doi:10.1111/evo.13386

Integrating phylogenomic and population genomic patterns in avian lice provides a more complete picture of parasite evolution

Andrew D. Sweet,[1,2,3] Bret M. Boyd,[1,4] Julie M. Allen,[1,5] Scott M. Villa,[6] Michel P. Valim,[7] Jose L. Rivera-Parra,[8] Robert E. Wilson,[9] and Kevin P. Johnson[1]

[1]*Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, Illinois 61820*

[2]*Program in Ecology, Evolution, and Conservation Biology, School of Integrative Biology, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820*

[3]*E-mail: adsweet2@illinois.edu*

[4]*Department of Entomology, University of Georgia, Athens, Georgia 30602*

[5]*Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611*

[6]*Department of Biology, University of Utah, Salt Lake City, Utah 84112*

[7]*Biotério da Universidade Iguaçu, Av. Abílio Augusto Távora, 2134, RJ 26275, Brazil*

[8]*Departamento de Petroleos, Facultad de Geologia y Petroleos, Escuela Politecnica Nacional, Quito, Ecuador*

[9]*Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska 99775*

Received August 14, 2017
Accepted October 23, 2017

Parasite diversity accounts for most of the biodiversity on earth, and is shaped by many processes (e.g., cospeciation, host switching). To identify the effects of the processes that shape parasite diversity, it is ideal to incorporate both deep (phylogenetic) and shallow (population) perspectives. To this end, we developed a novel workflow to obtain phylogenetic and population genetic data from whole genome sequences of body lice parasitizing New World ground-doves. Phylogenies from these data showed consistent, highly resolved species-level relationships for the lice. By comparing the louse and ground-dove phylogenies, we found that over long-term evolutionary scales their phylogenies were largely congruent. Many louse lineages (both species and populations) also demonstrated high host-specificity, suggesting ground-dove divergence is a primary driver of their parasites' diversity. However, the few louse taxa that are generalists are structured according to biogeography at the population level. This suggests dispersal among sympatric hosts has some effect on body louse diversity, but over deeper time scales the parasites eventually sort according to host species. Overall, our results demonstrate that multiple factors explain the patterns of diversity in this group of parasites, and that the effects of these factors can vary over different evolutionary scales. The integrative approach we employed was crucial for uncovering these patterns, and should be broadly applicable to other studies.

KEY WORDS: Diversification, doves, parasites, *Physconelloides*, Phthiraptera.

Understanding how host evolution and ecology shapes parasite diversity is a key question in evolutionary biology. Traditionally, these host influences were considered the primary drivers behind parasite evolutionary patterns, particularly for parasites that are closely associated with their hosts (Fahrenholz 1913; Harrison 1914; Eichler 1948). However, other factors, such as biogeography and parasite ecology, have been shown to be important forces shaping parasite evolution and host–parasite interactions (Johnson and Clayton 2003a; Weckstein 2004; du Toit et al. 2013; Jirsová et al. 2017).

A widely used approach for addressing questions related to host-parasite evolutionary dynamics is cophylogenetic analysis,

which compares the evolutionary trees of parasites to that of their hosts to test for congruence or cospeciation (Page 1994; Page and Charleston 1998; de Vienne et al. 2013; Clayton et al. 2016). In cases where a parasite phylogeny is highly congruent with the host phylogeny, host divergence (and cospeciation) is thought to be the primary factor shaping parasite diversification. In cases where the parasite phylogeny is incongruent with the host phylogeny, other factors (e.g., host switching) may drive parasite divergence (Page et al. 2004; Peterson et al. 2010; Sweet et al. 2016a). However, cophylogenetic approaches do not consider processes within a species (i.e., populations), and because divergence is typically initiated at the population level, it is important to consider population patterns for a group of parasites (Bush 1975; Templeton 1981; Criscione et al. 2005; Kochzius et al. 2009). Integrating both phylogenetic and population-level approaches can give valuable insight into host-parasite evolution over multiple time scales, and ultimately help to link macroevolutionary patterns to ecological (i.e., microevolutionary) processes (Nadler 1995; Harrison 1998; Clayton and Johnson 2003; Huyse et al. 2005; Carling and Brumfield 2008; Criscione 2008).

To consider both phylogenetic and population patterns, it is important to first identify population-species boundaries by determining the number of Operational Taxonomic Units (OTUs) in a system (Refrégier et al. 2008; de Vienne et al. 2013; Martínez-Aquino 2016). Comparing phylogenies at different taxonomic scales can bias the results. For example, oversplitting parasites relative to their hosts–effectively comparing parasite populations to host species–can incorrectly force phylogenetic congruence and increase estimates of cospeciation (de Vienne et al. 2013). Many types of parasites have reduced and cryptic morphologies, making species delimitation difficult (Nadler and De Leon 2011). Genetic data, such as DNA barcoding, has been used to more objectively define parasite OTUs (Hebert and Gregory 2005; Smith et al. 2006). This approach is useful, but using single short genetic fragments provides limited phylogenetic or population level information or reflects bias of the evolutionary forces on that gene (Brower 2006). Moving beyond simple barcoding, next generation sequencing facilitates the parallel collection of population (e.g., single nucleotide polymorphisms [SNPs]) and species (e.g., multiple nuclear or mitochondrial genes) level data. Additionally, the decrease in cost of NGS makes studies of nonmodel organisms available and cost effective (Yang and Rannala 2010). For example, genome-wide SNPs can indicate structure within a species, perhaps evidence of overlooked cryptic speciation (Leaché et al. 2014). Likewise, species trees estimated from many gene trees under the coalescent model can provide evidence for population-species boundaries (Edwards 2009; Fujita et al. 2012). Using multiple mitochondrial genes can also provide significant information for identifying OTUs (Pons et al. 2006; Sloan et al. 2016). Together, these various data-types can corroborate each other to robustly assess parasite (or host) OTUs for downstream cophylogenetic analysis.

Discerning population-species boundaries is important for cophylogenetic analysis, but population-level patterns also provide insights into the processes driving parasite divergence and host-parasite relationships (McCoy et al. 2005; Criscione 2008; Bruyndonckx et al. 2009). For example, many parasites exhibit phylogeographic structure (e.g., Whipps and Kent 2006; Whiteman et al. 2007; Morand 2012). Other parasites show population-level host-specificity, patterns that would not have been apparent with less dense sampling (i.e., species-level sampling; McCoy et al. 2001; Poulin et al. 2011). As with OTU analysis, population questions have been primarily addressed with short genetic fragments (e.g., COI mitochondrial locus) or microsatellite data, which are useful but contain limited information. NGS data, such as SNPs, can reveal more fine-scale structure within populations (Luikart et al. 2003; Hohenlohe et al. 2010).

When using NGS data to explore both species-level and population-level questions, there is an advantage to using full genome data from shotgun sequencing. Genome-reduction methods, such as Ultra-Conserved Elements (UCEs) or anchored hybrid enrichment, are useful for phylogenetic inference but data sets are restricted to the targeted loci (Faircloth et al. 2012; Lemmon et al. 2012). Methods useful for population-level questions, such as restriction-site associate DNA sequencing (RAD-seq), are less useful for phylogenetic estimation, especially at deeper time scales (Rubin et al. 2012; Manthey et al. 2016). Full-genome shotgun sequencing produces appropriate data for both phylogenetic and population genetic questions, given that locus assembly and SNP calling protocols are available. For organisms with relatively small genomes (<1 GB), this approach can be extremely cost effective through multiplexing (Boyd et al. 2017).

In this study, we focus on the body lice (Insecta: Phthiraptera: Ischnocera) from small New World ground-doves (Aves: Columbidae: Claravinae). Dove lice are obligate and permanent ectoparasites that feed on their hosts' downy feathers (Johnson and Clayton 2003b). There are three recognized species of ground-dove body lice that form a monophyletic group within the genus *Physconelloides* (*P. emersoni*, *P. eurysema*, and *P. robbinsi*), although there are likely several additional cryptic species (Clayton and Johnson 2003; Price et al. 2003; Johnson et al. 2011b). Past work has demonstrated significant phylogenetic congruence and cospeciation between doves and their body lice (Clayton and Johnson 2003; Johnson and Clayton 2003a, 2004). Patterns of congruence are perhaps reinforced by the inability of body lice to effectively disperse between different host species using hippoboscid flies (phoresis), a behavior utilized by other types of avian lice (Harbison et al. 2008, 2009). However, past phylogenetic and cophylogeneitc studies were on broad taxonomic scales (across Columbidae) and had relatively sparse sampling.

Ground-doves and their body lice are an advantageous system for understanding patterns of parasite diversification and host–parasite evolution using genomic data. These lice have relatively small genomes (~200 Mbp), and are associated with a moderately diverse, yet widespread host group (Johnson et al. 2011a,b; Sweet and Johnson 2015). Small New World ground-doves are a monophyletic subfamily (Clarvinae) of four genera and 17 species within the dove family Columbidae (Johnson and Clayton 2000; Pereira et al. 2007). They inhabit a wide geographic range extending from the southern United States to southern South America (Gibbs et al. 2001). Additionally, focusing on a relatively small monophyletic group of parasites is ideal for pursuing both phylogenetic and population-level questions, because it is feasible to obtain multiple samples from all species in the clade. Including multiple representatives of each species is also necessary for identifying cryptic species. Here, we include samples spanning the geographic ranges of each of the three ground-dove body louse species.

Whereas most studies of host–parasite evolution focus on either phylogenetic or population genetic patterns, here we integrate both scales by using full genome sequences of ground-dove body lice to identify species-population boundaries and assess the genetic structure within and between species. We accomplish this by developing a novel workflow to assemble genes and call SNPs from the same data source. In particular, we are interested in how the patterns in these lice relate to (a) their hosts' phylogenetic structure and (b) their geographic distributions. Do the lice exhibit similar patterns of host congruence and/or host specificity at both phylogenetic and population scales? Is there biogeographic/phylogeographic structure at both scales? Using our approach to address these questions will provide great insight into how parasites diversify over time and indicate which factors (e.g., host evolution or biogeography) might be driving parasite diversification at different points in time.

Materials and Methods

DNA EXTRACTION

Louse samples were collected in the field from their hosts using either the ethyl acetate fumigation or pyrethrin powder dusting methods (Clayton and Drown 2001) and were immediately submerged in 95% ethanol and stored at –80°C. Before DNA extraction, each specimen was photographed at the University of Utah as a voucher. Whole lice were ground up individually (34 individual specimens in total) in 1.5 mL tubes and genomic DNA was isolated using standard protocols and reagents of the Qiagen QIAamp DNA Micro Kit (Qiagen, Valencia, CA, USA). Our only modification of the Qiagen protocol was to incubate our specimens in ATL buffer and proteinase *K* at 55°C for 48 hours instead of the recommended 1–3 hours. This was done to ensure maximal yield of DNA from the louse remains. Following DNA extractions, we quantified each extraction with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using the manufacturer's recommended protocols and reagents.

