UNTANGLING THE COEVOLUTIONARY HISTORY BETWEEN DOVES AND THEIR PARASITIC LICE

BY

ANDREW DONALD SWEET

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Ecology, Evolution, and Conservation Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Dr. Kevin Johnson, Chair, Director of Research
Professor May Berenbaum
Professor James Whitfield
Associate Professor Brian Allan
Assistant Professor Julian Catchen

ABSTRACT

In host-parasite systems, any given host species can be associated with multiple types of parasites, each of which can have a unique ecological relationship with the host. However, it remains unclear how these ecological differences link to evolutionary patterns. What shapes the dynamics of a host-parasite interaction over evolutionary time? An ideal approach for addressing this question is to compare multiple lineages of similar parasites that are associated with the same group of hosts but have distinct ecological differences – or "ecological replicates." For my dissertation, I applied this strategy by focusing on the wing and body lice of doves. These two "ecomorphs" of lice are not closely related yet exclusively parasitize the same group of hosts. Notably, wing lice have a greater capability for dispersal than body lice. Dispersal is an important ecological component of host-parasite interactions and speciation in general.

The first part of my dissertation examined broad cophylogenetic patterns across the dove-louse system. I found that wing and body lice did not have correlated patterns, and body lice showed more cospeciation with their hosts. This pattern agreed with previous studies, the results of which suggested that the increased cospeciation in body lice was due to differences in dispersal ability. In contrast with previous work, I also found that both wing and body louse phylogenies are statistically congruent with the host phylogeny. However, the previous studies had limited taxon sampling compared to my study, indicating that taxon sampling can have a significant impact on the results of cophylogenetic comparisons, and that there can be variable cophylogenetic patterns

within a host-parasite system. Cophylogenetic variation in dove lice was further highlighted by my study on lice from phabine doves, a clade native to Australia and Southeast Asia. In this system, wing lice have higher levels of cospeciation with their hosts than did body lice, which is the opposite pattern found in other dove louse systems.

The second part of my dissertation focused on the wing and body lice of New World ground-doves. All three groups (wing lice, body lice, and doves) are monophyletic and have relatively few species, which makes the system ideal for obtaining a comprehensive taxonomic sample. As a group that straddles the population-species boundary, ground-dove lice are also useful for gaining insight into host-parasite evolution at phylogenetic and population scales. I used Sanger or whole-genome sequencing data to estimate phylogenetic and/or population patterns of the ground-dove hosts and both groups of lice. For the louse genomes, I developed a novel pipeline to assemble nuclear genes for phylogenetic analysis and call SNPs for population analysis. My results indicate that dispersal is a key factor in shaping the evolution of this host-parasite system. Body lice had higher levels of cospeciation with their hosts, were more host-specific, and had higher rates of divergence than wing lice. At the population level, some body lice also showed host-specific structure, whereas wing lice did not. Body lice also had lower levels of heterozygosity than wing lice, suggesting higher levels of inbreeding. However, dispersal is likely not the only factor that shapes this host-parasite system. Host phylogeny appears to have a significant effect as well. Both wing and body louse phylogenies were statistically congruent with the host phylogeny, and the congruence metrics for individual associations were correlated between the two types of lice. Biogeography may also dictate host-parasite interactions. The wing louse phylogeny was

significantly structured according to biogeographic region, and both wing and body lice also showed some biogeographic structure at the population level. Together, these results show that host-parasite interactions can be dictated by many ecological factors over evolutionary time, even in the presence of a primary, dominant factor (e.g., parasite dispersal).

ACKNOWLEDGEMENTS

I have been a tree amid the wood

And many a new thing understood

That was rank folly to my head before.

-Ezra Pound, "The Tree"

Although my name is on the title page, this dissertation is the culmination of work that has been supported by myriads. First, I am indebted to my advisor, Kevin Johnson. He took me on as a green student with relatively little research experience, and helped me to become a better scientist and scholar. I am grateful for his frequent guidance on my research, honest and constructive criticism of my writing, and help with organizing fieldwork. One day, I hope he is able to bask in the glory of a Twins victory over the Yankees in a playoff series.

I also want to thank my dissertation committee present and past: Brian Allan, May Berenbaum, Julian Catchen, Jim Whitfield, and Zac Cheviron. Their guidance was invaluable as my dissertation took shape, and several of them wrote me recommendation letters for various award applications. A special thanks to Brian Allan for serving as the chair for my preliminary exam.

Collaborators from all over the world, some of whom I have never met in person, have been foundational for my research. Thank you to Vitor de Q. Piacentini and Michel

Valim for collecting many louse specimens for me in Brazil, especially after my plans to collect in person fell through. Thank you to the many people who have over the years collected lice I used in my dissertation: Sarah Bush, Dale Clayton, Jeff DaCosta, Robert Faucett, John Klicka, Peggy Guttan-Mayerma, Ian Mason, Kevin McCracken, Matthew Miller, Robert Palmer, Garth Spellman, and Robert Wilson. Also thank you to Terry Chesser for collaborating on the phabine project. For my fieldwork in Ecuador, thank you to Jose Luis Rivera-Parra for helping to organize the trip, arrange field sites, etc. I was able to collect important samples, but the trip would not have been possible without Jose's effort. I am especially grateful for two individuals who helped me in the field: Edith Montalvo and Edison Auqui. They were instrumental in collecting the samples I needed, while simultaneously putting up with my terrible Spanish and marginal mist-netting "skills." Also, thank you to Henry Pollock for showing me the ropes on mist-netting, and for allowing me to bring some of his nets with me to Ecuador.

I am grateful for the many sources of financial support I have received while at Illinois. Thank you to the Program in Ecology, Evolution, and Conservation for several summers of RA support and for travel grants to attend conferences and workshop. I am also thankful for the TA positions that were available through my program. I received some summer support through the Clark Summer Fellowship. The Graduate College Dissertation Travel Grant gave me generous funding to conduct fieldwork. I also received support for fieldwork through an American Ornithological Society student research award, the Herbert Holdsworth Ross Award, and the Francis M. and Harlie M. Clark Research Support Grant. The Graduate College Dissertation Completion Fellowship allowed me to focus on finishing my dissertation work during my final year.

I am also thankful for the support of my friends throughout my graduate studies. Julie Allen and Bret Boyd have been two of my greatest mentors. They are amazing scientists and wonderful teachers, and I am glad to have overlapped with them in the Johnson Lab. Therese Catanach has been a good friend, and we have had some good times chasing doves in Texas, watching World Series games at Crane Alley, and of course playing catch whenever possible. Scott Shreve was a great mentor during my first year, and as a senior graduate student was very helpful during my transition to graduate school life. Thank you also to the various graduate and undergraduate students in the Johnson Lab for helpful discussions and assistance in obtaining data: Patrick Grady, Robert De Moya, Stephany Herrera, Gabriela Escalante, Brandon Jones, Ashley Markazi, Jennifer Nowak, and Matthew Mason. I took part in many good paper discussions with the Systematics Discussion Group, one of which spurred a fun collaborative side project on human hepatitis B virus. The two seasons with the Softballogists provided some healthy competition and good times with fellow biologists and friends.

I feel compelled to briefly acknowledge some crucial sources of inspiration and mental well-being. As I began graduate school, I quickly discovered the importance of satisfying the right hemisphere of my brain. And so, thank you to Frost and Faulkner, Strand and Hemingway, Harding and Robinson, Beethoven and Pärt, Saint-Saëns and Barber, Brahms and Mozart, Pollock and Picasso, Vermeer and Ruisdael, Raphael and Van Eyck. I am also grateful the Krannert Art Museum was across the street from my office.

Finally, I would like to thank my family. I doubt these last six years would have been sustainable without their love and support. Thank you to my parents, who have

always encouraged my zeal for science. I am grateful for their enthusiasm to learn about my dissertation research, and for their excitement to read my first published papers. They have been, and will continue to be, big role models in my life. Thanks to my siblings – Jeremiah, Hannah, and Jerushah. I am glad they made several visits to central Illinois. Thank you also to all four of my grandparents, who have always supported and expressed great interest in their grandchildren's endeavors. To conclude, thank you to my wife Jamie. I am grateful for her love, for joining me in Illinois, for listening to hours of practice talks, and for reading many of my drafts. She has provided so much encouragement, while also being willing to call me out when I begin to take myself too seriously.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: COPHYLOGENETIC PATTERNS ARE UNCORRELATED BETWEEN TWO LINEAGES OF PARASITES ON THE SAME HOSTS 11

CHAPTER 3: COMPARATIVE COPHYLOGENETICS OF AUSTRALIAN PHABINE PIGEONS AND DOVES (AVES: COLUMBIDAE) AND THEIR FEATHER LICE (INSECTA: PHTHIRAPTERA) 51

CHAPTER 4: PATTERNS OF DIVERSIFICATION IN SMALL NEW WORLD GROUND DOVES ARE CONSISTENT WITH MAJOR GEOLOGIC EVENTS 85

CHAPTER 5: COPHYLOGENETIC ANALYSIS OF NEW WORLD GROUND-DOVES (AVES: COLUMBIDAE) AND THEIR PARASITIC WING LICE (INSECTA: PHTHIRAPTERA: COLUMBICOLA) 122

CHAPTER 6: INTEGRATING PHYLOGENOMIC AND POPULATION GENOMIC PATTERNS IN AVIAN LICE PROVIDES A MORE COMPLETE PICTURE OF PARASITE EVOLUTION 163

CHAPTER 7: THE ROLE OF PARASITE DISPERSAL IN SHAPING A HOST-PARASITE SYSTEM AT MULTIPLE EVOLUTIONARY TIMESCALES 215

APPENDIX A: SUPPLEMENTARY FIGURES 259

APPENDIX B: SUPPLEMENTARY TABLES 286

CHAPTER 1: INTRODUCTION

Comparative phylogenetic and population genetic analyses are powerful approaches for learning about the processes that govern living systems (Harvey and Pagel, 1991). This is particularly true for host-parasite systems, or host-symbiont systems more generally, which involve some level of dependent interactions between two or more groups of organisms. For example, comparing the phylogeny of a group of parasites to that of their hosts can indicate co-divergence (i.e., cospeciation) or host switching (Hafner and Nadler, 1988; Page, 1994; de Vienne et al., 2013). Likewise, comparing population structures of parasites and hosts can indicate what ecological factors (e.g., parasite dispersal, climate, host habitat use) are shaping the interaction at a microevolutionary scale (McCoy et al., 2005; Whiteman et al., 2007; Criscione, 2008). Integrating phylogenetic and population genetic approaches in a single system is particularly useful, because these approaches provide snapshots of different evolutionary timescales (Cutter, 2013). However, to connect phylogenetic and population patterns, and therefore establish a link between evolutionary pattern and ecological mechanisms, an effective approach would focus on "ecological replicates" – multiple groups of similar parasites/symbionts that are associated with a single group of hosts but have some ecological difference of interest (Clayton and Johnson, 2003; Johnson and Clayton, 2003). For example, Weiblen and Bush (2002) compared *Ficus* figs to their associated mutualist and parasitic fig wasps. They found that mutualist wasps showed strong patterns of co-divergence with the figs, whereas the parasitic wasps had a history of frequent host switches between different fig species. These results clearly link the ecology of fig wasps (life history) to evolutionary patterns in the fig-fig wasp system.

Doves and their ectoparasitic lice are a well-studied ecological replicate system (Johnson and Clayton, 2003; Clayton et al., 2016). There are two types or "ecomorphs" of lice associated with doves – wing and body lice. The two ecomorphs primarily live on different niches of the host (wing and body feathers, respectively) and are not closely related (Cruiskshank et al., 2001; Johnson et al., 2007; Johnson et al., 2011). They also use different strategies to avoid host preening behavior (Clayton, 1991). Wing lice are elongate insects that insert themselves between feather barbs to avoid being removed by their host (Stenram, 1956). Translocation studies among different-sized hosts have shown that if a wing louse is too small or large to fit between the barbs, the louse is likely to be removed by preening or host movement (Bush and Clayton, 2006). In contrast, body lice have a rounded shape and avoid preening by burrowing into the host's downy feathers (a "run and hide" strategy) (Clayton et al., 1999). The drastic morphological differences in wing and body lice are likely driven by a combination of selective pressure from preening and inter-louse competition (Clay, 1949; Bush and Malenke, 2008). Similar ecomorphs have arisen multiple times independently across avian lice (Johnson et al., 2012). Wing and body lice also have different capabilities of dispersing among host individuals. Both are primarily transmitted vertically between host parents and offspring, and horizontally by direct host contact (e.g., mating, shared roosts) (Rothschild and Clay, 1952; Clayton and Tompkins, 1994). However, wing lice are able to effectively use winged hippoboscid flies, generalist blood-feeding parasites, to move among hosts – a behavior known more generally as "phoresis" (Keirans, 1975; Harbison et al., 2008; Harbison et al., 2009; Bartlow et al., 2017). The lice use their mandibles and legs to grab hold of the fly, which allows them to remain attached as the fly moves to another host. Body lice are rarely

phoretic, likely because their short legs inhibit their ability to grasp to a fly (Harbison et al., 2011; but see Couch, 1962).

There are also notable similarities between dove wing and body lice. Like other avian lice, both ecomorphs are permanent and obligate parasites; they spend their entire lifecycle on the host and cannot survive for more than 2-3 days away from the host (Marshall, 1981; Tompkins and Clayton, 1999). Both are relatively host specific; neither parasitizes birds outside of the dove family (Columbidae), and many species are specific to genus or species of host (Price et al., 2003). Despite living primarily on separate parts of the host, both ecomorphs eat downy feathers on their host's body (Nelson and Murray, 1971). Wing lice will migrate to the body to feed, likely when the host is inactive or sleeping, and then return to the wing feathers when the host is active (Harbison and Boughton, 2014).

Because of this abundant ecological knowledge, dove lice are an excellent group for using a comparative approach to understand how ecological mechanisms shape evolutionary history in host-parasite systems. Clayton and Johnson (2003) used dove lice in this framework by comparing wing and body louse phylogenies to the phylogeny of some New World doves. They found that body lice had a highly congruent evolutionary relationship with their hosts, whereas wing lice did not have as much congruence with the hosts and showed evidence for rampant host switching. They hypothesized that the difference in dispersal ability accounted for the observed cophylogenetic patterns. Because wing lice can effectively use phoresis, this would promote host switching and uncouple the louse and host phylogenies over evolutionary time.

My dissertation also concerns the evolutionary relationships between doves and their lice. Like Clayton and Johnson (2003), my primary objective is to link ecological (microevolutionary) mechanisms to macroevolutionary patterns, but I do so by integrating over multiple scales. In particular, I focus on three levels: family-wide phylogenetic patterns, phylogenetic patterns in subsets of taxa, and population-level patterns.

First, I expand upon Clayton and Johnson's (2003) study by comparing a more globally-sampled dove phylogeny to the phylogenies of their associated wing and body lice (Chapter 2). The results from this study emphasize the importance of taxonomic sampling in a cophylogenetic study and suggest that focusing on smaller, monophyletic subsets of taxa (i.e., species that share a common ancestor) is the optimal approach. A focused approach allows for a more complete phylogenetic representation (Jackson et al., 2008). In Chapter 3, I use the Australian phabine doves and their lice to show there can be incredible cophylogenetic variability within a broader (e.g., across a host family) host-parasite system. This further highlights the utility of more taxonomically focused comparative studies. The remainder of my dissertation focuses on another subset of taxa: New World ground-doves and their lice. All three groups (the doves, wing lice, and body lice) are in monophyletic groups, and there are few enough taxa that near-complete sampling is possible. There are 17 known species of ground-doves (Gibbs et al., 2003), three species of wing lice (*Columbicola*), and three species of body lice (*Physconelloides*) (Price et al., 2003). Because there are few parasite species, and possible cryptic species in each ecotype, this is a good example of a system that straddles the population-species boundary and is therefore suitable for assessing both phylogenetic

and population-level patterns. In Chapter 4, I estimate the phylogenetic relationships of ground-doves, their divergence times, and their biogeographic history. Clarifying the evolutionary history of the hosts is a crucial first step in any comparative host-parasite cophylogenetic study. I then use Sanger sequencing data to estimate phylogenetic patterns in the wing lice and compare this phylogenetic hypothesis to the host phylogeny (Chapter 5). In this chapter, I also estimate population-level patterns among the most widespread (i.e., associated with the most host species) wing louse species. In Chapter 6, I use whole genome sequencing of body lice to estimate their phylogenetic and population genetic patterns. I also develop and describe a novel bioinformatic pipeline for assembling appropriate data to accomplish these goals. Chapter 7 is the culmination of my ground-dove wing and body louse comparison. Using the pipeline from Chapter 6, I obtain genomic-level data for both ecomorphs of lice and compare their phylogenomic and population genomic patterns.

Each chapter in this dissertation has standalone merit as a scientific effort, but the aggregate contribution of these studies provides considerable insights into host-parasite coevolutionary relationships. They tell a story about a single, albeit complex, host-parasite system, but the findings have implications for host-parasite systems in general. In particular, I argue that parasite dispersal and host association are two ecological factors that can drive evolutionary patterns in a host-parasite system. As with many intensive studies, this dissertation also generates other questions which I could not address here. I hope my contribution encourages future work in this and other systems, and helps to further our understanding of how hosts and their parasites evolve together through time.

LITERATURE CITED

Bartlow A.W., Villa S.M., Thompson M.W., Bush S.E. 2016. Walk or ride? Phoretic behaviour of amblyceran and ischnoceran lice. *International Journal for parasitology* 46: 221–227.

Bush S.E., Clayton D.H. 2006. The role of body size in host specificity: reciprocal transfer experiments with feather lice. *Evolution* 60: 2158–2167.

Bush S.E., Malenke J.R. 2008. Host defence mediates interspecific competition in ectoparasites. *The Journal of animal ecology* 77: 558–64.

Clay T. 1949. Some problems in the evolution of a group of ectoparasites. *Evolution* 3: 279–299.

Clayton D.H., Lee P.L.M., Tompkins D.M., Brodie E.D. 1999. Reciprocal natural selection on host-parasite phenotypes. *The American Naturalist* 154: 261–270.

Clayton D.H., Johnson K.P. 2003. Linking coevolutionary history to ecological process: doves and lice. *Evolution* 57: 2335–41.

Clayton D.H., Tompkins D.M. 1994. Ectoparasite virulence is linked to mode of transmission. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 256: 211–217.

Clayton D.H. 1991. Coevolution of avian grooming and ectoparasite avoidance. Pp. 258–289 *in* J. E. Loye and M. Zuk, eds. Bird-parasite interactions: ecology, evolution, and behavior. Oxford University Press.

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

Couch A.B. 1962. Phoretic mallophagans from hippoboscids of mourning doves Zenaidura macroura. *The Journal of Parasitology* 48: 497.

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

Cruickshank R.H., Johnson K.P., Smith V.S., Adams R.J., Clayton D.H., Page R.D.M. 2001. Phylogenetic analysis of partial sequences of elongation factor 1 alpha identifies major groups of lice (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 19: 202-215.

Cutter A.D. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution* 69: 1172–1185.

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

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

Hafner M.S., Nadler S.A. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 332: 258-259.

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

Harbison C.W., Clayton D.H. 2011. Community interactions govern host-switching with implications for host-parasite coevolutionary history. *Proceedings of the National Academy of Sciences* 108: 9525–9.

Harbison C.W., Boughton R.M. 2014. Thermo-orientation and the movement of feather-feeding lice on hosts. *Journal of Parasitology* 100: 433–441.

Harbison C.W., Jacobsen M.V., Clayton D.H. 2009. A hitchhiker's guide to parasite transmission: The phoretic behaviour of feather lice. *International Journal for Parasitology* 39: 569–575.

Harvey P.H., Pagel M.D. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Jackson A.P., Machado C.A, Robbins N., Herre E.A. 2008. Multi-locus phylogenetic analysis of neotropical figs does not support co-speciation with the pollinators: The importance of systematic scale in fig/wasp cophylogenetic studies. *Symbiosis* 45:16.

Johnson K.P., Clayton D.H. 2003. 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. University of Chicago Press, Chicago.

Johnson K.P., Reed D.L., Hammond Parker S.L., Kim D., Clayton D.H. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for Columbicola (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 45: 506–18.

Johnson K.P., Shreve S.M., Smith V.S. 2012. Repeated adaptive divergence of microhabitat specialization in avian feather lice. *BMC Biology* 10: 52.

Johnson K.P., Weckstein J.D., Meyer M.J., Clayton D.H. 2011. There and back again: switching between host orders by avian body lice (Ischnocera: Goniodidae). *Biological Journal of the Linnean Society* 102: 614–625.

Keirans J.E. 1975. A review of the phoretic relationship between Mallophaga (Phthiraptera: Insecta) and Hippoboscidae (Diptera: Insecta). *Journal of Medical Entomology* 12: 71–76.

Marshall A.G. 1981. The ecology of ectoparasitic insects. Academic Press.

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

Nelson B.C., Murray M.D. 1971. The distribution of Mallophaga on the domestic pigeon (Columba livia). *International Journal for Parasitology* 1: 21-19.

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

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

Rothschild M., Clay T. 1952. Fleas, flukes, and cuckoos: a study of bird parasites. 3rd ed. Collins, London.

Stenram H. 1956. The ecology of *Columbicola columbae* L. *Opuscula Entomologica* 21: 170–190.

Tompkins D.M., Clayton D.H. 1999. Host resources govern the specificity of swiftlet lice: Size matters. *Journal of Animal Ecology* 68: 489–500.

Weiblen G.D., Bush G.L. 2002. Speciation in fig pollinators and parasites. *Molecular Ecology* 11:1573–1578.

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

CHAPTER 2: COPHYLOGENETIC PATTERNS ARE UNCORRELATED BETWEEN TWO LINEAGES OF PARASITES ON THE SAME HOSTS[1]

INTRODUCTION

Parasitic organisms are among the most abundant and diverse group of organisms on earth (Windsor, 1998; Poulin and Morand, 2000; Poulin and Morand, 2004; Dobson et al., 2008; Mora et al., 2011). One of the mechanisms that contributes to this diversity is cospeciation, the parallel speciation of two organisms with dependent life histories (Hafner and Nadler, 1990; Hafner et al., 1994; Hafner and Page, 1995; Page, 2003; de Vienne et al., 2013). Parasites that cospeciate with their hosts should exhibit congruent diversification patterns (Fahrenholz, 1913; Eichler, 1948). While this congruence has been found in some instances (Hafner and Nadler, 1988; Page et al., 2004; Hughes et al., 2007), many host-parasite systems show discordant patterns. This indicates evolutionary processes that promote diversification in parasites independently of their hosts (Paterson et al., 2000; Johnson et al., 2002; Brudydonckx et al., 2009). For example, host switching

---

[1] This is a pre-copyedited, author-produced version of an article accepted for publication in *Biological Journal of the Linnean Society* following peer review. The version of record: Sweet, A.D., B.M. Boyd, and K.P. Johnson. 2016. Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. *Biological Journal of the Linnean Society*. 118(4): 813-828 is available online at: https://doi.org/10.1111/bij.12771.

and parasite duplication (speciation within a host) may result in incongruent diversification patterns between hosts and their parasites (Page, 2003). Additionally, geography (Weckstein, 2004; Johnson et al., 2007), host preference (Johnson et al., 2005; Gorrell and Schulte-Hostedde, 2008), host-imposed selective pressures (Clayton et al., 1999; Clayton and Walther, 2001; Waite et al., 2012), competition between parasites (Poulin, 2007; Bush and Malenke, 2008; Johnson et al., 2009), and opportunities for host switching may influence the parasite diversification. Here we generally refer to diversification patterns between hosts and their parasites, either congruent or incongruent, as "cophylogenetic patterns."

Free-living organisms often host many lineages of closely related parasites (Poulin, 1997). Comparisons of phylogenies of multiple parasite lineages with those of their hosts can address fundamental questions in host-parasite coevolution. For example, it is important to understand how different parasite lineages respond to host speciation events. Additionally, host ecology may shape cophylogenetic patterns in different ways for different parasite lineages (Page, 1994; Johnson and Clayton, 2003). The ectoparasitic lice (Insecta: Phthiraptera) parasitizing pigeons and doves (Aves: Columbidae) are ideal subjects for addressing such questions. Pigeons and doves are parasitized by two groups of feather lice: wing and body lice (Johnson and Clayton, 2003; Johnson et al., 2012). While both feed on abdominal downy feathers, members of these two groups have different mechanisms for escaping host preening (Rothschild and Clay, 1952; Nelson and Murray, 1971; Clayton et al., 2005; Clayton et al., 2010). Wing lice are elongate and insert themselves between the barbs of the wing feathers to escape preening, whereas the rounder body lice burrow into feather down to escape preening (Clayton, 1991; Clayton

et al., 1999; Figure 1.1). Although both of these groups of lice are in the same family (Philopteridae), wing and body lice parasitized doves independently, being relatively distantly related to each other (Cruickshank et al., 2001). These two lineages can be treated as "ecological replicates" that have different environmental limitations (Johnson and Clayton, 2003). Additionally, pigeons and doves are distributed worldwide and occupy a variety of ecological niches. Some groups, such as ground-doves, exhibit terrestrial lifestyles and primarily feed on seeds. Other groups, such as the fruit doves, are primarily arboreal and feed on fruits (Goodwin, 1983; Gibbs et al., 2001). Since both groups of dove lice are found on most host species (Price et al., 2003), it is possible to obtain a geographically extensive sample across the range of host niches for both groups of lice.

Despite both wing and body lice being distributed worldwide on many species of doves, wing lice appear to be more likely than body lice to switch between host species because of ecological differences in dispersal capability. While both are obligate parasites, wing lice are more mobile than the more host specific body lice (Johnson et al., 2002; Price et al., 2003). Wing lice have been shown to "hitchhike" on hippoboscid flies, generalist ectoparasites that often target doves (Harbison et al., 2008; Harbison and Clayton, 2011). This hitchhiking behavior, known as phoresy, may allow wing lice to rapidly move between hosts that may not normally interact. Body lice do not appear to utilize phoresy, so they are unlikely to disperse between host individuals in this way (Harbison et al., 2009). However, body lice do show some evidence of host switching, which appears to be facilitated by host behaviors. For example, gregarious roosting and foraging bring different species of doves into contact and may facilitate exchange of both

wing and body lice (Harbison et al., 2008; Johnson et al., 2011a). Given this knowledge, we expect that wing lice will be more likely to show phylogenetic patterns incongruent with their hosts. Previous taxonomically or geographically limited cophylogenetic studies have shown this to be the case (Clayton and Johnson, 2003; Johnson and Clayton, 2004). A study with broader sampling is needed to more thoroughly evaluate these patterns.

Here we combined new and existing data from multiple studies to compare cophylogenetic patterns of wing and body lice on a worldwide scale. From this data set we estimated phylogenetic trees for the doves and their associated wing and body lice. We used the resulting trees in cophylogenetic analyses, under both topology-based and event-based approaches.

MATERIALS AND METHODS

Taxon and marker selection

We obtained sequence data from NCBI-GenBank deposited in previous studies. This includes pigeon and dove data from Johnson and Clayton (2000), Johnson (2004), Johnson and Weckstein (2011), Pereira et al. (2007), Sweet and Johnson (2015), and Johnson et al. (2001); wing louse data from Johnson et al. (2007) and Johnson and Clayton (2004); and body louse data from Johnson et al. (2011a, b), Johnson and Clayton (2004), and Johnson, Adams, and Clayton (2001) (Supplementary Table 2.1). In instances where no GenBank data were available, we sequenced samples according to methods outlined in Johnson and Clayton (2000), Johnson et al. (2007), and Johnson et al. (2011b). Wing lice in this study belong to the genus *Columbicola,* while body lice are spread across the genera *Auricotes, Coloceras, Campanulotes,* and *Physconelloides.* We

used *Aerodramus salangana* (swiflet) as the outgroup for the doves following the rooting of Johnson and Clayton (2000), *Oxylipeurus chiniri* (chachalaca louse) for wing lice following the rooting of Johnson et al. (2007), and *Stronglyocotes orbicularis* (tinamou louse), *Goniocotes talegallae* (brushturkey louse), and *Goniodes assimilis* (partridge louse) for body lice following the rooting of Johnson et al. (2011b).

For the doves, we used the mitochondrial loci cytochrome oxidase subunit I (COI), ATP synthase F0 subunit 8 (ATP8), NADH dehydrogenase subunit 2 (ND2), and cytochrome b (Cytb), and nuclear locus beta-fibrinogen intron 7 (FIB7). For wing lice, we used mitochondrial loci COI and 12S ribosomal RNA (12S), and nuclear locus elongation factor 1-alpha (EF-1α). For body lice, we used mitochondrial loci COI and 16S ribosomal RNA (16S), and nuclear locus EF-1α. These markers were chosen because the majority of our targeted taxa have this sequence information, therefore minimizing missing sequences in our final data matrix. We also excluded lice for which we did not have host DNA sequence data and vice-versa. Thus, each host taxon had data for at least one associated wing and body louse.

Phylogenetic analysis

We aligned sequences for each locus and in each taxonomic group (doves, wing lice, and body lice) independently. All alignments were done using MUSCLE (Edgar, 2004) and visualized alignments using Seaview v4 (Gouy et al., 2010). After inspecting the alignments we concatenated the locus based alignments into a single alignment for each group (doves, wing lice, and body lice) in Seaview. Using the concatenated data sets for each group, we estimated maximum likelihood (ML) trees in RAxML v7.0.4 (Stamatakis, 2006) using GTR + I + Γ model of sequence evolution and 500 bootstrap

replicates. We also estimated ultrametric Bayesian trees using BEAST v1.7.5 (Drummond et al., 2012). For the BEAST analyses, we partitioned each concatenated alignment by locus and used jModelTest2 (Darriba et al., 2012) to estimate the best-fitting substitution models for each locus according to the corrected Akaike Information Criterion (AICc; Sugiura, 1978). We treated all mitochondrial loci as a single locus in all three alignments. For wing lice and body lice, we applied a GTR + I + Γ model to the mitochondrial data, and a K80 + I + Γ model to EF-1α. For the doves, we applied separate GTR + I + Γ models to the mitochondrial data and FIB7. In BEAST we used a lognormal relaxed clock and a Yule speciation tree prior for all three partitioned data sets, and ran analyses for 20 million MCMC generations with sampling every 1,000. We checked resulting .log files in Tracer v1.4 (Rambaut and Drummond, 2007) and from the trace plots found each analysis reached stationarity and had Effective Sample Size (ESS) values >>200. Based on the trace files we discarded the first 2,000 trees (10%) as burnin.

Cophylogenetic analysis

Preparing trees for analysis. For phylogenetic analysis, we included multiple louse samples of the same species but that are associated with different host species. However, because in some cases there was no evidence that these multi-host parasites were genetically distinct (Supplementary Figures 2.1A-C), we collapsed these down to a single terminal taxon for cophylogenetic analysis using Mesquite v2.75 (Maddison and Maddison, 2011). We did this to avoid bias due to taxon duplication in our data set. We also removed outgroup taxa since their inclusion was for rooting the phylogenetic trees and not for cophylogenetic analysis. We used these trimmed trees for all subsequent

analyses. In particular, we analyzed our data with both topology-based and event-based methods (de Vienne et al., 2013).

Topology-based approach to test for cophylogenetic signal. For a topology-based comparison, we used ParaFit (Legendre et al., 2002) in the "ape" package of R (Paradis et al., 2004). ParaFit takes the host phylogeny, parasite phylogeny, and association matrix as input and tests for random association between the two groups of taxa by randomizing the association matrix. ParaFit also tests for the contribution of each host-parasite association to the global statistic through two individual link statistics: ParaFitLink1 ("F1") and ParaFitLink2 ("F2"). F1 is a more conservative test and is generally preferred, however F2 has greater power in some cases (Legendrew et al., 2002). We ran ParaFit comparing the wing louse tree to the host tree and comparing the body louse tree to the host tree, and also for both the ML trees from RAxML and the ultrametric trees from BEAST. We first converted our trees to patristic distance matrices using "ape," and ran ParaFit for 100,000 random permutations using the "lingoes" correction for negative eigenvalues. We also used the alternative correction – "calliez" – but the results were nearly identical. Therefore, we used "lingoes" results in all subsequent analyses. In all ParaFit analyses, we computed the F1 and F2 statistics for individual links.

To test whether cophylogenetic patterns may be correlated between the wing lice and body lice, we used contingency tables to partition the results of the individual link (i.e. host-parasite association) tests for each ParaFit analysis. The contingency tables were 2 x 2 matrices, with wing lice results on the rows and body lice results on the columns. Each cell indicated whether a particular host had a significant linkage with its

parasite species (indicating this association contributes to topological similarity between the trees). In instances when the links for both the wing and body lice of a particular host were significant, we counted those associations as a single decision in the appropriate cell. If, on the other hand, the body louse had a significant linkage but the wing louse did not, we counted the associations as a single decision in a different cell. For instances where a host had multiple links for one louse type but did not have multiple links for the other louse type (e.g. one host species has multiple wing louse species but only one body louse species associated with it), we counted the single species host-parasite link to match the number of links in the corresponding louse type link. If a host species had one wing louse species but multiple body louse species associated with it (or vice versa), we counted the wing louse link twice to correspond to each of the body louse links.

ParaFit produces p-values for each individual link test to provide a level of support for the contribution of that host-parasite association to the global statistic testing for random association between a group of hosts and their parasites. To correct for false discovery with multiple tests, we used the Benjamini-Hochberg control of false discovery rate (Benjamini and Hochberg, 1995). We did corrections in R assuming $\alpha = 0.05$. Using the corrected p-values, we tallied the individual test links in the cells of our contingency tables and used a Pearson's chi-square test for independence for each contingency table to test for potentially correlated cophylogenetic patterns between the wing and body lice. A significant chi-square result would indicate that cophylogenetic patterns in wing and body lice are correlated. That is, we tested the null hypothesis of whether the significant linkages of wing lice were independent of those for body lice over the same group of hosts. Since ParaFit produces two individual link test statistics, we tallied the results and

used a chi-square test for both statistics. We also used a Fisher's exact test for each contingency table to test whether small sample sizes may affect the chi-square results. We performed the chi-square tests and Fisher's exact tests in R.

Event-based approach to test for cophylogenetic signal. For an event-based approach, we used Jane v4.01 (Conow et al., 2010). Jane uses *a priori* event costs to reconcile host and parasite phylogenies by minimizing the overall cost. We used this method for both the wing and body louse data sets, using the ultrametric trees we generated from BEAST. We ran Jane with the Genetic Algorithm parameters set at 100 generations and with a population size of 100, and set the costs as default: 0 for cospeciation, 1 for duplication, 2 for duplication and host switch, 1 for loss, and 1 for failure to diverge. To test whether the resulting reconstruction cost is significantly lower than by chance, we randomized the tip associations 999 times. A significant result from this statistical test would indicate some level of phylogenetic congruence between host and parasite. Finally, we tested for the correlation of recovered cospeciation events from their placement on the host tree using a contingency table (following the procedure outlined by Johnson and Clayton, 2003).

Testing for taxonomic bias. Since our sample has a high proportion (10/15 representatives) of small New World ground-doves (*Columbina*, *Claravis*, *Uropelia*, and *Metriopelia*) relative to other clades, our cophylogenetic analyses could potentially be affected by a taxonomic/clade representative bias. To test this idea, we removed the small New World ground-dove links in "ape." Using this reduced data set, we ran ParaFit for 100,000 iterations for both the phylogram and ultrametric trees, and applied both the F1 and F2 individual link tests. From the results of the individual link tests we tested for

correlated cophylogenetic patterns between wing and body lice using contingency tables and Pearson's chi-square tests as described above.

RESULTS

Final data matrix

Based on our criterion of only including host samples with both associated wing and body louse data, we had a finalized matrix of 52 different dove species, along with 43 associated wing and 49 body louse taxa (Supplementary Table 2.1). NCBI data yielded complete or near-complete sampling of loci in the host, wing louse, and body louse data sets. For the loci Cytb, COI, ND2, and FIB7 in the birds, there were seven instances of missing data for a gene (4% of entire matrix). However, for the ATP8 locus there were eighteen instances of missing data (37%). There were four instances of missing data for the three loci in the wing louse data (~2%), and nine instances of missing data for the three loci in the body louse data (5%). Maximum likelihood and Bayesian phylogenetic analyses with RAxML and BEAST produced trees largely in agreement with previous studies using this data. However, several of the basal nodes for all dove and louse trees were not well supported.

Cophylogenetic analysis

The global ParaFit statistics were significant for both the wing and body lice data sets ($p < 0.001$, Table 2.1). This was true for patristic distances from both the phylogram and ultrametric trees. Although each dataset indicated strong support for a global non-random association between host and parasite trees, a subset of individual host-parasite links (i.e. host-parasite associations) contribute to this signal. Since ParaFit can also test

each link by recalculating the global ParaFit statistic with the link removed, we can get a better understanding of how certain links contribute to the global statistic. A significant individual link statistic means the global ParaFit statistic decreased in value when that particular linked was removed, and therefore indicates the link represents an important component of the overall host-parasite relationship (Legendre et al., 2002). ParaFit also produces two different individual link statistics (F1 and F2). Here we report results from both tests. The F1 phylogram results included 40 significant wing louse-host links and 33 significant body louse-host links after correcting for multiple comparisons, while the F2 phylogram results indicated 43 significant wing louse-host links and 55 significant body louse-host links after correction (Table 2.1). The F1 ultrametric results did not have any significant body or wing louse links after correction, while the F2 ultrametric results indicated 19 significant wing louse-host links after correction and no significant body louse-host links after correction (Table 2.1). Several links were significant before correction ($\alpha = 0.05$), but were not significant after correction. The specific host-parasite links and associated p-values of both individual link statistics for the phylogram and ultrametric trees are listed in Table 2.2.

Most of the chi-square tests of independence of significant linkages between wing and body lice performed on the contingency tables were not significant or were not applicable (Table 2.3). The only significant test was from the ParaFit phylogram F1 results ($p = 0.002$). The p-values from the other chi-square tests were all > 0.3. Fisher's exact tests yielded similar p-values.

Our Jane analyses recovered 14 nodes of cospeciation among the wing lice and their hosts, and 22 nodes of cospeciation among the body lice and their hosts (Table 2.4).

The specific nodes recovered as cospeciation events in both data sets are indicated in Figures 2 and 3. The placement of these events on the host tree is not correlated between wing and body lice (Table 2.3), suggesting these two parasite lineages diversify independently in response to host diversification. The total reconstruction cost was 84 for the wing lice and 79 for the body lice. In both analyses, none of the costs from 999 random tip associations were equal to or lower than these original reconstruction costs ($p = 0.0$).

Reduced cophylogenetic analysis

In our ParaFit analyses with the small New World ground-dove tips and links removed, our global statistics were significant in all cases ($p < 0.0001$). However, the corrected individual link statistics differed from the full data set results (Table 2.1). For the phylogram trees, wing lice had 12 significant links for both the F1 and F2 statistics, while the body lice had 27 and 33, respectively. The ultrametric trees also had 12 significant wing louse links for both the F1 and F2 statistics, while body lice had 30 and 31 respectively. Pearson's chi-square tests on the contingency tables were not significant ($p > 0.45$ in all cases, Table 2.3). The specific links and associated p-values from cophylogenetic analyses on the reduced data set are listed in Table 2.5.

DISCUSSION

The primary objective of this study was to determine if either or both wing and body lice have phylogenetic histories congruent with their dove hosts or with each other. If both types of lice are affected similarly by host speciation events, we might expect their cophylogenetic patterns to be similar. However, we failed to find significant

evidence that wing and body lice have similar phylogenetic histories. Despite a lack of correlated patterns between wing and body lice of specific host-parasite links, both the wing and body louse data sets individually showed evidence of cospeciation with their hosts.

The chi-square tests based on the contingency tables failed to reject the null hypothesis of independence of cophylogenetic patterns in wing and body lice in all but one case. These results indicate that dove wing and body lice have unique and independent evolutionary histories. This is consistent with previous smaller scale studies of both louse groups and can potentially be explained by differences in life history between wing and body lice (Clayton and Johnson, 2003; Johnson and Clayton, 2003; Johnson et al., 2003; Johnson and Clayton, 2004).

The ParaFit global statistic testing for random host-parasite association was significant for both wing and body louse phylogenies individually. Additionally, the Jane event reconstruction costs were significantly lower than by chance. This indicates that at some level both body and wing lice show congruent phylogenetic patterns with their hosts. Congruence between body lice and their dove hosts was expected. Previous studies based on event-based methods showed strong patterns of cospeciation between body lice and their hosts (Clayton and Johnson, 2003). However, the wing lice sampled in the present study also showed evidence for cospeciation with their hosts. While previous event-based results have recovered some cospeciation events within this group, the overall patterns pointed to a lack of cospeciation over larger time scales (Johnson et al., 2003). However, when taking into account a broad geographic and taxonomic sample, both wing and body lice appear to have undergone some level of cospeciation with their

hosts. Having a more extensive sample, and therefore more branches on phylogenetic trees provides greater statistical power. We suspect this allowed us to detect cophylogenetic signal that was obscured in studies with limited samples, an issue also discussed in Hughes et al. (2007). This could be the case particularly if the smaller samples are biased towards a particular geographic region or host group (Jackson et al., 2008).

Contrary to the global ParaFit statistics, which indicated overall host-parasite congruence in all cases, the individual link statistics of the lice differed among tree type (phylogram vs. ultrametric) and link statistic (F1 vs. F2). Neither wing nor body lice showed consistency in the number of significant links among the different analyses. For example, more wing louse links were significant in the phylogram F1 analysis, while more body louse links were significant in the phylogram F2 analysis. In the ultrametric F1 statistic, none of the links showed significance. Several links in this analysis initially showed significant p-values, but these became non-significant after we corrected for multiple tests (Table 2.2). The instances of more significant wing louse links than body louse links is somewhat surprising. As discussed above, past work has indicated that body lice have stronger phylogenetic congruence with their hosts, and so we might have expected them to have more significant individual links than wing lice.

Poorly resolved backbones of the phylogenies (Supplementary Figures 2.1A-C) could be a possible explanation for the varying individual link statistic results. This could particularly be a primary cause of the discord between the phylogram and ultrametric results. Since ParaFit takes topology and branch lengths (patristic distances) into

consideration, differences between ultrametric and non-ultrametric trees in relative patristic distances could account for these differences.

Alternatively, clade representation biases could be driving cophylogenetic signals. Our data set includes 10/15 representatives of the small New World ground-dove clade and their lice, which is the most thorough sampling representation of a clade in our data set. The hosts, their wing lice, and their body lice have all been shown to be monophyletic (Cruickshank et al., 2001; Pereira et al., 2007; Johnson et al., 2007, 2011b). In both the F1 and F2 ParaFit analysis, every link from this clade contributed to the overall pattern of non-random associations. Since the hosts and their lice are in monophyletic clades, and we were able to include strong taxon sample representation of these groups, the results are perhaps due to congruence between whole clades rather than between specific links within each clade. If the relationships between the clades are contributing significantly to the global statistic, removing a single host-parasite link from a clade would alter the global statistic. Since this is how ParaFit calculates the individual link statistics, each link in the small New World ground-dove clade could potentially be significant.

Our ParaFit analyses with the small New World ground-doves removed indicates some level of taxonomic bias may indeed be a reality in our data set. Although our ParaFit global statistics were once again significant in the reduced data set, results from the individual link tests were more consistent with previous studies. Body lice had at least twice as many significant links as wing lice in all scenarios (Table 2.1). Additionally, results were fairly consistent among tree types (phylogram and ultrametric) and test statistics (F1 and F2; Table 2.5). In general, the full data sets were not nearly as

consistent, which indicates the small New World ground-doves and associated lice were driving the results, perhaps due to a clade representation bias.

