phylogenetic tree estimation

For each species, a concatenated dataset of all DNA regions was used (gene tree/species tree methods were not used because both host genes are effectively part of the same locus, and for the endosymbiont, free recombination between loci cannot be assumed). To find the optimal partition scheme, searches were performed with PARTITIONFINDER 1.1.1 (Lanfear *et al.*, 2012) using the greedy algorithm, linked branch lengths, and model selection based on the Bayesian information criterion. For the host data, we provided six possible partitions, corresponding to each codon position for the two mtDNA genes. The optimal partitioning scheme had four partitions: first codon position of *COI* and third codon position of *COII*; second codon position of *COI*; third codon position of *COI*; and first and second codon positions of *COII*. For the bacterial endosymbiont data, we provided a maximum of four partitions (i.e. one for each non-protein-coding locus). Three partitions were selected in the optimal scheme, with one partition for 16S rRNA, one for *CP*78, and another combining *CP*50 and *CP*63. To infer the respective phylogenetic relationships among sampled hosts and bacterial endosymbionts, we used maximum likelihood, as implemented in RAxML 8.2.0 (Stamatakis, 2014). For each dataset, we used a GTR+G+I model of evolution with four discrete rate categories; each partition (as identified by PARTITIONFINDER, above) was allowed an independent implementation of this model. Node support was assessed using 1000 bootstrap pseudoreplicates.

For the host phylogeny, *C. relictus* was used as an outgroup as this species is the most closely related taxon for which both *COI* and *COII* data are available. Sequences were obtained from Cameron *et al.* 2012; GenBank accessions: NC_018120 and NC_018132. For the bacterial endosymbiont phylogeny, *Blattabacterium* sp. str. BPLAN (endosymbiont of *Periplaneta americana*; Sabree *et al.*, 2009; GenBank accession: CP001429.2) was used as an outgroup, as it is the most closely related taxon for which homologous sequences

of all four loci are available. In both cases, major clades within the ingroup were identified on the basis of well-supported monophyly (i.e. bootstrap values $\geq$ 70%, following Hillis & Bull, 1993) coupled with geographic cohesiveness, and then their spatial distributions were mapped.

Pattern-based congruence analyses

To test for cophylogenetic structure between host and endosymbiont haplotypes, we used PARAFIT (Legendre *et al.*, 2002). This analysis assesses congruence globally across the whole tree, and locally by testing the contribution of individual host-endosymbiont links to the global cophylogenetic structure (if the former is significant). PARAFIT was implemented in COPYCAT 2.04 (Meier-Kolthoff *et al.*, 2007), which runs the AxParafit and AxPcoords algorithms (Stamatakis *et al.*, 2007). Three different types of distances were considered: patristic distances either (1) without or (2) with branch length information included (i.e. unweighted or weighted, respectively), and (3) phenetic distances. Following Stamatakis *et al.* (2007), *P*-values were interpreted within the context of a range of alpha values. However, given that correction for multiple testing in PARAFIT may be warranted (Irwin *et al.*, 2012), we focused our interpretation at the $P < 0.02$ level (see Appendix S5 for details). Non-significant individual host-endosymbiont links were examined in three ways. First, their phylogenetic distribution was visually assessed via a tanglegram, estimated in DENDROSCOPE 3.0 (Huson & Scornavacca, 2012). Second, geographic clustering of sampling sites with non-significant individual links was evaluated using Cuzick & Edwards' (1990) test, implemented in CLUSTERSEER 2.5.2 (BioMedware; see Appendix S5 for details). Finally, differences in elevation between sampling sites with non-significant versus sampling sites with significant links were examined using *t*-tests (two-tailed, with equal variances as determined by an *F*-test).

For each of the five major clades identified for both species (see Results), correspondence between uncorrected *p*-distances calculated in PAUP* 4.0a146 (Swofford, 2002) were assessed using linear regression. The genetic distance between a pair of *C. punctulatus* individuals was treated as a predictor variable, and the genetic distance between corresponding *B. cuenoti* isolates was the response variable. To account for inherent differences in mutation rate, for each species the largest observed pairwise *p*-distance was re-scaled to equal one, with all other values multiplied by that scaling factor. Next, we fitted a regression line with the *y*-intercept constrained at zero, and calculated its slope. Under a scenario of perfect correspondence between host-endosymbiont pairwise genetic distances, slope = 1. Conversely, if some (but not all) moderately differentiated endosymbionts were associated with relatively undifferentiated hosts the slope would be > 1, whereas if some differentiated host pairs carry undifferentiated endosymbionts, the regression line would have a slope < 1. For comparison with clade-specific outcomes, the same procedure was repeated on data from all clades combined.

To identify sets of neighbouring geographic locations associated with unusually large genetic differences relative to the spatial distances between them, we analysed the DNA sequences of each species using BARRIER 2.2 (Manni *et al.*, 2004). When multiple barriers are calculated using this method, they are rank-ordered from strongest to weakest. To assess robustness of a given barrier, we analysed bootstrap genetic distance matrices, and examined the spatial locations and rankings of the top five barriers per species, as this represented the threshold beyond which additional barriers were weakly supported (i.e. bootstrap values < 50%; see Appendix S5 for details).