LIBRARY PREPARATION AND SEQUENCING

Total genomic DNA (gDNA) was fragmented on a Covaris M220 Focused-ultrasonicator (Covaris, Woburn, MA, USA) targeting a mean fragment size of 400 nt. gDNA libraries for each louse specimen were constructed for paired-end Illumina sequencing using the recommended protocols and reagents of the Kapa Library Preparation Kits (Kapa Biosystems, Wilmington, MA, USA). Six or 10 bp barcodes were added to each library sample so that 8–12 samples could be pooled and sequenced simultaneously on a single lane (Table S1). Three additional samples were sequenced on a single lane to obtain high-coverage genomes for methods development and assessing error (sequencing pool 4, Table S1). The pooled libraries were sequenced with 161 cycles on an Illumina HiSeq2500 instrument using the HiSeq SBS v. 4 sequencing kit, resulting in 160nt paired-end reads. Fastq files were generated from the sequence data using Casava v.1.8.2 or bcltofastq v.1.8.4 with Illumina 1.9 quality score encoding. All sequencing and fastq file generation was carried out at the W. M. Keck Center (University of Illinois, Urbana, IL, USA). Raw reads were deposited to the NCBI GenBank SRA database (SRP076185). We also obtained raw genomic reads from *Campanulotes compar* (NCBI BioProject PRJNA374052, ID 374052) as an outgroup taxon.

SEQUENCE QUALITY CONTROL

We analyzed raw Illumina sequence data using Fastqc v.0.10.1 (Babraham Bioinformatics) to check for unusual sequence patterns or errors. For quality control measures, we first removed duplicated sequence read pairs using the fastqSplitDups.py script (https://github.com/McIntyre-Lab/mcscript and https://github.com/McIntyre-Lab/mclib). Second, we removed the 5' and 3' Illumina sequencing adapters using Fastx_clipper v.0.014 from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). We then removed the first 5 nt from the 5' ends using Fastx_trimmer v.0.014 ("hard" trimming). Finally, we "soft" trimmed the 3' end of reads by removing bases with phred scores less than 28 using Fastq_quality_trimmer v.0.014 and a trimming window = 1 nt. After these quality control steps, we removed any reads less than 75 nt from the fastq files. We then reanalyzed our cleaned libraries using Fastqc to check for errors not removed by quality control.

DATA ASSEMBLY AND MAPPING

To obtain orthologous sequences for downstream analysis, we developed a novel approach to assemble and map the cleaned

genomic data. First, we used aTRAM (Allen et al. 2015) to assemble exons for the *Physconelloides emersoni* library sequenced at a higher depth. Of the three higher coverage libraries, this sample had the highest predicted depth after quality control (Sample 1, Table S1). We prepared the *P. emersoni* reads into a BLAST-formatted database using the format_sra.pl script from the aTRAM package. We then ran aTRAM for three iterations with ABySS (Simpson et al. 2009) for de novo assembly, using 1107 protein coding genes from the human body louse genome (*Pediculus humanus humanus*; Kirkness et al. 2010) as target sequences. These genes were identified by Johnson et al. (2013) as being 1:1 orthologs across nine insect genomes using OrthoDB v5 (Waterhouse et al. 2011). We used the resulting best contigs from aTRAM in a postprocessing pipeline from Allen et al. (2017) to identify exons. The pipeline uses Exonerate v.2.2.0 (Slater and Birney 2005) to identify exon/intron boundaries and custom scripts to stitch together the exon regions of each locus assembled in aTRAM. We then performed BLAST searches with blastx (Altschul et al. 1997) between our assembled loci and the *P. h. humanus* translated proteins. If the best hit for an assembled locus was not the corresponding *P. h. humanus* gene we removed that locus from the assembly.

Because aTRAM is most effective for assembling loci from high-coverage libraries and does not code heterozygous sites or call SNPs, we used Bowtie2 (Langmead and Salzberg 2012) to map our lower coverage libraries (avg. 13X) to the *P. emersoni* aTRAM-assembled loci. We also mapped all three higher coverage genomes and the outgroup taxon to the reference loci. Before mapping, we created an index file using Samtools (Li et al. 2009) and a dictionary file using CreateSequenceDictionary in Picard v.2.0.1 (https://broadinstitute.github.io/picard/) for the reference sequence. After mapping with Bowtie2, we sorted the BAM files and created pileup files using Samtools. Bcftools then converted the pileup files to VCF files (Li et al. 2009). We then filtered sites with depth <5 or >150, or with Phred quality scores <28 using Samtools and the Genome Analysis Toolkit from GATK (McKenna et al. 2010). We converted filtered VCF files to consensus FASTQ files using vcf2fq in vcfutils.pl from Samtools. All analyses were carried out on a 4 AMD Opteron with 16 2.4 Ghz processors and 64 CPU cores, maintained by the UIUC Life Sciences Computing Services (University of Illinois, Urbana, IL, USA). Details about the mapping and filtering steps and all relevant scripts are available at https://github.com/adsweet/louse_genomes.git.

MITOCHONDRIAL GENE ASSEMBLY

We also assembled mitochondrial protein coding genes using aTRAM. For target sequences, we used the translated sequences from the *Campanulotes compar* mitochondrial genome (NCBI # PRJNA16411). We ran aTRAM with a single iteration. Because mitochondrial reads are likely present at a much greater depth than nuclear reads, we tested several library fractions for assembly (1.5%, 4.7%, 7.8%, 15.6%, 50%, and 100%). To determine which fraction to use for each library, we assembled BLAST reads from the aTRAM output against the *C. compar* reference in Geneious v. 8.1.2 (Biomatters, Ltd.), and chose the minimum library fraction with uniform coverage $\geq$20X.

SAMPLE VALIDATION

To validate the species identity and identify any potential contamination in our assembled sequences, we used the NCBI BLAST web interface to search our CO1 sequences assembled in aTRAM against the GenBank database. We determined a sample to be verified if the top BLAST result was within the same species as our query sequence (Table S1).

SNP CALLING

We called SNPs for population-level analysis for *Physconelloides eurysema* using the GATK Genome Analysis Toolkit following the "Best Practices" guide from the Broad Institute (Van der Auwera et al. 2013). We focused on *P. eurysema* because this louse species is associated with nine host species in our study, and there is evidence from previous work that there are several cryptic species within this lineage. The other ground-dove *Physconelloides* taxa (*P. emersoni* and *P. robbinsi*) are well defined from both morphological and molecular data. We called SNPs jointly for all *P. eurysema* samples, and filtered calls with QD (quality by depth) <2.0, FS (Fisher strand test) >60.0, MQ (mapping quality) <40.0, and MQRankSum (mapping quality rank sum test) <-12.5.

PHYLOGENETIC ANALYSIS

We aligned each gene using the – auto flag in MAFFT (Katoh et al. 2002). For each alignment, we removed columns only containing Ns, and masked sites containing $\geq$90% gaps using trimAL v. 1.4 (Capella-Gutiérrez et al. 2009). For the aligned data, we used both concatenation and coalescent tree estimation methods. For concatenation estimation, we first concatenated all gene files in Geneious. We tested for the best partitioning schemes and models by searching through RAxML models with PartitionFinder v.2.1.1 (Lanfear et al. 2017). We then used the rcluster search scheme with rcluster-max set to 1000 and rcluster-percent set to 10, and selected optimal partitions and models based on AIC (Akaike 1974; Lanfear et al. 2014). We used RAxML v.8.1.3 (Stamatakis 2006) to estimate the best likelihood tree from the partitioned concatenated alignment, using 10 different starting trees and a GTR + Γ model for each partition. We then ran 250 bootstrap replicates in RAxML and summarized support on the best tree. For the coalescent analysis, we estimate gene trees for each gene alignment file using 100 rapid bootstrap replicates in RAxML using a

GTR + Γ model for each gene. We summarized the gene trees using ASTRAL v.4.10.6 (Mirarab and Warnow 2015; Sayyari and Mirarab 2016) with quartet-based local posterior probability support for branches. We also summarized gene trees with ASTRID v.1.4 (Vachaspati and Warnow 2015).

We estimated a phylogeny from the mitochondrial data assembled with aTRAM. We aligned each protein coding gene using – auto in MAFFT. Because many of the assemblies had variable sequence lengths, we trimmed the alignments to the *Campanulotes compar* mitochondrial genome sequence from GenBank (also included as an outgroup taxon). We did not include data for ATP8, ND3, or ND6, because aTRAM was unable to assemble contigs for those genes in most samples, presumably because they are extremely divergent from the reference sequence. Therefore, the final mitochondrial data set included 10 mitochondrial genes. We ran PartitionFinder on the concatenated alignment to test for optimal partition and model schemes using the AIC, and based on this analysis ran RAxML on the concatenated matrix with six partitions of GTR + I + Γ models.

OTU ANALYSIS

To objectively assess the number of ground-dove *Physconelloides* Operation Taxonomic Units (OTUs), we used the automatic barcode discovery method (ABGD; Puillandre et al. 2012) and the Bayesian General Mixed Yule Coalescent Model (bGMYC; Reid and Carstens 2012). ABGD requires an alignment of a barcode gene as input, and detects gaps in the distribution of pairwise differences. These gaps indicate interspecific boundaries. For our data set, we used the COI alignment as input into the web version of ABGD (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html). We used default parameters (Pmin = 0.001, Pmax = 0.1, Steps = 10, Relative gap width = 1.5, Bins = 20) and uncorrected (p-distances), Jukes-Cantor (JC), and Kimura (K80) models for the distance matrix.

The bGMYC method uses Markov Chain Monte Carlo (MCMC) to estimate the transition from speciation to coalescent (within-species) events, and can be implemented over a distribution of trees. However, the method requires ultrametric trees from a single locus. To accommodate this, we estimated ultrametric trees from our concatenated mitochondrial alignment using BEAST v.2.4.4 (Bouckaert et al. 2014) on the CIPRES Science Gateway (Miller et al. 2010). We partitioned the alignment and assigned substitution models as we did with the RAxML analysis, and ran the MCMC for 100 million generations, sampling every 1000 generations. We used a Yule tree prior, default substitution priors, and a strict clock model for branch length estimation. We used Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/) to assess whether the parameters reached convergence based on ESS values, and based on this assessment we discarded the first 10% of MCMC samples as a burn-in. From the postburn-in distribution of trees, we randomly selected 100 trees for bGMYC. We ran bGMYC on a single tree to assess MCMC and burn-in length, checking parameter convergence with likelihood plots. Based on this initial run we ran bGMYC for all 100 trees for 20,000 iterations with a burn-in of 10,000 and thinning set to 10. We chose a conspecific probability cutoff of ≥0.05 to prevent oversplitting. R scripts for the bGMYC analysis are available at https://github.com/adsweet/OTU_analyses.

COPHYLOGENETIC ANALYSIS

We used both event-based and distance-based cophylogenetic methods. In all analyses, we used the *Physconelloides* RAxML tree trimmed to one representative for each OTU. We also removed the outgroup. For the host tree, we used the small New World ground-dove ML phylogeny modified from Sweet and Johnson (2015).

For an event-based approach, we used Jane v.4 (Conow et al. 2010), which uses a Genetic Algorithm (GA) to find optimal solutions of evolutionary events (cospeciation, host switching, etc.) that reconcile host and parasite trees. We set the GA parameters to 500 generations and a population size of 1000, and used default event costs (0 cospeciation, one duplication, two duplication and host switch, one loss, and one failure to diverge). We also forced host and parasite nodes to be in one of two time zones. After solving for the most optimal solutions, we randomized the tip associations 999 times to test for the statistical significance of our optimal score.