Signals of host-parasite cospeciation in a taxonomically biased sample may be primarily attributable to clade-limited host switching, where parasites utilizing a geographically, ecologically, and/or phylogenetically similar group of hosts preferentially switches within that particular host group. This can produce a false signal of host-parasite phylogenetic congruence (de Vienne et al., 2007). Similar effects have been observed in primate viruses (Charleston and Robertson, 2002) and brood parasitic finches (Sorenson et al., 2004). Small New World ground-doves are in a monophyletic group, are similar in size, and most forage for small seeds in brushy habitat (Gibbs et al., 2001; Sweet and Johnson, 2015). Because of these shared attributes, the wing and body lice of these doves may be able to switch within the host clade, but are limited in switching to hosts outside of the clade due to host body size or habitat proximity of the host species. Although these lice are switching hosts, the switching events are limited to the small New World ground-dove clade, perhaps contributing to host-parasite congruence in the absence of strict cospeciation.

The results from Jane differed from the ParaFit results (Figures 2.2 and 2.3; Tables 2.3 and 2.4). However, Jane is an event-based method, so the ParaFit results are not completely analogous. Event-based analyses reconcile host and parasite phylogenies by reconstructing cospeciation and duplication events at nodes and sorting and host-switching events along branches, rather than estimating the statistical significance of particular host-parasite associations. The Jane results are more consistent with previous research, with more cospeciation events recovered in the body louse analysis (22) than

the wing louse analysis (14). The Jane results are also more consistent with the ParaFit results from the analyses without small New World ground-doves and their lice. If the ground-dove/lice clades are indeed biasing ParaFit results, then the Jane (event-based) results might be giving a more accurate portrayal of the evolutionary history within these groups. It seems likely that event-based methods such as Jane are more resistant to clade representation biases, because Jane reconstructs events along every node and branch of the tree, even within clades.

External factors driving cophylogenetic patterns

Although we found no evidence of significantly correlated cophylogenetic patterns between dove wing and body lice, having worldwide sampling highlights external factors potentially associated with cophylogenetic patterns. For example, a stronger signal of cospeciation in most of the body louse data sets perhaps hints at phoresis behavior in wing lice, as has been described in previous work (Harbison et al., 2008; Harbison et al., 2009; Harbison and Clayton, 2011). Our results show that his phenomenon could be operating at a worldwide scale.

Many of the host species consistently showing evidence of cospeciation with both their wing and body lice are phabines native to Australia and/or New Guinea. The phabines are a clade that includes *Geopelia* doves, *Geophaps* pigeons, *Petrophassa* rock pigeons, *Phaps* (bronzewings), and *Ocyphaps lophotes* (crested pigeon). Although the hosts are native to the same region, geography alone does not explain these patterns, since some Australian species did not have evidence of cospeciation with their parasites (e.g. *Lopholaimus antarcticus* [topknot pigeon]). As with the small New World ground-doves and their lice, clade-limited host switching may play a role in generating these

patterns of cospeciation. Like small New World ground-doves, Australian phabines are small terrestrial foraging birds that prefer open, scrubby habitat (Gibbs et al., 2001). The combination of their hosts' small size and habitat preferences may limit opportunities for phabine lice to switch to hosts outside of the clade. However, in contrast to the small New World ground-doves and their lice, phabine body lice are not monophyletic. In addition, our event-based analyses recovered several nodes of cospeciation in the phabine clade, while only recovering a few nodes of cospeciation in the small New World ground-dove clade. Taken together, these two differences indicate that clade-limited host switching may be less of a factor in the phabine system, and that any signal of cospeciation comes from actual topological congruence between phabines and their lice.

Conclusion

Based on our results from both topology-based and event-based cophylogenetic analysis, this study does not find evidence of correlated cophylogenetic patterns between the wing and body lice of pigeons and doves. Despite finding no overall correlation, we did find potentially interesting patterns within smaller groups. Since neither the wing lice nor body lice showed perfect patterns of cospeciation with their hosts, we would expect external factors to shape the observed patterns of parasitism. As proposed in previous studies, differences in the ability to switch hosts because of differences in the use of hippoboscid flies for phoresis may be driving differences between wing and body lice. However geography, host life history, and host phylogeny are all important factors for shaping the relationship between host and parasite.

Unlike in previous studies, however, we found that both wing and body lice had evidence for cospeciation with their hosts and that body lice did not have substantially

more associations contributing to this signal than wing lice. However, when we removed the small New World ground-doves and their associated lice from the ParaFit analyses, the results seemed more in line with previous studies and predictions from ecological differences. The results were also more consistent across analyses, which was not the case with the ground-dove data included. These results highlight the importance of considering phylogenetic scale and taxa representation in cophylogenetic analysis. Results drawn from subsets of these taxa may show varying patterns dependent on the sampling level.

Host-parasite interactions are complex systems. Understanding how different factors influence the dynamics of host-parasite relationships may ultimately depend on the scale and density of taxonomic sampling. With a large and geographically extensive data set of pigeons and doves and their wing and body lice, we were able to reveal cophylogenetic patterns previously hidden by less representative sampling, and in doing so further our understanding of possible life history and geographic factors driving the patterns. In addition, we highlight possible pitfalls of cophylogenetic analyses and provide insight into the importance of identifying the proper level of taxon sampling and relative clade representation in such studies.

FIGURES

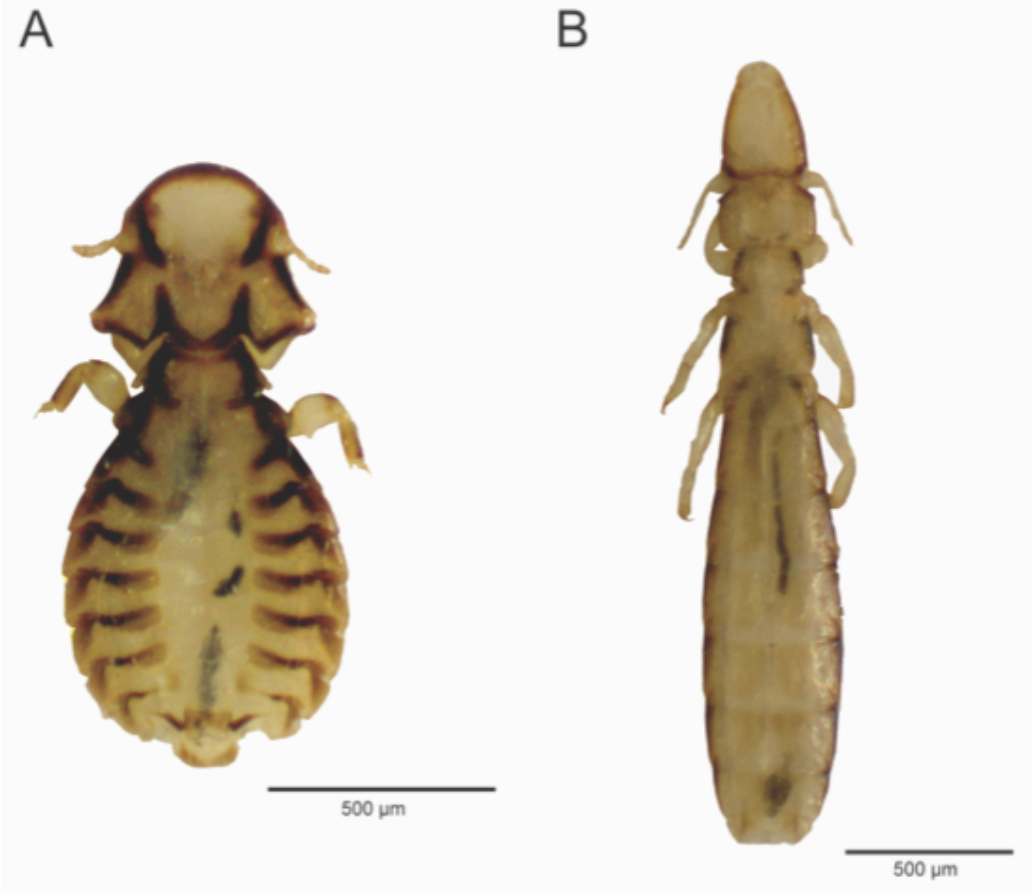


Figure 2.1. Photographs of A) a body louse (*Physconelloides emersoni*) and B) a wing louse (*Columbicola drowni*) from a black-winged ground-dove (*Metriopelia melanoptera)*. Scale indicated to the bottom right of each photograph.

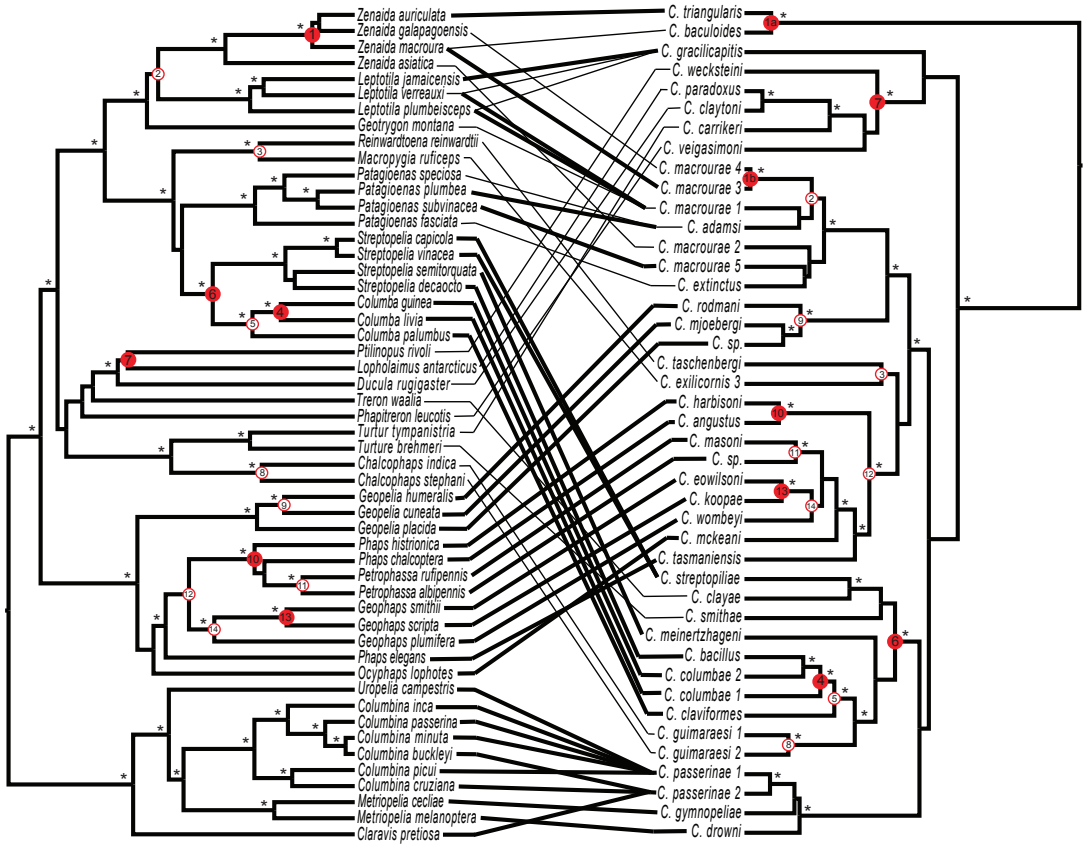


Figure 2.2. Tanglegram showing the associations between dove wing lice (right) and their hosts (left). Phylogenies were generated using BEAST v1.7.5 (Drummond et al., 2012). Asterisks (*) indicate posterior probabilities (PP) ≥0.95. Circles at nodes indicate cospeciation events as recovered by Jane v4 (Conow et al., 2010). Cospeciation events are numbered starting from the top of the host phylogeny, with matching numbers on corresponding speciation events indicated on the wing louse phylogeny. Open circles indicate recovered cospeciation events shared by wing and body lice. Bold lines between host and parasite indicate a significant link as recovered by the ParaFit (Legendre et al., 2002) F1 statistic using the phylogram topology.

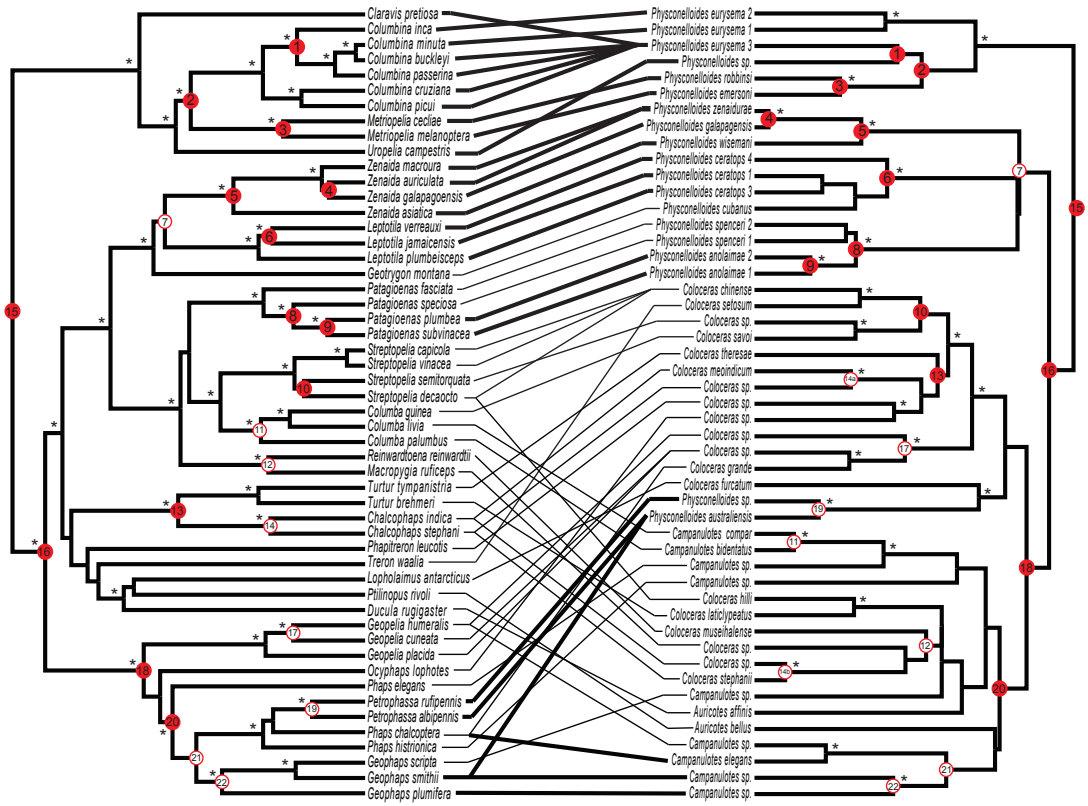


Figure 2.3. Tanglegram showing the associations between dove body lice (right) and their hosts (left). Phylogenies were generated using BEAST v1.7.5 (Drummond et al., 2012). Asterisks (*) indicate posterior probabilities (PP) ≥0.95. Circles at nodes indicate cospeciation events as recovered by Jane v4 (Conow et al., 2010). Cospeciation events are numbered starting from the top of the host phylogeny, with matching numbers on corresponding speciation events indicated on the body louse phylogeny. Open circles indicate recovered cospeciation events shared by wing and body lice. Bold lines between host and parasite indicate a significant link as recovered by the ParaFit (Legendre et al., 2002) F1 statistic using the phylogram topology.

TABLES

Table 2.1. Summary of ParaFit results for the full wing and body louse data set and the partial (excluding small New World ground-doves) data set. The ParaFitGlobal statistic and associate p-value are indicated for the ParaFit results. F1 and F2 Links refer to the number of significant ParaFitLink1 and ParaFitLink2 statistics (respectively) after correcting for false discovery rate with the Benjamini-Hochberg correction.

	ParaFit Full	
	Phylogram	**Ultrametric**
Wing	ParaFitGlobal = 1.947	ParaFitGlobal = 6.043
	P = 0.00001	P = 0.00001
F1 Links	40	0
F2 Links	43	19
Body	ParaFitGlobal = 0.276	ParaFitGlobal = 6.138
	P = 0.00001	P = 0.00007
F1 Links	33	0
F2 Links	55	0
	ParaFit Partial	
	Phylogram	**Ultrametric**
Wing		
	ParaFitGlobal = 471.8	ParaFitGlobal = 4.219
	P = 0.00002	P = 0.00003
F1 Links	12	12
F2 Links	12	12
Body		
	ParaFitGlobal = 0.132	ParaFitGlobal = 5.134
	P = 0.00001	P = 0.00001
F1 Links	27	30
F2 Links	33	31

Table 2.2. ParaFit individual link statistic p-values for both the ParaFitLink1 (F1) and ParaFitLink2 (F2) statistics of the full data set. Values listed with asterisks indicate significance after correcting for false discovery rate ($\alpha = 0.05$). Hosts are listed left of the p-values. Hosts listed more than once indicate multiple species of lice associated with that particular host.

Wing Lice					**Body Lice**				
	Phylogram		**Ultrametric**			**Phylogram**		**Ultrametric**	
	F1	F2	F1	F2		F1	F2	F1	F2
Claravis pretiosa	0.00008*	0.00007*	0.02570	0.02190	*Claravis pretiosa*	0.00011*	0.00001*	0.02856	0.02592
Uropelia campestris	0.00001*	0.00001*	0.01579	0.01328*	*Uropelia campestris*	0.00001*	0.00001*	0.01723	0.01527
Metriopelia cecliae	0.00002*	0.00002*	0.02864	0.02461	*Metriopelia cecliae*	0.00001*	0.00001*	0.01943	0.01699
Metriopelia melanoptera	0.00002*	0.00002*	0.02841	0.02421	*Metriopelia melanoptera*	0.00001*	0.00001*	0.01714	0.01495
Columbina cruziana	0.00001*	0.00001*	0.04662	0.04067	*Columbina cruziana*	0.00001*	0.00001*	0.02524	0.02303
Columbina picui	0.00001*	0.00001*	0.01189	0.01011*	*Columbina picui*	0.00001*	0.00001*	0.02481	0.02277
Columbina inca	0.00001*	0.00001*	0.01152	0.00987*	*Columbina inca*	0.00001*	0.00001*	0.06328	0.05695
Columbina passerina	0.00001*	0.00001*	0.01110	0.00958*	*Columbina passerina*	0.00001*	0.00001*	0.02342	0.02106
Columbina minuta	0.00001*	0.00001*	0.01074	0.00918*	*Columbina minuta*	0.00001*	0.00001*	0.03633	0.03301
Columbina buckleyi	0.00001*	0.00001*	0.08411	0.07502	*Columbina buckleyi*	0.00001*	0.00001*	0.02381	0.02178
Geopelia placida	0.01956*	0.01421*	0.07200	0.06285	*Geopelia placida*	0.03328	0.00008*	0.51637	0.50952
Geopelia humeralis	0.02365*	0.01787*	0.07796	0.06852	*Geopelia humeralis*	0.06950	0.00008*	0.52301	0.51534
Geopelia cuneata	0.01475*	0.01006*	0.98455	0.98682	*Geopelia humeralis*	0.06891	0.00023*	0.56601	0.55829
Ocyphaps lophotes	0.00108*	0.00053*	0.00502	0.00399*	*Geopelia cuneata*	0.02827	0.00006*	0.38403	0.37433
Geophaps plumifera	0.00056*	0.00030*	0.00223	0.00170*	*Ocyphaps lophotes*	0.55195	0.17624	0.69007	0.68750
Geophaps smithii	0.00022*	0.00007*	0.00231	0.00179*	*Geophaps plumifera*	0.02017*	0.00001*	0.03660	0.03291
Geophaps scripta	0.00028*	0.00015*	0.00324	0.00256*	*Geophaps smithii*	0.01292*	0.00001*	0.03678	0.03353
Phaps elegans	0.00003*	0.00002*	0.00569	0.00460*	*Geophaps smithii*	0.01096*	0.00001*	0.29318	0.27887
Phaps historionica	0.00205*	0.00114*	0.00566	0.00459*	*Geophaps scripta*	0.03457	0.00001*	0.28342	0.27100
Phaps chalcoptera	0.00013*	0.00008*	0.01099	0.00929*	*Phaps elegans*	0.03434	0.00003*	0.33949	0.32789

Table 2.2. Continued.

Petrophassa albipennis	0.00001*	0.00001*	0.09712	0.08609	*Phaps historionica*	0.36331	0.02395*	0.85158	0.85319
Petrophassa rufipennis	0.00003*	0.00001*	0.47186	0.45759	*Phaps chalcoptera*	0.00771*	0.00001*	0.08144	0.07461
Turtur tympanistria	0.11891	0.09854	0.28731	0.27249	*Phaps chalcoptera*	0.12494	0.00118*	0.62715	0.62295
Turtur brehmeri	0.04250	0.02994*	0.35844	0.34415	*Petrophassa albipennis*	0.00020*	0.00001*	0.22198	0.21032
Chalcophaps indica	0.06037	0.04672	0.20758	0.19174	*Petrophassa rufipennis*	0.00069*	0.00001*	0.20205	0.19010
Chalcophaps stephani	0.06646	0.05275	0.21715	0.20112	*Turtur tympanistria*	0.17887	0.00470*	0.26607	0.25424
Phapitreron leucotis	0.04800	0.03508*	0.08231	0.07290	*Turtur brehmeri*	0.13412	0.00330*	0.53348	0.52524
Treron waalia	0.42768	0.38297	0.48008	0.47061	*Chalcophaps indica*	0.04962	0.00009*	0.20788	0.19813
Lopholaimus antarcticus	0.11849	0.10004	0.04238	0.03614	*Chalcophaps indica*	0.12411	0.00203*	0.34677	0.33532
Ducula rufigaster	0.08872	0.07267	0.04223	0.03619	*Chalcophaps stephani*	0.05824	0.00015*	0.20217	0.19201
Ptilinopus rivoli	0.08741	0.07161	0.98733	0.98908	*Chalcophaps stephani*	0.14097	0.00300*	0.34242	0.33096
Geotrygon montana	0.11946	0.10241	0.01893	0.01615	*Phapitreron leucotis*	0.17388	0.00436*	0.19630	0.18541
Leptotila plumbiscens	0.01707*	0.01231*	0.11572	0.10485	*Treron waalia*	0.41405	0.05609	0.33430	0.32301
Leptotila plumbiscens	0.09522	0.07936	0.01835	0.01564*	*Lopholaimus antarcticus*	0.62149	0.04019*	0.17732	0.16627
Leptotila jamaicensis	0.01413*	0.01003*	0.10823	0.09812	*Ducula rufigaster*	0.24254	0.00200*	0.24748	0.23622
Leptotila verreauxi	0.01579*	0.01117*	0.14646	0.13503	*Ptilinopus rivoli*	0.25603	0.00314*	0.29547	0.28400
Leptotila verreauxi	0.09957	0.08351	0.01861	0.01602*	*Geotrygon montana*	0.06264	0.00040*	0.00711	0.00591
Zenaida asiatica	0.09108	0.07665	0.05169	0.04492	*Leptotila plumbiscens*	0.00975*	0.00001*	0.00565	0.00479
Zenaida macroura	0.03140	0.02398*	0.25981	0.24584	*Leptotila jamaicensis*	0.00930*	0.00001*	0.00580	0.00489
Zenaida macroura	0.10867	0.09123	0.08665	0.07844	*Leptotila verreauxi*	0.01108*	0.00001*	0.00966	0.00855
Zenaida auriculata	0.02731*	0.02027*	0.17294	0.15933	*Zenaida asiatica*	0.01177*	0.00001*	0.05639	0.05150
Zenaida galapagoensis	0.09795	0.08200	0.07817	0.07110	*Zenaida macroura*	0.00088*	0.00001*	0.05379	0.04903
Reinwardtoena reinwardtii	0.68163	0.63206	0.95386	0.95765	*Zenaida auriculata*	0.00087*	0.00001*	0.06820	0.06275
Macropygia ruficeps	0.51641	0.45329	0.95706	0.96126	*Zenaida galapagoensis*	0.00097*	0.00001*	0.08349	0.07734
Patagioenas fasciata	0.06957	0.05661	0.10090	0.09013	*Reinwardtoena reinwardtii*	0.39229	0.02314*	0.04714	0.04294
Patagioenas speciosa	0.03658	0.02829*	0.01522	0.01241*	*Macropygia ruficeps*	0.36144	0.02511*	0.04741	0.04286

Table 2.2. Continued.

Patagioenas subvinacea	0.02743*	0.02096*	0.35100	0.33643	*Patagioenas fasciata*	0.04900	0.00009*	0.28717	0.27382
Patagioenas plumbea	0.02122*	0.01559*	0.01519	0.01262*	*Patagioenas speciosa*	0.02749	0.00002*	0.89469	0.89269
Columba palumbus	0.01750*	0.01290*	0.00771	0.00584*	*Patagioenas subvinacea*	0.01469*	0.00001*	0.95666	0.95928
Columba livia	0.01276*	0.00883*	0.00718	0.00542*	*Patagioenas plumbea*	0.01227*	0.00001*	0.95821	0.96038
Columba guinea	0.01806*	0.01303*	0.16711	0.15326	*Columba palumbus*	0.50342	0.10344	0.18510	0.17503
Streptopelia semitorquata	0.03015*	0.02086*	0.03887	0.03341	*Columba livia*	0.46886	0.10415	0.18246	0.17266
Streptopelia decaocto	0.01284*	0.00881*	0.14090	0.12848	*Columba guinea*	0.07331	0.00052*	0.02042	0.01810
Streptopelia vinacea	0.01850*	0.01367*	0.05554	0.04879	*Streptopelia semitorquata*	0.01911*	0.00002*	0.20881	0.19817
Streptopelia capicola	0.01685*	0.01239*	0.05499	0.04831	*Streptopelia decaocto*	0.84719	0.83298	0.79868	0.79914
					Streptopelia decaocto	0.00641*	0.00001*	0.00549	0.00484
					Streptopelia vinacea	0.00368*	0.00001*	0.00556	0.00488
					Streptopelia capicola	0.00320*	0.00001*	0.00747	0.00656

Table 2.3. Summary of the contingency table results from ParaFit individual link statistics and cospeciation events recovered in Jane for the full data set and from ParaFit statistics for the partial (excluding small New World ground-doves) data set. Both ParaFitLink1 (F1) and ParaFitLink2 (F2) individual link statistics are reported for ParaFit. Values indicate total tallies for a particular cell of the contingency table. ParaFit values indicate the number of individual link statistics in that category after correcting for false discovery rate. Jane values indicate the number of cospeciation and/or non-cospeciation events as recovered on the host phylogeny. Results from Pearson's chi-square tests for each contingency table are listed in the right column.

Full					
wing/body	**no/no**	**no/yes**	**yes/no**	**yes/yes**	**Chi-square p-value**
Phylogram F1	14	5	12	28	0.002
Phylogram F2	1	16	4	39	1
Ultrametric F1	61	0	0	0	NA
Ultrametric F2	42	0	19	0	NA
Jane	25	14	6	8	0.286
Partial					
wing/body	**no/no**	**no/yes**	**yes/no**	**yes/yes**	
Phylogram F1	20	19	4	8	0.4481
Phylogram F2	14	25	4	8	1
Ultrametric F1	17	22	4	8	0.767
Ultrametric F2	16	23	4	8	0.8893

Table 2.4. Summary of Jane results for the wing and body louse data sets. Numbers listed are the number of events that resulted in the lowest reconstruction cost, based on the default cost parameters. Specific events are listed in the top row.

	Cospeciations	**Duplications**	**Duplications and host switches**	**Losses**	**Failures to diverge**
Wing	14	4	23	22	12
Body	22	1	25	19	9

Table 2.5. ParaFit individual link statistic p-values for both the ParaFitLink1 (F1) and ParaFitLink2 (F2) statistics of the partial (excluding small New World ground-doves) data set. Values listed with asterisks indicate significance after correcting for false discovery rate ($\alpha = 0.05$). Hosts are listed left of the p-values. Hosts listed more than once indicate multiple species of lice associated with that particular host.

Wing Lice					**Body Lice**				
	Phylogram		**Ultrametric**			**Phylogram**		**Ultrametric**	
	F1	F2	F1	F2		F1	F2	F1	F2
Geopelia placida	0.00689*	0.00689*	0.00044*	0.00030*	*Geopelia placida*	0.00159*	0.00014*	0.00017*	0.00010*
Geopelia humeralis	0.00882*	0.00881*	0.00062*	0.00048*	*Geopelia humeralis*	0.01493*	0.00296*	0.00038*	0.00025*
Geopelia cuneata	0.00529*	0.00529*	0.97899	0.98119	*Geopelia humeralis*	0.00322*	0.00046*	0.00028*	0.00021*
Ocyphaps lophotes	0.00806*	0.00805*	0.00069*	0.00057*	*Geopelia cuneata*	0.00106*	0.00015*	0.00015*	0.00012*
Geophaps plumifera	0.01338	0.01337	0.00033*	0.00025*	*Ocyphaps lophotes*	0.98095	0.98148	0.00155*	0.00107*
Geophaps smithii	0.00367*	0.00367*	0.00035*	0.00028*	*Geophaps plumifera*	0.00993*	0.00162*	0.00078*	0.00059*
Geophaps scripta	0.00321*	0.00321*	0.00056*	0.00049*	*Geophaps smithii*	0.00548*	0.00088*	0.14195	0.11651
Phaps elegans	0.00122*	0.00121*	0.00061*	0.00050*	*Geophaps smithii*	0.00175*	0.00020*	0.00069*	0.00052*
Phaps historionica	0.01011*	0.01011*	0.00068*	0.00058*	*Geophaps scripta*	0.54828	0.39861	0.50612	0.49294
Phaps chalcoptera	0.98669	0.98669	0.92363	0.92597	*Phaps elegans*	0.29108	0.14308	0.43060	0.41543
Petrophassa albipennis	0.00036*	0.00036*	0.00091*	0.00073*	*Phaps historionica*	0.90047	0.92751	0.73173	0.72954
Petrophassa rufipennis	0.00083*	0.00083*	0.00653*	0.00547*	*Phaps chalcoptera*	0.00020*	0.00004*	0.00212*	0.00145*
Turtur tympanistria	0.10294	0.10293	0.76565	0.76732	*Phaps chalcoptera*	0.00240*	0.00033*	0.00147*	0.00106*
Turtur brehmeri	0.11686	0.11684	0.28719	0.27598	*Petrophassa albipennis*	0.00001*	0.00001*	0.00158*	0.00131*
Chalcophaps indica	0.03282	0.03280	0.06931	0.06326	*Petrophassa rufipennis*	0.00011*	0.00002*	0.00354*	0.00266*
Chalcophaps stephani	0.03989	0.03989	0.06940	0.06321	*Turtur tympanistria*	0.37395	0.18811	0.01058*	0.00809*
Phapitreron leucotis	0.10146	0.10144	0.16257	0.15303	*Turtur brehmeri*	0.05223	0.01466*	0.43786	0.41939
Treron waalia	0.42496	0.42495	0.54868	0.54396	*Chalcophaps indica*	0.03481	0.00862*	0.16752	0.15235

Table 2.5. Continued.

Lopholaimus antarcticus	0.21807	0.21805	0.09336	0.08558	*Chalcophaps indica*	0.26451	0.11194	0.89698	0.90062
Ducula rufigaster	0.13632	0.13632	0.09364	0.08582	*Chalcophaps stephani*	0.04826	0.01226*	0.23149	0.21464
Ptilinopus rivoli	0.13266	0.13265	0.17689	0.16746	*Chalcophaps stephani*	0.25180	0.11168	0.89360	0.89649
Geotrygon montana	0.21255	0.21254	0.07552	0.06918	*Phapitreron leucotis*	0.12873	0.04992	0.59081	0.58004
Leptotila plumbiscens	0.08902	0.08899	0.25736	0.24790	*Treron waalia*	0.78962	0.79540	0.99065	0.99206
Leptotila plumbiscens	0.14086	0.14084	0.09413	0.08715	*Lopholaimus antarcticus*	0.70867	0.67833	0.50855	0.49539
Leptotila jamaicensis	0.07381	0.07377	0.18292	0.17346	*Ducula rufigaster*	0.21583	0.10358	0.21074	0.19447
Leptotila verreauxi	0.08507	0.08506	0.28267	0.27358	*Ptilinopus rivoli*	0.22241	0.10969	0.24347	0.22563
Leptotila verreauxi	0.14746	0.14744	0.13489	0.12628	*Geotrygon montana*	0.05016	0.01240*	0.00953*	0.00706*
Zenaida asiatica	0.94321	0.94321	0.22428	0.21358	*Leptotila plumbiscens*	0.00698*	0.00085*	0.00603*	0.00409*
Zenaida macroura	0.18222	0.18221	0.37579	0.36634	*Leptotila jamaicensis*	0.00467*	0.00060*	0.00677*	0.00498*
Zenaida macroura	0.13859	0.13857	0.03719	0.03284	*Leptotila verreauxi*	0.01772*	0.00285*	0.01842*	0.01434*
Zenaida auriculata	0.14541	0.14541	0.32555	0.31615	*Zenaida asiatica*	0.06730	0.01925*	0.11565	0.10233
Zenaida galapagoensis	0.12229	0.12228	0.03254	0.02892	*Zenaida macroura*	0.00106*	0.00007*	0.00699*	0.00537*
Reinwardtoena reinwardtii	0.65526	0.65524	0.91886	0.92173	*Zenaida auriculata*	0.00095*	0.00009*	0.00712*	0.00560*
Macropygia ruficeps	0.18765	0.18763	0.66010	0.65607	*Zenaida galapagoensis*	0.00641*	0.00080*	0.03422	0.02821*
Patagioenas fasciata	0.15042	0.15042	0.07117	0.06465	*Reinwardtoena reinwardtii*	0.33631	0.19851	0.64246	0.63398
Patagioenas speciosa	0.10359	0.10358	0.01453	0.01255	*Macropygia ruficeps*	0.33480	0.19545	0.85180	0.85395
Patagioenas subvinacea	0.97986	0.97986	0.10580	0.09755	*Patagioenas fasciata*	0.02949	0.00717	0.00037*	0.00029*
Patagioenas plumbea	0.07594	0.07594	0.02538	0.02205	*Patagioenas speciosa*	0.02108*	0.00382*	0.00046*	0.00028*
Columba palumbus	0.14230	0.14227	0.08380	0.07309	*Patagioenas subvinacea*	0.00508*	0.00077*	0.00069*	0.00050*
Columba livia	0.04722	0.04722	0.07566	0.06580	*Patagioenas plumbea*	0.00399*	0.00068*	0.00217*	0.00150*
Columba guinea	0.05858	0.05855	0.52160	0.51535	*Columba palumbus*	0.78127	0.72592	0.19800	0.18036
Streptopelia semitorquata	0.03934	0.03934	0.23935	0.22740	*Columba livia*	0.99839	0.99981	0.18988	0.17219
Streptopelia decaocto	0.05605	0.05605	0.50478	0.49799	*Columba guinea*	0.12180	0.04572	0.00140*	0.00099*
Streptopelia vinacea	0.03009	0.03009	0.13042	0.12199	*Streptopelia semitorquata*	0.07935	0.02283*	0.02707*	0.02190*

Table 2.5. Continued.

Streptopelia capicola	0.02854	0.02853	0.13205	0.12331	*Streptopelia decaocto*	0.70843	0.67464	0.92208	0.92600
					Streptopelia decaocto	0.00650*	0.00100*	0.00012*	0.00007*
					Streptopelia vinacea	0.00387*	0.00052*	0.00010*	0.00009*
					Streptopelia capicola	0.00307*	0.00041*	0.00011*	0.00007*

LITERATURE CITED

Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57: 289-300.

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

Bush S.E., Malenke J.R. 2007. Host defence mediates interspecific competition in ectoparasites. *Journal of Animal Ecology* 77: 558–564.

Charleston M.A., Robertson D.L. 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Systematic Biology* 51: 528–535.

Clayton D.H. 1991. Coevolution of avian grooming and ectoparasite avoidance. In: Loye JE, Zuk M, eds. Bird-parasite interactions: ecology, evolution, and behavior. Oxford: Oxford University Press, 258-289.

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

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

Clayton, D.H., Koop, J.A.H., Harbison C.W., Moyer B.R., Bush S.E. 2010. How birds combat ectoparasites. *The Open Ornithology Journal* 3: 41–71.

Clayton D.H., Moyer B.R., Bush S.E., Jones T.G., Gardiner D.W., Rhodes B.B., Goller F. 2005. Adaptive significance of avian beak morphology for ectoparasite control. Proceedings. *Biological Sciences / The Royal Society* 272: 811–817.

Clayton D.H., Walther B.A. 2001. Influence of host ecology and morphology on the diversity of Neotropical bird lice. *Oikos* 94: 455–467.

Clayton D.H., Lee P., Tompkins D., Brodie I. 1999. Reciprocal natural selection on host-parasite phenotypes. *The American Naturalist* 154: 261–270.

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.

Cruickshank R.H., Johnson K.P., Smith V.S., Adams R.J., Clayton D.H., Page R.D. 2001. Phylogenetic analysis of partial sequences of elongation factor 1alpha identifies major groups of lice (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 19: 202–215.

Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

de Vienne D.M., Giraud T, Shykoff J.A. 2007. When can host shifts produce congruent host and parasite phylogenies? A simulation approach. *Journal of Evolutionary Biology* 20: 1428–1438.

de Vienne D.M., Refregier G., Lopez-Villavicencio M., Tellier A., Hood M.E., Giraud T. 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist* 198: 347–385.

Dobson A., Lafferty K.D., Kuris A.M., Hechinger R.F., Jetz W. 2008. Colloquium paper: homage to Linnaeus: how many parasites? How many hosts? *Proceedings of the National Academy of Sciences* 105: 11482–11489.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Eichler W.D. 1948. Some rules in ectoparasitism. *The Annals and Magazine of Natural History* 1: 588–598.

Fahrenholz H. 1913. Ectoparasiten und abstammungslehre. *Zoologischer Anzeiger* 41: 371–374.

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

Goodwin D. 1983. Pigeons and doves of the world (3rd ed.). Ithaca, NY: Cornell University Press.

Gorrell J.C., Schulte-Hostedde A.I. 2008. Patterns of parasitism and body size in red squirrels (*Tamiasciurus hudsonicus*). *Canadian Journal of Zoology* 86: 99–107.

Gouy M., Guindon S., Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27: 221–224.

Hafner M.S., Nadler S.A. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 332: 258-259.

Hafner M.S., Nadler S.A. 1990. Cospeciation in Host-Parasite Assemblages: Comparative Analysis of Rates of Evolution and Timing of Cospeciation Events. *Systematic Biology* 39: 192–204.

Hafner M.S., Page R.D. 1995. Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philosophical Transactions of the Royal Society* 349: 77–83.

Hafner M.S., Sudman P.D., Villablanca F.X., Spradling T.A., Demastes J.W., Nadler S.A. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265: 1087–1090.

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

Harbison C.W., Clayton D.H. 2011. Community interactions govern host-switching with implications for host-parasite coevolutionary history. *Proceedings of the National Academy of Sciences* 108: 9525–9529.

Harbison C.W., Jacobsen M.V., Clayton D.H. 2009. A hitchhiker's guide to parasite transmission: The phoretic behaviour of feather lice. *International Journal for Parasitology* 39: 569–575.

Hughes J., Kennedy M., Johnson K.P., Palma R.L., Page R.D.M. 2007. Multiple cophylogenetic analyses reveal frequent cospeciation between pelecaniform birds and Pectinopygus lice. *Systematic Biology* 56: 232–251.

Jackson A.P., Machado C.A., Robbins N., Herre E.A. 2008. Multi-locus phylogenetic analysis of neotropical figs does not support co-speciation with the pollinators :

The importance of systematic scale in fig / wasp cophylogenetic studies. *Symbiosis* 45: 16.

Johnson K.P., Clayton D.H. 2004. Untangling coevolutionary history. *Systematic Biology* 53: 92–94.

Johnson K.P. 2004. Deletion bias in avian introns over evolutionary timescales. *Molecular Biology and Evolution* 21: 599–602.

Johnson K.P., Adams R.J., Clayton D.H. 2001. Molecular systematics of Goniodidae (Insecta: Phthiraptera). *The Journal of Parasitology* 87: 862–869.

Johnson K.P., Adams R.J., Page R.D.M., Clayton D.H. 2003. When do parasites fail to speciate in response to host speciation? *Systematic Biology* 52: 37–47.

Johnson K.P., Bush S.E., Clayton D.H. 2005. Correlated evolution of host and parasite body size: tests of Harrison's Rule using birds and lice. *Evolution* 59: 1744–1753.

Johnson K.P., Clayton D.H. 2000. A molecular phylogeny of the dove genus Zenaida: mitochondrial and nuclear DNA sequences. *The Condor* 102: 864–870.

Johnson K.P., Clayton D.H. 2003. Coevolutionary history of ecological replicates: comparing phylogenies of wing and body lice to Columbiform hosts. In: Page R.D.M., ed. Tangled trees: phylogeny, cospeciation, and coevolution. Chicago: University of Chicago Press. 260-286.

Johnson K.P., Kort D.S., Dinwoodey K., Mateman A.C., Cate C.T., Lessells C.M., Clayton D.H. 2001. A molecular phylogeny of the dove genera *Streptopelia* and *Columba*. *The Auk* 118: 874–887.

Johnson K.P., Malenke J.R., Clayton D.H. 2009. Competition promotes the evolution of host generalists in obligate parasites. *Proceedings. Biological Sciences / The Royal Society* 276: 3921–3926.

Johnson K.P., Reed D.L., Hammond Parker S.L., Kim D., Clayton D.H. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for Columbicola (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 45: 506–518.

Johnson K.P., Shreve S.M., Smith V.S. 2012. Repeated adaptive divergence of microhabitat specialization in avian feather lice. *BMC Biology* 10: 52.

Johnson K.P., Weckstein J.D. 2011. The Central American land bridge as an engine of diversification in New World doves. *Journal of Biogeography* 38: 1069–1076.

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

Johnson K.P., Weckstein J.D., Meyer M.J., Clayton D.H. 2011b. There and back again: switching between host orders by avian body lice (Ischnocera: Goniodidae). *Biological Journal of the Linnean Society* 102: 614–625.

Johnson K.P., Williams B.L., Drown D.M., Adams R.J., Clayton D.H. 2002. The population genetics of host specificity: genetic differentiation in dove lice (Insecta: Phthiraptera). *Molecular Ecology* 11: 25–38.

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

Maddison W.P., Maddison D.R. 2011. Mesquite: a modular system for evolutionary analysis. Available at http://mesquiteproject.org

Mora C., Tittensor D.P., Adl S., Simpson A.G.B., Worm B. 2011. How many species are there on earth and in the ocean? *PLoS Biology* 9: 1–8.

Nelson B.C., Murray M.D. 1971. The distribution of Mallophaga on the domestic pigeon (*Columba livia*). *International Journal for Parasitology* 1: 21–29.

Page R.D.M. 1994. Parallel Phylogenies: Reconstructing the History of Host-Parasite Assemblages. *Cladistics* 10: 155–173.

Page R.D.M. 2003. Tangled trees: phylogeny, cospeciation, and coevolution. Chicago: The University of Chicago Press.

Page R.D.M, Cruickshank R.H., Dickens M., Furness R.W., Kennedy M., Palma R.L., Smith V.S. 2004. Phylogeny of "*Philoceanus* complex" seabird lice (Phthiraptera: Ischnocera) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 30: 633–52.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.

Paterson A.M., Wallis G.P., Wallis L.J., Gray R.D. 2000. Seabird and louse coevolution: complex histories revealed by 12S and rRNA sequences and reconciliation analyses. *Systematic Biology* 49: 383-399.