Process-based tree reconciliation analyses

We used the event-cost method implemented in JANE 4.0 (Conow *et al.*, 2010) to examine the relative importance of different processes that promote or impede the evolution of cophylogenetic structure. Five types of events are considered: codivergence (concerted lineage splitting), duplication (divergence of the symbiont within an undifferentiated host), host-switching (lateral transfer of the symbiont to an unrelated host, usually following duplication), lineage sorting/loss, and failure to diverge. The latter two event types involve divergence of the host that is unaccompanied by that of the symbiont but differ in whether the symbiont then persists in only one versus both of the host sister lineages, respectively. Given that tree reconciliation solutions are sensitive to the cost schemes implemented, we explored a set of 14 alternatives (see Appendix S6 for details), with host-switching considered the most improbable, since *B. cuenoti* is believed to be strictly maternally-inherited (Brooks, 1970).

RESULTS

Sequencing, alignment and genetic marker polymorphism

Consistent with true mtDNA, *C. punctulatus* sequences were A+T rich, contained no premature stop codons, and were free of indels. The concatenated host dataset was comprised of 76 haplotypes. The 1125-bp alignment contained 315 variable positions, of which 273 were parsimony-informative. The *B. cuenoti* 16S rRNA sequences had an unbiased nucleotide composition, whereas the three IGR loci were A+T rich. Half of all variable positions in the 16S rRNA alignment were parsimony-informative, whereas this proportion ranged from 67% to 88% for the IGR loci. The concatenated endosymbiont dataset (2686-bp alignment with 254 variable and 188 parsimony-informative positions) included numerous multi-site indels, which contributed to high haplotypic diversity (94 haplotypes; see Table S3.3 in Appendix S3).

Phylogenetic tree estimation

The *C. punctulatus* maximum likelihood tree ($-\ln L = 6355.77$) contained five major clades (H1–H5,

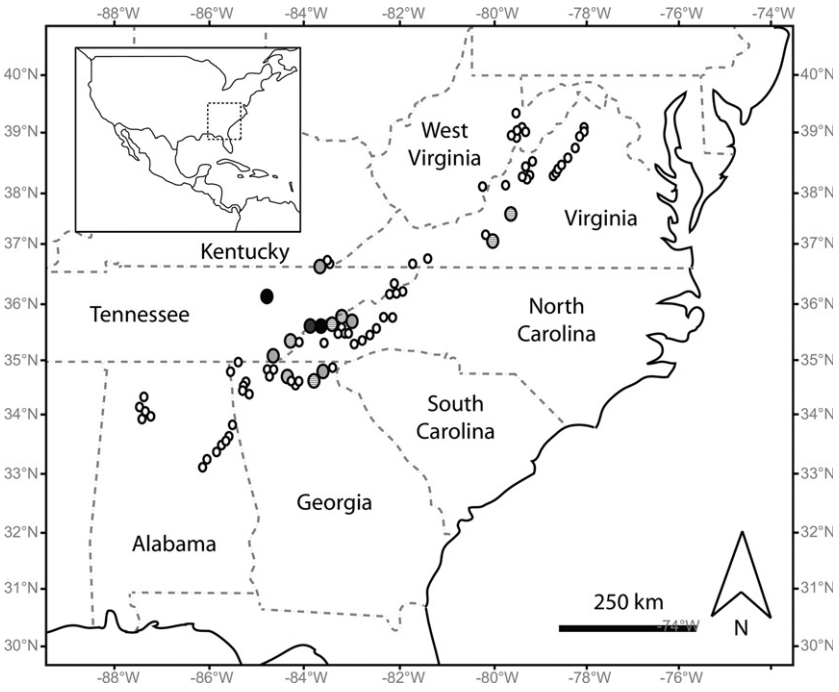

Figure 2 Geographic distribution of sites with non-significant individual host-endosymbiont links in the PARAFIT analysis of cophylogenetic structure, conducted using unweighted patristic distances. Inset shows location of study area within North America. Shading of sampling sites (larger circles) represents non-significant individual links (i.e. those that do not contribute to the global cophylogenetic structure present in the data as a whole) that were detected at the following alpha levels that were considered when iteratively running the analysis: black, $P < 0.05$; dark grey, $P < 0.04$; light grey, $P < 0.03$, and striped, < 0.02). Smaller circles (white) are sites that did contribute to the global cophylogenetic structure, and therefore represent significant individual host-endosymbiont links.

where the prefix "H" indicates host; Fig. 1, left panel). The *B. cuenoti* phylogeny ($-\ln L = 7481.81$) also comprised five major clades (E1–E5, prefix "E" indicates endosymbiont; Fig. 1, right panel). Members of each endosymbiont clade matched those of the corresponding host clade. For both species, clades 1 and 2 were clearly sisters, but support for other higher-level relationships had generally weak bootstrap support. Nonetheless, the best estimates of the host and endosymbiont trees had the same overall branching pattern among major clades. Geographically, there was a stepwise ordering of clades along the approximately north-south axis of the southern Appalachian Mountains, with a zone of parapatry in the central region (Fig. 1, middle panel).