For distance-based approaches, we used both ParaFit (Legendre et al. 2002) and PACo (Balbuena et al. 2013). ParaFit tests for random association between host and parasite trees through a global statistic, and tests the relative contribution of each host-parasite association (link) to the overall congruence. Before running ParaFit, we converted the host and parasite phylogenies to patristic distance matrices in APE and sorted each matrix to be consistent with the order of the association matrix. We then ran ParaFit for 100,000 iterations using the Cailliez correction for negative eigenvalues, and tested for the contribution of individual links with the ParaFitLink1 and ParaFitLink2 tests. Because the ParaFitLink tests are multiple tests, we corrected resulting *P*-values with the Benjamini–Hochberg control for false discovery rate (Benjamini and Hochberg 1995). PACo also assess congruence between host and parasite phylogenies, but by testing the dependence of the parasite phylogeny on the host phylogeny through a Procrustes superimposition. We ran PACo for 1000 iterations using the PACO R package (Hutchinson et al. 2017), and estimated the squared residuals for each association using the PACo jackknife method. A low value indicates congruence between a host and its associated parasite. We then tested whether the squared residual values for links from sister taxa with corresponding cospeciation events (from Jane) were significantly

lower than the other links. We compared the two sets of values with a Welch's *t*-test. We also compared the squared residual values of links that had significant ParaFitLink1 results to all other links (Pérez-Escobar et al. 2015).

TESTING FOR BIOGEOGRAPHIC STRUCTURE

To test for significant biogeographic structure in ground-dove body lice, we used the Maddison–Slatkin test on the concatenated alignment phylogeny (Maddison and Slatkin 1991). The Maddison–Slatkin test randomizes character states over a topology to test for significant phylogenetic structure for the given character, in this case biogeographic region. We assigned tips to biogeographic regions similarly to Sweet and Johnson (2016) – Andean, eastern South America (cis-Andean), western South America (trans-Andean and Andean slopes), southern Central America (from the Isthmus of Tehuantepec to Panama), or southern United States/northern Mexico. Before running the test, we removed duplicate taxa by collapsing two tips if the lice were from the same host species, biogeographic region, and were separated by short branch lengths. Including all tips could bias the results toward significance. The trimmed full phylogeny had 18 tips. We also tested for biogeographic structure within *P. eurysema* 3, the most widespread and diverse clade within *P. eurysema* (19 total samples). Because this application of the Maddison–Slatkin test was at the within-species level, we did not remove any taxa from the *P. eurysema* 3 clade. We ran the Maddison–Slatkin tests with an R script (available at https://github.com/juliema/publications/tree/master/BrueeliaMS) randomizing the biogeographic states 999 times.

ESTIMATING POPULATION STRUCTURE

For population-level analyses, we used the filtered SNPs called from GATK as input to STRUCTURE to assign individuals to clusters (Pritchard et al. 2000). However, linked SNPs can bias the STRUCTURE cluster estimates. To overcome this issue, we used a custom Python script to randomly select one SNP per locus from a VCF file (available at https://github.com/adsweet/population_genomic_scripts). A similar approach is taken in popular RAD-seq processing software STACKS (Catchen et al. 2011) and iPyrad (Eaton 2014). We generated three subsets of random SNPs for separate STRUCTURE runs. We then ran STRUCTURE with 20 independent runs of 100,000 MCMC iterations (after 100,000 burnin iterations) for $K = 2$–15. We determined the most likely value of K following the delta K method of Evanno et al. (2005) estimated across all runs in the web version of STRUCTURE Harvester v.0.6.94 (Earl and VonHoldt 2012). We summarized the runs using the greedy algorithm in CLUMPP v.1.1.2 (Jakobsson and Rosenberg 2007), and used the output from CLUMPP to construct STRUCTURE plots using distruct v.1.1 (Rosenberg 2004).

As an additional estimate of population structure, we used all *P. eurysema* SNPs to perform Discriminant Analysis of Principle Components (DAPC) in the R package ADEGENET (Jombart 2008). We also conducted Principal Component Analysis (PCA) in ADEGENET for *P. eurysema* 3. We subsampled the SNPs for *P. eurysema* 3, and also filtered out missing and monomorphic SNPs, with vcftools v. 0.1.14 (Danecek et al. 2011).

Finally, we constructed a Median-Joining (MJ) network in PopART v.1.7 (Leigh and Bryant 2015) for *P. eurysema* 3 using the concatenated mtDNA alignment. PopART does not allow missing data, so columns with missing data or ambiguities were masked by the program. We set epsilon to 0.

Results

GENOMIC SEQUENCING

Each of the three samples sequenced at a high coverage (seq. pool 4 in Table S1) produced an average of 34,986,920 reads per sample after cleanup steps, which amounts to an average predicted sequencing depth of 28X based on a 200 Mbp genome (Table S1). The *P. emersoni* sample subsequently used as an assembly reference had 48,122,466 reads and an estimated sequencing depth of 38X after cleanup (Sample 1, Table S1). Sequencing between 8 and 12 samples per Illumina lane produced an average of 16,302,251 reads and an average predicted depth of 13X per sample after cleanup steps (Table S1). BLAST searches on COI data assembled for each sample with aTRAM indicated none of the samples were cross-contaminated.

DATA ASSEMBLY AND SNP CALLING

aTRAM assembled 1095 nuclear loci from the high coverage *P. emersoni* genome library using 1107 *P. h. humanus.* (human body louse) reference loci. For 46% of the assembled loci, aTRAM assembled greater than 90% of the target sequence length. Seventy-one percent of the loci retained greater than 50% of the target length, and all loci retained more than 10% of the target length (Table S1). Thirty-seven loci were removed from the reference set based on the reciprocal-best-BLAST test, leaving 1058 assembled loci as a reference set for subsequent reference-based mapping. Using the 1058 target loci as references, Bowtie2 assembled an average of 1055 orthologous loci for each high and low coverage sample (Table S1). This value includes loci that were both successfully assembled and successfully passed through the filtering pipeline. Some assemblies were involuntarily filtered out because of low coverage and/or low Phred scores. In total, however, 99.7% of the target loci were mapped and retained for subsequent analysis (only 0.3% missing data).

We used aTRAM to assemble 10 protein coding mitochondrial genes, using a median library fraction of 6.25% for the

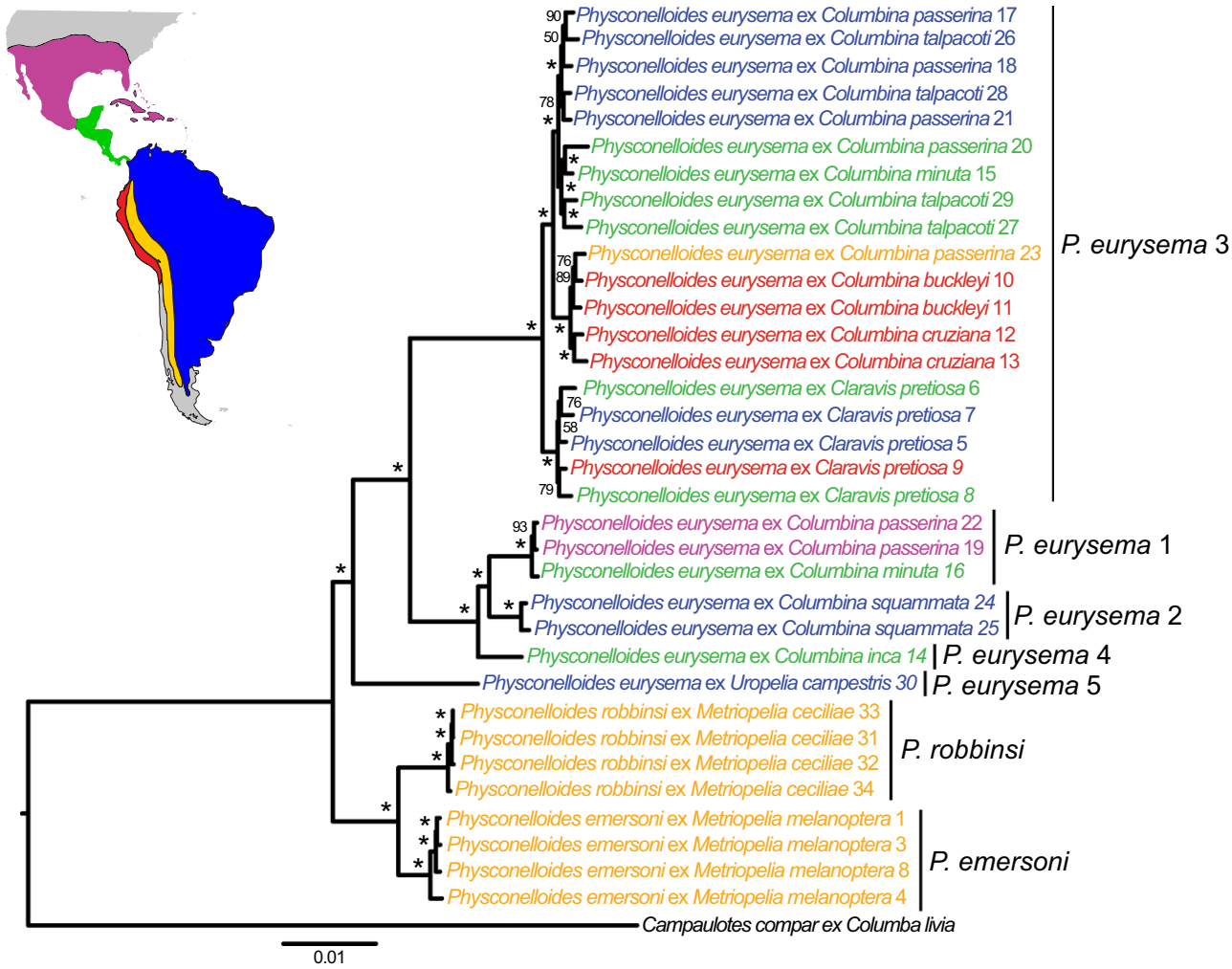

Figure 1. Evolutionary history of New World ground-dove body lice (*Physconelloides*) presented as a maximum likelihood phylogeny based on a concatenated sequence alignment of 1058 nuclear genes. Bootstrap support values are indicated at each node, and asterisks (*) indicate 100% bootstrap support. Branch lengths, as indicated by the scale bar below the phylogeny, are scaled to nucleotide substitutions per site. Vertical lines to the right of the tip labels indicate the taxa recovered from OTU analyses. Tip labels are colored according to biogeographic region, as indicated by the map in the upper left.

assemblies. The ATP8, ND3, and ND6 genes only assembled for an average of six libraries, presumably because their sequences are highly diverged from the reference, even when using a full library. These genes were excluded from future analysis. The 10 successful genes assembled for an average of 33 ingroup libraries.

The GATK software called 56,232 SNPs in the *P. eurysema* samples after filtering. However, we randomly selected three independent sets of a single SNP per locus for STRUCTURE analyses. This resulted in sets of 899, 908, and 880 SNPs. The number of SNPs did not equal the number of loci because some loci did not have had any SNPs.