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

Poulin R. 1997. Species richness of parasite assemlages: evolution and patterns. *Annual Review of Ecology and Systematics* 28: 341–358.

Poulin R. 2007. Are there general laws in parasite ecology? *Parasitology* 134: 763–776.

Poulin R., Morand S. 2000. The diversity of parasites. *The Quarterly Review of Biology*, 75: 277–293.

Poulin R., Morand S. 2004. Parasite biodiversity. Washington, D.C.: Smithsonian Institution Press.

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

Rambaut A., Drummond A.J. 2007. Tracer v1.4. Available at http://tree.bio.ed.ac.uk/software/tracer/

Rothschild M., Clay T. 1952. Fleas, flukes, and cuckoos: a study of bird parasites (3rd ed.). London: Collins.

Sorenson M.D., Balakrishnan C.N., Payne R.B. 2004. Clade-limited colonization in brood parasitic finches (Vidua spp.). *Systematic Biology* 53: 140–153.

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

Sugiura N. 1978. Further analysts of the data by akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods* 7: 13–26.

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

Waite J.L., Henry A.R., Clayton D.H. 2012. How effective is preening against mobile ectoparasites? An experimental test with pigeons and hippoboscid flies. *International Journal for Parasitology* 42: 463–467.

Weckstein J.D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology* 53: 154–164.

Windsor D.A. 1998. Most of the species on Earth are parasites. *International Journal for Parasitology* 28: 1939–1941.

CHAPTER 3: COMPARATIVE COPHYLOGENETICS OF AUSTRALIAN PHABINE PIGEONS AND DOVES (AVES: COLUMBIDAE) AND THEIR FEATHER LICE (INSECTA: PHTHIRAPTERA)[2]

INTRODUCTION

Parasitic organisms are ubiquitous in most biological systems. Their ability to occupy a variety of niches has resulted in great diversity and many independent transitions from free-living to parasitic lifestyles (Poulin and Morand, 2000; Poulin, 2011; Poulin and Randhawa, 2015). Some organisms parasitize many different hosts throughout their life cycles, and may even have a free-living life stage (Gandon, 2004; Banks and Paterson, 2005; Belzile and Gosselin, 2015). Other parasites are more tightly associated with their hosts, spending their entire life cycle on a single host and being limited to a particular species or group of hosts (Rohde, 1979; Hafner et al., 1994; Hafner and Page, 1995; Proctor, 2003). In cases in which parasite reproduction is heavily linked to the host, the diversification patterns (phylogenies) of these obligate parasites may mirror those of their hosts. In these cases, when a host undergoes speciation, its obligate parasites may also cospeciate, causing the parasite phylogeny to be congruent with the host phylogeny (Fahrenholz, 1913; Eichler, 1948). However, this expectation

[2] Reprinted, with permission, from: Sweet, A.D., R.T. Chesser, and K.P. Johnson. 2017. Comparative cophylogenetics of Australian phabine pigeons and doves (Aves: Columbidae) and their feather lice (Insecta: Phthiraptera). *International Journal for Parasitology*. 47: 347-356. https://doi.org/10.1016/j.ijpara.2016.12.003.

is rarely observed in nature. Although some obligate parasite groups exhibit patterns of congruence with their host's phylogeny, most exhibit some level of incongruence generated by host switching, duplication or sorting events during their evolutionary history with their hosts (Page, 1994; Page and Charleston, 1998). The degree of incongruity can vary among different host groups, and even among different groups of parasites associated with the same group of hosts (Whiteman et al., 2007; Toon and Hughes, 2008; Bueter et al., 2009; Stefka et al., 2011).

The feather lice (Insecta: Phthiraptera: Philopteridae) of pigeons and doves (Aves: Columbidae) are an example of obligate parasites that have varying levels of congruence between host and parasite phylogenies. Pigeons and doves harbor two types (ecomorphs) of feather lice: wing and body lice (Johnson et al., 2012). These two groups are not closely related, and their morphologies differ dramatically (Cruikshank et al., 2001). Wing lice are long and slender, and insert themselves between wing and tail feather barbs to avoid removal by host preening. In contrast, body lice are round and escape preening by burrowing into the downy feathers close to their host body (Clayton, 1991; Clayton et al., 2005, 2010). However, both types of lice eat the downy feathers of their hosts (Nelson and Murray, 1971). Comparative cophylogenetic analysis of wing and body lice from New World pigeons and doves indicates that body lice have a fairly congruent evolutionary history with their hosts, whereas wing lice exhibit less congruence and do not show evidence for cospeciation (Clayton and Johnson, 2003; Johnson and Clayton, 2004). The body lice of pigeons and doves are also more host-specific than wing lice, meaning that wing louse species are more often associated with multiple host species (Johnson et al., 2002). This difference may be due, in part, to the greater ability and incidence of wing lice using hippoboscid flies for transport (phoresis) within and among host species (Keirans, 1975; Harbison et al., 2008; Harbison and Clayton, 2011). Experimental studies have

indicated that wing lice are much more likely than body lice to successfully switch hosts using this behavior (Harbison et al., 2009). Globally across Columbidae, both groups of lice do show significant congruence with the host phylogeny; however, it is unclear how much of this congruence is due to shared biogeographic patterns (Sweet et al., 2016). It is important, therefore, to examine congruence within additional biogeographic regions to determine whether patterns observed within New World taxa also hold for other regional host-parasite faunas.

In this study we focus on the wing and body lice of phabine pigeons and doves, a monophyletic group of birds from Australia and southeastern Asia (Johnson and Clayton, 2000b; Pereira et al., 2007). By exploring the cophylogenetic patterns of a distinct group of birds and their lice, we can test whether the patterns these taxa exhibit are similar to those exhibited by New World taxa. Phabines are a monophyletic group of 15 species in the genera *Phaps*, *Geophaps*, *Ocyphaps*, *Petrophassa*, *Geopelia* and *Leucosarcia* (Pereira et al., 2007). Most representatives are primarily terrestrial and prefer arid, open scrub, or dry forest habitats (Goodwin, 1983; Gibbs et al., 2001). However, some species (*Leucosarcia melanoleuca* and *Geopelia humeralis*) occupy more humid, wetter habitats. As with other terrestrial doves, phabines primarily forage on small seeds and fruits. All phabine wing lice belong to the genus *Columbicola* (Price et al., 2003) whereas phabine body lice are classified into three genera (*Campanulotes, Coloceras* and *Physconelloides*).

We sampled most representatives of phabines together with their wing and body lice, focusing particularly on species from continental Australia. From these samples we sequenced or used existing sequences of multiple molecular loci, and used these sequences to estimate molecular phylogenies for all three groups. We then performed several cophylogenetic analyses

to test for congruence between the phylogenies of phabine pigeons and doves, and those of their wing and body lice.

MATERIALS AND METHODS

Sampling and sequencing

We obtained samples for 12 species of Australian phabine pigeons and doves and their wing (12 samples) and body lice (15 samples). For outgroup taxa, we used available GenBank sequences of *Columbina passerina, Zenaida macroura, Ptilinopus rivoli* and *Chalcophaps indica* for hosts, *Columbicola passerinae* (ex. Inca dove (*Columbina inca*)) for wing lice, and *Goniocotes talegallae* (ex. black-billed brushturkey (*Talegalla fuscirostris*)) for body lice. Muscle tissue was extracted from birds collected in the field and stored at -80 °C. Lice were collected in the field with pyrethrin powder or fumigation protocols (Clayton and Drown, 2001) and stored in 95% ethanol at -80 °C. DNA was extracted from bird tissue using a Qiagen Blood and Tissue Kit (Qiagen, Valencia, CA, USA) with standard protocols. DNA was extracted from individual louse specimens using a modified Qiagen protocol, with louse specimens incubating in a proteinase K/buffer solution at 55 °C for ~48 h. PCR was used to target genes for Sanger sequencing, using a Promega taq kit (Promega, Madison, WI, USA) according to recommended protocols. PCR products were purified with a Qiagen PCR Purification Kit according to standard protocols. For birds, 381 bp of the mitochondrial gene cytochrome oxidase subunit 1 (*Cox*1), 1,074 bp of NADH dehydrogenase subunit 2 (ND2), and 1,172 bp of the nuclear gene beta-fibrinogen intron 7 and flanking exon regions (FIB7) were sequenced. For wing lice, 383 bp of *Cox*1, 379 bp of 12S rRNA (12S), and 360 bp of the nuclear gene elongation factor 1α (EF-1α) were sequenced. For body lice, 383 bp of *Cox*1, 362 bp of EF-1α, and 553 bp of 16S rRNA

(16S) were sequenced. Sequencing primers and amplification protocols were used as outlined in Johnson and Clayton (2000a, b), and Johnson et al. (2007, 2011b). Resulting PCR products were sequenced with an ABI Prism BigDye Terminator kit (Applied Biosystems, Foster City, CA, USA), and fragments were run on an AB 3730x capillary sequencer at the University of Illinois Roy J. Carver Biotechnology Center (Champaign, IL, USA). Resulting complementary chromatograms were manually resolved and primer sequences removed in Sequencher v5.0.1 (Gene Codes, Ann Arbor, MI, USA) or Geneious v8.1.2 (Biomatters, Auckland, NZ). We submitted all resulting sequence files to GenBank (Supplementary Figure 3.1).

Phylogenetic analysis

All genes were aligned using the default parameters of the MAFFT plugin in Geneious (Katoh et al., 2002) and each resulting alignment was checked manually. For protein coding loci, alignments were trimmed to be within reading frame. Maximum-likelihood (ML) phylogenies were estimated using RAxML v.8.1.17 (Stamatakis, 2014) for each gene alignment, using 200 rapid bootstrap replicates (-f a) and GTR + Γ (GTRGAMMA) nucleotide substitution models. Finally, for each data set (doves, wing lice, and body lice) the gene alignments were concatenated in Geneious. PartitionFinder v1.1.1 (Lanfear et al., 2012) was used to test for appropriate partitioning schemes and substitution models for the concatenated alignments, setting up potential partition schemes according to genes and using the corrected Akaike Information Criterion (AICc) to test for the best fitting substitution models (Sugiura, 1978). PartitionFinder searched through all possible models, and again only through models available in MrBayes. Partition and model results are listed in Figure 3.1.

Partitioned ML and Bayesian analyses were run for the concatenated alignments in all three data sets. ML estimations were run in Garli v2.0 (Zwickl, D.J., 2006. Genetic algorithm

approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin, USA) with two searches of 500 bootstrap replicates, and summarizing the bootstrap trees using SumTrees in the DendroPy package (Sukumaran, J., Holder, M., 2008. SumTrees: Summarization of split support on phylogenetic trees v1.0.2). Best trees for each concatenated alignment were also estimated using Garli. Bayesian estimations were run in MrBayes v3.2 (Ronquist and Huelsenbeck, 2003). Two runs of four Markov chain Monte Carlo (MCMC) chains were run for 20 million generations, sampling every 1,000 trees. Resulting .p files were viewed in Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/) to assess parameter convergence, and .t files were analyzed in RWTY v1.0.0 (Warren, D., Geneva, A., Lanfear, R., 2016. rwty: R We There Yet? Visualizing MCMC convergence in phylogenetics. v1.0.0) to assess topological convergence. Based on these assessments the first 10% (2,000) trees were discarded as a burnin.

Taxonomic assessment

Properly defining host and parasite taxonomic units is critical for cophylogenetic analysis. Because avian lice often harbor cryptic species, it was necessary to more objectively evaluate the number of species in the wing and body louse data sets. To this end we used distances values and the Automatic Barcode Gap Discovery (ABGD) method for wing and body louse *Cox1* sequence data (Puillandre et al., 2012). Uncorrected pairwise distance matrices were generated from the wing and body louse *Cox*1 alignments using the "dist.dna" command in the R package *ape* (Paradis et al., 2004; R Development Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing.). The *Cox*1 alignment was then used as input for the web version of ABGD (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html), applying default Pmin, Pmax, and

Steps parameters and adjusting the relative gap width to 1.0. ABGD was run with both Jukes-Cantor (JC) and Kimura80 (K80) models. Based on these taxonomic assessments, the louse trees were trimmed accordingly before using those as input for cophylogenetic analyses.

Cophylogenetic analysis

We implemented both distance-based and event-based cophylogenetic methods. For distance-based analysis we used ParaFit (Legendre et al., 2002). This method takes host and parasite distances matrices and a host-parasite association matrix as input to test for congruence between the two trees. It also tests for the contribution of each individual link to the global patterns. For ParaFit, the best host and parasite trees from the Garli analyses were converted to patristic distance matrices using the "cophenetic" command in *ape*, and ordered the resulting distance matrices according to the association matrix. ParaFit was then run for 999 permutations with the Cailliez correction for negative eigenvalues and tested for the contribution of each individual link to the global metric with the ParaFitLink1 and ParaFitLink2 statistics. Because the individual link statistics are multiple tests, false discovery rate was corrected for using the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995). To account for poorly supported relationships, ParaFit was also run as described above with 50% majority-rule host and parasite consensus trees. SumTrees generated consensus trees from the Garli bootstrap analyses.

For the event-based analysis, we used Jane v4 (Conow et al., 2010). Instead of testing for global congruence and individual link contributions, Jane is a genetic algorithm (GA) that seeks to reconstruct evolutionary events (e.g. cospeciation, host switches) at the nodes and branches of the host and parasite trees. Jane reconstructed events using the recommended GA parameters of population size twice the number of generations (number of generations = 500, population size = 1,000) and default event costs (cospeciation = 0, duplication = 1, duplication and host switch = 2,

loss = 1, and failure to diverge = 1). Jane also randomized the tip mappings 999 times to test for the probability of obtaining our observed overall cost. An observed cost significantly lower than the randomized costs would indicate global host-parasite congruence.

To test whether lice switch between hosts of similar size, we reconstructed ancestral host body size using the "ace" command in *ape*, implementing the ML method under the Brownian motion model for continuous traits. The best ML phabine phylogeny from Garli was used as the input tree. Body sizes were assigned to the tree tips as the average mass (g) from Gibbs et al. (2001). After running the state reconstruction, the absolute difference in average host size was calculated between two host nodes/tips involved in a host switch based on the Jane analyses. These values were then averaged separately for the wing and body louse switches.

RESULTS

The body louse matrix was 88% complete, with three samples missing EF-1α and 16S sequences. There were no missing data for the wing lice or phabine hosts (Supplementary Figure 3.1). Statistics on individual loci are indicated in Figure 3.2. The concatenated dove alignment was 2,627 bp, concatenated body louse alignment 1,298 bp, and concatenated wing louse alignment 1,122 bp.

Phylogenetic analysis

Likelihood and Bayesian analyses of the birds (Figure 3.1) provided strong support for monophyly of the phabines and of most genera (*Geophaps*, *Petrophassa* and *Geopelia*), and moderate support for monophyly of *Phaps*. Relationships among genera were weakly supported. The wing louse phylogenies (Figure 3.2) indicated, with the exception of *Geophaps*, that lice from the same host genus formed monophyletic groups, with support ranging from rather weak

(*Phaps*) to strong (*Geopelia*, *Petrophassa*). Relationships among these major groups of wing lice were generally poorly supported. In contrast to the phylogeny of wing lice, the phylogeny of body lice (Figure 3.3) did not contain any monophyletic groups of lice confined to a single host genus. In addition, even within clades, the phylogeny of body lice was relatively uncorrelated with host taxonomy.

All individual gene trees from each data set did not have any well-supported conflicting nodes (Supplementary Figures 3.1-3.3). For all three concatenated data sets, the partitioned ML and Bayesian analyses estimated similar tree topologies. In all cases, the MCMC chains from the Bayesian analyses had parameter effective sample size (ESS) values >200 and average S.D. of split frequencies <0.01, indicating that the analyses converged to stationarity.

Cophylogenetic analysis

Analysis of the uncorrected p-distances of *Cox*1 and ABGD indicated in two cases that two body louse samples should be collapsed to a single taxon: *Coloceras* sp. from *Geopelia placida* and *Coloceras* sp. from *Geopelia cuneata*, and *Physconelloides australiensis* from *Petrophasa albipennis* and *Geophaps smithii*. The latter result agrees with current taxonomic treatment of these lice (Price et al., 2003; Johnson et al., 2011b). The *Cox*1 sequences within each pair were identical (Supplementary Figure 3.2), and the ABGD analysis likewise indicated that each pair should be considered a single taxon. We did not find any support for collapsing tips of the wing louse phylogeny. Mean uncorrected p-distance between all pairs of taxa was >12% (Supplementary Figure 3.3). Most species of wing lice in our dataset have been previously described (only two samples are undescribed species), in contrast to our body louse data set (11 samples are undescribed species).

Using the most likely host and parasite species phylogenies as input, ParaFit did not reject the null hypothesis of a random association between phabines and their body lice (ParaFitGlobal = 0.041, $P = 0.069$). The individual link tests (ParaFitLink1 and ParaFitLink2) did not recover any significant links after correcting for false discovery rate with the Benjamini-Hochberg correction (Figure 3.3). ParaFit also indicated random association using the 50% majority rule consensus trees (ParaFitGlobal = 0.023, $P = 0.081$), and did not recover any significant links after correction (Supplementary Figure 3.4). The Jane event-based reconstruction recovered only three potential cospeciation events between phabines and their body lice: one cospeciation event at the *Geopelia cuneata*/*Geopelia humeralis* split, one at the *Geophaps scripta*/*Geophaps smithii* split, and a third at the *Phaps/Geophaps* split (Figure 3.4). Jane also recovered eight host-switching events, one duplication, four losses and two failures to diverge, for a total cost of 23 (Figure 3.4). Other reconstructions with an identical total cost recovered zero duplications, nine host switches, three losses and two failures to diverge. The Jane randomization test indicated the observed cost was not lower than by chance ($P = 0.161$), suggesting no congruence between the phylogenetic tree of phabine body lice and their hosts. From the ancestral state reconstruction of host size, the average absolute difference in host size between phabine nodes/tips involved in body louse host switches was 111.69 g.

Comparing the most likely phabine host and wing louse phylogenies, ParaFit indicated global phylogenetic congruence between the two groups. The ParaFitGlobal test indicated a non-random host-parasite association (ParaFitGlobal = 0.322, $P = 0.005$), and the ParaFit individual link tests included three significant links after correction, all between *Geopelia* doves and their wing lice (Figure 3.5). ParaFit also indicated a significant global association between the 50% majority-rule consensus trees (ParaFitGlobal = 0.496, $P = 0.004$), and significant links between

Geopelia and their wing lice. In addition, the consensus analysis recovered a significant link between *Ocyphaps lophotes* and its wing louse (Supplementary Figure 3.5). Similarly, Jane recovered eight cospeciation events between phabines and their wing lice (Figure 3.5). The reconstruction recovered cospeciation events at both *Geopelia* splits, one at the *Phaps chalcoptera*/*Phaps elegans* split, at the *Geophaps smithii*/*Geophaps scripta* split, at the *Petrophassa rufipennis*/*Petrophassa albipennis* split, the *Petrophassa*/*Ocyphaps* split, and at the two deepest nodes. Jane also recovered three host-switching events and two losses between phabines and their wing lice, for a total cost of 8 (Figure 3.4). The average absolute difference in host size for phabine nodes/tips involved in wing louse host switches was 41.67 g. The Jane randomization test indicated that the observed cost was significantly lower than by chance ($P < 0.001$), suggesting a history of repeated cospeciation in this group of lice.

DISCUSSION

Comparisons of molecular phylogenies for Australian phabine pigeons and doves and their wing and body louse parasites revealed that the phylogeny of wing lice was highly congruent with that of their hosts, whereas the phylogeny of body lice was not. These results were consistent with both best and consensus trees, indicating the pattern is not an artifact of poor topology support. This result stands in dramatic contrast to patterns found for New World pigeons and doves and their lice (Clayton and Johnson, 2003; Johnson and Clayton, 2004), in which the phylogeny of body lice generally matched that of their hosts while the phylogeny of wing lice did not. However, in a study with a worldwide sample of pigeons and doves, both wing and body lice showed evidence of cophylogenetic congruence with their hosts (Sweet et al., 2016). The differences between the New World and Australian studies suggest that

biogeographic differences may exist in factors that promote congruence in wing and body lice. Thus, local congruence in some cases may be driving congruence at the global scale. Pigeons and doves are widespread birds, thriving in a variety of ecosystems in every continent other than Antarctica. Due to this geographical and ecological diversity, the evolutionary patterns exhibited by pigeons and doves, and their parasitic lice, may differ among different groups of hosts, especially because parasite diversification can be heavily affected by external factors (e.g., ecology or geography) (Paterson et al., 2000; Weckstein, 2004; Bush and Malenke, 2008; Bruyndonchx et al., 2009).

Regional differences in congruence may reflect regional differences in the abundance of hippoboscid flies, which wing lice can use to disperse between host species (Harbison et al., 2008). Phabine pigeons and doves are a well-defined group within Columbidae, and most species live in arid scrub or forest on the Australian continent. A reduced abundance of hippoboscid flies in arid rather than humid regions could explain the congruence of the wing louse phylogeny with hosts in Australia compared with the incongruence with hosts in the New World. It is also possible that hippoboscid flies rarely parasitize phabines in arid Australia. Although hippoboscid flies have been recorded from other Australian birds and from phabine hosts in the Philippines, we are unaware of any published records of hippoboscid flies associated with Australian phabines, while there are many records from New World pigeons and doves (Maa, 1963, 1969, 1980; Proctor and Jones, 2004; Toon and Hughes, 2008). This difference may be due to sampling effort, so it will be important to sample additional parasites on pigeons and doves in Australia.

In addition to ecological factors, geography may be an important factor governing diversification of phabine wing lice. In particular, there are several cases of clear allopatric codivergences of wing lice with their phabine hosts. For example, two pairs of sister species of

phabines (*Geophaps smithii* + *Geophaps scripta* and *Petrophassa rufipennis* + *Petrophassa albipennis*) are allopatric and appear to have speciated in response to biogeographic barriers. Their wing lice, which are host-specific, also cospeciated according to this allopatric divergence. Host specificity and the lack of dispersal to other host species in the same regions reinforces the pattern of cospeciation in this case. Biogeographic barriers are important for determining cophylogenetic structure, and can either promote congruence, as in phabine wing lice, or promote parasite diversification independent of host speciation. For example, the Andes mountains (Sweet and Johnson, 2016) and Amazonian rivers (Weckstein, 2004) can explain diversification patterns in various groups of bird lice, despite incongruent patterns between many of the host-parasite associations.

Host body size may also be an important factor in reinforcing cospeciation of wing lice with their hosts by limiting host-switching (Clayton et al., 2003, 2015). The body size of wing lice is closely correlated with that of their hosts, whereas the size of body lice is not (Johnson et al., 2005). Host preening defenses prevent wing lice from switching to hosts much larger or smaller than their usual host. In particular, wing lice must be of the appropriate size to fit between the feather barbs of the primary feathers to escape from host preening. This constraint may have been important in the codivergence of wing lice associated with the genus *Geopelia*. The three *Geopelia* doves represented in our data set have overlapping geographic distributions, yet vary in body size. The sister species *Geopelia humeralis* (110-160 g) and *Geopelia cuneata* (23-37 g) exhibit the greatest difference in size. *Geopelia placida* (36-60 g), which is sister to the other two species, is intermediate in body size (Gibbs et al., 2001). This variability in host size may reinforce phylogenetic congruence between *Geopelia* doves and their wing lice, as lice may not be able to switch to a related host even if the species co-occur.

While there is considerable evidence for cospeciation between phabines and their wing lice, there is also evidence for host-switching. Such events were likely facilitated by similarities in host size and by host geographic overlap, both of which can make it easier for lice to switch host species. For example, Jane recovered a host-switching event from *Phaps chalcoptera* to *Phaps histrionica*. Body sizes (*Phaps chalcoptera*: 230-390 g, *Phaps histrionica:* 260-320 g) and geographic ranges (*P. chalcoptera* widespread throughout Australia; *P. histrionica* primarily in the northern interior of Australia; Gibbs et al., 2001) of these two species overlap considerably.

While ecological and geographic factors may be important for generating congruence between Australian phabines and their wing lice, the same is not true for these hosts and their body lice, which do not appear to have a congruent evolutionary history. Body lice are not known to switch host effectively using phoresy on hippoboscid flies (Harbison et al., 2009), and are more often shared among host species that forage on the ground than among those that forage in the canopy (Johnson et al., 2011a). It may be that dispersal among hosts on the ground is the primary mode of host-switching for phabine body lice, particularly since their hosts are primarily terrestrial. Two species of body lice were found on two different host species (Figure 3.4), suggesting that these lice are able to disperse in ecological time among different host species.

Across species, there is much less of a match between the size of body lice and that of their hosts. For example, Jane recovered body lice from the small *Geopelia cuneata* switching to the considerably larger *Phaps chalcoptera.* This is consistent with previous research, which found that body louse size is not correlated with host size (Johnson et al., 2005). The average differences in host body size between pairs of hosts involved in host switches support this notion. Even when including inferred ancestral host sizes, hosts had a much higher absolute average difference in body size for body louse switches compared with wing louse switches. Unlike wing

lice, body lice burrow through the downy feathers to avoid preening, so their mechanism of escape is less tied to host body size. This may facilitate host-switching if there is a dispersal opportunity. A species of phabine can host multiple species of body lice that differ dramatically in size. As with wing lice, host distributional overlap may be an important factor for host-switching by body lice. Jane recovered several host switches along the lineage of body lice from *Geophaps smithii*, which has a relatively small distribution in Australia. All of the host switches, however, involve other species of phabines (*Petrophassa rufipennis*, *Ocyphaps lophotes*, *Geophaps plumifera* and *Geopelia humeralis*) whose ranges overlap that of *Geophaps smithii*. If body lice are indeed switching hosts primarily via ground contact, geographic proximity is necessary for dispersal to a new host species.

Previous studies of the wing and body lice of pigeons and doves in the New World have indicated that body lice exhibit more congruent cophylogenetic patterns with their hosts than do wing lice. However, our study revealed the opposite pattern, with wing lice of Australian phabine pigeons and doves exhibiting more phylogenetic congruence with their hosts than phabine body lice. This result highlights the importance of focusing cophylogenetic analyses on specific groups and biogeographic regions. A broader taxonomic and geographic focus, such as the entire pigeon and dove family (Columbidae) and its lice, can provide insight into general patterns in a group, but will mask narrower patterns if sampling is limited.

The drastic variation in cophylogenetic patterns between the New World dove and Australian phabine systems suggest regional differences in factors that shape these host-parasite interactions. For example, the lack of rampant host-switching in phabine wing lice may indicate that their hosts lack associated parasitic hippoboscid flies that wing lice of other species of pigeons and doves use as a means to switch hosts. This should be investigated with further

sampling. Other factors including climate, host body size and host distribution may also influence cophylogenetic patterns. Although phabines are only a moderately diverse group of Columbidae confined to a particular geographic region, comparisons of their phylogeny with those of their lice provide valuable insight into the processes of parasite diversification and host-parasite coevolution.

FIGURES

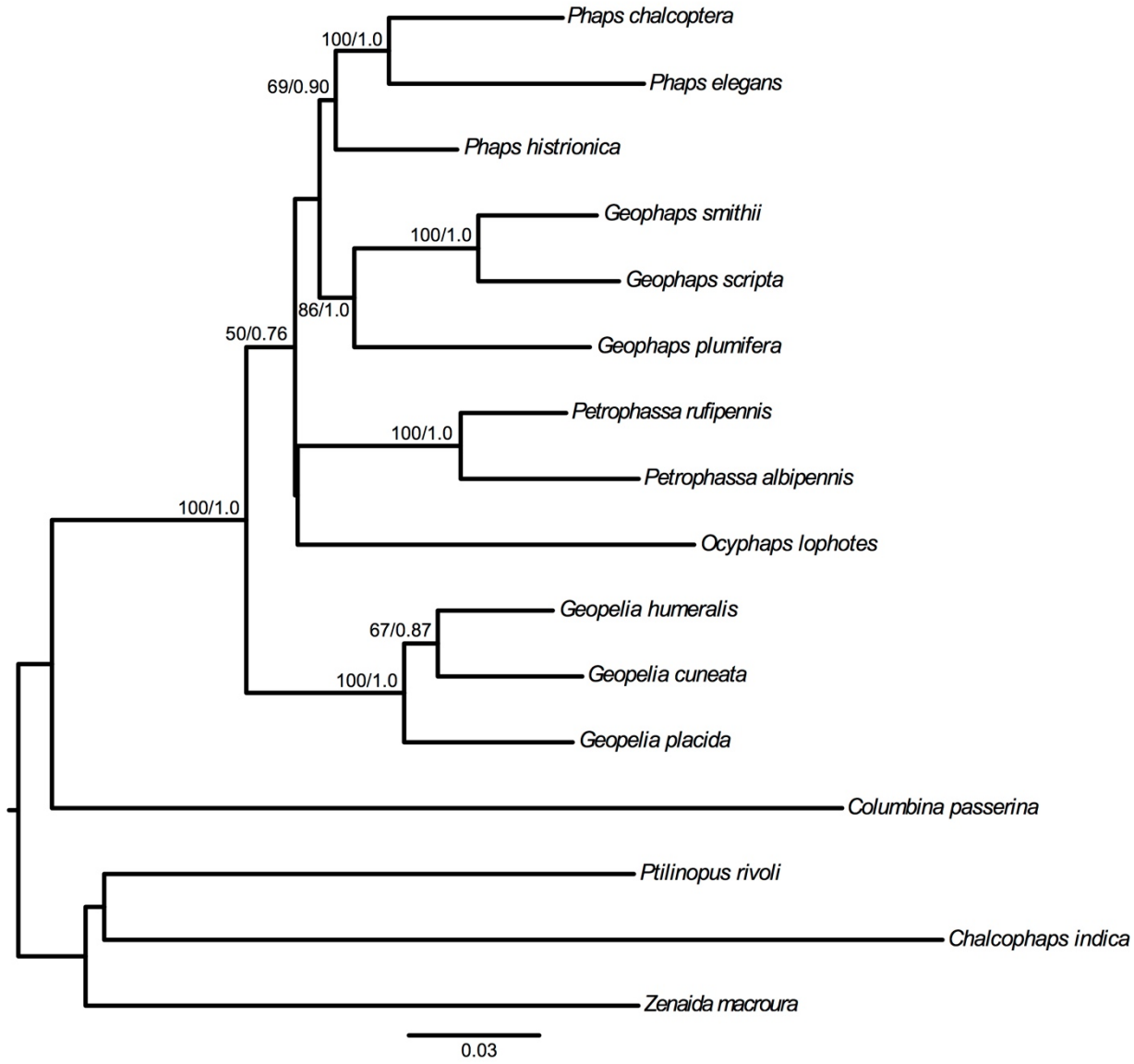


Figure 3.1. Best maximum likelihood (ML) phylogeny of phabine pigeons and doves. Values at nodes are bootstrap (BS) values from Garli and posterior probability (PP) values from MrBayes (BS/PP). Only values >50 BS/>0.50 PP are indicated. Branch lengths are nucleotide substitutions per site, as indicated by the scale bar.

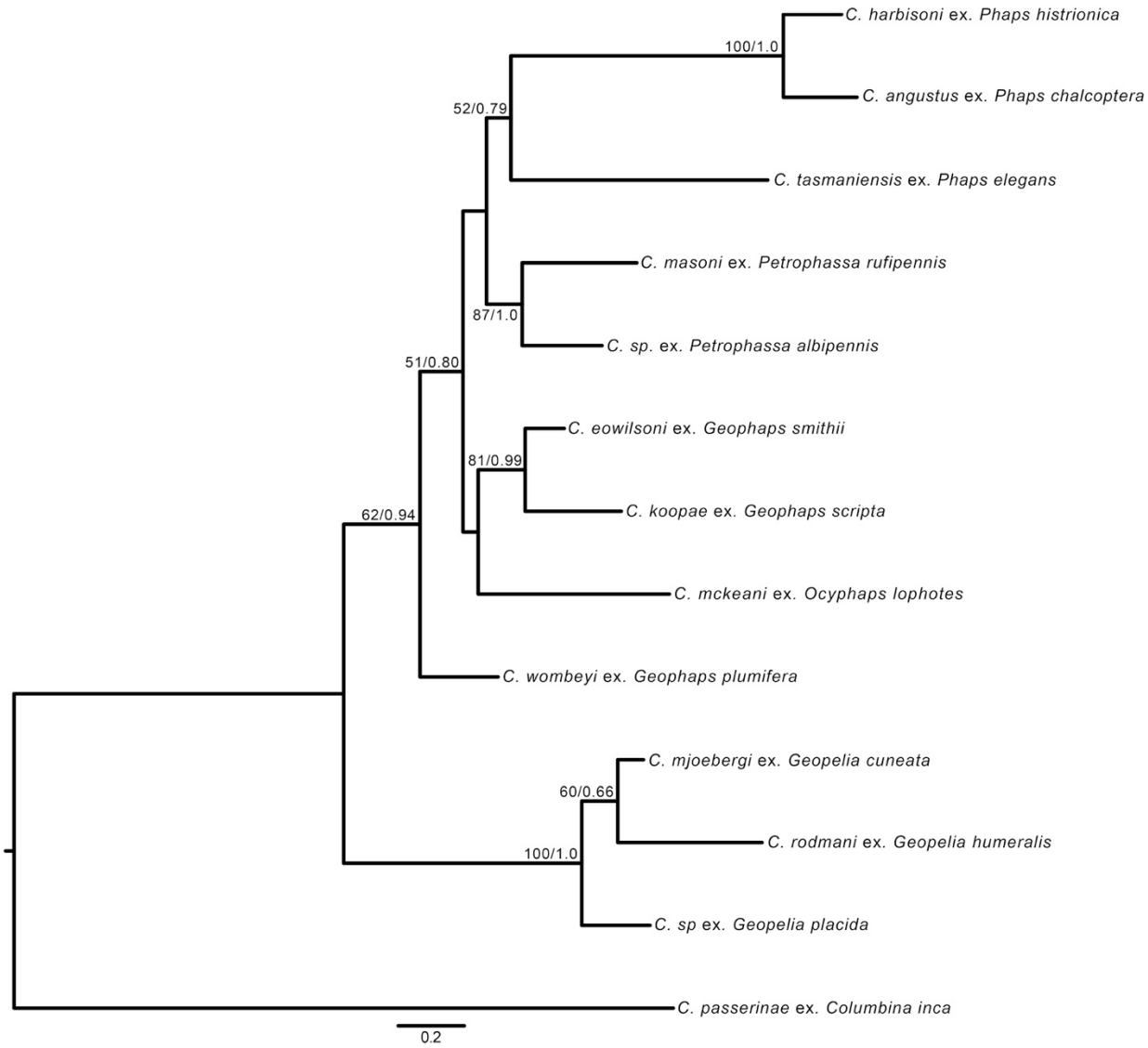


Figure 3.2. Best maximum likelihood (ML) phylogeny of wing lice from phabine pigeons and doves (*C.* = *Columbicola*). Values at nodes are bootstrap (BS) values from Garli and posterior probability (PP) values from MrBayes (BS/PP). Only values >50 BS/>0.50 PP are indicated. Branch lengths are nucleotide substitutions per site, as indicated by the scale bar.

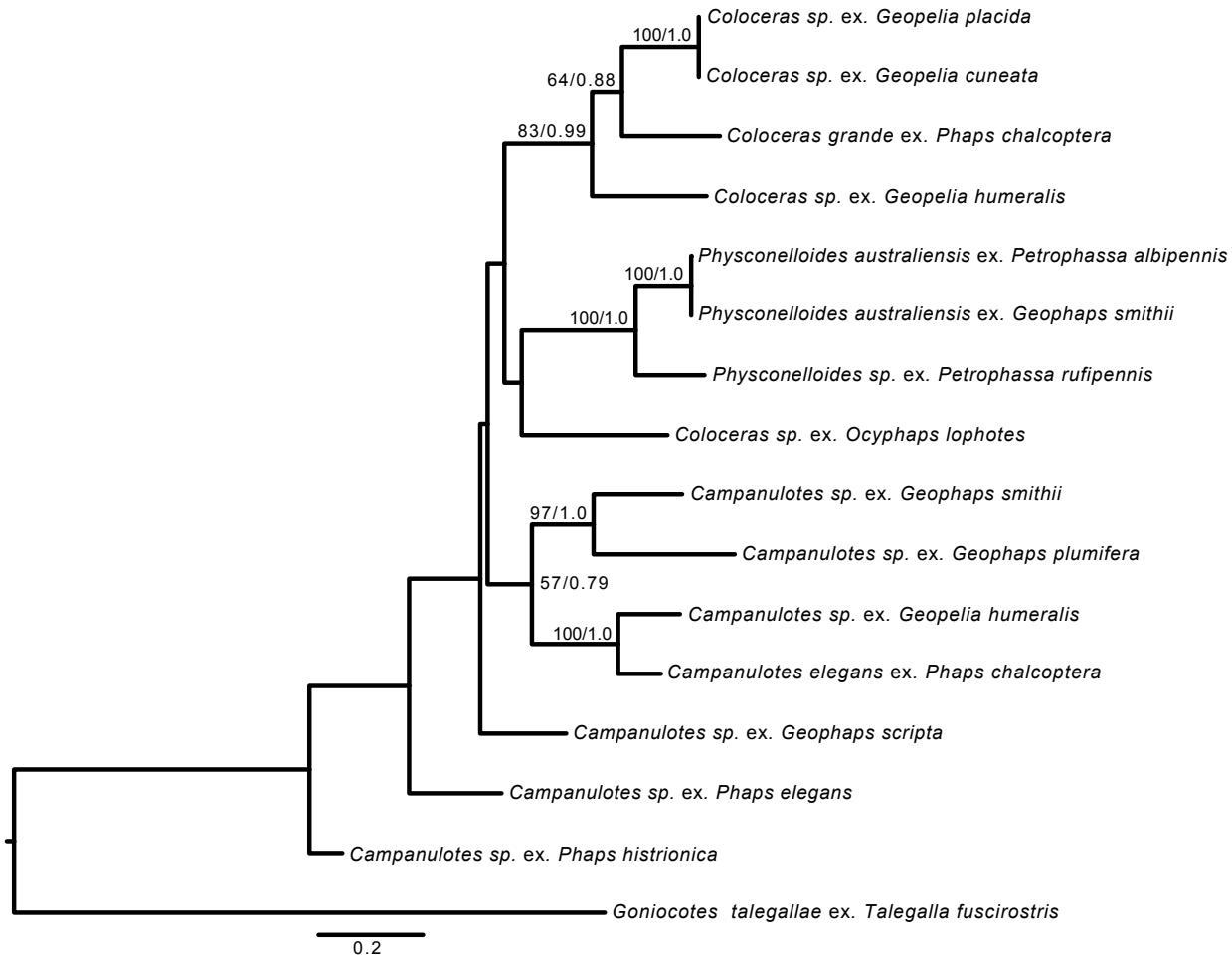


Figure 3.3. Best maximum likelihood (ML) phylogeny of body lice from phabine pigeons and doves. Values at nodes are bootstrap (BS) values from Garli and posterior probability (PP) values from MrBayes (BS/PP). Only values >50 BS/>0.50 PP are indicated. Branch lengths are nucleotide substitutions per site, as indicated by the scale bar.

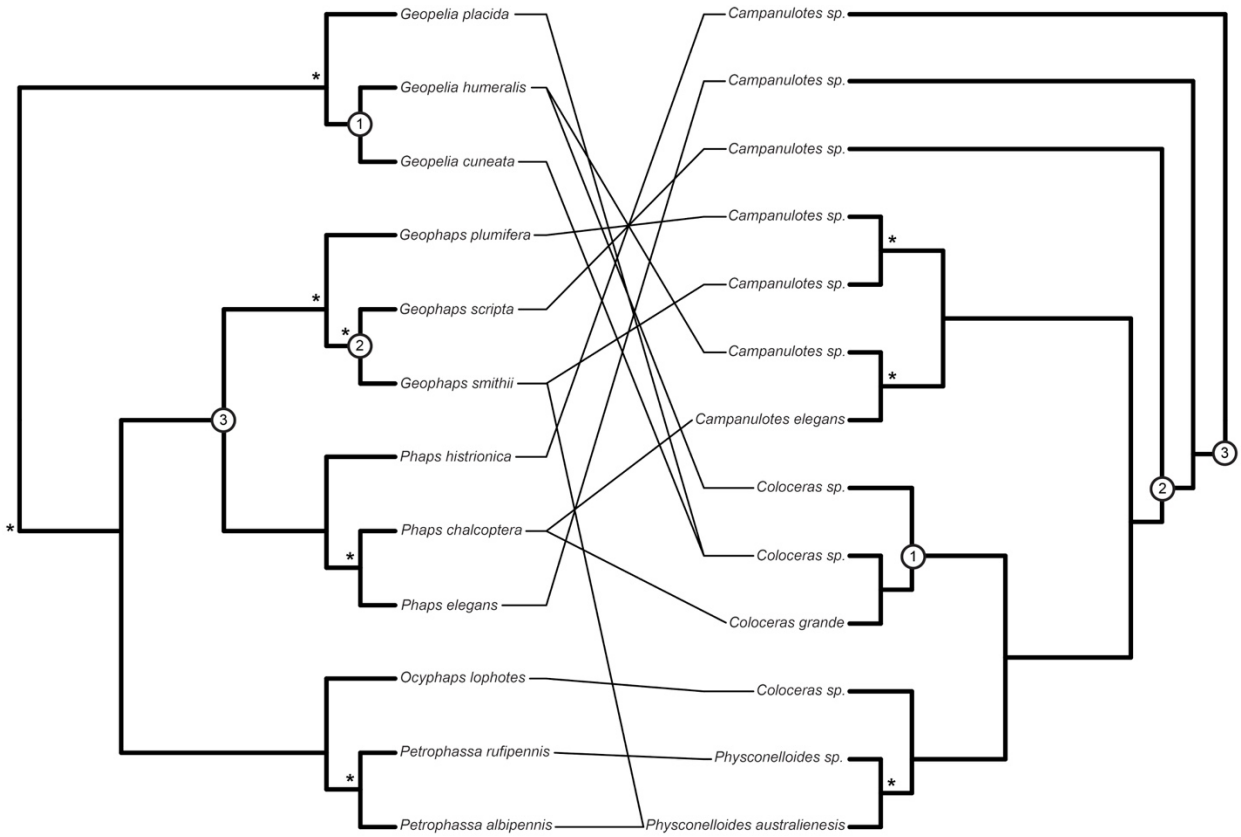


Figure 3.4. Tanglegram between phabine pigeons and doves (left) and their body lice (right). Topologies are the best maximum likelihood (ML) trees from Garli. Branches with >75 BS/0.95 PP support are indicated with asterisks (*). Circles over nodes indicate cospeciation events as recovered by Jane, with matching numbers indicating corresponding events in the host and parasite phylogenies.