Pattern-based congruence analyses

All PARAFIT analyses identified significant global cophylogenetic structure ($P = 0.001$), irrespective of the genetic distance measure used. All 95 individual host-endosymbiont links were significant ($P < 0.02$) when using either weighted patristic distances or phenetic distances. However, when using unweighted patristic distances, there was between two and 14 non-significant individual links, depending on the alpha level used (i.e. $n = 2$, 3, 7 or 14 non-significant links, at alpha levels of $P < 0.05$, < 0.04, < 0.03 or < 0.02, respectively). Based on the optimal tanglegram, several of these non-significant links appeared to be phylogenetically clustered (see Fig. S5.1 in Appendix S5). For example, a four-taxon clade (associated with sites A38, A39, A75 and A80) and a five-taxon clade (sites A28, A36, A37, A94 and A96; see Table S.1 in Appendix S1 for site information) together accounted for 43% of all non-significant links detected at the $P < 0.02$ level. Notably, however, none of these instances of topological discord traversed the boundaries of major clades; instead, all were restricted within clades. Geographically, most non-significant individual links occurred at locations in and around the central region of the study area (Fig. 2). Based on Cuzick & Edwards' (1990) test, spatial clustering of the 14 non-significant individual links identified at the $P < 0.02$ level was evident after Bonferroni correction when using a neighbourhood size of $k = 1$ ($Tk = 6$, $P = 0.008$), and before (but not after) Bonferroni correction when using $k = 2$ ($Tk = 11$, $P = 0.046$). However, when considering a larger neighbourhood size ($k = 3$), support for spatial clustering was weaker ($Tk = 15$, $P = 0.146$). Mean elevation of the geographic locations associated with the 14 non-significant individual links (752 m) was not statistically different from that of the 81 locations that contributed significant links (707 m; $t = -0.462$, $d.f. = 93$, $P = 0.645$).

Based on raw uncorrected *p*-distances, the inherent mutation rate of *C. punctulatus* mtDNA was on average 4× faster than that of DNA regions sequenced from *B. cuenoti* (median = 3.8 ×, interquartile range = 3.3–4.4×). Within each major clade, the correlation between re-scaled uncorrected *p*-distances between pairs of hosts and their corresponding endosymbionts was significant and positive. However, all relationships consistently deviated from the expectation of perfect congruence, in the same direction (i.e. slope < 1; Fig. 3). Clades 1–3 showed similar relationships to one another (slopes = 0.622–0.703, $R^2 = 0.814$–0.878, $P < 0.0001$; Fig. 3), and although clade 4 had too few samples to yield a robust regression line, the observed relationship (slope = 0.562, $R^2 = 0.535$, $P = 0.001$; Fig. 3) did not strongly differ from those above. Conversely, clade 5 differed markedly (Fig. 3): its regression line was flatter (slope = 0.225), and the spread of data points was more

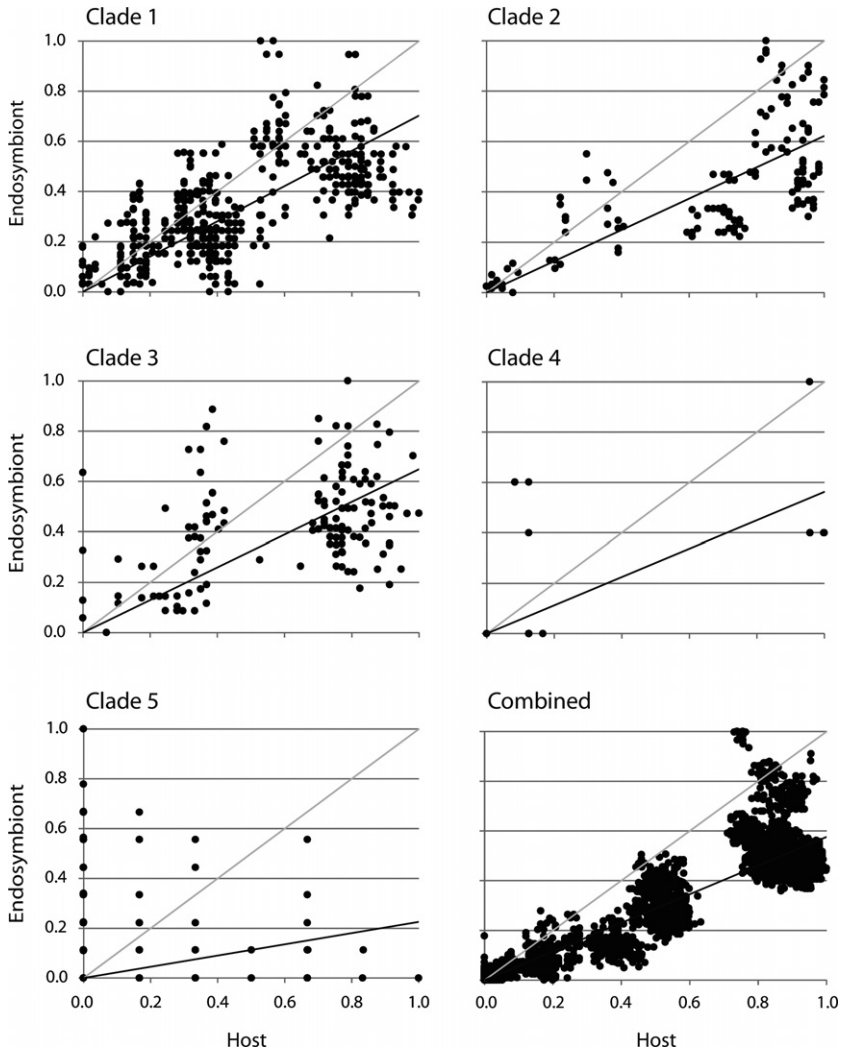

Figure 3 Correlation between host and endosymbiont pairwise uncorrected *p*-distances that have been rescaled so that the maximum observed value for each species equals one. Each black dot is a data point from a single host-endosymbiont pair, and these are organized either separately for each of the five major clades, or for all samples combined. On all panels, the pale grey regression line represents the relationship that is expected when congruence is perfect (i.e. slope = 1), and the black regression line represents the observed relationship. The number of data points on which regressions were based is as follows: clade 1, $n = 528$; clade 2, $n = 136$; clade 3, $n = 153$; clade 4, $n = 15$; clade 5, $n = 210$; and combined, $n = 4465$.

diffuse ($R^2 = 0.041$). For all clades combined, a significant and positive relationship was evident, but as above, the regression line differed from a scenario of perfect correspondence in a direction that indicated some, but not all, moderately differentiated host pairs carry relatively undifferentiated endosymbionts (slope = 0.576, $R^2 = 0.951$, $P < 0.0001$; Fig. 3).