PHYLOGENETIC ANALYSIS

The concatenated alignment of nuclear loci was 1,553,983 bp in length, and contained 7.8% gaps or ambiguous (N) characters (i.e., missing data). PartitionFinder indicated the concatenated alignment should be partitioned into 681 subsets. The partitioned ML phylogenetic analysis in RAxML estimated a well-supported tree, with most edges receiving 100% bootstrap support (BS). Only nine of 33 internal edges received BS support <100, and only two <75 BS (Fig. 1). The ASTRAL and ASTRID trees generated from individual gene trees mostly agreed with the topology estimated from the concatenated alignment. The ASTRAL tree was very highly supported, with most edges receiving local posterior probability support of 1.0. Any well-supported conflicts among the concatenated and coalescent trees were at short branches near the tips of the phylogenies (Figs. S1–S2).

The concatenated mitochondrial alignment was 9121 bp long. The RAxML analysis on this alignment estimated a generally well-supported tree (Fig. S3), with most edges receiving >75

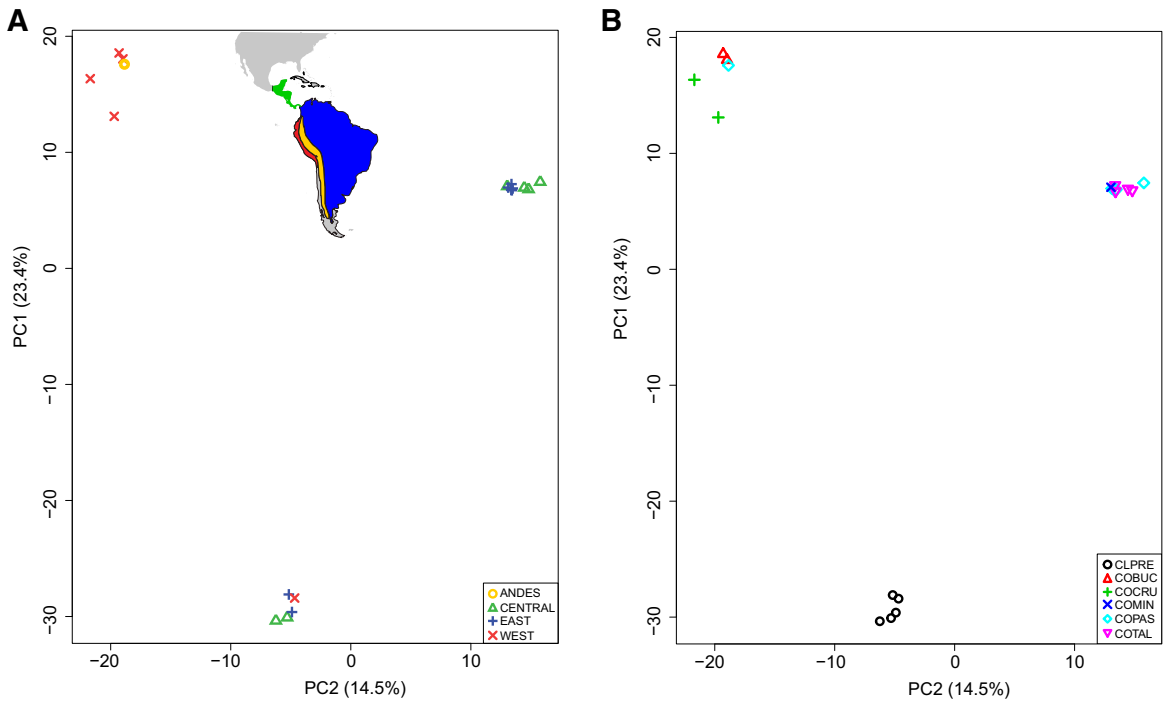

Figure 2. Population structure of *Physconelloides eurysema* 3 lice with Principal Component Analysis (PCA) based on 18,912 SNPs. Points represent individual lice, and are colored according to (A) biogeographic region or (B) host species. Host species codes are as follows: CLPRE = *Claravis pretiosa*, COBUC = *Columbina buckleyi*, COCRU = *Columbina cruziana*, COMIN = *Columbina minuta*, COPAS = *Columbina passerina*, COTAL = *Columbina talpacoti*.

BS support. However, the mtDNA tree was not as well supported as the trees based on nuclear loci, particularly at the deepest edges of the tree (<50% BS). Importantly, the mitochondrial and nuclear trees did not have any well-supported differences at deeper nodes or long branches. There were well-supported relationship differences at shorter edges near the tips of the phylogenies (i.e., within OTUs, Fig. S3–S4).

OTU ANALYSIS

Formal OTU analysis with the mitochondrial data indicated several cryptic lineages within *P. eurysema*. The ABGD method, based on COI pairwise distances, suggested seven total OTUs in the group regardless of the distance model: the two species from *Metriopelia* doves (*P. emersoni* and *P. robbinsi*) and five OTUs within *P. eurysema*. The bGMYC analysis, based on the BEAST ultrametric tree from all the mitochondrial data, also estimated seven total OTUs (two *Metriopelia* lice OTUs and five *P. eurysema* OTUs) at the 5% conspecific posterior probability cutoff. At the 95% cutoff, the analysis estimated seven total *P. eurysema* OTUs (nine total). However, the 5% cutoff is a more conservative approach to splitting taxa and perhaps more appropriate in this case.

BIOGEOGRAPHIC STRUCTURE

The Maddison–Slatkin randomization test for biogeographic structure was not significant across the phylogeny ($P = 0.154$). There were nine observed transitions on the tree after collapsing identical taxa, and ten median transitions from the character state randomizations (Fig. S5). In contrast, the randomization test for *P. eurysema* 3 was significant ($P = 0.004$), indicating the phylogeny within this clade is significantly structured according to biogeography (Fig. S6). Analyses using nuclear data grouped *P. eurysema* 3 from west of the Andes in a very distinctive cluster (Figs. 1 and 2). One oddity in this cluster is an individual louse from *Columbina passerina* sampled from a high elevation site (>2000 m.) in the Andes. Nevertheless, this is likely a "western" *P. eurysema* 3 louse. It may be that *C. passerina* have recently dispersed into higher elevation sites with agricultural development, as has been documented in other ground-dove species (Pearson 1975). *P. eurysema* 3 from east of the Andes and Central America also formed distinct clusters. The MJ network from mitochondrial data showed similar biogeographic structure for *P. eurysema* 3, except for lice from *C. passerina* sampled from eastern South America (Fig. 3). These clustered separately from other lice sampled from the same region.

COPHYLOGENETIC ANALYSIS

Jane recovered three cospeciation events between ground-doves and their body lice: at the *Metriopelia* split, at the *Metriopelia*/*Columbina* split, and at the *Columbina squammata*/*C. inca* split (Fig. 4; Fig. S7). However, the latter cospeciation event had

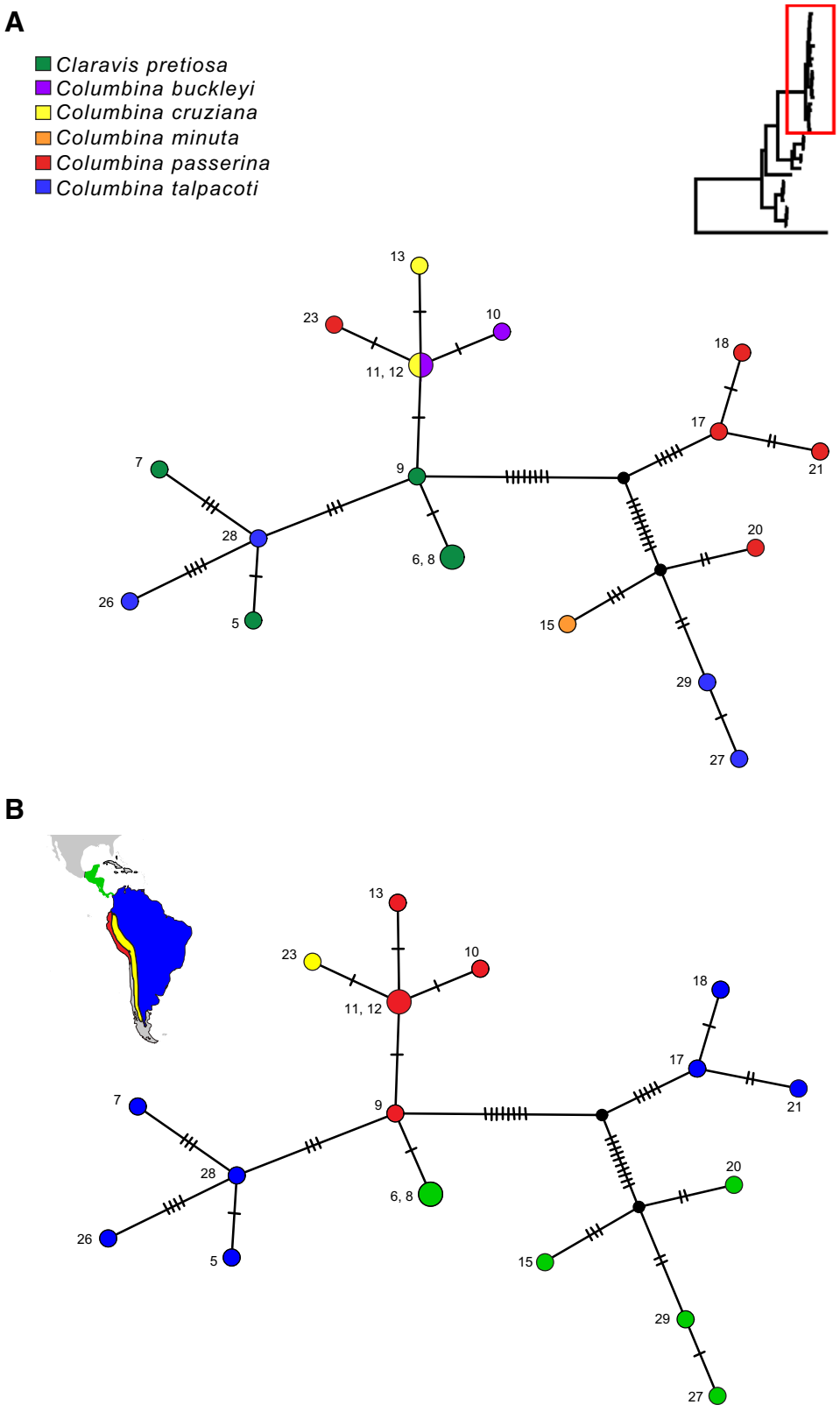

Figure 3. Population structure of *Physconeloides eurysema* 3 presented as a median-joining network generated from ten mitochondrial genes. Taxa are indicated with the red box on the phylogeny (from the concatenated nuclear data) in the upper right. Node size is proportional to the number of individuals in a haplotype. Numbers adjacent to each node represent individuals as indicated in Table S1. Tick marks indicate the number of steps between haplotypes. Nodes are colored according to (A) host species and (B) biogeographic region.