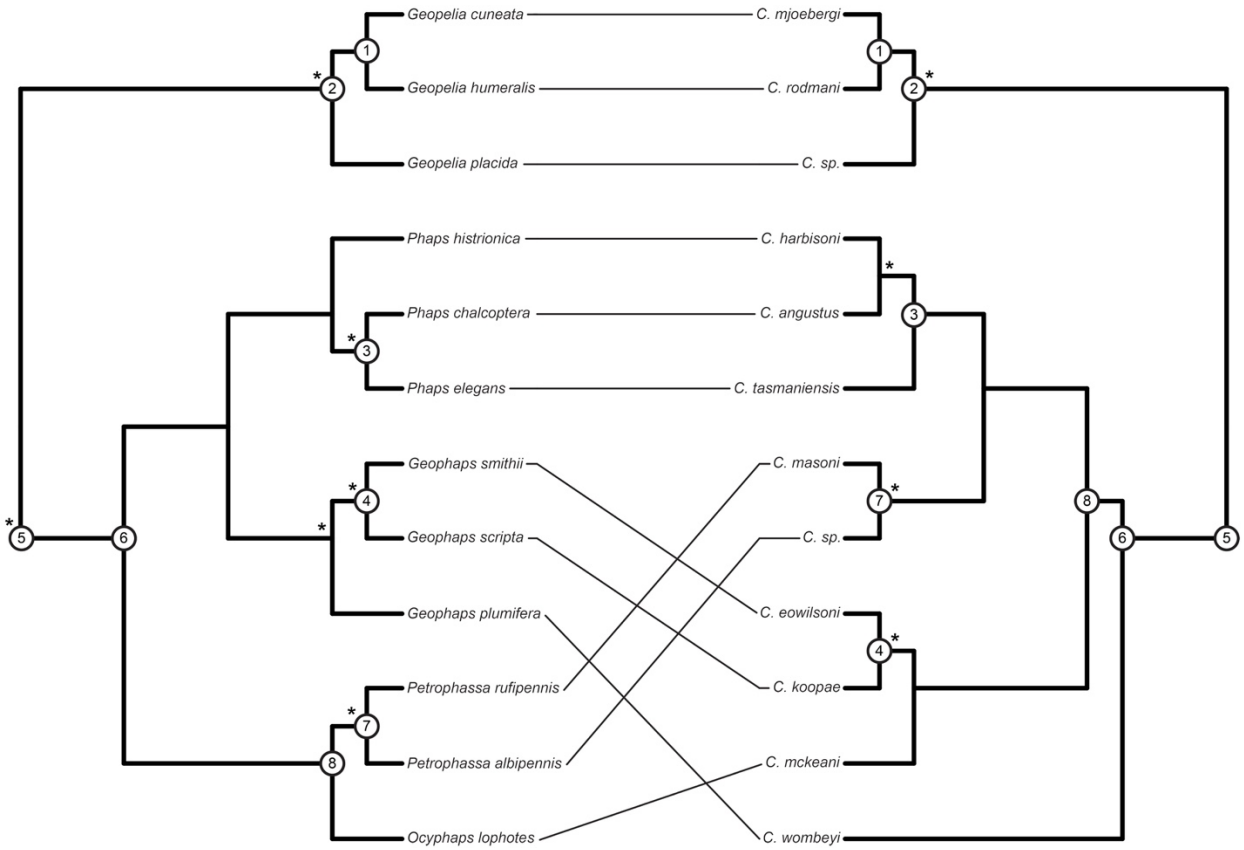


Figure 3.5. Tanglegram between phabine pigeons and doves (left) and their wing lice (right). Topologies are the best maximum likelihood (ML) trees from Garli. Branches with >75 BS/0.95 PP support are indicated with asterisks (*). Circles over nodes indicate cospeciation events as recovered by Jane, with matching numbers indicating corresponding events in the host and parasite phylogenies.

TABLES

Table 3.1. Best fitting substitution models for each partition as estimated by PartitionFinder.

	ML		MrBayes	
	Partition	**Model**	**Partition**	**Model**
Phabine doves				
	*Cox*1	GTR + I + Γ	*Cox*1	GTR + I + Γ
	ND2	GTR + I + Γ	ND2	GTR + I + Γ
	FIB7	HKY + I + Γ	FIB7	HKY + I + Γ
Wing lice				
	12S	GTR + Γ	12S	GTR + Γ
	*Cox*1	TrN + I + Γ	*Cox*1	GTR + I + Γ
	EF-1α	K80 + Γ	EF-1α	K80 + Γ
Body lice				
	16S	GTR + Γ	16S	GTR + Γ
	*Cox*1	TIM + I + Γ	*Cox*1	GTR + Γ
	EF-1α	HKY + I	EF-1α	HKY + I

ML, maximum likelihood; *Cox*1, mitochondrial cytochrome oxidase subunit 1; ND2, NADH dehydrogenase subunit 2; FIB7, nuclear beta-fibrinogen intron 7; EF-1α, elongation factor 1α; 12S, 12S rRNA; 16S, 16S rRNA.

Table 3.2. Summary statistics for each locus sequenced for phabine pigeons and doves, phabine wing lice, and phabine body lice.

	Locus	Length (bp)	Variable sites	Parsimony-informative sites
Phabine doves				
	*Cox*1	381	121	92
	ND2	1074	501	343
	FIB7	1172	248	56
Wing lice				
	12S	379	170	107
	*Cox*1	383	157	134
	EF-1α	360	84	39
Body lice				
	16S	553	221	144
	*Cox*1	383	142	117
	EF-1α	362	29	14

*Cox*1, mitochondrial cytochrome oxidase subunit 1; ND2, NADH dehydrogenase subunit 2; FIB7, nuclear beta-fibrinogen intron 7; EF-1α, elongation factor 1α ; 12S, 12S rRNA; 16S, 16S rRNA.

Table 3.3. ParaFit individual link test statistics and *P* values for phabine pigeons and doves and their body lice.

Host	**Body louse**	**ParaFitLink1 Stat**	**ParaFitLink1 *P* value**	**ParaFitLink2 Stat**	**ParaFitLink2 *P* value**
Geopelia cuneata	*Coloceras* sp.	0.0089	0.0185	0.0057	0.0186
Geopelia humeralis	*Campanulotes* sp.	-0.0012	0.9443	-0.0008	0.9448
Geopelia humeralis	*Coloceras* sp.	0.0074	0.0397	0.0047	0.0396
Geopelia placida	*Coloceras* sp.	0.0086	0.0207	0.0055	0.0205
Geophaps plumifera	*Coloceras* sp.	0.0048	0.2435	0.0031	0.2398
Geophaps scripta	*Campanulotes* sp.	0.0009	0.2649	0.0006	0.2615
Geophaps smithii	*Physconelloides australiensis*	<0.0001	0.7487	<0.0001	0.7487
Geophaps smithii	*Campanulotes* sp.	0.0030	0.2890	0.0019	0.2869
Ocyphaps lophotes	*Coloceras* sp.	0.0039	0.0398	0.0025	0.0381
Petrophassa albipennis	*Physconelloides australiensis*	0.0067	0.0446	0.0043	0.0442
Petrophassa rufipennis	*Physconelloides* sp.	0.0059	0.1090	0.0038	0.1063
Phaps chalcoptera	*Campanulotes elegans*	0.0004	0.6990	0.0003	0.6983
Phaps chalcoptera	*Coloceras grande*	-0.0019	0.9486	-0.0012	0.9493
Phaps elegans	*Campanulotes* sp.	0.0036	0.1527	0.0023	0.1503
Phaps histrionica	*Campanulotes* sp.	0.0015	0.6804	0.0010	0.6785

Table 3.4. Results from the Jane event-based cophylogenetic reconstruction between phabine pigeons and doves and their lice.

	Cospeciations	Duplications	Host Switches	Losses	Failures to Diverge	Total Cost
Body Lice	3	1	8	4	2	23
	3	0	9	3	2	23
Wing Lice	8	0	3	2	0	8

Table 3.5. ParaFit individual link test statistics and *P* values for phabine pigeons and doves and their wing lice (*Columbicola)*.

Host	**Wing louse**	**ParaFitLink1 Stat**	**ParaFitLink1 *P* value**	**ParaFitLink2 Stat**	**ParaFitLink2 *P* value**
Geopelia cuneata	*C. mjoebergi*	0.1074	0.0044[a]	0.0012	0.0044[a]
Geopelia humeralis	*C. rodmani*	0.1253	0.0095[a]	0.0014	0.0094[a]
Geopelia placida	*C.* sp.	0.0995	0.0096[a]	0.0011	0.0096[a]
Geophaps plumifera	*C. wombeyi*	0.0020	0.4710	<0.0001	0.4700
Geophaps scripta	*C. koopae*	0.0178	0.0444	0.0002	0.0442
Geophaps smithii	*C. eowilsoni*	0.0116	0.0551	0.0001	0.0544
Ocyphaps lophotes	*C. mckeani*	0.0232	0.0514	0.0003	0.0511
Petrophassa albipennis	*C.* sp.	0.0156	0.0453	0.0002	0.0447
Petrophassa rufipennis	*C. masoni*	0.0155	0.1072	0.0002	0.1064
Phaps chalcoptera	*C. angustus*	0.0388	0.1699	0.0004	0.1691
Phaps elegans	*C. tasmaniensis*	0.0352	0.0548	0.0004	0.0540
Phaps histrionica	*C. harbisoni*	0.0195	0.4623	0.0002	0.4611

[a]Statistically significant after Benjamini-Hochberg correction ($\alpha = 0.05$)

LITERATURE CITED

Banks J.C., Paterson A.M. 2005. Multi-host parasite species in cophylogenetic studies. *International Journal of Parasitology* 35: 741-746.

Belzile C., Gosselin M. 2015. Free-living stage of the unicellular algae *Coccomyxa* sp. parasite of the blue mussel (*Mytilus edulis*): Low-light adaptation, capacity for growth at a very wide salinity range and tolerance to low pH. *Journal of Invertebrate Pathology* 132: 201–207. http://doi.org/10.1016/j.jip.2015.10.006

Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistics Society Series B Statistical Methodology* 57: 289–300.

Bruyndonckx N., Dubey S., Ruedi M., Christe P. 2009. Molecular cophylogenetic relationships between European bats and their ectoparasitic mites (Acari, Spinturnicidae). *Molecular Phylogenetics and Evolution* 51: 227–237. http://doi.org/10.1016/j.ympev.2009.02.005

Bueter C., Weckstein J., Johnson K.P., Bates J.M., Gordon C.E. 2009. Comparative phylogenetic histories of two louse genera found on Catharus thrushes and other birds. *Journal of Parasitology* 95: 295–307. http://doi.org/10.1645/GE-1642.1

Bush S.E., Malenke J.R. 2008. Host defence mediates interspecific competition in ectoparasites. *Journal of Animal Ecology* 77: 558–64. http://doi.org/10.1111/j.1365-2656.2007.01353.x

Clayton D.H., Bush S.E., Goates B.M., Johnson K.P. 2003. Host defense reinforces host-parasite cospeciation. *Proceedings of the National Academy of Science*. 100: 15694–15699.

Clayton D.H., Drown D.M. 2001. Critical evaluation of five methods for quantifying chewing lice (Insecta: Phthiraptera). *Journal of Parasitology* 87: 1291-1300.

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

Clayton D.H., Koop J.A.H., Harbison C.W., Moyer B.R., Bush S.E. 2010. How birds combat ectoparasites. *Open Ornithology Journal* 3: 41–71. http://doi.org/10.2174/1874453201003010041

Clayton D.H., Moyer B.R., Bush S.E., Jones T.G., Gardiner D.W., Rhodes B.B., Goller F. 2005. Adaptive significance of avian beak morphology for ectoparasite control. *Proceedings of the Royal Society* 272(1565): 811–817. http://doi.org/10.1098/rspb.2004.3036

Clayton D.H. 1991. Coevolution of avian grooming and ectoparasite avoidance. In: J. E. Loye and M. Zuk, eds. Bird-parasite interactions: ecology, evolution, and behavior. Oxford University Press, Oxford, UK, pp. 258–289.

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

Conow C., Fielder D., Ovadia Y., Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms in Molecular Biology* 5: 16. http://doi.org/10.1186/1748-7188-5-16

Cruickshank R.H., Johnson K.P., Smith V.S., Adams R.J., Clayton D.H., Page R.D. 2001. Phylogenetic analysis of partial sequences of elongation factor 1alpha identifies major groups of lice (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 19: 202–215. http://doi.org/10.1006/mpev.2001.0928

Eichler W.D. 1948. Some rules in ectoparasitism. Annalual Magazine of Natural History 1: 588–598.

Fahrenholz H. 1913. Ectoparasiten und abstammungslehre. *Zoologischer Anzeiger* 41: 371–374.

Gandon S., Poulin R. 2004. Evolution of multihost parasites. *Evolution* 58: 455–469. http://doi.org/10.1554/03-390

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

Goodwin D. 1983. Pigeons and doves of the world (3rd ed.). Cornell University Press, Ithaca, New York, USA.

Hafner M.S., Page R.D. 1995. Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philosophical Transactions of the Royal Society London, B, Biological Sciences* 349(1327): 77–83. http://doi.org/10.1098/rstb.1995.0093

Hafner M.S., Sudman P.D., Villablanca F.X., Spradling T.A., Demastes J.W., Nadler S.A. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science.* 265: 1087–1090.

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

Harbison C.W., Clayton D.H. 2011. Community interactions govern host-switching with implications for host-parasite coevolutionary history. *Proceedings of the National Academy of Sciences* 108: 9525–9529. http://doi.org/10.1073/pnas.1102129108

Harbison C.W., Jacobsen M.V., Clayton D.H. 2009. A hitchhiker's guide to parasite transmission: The phoretic behaviour of feather lice. *International Journal for Parasitology* 39: 569–575. http://doi.org/10.1016/j.ijpara.2008.09.014

Johnson K.P., Clayton D.H. 2000a. Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). *Molecular Phylogenetics and Evolution* 14: 141–151.

Johnson K.P., Clayton D.H. 2000b. A Molecular phylogeny of the dove genus *Zenaida*: mitochondrial and nuclear DNA sequences. *The Condor* 102: 864–870.

Johnson K.P., Reed D.L., Hammond Parker S.L., Kim D., Clayton D.H. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for *Columbicola* (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 45: 506–518. http://doi.org/10.1016/j.ympev.2007.07.005

Johnson K.P., Shreve S.M., Smith V.S. 2012. Repeated adaptive divergence of microhabitat specialization in avian feather lice. BMC Biology 10: 52. http://doi.org/10.1186/1741-7007-10-52

Johnson K.P., Weckstein J.D., Bush S.E., Clayton D.H. 2011a. The evolution of host specificity in dove body lice. *Parasitology* 138: 1730–1736. http://doi.org/10.1017/S0031182010001770

Johnson K.P., Williams B.L., Drown D.M., Adams R.J., Clayton D.H. 2002. The population genetics of host specificity: genetic differentiation in dove lice (Insecta: Phthiraptera). *Molecular Ecology* 11: 25–38.

Johnson K.P., Bush S.E., Clayton D.H. 2005. Correlated evolution of host and parasite body size: tests of Harrison's Rule using birds and lice. *Evolution*. 59: 1744–1753.

Johnson K.P., Weckstein J.D., Meyer M.J., Clayton D.H. 2011b. There and back again: switching between host orders by avian body lice (Ischnocera: Goniodidae). *Biological Journal of the Linnean Society* 102: 614–625. http://doi.org/10.1111/j.1095-8312.2010.01612.x

Johnson K., Clayton D. 2004. Untangling coevolutionary history. *Systematic Biology* 53: 92–94. http://doi.org/10.1080/10635150490264824

Katoh K., Misawa K., Kuma K., and Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066. http://doi.org/10.1093/nar/gkf436

Keirans J.E., 1975. A review of the phoretic relationship between Mallophaga (Phthiraptera: Insecta) and Hippoboscidae (Diptera: Insecta). *Journal of Medical Entomology* 12: 71–76.

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substituion 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.

Maa T.C. 1963. Genera and species of Hippoboscidae (Diptera): Types, synonymy, habitats and natural groupings. *Pacific Insects* 6: 1–186.

Maa T.C. 1969. A revised checklist and consise host index of Hippoboscidae (Diptera). *Pacific Insects* 20: 261–299.

Maa T.C. 1980. A preliminary revision of the genus *Myopthiria* Rondani (Dipera: Hippoboscidae). *Review Suisse Zoology* 87: 835-861.

Nelson B.C., Murray M.D. 1971. The distribution of Mallophaga on the domestic pigeon (*Columba livia*). *Internation Journal for Parasitology* 1: 21–29. http://doi.org/10.1016/0020-7519(71)90042-7

Page R.D.M. 1994. Parallel phylogenies: reconstructing the history of host-parasite assemblages. Cladistics 10: 155–173. http://doi.org/10.1111/j.1096-0031.1994.tb00170.x

Page R.D.M., Charleston M.A. 1998. Trees within trees: phylogeny and historical associations. *Trends in Ecology and Evolution* 13: 356–359. http://doi.org/10.1016/S0169-5347(98)01438-4

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290. http://doi.org/10.1093/bioinformatics/btg412

Paterson A.M., Wallis G.P., Wallis L.J., Gray R.D. 2000. Seabird and louse coevolution: complex histories revealed by 12S rRNA sequences and reconciliation analyses. Systematic Biology 49: 383–399.

Pereira S.L., Johnson K.P., Clayton D.H., Baker A.J. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleocene. *Systematic Biology* 56: 656–672. http://doi.org/10.1080/10635150701549672

Poulin R. 2011. Evolutionary ecology of parasites (2nd ed.). Princeton University Press, Princeton, USA.

Poulin R., Morand S. 2000. The diversity of parasites. *Quarterly Review of Biology* 75: 277–293.

Poulin R., Randhawa H.S. 2013. Evolution of parasitism along convergent lines: from ecology to genomics. *Parasitology* 142(S1): S6–S15. http://doi.org/10.1017/S0031182013001674

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

Proctor H.C., Jones D.N. 2004. geographical structuring of feather mite assemblages from the Australian brush-turkey (Aves: Megapodiidae). *Journal of Parasitology* 90: 60–66. http://doi.org/10.1645/GE-57R

Proctor H.C. 2003. Feather mites (Acari: Astigmata): ecology, behavior, and evolution. *Annual Review of Entomology* 48: 185–209. http://doi.org/10.1146/annurev.ento.48.091801.112725

Puillandre N., Lambert A., Brouillet S., Achaz G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21: 1864–1877. http://doi.org/10.1111/j.1365-294X.2011.05239.x

Rohde K. 1979. A critical evaluation of intrinsic and extrinsic factors responsible for niche restriction in parasites. American Naturalist 114: 648–671.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. http://doi.org/10.1093/bioinformatics/btl446

Stefka J., Hoeck P.E., Keller L.F., Smith V.S. 2011. A hitchhikers guide to the Galapagos: co-phylogeography of Galapagos mockingbirds and their parasites. BMC Evolutionary Biology 11: 284. http://doi.org/10.1186/1471-2148-11-284

Sugiura N. 1978. Further analysts of the data by Akaike's Information Criterion and the finite corrections. Communications in Statistcs – Theory and Methods 7: 13–26. http://doi.org/10.1080/03610927808827599

Sweet A.D., Boyd B.M., Johnson K.P. 2016. Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. *Biological Journal of the Linnean Society* 118: 813-828. http://doi.org/10.1111/bij.12771

Sweet A.D., Johnson K.P. 2016. Cophylogenetic analysis of New World ground-doves (Aves: Columbidae) and their parasitic wing lice (Insecta: Phthiraptera: *Columbicola*). *Molecular Phylogenetics and Evolution* 103: 122-132.

Toon A., Hughes J.M. 2008. Are lice good proxies for host history? A comparative analysis of the Australian magpie, *Gymnorhina tibicen*, and two species of feather louse. *Heredity* 101: 127–135. http://doi.org/10.1038/hdy.2008.37

Weckstein J.D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology* 53: 154–164. http://doi.org/10.1080/10635150490265085

Whiteman N.K., Kimball R.T., Parker P.G. 2007. Co-phylogeography and comparative population genetics of the threatened Galápagos hawk and three ectoparasite species: Ecology shapes population histories within parasite communities. *Molecular Ecology* 16: 4759–4773. http://doi.org/10.1111/j.1365-294X.2007.03512.x

CHAPTER 4: PATTERNS OF DIVERSIFICATION IN SMALL NEW WORLD GROUND DOVES ARE CONSISTENT WITH MAJOR GEOLOGIC EVENTS[3]

INTRODUCTION

Since breaking off from the Gondwanan supercontinent 140-160 mya (Jokat et al., 2003; Upchurch, 2008), South America has undergone several major geologic events that have helped define the diversity of the continent's biota. Speciation and diversification patterns in New World birds have historically been the foci of many studies (Wallace, 1889; Chapman, 1917; Chesser, 2004), which is largely attributable to the large concentration of species in the area (~35% of all bird species are endemic to the Neotropics). Because of their diversity and geographic range, birds are excellent organisms for studying how major New World geologic events influenced regional diversification patterns. Two events that have particularly impacted New World avian diversity are Andean uplift and the Panamanian Land Bridge formation.

Andean uplift had a dramatic impact on South American geology, climate, and biogeography. Uplift in the south and central Andes began >60 mya and continued a south-north elevation increase. By 25 mya the Western Cordillera of the central Andes were at 50% current elevation, but underwent a drastic increase in elevation (2000-3500

[3] Reprinted, with permission, from: Sweet, A.D. and K.P. Johnson. 2015. Patterns of diversification in small New World ground doves are consistent with major geologic events. *The Auk.* 132: 300-312. https://doi.org/10.1642/AUK-14-193.1.

m) between 6-10 mya (Gregory-Wodzicki, 2000; Garzione, 2008), although Garzione et al. (2014) presented evidence for a slightly earlier rapid uplift (13-16 mya and 9-13 mya). The Eastern Cordillera of the Northern Andes uplift occurred more recently, with evidence of a rapid elevation increase between 2.5-7 mya (Hoorn et al., 1995; Hoorn et al., 2010). These gradual and rapid uplift events coupled with resulting changes in weather, climate, and ecology most likely had an incredible impact on avian diversification patterns and rates throughout the continent (Fjeldsa, 1994). Many attribute Andean uplift as a key factor contributing to the patterns of speciation that underlay the current avifauna diversity of the region (Vuilleumier, 1969; Cracraft and Prum, 1988; Bates et al., 1998; Cheviron et al., 2005; Fjeldsa and Rahbek, 2006; Fjeldsa and Irestedt, 2009; Quinetero et al., 2012).

The dry Puna grasslands of the Central Andes' Altiplano plateau are an example of novel habitat formed as a direct result of Andean uplift. As the Andes gradually increased in elevation, the Altiplano plateau became separated from the surrounding lowlands and was probably isolated around 15 mya. The plateau was at 50% current elevation 10 mya, followed by a recent rapid uplift to its current elevation of 3200-3700 m (Gregory-Wodzicki, 2000; Hoke and Garzione, 2008; Vandervoort et al., 1995). Between 10-15 mya, climate changes in the Central Andes resulted in a drier and cooler environment, which directly altered the biota of the Puna grasslands (Ehlers and Poulsen, 2009; Simpson et al., 1975). High elevation and extreme climatic conditions created a unique, isolated environment to which specific flora and fauna adapted (Szumik et al., 2012). Avian taxa in the region would have become particularly isolated from taxa in surrounding areas, perhaps specifically from the southern open lowlands (Fjeldsa et al.,

2012). This isolation probably contributed to the currently low species overlap between the Puna-Altiplano and adjacent lowland and cloud forest habitats (Lloyd and Marsden, 2008; Lloyd et al., 2010).

A second major geologic event greatly impacting New World biogeography was the Panamanian Land Bridge formation, which occurred around 2.5-3.2 mya (Keigwin, 1978; Coates and Obando, 1996; Coates et al., 1992; Wegner et al., 2010; Leigh et al., 2013), although an earlier formation has been proposed more recently (Farris et al., 2011; Montes et al., 2012; and Bacon et al., 2013). Before this terrestrial connection between North and South America, faunas endemic to these two continents were separated by water. With the formation of the land bridge, flora and fauna could freely move between continents, which some have dubbed the "Great American Interchange" (Simpson, 1950; 1980; Stehli and Webb, 1985). Mammalian fossil records from both continents indicate dispersal events close to the time of land bridge formation, with species interchanging from both continents (Marshall, 1988; Stehli and Webb, 1985). Due to their flight ability, avian dispersal between continents was sometimes suggested to be uninhibited by the pre-land bridge water barrier (Voelker, 1999; Lomolino et al., 2006), and the sparse avian fossil record has made it difficult to prove otherwise (Vuilleumier, 1985). However, several historical biogeographic reconstructions of Neotropical and Nearctic birds based on molecular data have indicated that many species were hindered from dispersing between continents due to the water barrier (Barker, 2007; Burns and Racicot, 2009; Sedano and Burns, 2010; Johnson and Weckstein, 2011; Pulgarin-R et al., 2013). Instead the timing of dispersal events appears similar to that of mammals, soon after the land bridge formation (Smith and Klicka, 2010). Also, there are more instances of North

American species successfully dispersing into South America rather than vice versa, perhaps because a transition from Nearctic to Neotropical climate is easier than the reverse (Smith and Klicka, 2010).

To understand the impact of these two major New World geologic events on avian evolutionary history, an ideal study group would be one having representatives in both North and South America, as well as lowland and Andean zones. One such group is the small New World ground dove clade (Aves: Columbidae). This group of 17 species and 45 subspecies contains the genera *Metriopelia*, *Claravis, Uropelia,* and *Columbina.* Two species have often been placed within a fifth genus, *Scardafella* (e.g. Goodwin, 1983; Gibbs et al., 2001; del Hoyo et al., 1997), but are otherwise included within *Columbina* (American Ornithologists' Union, 1998). Two of these 17 species have not been detected recently (*Columbina cyanopis* and *Claravis godefrida*) and may be extinct. The overall geographic ranges of this group covers the southern United States, through Central America (including the Caribbean Islands), and throughout most of South America (Figure 3). Several phylogenies have indicated this group forms a monophyletic clade, although the placement of the clade within Columbidae is still unclear. Some phylogenies placed the small New World ground doves as sister to the rest of the family (Johnson and Clayton, 2000; Johnson, 2004; Johnson et al., 2010), but other work has placed the group as nested within the family (Pereira et al., 2007; Shapiro et al., 2002; Gibb and Penny, 2010). Relationships among the species within this clade, however, have yet to be addressed in detail. Furthermore, given the widespread distribution of this clade throughout Central and South America, this group could also provide great insight into how past geologic events in this region has influenced current biological patterns of

distribution and speciation. Here we reconstruct the phylogeny of the small New World ground dove clade based on several gene regions. We then use this phylogeny in a molecular dating analysis to evaluate the effects of the Andean uplift and appearance of the Panamanian land bridge on the timing of divergence in this group. If Andean uplift events did not have a major effect on the speciation patterns of these clades, then our estimated timing of relevant speciation events are not expected to correlate with the timing of these events. In particular, we would not expect high-Andean species nor sister taxa currently separated by the Andean range to have divergence estimates correlated to Andean uplift events. Similarly, if the Isthmus formation did not have a significant effect on the speciation patterns of small New World ground doves, then we expect our results to indicate dispersal events not consistent with the timing of the land bridge formation. More specifically, we would expect the recovery of dispersal events occurring before the estimated age of the land bridge.

MATERIALS AND METHODS

DNA sequencing

We extracted DNA using a Qiagen Blood and Tissue Kit (Qiagen, Velencia, CA, cat. # 69506) from tissues and feather samples of 15 extant species of the small New World ground dove clade, including 16 subspecies from 10 species, for a total of 26 ingroup samples (Table 4.1). The remaining 10 samples were either monotypic species or duplicate samples of a subspecies. Using PCR, we amplified portions of four mitochondrial loci – cytochrome (Cytb), NADH dehydrogenase subunit 2 (ND2), cytochrome oxidase subunit 1 (CO1), and ATP synthase 8 (ATP8) – and one nuclear

locus, beta-fibrinogen intron (FIB7), which have been used successfully in previous studies of the phylogeny of Columbiformes (Johnson and Clayton, 2000; Johnson et al., 2001; Periera et al., 2007). For Cytb we used the primers L14841 and H4a to amplify the gene and primers with L15517 and H15299 as internal for sequencing (Kocher et al., 1989; Harshman 1996). To amplify ND2 we used the primers L5215 and H6313 (Johnson and Sorenson, 1998) and L5758s and H5766s (Price et al., 2004) internally for sequencing. To amplify and sequence CO1 we used the primers L6625 and H7005 (Hafner et al., 1994). To amplify and sequence ATP8 we used the primers CO2GQL and A6MNH (Lovette et al., 1998). For the nuclear intron FIB7 we amplified using the primers FIBB17U and FIBB17L (Prychitko and Moore, 1997) with the internal sequencing primers FIBDOVEF and FIBDOVER (Johnson and Clayton, 2000). We amplified selected loci using PCR on a PTC 100 Thermal Cycler according to previously used protocols for each locus (Johnson, 2004; Pereira et al., 2007). Resulting amplified products were purified with a Qiagen PCR Purification kit (cat. # 28106) and sequenced using ABI Prism BigDye Terminators and Sanger DNA sequencing on an AB 3730xl DNA Analyzer (University of Illinois Roy J. Carver Biotechnology Center, Champaign, IL). We reconciled resulting complementary chromatograms and trimmed primer sequences by eye using Sequencher v. 5.0.1 (Gene Codes, Ann Arbor, MI). For outgroups, we selected representatives outside the New World ground dove clade from two genera from each of five monophyletic clades within Columbidae identified in Pereira et al. (2007), using previously published sequences (Table 4.1). In one case we were unable to amplify and sequence a gene for one extract, so this was coded as missing data.

Phylogenetic analysis

We aligned the edited sequences for each of the five loci using the default gap opening and gap extend parameters in MUSCLE (Edgar, 2004) and checked each alignment by eye using SeaView v4 (Gouy et al., 2010). To check for major discordance between individual gene trees, we created neighbor-joining and majority-rule maximum parsimony trees (100 random sampling replicates, TBR branch swapping, 100 bootstrap replicates) for each gene alignment in PAUP* (Swofford, 2003). Since these gene trees did not have any nodes that strongly conflicted in bootstrap support (>75%), we concatenated the data using SeaView. We also computed the pairwise distance values for the mitochondrial data using PAUP* (Supplementary Table 4.1).

Using the concatenated data set, we performed Bayesian and Maximum Likelihood (ML) analysis using mixed models. We determined the appropriate substitution model for each gene partition by calculating the AIC (Akaike Information Criterion) values for 88 different models in jModelTest2 (Akaike 1974, Darriba et al. 2012). Based on the AIC results, we applied a GTR + I + G model to the mitochondrial loci (CO1, Cytb, ND2, ATP8), and a GTR + G model to the nuclear locus (FIB7).

We preformed our ML analysis on the concatenated dataset using Garli v2.0 (Zwickl, 2006), applying the appropriate models for each gene partition and running 500 bootstrap replicates. We obtained a 50% majority-rule consensus tree from the bootstrap reps using SumTrees (Sukumaran and Holder, 2008) and edited the resulting tree in FigTree v1.4 (Rambault, 2012). We also created a concatenated dataset of the four mitochondrial loci in order to compare the resulting tree with the tree for the full dataset.

We generated the mitochondrial tree in Garli 2.0 using the same methods as in the full dataset.

For the Bayesian analysis, we used MrBayes v3.2 (Ronquist and Huelsenbeck, 2003). As with the ML analysis, we used a mixed model analysis and assigned appropriate models to the gene partitions based on the AIC results. We ran 4 runs with 4 chains for 20 million generations under MCMC sampling every 1000 trees, and viewed the trace files in Tracer v1.4 to ensure chain mixture and stationarity of the MCMC. (Rambaut and Drummond, 2007). Based on the trace files, we discarded the first 2 million generations (10%) as burnin and edited the resulting 50% majority-rule consensus tree in FigTree v1.4 (Rambaut, 2012).

Divergence time estimation

In order to estimate divergence times, we created a chronogram using BEAST v1.7.5 (Drummond et al., 2012). We partitioned the data into mitochondrial and nuclear loci, and applied a strict molecular clock estimate of 1.96 ± 0.1%/My divergence between two taxa (i.e. 0.0098 ± 0.0005 substitutions/site/lineage/My) under a normal distribution for the mitochondrial partition and a Yule speciation process model. This estimate is based on Weir and Schluter (2008), who showed that a molecular clock of 2%/My accumulated pairwise divergence between lineages could be used in dating avian phylogenies. Several avian phylogenetic studies have used this estimate to infer seemingly accurate divergence estimates (Milá et al., 2009; Qu et al., 2010; Sedano and Burns, 2010). In particular, Weir and Schluter (2008) determined an average rate of pairwise divergence between two taxa at 1.96 ± 0.10 %/My for Columbiformes. We ran our MCMC for 20 million generations in BEAST, sampling every 1000 trees, and

discarding the first 2 million (10%) generations as burnin based on the trace plot in Tracer.

Biogeographic analysis

Since one of our main historical biogeographic questions centers on the formation of the Panamanian land bridge, we reconstructed ancestral geographic ranges with a focus on whether particular species in the clade are currently found in North or South America. We primarily used both parsimony reconstruction and likelihood character mapping over the BEAST tree since our focus was on the dispersal events between North and South America after the formation of the Panamanian land bridge, rather than vicariance events. In this biogeographic scenario, methods such as dispersal-vicariance analysis (DIVA) that assume vicariance as the null model are inappropriate. Such models are biased toward vicariance events, and could therefore incorrectly attribute a speciation event to vicariance rather than to dispersal (see Johnson and Weckstein, 2011 and Bess et al., 2014 for further rationale). In this case, North America and South America came into contact rather than separating from each other, so scenarios that posit vicariance are not biogeographically plausible. For the parsimony analysis, we coded each species as one of two character states: having a primarily North American range or a primarily South American range. Species that are widespread in both continents were given a polymorphic character state. For the likelihood analysis, we coded species with ranges spanning both continents as having a third character state, rather than being polymorphic (because current implementations of these maximum likelihood reconstructions do not allow for more than two character states). We implemented the character reconstruction and mapping in Mesquite v2.75 (Maddison and Maddison, 2011). For the purpose of

comparison, we analyzed our data using methods in which vicariance is the null model such as S-DIVA (Yu et al., 2010) and Bayesian binary MCMC (BBM) as implemented in RASP v2.1b (Yu et al., 2013), as well as the dispersal-extinction-cladogenesis (DEC) model in Lagrange (Ree and Smith, 2008). All three analyses used the same geographic coding as described above in our parsimony reconstruction analysis. For the BBM model, we ran the MCMC for 5 million generations, sampling every 1000 trees, and discarding the first 500 trees (10%) as burnin.

RESULTS

Phylogenetic analysis

The final MUSCLE alignment of the concatenated dataset was 4,018 characters, with a >95% complete matrix (only ~5% gaps or missing data). Many of the gaps came as a result of a large (665 bp) indel in the FIB7 gene for both *Claravis pretiosa* specimens. Both the ML and Bayesian analyses generated similar trees (Figure 4.1), with support for the Bayesian analysis reaching stationarity and convergence based on the trace plots and Effective Sample Sizes >200 for all paramters. In addition, the ML mitochondrial tree generated in GARLI did not have any major conflicting nodes (<75% BS) with the fully concatenated tree. The concordant gene trees, mitochondrial/full trees, and ML/Bayesian trees support our decision to concatenate our data, and gives credence to the robustness of subsequent results. The majority of ingroup nodes (19/23) received high support (>90 bootstrap/>0.95 posterior probability) from both methods. There is modest support (67 BS, 0.92 PP) for the clade comprising *Claravis mondetoura*, *Metriopelia,* and *Columbina*. However, a clade comprising *Claravis mondetoura,*

Metriopelia, Columbina, and *Uropelia*, to the exclusion of *Claravis pretiosa*, is highly supported (100 BS, 1.0 PP). The BEAST tree places *Claravis mondetoura* sister to *Metriopelia*, but with low posterior probability (<0.7). These results indicate that the genus *Claravis* is paraphyletic. All trees place *Claravis pretiosa* as sister to all other small New World ground doves. The monotypic genus *Uropelia* also appears to be highly divergent from other taxa, being placed as sister to the rest of the group excluding *Claravis pretiosa*. This placement of *Uropelia* and *Claravis pretiosa* is consistent with past phylogenies constructed with fewer species represented (Johnson and Clayton, 2000; Pereira et al., 2007).

Divergence time estimation

BEAST produced a chronogram consistent with the GARLI and MrBayes analyses, with support for convergence based on the trace files. The only major difference between the BEAST tree and the ML/Bayesian trees is the placement of *Claravis mondetoura* as sister to *Metriopelia* (Figure 4.2). Based on the 95% credibility intervals, the small New World ground dove clade diverged from other pigeons and doves between around 19-26 mya, and this clade began to radiate between around 13-18 mya. Some species have diverged quite recently, for example, *Columbina squammata* and *Columbina inca* (<2.5 mya) and *Columbina talpacoti, Columbina buckleyi,* and *Columbina minuta* (<2 mya). *Metriopelia* and *Columbina* diverged from each other around 11-14 mya, with divergences within *Metriopelia* beginning around 9 mya and within *Columbina* beginning around 7.5 mya. The dates we estimate here are considerably younger than the estimates of Pereira et al. (2007), who recover a divergence time of >50 mya for the small New World ground dove clade, with

divergences within the clade beginning >30 mya. However, Pereira et al. (2007) used several external calibrations on deep and highly-divergent nodes, and an internal minimum age constraint based on the oldest Columbiformes fossil for divergence estimates. Using solely external calibrations on such deep nodes can be misleading (Ho et al., 2008).

Biogeographic analysis

Both parsimony and likelihood reconstruction of historical biogeographic regions indicate an ancestral origination in South America with multiple colonization events of North America (Figure 4.2). All of these colonization events are inferred to have occurred after around 2 mya. As expected, the ancestral area reconstructions implemented in S-DIVA and Lagrange seemed to be biased toward vicariance events and produced results that are unlikely. For example, while both analyses recovered a South American origin for the clade, both recovered the ancestor of *C. inca* and *C. squammata* as being widespread in both North and South America indicating a subsequent separation by vicariance. This is in contrast to the parsimony and likelihood character reconstruction, which recover the *C. inca/C. squammata* ancestor as present in South America with subsequent dispersal into North America. The BBM model recovers an identical scenario, as the results from the MCMC chain produced posterior probabilities nearly identical to the likelihood values at each node over the entire tree (likelihood values recorded as pie charts over each node in Figure 4.2).

DISCUSSION

Phylogenetic relationships of small New World ground-doves

Phylogenetic relationships among small New World ground doves based on nuclear and mitochondrial gene regions, with comprehensive species level sampling, are generally well resolved and supported. This tree is broadly in agreement with previous, less exhaustive phylogenetic analyses that included some members of this clade (Johnson et al. 2001, Shapiro et al. 2002, Pereira et al. 2007); however some important novel results emerged from our comprehensive analysis. Similar to prior results, we recovered *Claravis pretiosa* as sister to the rest of the clade, with *Uropelia* recovered as sister to *Metriopelia* plus *Columbina* (Table 4.1). Both *Metriopelia* and *Columbina* were recovered as monophyletic with high support (100 bootstrap, 1.0 PP). The molecular tree also places *Columbina inca* and *C. squammata* within *Columbina*. These two species are often placed in a separate genus (Goodwin, 1983; Gibbs et al., 2001; del Hoyo et al., 1997), *Scardafella*, but the AOU recognizes them as members of *Columbina* (Lack, 2003, American Ornithologists' Union, 1998). Since the clade is nested within *Columbina* based on comprehensive sampling, this provides further support for the inclusion of this clade within the genus. Recognizing these species as a separate genus (*Scardafella)* would render *Columbina* paraphyletic.

We also find that some species are very recently diverged from each other. The divergence among *C. talpacoti*, *C. buckleyi*, and *C. minuta* is relatively recent (between 1-2 mya). Although the ranges of *C. buckleyi* and *C. minuta* overlap to some degree, *C. buckleyi* and *C. talpacoti* do not tend to overlap. *C. buckleyi* has a limited range along coastal Ecuador and Peru, while *C. talpacoti* has a more widespread range, but is found

east of the Andes range and in Central America (IUCN). While the recovery of *C. buckleyi* and *C. talpacoti* as allopatric sister species is consistent with previous work (Gibbs et al., 2001), this very recent divergence seems inconsistent with theories on separation by the Andean uplift (see discussion below), and indicates a more recent separation event.

Finally, and perhaps most surprisingly, our phylogeny indicates that the genus *Claravis* is paraphyletic. While the placement of *Claravis pretiosa* is consistent with previous work (as sister to the rest of the clade), *Claravis mondetoura* is recovered together in a clade with *Uropelia, Metriopelia,* and *Columbina*. While the exact placement of *C. mondetoura* within this clade is uncertain, the exclusion of *Claravis pretiosa*, and thus the paraphyly of the genus *Claravis*, is very highly supported (100 BS, 1.0 PP). Members of the genus *Claravis* are unique among small New World ground doves in that males have mostly blue-gray plumage coloration. Females, however are brownish, similar to most other small New World ground doves. It may be that blue colored males was the ancestral condition in this clade and later was lost in other lineages, with males evolving a more similar plumage coloration to females. Another genus of New World doves, *Geotrygon*, was also shown to be paraphyletic despite strong morphological similarities (Johnson and Weckstein, 2011; Banks et al., 2013). Therefore, such a finding is not unprecedented among pigeons and doves.

Divergence time estimation with respect to major geologic events

Andean uplift. Since the small New World ground dove clade is wide spread throughout South America, the effect of the Andean uplift on the radiation of this group is expected to be pronounced. There are three clades that are of particular interest: the

genus *Metriopelia*, the *Columbina cruziana/picui* clade, and the *Columbina talpacoti*/*C. buckleyi* clade. Species in *Metriopelia* are found in the Puna grasslands of the high Andes. *Columbina cruziana* is found on the coasts of western Peru and Ecuador, while *C. picui* is primarily found in the lowland semi-arid scrub and grasslands east of the Andes, though it does occur in the lowlands of southern Chile. *Columbina buckleyi* and *C. talpacoti* show a similar geographic structure, with *C. buckleyi* found exclusively on the Ecuadorian and northern Peruvian coasts west of the Andes, and *C. talpacoti* found throughout the Amazonian lowlands and Central America.

Metriopelia is estimated to have diverged from its ground-dove ancestor around 11-14 mya. This estimate is somewhat consistent with the timing of the uplift of the Central Andes – the current range of species within the genus – reaching >50% of current elevation (Gregory-Wodzicki, 2000). This divergence time estimation could also coincide with the ecological and geographical isolation of the Puna grasslands – the primary habitat for birds in this genus – due to geographic and climatic changes. Vandervoort et al. (1995) and Gregory-Wodziki (2000) present data indicating internal draining for the region was established around 15 mya, which indicates isolation from the surrounding lowlands. There is also indication of a climatic shift to a much drier environment. Most of the speciation events within *Metriopelia* do not appear to occur until much later, however. This diversification coincides with the traditional age estimate of rapid elevation increase in the Altiplano plateau and Eastern Cordillera within the last 10 my. Garzione et al. (2008) estimated that the central Andean plateau rose from ~2000 m to the current ~4000 m between 6.5-10 mya. Alternatively, the patterns seen in *Metriopelia* could be consistent with the more recent estimates of rapid elevational uplift in the Puna-

Altiplano region of Garzione et al. (2014). In this scenario, the initial divergence of the group was perhaps initiated by a rapid uplift of ~1900 km and subsequent climatic changes between 13-16 mya rather than as a result of climatic changes over a more gradual uplift period. Similarly, the later increase in speciation events within the genus could be related to the second rapid uplift event of ~700 m between 9-13 mya. While it is difficult to form strong conclusions about the effect of Andean uplift on this genus – primarily because both the group's diversification and the uplift events occurred over a long period of time – there are nonetheless indications that diversification in the genus coincided with a rapid increase (or increases) in Andean elevation. A similar pattern of increased diversification associated with rapid elevation increase has been documented in tanagers endemic to the Northern Andes (Sedano and Burns, 2010).