Geographic locations of abrupt genetic discontinuities among *C. punctulatus* were similar to those for *B. cuenoti* (Fig. 4). Four of the five strongest barriers identified among host collection sites (barriers "b"–"e") were shared by its bacterial endosymbiont. Of these shared barriers, one of them formed an enclosed loop around a single sampling site in North Carolina (barrier "b" surrounding site A94; see Table S1.1 in Appendix S1 for site information). Conversely, the other three barriers were each approximately north-south oriented linear extensions that bisected groups of neighbouring sites. Barrier "c" extended across multiple states (Kentucky, Virginia, Tennessee and North Carolina). Barrier "d" was less expansive; it occurred in an area of sparse sampling in Virginia, between sites A51 and A72. Barrier "e" was also restricted to Virginia, and rendered all sites east of site A54 a distinct group. The two species-specific barriers had similar configurations, but occurred in different locations: host barrier "a" formed an enclosed loop around site A28 in North Carolina, whereas endosymbiont barrier "f" enclosed site A105 in West Virginia.

Rank-ordering of the five major barriers differed markedly between the two species (Fig. 4). For example, the strongest barrier for the host (barrier "b") was a relatively minor barrier for the bacterial endosymbiont. The reverse was also true in that barrier "d" was ranked first for *B. cuenoti*, but fourth for *C. punctulatus*. For the four shared barriers ("b"–"e"), the average difference in rank was 2.75 (range: 1–4).

Process-based tree reconciliation analyses

Significant cophylogenetic structure ($P < 0.01$) between the host and its endosymbiont was detected in nine out of 14 event-cost schemes assessed using JANE (see Table S6.5 in Appendix S6). For the remaining five schemes that did not detect significant congruence between molecular phylogenies,

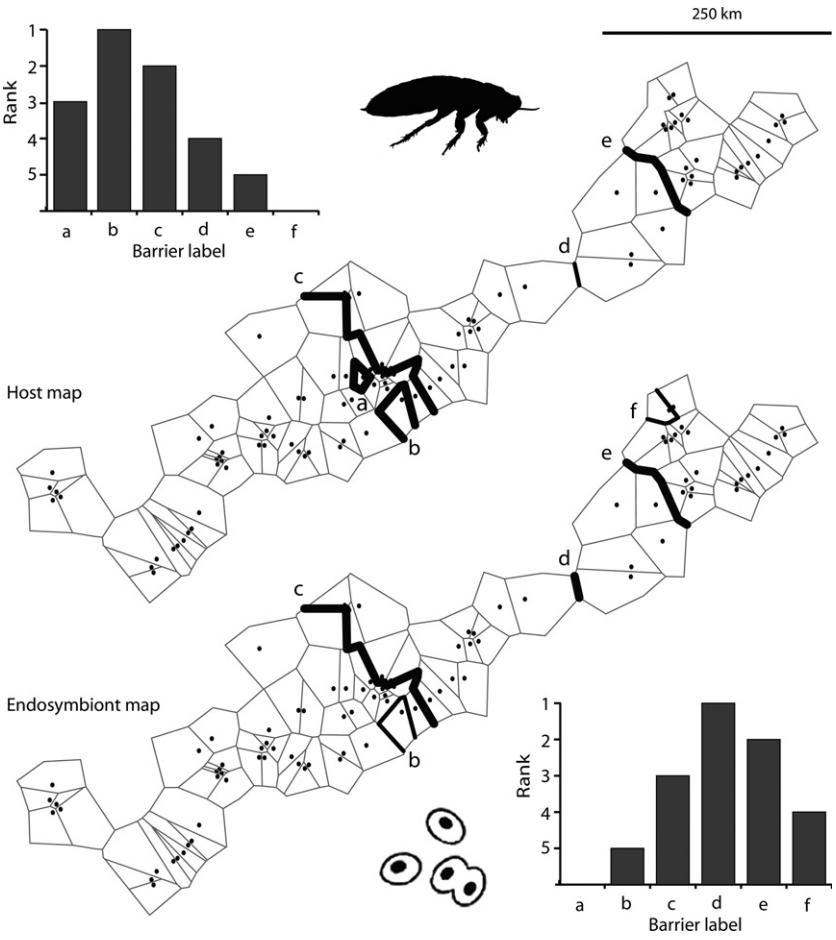

Figure 4 Geographic zones of abrupt spatial-genetic discontinuities for *Cryptocercus punctulatus* (top) and *Blattabacterium cuenoti* str. *punctulatus* (bottom), identified via BARRIER analyses. For each species, bar charts show the rank-ordering of the top five barriers, where the strongest is ranked "1" (*y*-axis), and the barrier label (*x*-axis) corresponds to contiguous sets of edges shown on the maps as thick black lines. Each map displays sampling sites (n = 95 solid black dots; see Figs. 1 and 2 for associated state borders), which are the centroids of Voronoï tessellation (grey polygons). The thickness of barriers corresponds to their robustness, as assessed via analyses of bootstrap genetic distance matrices (thickest lines, > 90% support; thinner lines, 70–80% support).