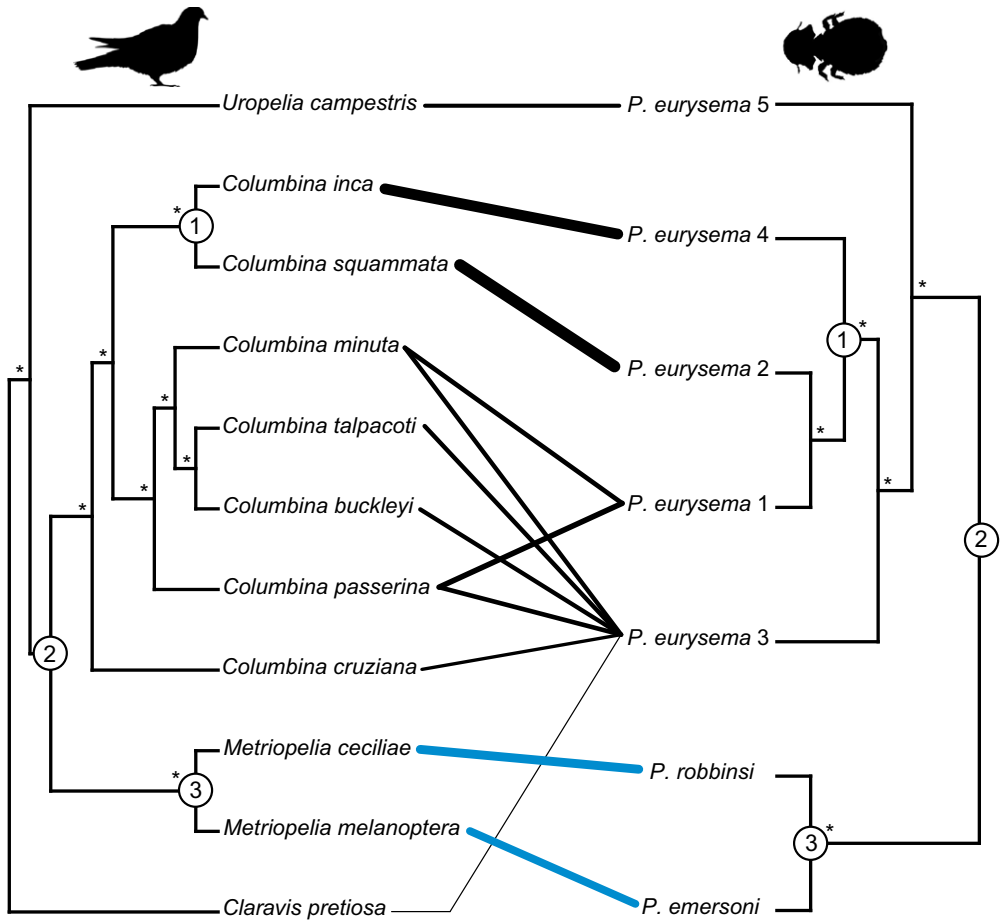

Figure 4. **Tanglegram comparing the evolutionary histories of small New World ground-doves (left) and their *Physconelloides* body lice (right). The host phylogeny is adapted from Sweet and Johnson (2015). The louse phylogeny is the species tree recovered from OTU analyses. Relationships with significant support (>75 bootstrap) are indicated with asterisks (*). Host-parasite link thickness is inversely proportional to the PACo jackknifed squared residuals (i.e., thicker links indicate a higher contribution to congruence). Blue links indicate significant ParaFitLink tests after correction ($\alpha = 0.05$). Circles above nodes indicate cospeciation events recovered from Jane. Numbers inside the circles indicate corresponding speciation events. Dove silhouette from Phylopic (http://phylopic.org/) courtesy of Luc Viatour and Andreas Plank.**

an equally parsimonious placement at the split of all *Columbina* minus *C. cruziana*. Jane also recovered two host switches: one from the common ancestor of *Columbina* to *Uropelia*, and a second from *C. squammata* to the ancestor of *C. minuta*, *C. buckleyi*, *C. talpacoti*, and *C. passerina.* Finally, Jane recovered one duplication event (at the base of *Columbina)*, five losses, and six failures to diverge with the hosts (Fig. S7). The randomization test indicated the best cost was lower than expected ($P = 0.002$), suggesting the host and parasite phylogenies are overall significantly congruent.

Both the ParaFit (ParaFitGlobal = $9.86e^{-5}$, $P = 0.002$) and PACo ($m^2 = 0.06$, $P = 0.005$) global tests indicated significant congruence between the host and parasite phylogenies (Table 1). Two links were significant from the ParaFit individual link tests after correction for multiple tests ($\alpha = 0.05$): *Metriopelia melanoptera* and *P. emersoni* ($P = 0.007$), and *Metriopelia ceciliae* and *P. robbinsi* ($P = 0.006$). The ParaFitLink1 and ParaFitLink2 statistics gave similar *P* values. The PACo jackknife test for individual link contribution indicated the links for sister taxa with possible cospeciation events (*Metriopelia* and the *C. squammata*/*C. inca* split) had significantly lower squared residuals than the other links in the group ($t = -3.32$, $P = 0.008$; Fig. S8). These four associations had the lowest squared residual values. The squared residual values for significant ParaFitLink1 links were also significantly lower than the other links ($t = -2.27$, $P = 0.045$; Fig. S9).

POPULATION STRUCTURE

Population-level analysis of *P. eurysema* indicated significant structure from both nuclear and mitochondrial data. Based on

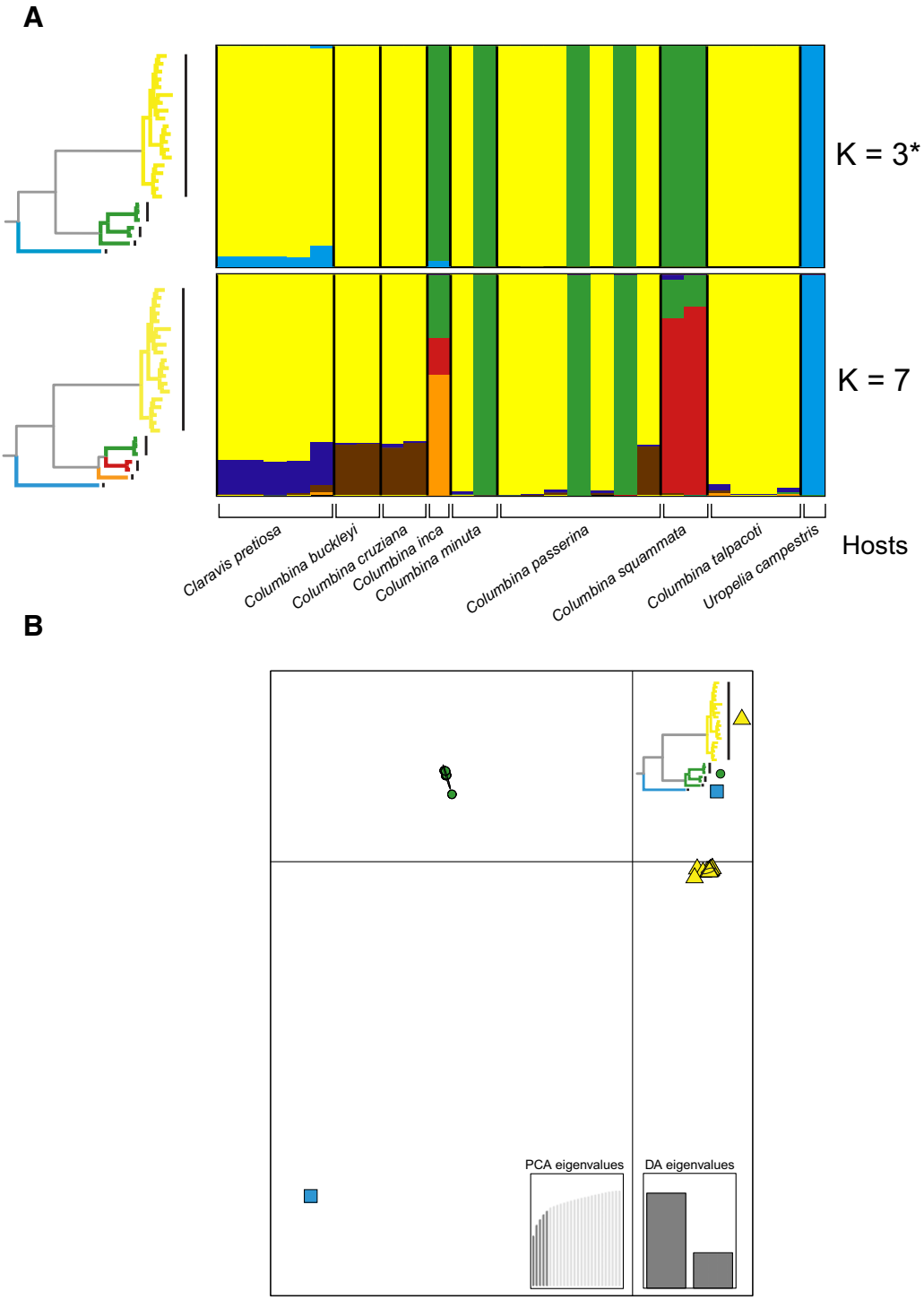

Figure 5. Population structure of *Physconelloides eurysema* lice from small New World ground-doves based on genome-wide SNPs. (A) STRUCTURE plot from 908 randomly sampled unlinked SNPs and (B) discriminant analysis of principal components (DAPC) plot based on 56,232 SNPs. For the STRUCTURE plot, individual lice are grouped according to host species, and colored according to the likelihood of being in a particular cluster. Phylogenies to the left of the STRUCTURE plots are modified from the *Physconelloides* concatenated phylogeny, and are colored according to the clusters from the STRUCTURE plot. Vertical lines to the right of the phylogenies indicate taxa recovered from the OTU analyses. *K* (number of clusters) values are indicated to the right of the STRUCTURE plots. The asterisk (*) indicates the most optimal *K* value. Points on the DAPC plot indicate individual lice. The colors and shapes indicate clusters, in accordance with the phylogeny in the upper right. The phylogeny is the same as in (A). PCA and discriminant functions used for the DAPC are indicated in the bottom left of (B).

Table 1. Results for the ParaFit analysis for small New World ground-doves and their body lice *Physconelloides*.

Host	Parasite	PF1 Statistic	PF1 *P*-value	PF2 Statistic	PF2 *P*-value
Claravis pretiosa	*Physconelloides eurysema* 3	-1.39×10^{-5}	0.964	-4.368×10^{-3}	0.966
Uropelia campestris	*P. eurysema* 5	1.85×10^{-5}	0.014	5.790×10^{-3}	0.013
Metriopelia melanoptera	*P. emersoni*	2.58×10^{-5}	0.007†	8.073×10^{-3}	0.006†
Metriopelia ceciliae	*P. robbinsi*	2.80×10^{-5}	0.006†	8.773×10^{-3}	0.006†
Columbina cruziana	*P. eurysema* 3	3.43×10^{-6}	0.309	1.074×10^{-3}	0.306
Columbina squammata	*P. eurysema* 2	8.46×10^{-6}	0.084	2.650×10^{-3}	0.081
Columbina inca	*P. eurysema* 4	8.11×10^{-6}	0.080	2.541×10^{-3}	0.076
Columbina minuta	*P. eurysema* 1	9.08×10^{-6}	0.078	2.845×10^{-3}	0.076
Columbina minuta	*P. eurysema* 3	1.35×10^{-5}	0.039	4.240×10^{-3}	0.037
Columbina buckleyi	*P. eurysema* 3	1.34×10^{-5}	0.046	4.205×10^{-3}	0.043
Columbina talpacoti	*P. eurysema* 3	1.35×10^{-5}	0.042	4.235×10^{-3}	0.039
Columbina passerina	*P. eurysema* 1	8.23×10^{-6}	0.099	2.578×10^{-3}	0.095
Columbina passerina	*P. eurysema* 3	1.16×10^{-5}	0.068	3.642×10^{-3}	0.065

†Significant after the Benjamini–Hochberg correction ($\alpha = 0.05$).