Columbina cruziana and *C. picui* are recovered as sister species, but their geographic ranges are separated by the Andes. Thus, it might be expected that these species would have diverged around the time of the Andean uplift; however, divergence time estimation indicates these taxa diverged 3-7 mya, which is significantly more recent than the divergence estimate of *Metriopelia* and therefore the southern/central Andean uplift. In this case, however, *C. cruziana* and *C. picui* are separated by the northern range of the Andes, which is estimated to have formed 2.5-10 mya. By 2.5-7 mya, coastal Peru would have been cut off from the Amazonian lowlands to the east by the rise of the Eastern Cordillera of Colombia. This event coincides with the divergence estimate between *C. cruziana* and *C. picui*, and strongly suggests the northern Andean uplift as a cause of vicariance and subsequent divergence. The geographic range of *C. picui* does extend to the western side of the Andes in central Chile, but does not extend further north

than the Atacama Desert. A large body of research indicates that the Atacama underwent a rapid increase in aridity around 6 mya and subsequently developed towards its current extreme conditions (Hartley and Chong, 2002). The timing of this environmental change coincides with the development of the Northern Andes and rapid elevation increase in the Central Andes, probably due to the climatic changes associated with western South America being closed off to the rest of the continent by the entirety of the Andes mountain range (Hartley, 2003). The timing also coincides with the estimated *C. cruziana*-*C. picui* divergence time. This could indicate that the speciation event was the direct result of either a north-south divergence due to the formation of the Atacama, an east-west divergence due to the rapid elevation increase in the Northern Andes, or a combination of both geologic events. In other words, *C. cruziana* could have been isolated by both the Andes to the east and the Atacama Desert to the south. Other research has indicated that the Atacama formed much earlier, reaching a point of extreme aridity around 14 mya (Houston and Hartley, 2003; Dunai et al., 2005). If this is the case, the *C. cruziana*-*C. picui* ancestor would have already been hindered from spreading north along the Chilean Pacific coast by around 14 mya, and the subsequent speciation event could be more directly attributed to the rise of the Northern Andes.

The sister species *Columbina talpacoti* and *C. buckleyi* show a similar distributional pattern to *C. cruziana* and *C. picui*, but are estimated to have diverged more recently (<1 mya, and probably <50 kya), which does not coincide with major Andean uplift events. The mitochondrial uncorrected pairwise distance between the two species is also relatively small (0.3-0.8%) consistent with a very recent speciation event. This pattern of closely related species on either side of the Andes has been documented in

several other cases in birds (Brumfield and Capparella, 1996; Miller et al., 2008; Weir and Price, 2011; Lougheed et al., 2013). Some of these instances were shown to be the probable result of trans-Andean gene flow (Miller et al., 2008), but this seems unlikely with ground doves given their poor long-distance flight ability. Dispersal of the ancestors of this group over the Andes seems unlikely. Another possibility is that the ancestor of *C. talpacoti* and *C. buckleyi* was distributed across the Northern Andes throughout forested glacial refugia (Haffer, 1969). These corridors would have provided suitable habitat for dispersing around the Northern Andes, therefore allowing for a continuous distribution across each side of the Andes. It seems plausible that, when these corridors disappeared with the glacial retreat of the Pleistocene, patches of habitat (and presumably representatives of the *C. talpoacoti/C. buckleyi* ancestor) remained isolated on either side of the range, leading to a speciation event.

Panamanian Land Bridge Formation. In many cases the focus of biogeographic studies is on vicariance events, the separation of two previously connected areas (Phillips et al., 2013; Bauza-Ribot et al., 2012; Maderspacher, 2012). However, North and South America represent a case of two continents connecting after millennia of separation. We therefore modeled our biogeographic comparison under the assumption of dispersal (Christenhusz and Chase, 2013), and estimated the origin and direction of dispersal of small New World ground doves, which now occur on both continents. Our results indicate several dispersal events into North America from South America, and the timing of the colonization events appear to coincide with the Panamanian land bridge formation (Figure 4.2). Both the parsimony and maximum likelihood biogeographic reconstructions recover only South to North America dispersal events. Likewise, the posterior

probabilities at this node from the MCMC chain of the BBM model were very similar to the likelihood values from the ML character reconstruction. The BBM results indicate that the ML results are robust to our use of a third character state (present in both North and South America) rather than using a polymorphic character state (present in North America/present in South America). Using a third character state over a polymorphic state is not ideal, but is required in current implementations of likelihood ancestral character reconstructions. Nevertheless, our BBM results – an ancestral area reconstruction model that allows polymorphic states – indicate that our ML analysis produced consistent results.

Based on all three analyses, all South to North American dispersal events occur after the formation of the land bridge. The lack of pre-land bridge dispersal is not surprising given the strong support of South American origin and the unlikely possibility of these ground doves flying across a significantly large water barrier. It is possible the doves could have dispersed into North America by "island hopping" on small land masses thought to have existed between the two continents (Stehli and Webb, 1985; DaCosta and Klicka, 2008), as *C. passerina* seems to have done in colonizing Caribbean islands, but our results strongly suggest otherwise. Of particular significance are the sister taxa *Columbina squammata* and *C. inca*. Although the geographic reconstruction recovers their ancestor as a South American species, *C. inca*'s current range is exclusively in North America while *C. squammata* is a South American species. Their estimated divergence time is around 2-2.5 mya, which indicates a dispersal and speciation event shortly after the land bridge formation. Species that currently have a distribution on

both continents are reconstructed as having originated in South America, with recent range expansion into North America.

Conclusion

Through sampling representatives from each extant species of small New World ground doves, we were able to reconstruct a fairly well-resolved and well-supported phylogeny of this group. More importantly, we were able to use a dated phylogeny to understand the timing of diversification in this group as it relates to historic biogeographic events. Due to their range throughout the New World, we were able to test hypotheses regarding the Andean uplift and formation of the Panamanian land bridge. If neither Andean uplift nor land bridge formation had a major effect on New World ground dove speciation patterns, we would expect the estimated divergence times and ancestral area reconstructions among relevant species to not coincide with either of these geologic events. In particular, we would not expect the divergence estimates of relevant clades (e.g. sister taxa separated by the Andes) to coincide with Andean uplift events. Furthermore, we would not expect the timing of dispersal events between North and South America to coincide with the Isthmus closure. However, our results in this study support several divergence time estimates that are consistent with Andean uplift events, as well as biogeographic reconstructions consistent with dispersal events from South to North America occurring near or after the land bridge formation. These results suggest that Andean uplift and the Panamanian Land Bridge formation were important events in the evolutionary history of small New World ground doves, and provide further insight into how these events contributed to the diversification of New World birds.

FIGURES

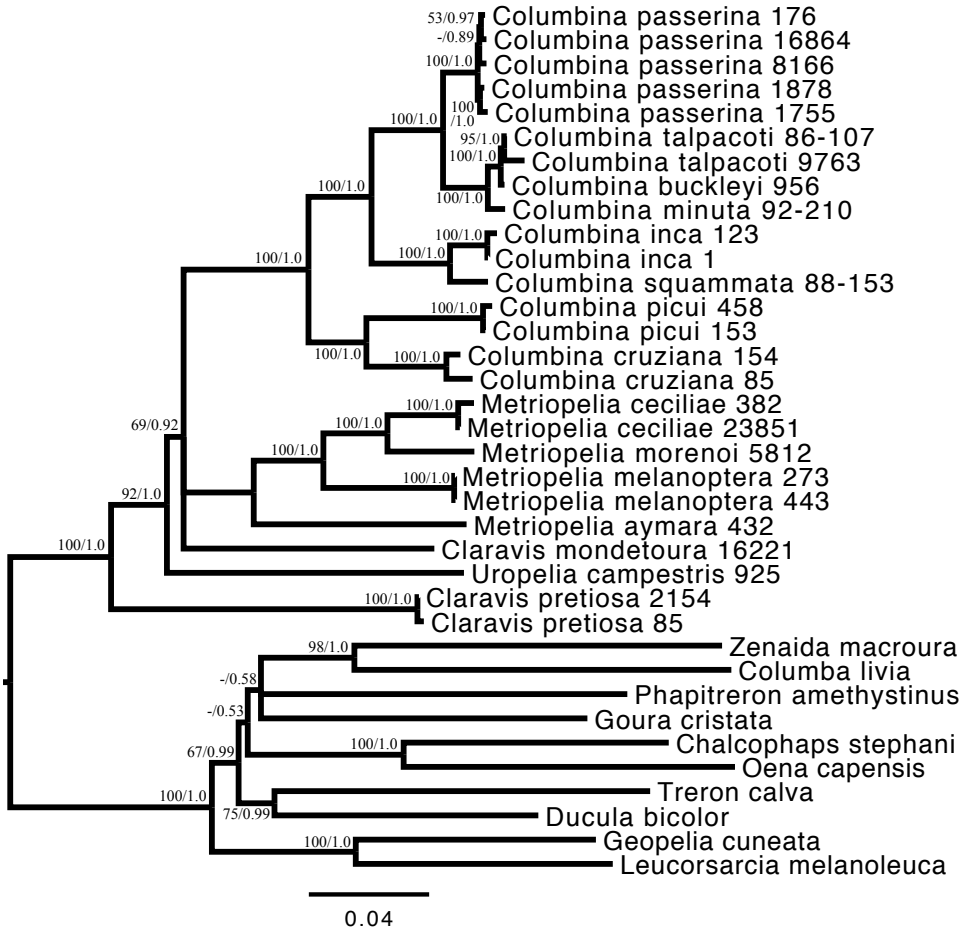


Figure 4.1. Maximum Likelihood and Bayesian tree. For the support values indicated at each node, bootstrap values appear first followed by posterior probability values. Dashes indicate that particular node was not recovered in the appropriate analysis. Letter code at the end of each taxon name indicates country origination of the sample. They are as follows: ARG: Argentina, BAH: Bahamas, BOL: Bolivia, BRA: Brazil, CAP: captive, CR: Costa Rica, ECU: Ecuador, GUY: Guyana, MEX: Mexico, PAR: Paraguay, PER: Peru, USA: United States, VEN: Venezuela.

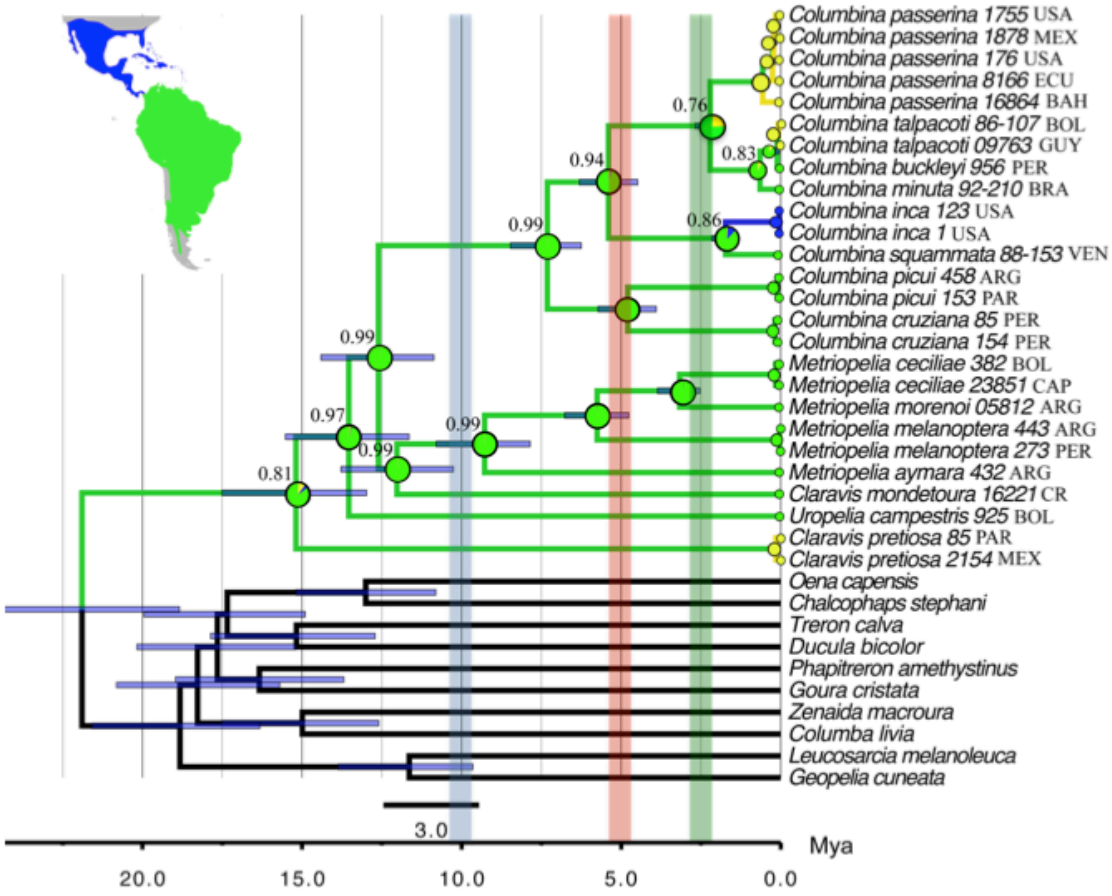


Figure 4.2. Chronogram generated by BEAST. Time along the bottom axis is listed in millions of years before present, and blue error bars over each node indicates the 95% credibility intervals. Colored branches indicate results of the parsimony biogeographic reconstruction, and the pie charts over each node indicate the likelihood that a region is the ancestral area for that particular clade. Values to the upper-left of each pie chart are the marginal probabilities for the most likely ancestral area at each respective node. Marginal probabilities of >0.99 for a particular area at a node are not indicated. Color indication for each region are as follows: Green – South America, Blue – North America, Yellow – widespread in both continents. Columns indicate the approximate estimatin timing for three major geologic events. The blue column indicates the approximate timing of the rapid elevation increase in the central Andean plateau, the red column

(Figure 4.2. continued)

indicates the approximate rapid elevation increase in the Northern Andes, and the green column indicates the approximate estimate for the Panamanian Land Bridge formation. Taxon names are as in Figure 4.1.

TABLES

Table 4.1. Samples included in the study. GenBank accession numbers indicated.

Ingroup									
Genus	**species**	**Extract code**	**Voucher**	**Locality**	**Cytb**	**ND2**	**COI**	**ATP8**	**FIB7**
Columbina	*passerina*	Copas.1.26.1998.1	KUMNH B1755	USA: Missouri	KJ639102	KJ645745	KJ630887	KJ630861	KJ668695
Columbina	*passerina*	CopasTX.9.8.1998.10	176	USA: Texas	KJ639082	KJ645725	KJ630867	KJ630841	KJ668676
Columbina	*passerina*	Copas1878.10.28.1998.1	KUMNH B1878	Mexico	KJ639097	KJ645740	KJ630882	KJ630856	KJ668690
Columbina	*passerina*	Copas8166.8.1.2006.9	ANSP8166	Ecuador	KJ639095	KJ645738	KJ630880	KJ630854	KJ668688
Columbina	*passerina*	Copas16864.8.1.2006.10	LSU B16864	Bahamas	KJ639091	KJ645734	KJ630876	KJ630850	KJ668684
Columbina	*minuta*	Comin.5.6.1998.3	DFS92-210	Brazil	KJ639100	KJ645743	KJ630885	KJ630859	KJ668693
Columbina	*buckleyi*	Cobuc956.6.8.2001.4	LSU RCF956	Peru	KJ639079	KJ645722	KJ630864	KJ630838	KJ668673
Columbina	*talpacoti*	Cotal1504.10.28.1998.8	FMNH SML86-107	Bolivia	KJ639088	KJ645731	KJ630873	KJ630847	KJ668681
Columbina	*talpacoti*	Cotal9763.10.6.1998.3	NMNH B09763	Guyana	KJ639101	KJ645744	KJ630886	KJ630860	KJ668694
Columbina	*picui*	Copic.1.26.1998.5	KUMNH B153	Paraguay	KJ639094	KJ645737	KJ630879	KJ630853	KJ668687
Columbina	*picui*	Copic458.8.1.2006.5	KGM458	Argentina	KJ639080	KJ645723	KJ630865	KJ630839	KJ668674
Columbina	*cruziana*	Cocru154.11.26.2003.5	REW154	Peru	KJ639084	KJ645727	KJ630869	KJ630843	KJ668678
Columbina	*cruziana*	Cocru85.8.1.2006.11	REW85	Peru	KJ639089	KJ645732	KJ630874	KJ630848	KJ668682
Claravis	*pretiosa*	Clpre.1.26.1998.3	KUMNH B85	Paraguay	KJ639096	KJ645739	KJ630881	KJ630855	KJ668689
Claravis	*pretiosa*	Clpre2154.10.28.1998.2	KUMNH B2154	Mexico	KJ639087	KJ645730	KJ630872	KJ630846	KJ668680
Claravis	*mondetoura*	Clmon16221.8.1.2006.4	LSU B16221	Costa Rica	KJ639093	KJ645736	KJ630878	KJ630852	KJ668686
Metriopelia	*ceciliae*	Mecec.4.23.1998.4	LSU B23851	captive	KJ639085	KJ645728	KJ630870	KJ630844	KJ668679
Metriopelia	*ceciliae*	Mecec382.6.8.2001.3	LSU CCW382	Bolivia	KJ639081	KJ645724	KJ630866	KJ630840	KJ668675
Meteriopelia	*morenoi*	Memor.10.6.1998.5	NMNH B05812	Argentina	KJ639083	KJ645726	KJ630868	KJ630842	KJ668677
Metriopelia	*melanoptera*	Memel273.5.15.2003.15	REW273	Peru	KJ639086	KJ645729	KJ630871	KJ630845	-
Metriopelia	*melanoptera*	Memel443.8.1.2006.6	KGM443	Argentina	KJ639092	KJ645735	KJ630877	KJ630851	KJ668685
Metriopelia	*aymara*	Meaym432.8.1.2006.12	KGM432	Argentina	KJ639099	KJ645742	KJ630884	KJ630858	KJ668692
Columbina	*inca*	Coinc1.9.16.1997.1	1	USA: Arizona	KJ639103	KJ645746	KJ630888	KJ630862	KJ668696
Columbina	*inca*	CoincTX.9.8.1998.8	123	USA: Texas	KJ639090	KJ645733	KJ630875	KJ630849	KJ668683
Columbina	*squammata*	Scsqu.5.6.1998.11	SML88-153	Venezuela	KJ639104	KJ645747	KJ630889	KJ630863	KJ668697
Uropelia	*campestris*	Urcam925.6.8.2001.5	LSU CCW925	Bolivia	KJ639098	KJ645741	KJ630883	KJ630857	KJ668691
Outgroups									
Columba	*livia*	Coliv423	UT 423	USA: Utah	AF182694	AF353433	EF373367	EF373446	AF182661
Chalcophaps	*stephani*	Chste	NMNH B4013	Papua New Guinea	AY443673	EF373328	EF373365	EF373439	EF373477

Table 4.1. Continued.

Ducula	*bicolor*	Dubic	LSU B19214	captive	AF182705	KF446740	KJ630891	AY443632	AF182672
Geopelia	*cuneata*	Gecun	KUMNH B1586	captive	AF182711	KC484595	KJ630890	AY443645	AF182678
Goura	*cristata*	Gocri	KUMNH B1588	captive	AF182709	EF373336	EF373374	EF373453	AF182676
Leucosarcia	*melanoleuca*	Lemel	LSU B20539	captive	AF182712	EF373341	EF373379	EF373458	AF182679
Oena	*capensis*	Oecap	FMNH SMG-4180	Madagascar	AF182707	EF373345	EF373383	EF373462	AF182674
Phapitreron	*amethystinus*	Phame	FMNH ATP92-109	Philippines	AF182706	EF373349	EF373387	EF373466	AF182673
Treron	*calva*	Trcal	AMNH ALP80	Cent. Aft. Rep.	AY443674	EF373354	EF373392	EF373471	AY443696
Zenaida	*macroura*	Zemac5	UT 5	USA: Arizona	AF182703	EF373359	EF373397	EF373476	AF258321

LITERATURE CITED

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

American Ornithologists' Union. 1998. Check-list of North American birds, 7th ed. Washington, D.C.: American Ornithologists' Union.

Bacon C.D., Mora A., Wagner W.L., Jaramilo C.A. 2013. Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical Journal of the Linnean Society* 171:287-300.

Banks R.C., Weckstein J.D., Remsen J.V., Johnson K.P. 2013. Classification of a clade of New World doves (Columbidae: Zenaidini). *Zootaxa* 3669: 184–188.

Barker F.K. 2007. Avifaunal interchange across the Panamanian isthmus: insights from *Campylorhynchus* wrens. *Biological Journal of the Linnaen Society* 90: 687-702.

Bates J.M., Hackett S.J., Cracraft J. 1998. Area-relationships in the Neotropical lowlands: An hypothesis based on raw distributions of Passerine birds. *Journal of Biogeography* 25: 783–793.

Bauzà-Ribot M.M., Juan C., Nardi F., Oromí P., Pons J., Jaume D. 2012. Mitogenomic phylogenetic analysis supports continental-scale vicariance in subterranean thalassoid crustaceans. *Current Biology* 22: 2069–2074.

Bess E.C., Catanach T.A., Johnson K.P. 2014. The importance of molecular dating analyses for inferring Hawaiian biogeographical history: A case study with bark lice (Psocidae: *Ptycta*). *Journal of Biogeography* 4: 158–167.

Brumfield R.T., Capparella A.P. 1996. Historical diversification of birds in northwest South America: a molecular perspective on the role of vicariant events. *Evolution* 50: 1607–1624.

Burns K.J., Racicot R.A. 2009. Molecular Phylogenetics of a clade of lowland tanagers: Implications for avian participation in the Great American Interchange. *The Auk* 12: 635–648.

Chapman F.M. 1917. The distribution of bird-life in Colombia. *Bulletin of the American Museum of Natural History* 36: 1–729.

Chesser R.T. 2004. Molecular systematics of New World suboscine birds. *Molecular Phylogenetics and Evolution* 32: 11–24.

Cheviron Z.A., Hackett S.J., Capparella A.P. 2005. Complex evolutionary history of a Neotropical lowland forest bird (*Lepidothrix coronata*) and its implications for historical hypotheses of the origin of Neotropical avian diversity. *Molecular Phylogenetics and Evolution* 36: 338–357.

Christenhusz M.J.M., Chase M.W. 2013. Biogeographical patterns of plants in the Neotropics - dispersal rather than plate tectonics is most explanatory. *Botanical Journal of the Linnean Society* 171: 277–286.

Coates A.G., Jackson B.C., Collins L.S., Cronin T.M., Dowsett H.J., Bybell L.M., Obando J.A. 1992. Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin* 104: 814–828.

Coates A.G., Obando J.A. 1996. Evolution and Environment in Tropical America. In Jackson J.B.C., Budd A.F., Coates A.G., eds. Evolution and Environment in Tropical America. University of Chicago Press, Chicago, IL.

Cracraft J., Prum R.O. 1988. Patterns and processes of diversification: speciation and historical congruence in some Neotropical birds. *Evolution* 42: 603–620.

DaCosta J.M., Klicka J. 2008. The Great American Interchange in birds: a phylogenetic perspective with the genus *Trogon*. *Molecular Ecology* 17: 1328–1343.

Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

del Hoyo J., Elliott A., Boles W. 1997. Handbook of Birds of the World - Vol 4: Sandgrouse to Cuckoos. Barcelona: Lynx.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Dunai T.J., González López G.A., Juez-Larré J. 2005. Oligocene–Miocene age of aridity in the Atacama Desert revealed by exposure dating of erosion-sensitive landforms. *Geology* 33: 321.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Ehlers T., Poulsen C.J. 2009. Influence of Andean uplift on climate and paleoaltimetry estimates. *Earth and Planetary Science Letters* 281: 238–248.

Farris D.W, Jaramillo C., Bayona G., Restrepo-Moreno S.A., Montes C., Cardona A., Mora A., Speakman R.J., Glascock M.D., Valencia V. 2011. Fracturing of the

Panamanian Isthmus during initial collision with South America. *Geology* 39: 1007-1010.

Fjeldså J. 1994. Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodiversity and Conservation* 3: 207–226.

Fjeldså J., Bowier R.C.K., Rahbek C. 2012. The role of mountain ranges in the diversification of birds. *Annual Review of Ecology, Evolution, and Systematics* 43: 249-265.

Fjeldså J., Irestedt M. 2009. Diversification of the South American avifauna: patterns and implications for conservation in the Andes. *Annals of the Missouri Botanical Garden* 96: 398-409.

Fjeldså J., Rahbek C. 2006. Diversification of tanagers, a species rich bird group, from lowlands to montane regions of South America. *Integrative and Comparative Biology* 46: 72–81.

Garzione C.N., Auerbach D.J., Smith J.J., Rosario J.J., Passey B.H., Jordan T.E., Eiler J.M. 2014. Clumped isotope evidence for diachronous surface cooling of the Altiplano and pulsed surface uplift of the Central Andes. *Earth and Planetary Science Letters* 393:173-181.

Garzione C.N., Hoke G.D., Libarkin J.C., Withers S., MacFadden B., Eiler J., Ghosh P., Mulch A. 2008. Rise of the Andes. *Science* 320: 1304–1307.

Gibb G.C., Penny D. 2010. Two aspects along the continuum of pigeon evolution: A South-Pacific radiation and the relationship of pigeons within Neoaves. *Molecular Phylogenetics and Evolution* 56: 698-706.

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

Goodwin D. 1983. Pigeons and Doves of the World, 3rd ed. Ithaca, New York: Cornell University Press.

Gouy M., Guindon S., Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27: 221–224.

Gregory-Wodzicki K. M. 2000. Uplift history of the Central and Northern Andes: A review. *Geological Society of America Bulletin* 112: 1091–1105.

Haffer J. 1969. Speciation in Amazonian forest birds. *Science* 165: 131–137.

Hafner M.S., Sudman P.D., Villablanca F.X., Spradling T.A., Demastes J.W., Nadler S.A. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265: 1087–1890.

Harshman J. 1996. Phylogeny, evolutionary rates, and ducks. University of Chicago, Chicago, Illinois.

Hartley A. 2003. Andean uplift and climate change. *Journal of the Geological Society* 160: 7–10.

Hartley A.J., Chong G. 2002. Late Pliocene age for the Atacama Desert: Implications for the desertification of western South America. *Geology* 30: 43–46.

Ho S.Y.W., Saarma U., Barnett R., Haile J., Shapiro B. 2008. The effects of inappropriate calibration: three case studies in molecular ecology. *PLoS ONE* 3:e1615.

Hoke G.D., Garzione C.N. 2008. Paleosurfaces, paleoelevation, and the mechanisms for the late Miocene topographic development of the Altiplano plateau. *Earth and Planetary Science Letters* 271: 192–201.

Hoorn C., Guerrero J., Sarmiento G.A., Lorente M.A. 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. *Geology* 23: 237–240.

Hoorn C., Wesselingh F.P., Steege H., Bermudez M.A., Mora A., Sevink J., Sanmartin I., Anderson C.L., Figueirdo J.P., Jaramillo C., Riff D., Negri F.R., Hooghiemstra H., Lundberg J., Stadler T., Sarkinen T., Antonelli A. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330: 927–931.

Houston J., Hartley A.J. 2003. The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert. *International Journal of Climatology* 23: 1453–1464.

Johnson K.P. 2004. Deletion bias in avian introns over evolutionary timescales. *Molecular Biology and Evolution* 21: 599–602.

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

Johnson K.P., Clayton D.H., Dumbacher J.P., Fleischer R.C. 2010. The flight of the passenger pigeon: phylogenetics and biogeographic history of an extinct species. *Molecular Phylogenetics and Evolution* 57: 455–458.

Johnson K.P., Sorenson M.D. 1998. Comparing molecular evolution in two mitochondrial protein coding genes (cytochrome b and ND2) in the dabbling ducks (Tribe: Anatini). *Molecular Phylogenetics and Evolution* 10: 82–94.

Johnson K.P., Weckstein J.D. 2011. The Central American land bridge as an engine of diversification in New World doves. *Journal of Biogeography* 38: 1069–1076.

Johnson K.P, De Kort S., Dinwoodey K., Mateman A.C., Cate C.T., Lessells C.M., Clayton D.H. 2001. A molecular phylogeny of the dove genera *Streptoplia* and *Columba*. *The Auk* 118: 874–887.

Jokat W. 2003. Timing and geometry of early Gondwana breakup. *Journal of Geophysical Research* 108: 2428.

Keigwin L.D. 1978. Pliocene closing of the Isthmus of Panama, based on biostratigraphic evidence from nearby Pacific Ocean and Caribbean Sea cores. *Geology* 6: 630.

Kocher, T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S., Villablanca F.X., Wilson A.C. 1989. Dynmaics of mitochontrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science* 88: 6196–6200.

Lack P. 2003. Pigeons and Doves. In Perrins C., ed. The New Encylopedia of Birds. Oxford: Oxford University Press. pp. 288–295.

Leigh G.E., O'Dea A., Vermeij G.J. 2013. Historical biogeography of the Isthmus of Panama. *Biological Reviews* 89: 148-172.

Lloyd H., Marsden S.J. 2008. Bird community variation across *Polylepis* woodland fragments and matrix habitats: implications for biodiversity conservation within a high Andean landscape. *Biodiversity and Conservation* 17: 2645–2660.

Lloyd H., Sevillano Ríos S., Marsden S.J., Valdés-Velásquez A. 2010. Bird community composition across an Andean tree-line ecotone. *Austral Ecology* 37: 470–478.

Lomolino M.V., Riddle B.R., Brown J.H. 2006. Biogeography. Sunderland, MA: Sinauer Assusociates.

Lougheed S.C., Campagna L., Dávila J.A., Tubaro P.L., Lijtmaer D.A., Handford P. 2013. Continental phylogeography of an ecologically and morphologically diverse Neotropical songbird, *Zonotrichia capensis*. *BMC Evolutionary Biology* 13: 58.

Lovette I.J., Bermingham E., Seutin G., Ricklefs R.E. 1998. Evolutionary differentiation in three endemic West Indian warblers. *The Auk* 115: 890-903.

Maddison W.P., Maddison D.R. 2011. Mesquite: a modular system for evolutionary analysis, v.2.75. http://mesquiteproject.org.

Maderspacher F. 2012. Evolution: drift will tear us apart. *Current Biology* 22: R909–12.

Marshall L.G. 1988. Land mammals and the Great American Interchange. *American Scientist* 76: 380–388.

Milá B., Wayne R.K., Fitze P., Smith T.B. 2009. Divergence with gene flow and fine-scale phylogeographical structure in the wedge-billed woodcreeper, *Glyphorynchus spirurus*, a Neotropical rainforest bird. *Molecular Ecology* 18: 2979–2995.

Miller M.J., Bermingham E., Klicka J., Escalante P., do Amaral F.S.R., Weir J.T., Winker K. 2008. Out of Amazonia again and again: episodic crossing of the Andes promotes diversification in a lowland forest flycatcher. *Proceedings of the Royal Society* 275: 1133–1142.

Montes C., Cardona A., McFadden R., Moron S.E., Silva C.A., Restrepo-Moreno S., Ramirez D.A., Hoyos N., Wilson J., Bayona G.A., Jaramillo C.A., Valencia V.,

Bryan J., Flores J.A. 2012. Evidence for middle Eocene and younger land emergence in central Panama: Implications for Isthmus closure. *GSA Bulletin* 124: 780-799.

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

Phillips M.J., Page T.J., de Bruyn M., Huey J.A., Humphreys W.F., Hughes J.M., Santos S.R., Schmidt D.J., Waters J.M. 2013. The linking of plate tectonics and evolutionary divergence. *Current Biology* 23: R603–R605.

Price J.J., Johnson K.P., Clayton D.H. 2004. The evolution of echolocation in swiftlets. *Journal of Avian Biology* 35: 135–143.

Prychitko T.M., Moore W.S. 1997. Comparative evolution of the mitochondrial cytochrome b gene and nuclear -fibrinogen intron 7 in woodpeckers. *Molecular Biology and Evolution* 17: 1101–1111.

Pulgarin-R P.C., Smith B.T., Bryson, Jr. R.W., Spellman G.M., Klicka J. 2013. Multilocus phylogeny and biogeography of the New World *Pheiticus* grosbeaks (Aves: Cardinalidae). *Molecular Phylogenetics and Evolution* 69: 1222-1227.

Qu Y., Lei F., Zhang R., Lu X. 2010. Comparative phylogeography of five avian species: implications for Pleistocene evolutionary history in the Qinghai-Tibetan plateau. *Molecular Ecology* 19: 338–51.

Quintero E., Ribas C.C., Cracraft J. 2012. The Andean *Hapalopsittaca* parrots (Psittacidae, Aves): an example of montane-tropical lowland vicariance. *Zoologica Scripta* 42: 28–43.

Rambaut A. 2012. FigTree v1.4. Edinburgh, UK: University of Edinburgh.

Rambaut A., Drummond A.J. 2007. Tracer v1.4.

Ree R.R., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4-14.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics Applications Note* 19: 1572–1574.

Sedano R.E., Burns K.J. 2010. Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). *Journal of Biogeography* 37: 325–343.

Shapiro B., Sibthorpo D., Rambaut A., Austin J., Wrag G.M., Biminda-Emonds O.R.P., Lee P.L.M., Cooper A. 2002. Flight of the Dodo. *Science* 295: 1683.

Simpson B.B. 1975. Pleistocene changes in the flora of the high tropical Andes. *Paleobiology* 1: 273–294.

Simpson G.G. 1950. History of the fauna of Latin America. *American Scientist* 38:261-389.

Simpson G.G. 1980. Splendid isolation: the curious history of South American mammals. Yale University Press.

Smith B.T., Klicka J. 2010. The profound influence of the Late Pliocene Panamanian uplift on the exchange, diversification, and distribution of New World birds. *Ecography* 33: 333-342.

Stehli F.G., Webb S.D. 1985. The Great American Biotic Interchange. New York: Plenum Press.

Sukumaran J., Holder M. 2008. SumTrees: Summarization of split support on phylogenetic trees v1.0.2.

Swofford D. 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10. Sunderland, MA: Sinauer Assusociates.

Szumik C., Aagesen L., Casagranda D., Arzamendia V., Giacomo D., Giraudo A., Goloboff P., Gramajo C., Kopuchian C., Kretzshmar S., Lizarralde M., Molina A., Mollerach M., Navarro F., Nomdedeu S., Panizza A., Pereyra V.V., Sandoval M., Scrocchi G., Zuolaga F.O. 2012. Detecting areas of endemism with a taxonomically diverse data set: plants, mammals, reptiles, amphibians, birds, and insects from Argentina. *Cladistics* 28: 317–329.

Upchurch P. 2008. Gondwanan break-up: legacies of a lost world? *Trends in Ecology and Evolution* 23: 229–236.

Vandervoort D.S., Jordan T.E., Zeitler P.K., Alonso R.N. 1995. Chronology of internal drainage development and uplift, southern Puna plateau, Argentine central Andes. *Geology* 23: 145.

Voelker G. 1999. Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (Anthus: Motacillidae). *Evolution* 53: 1536–1552.

Vuilleumier F. 1969. Pleistocene speciation in birds living in the high Andes. *Nature* 223: 1179–1180.

Vuilleumier F. 1985. Forest birds of Patagonia: Ecological geography, speciation, endemism, and faunal history. *Ornithological Monographs* 36: 255–304.

Wallace A.R. 1889. Travels on the Amazon. New York: Harper and Row.

Wegner W., Worner G., Harmon R.S., Jicha B.R. 2010. Magmatic history and evolution of the Central American Land Bridge in Panama since Cretaceous times. *Geological Society of America Bulletin* 123: 703–724.

Weir J.T., Price M. 2011. Andean uplift promotes lowland speciation through vicariance and dispersal in *Dendrocincla* woodcreepers. *Molecular Ecology* 20: 4550–4563.

Weir J.T., Schluter D. 2008. Calibrating the avian molecular clock. *Molecular Ecology* 17: 2321–2328.

Yu Y., Harris A.J., He X.J. 2013. RASP (Reconstruct Ancestral State in Phylogenies) 2.1 beta. Available at http://mnh.scu.edu.cn/soft/blog/RASP.

Yu Y., Harris A.J., He X.J. 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution* 56: 848-850.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, The University of Texas, Austin, Texas.

CHAPTER 5: COPHYLOGENETIC ANALYSIS OF NEW WORLD GROUND-DOVES (AVES: COLUMBIDAE) AND THEIR PARASITIC WING LICE (INSECTA: PHTHIRAPTERA: COLUMBICOLA)[4]

INTRODUCTION

Parasites and their hosts form intricate and often complex evolutionary relationships. Untangling the narrative of how hosts and parasites interact and what factors are important for shaping their evolutionary patterns is a challenging task. Many parasites utilize multiple unrelated hosts, or are associated with different hosts during different life stages (Morand et al., 1995; Bartholomew et al., 1997; Parker et al., 2003). Hosts also often harbor multiple types of closely related parasites (Poulin, 1997; Bruydonckx et al., 2009; Prugnolle et al., 2010; Colinet et al., 2013). However, reconstructing the evolutionary history of these interactions can provide novel biological insight (Poulin, 2011). Comparing the evolutionary trees of hosts and associated parasites is a way to test for factors influencing joint patterns of host and parasite diversification (Page, 2003). Congruence between host and parasite trees indicates that cospeciation may be important, suggesting parasites are strongly associated with their hosts, whereas incongruence is a sign of host switching or other cophylogenetic events (e.g. parasite

[4] Reprinted, with permission, from: Sweet, A.D. and K.P. Johnson. 2016. Cophylogenetic analysis of New World ground-doves (Aves: Columbidae) and their parasitic wing lice (Insecta: Phthiraptera: Columbicola). *Molecular Phylogenetics and Evolution*. 103: 122-132. https://doi.org/10.1016/j.ympev.2016.07.018.

duplication; Page, 1994). The cophylogenetic patterns revealed by these comparisons can also provide a starting point for testing hypotheses about what biotic and/or abiotic factors dictate the observed patterns of congruence. For example, host habitat (Krasnov et al., 1997), host/parasite behavior (Clayton et al., 2010), biogeography (Weckstein, 2004), and climate (Feder et al., 1993) could all play important roles in shaping an interaction.

Targeting host-parasite systems where the parasites have permanent and obligate relationships to a group of hosts can simplify the "untangling" process (Fahrenholz, 1913; Eichler, 1948; Hafner and Page, 1995). Ectoparasitic feather lice (Insecta: Ischnocera) of pigeons and doves (Aves: Columbidae) (hereafter referred to only as "doves") are an ideal system for this purpose (Johnson and Clayton, 2004). Dove lice are widespread, fairly host specific (an average of 1.7 louse species per dove host taxon), and spend their entire life cycle on the host (Price et al., 2003; Marshall, 1981). Doves are host to two types of distantly related feather lice – wing lice and body lice. Dove wing lice are in a single genus (*Columbicola*), whereas dove body lice are in multiple genera (*Auricotes, Coloceras, Campanulotes, Kodocephalon,* and *Physconelloides*). Although both types of lice parasitize the same group of hosts and are often found together on a single individual, previous analysis indicates the two groups have different evolutionary histories with their hosts (Clayton and Johnson, 2003; Johnson and Clayton, 2004; Johnson et al., 2002). Body lice showed strong phylogenetic congruence with their hosts, which is expected for host-specific parasites and implies cospeciation between the two groups of organisms. Wing lice did not exhibit a similar pattern of congruence, instead showing evidence of multiple host-switches between dove hosts. More frequent host switching may be due to wing lice using phoresis ("hitchhiking") behavior with generalist

parasitic hippoboscid flies to move between host species (Harbison et al., 2008; Harbison and Clayton, 2011).

However, these initial cophylogenetic studies had limited taxonomic and geographic sampling. A recent study with a more extensive taxonomic representation of both doves and their lice showed different patterns, with wing lice showing strong evidence for overall phylogenetic congruence with their hosts (Sweet et al., 2016). These results imply that taxonomic and geographic scale of sampling could greatly affect the results of a cophylogenetic study. Clade-limited host switching, when parasites preferentially switch among closely related hosts, can also mislead results by producing a false signal of phylogenetic congruence (Charleston and Robertson, 2002; Sorenson et al., 2004; de Vienne et al., 2007). In consideration of these issues, it is therefore important to study systems with as complete a taxonomic representation as possible. Since this may be less feasible for higher taxonomic groupings (e.g. the over 300 species of Columbidae), it is necessary to focus on cophylogenetic patterns in specific clades in order to obtain near complete taxonomic sampling. Targeting a specific clade of hosts and their parasites, with comprehensive sampling of multiple individuals per taxon, also provides the opportunity to sample from multiple host populations in different geographic locations.

To this end, we focus on the cophylogenetic patterns between small New World ground-doves and their wing lice (*Columbicola*). Small New World ground-doves are a clade of four genera (*Claravis*, *Columbina*, *Metriopelia*, and *Uropelia*) and 17 species within Columbidae (Johnson and Clayton, 2000; Shapiro et al., 2002; Pereira et al., 2007). Representatives of the clade are small-bodied birds that primarily forage on grass

seeds and prefer open scrubby habitat, although species in the genus *Claravis* are found in forested areas (Goodwin, 1983; Gibbs et al., 2003). The group has a broad geographic distribution, extending throughout the southern United States, Central America, and most of South America, although many species have more localized ranges. Four species of *Columbicola* are known to parasitize small New World ground-doves (*C. passerinae*, *C. altamimiae*, *C. drowni*, and *C. gymnopelia*). These lice form a monophyletic group within *Columbicola* (Johnson et al., 2007). However, previous phylogenetic studies on *Columbicola* have detected additional lineages within the ground-dove wing lice clade, perhaps indicative of cryptic species (Johnson et al., 2002; Johnson et al., 2007). Parasites often have simplified morphological features, which make cryptic species a relatively common phenomenon (Poulin and Morand, 2000; Jousson et al., 2000; Lafferty and Kuris, 2002; Miura et al., 2005; Detwiler et al., 2010). For cophylogenetic analyses it is important to properly identify parasite species, as misrepresenting the number of tips on a parasite (or host) tree can alter the outcome of an analysis (Refrégier et al., 2008; de Vienne et al., 2013; Martinez-Aquino, 2016).

In this study, we use mitochondrial and nuclear data from multiple geographic representatives of each species of small New World ground-dove *Columbicola* to infer a robust phylogeny of the clade. Based on our phylogenetic analysis we identify potential cryptic lineages/species in this group, and use these results to aid us in providing an adequate parasite species tree for subsequent cophylogenetic analyses. We compare this tree to a published tree for their hosts. We also explore the phylogeographic patterns of ground-dove *Columbicola,* particularly focusing on the widespread species *C. passerinae.*

We use this as a basis to test whether the ground-dove *Columbicola* phylogeny is significantly structured according to host biogeography rather than to host phylogeny.

METHODS

Data collection

Louse specimens were collected from hosts using fumigation or pyrethrin powdering protocols (Clayton and Drown, 2001), then immediately placed in 95% ethanol and stored long-term at -80 C. We extracted DNA from individual lice using a Qiagen Blood and Tissue Kit (Qiagen, Valencia, CA, USA) according to standard protocol and modified according to Johnson et al. (2003), with lice incubating in digestion buffer at 55 °C for ~48 hours. After DNA extraction all lice exoskeletons were slide-mounted and saved as voucher specimens. Using polymerase chain reaction (PCR), we targeted the mitochondrial locus cytochrome oxidase subunit 1 (CO1) and the nuclear loci elongation factor 1α (EF-1α), transmembrane emp24 domain-containing protein 6 (TMEDE6), and a hypothetical protein (HYP). For PCR reactions, we used NEB 5X Master Mix (New England Biolabs, Ipswich, MA, USA) and the manufacturer's protocol for 25 μL reactions (5 μL 5X Master Mix, 0.2 μM forward and reverse primers). We used primers H7005 and L6625 for CO1 (Hafner et al., 1994), Ef1 and Cho-10 for EF-1α (Danforth and Ji, 1998), BR69-295L and BR69-429R for TMEDE6, and BR50-181L and BR50-621R for HYP (Sweet et al., 2014). Our thermal cycler protocols followed Johnson et al. (2001) and Sweet et al. (2014). We purified the resulting PCR products with ExoSAP-IT (Affymetrix, Inc., Santa Clara, CA, USA) according to standard protocol, and sequenced them using an ABI Prism BigDye Terminator kit (Applied Biosystems,

Foster City, CA, USA). Fragments were then run on an AB 3730x capillary sequencer (Applied Biosystems, Foster City, CA, USA) at the University of Illinois Roy J. Carver Biotechnology Center (Champaign, IL, USA). We used Geneious v8.1.2 (Biomatters) to manually resolve the resulting complementary chromatograms and remove primer sequences. We submitted all novel sequences to GenBank (Table 5.1). We also utilized existing GenBank data for COI and EF-1α generated in Johnson et al. (2007) and Johnson et al. (2002). For an outgroup taxon, we included sequence data from the rock pigeon wing louse *C. columbae*. In total, we sequenced 51 *Columbicola* samples from 13 host species.