two common features were that they all had cost settings that maximized lineage sorting/loss events, and their total number of inferred events was always large (344–456, of which 250–362 were losses). Two additional event-cost schemes that maximized lineage sorting/loss events had significant cophylogenetic structure, yet still returned a tree reconciliation solution that contained many (344) independent events (see Table S6.5 in Appendix S6). Accordingly, we focused our interpretation on the seven event-cost schemes that showed strong cophylogenetic structure and invoked relatively few (144–178) events.

Regardless of the cost scheme used, codivergence and lineage sorting/loss were the two most frequently inferred events. Unexpectedly, among the seven focal event-cost schemes (see above), 18–28 host-switching events were inferred despite their biological improbability (i.e. given strict maternal transmission of *B. cuenoti*; Brooks, 1970). When only a single event type was maximized via cost assignments (cost schemes A, B and D in Table S6.5 in Appendix S6), the lowest global cost (171) was associated with maximizing codivergence, and 56–61 codivergence events were always inferred. Similar outcomes were seen when two different event types were jointly maximized (cost schemes E, G and I in Table S6.5 in Appendix S6): the lowest global costs were associated with schemes in which codivergence was maximized, and 56–61 of these events were inferred.

DISCUSSION

This study extends our understanding of codivergence between *C. punctulatus* wood-roaches and their obligate bacterial endosymbiont, *B. cuenoti*. The main advances on earlier work (Clark *et al.*, 2001; Clark & Kambhampati, 2003; Lo *et al.*, 2003; Maekawa *et al.*, 2005) include intensive population-level sampling (Fig. 1, centre panel), development of loci with improved phylogenetic information content, and the application of recently developed analytical methods. We found that strong congruence exists even over very fine spatial scales, but several notable (albeit sometimes subtle) differences also exist. Below we address our three primary questions (see Introduction), and then close by highlighting how the *Cryptocercus* and *Blattabacterium* may provide an excellent opportunity for investigating geographic scaling of processes that affect biogeographic patterns.

Cophylogenetic structure: topological discord is rare, but potentially non-negligible

Our analyses identified significant overall cophylogenetic structure between *C. punctulatus* and *B. cuenoti* (PARAFIT global $P = 0.001$). However, when considering unweighted patristic distances (i.e. tree topologies only, *cf.* topology plus

branch length information), up to 14 out of 95 individual host-endosymbiont pairs (14.7%) did not contribute to the overall pattern of matching branching order. These putatively incongruent lineage splitting events were all localized within major clades. Furthermore, these 14 apparent mismatches were spatially clustered within the central region of the study area (Fig. 2; Cuzick & Edwards' test, $P = 0.008$, assuming small neighbourhood size). This area contains the Great Smoky Mountains National Park, which is a deeply dissected topographically complex landscape. Such regions are known to promote lineage differentiation, particularly in low-mobility taxa (e.g. Garrick *et al.*, 2008; Thomas & Hedin, 2008; Walker *et al.*, 2009; Bull *et al.*, 2013). If population fragmentation is a recurrent process, then some lineage splitting events will be relatively young and are therefore unlikely to be faithfully tracked by gene trees of both species (Maddison, 1997; Templeton, 2002; Garrick *et al.*, 2010). In a previous study, Clark *et al.* (2001) also identified rare instances of topological conflict between *C. punctulatus* and *B. cuenoti*, but those authors did not consider them biologically meaningful. However, the number of sampling sites included in the present study was considerably larger (6.8×), as was the number of haplotypes resolved for both the host and endosymbiont (5.4 and 6.7×, respectively). Accordingly, we consider rare instances of topological discord to be potentially non-negligible, with the caveat that despite improved genetic and sampling, phylogenetic uncertainty may still play a role.

Findings of incongruent lineage splitting between an insect host and its maternally-inherited bacterial endosymbiont are not without precedent. For example, Symula *et al.* (2011) used population genetic approaches to investigate the tsetse fly/*Wigglesworthia* symbiosis in Uganda. Compared to studies of the well-characterized aphid/*Buchnera* system in which perfect topological congruence has been demonstrated (Funk *et al.*, 2000, 2001; Abbot & Moran, 2002), intraspecific sampling of tsetse fly hosts and their *Wigglesworthia* was much more extensive. Symula *et al.* (2011) found that while topological congruence was strong at a regional scale (as in the present study), there was significant discordance within each of two regions. The authors suggested discordance over finer spatial scales may be attributable to incomplete lineage sorting. Our study supports the notion that rare events will be more readily detected when sampling is dense, and we consider incomplete lineage sorting to be a plausible explanation for putative topological discord between *C. punctulatus* and *B. cuenoti*. However, to test this idea empirical data on effective population sizes for both the host and endosymbiont are needed; until they become available, incomplete lineage sorting represents a working hypothesis only.