PF1 and PF2 are the statistics and P-values for the ParaFitLink1 and ParaFitLink2 tests, respectively. Numbers next to the parasite species names indicate potentially cryptic species recovered from OTU analyses.

SNP data, STRUCTURE estimated populations that largely corresponded to the major branches in the phylogenetic trees (Fig. 1, Figs. S1–S3). For all three runs from randomly sampled unlinked SNPs, STRUCTURE estimated an optimal $K = 3$ based on the Evanno method. However, despite having lower delta K values, higher levels of K showed more structure corresponding to major branches from the phylogenetic analyses (Fig. 5A, Fig. S10–S11). Using all 56,232 SNPs, ADEGENET also estimated an optimal $K = 3$. The DAPC scatterplot showed clear distinction among all three clusters (Fig. 5B). DAPC for *P. eurysema* 3 estimated $K = 2$, and showed distinction between lice from *Claravis pretiosa* and lice from other host species (Fig. S12). PCA plots based on 18,912 SNPs showed further population-level differentiation within *P eurysema* 3, with several distinct clusters of taxa (Fig. 2). The MJ network of the mitochondrial sequences also showed several well-delimited groups within *P. eurysema* 3, including some differences with the nuclear data (Fig. 3). For example, lice from *C. pretiosa* are in a well-supported clade in the nuclear phylogenies (both concatenated and coalescent) and cluster together in the PCA, but these samples do not group together in the MJ network.

Discussion

DRIVERS OF DIVERSIFICATION AT PHYLOGENETIC AND POPULATION SCALES

Incorporating both phylogenetic and population perspectives provides more information for assessments of the diversification process (Cutter 2013). This is particularly important for studies focused on parasites, organisms with diversification patterns that can be heavily dependent on host and external (e.g., biogeography) factors. Thus, diversification of parasites can potentially differ between species and population scales (Bell et al. 2016). In this study, we integrated phylogenetic and population-scale patterns estimated using genome-wide loci and SNPs from a group of parasitic lice. For this endeavor, we developed and applied a novel workflow to assemble orthologous loci and call SNPs for use in both phylogenetic and population genetic analyses. Because we had shotgun genome sequencing reads available for each individual louse, we were also able to assemble most mitochondrial genes. Other forms of genomic-level data (e.g., UCEs, RADseq, anchored-hybrid enrichment) are incredibly useful, but are restricted to specific regions of the genome, or are more appropriate for either phylogenetic or population genetic analysis. Our approach allowed us to obtain multiple types of molecular data sets from the same raw sequence data. This could be applied beyond host-parasite or host-symbiont systems, and be used to uncover patterns of diversification in any group of organism. It will be particularly useful for groups with multiple individual samples of a few closely related species, as we have done here with dove body lice. These systems that straddle the population-species boundary are ideal for exploring diversification at multiple time scales (Russell et al. 2007). Another strength of this approach is the use of aTRAM to generate reference sequences, which is useful for groups that lack a closely related reference genome.

For this study on dove body lice (*Physconelloides*), our results suggest that host and biogeographic factors can have similar or varying effects on parasite diversification over time, patterns that would have been obscured using a traditional approach

focusing on only phylogenetic or only population genetic patterns. Host associations do appear to dictate parasite divergence patterns at both phylogenetic and population genetic scales. Three (50%) of the nodes within the body louse phylogeny are inferred to be cospeciation events, and five (of seven) louse taxa are host specific. Within species, some louse population clusters also appear to be confined to a single host species. For example, all lice from *Claravis pretiosa* cluster together, regardless of sampling location, in both full locus (Fig. 1) and SNP-based (Fig. 2B) analyses. Likely, this is because body lice are closely tied to their hosts and are not able to easily disperse among host species. Patterns of host-driven divergence have been observed in other host–parasite systems at both species (e.g., gopher lice, Hafner et al. 1994; teleost copepods, Paterson and Poulin 1999; avian malarial parasites, Ricklefs and Fallon 2002; bat mites, Brudydonckx et al. 2009) and population scales (e.g., snail trematodes, Dybdahl and Lively 1996; Galapagos hawk lice, Whiteman et al. 2007; rodent mites, Engelbrecht et al. 2016). Here, we show patterns occurring at both scales in the same system. In addition, this pattern of phylogenetic congruence has been observed in broader studies of dove body lice (Clayton and Johnson 2003; Sweet et al. 2016a). However, worldwide, other groups of dove body lice do not show phylogenetic congruence with their hosts, so there is certainly variability within the dove body louse system (Sweet et al. 2016b).

Not all population-level patterns in the body louse system exhibit congruence and specificity with their hosts. In several host-parasite systems, including wing lice from the same group of ground-dove host species, biogeography is a good predictor of diversification and codiversification patterns (e.g., toucan lice, Weckstein 2004; southern beech fungus, Peterson et al. 2010; rodent lice, du Toit et al. 2013; digeneans of freshwater fish, Martínez-Aquino et al. 2014; ground-dove wing lice, Sweet and Johnson 2016). At the phylogenetic timescale in ground-dove body lice, biogeography does not seem to dictate diversification. In contrast, ground-dove body lice are structured by biogeographic region within species. The structure within *P. eurysema* 3 was significantly associated with biogeography. Together with the patterns of host-specificity in *P. eurysema* 3 (i.e., lice from *C. pretiosa* as the earliest diverging lineage), this suggests populations of ground-dove body lice are initially structured according to biogeography, but over time eventually sort according to host species. A similar pattern of initial instability with subsequent lineage sorting has also been discussed at the cophylogenetic level in the fig/fig-wasp system (Cruaud et al. 2012). The discrepancy between phylogenetic and population patterns in our system have important implications for understanding parasite diversification, particularly for parasites with limited dispersal ability. It may be that parasites have some limited ability to disperse between sympatric host species, but over evolutionary time continued low dispersal and differential selection among host species results in host-specificity. For example, local adaptation to a given host species may prohibit parasites from successfully reproducing on a wide variety of hosts species, selecting for increased host specialization over time (Kaltz and Shykoff 1998; Gandon 2002; Clayton et al. 2003; Lively et al. 2004).

Other ecological factors, such as host species proximity or host species interactions, could also limit or promote diversification of parasites, as has been proposed in other systems (Desdevises et al. 2002; Hoberg and Brooks 2015; Bell et al. 2016). The doves associated with *P. eurysema* 1 and 3 are known to form mixed-species foraging flocks (Parker et al. 1995; Piratelli and Blake 2006). Foraging in proximity or sharing dust baths would provide an opportunity for lice to disperse among host species (Hoyle 1938; Martin and Mullens 2012). However, other ground-dove species with host-specific lice, such as *M. ceciliae* and *M. melanoptera,* also co-occur in parts of their ranges and do not appear to share lice. Perhaps more intimate relationships such as sharing nesting sites could also allow for louse dispersal (Clayton 1990; Johnson et al. 2002; Clayton et al. 2016). For example, *Columbina talpacoti* will build their nests on top of old nests from other bird species (Skutch 1956). If an individual builds a nest on the old nest of another ground-dove species, this could facilitate a host-switch if body lice are still present in the old nesting material. Finally, although body lice are not likely to use phoresis, it is possible that a low amount of phoresis might occur in this group. *Physconelloides* body lice from mourning doves (*Zenaida macroura*) have been found attached to hippoboscid flies (Couch 1962). If ground-dove body lice can disperse via phoresy, this could explain why some louse OTUs are more generalist.

DIVERSIFICATION PATTERNS AMONG GROUND-DOVE BODY LICE

Ground-dove body lice appear to be a much more diverse group than previously assumed, with evidence for seven different species (three species are currently described; Price et al. 2003). This agrees with previous molecular phylogenetic studies of dove body lice, which indicated at least two additional taxa within *P. eurysema* using limited ground-dove louse representatives (Clayton and Johnson 2003; Johnson et al. 2011b). Because most host species and geographic regions are represented, the diversity recovered in this study is likely robust to sampling. However, we cannot completely rule out that the host-specific louse OTUs are present on other host species, but at much lower prevalence.

The phylogenetic patterns also provide insight into the origin of this louse lineage. Lice from the *Metriopelia* doves (*P. robbinsi* and *P. emersoni*) are sister to the other ground-dove body lice. *Metriopelia* doves are high-Andean species, generally found >2,000 m. in open Paramo and Altiplano grasslands (Gibbs

et al. 2001). These birds diverged from other ground-doves and began diversifying ~11–14 mya, perhaps as a result of rapid elevational increases in the Andes (Sweet and Johnson 2015). However, *Metriopelia* doves are nested within the ground-dove phylogeny, and, unlike their lice, are not the earliest diverging lineage. The cospeciation event between the *Metriopelia-Columbina* split and the base of the body louse phylogeny suggests this parasite lineage diverged ~11–14 mya. At the very least, this is likely a minimum age for the group. Subsequent diversification into other ground-dove species then occurred after their divergence from the *Metriopelia* common ancestor. However, formal divergence time estimation for the lice is needed to confirm these hypotheses, which is challenging given the lack of fossil calibration points.

Phylogenetic hypotheses were highly consistent among different molecular data sets. However, there was limited contradiction between the mtDNA and nuclear data within OTUs. For example, lice from *C. pretiosa* did not group together in the mitochondrial MJ network, whereas analyses with nuclear data (both full loci and SNPs) clustered these lice together with high support (Figs. 1, 2B, and 3). It may be that the mtDNA, a single locus, has not fully sorted among populations for lice from *C. pretiosa,* whereas the signal from nuclear data has spread across >1000 loci and can detect limited current gene flow between louse populations on different host species (Pamilo and Nei 1988; McGuire et al. 2007; McKay and Zink 2010). Alternatively, this disparity between nuclear and mtDNA data may reflect dispersal differences between male and female lice. Lice from *C. pretiosa* are not randomly arranged in the MJ network, but show some phylogeographic structure. Lice sampled from Central America and western South America, two connected biogeographic regions, group with other lice from western South America (# 6, 8, and 9 in Fig. 3). Likewise, lice from eastern South America group with other lice from the same region (# 5 and 7 in Fig. 3). If female lice from *C. pretiosa* are more able to disperse than males, this could result in mtDNA phylogeographic structure not evident in nuclear data.

CONCLUSION

In this study, we used full genome sequence data to show that parasite diversification is shaped by multiple factors that have varying effects over time. In our system, ground-dove body lice, host association seems to be important at both deep and shallow time scales, whereas biogeography only explains patterns at shallow scales. Central to this result is the integration of phylogenetic and population genetic approaches using the same underlying data source. Excluding either approach would have masked patterns of host specificity or phylogeographic structure. We recommend that future studies interested in understanding host-parasite codiversification take a similar approach. Additionally, the utility and flexibility of whole genome sequencing made it possible to obtain various types of data sets (nuclear and mitochondrial loci, SNPs) from individual specimens, using a novel assembly workflow. Our approach has great promise for addressing questions in evolutionary biology with genomic data, particularly for groups of organisms along the population-species boundary or which do not have a closely related reference genome.