Phylogenetic analysis

We aligned each locus using default gap parameters in the Geneious MUSCLE plugin (Edgar, 2004), and checked each alignment by eye in Geneious. We concatenated all four alignments, and used PartitionFinder v1.1.1. (Lanfear, 2012) to search for the most appropriate gene partitions and substitution models under the corrected Akaike Information Criterion (AICc, Sugiura, 1978). We ran two PartitionFinder searches, one searching through all 56 models in PartitionFinder and another only searching through models applicable in MrBayes.

We then ran both maximum likelihood (ML) and Bayesian phylogenetic reconstruction methods. For ML, we ran a partitioned analysis in Garli v2.0. Based on our PartitionFinder results, we applied a TVM+G substitution model to COI, a TrN+I+G model to HYP, and a TrNef+G model to an EF-1α/TMEDE6 partition. We ran Garli using two searches of 500 bootstrap replicates, and summarized resulting bootstrap trees using Sumtrees v3.3.1 (Sukumaran and Holder, 2008). For the Bayesian analysis, we

used MrBayes v3.2 (Ronquist and Huelsenbeck, 2003). Based on our MrBayes-specific model search, we applied a GTR+G model to COI, a K80 model to EF-1α, and a GTR+I+G model to a HYP/TMEDE6 partition. We ran 20 million generations of Markov Chain Monte Carlo (MCMC) for 2 runs of 4 chains each, sampling every 1,000 trees. To assess parameter convergence we viewed trace files in Tracer v1.5 (Rambaut and Drummond, 2007), and ran .t files with the R package RWTY v1.0.0 (Warren et al., 2016) to assess topological congruence. Based on these assessments, we discarded the first 10% (2,000 trees) as a burnin.

OTU analysis

We used several methods for identifying potential cryptic lineages/species. First, we computed uncorrected pairwise distances for the COI sequences. We computed the distance matrix using the "dist.dna" command in the *ape* package (Paradis et al., 2004) in R (R Development Core Team 2015). Second, we used the COI sequences to infer a median-joining (MJ) network (Bandelt et al., 1999) in PopART v1.7 (Leigh and Bryant, 2015). We set epsilon = 0 and only included sequence data for *Columbicola passerinae*, since previous studies have identified this species as potentially harboring multiple cryptic lineages (Johnson et al., 2007). Third, we used the online version of the Automatic Barcode Gap Discovery (ABGD) method (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html, Puillandre et al., 2012) to partition the number of possible groups in our data set. We ran the ABGD analysis with the COI alignment applying both K2P and Jukes-Cantor (JC) distance models, and used default Pmin, Pmax, gap width, and steps values. Finally, we used the Generalized Mixed Yule Coalescent (GMYC) approach to species delimitation (Fujisawa and Barraclough,

2013). Since this approach requires ultrametric and bifurcating gene trees, we estimated the gene tree phylogeny of COI using BEAST v1.8.2 (Drummond et al., 2012). We ran the MCMC analysis for 20 million generations using a GTR+G substitution model, an uncorrelated lognormal relaxed clock, and a Yule speciation processes tree prior. We sampled trees every 1,000 generations, and discarded the first 10% as a burn-in based on plots and ESS values from Tracer. From the post-burn-in tree samples we constructed a maximum clade credibility (MCC) tree using TreeAnnotator. With the MCC COI tree as an input, we implemented GMYC in the R package *splits* (Ezard et al., 2009) using the single threshold setting.

Cophylogenetic analysis

Using our small New World ground-dove *Columbicola* phylogeny (see Figure 5.1) and information from our species/lineage analyses, we conducted both event-based and distance-based cophylogenetic analyses. For all methods, we used the host phylogeny generated by Sweet and Johnson (2015). We also pruned the louse phylogeny so that each species was represented by a single tip, and removed the outgroup taxon.

For the distance-based cophylogenetic analyses, we used ParaFit (Legendre et al., 2002) and PACo (Balbuena et al., 2013). Both methods assess overall congruence between the host and parasite phylogenies, as well as the relative contribution of individual host-parasite links (associations) to the overall congruence. However, ParaFit assesses whether or not parasites are randomly associated with their hosts, whereas PACo assess the dependence of the parasite phylogeny on the host phylogeny through a residual sum of square goodness-of-fit test. For both analyses, we converted the host and parasite phylogenies to patristic distance matrices using the "cophenetic" command in *ape*, and

sorted each distance matrix according to the host-parasite association matrix. We ran ParaFit for 100,000 permutations in the R package *ape* using the Cailliez correction for negative eigenvalues and testing for the contribution of each individual link using the ParaFitLink1 test. Since ParaFit runs multiple tests to calculate p-values for each individual link, it is necessary to correct the raw output. Using R, we corrected individual link p-values using the Benjamini-Hochberg correction for false discovery rate (Benjamini and Hochberg, 1995). We also ran PACo for 100,000 permutations with the R packages *ape* and *vegan* (Oksanen et al., 2016), and used the jackknife method to estimate the importance of each individual link to the overall sum of squares score.

For an event-based approach, we used Jane v4 (Conow et al., 2010). We used default settings for the Genetic Algorithm parameters (100 generations, population size of 100) and event costs (0 cospeciation, 1 duplication, 2 duplication and host switch, 1 loss, and 1 failure to diverge). After solving for the most optimal solutions, we tested whether our best score was lower than expected by chance by randomizing the tip mappings 999 times. If the randomization procedure indicates our best score from the data is lower than by chance, this would indicate some level of congruence between the host and parasite phylogenies.

Testing for biogeographic structure

To test if our inferred louse phylogeny is significantly structured according to host biogeography, we used the Maddison-Slatkin test (Maddison and Slatkin, 1991). We coded lice as being associated with hosts in one of the following regions: southern United States/northern Central America (north of the Isthmus of Tehuantepec), southern Central America (south of the Isthmus of Tehuantepec), Andes Mountains, South America west

of the Andes (trans-Andes), or South America east of the Andes (cis-Andes). We used the best ML tree as an input, but removed duplicate taxa to avoid biasing our test results. In this context, we considered two louse taxa as duplicate if they were from the same host species, geographic region (according to our coding), and were separated by short branch lengths (uncorrected CO1 distance values < 0.1%), being relatively genetically indistinct. We also removed the outgroup taxon. We implemented the Maddison-Slatkin procedure using an R script that randomly assigns character states (geographic regions) 999 times and calculates the parsimony score for each assignment (Bush et al., 2016, script from https://github.com/juliema/publications/tree/master/BrueeliaMS).

RESULTS

Phylogenetic analysis

We sequenced 378 bp of the CO1 locus (131 variable sites, 115 parsimony-informative sites), 345 bp of EF-1α (32 variable sites, 15 parsimony-informative sites), 407 bp of HYP (71 variable sites, 26 parsimony-informative sites), and 219 bp of TMEDE6 (28 variable sites, 4 parsimony-informative sites). The final concatenated alignment was 1,349 bp in length, with ~36% missing data (Table 5.1). The ML and Bayesian phylogenetic analyses inferred similar trees (Figure 5.1). After the 10% burn-in all parameters and topologies from the Bayesian analysis had ESS values <200 or average standard deviation of split frequencies <0.01, thus indicating the MCMC runs had converged to stationarity. Monophyly of all four *Columbicola* species were recovered with good support from both analyses (>80 bootstrap [BS] and >0.95 posterior probability [PP]).

OTU analyses

The OTU analyses indicated there are five taxa in this clade of *Columbicola*. This includes the three species found on *Metriopelia* ground-doves – *C. altamimiae*. *C. drowni*, and *C. gymnopeliae*. The uncorrected p-distances had an average distance of 13.43% between samples of *C. altamimiae* and all other ingroup samples, 14.31% between samples of *C. drowni* and all other ingroup samples, and 13.88% between samples of *C. gymnopeliae* and all other ingroup samples (Supplementary Table 5.1). The ABGD analysis also recovered those three taxa as separate taxonomic units. The GMYC analysis did not recover these three taxa as separate units, instead grouping all three into a single taxonomic unit. However, GMYC separated them from the fourth species in the clade, *C. passerinae.* All OTU analyses indicated *C. passerinae* should be considered two taxa (here labeled *C. passerinae* 1 and *C. passerinae* 2). Both the ABGD (with both K2P and JC distance models) and GMYC analyses recovered two separate taxonomic units within *C. passerinae*, and those two groups had an average uncorrected p-distance of 9.30% between them. In addition, the median-joining network indicated there are 23 steps between the two *C. passerinae* groups (Figure 5.2). There was some genetic distinctiveness between the *C. passerinae* parasitizing *Columbina passerina* and *Co. inca* from the United States/Mexico and *Co. talpacoti* from Panama, which together differed by an average of 1.80% in COI. However, the ABGD and GMYC analyses did not recover this group as a distinct unit. Additionally, the MJ network placed lice from *Co. talpacoti* from Panama as embedded within *C. passerinae* 1 (Figure 5.2).

Cophylogenetic analysis

Our cophylogenetic analyses based on five *Columbicola* taxa indicated some level of congruence between the host and parasite phylogenies. Both distance-based tests were significant across the entire data set (PACo global P = 0.002, ParaFit global P = 0.003), thus rejecting the independence of the host and parasite phylogenies. The ParaFitLink1 test recovered five host-parasite links as significantly contributing to the global score after correcting for multiple tests ($\alpha = 0.05$) (Figure 5.3, Table 5.2). ParaFit also calculates a second individual link statistic (ParaFitLink2), but ParaFitLink1 is better suited for scenarios with widespread parasites (Legendre et al., 2002, Dhami et al., 2013, Perez-Escobar et al., 2015 [Supporting information]). The individual jackknife link test in PACo recovered three host-parasite links with the 95% confidence intervals of their squared residuals lower than the median global squared residual (Figure 5.4). All three links were between *Metriopelia* ground-doves and their lice.

The event-based method of Jane also recovered a global signal of congruence across the whole data set. The observed cost was 27, which was significantly lower than by chance (P = 0.049). The event reconstruction recovered one cospeciation event, between *M. melanoptera* and *M. ceciliae* and their lice, *C. drowni* and *C. gymnopeliae* (Figure 5.3). Jane also recovered three duplications, fifteen losses, nine failures to diverge, and no host switches (Table 5.3).

Phylogeographic patterns and biogeographic structure

Lice from *C. passerinae* 1 group neatly according to biogeography in both the phylogeny and MJ network. There are distinct groups of lice from southern United States/northern Central America, southern Central America, and cis-Andean South

America (Figures 5.1 and 5.2). Lice from *C. passerinae* 2 also tend to group by biogeography, but the patterns are less well-defined than in *C. passerinae* 1. Lice from *Co. cruziana* and *Co. buckleyi,* both restricted to South America's northwest coast, fall within *C. passerinae* 2. However, all lice from *Cl. pretiosa* are also within this lineage, including samples from Mexico and Brazil. Lice parasitzing *Co. talpacoti* sampled from Brazil also fall within the lineage. In total, three of our four host biogeographic regions are represented in *C. passerinae* 2. However, lice from *C. passerinae* 2 seem to primarily parasitize hosts from trans-Andean South America, given that all lice from this region group with the lineage (Figures 5.1 and 5.2).

Randomization of biogeography over the *Columbicola* phylogeny with the Maddison-Slatkin test indicated significant phylogenetic conservation of biogeography. After trimming the louse phylogeny to remove duplicate samples, there were six observed character state transitions (Supplementary Figure 5.1). None of the randomizations resulted in equal or fewer state transitions than the observed value ($P < 0.001$).

DISCUSSION

Our phylogenetic analyses of ground-dove wing lice (*Columbicola*) revealed patterns of diversification concordant at some level with both host phylogeny and biogeography within this group. In particular, biogeographic distribution, separate from host phylogeny, appears to play a major role in patterns of louse diversification. More specifically, some lineages of lice are associated with multiple distantly related hosts, but these lice tend to occur together in the same biogeographic region.

Some of our phylogenetic results agreed with previous, less extensively sampled studies (Clayton and Johnson, 2003; Johnson et al., 2007). In particular, we recovered two distinct clades within *C. passerinae* (*C. passerinae* 1 and *C. passerinae* 2), and *C. gymnopeliae* + *C. drowni* as sister to *C. passerinae*. We also recovered *C. altamimiae* as sister to the rest of the clade. Increased taxon sampling indicated even more structure within the clade. We found some evidence for genetic differentiation within *C. passerinae* 1*,* with lice from *Co. passerina, Co. inca*, and *Co. talpacoti* forming a distinct group. However, our OTU analyses and MJ network did not recover these lice as representing a distinct taxon, so this likely represents population level structuring. We recovered *C. drowni* as the sister species to *C. gymnopeliae*, although this relationship was not well supported (53 BS/0.75 PP). All three species of *Metriopelia* doves included in this study have unique species of lice previously described based on morphological data (Eichler, 1953; Clayton and Price, 1999). Each louse species is generally associated with one species of *Metriopelia*. *Columbicola altamimiae* has been recorded on both *M. aymara* (the primary host) and *M. melanoptera* (Price et al., 2003), but we were unable to document this by our sampling. Our study provides further evidence for the genetic distinctiveness of each species of *Metriopelia* wing louse. All three species (*C. altamimiae*, *C. drowni*, and *C. gymnopeliae*) are separated from the rest of the ground-dove wing louse clade by long branches and have large CO1 uncorrected distances between them and other taxa (13.43% - 14.31%).

The phylogenetic structure and host specificity of *Metriopelia* wing lice is perhaps reflective of their hosts' life history. *Metriopelia* doves are high Andean species, living in paramo and puna grasslands generally above 2,000 meters. Due to the high

elevation and extreme conditions, these regions have different habitats compared to the neighboring lowlands and cloud forest (Szumik et al., 2012). Inhabitants of these regions are well-adapted to the environment, and have therefore become isolated from other taxa in adjacent regions (Lloyd et al., 2010). This lack of species overlap is especially evident in avian taxa (Fjeldsa et al., 2012). The isolation of *Metriopelia* from other genera of small New World ground-doves is reflected by the phylogenetic distinctiveness of their wing lice. While all other species of ground-doves share one or two species of wing lice with other lowland ground dove species, *Metriopelia* dove lice are in distinct lineages. Interestingly, we did not recover lice from *Metriopelia* as monophyletic, with lice from *M. aymara* coming out as sister to the rest of the ingroup. This indicates ground-dove *Columbicola* may have originated before major central Andean uplift (16-9 mya; Garzione et al., 2014), when *Metriopelia* likely became isolated from other ground-dove taxa (Sweet and Johnson, 2015).

Our results also indicate significant host-specificity among *Metriopelia* wing lice, consistent with known association records (Price et al., 2003). Each species of *Metriopelia* dove included in this study has a unique species of *Columbicola* associated with it, which indicates the lice have been isolated on their hosts for a long period of time. If this was not the case, we might expect the three louse species to be separated by short branch lengths and smaller COI pairwise distances, or for individual species to occur on multiple host species (i.e. not have fully sorted). All three dove species have range overlap, but have some variation in altitudinal preference. *M. aymara* tends to live at higher altitudes (2,800 - >5,000 meters) than the other two species. This habitat difference could result in hosts rarely coming in contact, therefore isolating their

parasites. Host size differences may also be important for explaining the phylogenetic patterns in these lice, particularly in lice from *M. melanoptera* and *M. ceciliae*. Experimental work has indicated that host size differences can limit host switching in dove wing lice (Clayton et al., 2003; Johnson et al., 2005). If a new host is too small or large, wing lice are not able to effectively avoid preening behavior and are thus not able to establish viable populations. *M. melanoptera* (113-125 grams) are considerably larger than *M. ceciliae* (51-67 grams) (Gibbs et al., 2001).

Cophylogenetic congruence

All three cophylogenetic analysis methods, including both topology-based and event-based methods, recovered a global signal of congruence between the host and parasite phylogenies. This indicates some level of cospeciation between the two groups of organisms, which might seem surprising given the general patterns of the phylogeny and host associations (Figure 5.3). These results also differ from earlier cophylogenetic studies on dove wing lice, which indicated a lack of congruence between the two groups (Clayton and Johnson, 2003). However, more recent studies with greater taxonomic representation also recovered evidence of cospeciation between doves and their wing lice (Sweet et al., 2016).

The signal of congruence is probably driven by the *Metriopelia* associations. The only cospeciation event recovered by Jane is between *M. melanoptera* + *M. ceciliae* and their lice, and ParaFit recovered both of those links as significantly contributing to the overall congruence. In the PACo analysis, the three *Metriopelia* host-parasite links had squared residual values that were lower than the other links and 95% confidence intervals below the median squared residual value, which indicates phylogenetic congruence

between those taxa. As discussed above, *Metriopelia* doves are generally isolated from other species in the ground-dove clade, and based on the one-to-one relationships with their wing lice are probably generally isolated from one another. Over evolutionary time, a lack of opportunity to switch hosts could not only result in distinct lineages of parasites, but also in cospeciation with their hosts. Interestingly, removing the *Columbicola passerinae* lineages from the louse phylogeny would result in a perfectly congruent relationship between *Metriopelia* and their wing lice. Furthermore, the basal lineages of the louse phylogeny (*C. altamimiae*) is associated with *M. aymara*, which is nested within the ground-dove phylogeny. This finding indicates ground-dove *Columbicola* possibly switched from ancestral *Metriopelia* doves to other ground-doves before Andean uplift isolated the groups, although Jane did not recover any host switching events. Since *M. aymara* is the basal lineage of *Metriopelia* and is usually found at higher elevations than other *Metriopelia*, it is possible this species first became isolated due to rapid uplift over several million years. This could explain why *C. altamimiae* is the earliest diverging species of ground-dove wing louse.

Conversely, the two lineages of *Columbicola passerinae* do not appear congruent with their hosts. We did not recover any cospeciation events between these lice and their hosts. Most of the individual links did not significantly contribute to the global ParaFit score and had high squared residuals from the PACo jackknife test. Taken together, these results suggest a lack of cospeciation between these two louse lineages and their hosts over evolutionary time. Both lineages are widespread, with *C. passerinae* 1 associated with seven ground-dove species and *C. passerinae* 2 associated with four. This distribution indicates recent host-switching or ongoing gene flow between the louse

populations on these different host species. Small New World ground-doves generally prefer open scrubby habitat, and many of the lowland species (non-*Metriopelia*) are known to forage in mixed flocks (Dias, 2006). Host proximity could allow for ongoing gene flow between louse populations on different host species that are geographically proximal.

Phylogeographic structure of *Columbicola passerinae*

Although most of the ground-dove *Columbicola* phylogeny does not appear to be predicted by the hosts' phylogeny, particularly within *C. passerinae*, there is still significant phylogenetic structure within the clade. Previous work in other host-parasite systems have found that host biogeography, rather than host phylogeny, can be a better predictor of parasite evolutionary patterns (Weckstein, 2004; Johnson et al., 2007; Bush et al., 2016). After assigning host biogeography to the tips of our parasite phylogeny and using the Maddison-Slatkin character randomization test, we found the louse phylogeny is significantly structured according to host biogeography. Given these results and the results of our cophylogenetic analysis, this indicates host biogeography is very important for shaping evolutionary patterns in ground-dove wing lice. A similar pattern was recovered in a broad phylogenetic study of the *Columbicola* genus, which indicates host biogeography is an important factor at both a global and local scale (Johnson et al., 2007). If parasites are able to switch hosts but are limited to a group of similar hosts, as is the case with ground-dove wing lice, then the lice are likely to switch to hosts in close proximity. This could especially be the case in mixed foraging flocks. Since ground-doves are non-migratory and generally do not travel long distances within their ranges,

over time the evolutionary patterns of their parasites would diverge according to geographic regions.

We see early indications of genetic divergence from our phylogenetic and network analyses. Ground-dove wing lice form distinct groups based on hosts primarily from southern United States/northern Central America, southern Central America, trans-Andean South America, cis-Andean South America, or Andean highlands. Further evidence for this phylogeographic structure comes from lice that parasitize the same host species falling in different clades, but being grouped with lice from the same geographic region rather than host species. For example, *C. passerinae* from *Co. talpacoti* in Panama form a distinct group, whereas lice off of *Co. talpacoti* from Brazil group with lice from *Cl. pretiosa* also from Brazil. Also, lice from *Co. passerina* from Brazil group with lice from other cis-Andean South American hosts, and *C. passerina* lice sampled from hosts in Mexico/United States group with other lice from that region. Lice from these widespread host species that group according to geographic region are likely from different subspecies of host. Over time, these host subspecies may continue to diverge from each other and show similar conserved phylogeographic structure. Lice have much shorter generation times and faster substitution rates than their hosts, so geographic structure can potentially be more easily detected in lice than in their hosts (Johnson et al. 2014). Future phylogeographic analysis of widespread ground-dove species is needed to explore these patterns.

Columbicola passerinae 2, however, does not conform as strictly to this phylogeographic pattern. Although there is some phylogeographic structure within the lineage, a louse from trans-Andean Peru (ex. *Co. buckleyi*) is imbedded within lice from

Brazil and Mexico (ex. *Co. talpacoti* and *Cl. pretiosa*), thus breaking up the conserved structure. The presence of both cis- and trans-Andean representatives within *C. passerinae* 2 indicates the lice are able to move around or across the Andes. Given the range restrictions of *Co. buckleyi* and *Co. cruziana*, perhaps the lice are using *Cl. pretiosa* as a "bridge species." *Claravis pretiosa* has a range that extends to both sides of the Andes and can move considerably during certain seasons (Gibbs et al., 2001; Piratelli and Blake, 2006). Although they are more arboreal and tend to forage in pairs, they are also known to forage at forest edges in open scrub close to other ground-dove species, including *Co. cruziana* and *Co. buckleyi* (Skutch, 1959; Parker et al., 1995; pers. obs.). *Claravis pretiosa* are not present in the high Andes, so wing lice in the *Co. passerinae* 2 clade are likely moving north around the Andes and into eastern South America via *Cl. pretiosa.* The lice are then able to switch to other hosts in that region of the continent. Interestingly, *Co. talpacoti* is the only other host species from east of the Andes with lice in *Co. passerinae* 2. It is possible that *C. passerinae* 2 is rare on eastern dove species, and that more thorough sampling will reveal other ground-dove species hosting this wing louse lineage. It is also possible that *C. passerinae* 2 are preferentially switching from *Cl. pretiosa* to *Co. talpacoti.* Alternatively, the existence of *C. passerinae* 2 on *Co. talpacoti* could be a reflection of the recent evolutionary history of this species. *Co. talpacoti* and *Co. buckleyi* are sister taxa that diverged very recently, with *Co. buckleyi* originally considered a subspecies of *Co. talpacoti* (Meyer de Schauensee, 1970). *Columbina talpacoti* likely shared lice with *Co. buckleyi* before the two species diverged, and could have retained this shared lineage until the present. Because wing lice from *Co. buckleyi* are so similar to lice from *Co. talpacoti*, this could be evidence that the two host species

still have some contact, which would explain why a louse from *Co. buckleyi* (a trans-Andean species) is parasitized by a cis-Andean louse.

Despite less consistent phylogeographic patterns in *C. passerinae* 2, the fact that there is any conserved phylogeographic structure within *C. passerinae* indicates barriers for dispersal. If the primary mode of host-switching/dispersal is via phoresis using parasitic hippoboscid flies, it is also possible the flies have limited ranges due to geographical or ecological barriers. Hippoboscid flies have been recorded from small New World ground-doves throughout the doves' range (Maa, 1969), but if the flies have a restricted range or limited gene flow this would also limit louse dispersal. Additional work focused on the phylogeography of hippoboscid flies is needed to test whether the flies exhibit similar patterns to the lice.

Conclusions

By focusing on a small clade of doves and their associated wing lice (also in a monophyletic group), we were able to sample lice from most host species, including multiple louse samples per host species. This approach allowed us to uncover cophylogenetic and phylogeographic patterns that would be obscured in broader-scale studies, thereby further untangling some of the evolutionary history of this host-parasite system. In particular, the results of this study indicate that biogeography and host life-history are important factors for shaping host-parasite evolutionary patterns, particularly for systems involving permanent parasites. Although permanent parasites are tightly tied to their hosts, host phylogeny is rarely the primary predictor of parasite diversification. Identifying what external factors are promoting parasite diversification is crucial for understanding host-parasite interactions.

FIGURES

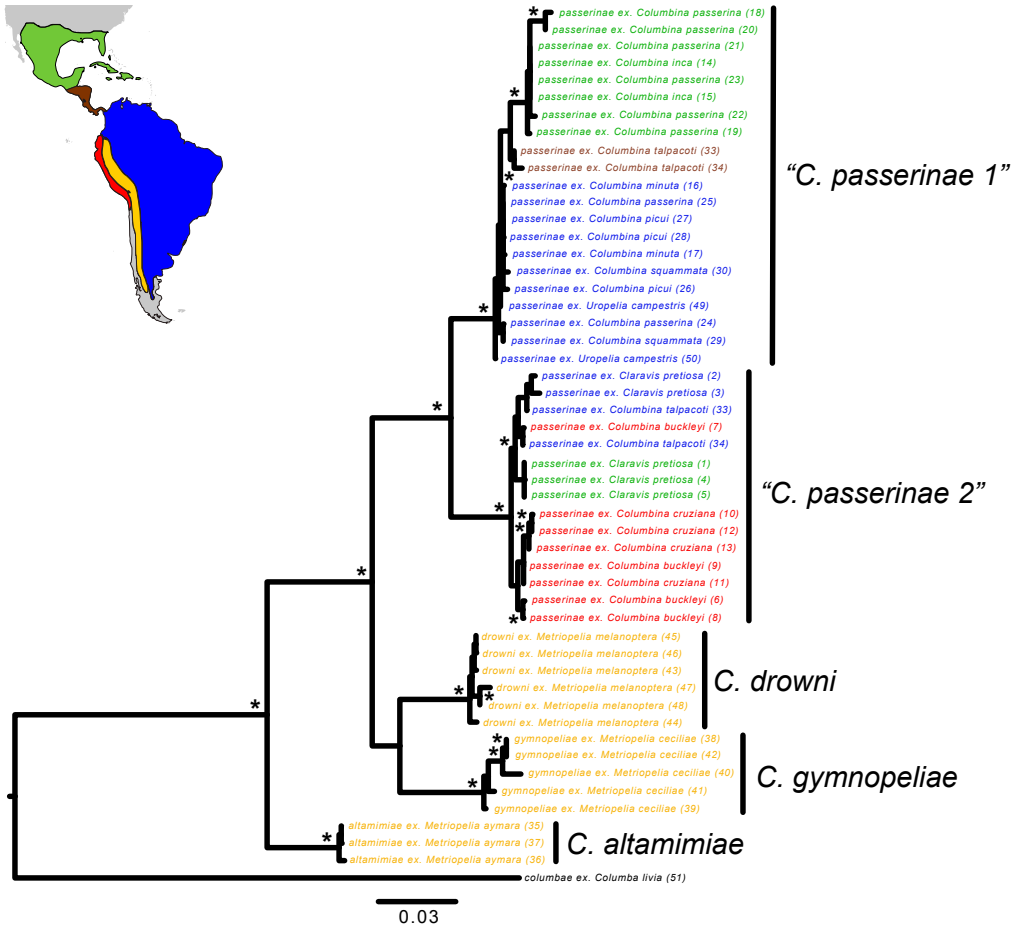


Figure 5.1**.** Maximum likelihood topology of New World ground-dove wing lice *Columbicola*, from a partitioned concatenated analysis of CO1, EF-1α, TMEDE6, and HYP sequences. Values at nodes indicate ML bootstrap (BS) support values followed by Bayesian posterior probabilities (PP). Support is only indicated on nodes with ≥50 BS/≥0.50 PP values, and nodes with ≥95 BS/≥0.95 PP support are indicated with asterisks (*). Scale bar indicates nucleotide substitutions per site. Species as recovered by OTU analysis are indicated to the right of the figure. Individual samples are colored according to the geographic region where the sample was collected, as indicated on the map.

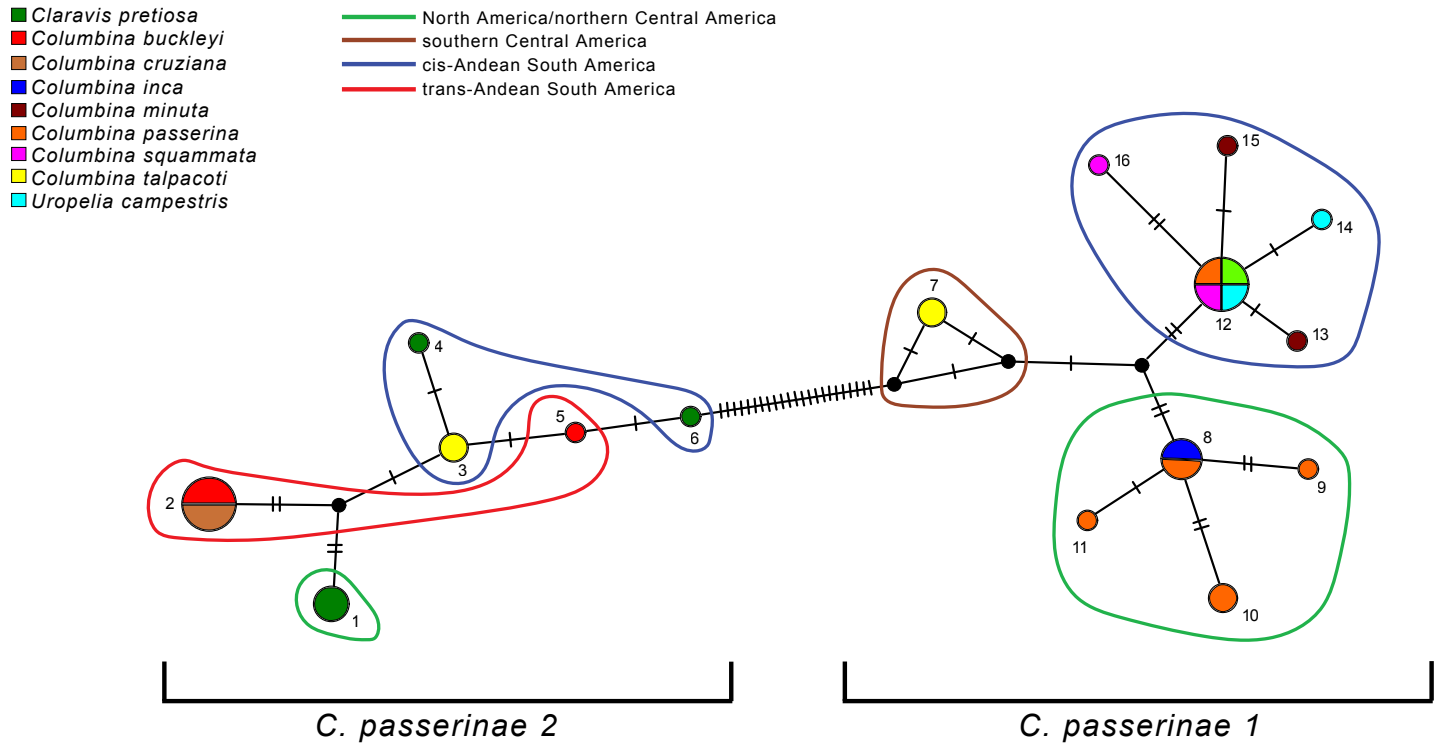


Figure 5.2. Median-joining (MJ) network estimated from CO1 sequences of *Columbicola passerinae* samples. The two distinct lineages within *C. passerinae* are indicated at the bottom of the figure. Nodes correspond to unique haplotypes, and are numbered according to Table 5.1. The size of each node indicates the relative number of individuals in each haplotype. Nodes are colored according to host species as indicated in the upper-left of the figure. Inferred ancestral nodes are colored black. Haplotypes are grouped with colored lines according to geographic region of sampling, and are colored as in Figure 5.1.

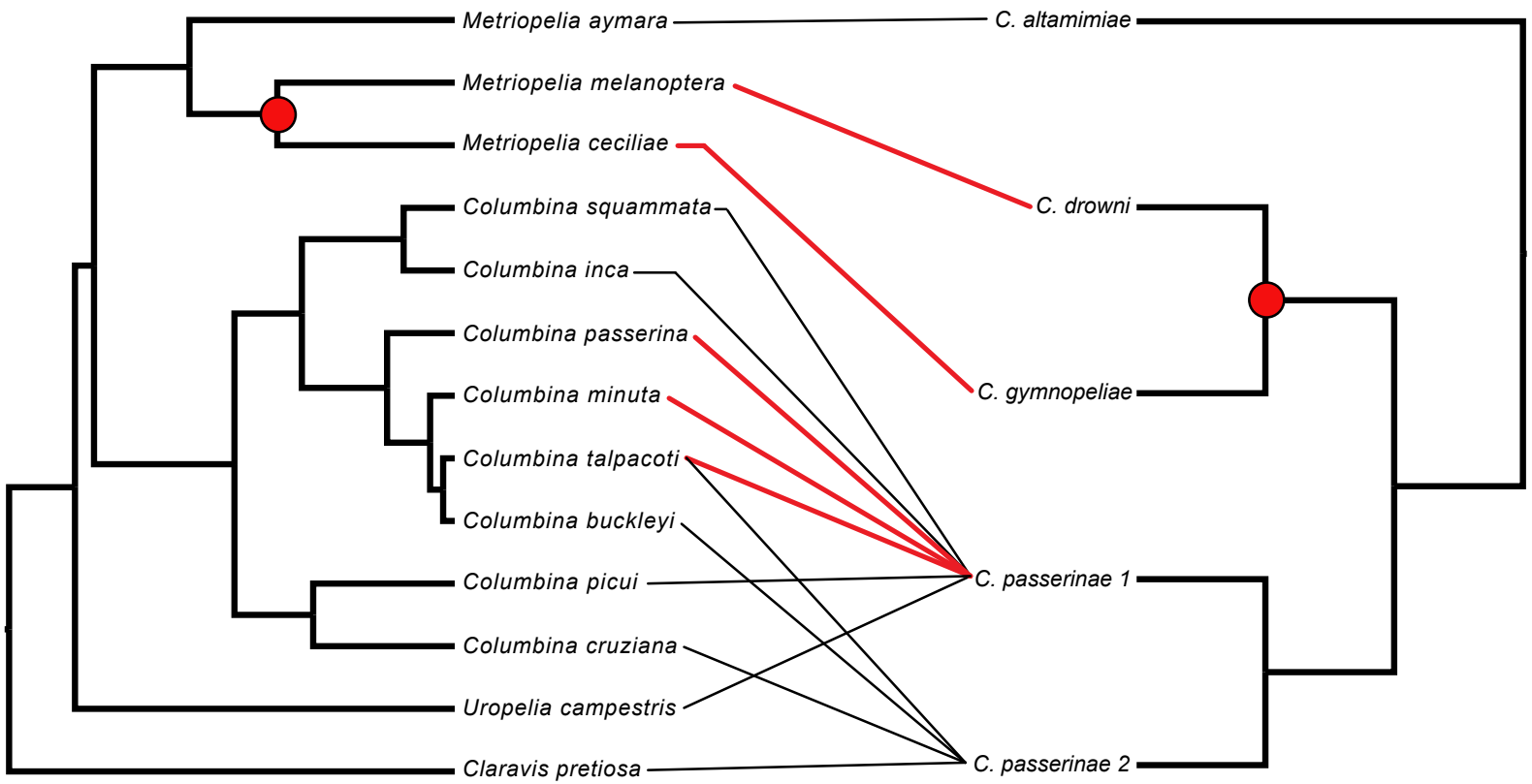


Figure 5.3. Tanglegram of small New World ground-doves and their associated *Columbicola* wing lice. Red circles indicate a cospeciation event as recovered by Jane. Red lines indicate significant host-parasite links estimated by the ParaFitLink1 test after correcting for multiple tests.

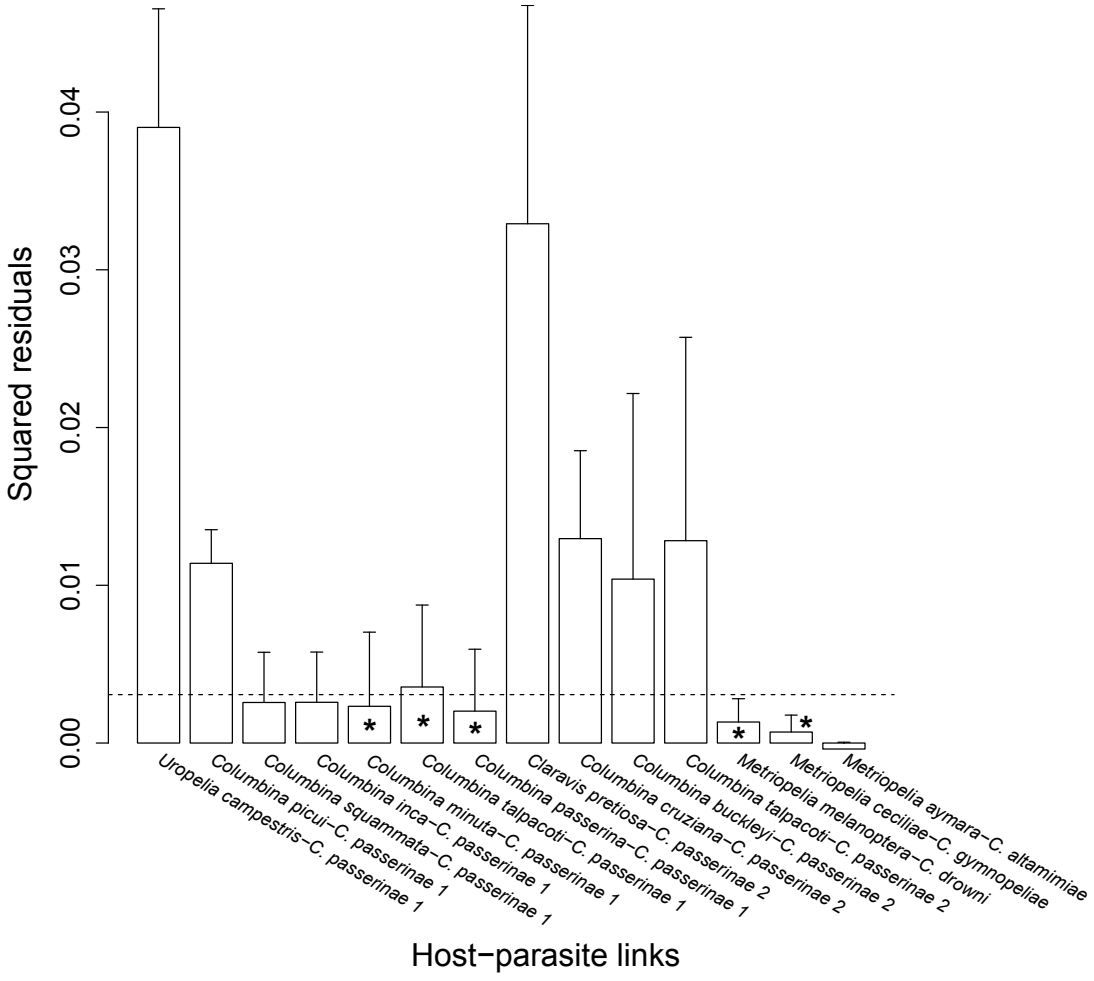


Figure 5.4. Squared residuals from the PACo jackknife test for individual host-parasite links. Median squared residual value indicated with the horizontal dotted line. Host-parasite associations that were significant according to the corrected ParaFitLink1 test results are indicated by asterisks (*).

TABLES

Table 5.1. *Columbicola* samples used in this study. Sequence data indicated by the associated GenBank accession numbers. Dashes (-) indicate missing data. Haplotypes for *C. passerinae* samples refer to Figure 5.2.

Number	*Columbicola* species	Haplotype	Voucher	Host	Locality	Host Voucher	CO1	Ef1a	HYP	TMEDE6
1	*passerinae* 2	1	Copsr.9.29.1998.3	*Claravis pretiosa*	Mexico	CO-23	KX528475	KX528509	-	-
2	*passerinae* 2	6	Cosp.Clpre.11.24.2014.10	*Claravis pretiosa*	Brazil	6413	KX528476	-	KX532199	-
3	*passerinae* 2	4	Cosp.Clpre.11.24.2014.9	*Claravis pretiosa*	Brazil	5178	KX528477	-	KX532200	-
4	*passerinae* 2	1	Cosp.Clpre.2.1.1999.2	*Claravis pretiosa*	Mexico	CO-14	AF414733[1]	-	-	-
5	*passerinae* 2	1	Cosp.Clpre.2.1.1999.6	*Claravis pretiosa*	Mexico	CO-23	KX528478	KX528511	-	-
6	*passerinae* 2	2	Copas.Cobuc.9.4.2013.10	*Columbina buckleyi*	Peru	REW169	KX528479	-	KX532192	KX532221
7	*passerinae* 2	5	Copas.Cobuc.9.4.2013.9	*Columbina buckleyi*	Peru	REW187	KX528466	-	KX532193	KX532222
8	*passerinae* 2	2	Cosp.Cobuc.10.27.2003.4	*Columbina buckleyi*	Peru	REW169	KX528467	-	KX532201	KX532228
9	*passerinae* 2	2	Cosp.Cobuc.7.27.2004.7	*Columbina buckleyi*	Peru	REW187	KX528480	KX528512	-	-
10	*passerinae* 2	2	Cosp.Cocru.10.27.2003.3	*Columbina cruziana*	Peru	REW86	KX528481	-	KX532202	-
11	*passerinae* 2	2	Cosp.Cocru.4.9.2014.4	*Columbina cruziana*	Peru	REW85	KX528482	KX528513	-	KX532229
12	*passerinae* 2	2	Cosp.Cocru.4.9.2014.5	*Columbina cruziana*	Peru	REW86	KX528483	KX528514	KX532203	KX532230
13	*passerinae* 2	2	Cosp.Cocru.7.27.2004.4	*Columbina cruziana*	Peru	REW87	KX528484	KX528515	-	-
14	*passerinae* 1	8	Copsr.9.21.1999.3	*Columbina inca*	USA	I136	AF414727[1]	-	-	-
15	*passerinae* 1	8	Copsr.9.29.1998.6	*Columbina inca*	USA	115	AF414725[1]	KX528510	-	-
16	*passerinae* 1	13	Copas.Comin.9.4.2013.1	*Columbina minuta*	Peru	JLK258	KX528468	-	KX532194	-
17	*passerinae* 1	15	Copas.Comin.9.4.2013.8	*Columbina minuta*	Peru	REW201	KX528469	-	KX532195	KX532223
18	*passerinae* 1	10	Copsr.11.24.2003.1	*Columbina passerina*	USA	299	KX528473	KX528506	KX532198	KX532227
19	*passerinae* 1	11	Copsr.2.1.1999.12	*Columbina passerina*	USA	93, 101	AF414729[1]	-	-	-
20	*passerinae* 1	10	Copsr.2.1.1999.3	*Columbina passerina*	Mexico	CO-9	AF414727[1]	-	-	-

Table 5.1 Continued.