Spatial-genetic structure: most breaks are in the same locations, but differ in strength

The geographic locations of abrupt genetic breaks in *C. punctulatus* and *B. cuenoti* were similar, but not identical, as both the host and endosymbiont had one unique break. Furthermore, although four of the five most pronounced spatial-genetic breaks were shared by the host and its endosymbiont, there were marked differences in the rank-ordering of strength of these breaks (Fig. 4). Thus, although major clades of the two species show the same geographic localization, the level of DNA sequence differences within and among those clades differs. One contributing factor may be accelerated rates of molecular evolution in *B. cuenoti* due to depletion of DNA repair genes, which has been documented for bacteria that transition from a free-living to endosymbiotic lifestyle (Moran *et al.*, 2008). However, elevated mutation rates alone do not fully explain our observations. For example, when we plotted standardized pairwise sequence divergences for *C. punctulatus* versus *B. cuenoti*, we found that relationships consistently indicated that some, but not all, moderately differentiated host pairs carry relatively undifferentiated endosymbionts (Fig. 3). This finding suggests that stochastic and/or location-specific processes may drive discordance in levels of sequence divergence between hosts and their symbionts. Such processes may include strong bottleneck-induced genetic drift that affects maternally-inherited endosymbiotic bacteria during mother-offspring transfer (Moran *et al.*, 2008), perhaps due to heterogeneity in host population sizes (Funk *et al.*, 2001; Abbot & Moran, 2002). Indeed, spatial and temporal variability in effective population sizes of *C. punctulatus* is quite likely, given that mid-to late Pleistocene glacial cycles caused recurrent range contractions and expansions of southern Appalachian forest biota (Soltis *et al.*, 2006).

Underlying processes: repeated codivergence due to host-tracking, at multiple levels

Our data show clear evidence of repeated codivergence, which is usually attributable to host-tracking (Page & Charleston, 1998). Event-cost tree reconciliation analyses indicated that the primary cause of occasional instances of putative topological discord was divergence of the host unaccompanied by divergence of the symbiont (i.e. lineage sorting/loss; see Table S6.5 in Appendix S6). These analyses also suggest that other processes such as host-switching could have played a role. However, host-switching by *B. cuenoti* seems biologically improbable, as this endosymbiont is thought to be strictly maternally inherited (Brooks, 1970). Indeed, even for facultative endosymbiotic bacteria, the mechanisms for horizontal transfer (i.e. paternal leakage, or transmission mediated by ectoparasites), and observed occurrences of such events, are few (Moran *et al.*, 2008; but see Duron & Hurst, 2013). Accordingly, as with Symula *et al.*'s (2011) tsetse fly/*Wigglesworthia* study (described above), incomplete lineage sorting may explain the rare cases of host-endosymbiont incongruence identified in the present study.

Our primary finding of strong codivergence is in agreement with previous studies that focused on *C. punctulatus* and *B. cuenoti* from the southern Appalachians (Clark *et al.*,

2001), and also with studies that assessed broader-scale biogeographic patterns exhibited by congeneric members of these taxa (Maekawa *et al.*, 2005; Che *et al.*, 2016). The extent of geographic scaling of codivergence seen in this group (i.e. from within just a few hundred meters up to intercontinental scales; this study; Maekawa *et al.*, 2005) has rarely been reported. We therefore suggest that *Cryptocercus* and *Blattabacterium* may represent an excellent study system for investigating links between micro- and macro-evolutionary processes at the population-level/species-level interface, and for understanding how biogeographic patterns form across different levels of spatial organization.

ACKNOWLEDGEMENTS

This project was funded by grants to R.C.G. from the American Philosophical Society (Franklin research grant), National Geographic Society (Committee for Research and Exploration grant # 9695-15), Conservation and Research Foundation, and University of Mississippi start-up funds. Scientific collecting permits were provided by the Alabama Department of Conservation and Natural Resources, Georgia Department of Natural Resources (permit # 29-WBH-12-16), United States Department of Agriculture Forest Service, and United States National Park Service (permit #'s GRSM-2012-SCI-2242, SHEN-2012-SCI-0015, and CUGA-2012-SCI-0008). Z.L.S. and B.C.J. were supported through start-up funds from The Ohio State University. R.J. Dyer, C. Hyseni, R.E. Symula and E.W. Collier assisted with some fieldwork, and J. Meier-Kolthoff and R. Libeskind-Hadas provided help with COPYCAT and JANE analyses, respectively. B.C. Carstens, R.E. Symula and two anonymous reviewers gave constructive comments on the manuscript.

REFERENCES

Abbot, P. & Moran, N.A. (2002) Extremely low levels of genetic polymorphism in endosymbionts (*Buchnera*) of aphids (*Pemphigus*). *Molecular Ecology*, **11**, 2649–2660.

Althoff, D.M., Segraves, K.A. & Johnson, M.T.J. (2014) Testing for coevolutionary diversification: linking pattern with process. *Trends in Ecology and Evolution*, **29**, 82–89.

Araújo, M.B. & Rozenfeld, A. (2014) The geographic scaling of biotic interactions. *Ecography*, **37**, 406–415.

Brooks, M.A. (1970) Comments on the classification of intracellular symbiotes of cockroaches and a description of the species. *Journal of Invertebrate Pathology*, **16**, 249–258.

Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology*, **12**, 168–180.

Bull, J.K., Sands, C.J., Garrick, R.C., Gardner, M.G., Tait, N.N., Briscoe, D.A., Rowell, D.M. & Sunnucks, P. (2013) Environmental complexity and biodiversity: the multi-layered evolutionary history of a log-dwelling velvet worm in montane temperate Australia. *PLoS One*, **8**, e84559.