DATA ARCHIVAL LOCATION

Raw genomic reads are deposited in the GenBank SRA database under accession SRP076185; phylogenetic tree files, sequence alignments, voucher photos, and other relevant files are deposted on Dryad https://doi.org/10.5061/dryad.02sk7.

AUTHOR CONTRIBUTIONS

A.D.S. and K.P.J. designed the study; A.D.S., M.P.V., J.L.R.P., and R.E.W. conducted field work; S.M.V. photographed specimens; A.D.S. generated genomic data; A.D.S., B.M.B., J.M.A., and K.P.J. analyzed data; and A.D.S., B.M.B., J.M.A., S.M.V., M.P.V., R.E.W., and K.P.J. wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank the following individuals and institutions for providing louse samples for this study and assisting in field collecting: Dale H. Clayton, Kevin G. McCracken, Robert Faucett, Edith Montalvo, Edison Auqui, Jeff DaCosta, Peggy Guttan-Mayerma, John Klicka, Matthew Miller, and Garth Spellman. Alvaro Hernandez and Chris Wright at the Roy J. Carver Biotechnology Center (University of Illinois) assisted with Illumina library preparation and sequencing. Jeff Haas and the Life Science Biocomputing Support Team (University of Illinois) helped with server and software maintenance. We also thank Mohamed Noor, Jessica Light, and two anonymous reviewers for helping to improve this manuscript. This work was supported by a UIUC Graduate College Dissertation Travel Grant and American Ornithologists' Union Student Research Award to A.D.S.; National Science Foundation grants DEB-1239788 and DEB-1342604 to K.P.J.; FAPESP– São Paulo Research Foundation 2011/11420-5 to M.P.V.; and by the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project number ILLU 875–952.

LITERATURE CITED

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. 19:716–723.

Allen, J. M., B. Boyd, N. P. Nguyen, P. Vachaspati, T. Warnow, D. I. Huang, P. G. Grady, K. C. Bell, Q. C. Cronk, L. Mugisha, et al. 2017. Phylogenomics from whole genome sequences using aTRAM. Syst. Biol. 66:786–798.

Allen, J. M., D. I. Huang, Q. C. Cronk, and K. P. Johnson. 2015. aTRAM—automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. BMC Bioinformatics 16:98.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Balbuena, J. A., R. Míguez-Lozano, and I. Blasco-Costa. 2013. PACo: a novel procrustes application to cophylogenetic analysis. PLoS One 8:e61048.

Bell, K. C., K. L. Calhoun, E. P. Hoberg, J. R. Demboski, and J. A. Cook. 2016. Temporal and spatial mosaics: deep host association and

shallow geographic drivers shape genetic structure in a widespread pinworm, *Rauschtineria eutamii* (Nematoda: Oxyuridae). Biol. J. Linn. Soc. 119:397–413.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57:289–300.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Boyd, B. M., J. M. Allen, N. Nguyen, A. D. Sweet, T. Warnow, M. D. Shapiro, S. M. Villa, S. E. Bush, D. H. Clayton, and K. P. Johnson. 2017. Phylogenomics using target-restricted assembly resolves intrageneric relationships of parasitic lice (Phthiraptera: *Columbicola*). Syst. Biol. 66:896–911.

Brower, A. V. Z. 2006. Problems with DNA barcodes for species delimitation: "Ten species" of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst. Biodivers. 4:127–132.

Bruyndonckx, N., S. Dubey, M. Ruedi, and P. Christe. 2009. Molecular cophylogenetic relationships between European bats and their ectoparasitic mites (Acari, Spinturnicidae). Mol. Phylogenet. Evol. 51:227–237.

Bush, G. L. 1975. Modes of animal speciation. Annu. Rev. Ecol. Syst. 6:339–364.

Capella-Gutiérrez, S., J. M. Silla-Martinez, and T. Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Carling, M. D., and R. T. Brumfield. 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. Genetics 178:363–377.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. *Stacks*: building and genotyping loci de novo from short-read sequences. Genes Genomes Genet. 1:171–182.

Clayton, D. H. 1990. Host specificity of Strigiphilus owl lice (Ischnocera: Philopteridae), with the description of new species and host associations. J. Med. Entomol. 27:257–265.

Clayton, D. H., S. E. Bush, B. M. Goates, and K. P. Johnson. 2003. Host defense reinforces host-parasite cospeciation. PNAS 100:15694–15699.

Clayton, D. H., S. E. Bush, and K. P. Johnson. 2016. Coevolution of life on hosts: integrating ecology and history. Chicago Univ. Press, Chicago.

Clayton, D. H., and D. M. Drown. 2001. Critical evaluation of five methods for quantifying chewing lice (Insecta: Phthiraptera). J. Parasitol 87:1291–1300.

Clayton, D. H., and K. P. Johnson. 2003. Linking coevolutionary history to ecological process: doves and lice. Evolution 57:2335–2341.

Conow, C., D. Fielder, Y. Ovadia, and R. Libeskind-Hadas. 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol. Biol. 5:16.

Couch, A. B. 1962. Phoretic mallophagans from hippoboscids of mourning doves *Zenaidura macroura*. J. Parasitol. 48:497.

Criscione, C. D. 2008. Parasite co-structure: broad and local scale approaches. Parasite 15:439–443.

Criscione, C. D., R. Poulin, and M. S. Blouin. 2005. Molecular ecology of parasites: elucidating ecological and microevolutionary processes. Mol. Ecol. 14:2247–2257.

Cruaud, A., Ronsted, N., Chantarasuwan, B., Chou, L. S., Clement, W. L., Couloux, A., Cousins, B., Genson, G., Harrison, R. D., Hanson, P. E., et al. 2012. An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. Syst. Biol. 61:1029–1047. https://doi.org/10.1093/sysbio/sys068

Cutter, A. D. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. Mol. Phylogenet. Evol. 69:1172–1185.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, et al. 2011. The variant call format and VCFtools.Bioinformatics 27:2156–2158.

de Vienne, D. M., G. Refrégier, M. López-Villavicencio, A. Tellier, M. E. Hood, and T. Giraud. 2013. Coespeciation vs hos-shift speciation: methodsfor testing, evidence from natural associations and ralation to coevolution. New Phytol. 198:347–385.

Desdevises, Y., S. Morand, O. Jousson, and P. Legendre. 2002. Coevolution between Lamellodiscus (Monogenea: Diplectanidae) and Sparidae (Teleostei): the study of a complex host-parasite system. Evolution 56:2459–2471.

du Toit, N., B. J. Van Vuuren, S. Matthee, and C. A. Matthee. 2013. Biogeography and host-related factors trump parasite life history: limited congruence among the genetic structures of specific ectoparasitic lice and their rodent hosts. Mol. Ecol. 22:5185–5204.

Dybdahl, M. F., and C. M. Lively. 1996. The geography of coevolution: comparative population structures for a snail and its trematode parasite. Evolution 50:2264–2275.

Earl, D. A., and B. M. VonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Eaton, D. A. R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Eichler, W. D. 1948. Some rules in ectoparasitism. Ann. Mag. Nat. Hist. 1:588–598.

Engelbrecht, A., S. Matthee, N. Du Toit, and C. A. Matthee. 2016. Limited dispersal in an ectoparasitic mite, *Laelaps giganteus*, contributes to significant phylogeographic congruence with the rodent host, Rhabdomys. Mol. Ecol. 25:1006–1021.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Fahrenholz, H. 1913. Ectoparasiten und abstammungslehre. Zool. Anz. 41:371–374.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726.

Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–488.

Gandon, S. 2002. Local adaptation and the geometry of host–parasite coevolution. Ecol. Lett. 5:246–256.

Gibbs, D. E., E. Cox, and J. Cox. 2001. Pigeons and doves: A guide to the pigeons and doves of the world. Pica Press, Sussex.

Harbison, C. W., S. E. Bush, J. R. Malenke, and D. H. Clayton. 2008. Comparative transmission dynamics of competing parasite species. Ecology 89:3186–3194.

Harbison, C. W., M. V. Jacobsen, and D. H. Clayton. 2009. A hitchhiker's guide to parasite transmission: the phoretic behaviour of feather lice. Int. J. Parasitol. 39:569–575.

Harrison, L. 1914. The Mallophaga as a possible clue to bird phylogeny. Aust. Zool. 1:7–11.

Harrison, R. G. 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. *In* D. J. Howard, and S.

H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford.

Hebert, P. D. N., and T. R. Gregory. 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54:852–859.

Hoberg, E. P., and D. R. Brooks. 2015. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. Philos. Trans. R Soc. B Biol. Sci. 370:20130553.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6:e1000862.

Hoyle, W. L. 1938. Transmission of poultry parasites by birds with special reference to the "English" or house sparrow and chickens. Trans. Kansas Acad. Sci. 41:379–384.

Hutchinson, M. C., E. F. Cagua, J. A. Balbuena, D. B. Stouffer, and T. Poisot. 2017. paco: implementing procrustean approach to cophylogeny in R. Methods Ecol. Evol. 8:932–940.

Huyse, T., R. Poulin, and A. Théron. 2005. Speciation in parasites: a population genetics approach. Trends Parasitol. 21:469–475.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Jirsová, D., J. Štefka, and M. Jirků. 2017. Discordant population histories of host and its parasite: a role for ecological permeability of extreme environment? PLoS One 12:e0175286.

Johnson, K., and D. Clayton. 2004. Untangling coevolutionary history. Syst. Biol. 53:92–94.

Johnson, K. P., R. J. Adams, and D. H. Clayton. 2002. The phylogeny of the louse genus Brueelia does not reflect host phylogeny. Biol. J. Linn. Soc. 77:233–247.

Johnson, K. P., and D. H. Clayton. 2000. Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). Mol. Phylogenet. Evol. 14:141–151.

———. 2003a. Coevolutionary history of ecological replicates: comparing phylogenies of wing and body lice to Columbiform hosts. Pp. 260–286 *in* R. D. M. Page, ed. Tangled trees: phylogeny, cospeciation, and coevolution. Chicago Univ. Press, Chicago.

———. 2003b. The biology, ecology, and evolution of chewing lice. *In* R. D. Price, R. A. Hellenthal, R. L. Palma, K. P. Johnson, and D. H. Clayton, eds. The chewing lice: World checklist and biological overview. Illinois Natural History, Illinois.

Johnson, K. P., K. K. O. Walden, and H. M. Robertson. 2013. Next-generation phylogenomics using a target restricted assembly method. Mol. Phylogenet. Evol. 66:417–422.

Johnson, K. P., J. D. Weckstein, S. E. Bush, and D. H. Clayton. 2011a. The evolution of host specificity in dove body lice. Parasitology 138:1730–1736.

Johnson, K. P., J. D. Weckstein, M. J. Meyer, and D. H. Clayton. 2011b. There and back again: switching between host orders by avian body lice (Ischnocera: Goniodidae). Biol. J. Linn. Soc. 102:614–625.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Kaltz, O., and J. A. Shykoff. 1998. Local adaptation in host–parasite systems. Heredity 81:361–370.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kirkness, E. F., B. J. Haas, W. Sun, H. R. Braig, M. A. Perotti, J. M. Clark, H. Lee, H. M. Robertson, R. C. Kennedy, D. Gerlach, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. PNAS 108:6335.