21	*passerinae* 1	8	Copsr.9.14.1999.7	*Columbina passerina*	USA	G181	KX528474	KX528507	-	-
22	*passerinae* 1	9	Copsr.9.21.1999.1	*Columbina passerina*	USA	G280	AF414731[1]	-	-	-
23	*passerinae* 1	8	Copsr.9.29.1998.2	*Columbina passerina*	Mexico	CO-11	AF414728[1]	KX528508	-	-
24	*passerinae* 1	12	Cosp.Copas.11.24.2014.7	*Columbina passerina*	Brazil	5049	KX528486	-	KX532204	-
25	*passerinae* 1	12	Cosp.Copas.11.24.2014.8	*Columbina passerina*	Brazil	5051	KX528487	-	-	-
26	*passerinae* 1	12	Copas.Copic.9.4.2013.4	*Columbina picui*	Argentina	KGM459	KX528470	-	-	KX532224
27	*passerinae* 1	12	Cosp.Copic.1.20.2003.7	*Columbina picui*	Argentina	KGM292	KX528488	KX528516	-	-
28	*passerinae* 1	12	Cosp.Copic.1.8.2003.3	*Columbina picui*	Argentina	KGM459	KX528489	KX528517	-	-
29	*passerinae* 1	12	Cosp.Cosqu.11.24.2014.1	*Columbina squammata*	Brazil	4073	KX528490	-	KX532205	-
30	*passerinae* 1	16	Cosp.Cosqu.11.24.2014.4	*Columbina squammata*	Brazil	4781	KX528491	-	-	-
31	*passerinae* 1	7	Copas.Cotal.4.9.2014.10	*Columbina talpacoti*	Panama	JMD732	KX528471	KX528504	KX532196	KX532225
32	*passerinae* 1	7	Copas.Cotal.4.9.2014.9	*Columbina talpacoti*	Panama	GMS1870	KX528472	KX528505	KX532197	KX532226
33	*passerinae* 2	3	Cosp.Cotal.11.24.2014.11	*Columbina talpacoti*	Brazil	6471	KX528492	KX528518	KX532206	-
34	*passerinae* 2	3	Cosp.Cotal.11.24.2014.12	*Columbina talpacoti*	Brazil	6473	KX528493	-	KX532207	-
35	*altamimiae*		Coalt.Meayr.9.4.2013.2	*Metriopelia aymara*	Argentina	KGM431	KX528463	KX528503	KX532188	KX532218
36	*altamimiae*		Cosp.Meaym.4.9.2014.12	*Metriopelia aymara*	Argentina	KGM1148	KX528494	KX528519	KX532208	KX532231
37	*altamimiae*		Cosp.Meaym.4.9.2014.7	*Metriopelia aymara*	Peru	REW291	KX528495	KX528520	KX532209	-
38	*gymnopeliae*		Cogym.10.5.1999.12	*Metriopelia ceciliae*	Peru	RCF952	AY151009[1]	AY151021[1]	-	-
39	*gymnopeliae*		Cogym.Mecec.9.4.2013.12	*Metriopelia ceciliae*	Peru	REW150	KX528464	-	KX532190	KX532219
40	*gymnopeliae*		Cogym.Mecec.9.4.2013.7	*Metriopelia ceciliae*	Peru	REW192	KX528465	-	KX532191	KX532220
41	*gymnopeliae*		Cosp.Mecec.4.9.2014.3	*Metriopelia ceciliae*	Peru	REW153	KX528496	-	KX532210	KX532232
42	*gymnopeliae*		Cosp.Mecec.7.27.2004.5	*Metriopelia ceciliae*	Peru	REW192	EF679008[1]	EF679143[1]	KX532211	-
43	*drowni*		Coalt.1.8.2003.4	*Metriopelia melanoptera*	Argentina	KGM511	EF679018[1]	EF679153[1]	KX532187	-
44	*drowni*		Cosp.Memel.1.8.2003.2	*Metriopelia melanoptera*	Argentina	KGM514	EF679017[1]	EF679152[1]	KX532212	-
45	*drowni*		Cosp.Memel.4.9.2014.6	*Metriopelia melanoptera*	Peru	REW282	KX528497	KX528521	KX532213	KX532233

Table 5.1. Continued.

46	*drowni*		Cosp.Memel.4.9.2014.8	*Metriopelia melanoptera*	Peru	REW319	KX528498	-	KX532214	KX532234
47	*drowni*		Cosp.Memel.9.4.2013.3	*Metriopelia melanoptera*	Argentina	KGM444	KX528499	-	KX532215	KX532235
48	*drowni*		Cosp.Memel.9.4.2013.6	*Metriopelia melanoptera*	Bolivia	KGM512	KX528500	-	KX532216	KX532236
49	*passerinae* 1	12	Cosp.Urcam.10.12.1999.5	*Uropelia campestris*	Bolivia	CCW925	KX528501	KX528522	KX532217	-
50	*passerinae* 1	14	Cosp.Urcam.11.24.2014.5	*Uropelia campestris*	Brazil	4789	KX528502	-	-	-
51	*columbae*		Cocol.6.29.1998.3	*Columba livia*		4 Col	EF678962[1]	EF679097[1]	KX532189	KX532237

[1] Previously published sequence

Table 5.2. Results from ParaFit analysis on ground-doves and their wing lice. The global test results and each individual link test results are indicated. For the individual link test, values with an asterisk (*) indicate significance after the Benjamini-Hochberg correction ($\alpha = 0.05$).

ParaFitGlobal = 20.663	**P-value = 0.003**	
Host	***Columbicola* species**	**ParaFitLink1 P-values**
Claravis pretiosa	*C. passerinae* 2	0.960
Columbina buckeyi	*C. passerinae* 2	0.057
Columbina cruziana	*C. passerinae* 2	0.200
Columbina inca	*C. passerinae* 1	0.024
Columbina minuta	*C. passerinae* 1	0.014*
Columbina passerinae	*C. passerinae* 1	0.016*
Columbina picui	*C. passerinae* 1	0.115
Columbina squammata	*C. passerinae* 1	0.024
Columbina talpacoti	*C. passerinae* 1	0.013*
Columbina talpacoti	*C. passerinae* 2	0.056
Metriopelia aymara	*C. altamimiae*	0.031
Metriopelia ceciliae	*C. gymnopeliae*	0.015*
Metriopelia melanoptera	*C. drowni*	0.011*
Uropelia campestris	*C. passerinae* 1	0.966

Table 5.3. Summary of the Jane v4 results for New World ground-doves and their wing lice. Cospeciation event indicated in Figure 5.3.

Cospeciations	**Duplications**	**Duplications and Host Switches**	**Losses**	**Failures to Diverge**
1	3	0	15	9

LITERATURE CITED

Balbuena J.A., Míguez-Lozano R., Blasco-Costa I. 2013. PACo: A Novel Procrustes Application to Cophylogenetic Analysis. *PLoS ONE* 8(4): e61048. doi:10.1371/journal.pone.0061048

Bandelt H.J., Forster P., Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48. doi:10.1093/oxfordjournals.molbev.a026036

Bartholomew J.L., Whipple M.J., Stevens D.G., Fryer J.L. 1997. The life cycle of *Ceratomyxa Shasta*, a Myxospore parasite of salmonids, requires a freshwater polychaete as an alternate host. *Journal of Parasitology* 83(5): 859-868.

Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistic Society* 57: 289-300.

Bruyndonckx N., Dubey S., Ruedi M., Christe P. 2009. Molecular cophylogenetic relationships between European bats and their ectoparasitic mites (Acari, Spinturnicidae). *Molecular Phylogenetics and Evolution* 51(2): 227–237. doi:10.1016/j.ympev.2009.02.005

Bush S.E., Weckstein J.D., Gustafsson D.R., Allen J., Diblasi E., Shreve S.M., Boldt R., Skeen H.R., Johnson K. P. 2016. Unlocking the black box of feather louse diversity: A molecular phylogeny of the hyper-diverse genus *Brueelia*. *Molecular Phylogenetics and Evolution* 94: 737–751. doi:10.1016/j.ympev.2015.09.015

Charleston M.A., Robertson D.L. 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Systematic Biology* 51: 528–535. doi:10.1080/10635150290069940

Clayton D.H., Bush S.E., Goates B.M., Johnson K.P. 2003. Host defense reinforces host-parasite cospeciation. *Proceedings of the National Academy of Sciences* 100: 15694–15699.

Clayton D.H., Drown D.M. 2001. Critical Evaluation of Five Methods for Quantifying Chewing Lice (Insecta: Phthiraptera). *Journal of Parasitology* 87: 1291–1300.

Clayton D.H., Johnson K.P. 2003. Linking coevolutionary history to ecological process: doves and lice. *Evolution* 57: 2335–41.

Clayton D.H., Koop J.A.H., Harbison C.W., Moyer B.R., Bush S.E. 2010. How birds combat ectoparasites. *Open Ornithology Journal* 3: 41–71. doi:10.2174/1874453201003010041

Clayton D.H., Price R.D. 1999. Taxonomy of New World *Columbicola* (Phthiraptera: Philopteridae) from the Columbiformes (Aves), with descriptions of five new species. *Annals of the Entomological Society of America*. 92(5): 675-685.

Colinet D., Deleury E., Anselme C., Cazes D., Poulain J., Azema-Dossat C., Belghazi M., Gatti J., Poirié M. 2013. Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: The case of Leptopilina parasitoids of Drosophila. *Insect Biochemistry and Molecular Biology* 43(7): 601–611. doi:10.1016/j.ibmb.2013.03.010

Conow C., Fielder D., Ovadia Y., Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms in Molecular Biology* 5: 16. doi:10.1186/1748-7188-5-16

Danforth B.N., Ji S. 1998. Elongation factor-1 alpha occurs as two copies in bees: implications for phylogenetic analysis of EF-1 alpha sequences in insects. *Molecular Phylogenetics and Evolution* 15(3): 225–235. doi:10.1093/oxfordjournals.molbev.a025920

de Vienne D.M., Refregier G., Lopez-Villavicencio M., Tellier A., Hood M.E., Giraud T. 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist* 198: 347-385.

de Vienne D.M., Giraud T., Shykoff J.A. 2007. When can host shifts produce congruent host and parasite phylogenies? A simulation approach. *Journal of Evolutionary Biology* 20(4): 1428–1438. doi:10.1111/j.1420-9101.2007.01340.x

Detwiler J.T., Bos D.H., Minchella D.J. 2010. Revealing the secret lives of cryptic species: Examining the phylogenetic relationships of echinostome parasites in North America. *Molecular Phylogenetics and Evolution* 55(2): 611–620. doi:10.1016/j.ympev.2010.01.004

Dhami M.K., Buckley T.R., Beggs J.R., Taylor M.W. 2013. Primary symbiont of the ancient scale insect family Coelostomidiidae exhibits strict cophylogenetic patterns. *Symbiosis* 61(2): 77–91. doi:10.1007/s13199-013-0257-8

Dias R.I. 2006. Effects of position and flock size on vigilance and foraging behavior of the scaled dove *Columbina squammata*. *Behavioral Processes* 73(3): 248-252.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Phylogenetics and Evolution* 29(8): 1969–73. doi:10.1093/molbev/mss075

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5): 1792–7. doi:10.1093/nar/gkh340

Eichler W., 1953. Mallophagen. In Niethammer, G. ed. *Zur Vogelwelt Boliviens* (pp. 195–303). Bonn, Germany: Bonn Zoological Bulletin.

Eichler W. 1948. Some rules in ectoparasitism. *Annual Magazine of Natural History* 1: 588–598.

Ezard T., Fujisawa T., Barraclough T. 2014. splits: SPecies' LImits by Threshold Statistics. Version 1.0-10/r51.

Fahrenholz H. 1913. Ectoparasiten und abstammungslehre. *Zoologischer Anzeiger* 41: 371–374.

Feder J.L., Hunt T.A., Bush L. 1993. Genetics of apple and hawthorn infesting races of *Rhagoletis pomonell*. *Entomologia Experimentalis et Applicata* 69: 117–135.

Fjeldså J., Bowie R.C. K., Rahbek C. 2012. The role of mountain ranges in the diversification of birds. *Annual Review of Ecology, Evolution, and* Syste*matics* 43(1): 249–265. doi:10.1146/annurev-ecolsys-102710-145113

Fujisawa T., Barraclough T.G. 2013. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology* 62(5): 707–724. doi:10.1093/sysbio/syt033

Garzione C.N., Auerbach D.J., Smith J.J., Rosario J.J., Passey B.H., Jordan T.E., Eiler J.M. 2014. Clumped isotope evidence for diachronous surface cooling of the

Altiplano and pulsed surface uplift of the Central Andes. *Earth and Planetary Science Letters* 393: 173-181.

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

Goodwin D. 1983. Pigeons and Doves of the World (3rd ed.). Ithaca, New York: Cornell University Press.

Hafner M.S., Page R.D. 1995. Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 349(1327): 77–83. doi:10.1098/rstb.1995.0093

Hafner M.S., Sudman P.D., Villablanca F.X., Spradling T.A., Demastes J.W., Nadler S.A. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265(5175): 1087–90.

Harbison C.W., Bush S.E., Malenke J.R., Clayton D.H. 2008. Comparative transmission dynamics of competing parasite species. *Ecology* 89(11): 3186–3194.

Harbison C.W., Clayton D.H. 2011. Community interactions govern host-switching with implications for host-parasite coevolutionary history. *Proceedings of the National Academy of Sciences* 108(23): 9525–9. doi:10.1073/pnas.1102129108

Johnson K.P., Adams R.J., Clayton D.H. 2001. Molecular systematics of Goniodidae (Insecta: Phthiraptera). *Journal of Parasitology* 87(4): 862–869.

Johnson K.P., Adams R.J., Page R.D.M., Clayton D.H. 2003. When do parasites fail to speciate in response to host speciation? *Systematic Biology* 52(1): 37-47.

Johnson K.P., Allen J.M., Olds B.P., Mugisha L., Reed D.L., Paige K.N., Pittendrigh B.R. 2014. Rates of genomic divergence in humans, chimpanzees and their lice. *Proceedings of the Biological Sciences* 281(1777): 20132174.

Johnson K.P., Clayton D.H. 2000. A molecular phylogeny of the dove genus *Zenaida*: Mitochondrial and nuclear DNA sequences. The Condor, 102(4): 864–870.

Johnson K.P., Reed D.L., Hammond Parker S.L., Kim D., Clayton D.H. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for *Columbicola* (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 45(2): 506–18. doi:10.1016/j.ympev.2007.07.005

Johnson K.P., Williams B.L., Drown D.M., Adams R.J., Clayton D.H. 2002. The population genetics of host specificity: genetic differentiation in dove lice (Insecta: Phthiraptera). *Molecular Ecology* 11(1): 25–38.

Johnson K.P., Bush S.E., Clayton D.H. 2005. Correlated evolution of host and parasite body size: tests of Harrison's Rule using birds and lice. *Evolution* 59(8): 1744–1753.

Johnson K.P., Clayton D.H. 2004. Untangling coevolutionary history. *Systematic Biology* 53(1): 92–94. doi:10.1080/10635150490264824

Jousson O., Bartoli P., Pawlowski J. 2000. Cryptic speciation among intestinal parasites (Trematoda: Digenea) infecting sympatric host fishes (Sparidae). Journal of Evolutionary Biology 13(5): 778–785. doi:10.1046/j.1420-9101.2000.00221.x

Krasnov B.R., Shenbrot G.I., Medvedev S.G., Vatschenok V.S., Khokhlova I.S. 1997. Host-habitat relations as an important determinant of spatial distribution of flea assemblages (Siphonaptera) on rodents in the Negev Desert. *Parasitology* 114: 159–173. doi:10.1017/S0031182096008347

Lafferty K.D., Kuris A.M. 2002. Trophic strategies, animal diversity and body size. *Trends in Ecology and Evolution* 17(11): 507–513. doi:10.1016/S0169-5347(02)02615-0

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substituion models for phylogenetic analyses. *Molecular Biology and Evolution* 29(6): 1695–1701.

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

Leigh J.W., Bryant D. 2015. Popart: Full-feature software for haplotype network construction. Method Ecol. Evol., 6(9): 1110–1116. doi:10.1111/2041-210X.12410

Lloyd H., Sevillano Ríos S., Marsden S.J., Valdés-Velásquez A. 2010. Bird community composition across an Andean tree-line ecotone. *Austral Ecology* 37(4): 470–478. doi:10.1111/j.1442-9993.2011.02308.x

Maa T.C. 1969. A revised checklist and concise host index of hippoboscidae (Diptera). *Pacific Insect Monographs* 20: 261-299.

Maddison W.P., Slatkin M. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* 45(5): 1184–1197.

Marhsall A. G. 1981. The ecology of ectoparasitic insects. Academic Press.

Martínez-Aquino A. 2016. Phylogenetic framework for coevolutionary studies: a compass for exploring jungles of tangled trees. *Current Zoology* 1-11. doi:10.1093/cz/zow018

Meyer de Schauensee R., Mack A.L. 1970. A guide to the birds of South America. Livingston Publishing Company.

Miura O., Kuris A.M., Torchin M.E., Hechinger R.F., Dunham E.J., Chiba S. 2005. Molecular-genetic analyses reveal cryptic species of trematodes in the intertidal gastropod, *Batillaria cumingi* (Crosse). *International Journal for Parasitology* 35(7): 793–801. doi:10.1016/j.ijpara.2005.02.014

Morand S., Robert F., and Conners V.A. 1995. Complexity in parasite life cycles: population biology of cestodes in fish. *Journal of Animal Ecology* 64(2): 256-264.

Oksanen J., Guillaume G.B., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos, P., Stevens M.N.H., Wagner H. 2016. vegan: Community ecology package. Version 2.3-4.

Page R.D.M. 1994. Parallel Phylogenies: Reconstructing the History of Host-Parasite Assemblages. *Cladistics* 10, 155–173. doi:10.1111/j.1096-0031.1994.tb00170.x

Page R.D.M. 2003. Tangled trees: phylogeny, cospeciation, and coevolution. Chicago, IL: University of Chicago Press.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2): 289–290. doi:10.1093/bioinformatics/btg412

Parker G.A., Chubb J.C., Roberts G.N., Michaud M., Milinski M. 2003. Optimal growth strategies of larval helminths in their intermediate hosts. Journal of Evolutionary Biology 16(1): 47–54. doi:10.1046/j.1420-9101.2003.00504.x

Parker T.A., Schulenberg T.S., Kessler M., Wust W.H. 1995. Natural history and conservation of the endemic avifauna in north-west Peru. *Bird Conservation International* 5: 201-231. doi:10.1017/S0959270900001015

Pereira S. L., Johnson K.P., Clayton D.H., Baker A.J. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-

driven radiation in the Paleocene. *Systematic Biology* 56(4): 656–672. doi:10.1080/10635150701549672

Pérez-Escobar O.A., Balbuena J.A., Gottschling M. 2015. Rumbling orchids: How to assess divergent evolution between chloroplast endosymbionts and the nuclear host. *Systematic Biology* 65(1): 51-65.

Piratelli A., Blake J.G., 2006. Bird communities of the southeastern Cerrado Region, Brazil. *Ornitologia Neotropical* 17(2): 213–225.

Poulin R. 2011. Evolutionary ecology of parasites (2nd ed.). Princeton University Press.

Poulin R. 1997. Species richness of parasite assemlages: evolution and patterns. Annual *Review of Ecology and Systematics* 28: 341–358.

Poulin R., Morand S. 2000. The diversity of parasites. *Quarterly Review of Biology* 75(3): 277–293.

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

Prugnolle F., Durand P., Neel C., Ollomo B., Ayala F. J., Arnathau C., Etienne L., Mpoudi-Ngole E., Nkoghe D., Leroy E., Delaporte E., Peeters M., Renaud F. 2010. African great apes are natural hosts of multiple related malaria species, including Plasmodium falciparum. *Proceedings of the National Academy of Sciences* 107(4): 1458–1463. doi:10.1073/pnas.0914440107

Puillandre N., Lambert A., Brouillet S., Achaz G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21(8): 1864–1877. doi:10.1111/j.1365-294X.2011.05239.x

R Development Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rambaut A., Drummond, A.J. 2007. Tracer v1.4.

Refrégier G., Le Gac M., Jabbour F., Widmer A., Shykoff J.A, Yockteng R., Hood M.E., Giraud T. 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 Evolutionary Biology* 8: 100. doi:10.1186/1471-2148-8-100

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics Applications Note* 19(12): 1572–1574.

Shapiro B., Sibthorpe D., Rambaut A., Austin J., Wrag G. M., Biminda-Emonds O. R. P., Lee, P.L.M., Cooper, A. 2002. Flight of the Dodo. *Science* 295: 1683.

Skutch A.F. 1959. Life history of the blue ground dove. *The Condor* 61(2): 65–74.

Sorenson M.D., Balakrishnan C.N., Payne R.B. 2004. Clade-limited colonization in brood parasitic finches (*Vidua* spp.). *Systematic Biology* 53(1): 140–153. doi:10.1080/10635150490265021

Sugiura N. 1978. Further analysts of the data by Akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods* 7(1): 13–26. doi:10.1080/03610927808827599

Sukumaran J., Holder M. 2008. SumTrees: Summarization of split support on phylogenetic trees. Version 1.0.2.

Sweet A.D., Allen J.M., Johnson K.P. 2014. Novel primers from informative nuclear loci for louse molecular phylogenetics (Insecta : Phthiraptera*). Journal of Medical Entomology* 51(6): 1122–1126.

Sweet A.D., Boyd B.M., Johnson K.P. 2016. Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. *Biological Journal of the Linnean Society* doi:10.1111/bij.12771

Sweet A.D., Johnson K.P. 2015. Patterns of diversification in small New World ground doves are consistent with major geologic events. *The Auk* 132(1): 300–312. doi:10.1642/AUK-14-193.

Szumik C., Aagesen L., Casagranda D., Arzamendia V., Giacomo D., Giraudo A., goloboff P., Gramajo C., Kopuchian C., Kretzschmar S., Lizarralde M., Molina A., Mollerach M., Navarro F., Nomdedeu S., Panizza A., Pereyra V., Sandoval M., Scrocchi G., Zuolaga F. O. 2012. Detecting areas of endemism with a taxonmically diverse data set: plants, mammals, reptiles, amphibians, birds, and insects from Argentina. *Cladistics* 28: 317–329.

Warren D., Geneva A., Lanfear R. 2016. rwty: R We There Yet? Visualizing MCMC convergence in phylogenetics. R package version 1.0.0.

Weckstein J. D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology* 53(1): 154–164. doi:10.1080/10635150490265085

CHAPTER 6: INTEGRATING PHYLOGENOMIC AND POPULATION GENOMIC PATTERNS IN AVIAN LICE PROVIDES A MORE COMPLETE PICTURE OF PARASITE EVOLUTION[5]

INTRODUCTION

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

[5] Reprinted, with permission, from: Sweet, A.D., Boyd, B.M., Allen, J.M., Villa, S.M., Valim, M.P., Rivera-Parra, J.L., Wilson, R.E., and Johnson, K.P. 2018. Integrating phylogenomic and population genomic patterns in avian lice provides a more complete picture of parasite evolution. *Evolution*. 72(1): 95-112. http://doi:10.1111/evo.13386.

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, over-splitting 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 non-model 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 (Price et al., 2003, Clayton and Johnson, 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). 6 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 (Supplementary Table 6.1). Three additional samples were sequenced on a single lane to obtain high-coverage genomes for methods development and assessing error (sequencing pool 4, Supplementary Table 6.1). The pooled libraries were sequenced with 161 cycles on an Illumina HiSeq2500 instrument using the HiSeq SBS v4 sequencing kit, resulting in 160nt paired-end reads. Fastq files were generated from the sequence data using Casava v1.8.2 or bcltofastq v1.8.4 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 v0.10.1 (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, Supplementary Table 6.1). 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 1,107 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 post-processing pipeline from Allen et al. (2017) to identify exons. The pipeline uses Exonerate v2.2.0 (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 v8.1.2 (Biomatters, Ltd.), and chose the minimum library fraction with uniform coverage ≥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 (Supplementary Table 6.1).

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 ≥90% gaps using trimAL v1.4 (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 v2.1.1 (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 v8.1.3 (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 v4.10.6 (Mirarab and Warnow, 2015; Sayyari and Mirarab, 2016) with quartet-based local posterior probability support for branches. We also summarized gene trees with ASTRID v1.4 (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 v2.4.4 (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 1,000 generations. We used a Yule tree prior, default substitution priors, and a strict clock model for branch length estimation. We used Tracer v1.5 (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 post-burn-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 over-splitting. 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 v4 (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 1,000, and used default event costs (0 cospeciation, 1 duplication, 2 duplication and host switch, 1 loss, and 1 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 1,000 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 v0.6.94 (Earl and VonHoldt, 2012). We summarized the runs using the *greedy* algorithm in CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007), and used the output from CLUMPP to construct STRUCTURE plots using distruct v1.1 (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 the *P. eurysema* 3. We subsampled the SNPs for *P. eurysema* 3, and also filtered out missing and monomorphic SNPs, with vcftools v0.1.14 (Danecek et al., 2011).

Finally, we constructed a Median-Joining (MJ) network in PopART v1.7 (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 Supplementary Table 6.1) 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 (Supplementary Table 6.1). The *Physconelloides emersoni* sample subsequently used as an assembly reference had 48,122,466 reads and an estimated sequencing depth of 38X after cleanup (Sample 1, Supplementary Table 6.1). Sequencing between 8 to 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 (Supplementary Table 6.1). BLAST searches on COI data assembled for each sample with aTRAM indicated all samples were not cross-contaminated.

Data assembly and snp calling

aTRAM assembled 1,095 nuclear loci from the high coverage *P. emersoni* genome library using 1,107 *Pediculus humanus humanus* (human body louse) reference loci. For 46% of the assembled loci, aTRAM assembled greater than 90% of the target sequence length. 71% of the loci retained greater than 50% of the target length, and all loci retained more than 10% of the target length (Supplementary Table 6.1). Thirty-seven loci were removed from the reference set based on the reciprocal-best-BLAST test, leaving 1,058 assembled loci as a reference set for subsequent reference-based mapping. Using the 1,058 target loci as references, Bowtie2 assembled an average of 1,055 orthologous loci for each high and low coverage sample (Supplementary Table 6.1). 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 median library fraction of 6.25% for the 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 (Figure 6.1). 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 (Supplementary Figures 6.1- 6.2).

The concatenated mitochondrial alignment was 9,121 bp long. The RAxML analysis on this alignment estimated a generally well-supported tree (Supplementary Figure 6.3), with most edges receiving >75 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, Supplementary Figure 6.3-6.4).

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 (Supplementary Figure 6.5). 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 (Supplementary Figure 6.6). Analyses using nuclear data grouped *P. eurysema* 3 from west of the Andes in a very distinctive cluster (Figures 6.1 and 6.2). One oddity in this cluster is an individual louse from *Columbina passerina* sampled from a high elevation site (>2,000 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 (Figure 6.3). 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 (Figure 6.4; Supplementary Figure 6.7). However, the latter cospeciation event had 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 (Supplementary Figure 6.7). 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 6.1). 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 ($t = -3.32$, $P = 0.008$; Supplementary Figure 6.8). 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 ($t = -2.27$, $P = 0.045$; Supplementary Figure 6.9).

Population structure

Population-level analysis of *P. eurysema* indicated significant structure from both nuclear and mitochondrial data. Based on SNP data, STRUCTURE estimated populations that largely corresponded to the major branches in the phylogenetic trees (Figure 6.1, Supplementary Figures 6.1- 6.3). 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 (Figure 6.5A, Supplementary Figure 6.10-6.11). Using all 56,232 SNPs, ADEGENET also estimated an optimal K = 3. The DAPC scatterplot showed clear distinction among all three clusters (Figure 6.5B). DAPC for *P. eurysema* 3 estimated K = 2, and showed distinction between lice from *Claravis pretiosa* and lice from other host species (Supplementary Figure 6.12). PCA plots based on 18,912 SNPs showed further population-level differentiation within *P eurysema* 3, with several distinct clusters of taxa (Figure 6.2). The MJ network of the mitochondrial sequences also showed several well-delimited groups within *P. eurysema* 3, including some differences with the nuclear data (Figure 6.3). 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 (Figure 6.1) and SNP-based (Figure 6.2B) 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, Weckstei,n 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 (Figures 6.1, 6.2B, and 6.3). 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 >1,000 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 Figure 6.3). Likewise, lice from eastern South America group with other lice from the same region (# 5 and 7 in Figure 6.3). 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.

FIGURES

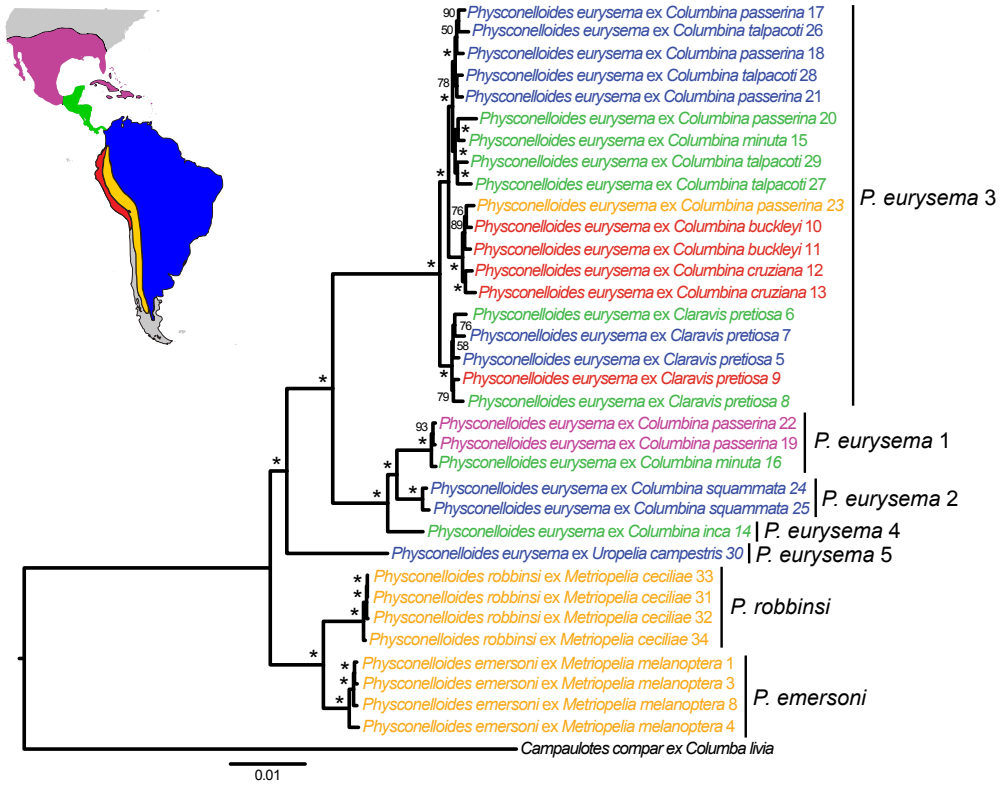


Figure 6.1. Evolutionary history of New World ground-dove body lice (*Physconelloides*) presented as a maximum likelihood phylogeny based on a concatenated sequence alignment of 1,058 nuclear genes. Bootstrap support values are indicated at each node, and asterisks (*) indicate 100% bootstrap support. 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.

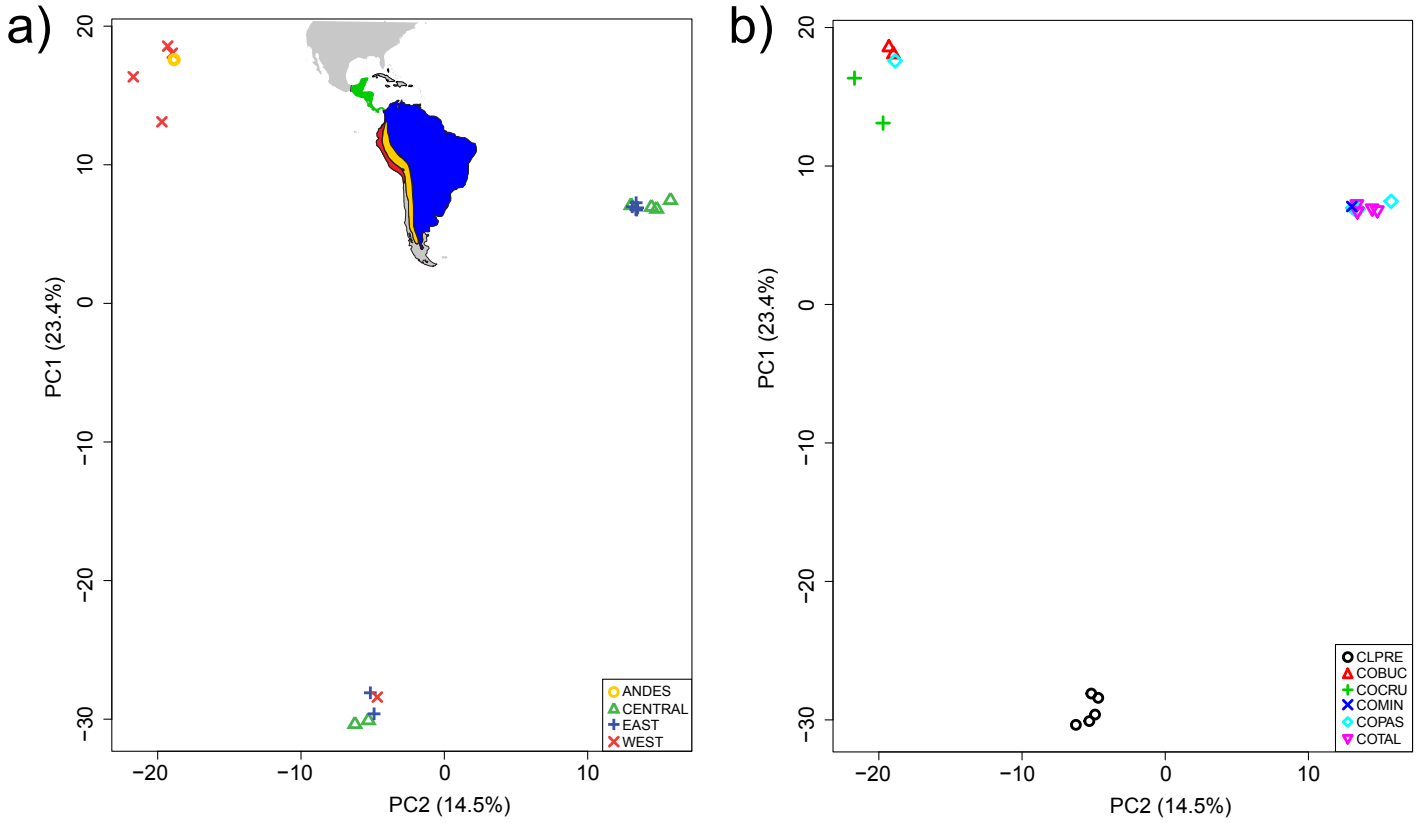


Figure 6.2. Population structure of *Physconelloides eurysema* 3 lice with Principal Component Analysis (PCA) based on 18,912 SNPs. Points represent individual lice, and are colored according to a) biogeographic region or b) host species. Host species codes are as follows: CLPRE = *Claravis pretiosa*, COBUC = *Columbina buckleyi*, COCRU = *Columbina cruziana*, COMIN = *Columbina minuta*, COPAS = *Columbina passerina*, COTAL = *Columbina talpacoti*.

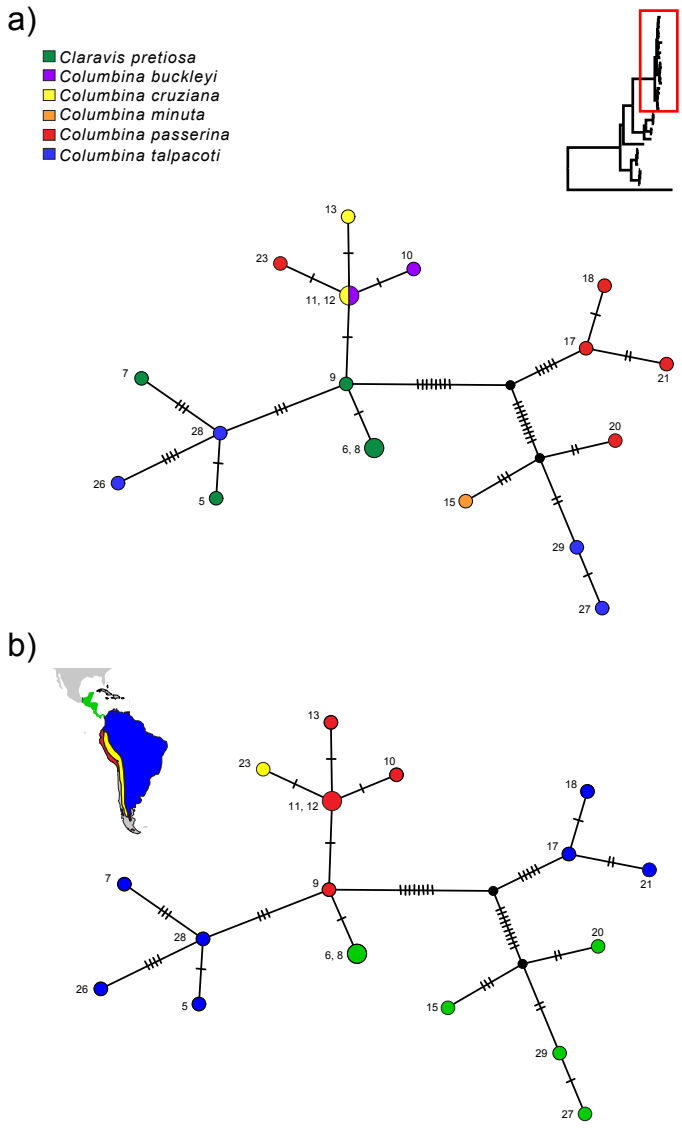


Figure 6.3. Population structure of *Physconeloides eurysema* 3 presented as a median-joining network generated from ten mitochondrial genes. Taxa are indicated with the red box on the phylogeny (from the concatenated nuclear data) in the upper-right. Node size is proportional to the number of individuals in a haplotype. Numbers adjacent to each node represent individuals as indicated in Supplementary Table 6.1. Tick marks indicate the number of steps between haplotypes. Nodes are colored according to a) host species and b) biogeographic region.

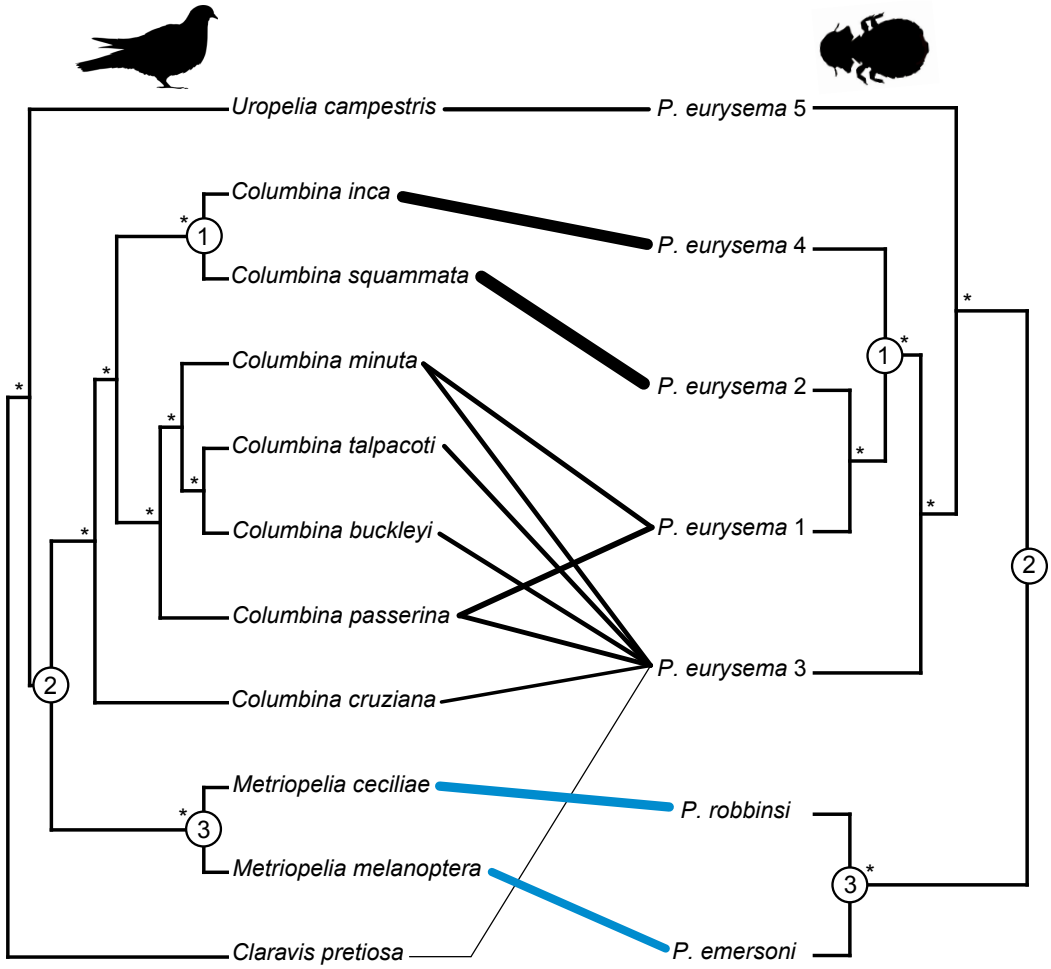


Figure 6.4. Tanglegram comparing the evolutionary histories of small New World ground-doves (left) and their *Physconelloides* body lice (right). The host phylogeny is adapted from Sweet and Johnson (2015). The louse phylogeny is the species tree recovered from OTU analyses. Relationships with significant 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.