Burnside, C.A., Smith, P.T. & Kambhampati, S. (1999) Three new species of the wood roach, *Cryptocercus* (Blattodea: Cryptocercidae), from the Eastern United States. *Journal of the Kansas Entomological Society*, **72**, 361–378.

Cameron, S.L., Lo, N., Bourguignon, T., Svenson, G.J. & Evans, T.A. (2012) A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): robust support for interfamilial relationships and molecular synapomorphies define major clades. *Molecular Phylogenetics and Evolution*, **65**, 163–173.

Che, Y., Wang, D., Shi, Y., Dua, X., Zhao, Y., Lo, N. & Wang, Z. (2016) A global molecular phylogeny and time-scale of evolution for *Cryptocercus* woodroaches. *Molecular Phylogenetics and Evolution*, **98**, 201–209.

Clark, J.W. & Kambhampati, S. (2003) Phylogenetic analysis of *Blattabacterium*, endosymbiotic bacteria from the wood roach, *Cryptocercus* (Blattodea: Cryptocercidae), including a description of three new species. *Molecular Phylogenetics and Evolution*, **26**, 82–88.

Clark, J.W., Hossain, S., Burnside, C.A. & Kambhampati, S. (2001) Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. *Proceedings of the Royal Society B: Biological sciences*, **268**, 393–398.

Conow, C., Fielder, D., Ovadia, Y. & Libeskind-Hadas, R. (2010) Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms for Molecular Biology*, **5**, 16.

Cruzan, M.B. & Templeton, A.R. (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution*, **15**, 491–496.

Cuzick, J. & Edwards, R. (1990) Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society Series B*, **52**, 73–104.

Duron, O. & Hurst, G.D.D. (2013) Arthropods and inherited bacteria: from counting the symbionts to understanding how symbionts count. *BMC Biology*, **11**, 45.

Everaerts, C., Maekawa, K., Farine, J.P., Shimada, K., Luykx, P., Brossut, R. & Nalepa, C.A. (2008) The *Cryptocercus punctulatus* species complex (Dictyoptera: Cryptocercidae) in the eastern United States: comparison of cuticular hydrocarbons, chromosome number, and DNA sequences. *Molecular Phylogenetics and Evolution*, **47**, 950–959.

Funk, D.J., Helbling, L., Wernegreen, J.J. & Moran, N.A. (2000) Intraspecifc phylogenetic congruence among multiple symbiont genomes. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 2517–2521.

Funk, D.J., Wernegreen, J.J. & Moran, N.A. (2001) Intraspecific variation in symbiont genomes: bottlenecks and the aphid-*Buchnera* association. *Genetics*, **157**, 477–489.

Garrick, R.C. (2011) Montane refuges and topographic complexity generate and maintain invertebrate biodiversity: recurring themes across space and time. *Journal of Insect Conservation*, **15**, 469–478.

Garrick, R.C., Rowell, D.M., Simmons, C.S., Hillis, D.M. & Sunnucks, P. (2008) Fine-scale phylogeographic congruence despite demographic incongruence in two low-mobility saproxylic springtails. *Evolution*, **62**, 1103–1118.

Garrick, R.C., Caccone, A. & Sunnucks, P. (2010) Inference of population history by coupling exploratory and model--driven phylogeographic analyses. *International Journal of Molecular Sciences*, **11**, 1190–1227.

Gillespie, R.G., Baldwin, B.G., Waters, J.M., Fraser, C.I., Nikula, R. & Roderick, G.K. (2012) Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology and Evolution*, **27**, 47–55.

Hillis, D.M. & Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182–192.

Huson, D.H. & Scornavacca, C. (2012) Dendroscope 3: an interactive viewer for rooted phylogenetic trees and networks. *Systematic Biology*, **61**, 1061–1067.

Irwin, N.R., Bayerlová, M., Missa, O. & Martínková, N. (2012) Complex patterns of host switching in new world arenaviruses. *Molecular Ecology*, **21**, 4137–4150.

Kambhampati, S., Luykx, P. & Nalepa, C.A. (1996) Evidence for sibling species in *Cryptocercus punctulatus*, the wood roach, from variation in mitochondrial DNA and karyotype. *Heredity*, **76**, 485–496.

Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Legendre, P., Desdevises, Y. & Bazin, E. (2002) A statistical test for host-parasite coevolution. *Systematic Biology*, **51**, 217–234.

Lo, N., Bandi, C., Watanabe, H., Nalepa, C. & Beninati, T. (2003) Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Molecular Biology and Evolution*, **20**, 907–913.

Lo, N., Luykx, P., Santoni, R., Beninati, T., Bandi, C., Casiraghi, M., Lu, W.H., Zakharov, E.V. & Nalepa, C.A. (2006) Molecular phylogeny of *Cryptocercus* wood-roaches based on mitochondrial COII and 16S sequences, and chromosome numbers in Palearctic representatives. *Zoological Science*, **23**, 393–398.

Losos, J.B. & Ricklefs, R.E. (2009) Adaptation and diversification on islands. *Nature*, **457**, 830–836.

Luykx, P. (1983) XO:XX sex chromosomes and Robertsonian variation in the autosomes of the wood-roach Cryptocercus punctulatus (Dictyoptera: Blattaria: Cryptocercidae). *Annals of the Entomological Society of America*, **76**, 518–522.