Kochzius, M., C. Seidel, J. Hauschild, S. Kirchhoff, P. Mester, I. Meyer-Wachsmuth, A. Nuryanto, and J. Timm. 2009. Genetic population structures of the blue starfish Linckia laevigata and its gastropod ectoparasite *Thyca crystallina*. Mar. Ecol. Prog. Ser. 396:211–219.

Lanfear, R., B. Calcott, D. Kainer, C. Mayer, and A. Stamatakis. 2014. Selecting optimal partitioning schemes for phylogenomic data sets. BMC Evol. Biol. 14:82.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Leaché, A. D., M. K. Fujita, V. N. Minin, and R. R. Bouckaert. 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63:534–542.

Legendre, P., Y. Desdevises, and E. Bazin. 2002. A statistical test for host–parasite coevolution. Syst. Biol. 51:217–234.

Leigh, J. W., and D. Bryant. 2015. Popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61:727–744.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Lively, C. M., M. F. Dybdahl, J. Jokela, E. E. Osnas, and L. F. Delph. 2004. Host sex and local adaptation by parasites in a snail-trematode interaction. Am. Nat. 164:S6–S18.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. 4:981–994.

Maddison, W. P., and M. Slatkin. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. Evolution 45:1184–1197.

Manthey, J. D., L. C. Campillo, K. J. Burns, and R. G. Moyle. 2016. Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). Syst. Biol. 65:640–650.

Martin, C. D., and B. A. Mullens. 2012. Housing and dustbathing effects on northern fowl mites (*Ornithonyssus sylviarum*) and chicken body lice (*Menacanthus stramineus*) on hens. Med. Vet. Entomol. 26:323–333.

Martínez-Aquino, A. 2016. Phylogenetic framework for coevolutionary studies: a compass for exploring jungles of tangled trees. Curr. Zool. 62:393–403.

Martínez-Aquino, A., F. S. Ceccarelli, L. E. Eguiarte, E. Vázquez-Domínguez, and G. Pérez-Ponce De León. 2014. Do the historical biogeography and evolutionary history of the digenean *Margotrema* spp. across central Mexico mirror those of their freshwater fish hosts (Goodeinae)? PLoS One 9:e101700.

McCoy, K. D., T. Boulinier, and C. Tirard. 2005. Comparative host-parasite population structures: disentangling prospecting and dispersal in the black-legged kittiwake *Rissa tridactyla*. Mol. Ecol. 14:2825–2838.

Mccoy, K. D., T. Boulinier, C. Tirard, and Y. Michalakis. 2001. Host specificity of a generalist parasite: genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. J. Evol. Biol. 14:395–405.

McGuire, J. A., C. W. Linkem, M. S. Koo, D. W. Hutchison, A. K. Lappin, D. I. Orange, J. Lemos-Espinal, B. R. Riddle, and J. R. Jaeger. 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. Evolution 61:2879–2897.

McKay, B. D., and R. M. Zink. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. Mol. Phylogen. Evol. 54:647–650.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Pp. 1–8 *in* Proceedings of the gateway computing environments workshop (GCE). New Orleans, LA.

Mirarab, S., and T. Warnow. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.

Morand, S. 2012. Phylogeography helps with investigating the building of human parasite communities. Parasitology 139:1966–1974.

Nadler, S. 1995. Microevolution and the genetic structure of parasite populations. J. Parasitol. 81:395–403.

Nadler, S. A., and G. P. P. De Leon. 2011. Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology. Parasitology 138:1688–1709.

Page, R. D. M. 1994. Parallel phylogenies: reconstructing the history of host-parasite assemblages. Cladistics 10:155–173.

Page, R. D. M., and M. A. Charleston. 1998. Trees within trees: phylogeny and historical associations. Trends Ecol. Evol. 13:356–359.

Page, R. D. M., R. H. Cruickshank, M. Dickens, R. W. Furness, M. Kennedy, R. L. Palma, and V. S. Smith. 2004. Phylogeny of "*Philoceanus* complex" seabird lice (Phthiraptera: Ischnocera) inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 30:633–652.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Parker, T. A., T. S. Schulenberg, M. Kessler, and W. H. Wust. 1995. Natural history and conservation of the endemic avifauna in north-west Peru. Bird Cons. Intl. 5:201–231.

Paterson, A. M., and R. Poulin. 1999. Have chondracanthid copepods co-speciated with their teleost hosts? Syst. Parasitol. 44:79–85.

Pearson, D. L. 1975. Range extensions and new records for bird species in Ecuador, Peru, and Bolivia. Condor 77:96–99.

Pereira, S. L., K. P. Johnson, D. H. Clayton, and A. J. Baker. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleocene. Syst. Biol. 56:656–672.

Pérez-Escobar, O. A., J. Antonio Balbuena, and M. Gottschling. 2015. Rumbling orchids: how to assess divergent evolution between chloroplast endosymbionts and the nuclear host. Syst. Biol. 65:51–65.

Peterson, K. R., D. H. Pfister, and C. D. Bell. 2010. Cophylogeny and biogeography of the fungal parasite Cyttaria and its host Nothofagus, southern beech. Mycologia 102:1417–1425.

Piratelli, A., and J. G. Blake. 2006. Bird communities of the southeastern Cerrado Region, Brazil. Ornitol. Neotrop. 17:213–225.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55: 595–609.

Poulin, R., B. R. Krasnov, and D. Mouillot. 2011. Host specificity in phylogenetic and geographic space. Trends Parasitol. 27:355–361.

Price, R. D., R. A. Hellenthal, R. L. Palma, K. P. Johnson, and D. H. Clayton. 2003. The chewing lice: world checklist and biological overview. Illinois Natural History Survey, Champaign, IL.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz. 2012. ABGD, automatic barcode gap discovery for primary species delimitation. Mol. Ecol. 21:1864–1877.

Refrégier, G., M. Le Gac, F. Jabbour, A. Widmer, J. A. Shykoff, R. Yockteng, M. E. Hood, and T. Giraud. 2008. Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. BMC Evol. Biol. 8:100.

Reid, N. M., and B. C. Carstens. 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol. Biol. 12: 196.

Ricklefs, R. E., and S. M. Fallon. 2002. Diversification and host switching in avian malaria parasites. Proc. R Soc. B Biol. Sci. 269:885–892.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.

Rubin, B. E. R., R. H. Ree, and C. S. Moreau. 2012. Inferring phylogenies from RAD sequence data. PLoS One 7:e33394.

Russell, A. L., J. Ranivo, E. P. Palkovacs, S. M. Goodman, and A. D. Yoder. 2007. Working at the interface of phylogenetics and population genetics: a biogeographical analysis of *Triaenops* spp. (Chiroptera: Hipposideridae). Mol. Ecol. 16:839–851.

Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33:1654–1668.

Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and İ. Birol. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19:1117–1123.

Skutch, A. F. 1956. Life history of the ruddy ground dove. Condor 58:188–205.

Slater, G. S. C., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31.

Sloan, D. B., J. C. Havird, and J. Sharbrough. 2016. The on-again-off-again relationship between mitochondrial genomes and species boundaries. Mol. Ecol. 2212–2236.

Smith, M. A., N. E. Woodley, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). PNAS 103:3657–3662.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Sweet, A. D., B. M. Boyd, and K. P. Johnson. 2016a. Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. Biol. J. Linn. Soc. 118:813–828.

Sweet, A. D., R. T. Chesser, and K. P. Johnson. 2016b. Comparative cophylogenetics of Australian phabine pigeons and doves (Aves: Columbidae) and their feather lice (Insecta: Phthiraptera). Int. J. Parasitol. 47:347–356.

Sweet, A. D., and K. P. Johnson. 2015. Patterns of diversification in small New World ground doves are consistent with major geologic events. Auk 132:300–312.

———. 2016. Cophylogenetic analysis of New World ground-doves (Aves: Columbidae) and their parasitic wing lice (Insecta: Phthiraptera: *Columbicola*). Mol. Phylogenet. Evol. 103:122–132.

Templeton, A. R. 1981. Mechanisms of speciation-a population genetic approach. Annu. Rev. Ecol. Syst. 12:23–48.

Vachaspati, P., and T. Warnow. 2015. ASTRID: accurate species TRees from internode distances. BMC Genomics 16(Suppl 10):S3.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, et al. 2013. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43:11.10.1–33.

Waterhouse, R. M., E. M. Zdobnov, F. Tegenfeldt, J. Li, and E. V Kriventseva. 2011. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 39:D283.

Weckstein, J. D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. Syst. Biol. 53:154–164.

Whipps, C. M., and M. L. Kent. 2006. Phylogeography of the cosmopolitan marine parasite *Kudoa thyrsites* (Myxozoa: Myxosporea). J. Eukaryot. Microbiol. 53:364–373.

Whiteman, N. K., R. T. Kimball, and P. G. Parker. 2007. Co-phylogeography and comparative population genetics of the threatened Galápagos hawk and three ectoparasite species: ecology shapes population histories within parasite communities. Mol. Ecol. 16:4759–4773.

Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. PNAS 107:9264–9269.

Associate Editor: J. Light
Handling Editor: M. Noor

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. ASTRAL phylogeny from gene trees of body lice from small New World ground-doves.
Figure S2. ASTRID cladogram from gene trees of body lice from small New World ground-doves.
Figure S3. Maximum likelihood phylogeny from mitochondrial sequence data of body lice from small New World ground-doves.
Figure S4. Summary of phylogenetic relationships among *Physconelloides* lice from small New World ground-doves.
Figure S5. Biogeographic states and Maddison-Slatkin randomization results of *Physconelloides* lice from small New World ground-doves.
Figure S6. Biogeographic states and Maddison-Slatkin randomization results of *Physconelloides eurysema* 3 lice from small New World ground-doves.
Figure S7. Reconciliation of phylogenetic trees of small New World ground-doves and their parasitic body lice.
Figure S8. Box-and-whisker plot of jackknifed Procrustes squared residuals from individual links between small New World ground-doves and their *Physconelloides* body lice. Lower residual values suggest a greater contribution to phylogenetic congruence.
Figure S9. Box-and-whisker plot of jackknifed Procrustes squared residuals from individual links between small New World ground-doves and their *Physconelloides* body lice. Lower residual values suggest a greater contribution to phylogenetic congruence. Links that had significant ParaFitLink1 statistics are represented by the left box (light blue).
Figure S10. STRUCTURE plot from 889 randomly sampled unlinked SNPs called for *Physconelloides* body lice from small New World ground-doves. Individual lice are grouped according to host species, and colored according to the likelihood of being in a particular cluster. Phylogenies to the left of the STRUCTURE plots are modified from the concatenated *Physconelloides* phylogeny, and are colored according to the clusters from the STRUCTURE plot.
Figure S11. STRUCTURE plot from 880 randomly sampled unlinked SNPs called for *Physconelloides* body lice from small New World ground-doves. Individual lice are grouped according to host species, and colored according to the likelihood of being in a particular cluster. Phylogenies to the left of the STRUCTURE plots are modified from the concatenated Physconelloides phylogeny, and are colored according to the clusters from the STRUCTURE plot.
Figure S12. DAPC density plot using SNPs from *Physconelloides eurysema* 3 ($K = 2$).
Table S1. Specimen information, extraction results, library preparation details, Illumina sequencing statistics, locus assembly, and raw sequence data deposition for *Physconelloides* body lice from small New World ground-doves.