Maddison, W. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Maekawa, K., Park, Y.C. & Lo, N. (2005) Phylogeny of endosymbiont bacteria harbored by the woodroach *Cryptocercus* spp. (Cryptocercidae: Blattaria): molecular clock evidence for a late Cretaceous-early Tertiary split of Asian and American lineages. *Molecular Phylogenetics and Evolution*, **36**, 728–733.

Manni, F., Guérard, E. & Heyer, E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

Meier-Kolthoff, J.P., Auch, A.F., Huson, D.H. & Göker, M. (2007) CopyCat: cophylogenetic analysis tool. *Bioinformatics*, **23**, 898–900.

Moran, N.A., McCutcheon, J.P. & Nakabachi, A. (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165–190.

Nalepa, C.A., Luykx, P., Klass, K.-D. & Deitz, L.L. (2002) Distribution of karyotypes of the *Cryptocercus punctulatus* species complex (Dictyoptera: Cryptocercidae) in the Southern Appalachians: relation to habitat and history. *Annals of the Entomological Society of America*, **95**, 276–287.

Neef, A., Latorre, A., Pereto, J., Silva, F.J., Pignatelli, M. & Moya, A. (2011) Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biology and Evolution*, **3**, 1437–1448.

Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K., Hongoh, Y., Constantino, R., Uys, V., Zhong, J., Kudo, T. & Ohkuma, M. (2007) Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology*, **16**, 1257–1266.

Page, R.D.M. & Charleston, M.A. (1998) Trees within trees: phylogeny and historical associations. *Trends in Ecology and Evolution*, **13**, 356–359.

Patiño-Navarrete, R., Moya, A., Latorre, A. & Peretó, J. (2013) Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biology and Evolution*, **5**, 351–361.

Riddle, B., Dawson, M.N., Hadly, E.A., Hafner, D.J., Hickerson, M.J., Mantooth, S.J. & Yoder, A.D. (2008) The role of molecular genetics in sculpting the future of integrative biogeography. *Progress in Physical Geography*, **32**, 173–202.

Sabree, Z.L., Kambhampati, S. & Moran, N.A. (2009) Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proceedings of the National Academy of Sciences USA*, **106**, 19521–19526.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S. & Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stamatakis, A., Auch, A.F., Meier-Kolthoff, J. & Göker, M. (2007) AxPcoords & parallel AxParafit: statistical co-phylogenetic analyses on thousands of taxa. *BMC Bioinformatics*, **8**, 405.

Suh, S.O., Marshall, C.J., McHugh, J.V. & Blackwell, M. (2003) Wood ingestion by passalid beetles in the presence of xylose-fermenting gut yeasts. *Molecular Ecology*, **12**, 3137–3145.

Swofford, D.L. (2002) *PAUP*: phylogenetic analysis using parsimony (*and other methods)*. Sinauer, Sunderland, MA, USA.

Symula, R.E., Marpuri, I., Bjornson, R.D., Okedi, L., Beadell, J., Alam, U., Aksoy, S. & Caccone, A. (2011) Influence of host phylogeographic patterns and incomplete lineage sorting on within-species genetic variability in

Wigglesworthia species, obligate symbionts of tsetse flies. *Applied and Environmental Microbiology*, **77**, 8400–8408.
Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
Templeton, A.R. (2002) Out of Africa again and again. *Nature*, **416**, 45–51.
Thomas, S.M. & Hedin, M. (2008) Multigenic phylogeographic divergence in the paleoendemic southern Appalachian opilionid *Fumontana deprehendor* Shear (Opiliones, Laniatores, Triaenonychidae). *Molecular Phylogenetics and Evolution*, **46**, 645–658.
Thompson, J.N. (2005) *The geographic mosaic of coevolution*. University of Chicago Press, Chicago, IL, USA.
Ulyshen, M.D. (2016) Wood decomposition as influenced by invertebrates. *Biological Reviews*, **91**, 70–85.
Walker, M.J., Stockman, A.K., Marek, P.E. & Bond, J.E. (2009) Pleistocene glacial refugia across the Appalachian Mountains and coastal plain in the millipede genus *Narceus*: evidence from population genetic, phylogeographic, and paleoclimatic data. *BMC Evolutionary Biology*, **9**, 25.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Population sampling.
Appendix S2 DNA isolation and genetic markers.
Appendix S3 DNA sequence accession numbers.
Appendix S4 Genetic marker polymorphism.
Appendix S5 Pattern-based congruence analyses.
Appendix S6 Process-based tree reconciliation analyses.

DATA ACCESSIBILITY

DNA sequences and alignments are available from GenBank under the following accession numbers: host *COI*: KU609620, KX944873-KX945112 and KY241441-KY241456; host *COII*: KU609623, KX945116-KX945355 and KY241457-KY241472; symbiont 16S rRNA: KY226232-KY226323; symbiont *CP50*: KY226324-KY226418; symbiont *CP63*: KY226419-KY226513; and symbiont *CP78*: KY226514-KY226608.

BIOSKETCHES

Garrick's lab focuses on understanding processes that generate and maintain biodiversity within and among species, with an emphasis on montane forest biota.

Sabree's lab focuses on genomic consequences of transitions to symbiosis, and population genomics of host-associated bacteria.

Author contributions: R.C.G. and Z.L.S. conceived the study; R.C.G., B.C.J. and Z.L.S. developed genetic markers and/or generated the data; R.C.G., B.C.J. and J.C.O. analysed the data; R.C.G. drafted the manuscript and all authors contributed to revisions.

Editor: Şerban Procheş