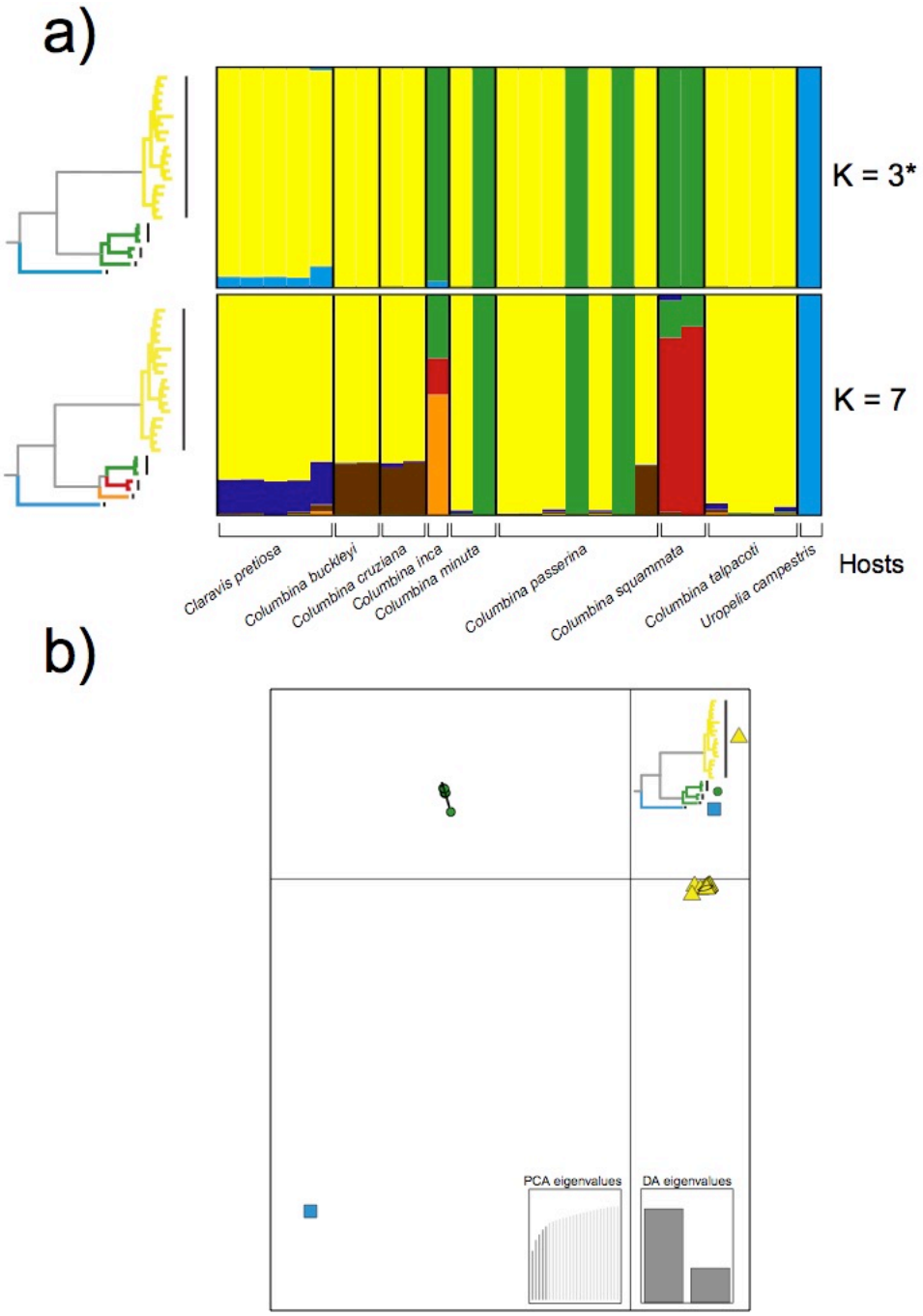


Figure 6.5. Population structure of *Physconelloides eurysema* lice from small New World ground-doves based on genome-wide SNPs. a) STRUCTURE plot from 908 randomly sampled unlinked SNPs and b) Discriminant Analysis of Principal Components (DAPC) plot based on 56,232 SNPs. For the STRUCTURE plot, individual lice are grouped

(Figure 6.5. continued)

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).

TABLES

Table 6.1. Results for the ParaFit analysis for small New World ground-doves and their body lice *Physconelloides*. 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.

Host	Parasite	PF1 Statistic	PF1 P-value	PF2 Statistic	PF2 P-value
Claravis pretiosa	*Physconelloides eurysema* 3	-1.39E-05	0.964	-4.368E-03	0.966
Uropelia campestris	*P. eurysema* 5	1.85E-05	0.014	5.790E-03	0.013
Metriopelia melanoptera	*P. emersoni*	2.58E-05	0.007†	8.073E-03	0.006†
Metriopelia ceciliae	*P. robbinsi*	2.80E-05	0.006†	8.773E-03	0.006†
Columbina cruziana	*P. eurysema* 3	3.43E-06	0.309	1.074E-03	0.306
Columbina squammata	*P. eurysema* 2	8.46E-06	0.084	2.650E-03	0.081
Columbina inca	*P. eurysema* 4	8.11E-06	0.080	2.541E-03	0.076
Columbina minuta	*P. eurysema* 1	9.08E-06	0.078	2.845E-03	0.076
Columbina minuta	*P. eurysema* 3	1.35E-05	0.039	4.240E-03	0.037
Columbina buckleyi	*P. eurysema* 3	1.34E-05	0.046	4.205E-03	0.043
Columbina talpacoti	*P. eurysema* 3	1.35E-05	0.042	4.235E-03	0.039
Columbina passerina	*P. eurysema* 1	8.23E-06	0.099	2.578E-03	0.095
Columbina passerina	*P. eurysema* 3	1.16E-05	0.068	3.642E-03	0.065

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

LITERATURE CITED

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

Allen J.M., Boyd B, Nguyen N.P., Vachaspati P., Warnow T., Huang D.I., Grady P.G., Bell K.C., Cronk Q.C., Mugisha L., Pittendrigh B.R., Soledad Leonardi M., Reed D.L., Johnson K.P. 2017. Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology* 66: 786-798.

Allen J.M., Huang D.I., Cronk Q.C., Johnson K.P. 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., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.

Balbuena J.A., Míguez-Lozano R., Blasco-Costa I. 2013. PACo: A Novel Procrustes Application to Cophylogenetic Analysis. *PLoS One* 8: e61048.

Bell K.C., Calhoun K.L., Hoberg E.P., Demboski J.R., Cook J.A. 2016. Temporal and spatial mosaics: Deep host association and shallow geographic drivers shape genetic structure in a widespread pinworm, *Rauschtineria eutamii* (Nematoda: Oxyuridae). *Biological Journal of the Linnean Society* 119: 397–413.

Benjamini Y., Hochberg Y.. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statatistics Society Series B* 57: 289–300.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A.,

Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Boyd B.M., Allen J.M., Nguyen N., Sweet A.D., Warnow T., Shapiro M.D., Villa S.M., Bush S.E., Clayton D.H., Johnson K.P. 2017. Phylogenomics using Target-restricted Assembly Resolves Intra-generic Relationships of Parasitic Lice (Phthiraptera: *Columbicola*). *Systematic Biology* 66: 896-911.

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

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

Bush G.L. 1975. Modes of Animal Speciation. Annu. Rev. Ecol. Syst. 6: 339–364.

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

Carling M.D., Brumfield R.T. 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., Amores A., Hohenlohe P., Cresko W., Postlethwait J.H. 2011. Stacks: Building and genotyping loci *de novo* from short-read sequences. *Genes, Genomes, Genetics* 1: 171–182.

Clayton D.H. 1990. Host specificity of Strigiphilus owl lice (Ischnocera: Philopteridae),

with the description of new species and host associations. *Journal of Medical Entomology* 27: 257–265.

Clayton D.H., Bush S.E., Goates B.M., Johnson K.P. 2003. Host defense reinforces host-parasite cospeciation. *Proceedings of the National Academy of Sciences* 100: 15694-15699.

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

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

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

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

Couch A.B. 1962. Phoretic mallophagans from hippoboscids of mourning doves Zenaidura macroura. *Journal of Parasitology* 48: 497.

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

Criscione C.D., Poulin R., Blouin M. 2005. Molecular ecology of parasites: Elucidating ecological and microevolutionary processes. *Molecular Ecology* 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., Hossaert-Mckey M., Jabbour-Zahab R., Jousselin E., Kerdelhué C., Kjellberg F., Lopez-Vaamonde C., Peebles J., Peng Y.Q., Santinelo Pereira R.A., Schramm T., Ubaidillah R., Van Noort S.,

Weiblen G.D., Yang D.R., Yodpinyanee A., Libeskind-Hadas R., Cook J.M., Rasplus J.Y., Savolainen V. 2012. An extreme case of plant-insect codiversification: Figs and fig-pollinating wasps. *Systematic Biology* 61: 1029–1047. doi:10.1093/sysbio/sys068.

Cutter A.D. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution* 69: 1172–1185.

Danecek P., Auton A., Abecasis G.,. Albers C.A, Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R., 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

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

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

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

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

2275.

Earl D.A., VonHoldt B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* 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. *Annual Magazine of Natural History* 1: 588–598.

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

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

Fahrenholz H. 1913. Ectoparasiten und abstammungslehre. *Zoologischer Anzeiger* 41: 371–374.

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

Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and*

Evolution 27: 480–488.

Gandon S. 2002. Local adaptation and the geometry of host – parasite coevolution. *Ecology Letters* 5: 246–256.

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

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

Harbison C.W., Jacobsen M.V., Clayton D.H. 2009. A hitchhiker's guide to parasite transmission: The phoretic behaviour of feather lice. *International Journal for Parasitology* 39: 569–575.

Harrison L. 1914. The Mallophaga as a possible clue to bird phylogeny. *Austral Zoology* 1: 7–11.

Harrison R.G. 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. *in* Howard, D. J., and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford University Press.

Hebert P.D.N., Gregory T.R. 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54: 852–859.

Hoberg E.P., Brooks D.R. 2015. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Philosophical Transactions of the Royal Society B Biological Sciences* 370: 20130553.

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

Hoyle W.L. 1938. Transmission of poultry parasites by birds with special reference to the "English" or house sparrow and chickens. *Transactions of the Kansas Academy of Sciences* 41: 379–384.

Hutchinson M.C., Cagua E.F., Balbuena J.A., Stouffer D.B., Poisot T. 2017. paco: implementing Procrustean Approach to Cophylogeny in R. *Methods in Ecology and Evolution* 8: 932–940.

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

Jakobsson M., Rosenberg N.A. 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., Štefka J., Jirků M. 2017. Discordant population histories of host and its parasite : A role for ecological permeability of extreme environment ? *PLoS One* 12: e0175286.

Johnson, K.P., Clayton D.H. 2004. Untangling coevolutionary history. *Systematic Biology* 53: 92–94.

Johnson K.P., Adams R.J., Clayton D.H. 2002. The phylogeny of the louse genus Brueelia does not reflect host phylogeny. *Biological Journal of the Linnean Society* 77: 233–247.

Johnson K.P., Clayton D.H. 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. University of Chicago Press, Chicago.

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

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

Johnson K.P., Walden K.K.O., Robertson H.M. 2013. Next-generation phylogenomics using a Target Restricted Assembly Method. *Molecular Phylogenetics and Evolution* 66: 417–422.

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

Johnson K.P., Weckstein J.D., Meyer M.J., Clayton D.H. 2011b. There and back again: switching between host orders by avian body lice (Ischnocera: Goniodidae). *Biological Journal of the Linnean Society* 102: 614–625.

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

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

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

Kirkness E.F., Haas B.J., Sun W., Braig H.R., Perotti M.A., Clark J.M., Lee H., Robertson H.M., Kennedy R.C., Gerlach D., Kriventseva E.V., Elsik G., Graur D.,

Hill C.A.,. Veenstra J.A, Tubío J.M.C., Ribeiro J.M.C., Rozas J., Johnston J.S., Reese J.T., Popadic A., Raoult D., Reed D.L., Tomoyasu Y., Mittapalli O., Margam V.M., Li H., Meyer J.M., Johnson R.M., Romero-severson J., Vanzee J.P., Alvarez-ponce D., Vieira F.G., Guirao-rico S., Anzola J.M., Yoon S., Strycharz J.P., Unger M.F., Christley S., Lobo N.F., Seufferheld M.J., Wang N., Dasch A., Struchiner C.J., Madey G., Hannick L.I., Bidwell S., Joardar V., Caler E., Shao R., Barker S.C., Cameron S., Bruggner R.V., Regier A., Johnson J., Viswanathan L., Utterback R., Sutton G.G., Lawson D., Robert M., Venter J.C., Strausberg R.L., Berenbaum M.R., Collins F.H., Zdobnov E.M. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences* 108: 6335.

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

Lanfear R., Calcott B., Kainer D., Mayer C., Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14: 82.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

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

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

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

Leigh J.W., Bryant D. 2015. Popart: Full-Feature Software for Haplotype Network Construction. *Methods in Ecology and Evolution* 6: 1110–1116.

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727-744.

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

Lively C.M., Dybdahl M.F., Jokela J., Osnas E.E., Delph L.F. 2004. Host Sex and Local Adaptation by Parasites in a Snail-Trematode Interaction. *American Naturalist* 164: S6–S18.

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

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

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

Martin C.D., Mullens B.A. 2012. Housing and dustbathing effects on northern fowl mites

(*Ornithonyssus sylviarum*) and chicken body lice (*Menacanthus stramineus*) on hens. *Medical and Vetinary Entomology* 26: 323–333.

Martínez-Aquino A. 2016. Phylogenetic framework for coevolutionary studies: a compass for exploring jungles of tangled trees. *Current Zoology* 62: 393–403.

Martínez-Aquino A., Ceccarelli F.S., Eguiarte L.E., Vázquez-Domínguez E., Pérez-Ponce De León G. 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., Boulinier T., Tirard C. 2005. Comparative host-parasite population structures: Disentangling prospecting and dispersal in the black-legged kittiwake *Rissa tridactyla*. *Molecular Ecology* 14: 2825–2838.

Mccoy, K.D., Boulinier T., Tirard C., Michalakis Y. 2001. Host specificity of a generalist parasite: Genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. *Journal of Evolutionary Biology* 14: 395–405.

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

McKay B.D., Zink R.M. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. *Molecular Phylogenetics and Evolution* 54: 647-650.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A.. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA

sequencing data. *Genome Research* 20: 1297–1303.

Miller M.A., Pfeiffer W., Schwartz T. 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., Warnow T. 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. *Journal of Parasitology* 81: 395–403.

Nadler S.A., De Leon G.P.P. 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., Charleston M.A. 1998. Trees within trees: Phylogeny and historical associations. *Trends in Ecology and Evolution* 13: 356–359.

Page R.D.M., Cruickshank R.H., Dickens M., Furness R.W., Kennedy M., Palma R.L., Smith V. S. 2004. Phylogeny of "*Philoceanus* complex" seabird lice (Phthiraptera: Ischnocera) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 30: 633–52.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.

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

Paterson A.M., Poulin R. 1999. Have chondracanthid copepods co-speciated with their teleost hosts? *Systematic Parasitology* 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., Johnson K.P., Clayton D.H., Baker A.J. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleocene. *Systematic Biology* 56: 656–672.

Pérez-Escobar O.A., Antonio Balbuena J., Gottschling M. 2015. Rumbling orchids: How to assess divergent evolution between chloroplast endosymbionts and the nuclear host. *Systematic Biology* 65: 51-65.

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

Piratelli A., Blake J.G. 2006. Bird communities of the southeastern Cerrado Region, Brazil. *Ornitologia Neotropical* 17: 213–225.

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

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

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

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945-959.

Puillandre N., Lambert A., Brouillet S., Achaz G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21: 1864–1877.

Refrégier G., Le Gac M., Jabbour F., Widmer A., Shykoff J.A., Yockteng R., Hood M.E., Giraud T. 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 Evolutionary Biology* 8: 100.

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

Ricklefs R.E., Fallon S.M. 2002. Diversification and host switching in avian malaria parasites. *Proceedings Royal Society B Biological Sciences* 269: 885–892.

Rosenberg N.A. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.

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

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

Sayyari, E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.

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

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

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

Sloan D.B., Havird J.C., Sharbrough J. 2016. The on-again-off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* 26: 2212–2236.

Smith M.A., Woodley N.E., Janzen D.H., Hallwachs W., and Hebert P.D.N. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences* 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., Boyd B.M., Johnson K.P. 2016a. Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. *Biological Journal of the Linnean Society* 118: 813-828.

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

Sweet A.D., Johnson K.P. 2016. Cophylogenetic analysis of New World ground-doves (Aves: Columbidae) and their parasitic wing lice (Insecta: Phthiraptera: *Columbicola*). *Molecular Phylogenetics and Evolution* 103: 122–132.

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

Templeton A.R. 1981. Mechanisms of speciation-a population genetic approach. *Annuual Review of Ecology and Systematics* 12: 23–48.

Vachaspati P., Warnow T. 2015. ASTRID: Accurate Species TRees from Internode Distances. *BMC Genomics* 16(Suppl 10): S3.

Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., Del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K.V, Altshuler D., Gabriel S., DePristo M.A.. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 43: 11.10.1-33.

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

Weckstein J.D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology* 53: 154–164.

Whipps C.M., Kent M.L. 2006. Phylogeography of the cosmopolitan marine parasite *Kudoa thyrsites* (Myxozoa: Myxosporea). *Journal of Eukaryotic Microbiology* 53: 364–373.

Whiteman N.K., Kimball R.T., Parker P.G. 2007. Co-phylogeography and comparative

population genetics of the threatened Galápagos hawk and three ectoparasite species: Ecology shapes population histories within parasite communities. *Molecular Ecology* 16: 4759–4773.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences* 107: 9264-9269.

CHAPTER 7: THE ROLE OF PARASITE DISPERSAL IN SHAPING A HOST-PARASITE SYSTEM AT MULTIPLE EVOLUTIONARY TIMESCALES

INTRODUCTION

Parasite dispersal is a key ecological mechanism that shapes host-parasite interactions and can cause effects at both population (micro) and species-level (macro) scales (Price, 1980; McCoy et al., 1999, 2003; Poulin, 2007, 2011; Criscione, 2008; Stefka et al., 2011). Parasites that are able to disperse effectively often have little population structure among different host species or populations (Dybdahl and Lively, 1996; McCoy et al., 2005; Kochzius et al., 2009). Dispersal can also result in parasite lineages switching between different host species (Page and Charleston, 1998; Clayton et al., 2004; Hoberg and Brooks, 2008). Despite its importance, dispersal is not the only factor that can shape a host-parasite system. For example, host diversification, host ecology, and biogeography can all potentially drive host-parasite evolution in some systems (Weckstein, 2004; Whiteman et al., 2007; Vinarski et al., 2007; Barrett et al., 2008). Although there are many studies focused on these topics at either the microevolutionary or macroevolutionary scale (Criscione et al., 2005; de Vienne et al., 2013; Cruaud and Rasplus, 2016), few studies have examined the effects of dispersal on both micro- and macroevolutionary patterns simultaneously in the same host-parasite system (Huyse et al., 2005; du Toit et al., 2013; Bell et al., 2016). Ideally, such an approach would compare "ecological replicate" parasites; i.e., different lineages of parasites with similar life histories that are associated with the same group of hosts, but have some ecological variable (e.g., dispersal ability) that differs among the parasites (Weiblen and Bush, 2002; Clayton and Johnson, 2003; Marussich and Machado, 2007). Because hosts commonly harbor multiple types of similar parasites, there are many potential

examples of ecologically replicate systems, including figs and fig-wasps (Weiblen and Bush, 2002; Marussich and Machado, 2007), parasitoid wasps (Hackett-Jones et al., 2009), avian malarial parasites (Ricklefs et al., 2004), and helminth worms of mammals (Bordes and Morand, 2009). The framework can also extend beyond host-parasite relationships to systems such as endosymbiotic bacteria of insects (Moran and Baumann, 2000) or plant-herbivore interactions (Ehrlich and Raven, 1964).

One model ecological replicate system are the wing and body lice (Insecta: Phthiraptera) of doves (Aves: Columbidae) (Clayton and Johnson, 2003; Johnson and Clayton, 2004; Clayton et al., 2016). Both louse "ecomorphs" only parasitize doves, spend their entire lifecycles on the host, and consume downy feathers (Nelson and Murray, 1971). However, the two ecomorphs are not closely related (Johnson et al., 2007; Johnson et al., 2011; Johnson et al. 2012), and they use different strategies for avoiding host preening. Wing lice have evolved an elongated morphology that allows them to insert themselves between barbs in wing and tail feathers, whereas body lice burrow into the downy feathers close to the host's body to avoid being removed (Clayton et al., 1999). Importantly, the two ecomorphs of lice also differ in their dispersal abilities. Both are primarily transmitted vertically (from parent to offspring) or horizontally by direct contact (e.g., mating) (Rothschild and Clay, 1952; Clayton and Tompkins, 1994). However, wing lice can also use winged hippoboscid flies to disperse between host individuals or host species, a behavior known as phoresy (Keirans, 1975; Harbison et al., 2008; Harbison et al., 2009). Hippoboscid flies are generalist blood-feeding parasites, with many individual fly species recorded from multiple dove genera (Maa, 1969). Multiple wing lice can grasp to a single fly with their legs and mandibles and can then be transported by the fly to another host individual, perhaps resulting in the establishment of a new louse population. Although there is at least one record of body lice

attached to hippoboscid flies in the wild (Couch, 1962), phoresy appears to be extremely rare in this ecomorph. Body lice have short legs that inhibit them from grasping to the flies. In an experimental study comparing the phoretic ability of wing and body lice from captive pigeons, no body lice were found attached to hippoboscid flies, whereas wing lice were attached to several flies (Harbison et al., 2011).

The difference in dispersal ability between wing and body lice appears to be reflected in their cophylogenetic and population genetic patterns. Clayton and Johnson (2003) showed that wing lice have little cospeciation with their hosts and high levels of host switching compared to body lice from the same host species. Focusing within a few louse species, Johnson et al. (2002) used a portion of the COI mitochondrial gene to show that body lice have more population structure and are more host-specific than wing lice. To build on this work, an ideal approach would integrate phylogenetic and population genetic patterns for the same louse taxa, thus simultaneously providing macro- and microevolutionary perspectives of the dove louse system. Additionally, such an approach should consider two data sampling issues. First, a phylogenetic comparison should utilize comprehensive taxonomic representation from a subset of taxa. A host or parasite phylogeny that is missing key lineages can result in misleading cophylogenetic patterns (Paterson et al., 2000; Sweet et al., 2016). Similarly, phenomena such as clade-limited host switching can produce seemingly congruent host and parasite phylogenies at broader (e.g., family-wide) taxonomic scales (Sorenson et al., 2004; Jackson et al., 2008; Demastes et al., 2012). Second, using many genetic markers, rather than a single gene or set of a few genes, provides more power for phylogenetic and population genetic analyses (Luikart et al., 2003; Delsuc et al., 2005). Whole genomic sequence data can be particularly useful, as it is possible to obtain markers for both levels of analysis from the same underlying data source (Cutter, 2013).

Here we focus on the wing and body lice of small New World ground-doves, a monophyletic group of 17 known dove species distributed from the southern United States to southern South America (Gibbs et al., 2003; Pereira et al., 2007; Sweet and Johnson, 2015). There are three described species of both wing (genus *Columbicola*) and body lice (genus *Physconelloides*), although there are likely additional cryptic species (Price et al., 2003; Sweet and Johnson, 2016; Sweet et al., 2018). Both types of lice also form monophyletic groups within their respective genera (Johnson et al., 2007; Johnson et al., 2011), which makes interpretation of evolutionary history straightforward. Obtaining genomic-level data is very feasible for these lice, as recently published genomic studies on avian lice have established pipelines for assembling data appropriate for both phylogenetic and population genetic analysis (Allen et al., 2017; Boyd et al., 2017; Sweet et al., 2018).

Over macroevolutionary timescales, we focus on two types of patterns in ground-dove lice: phylogenetic congruence and the relative timing of divergence between species. If dispersal is a major driver of host-parasite evolution, then we expect body lice to show more phylogenetic congruence and cospeciation with their hosts than do wing lice (Brooks and McClennan, 1991; Clayton and Johnson, 2003). Similarly, dispersal can influence the rate at which lineages of parasites diverge once two host lineages have speciated. If dispersal between host species is highly limited, as in the case for body lice, then we would expect these louse lineages to diverge and speciate at the same time as host lineages (Page, 1993; Hafner and Page, 1995). However, if dispersal between host species is less limited, then there may be a time lag between divergence and speciation of host lineages compared to their associated parasite lineages (Hafner et al., 1994; Banks and Paterson, 2005; Light and Hafner, 2007). Although this might be difficult to detect directly, we can predict that for an equivalent divergence event (i.e., two daughter lineages

of wing and body lice on sister species of hosts), wing lice may show less overall genetic divergence than body lice, because their divergence is more likely to lag behind that of their hosts (Vogwill et al., 2008). It may also be the case that host speciation strongly influences diversification of both parasite ecomorphs, irrespective of dispersal differences. If measures of phylogenetic congruence are correlated between wing and body lice from the same host species, this could indicate that host diversification influences wing and body lice in similar ways.

On an ecological (microevolutionary) timescale, comparative population structure and genetic diversity (heterozygosity) are informative for assessing the role of dispersal in host-parasite evolution. Parasite populations could be structured across several scales – for example among host species or among biogeographic regions (McCoy et al., 2001; Nieberding et al., 2008; Stefka et al., 2011; Sweet and Johnson, 2016). Because dispersal is local, we predict that differences in dispersal between wing and body lice should manifest themselves as population structure between host species but not necessarily as structure between biogeographic regions. Host dispersal itself is likely to be the strongest factor allowing lice to disperse between biogeographic regions, because lice are intimately tied to their hosts. In this case, host dispersal should affect parasite dispersal in similar ways between wing and body lice. In particular, we predict that body lice should show more population genetic structure among host species than wing lice (Johnson et al., 2002), but not necessarily more population structure among biogeographic regions.

Dispersal may also contribute to genetic structure between louse populations (infrapopulations) on different host individuals of the same host species. In this case, because dispersal is likely to be more limited for body lice, their infrapopulations are more likely to be highly inbred (Nadler, 1995). Thus, we predict that body lice will show lower levels of

heterozygosity, because their more limited dispersal among host individuals should lead to higher inbreeding.

In this study, we address the question of whether the impacts of dispersal differences can be observed at both macro- and microevolutionary timescales by comparing phylogenomic and population genomic patterns between wing and body lice sampled from across the diversity of small New World ground-doves. To estimate these patterns, we use genome sequence data from multiple individuals of each wing and body louse species. These data include assemblies of over 1,000 nuclear genes and tens of thousands of single-nucleotide polymorphisms (SNPs) called from these same genes. The results provide important insight into how dispersal affects host-parasite interactions, and ultimately how ecological mechanisms link to evolutionary patterns.

MATERIALS AND METHODS

Sampling

Samples of lice were collected from ground-doves in the field using pyrethrin powder dusting or fumigation methods as outlined in Clayton and Drown (2001). All collected specimens were immediately place in 95% ethanol and stored long-term at -80 °C. A total of 31 wing louse (*Columbicola*) and 34 body louse (*Physconelloides*) specimens were selected for whole genome sequencing. These represent all described ground-dove louse species, several potential cryptic species, and most host species and biogeographic areas (Supplementary Table 7.1). Each louse was then photographed as a voucher. Individual lice were then ground up in a 1.5 mL tube, and genomic DNA (gDNA) was extracted using reagents and a modified protocol of the Qiagen QIAamp DNA Micro Kit (Qiagen, Valencia, CA, USA). Our modification extended the duration of the incubation step to 48 hours, instead of the recommended 1-3 hours.

The extractions were then quantified with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using the manufacturer's recommended protocols and reagents.

Library preparation and sequencing

Total gDNA was fragmented on a Covaris M220 Focused-ultrasonicator (Covaris, Woburn, MA, USA) targeting a mean fragment size of 400 nt. The fragmented gDNA of each specimen was then constructed into a library for paired-end Illumina whole genome shotgun sequencing using a Hyper Library Preparation Kit (Kapa Biosystems, Wilmington, MA, USA). A 10-nt barcode was adapted to each library so that up to 16 individual libraries could be pooled and sequenced on a single Illumina lane (two lanes in total). The libraries were sequenced with the HiSeq4000 v1 sequencing kit for 151 cycles on an Illumina HiSeq4000 instrument. The sequencing resulted in 150 bp paired-end reads in fastq files generated in bcl2fastq v2.17.1.14. All library preparation and sequencing was carried out at the Roy J. Carver Biotechnology Center (University of Illinois, Urbana, IL, USA). For wing lice, we deposited the raw reads on the NCBI SRA database (accession SRP116697; BioProject PRJNA400795). We obtained additional raw genomic read data from NCBI's SRA database for our ingroup (SRR3161921-SRR3161923, SRR3161930- SRR3161931) and outgroup *Columbicola* taxa (*C. columbae*: SRR3161917, *C. gracilicapitis*: SRR3161913, *C. macrourae*: SRR3161953, *C. veigasimoni*: SRR3161919) (Boyd et al., 2017). Raw sequence reads for 34 body lice (*Physconelloides*) were deposited previously (Sweet et al., 2018; SRP076185).

We ran several quality control measures on the raw Illumina data. First, we removed duplicate read pairs using the fastqSplitDups script (https://github.com/McIntyre-Lab/mcscript and https://github.com/McIntyre-Lab/mclib). We then removed the Illumina sequencing adapters with Fastx_clipper v0.014 from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit).

Finally, we removed the first 5 nt of the 5' ends of reads using Fastx_trimmer v0.014 and removed 3' ends of reads until reaching a base with a phred score ≥28 using Fastq_quality_trimmer v0.014. Following quality control, we removed any reads less than 75 nt and analyzed the cleaned libraries with Fastqc v0.11.5 (Babraham Bioinformatics) to check for additional errors.

Sequence assembly

To assemble nuclear loci from genomic reads, we used an approach similar to the one detailed in Sweet et al. (2018), which maps lower coverage, multiplexed genomic data to reference loci from a closely related taxon. For our reference set of nuclear loci for wing lice, we used 1,039 exons of *Columbicola drowni* generated in Boyd et al. (2017) (raw data: SRR3161922). This data set was assembled *de novo* in aTRAM (Allen et al., 2015) using orthologous protein-coding genes from the human body louse genome (*Pediculus humanus humanus*; Kirkness et al., 2010) as a set of target sequences. We mapped our newly generated *Columbicola* reads and the reads obtained from GenBank to the *C. drowni* references using Bowtie2 (Langmead and Salzberg, 2012). We then created VCF files for each sample using Samtools and Bcftools (Li et al., 2009), and filtered out sites according to sequencing depth (<5 or >150) and quality (phred scores <28) using Samtools and the Genome Analysis Toolkit v3.7 (GATK; McKenna et al., 2010). The entire read mapping pipeline is detailed at https://github.com/adsweet/louse_genomes.git. For body lice, nuclear data were obtained using the same pipeline and software parameters, except that 1,095 loci from *P. emersoni* were used as the references for mapping.

In addition to the nuclear exons, we used aTRAM to assemble mitochondrial genes for ground-dove lice. To generate a set of target genes for wing lice, we mapped cleaned Illumina

reads from *C. passerinae* (SRA accession SRR3161930) to annotated mitochondrial protein-coding genes of *Campanulotes compar* (pigeon body louse; GenBank accession AY968672) in Geneious v8.1.2 (Biomatter, Ltd., Auckland, NZ). Our preliminary analyses indicated the *Campanulotes compar* mitochondrial genes were too divergent from *Columbicola* to be useful as target genes in aTRAM. Based on the mapping, we identified the *Columbicola* mitochondrial genes, extracted these regions, and translated the sequences to amino acids. We used these protein sequences as our target set in aTRAM. We ran aTRAM for a single iteration using ABySS (Simpson et al., 2009) for *de novo* assembly. We also used one of several library fractions (1.5%, 4.7%, 15.6%, 50%, or 100%), and for each library chose the minimum fraction that had uniform coverage above 20X. Because *Campanulotes compar* is more closely related to *Physconelloides*, we were able to obtain mitochondrial sequences of *Physconelloides* that were assembled in aTRAM using *Campanulotes compar* target sequences.

Calling SNPs in widespread lice

To compare population structure between wing and body lice across multiple host species, we focused on the most widespread (i.e., least host-specific) species of wing and body lice. For wing lice, we called SNPs jointly for *C. passerinae* with GATK following the "Best Practices" guide (Van der Auwera et al., 2013; https://software.broadinstitute.org/gatk/best-practices/). We used *C. drowni* loci as a reference and filtered out SNP 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. SNPs were called for body lice with the same approach using *P. emersoni* as the reference.

Phylogenetic estimation

We applied similar approaches for estimating phylogenetic relationships in both wing and body lice. First, we aligned each nuclear locus in MAFFT (--auto; Katoh et al., 2002), and removed columns with only ambiguous sequences ("N"). We concatenated all alignments using SequenceMatrix (Vaidya et al., 2010) and tested for optimal partitioning schemes and substitution models with the rcluster search in PartitionFinder v2.1.1 (Lanfear et al., 2014, 2017). We selected optimal partitions based on the Akaike Information Criterion (AIC; Akaike, 1974). From the partitioned concatenated alignment, we estimated the best likelihood tree and 250 rapid bootstrap replicates in RAxML v8.1.3 (Stamatakis, 2006). We also estimated phylogenies using coalescent-based methods, which account for discrepancies between gene and species trees due to Incomplete Lineage Sorting. For these analyses, we first estimated gene trees in RAxML with a GTR + Γ substitution model for each gene alignment. We then summarize the gene trees in ASTRAL v4.10.6 with local posteriori probability branch support (Mirarab and Warnow, 2015; Sayyari and Mirarab, 2016).

We also estimated mitochondrial phylogenies from the assembled mitochondrial genes. As with the nuclear data, we aligned the mitochondrial genes in MAFFT, tested for optimal partitioning and model schemes based on the AIC in PartionFinder, and estimated a phylogeny from the concatenated alignment with 250 rapid bootstrap replicates in RAxML.

We also used the mitochondrial data to estimate the number of Operational Taxonomic Units (OTUs). First, we used the COI alignment in the web version of the automatic barcode discovery method (ABGD; http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html; Puillandre et al., 2012), which tests for interspecific boundaries based on the distribution of genetic distances from a barcode gene. We used default parameters (Pmin – 0.001, Pmax = 0.1, Steps =

10, Relative gap width = 1.5, Bins = 20) and three distance models (uncorrected, Jukes-Cantor, and Kimura) for our ABGD analysis. We also tested for OTUs using the Bayesian General Mixed Yule Coalescent Model (bGMYC; Reid and Carstens, 2012). Because this method requires ultrametric trees, we estimated trees with our concatenated mitochondrial alignment in BEAST v2.4.4 (Bouckaert et al., 2014) on the CIPRES Science Gateway (Miller et al., 2010). We set the alignment partitions and substitution models in accordance with the RAxML analysis, used a Yule tree prior, default substitution priors, and a strict molecular clock. We ran the MCMC for 50 million generations, sampling every 10,000 generations and discarding the first 10% of MCMC samples as a burnin based on ESS values viewed in Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). We then randomly selected 100 trees from the post-burnin distribution of trees, and used these for our bGMYC analysis. For all 100 trees, we ran bGMYC for 20,000, with a burnin of 10,000, thinning = 10, and a conspecific probability cutoff ≥ 0.05.

Cophylogenetic patterns in ground-dove lice

We tested for phylogenetic congruence between the putative louse species trees (trimmed to one representative per OTU) and the small New World ground-dove phylogeny from Sweet and Johnson (2015). First, we used the distance-based methods ParaFit (Legendre et al., 2002) and PACo (Balbuena et al., 2013). We converted the host and parasite trees to patristic distance matrices and ran ParaFit for 100,000 iterations in the R package APE (Paradis et al., 2004), using the Cailliez correction for negative eigenvalues and testing for the contribution of individual links with both ParaFit link tests (ParaFitLink1 [PF1] and ParaFitLink2 [PF2]). We corrected the resulting p-values for the individual link tests with the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995). For PACo, we used the same patristic distance

matrices and ran 1,000 iterations in the PACo R package (Hutchinson et al., 2017). We also used the jackknife approach in PACo to calculate the squared residual values for each host-parasite association. Second, we tested for specific "coevolutionary events" between ground-doves and their wing lice using the event-based method Jane v4 (Conow et al., 2010). We set generations to 500 and population size to 1,000 for the Genetic Algorithm, and randomized the tip associations 999 times to test for the statistical significance of our optimal score.

To test for a correlation between the cophylogenetic patterns of ground-dove wing and body lice, we compared the PACo residuals, PF1, and PF2 values associated with each host species. We did not include information for wing lice from *Metriopelia aymara* or *Columbina picui*, because there were no body lice associated with those host species. We used average values for host species with multiple louse associations. For all three metrics, we used the Spearman's Rank Coefficient in R to test for a correlation between wing and body lice.

Comparing divergence rates

Two pairs of sister species, *C. drowni* and *C. gymnopeliae* (wing lice) and *P. emersoni* and *P. robbinsi* (body lice), are associated with the same two host species (*M. ceciliae* and *M. melanoptera*), and both pairs likely co-diverged with their hosts. This implies both louse species pairs diverged in response to the same host speciation event, and comparing their genetic distances can provide an estimate of relative divergence rates between the two groups of lice. For each aligned nuclear gene, we calculated the uncorrected genetic distances between each species pair in APE. We used data from one representative of each species: *C. drowni* and *C. gymnopeliae* sequenced by Boyd et al. (2017) and the higher-coverage *P. emersoni* and *P. robbinsi* from Sweet et al. (2018). We excluded genes not present in both wing and body louse data sets. Based on our initial assessment of the distribution of distances, we also removed 11

genes with outlier distances (higher than 5%) in either wing or body lice. After these filtering steps, we were able to calculate distances for 1,006 genes. We also used a χ^2 test to compare the proportion of total differences across all aligned genes between the two pairs of wing and body louse sister species.

We also calculated the uncorrected distances between mitochondrial sequences in each species pair. Although we treated the mitochondrial data as a single locus (i.e., we calculated distances from concatenated gene alignments), we used only the 6 mitochondrial genes available for both wing and body lice. However, rather than use single representatives of each species, we compared distances among all samples of each species pair (3 samples of *C. drowni* and *C. gymnopeliae*; 4 samples of *P. emersoni* and *P. robbinsi*).

Population genomic analysis

We assessed the population structure of widespread louse species using STRUCTURE, Discriminant Analysis of Principal Components (DAPC), and Principal Component Analysis (PCA). Our approach followed the analysis of the body louse *P. eurysema* in Sweet et al. (2018). For STRUCTURE analyses we randomly selected one SNP per assembled gene, which ensures that individual SNPs are unlinked. For wing lice, we ran STRUCTURE 20 times on these subsets of SNPs with 50,000 MCMC iterations and 25,000 burnin iterations for K = 2-8. We then used the ΔK method (Evanno et al., 2005) in STRUCTURE Harvester v0.6.94 (Earl and VonHoldt, 2012) to determine the optimal number of clusters. We summarized all STRUCTURE runs in CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) and visualized the results by constructing plots with distruct v1.1 (Rosenberg 2004). We also ran STRUCTURE analyses for two possible cryptic wing louse taxa (*C. passerinae* 1 and *C. passerinae* 2) identified in previous phylogenetic studies of the genus (Johnson et al., 2007; Sweet et al., 2016). We once again randomly selected

SNPs and ran STRUCTURE as detailed above. We performed DAPC in the R package ADEGENET (Jombart, 2008) using all SNPs for *C. passerinae*. For PCA we subsampled SNPs for the cryptic taxa *C. passerinae* 1 and *C. passerinae* 2 using vcftools v0.1.14 (Danacek et al., 2011) and analyzed them separately in ADEGENET.

Heterozygosity can be a useful measure of host specificity and the effect of ecological dynamics (e.g., dispersal) on parasite populations. We estimated heterozygosity for wing and body louse individuals using two approaches. First, we estimated the scaled population mutation rate (θ), an indicator of heterozygosity, for individuals using mlRho v2.9, which uses a maximum likelihood approach to estimate population parameters for diploid individuals (Haubold et al, 2010). We converted pileup files generated from Samtools to "profile" files and formatted these for mlRho using the auxiliary software for the program (available at http://guanine.evolbio.mpg.de/mlRho/). For each individual, we ran mlRho with maximum distance (-M) set to 0. Second, we calculated raw heterozygosity for individuals by dividing the number of heterozygous sites by the number of total sites.

RESULTS

Sequencing and assembly

Paired-end Illumina sequencing of 31 *Columbicola* specimens yielded an average ~44.9 million raw reads per specimen (Supplementary Table 7.1). After cleanup steps, there were on average ~33.4 million reads per specimen, which translates to an average predicted sequencing depth of ~25X per specimen (based on a 200 Mbp genome size). Including the 9 additional samples from GenBank (5 outgroup taxa and 4 previously sequenced ingroup samples), an average of 1,036 genes per library mapped against the *C. drowni* reference (1,039 genes

targeted). For body lice, we obtained an average of 1,055 genes per library mapped against the *P. emersoni* reference.

For wing lice, we assembled 7 mitochondrial protein-coding genes for most samples (CO1, CO2, CO3, Cytb, ND1, ND3, and ND5), using an average library fraction of 57.1%. The other targeted mitochondrial protein-coding genes (ATP6, ATP8, ND2, ND4, ND4L, and ND6) assembled for none or only a few samples, and so we excluded those genes from downstream analyses. For the 7 "successful" genes, aTRAM assembled data for all 40 samples in all but one gene (ND5), which assembled for 38 samples. By comparison, we obtained 10 body louse mitochondrial genes (all but ATP8, ND3, and ND6) assembled with aTRAM using an average library fraction of 15.7%.

The GATK pipeline called 25,952 SNPs for *C. passerinae* after filtering. This included 15,225 SNPs for *C. passerinae* 1 and 14,456 SNPs for *C. passerinae* 2. Selecting one random SNP per gene for STRUCTURE analyses resulted in 773 SNPs for *C. passerinae*, 635 SNPs for *C. passerinae* 1, and 636 SNPs for *C. passerinae* 2. We obtained 56,232 SNPs from *P. eurysema*, including 880-908 SNPs for STRUCTURE.

Phylogenetic analysis

The concatenated nuclear alignment for wing lice was 1,104,066 bp long, 3.8% of which consisted of gaps or ambiguous characters (missing data). The best partitioning scheme of the concatenated alignment estimated in PartitionFinder consisted of 345 subsets. The resulting phylogeny from RAxML was very well supported. All OTUs received 100% bootstrap (BS) support, and many of the branches within OTUs received high support (>75 BS). The species-level relationships agreed with other phylogenetic assessments of this group (Johnson et al., 2007; Sweet and Johnson, 2016). *Columbicola altamimiae* (ex *Metriopelia aymara*) was sister to

the rest of the ingroup, and *C. gymnopeliae* (ex *M. ceciliae*) and *C. drowni* (ex *M. melanoptera*) were sister to *C. passerinae* (Figure 7.1, Supplementary Figure 7.1). The coalescent phylogenies estimated from individual gene trees in ASTRAL were also well supported and largely agreed with the concatenated phylogeny (Supplementary Figures 7.2). In the ASTRAL phylogeny, all species-level relationships received 100% local posterior probability support. Finally, the concatenated mitochondrial alignment was 5,535 bp long and contained 13.4% missing data. The resulting mt phylogeny also exhibited well-supported species relationships in agreement with the nuclear phylogenies (Supplementary Figure 7.3). The only topological differences among the concatenated, coalescent, and mitochondrial phylogenies were all within species. Concatenated and coalescent phylogenetic estimates of body lice, based on 1,553,983 nuclear bp and 9,121 mitochondrial bp, also produced consistent, well-supported hypotheses. *Physconelloides emersoni* and *P. robbinsi* were recovered as sister to *P. eurysema*.

The OTU assessments indicated there are five ingroup species of wing lice. In the ABGD analysis based on the COI alignment, all distances models supported five distinct taxa. Likewise, the bGMYC analysis based on 100 mitochondrial trees sampled from a posterior distribution supported five taxa at the 0.05 conspecific cutoff. The supported taxa include the three species that parasitize *Metriopelia* doves (*C. altamimiae*, *C. drowni*, and *C. gymnopeliae*) and two species within *C. passerinae* ("*C. passerinae* 1" and "*C. passerinae* 2"). The presence of two possible cryptic species within *C. passerinae* supports the results of previous work (Johnson et al., 2007). By comparison, assessments of body lice recovered 5 cryptic OTUs within *P. eurysema* (7 total OTUs).

Cophylogenetic analysis

Both ParaFit (ParaFitGlobal = 1.97, P = 0.005) and PACo (m^2 = 0.078, P = 0.003) indicated the ground-dove and wing louse phylogenies were congruent overall. However, none of the individual links were significant in the ParaFit test (Supplementary Table 7.2). The Jane reconciliation recovered a single cospeciation event between ground-doves and their wing lice (between *M. melanoptera*/*M. ceceiliae* and their lice), along with 2 duplications, no host switches, 10 losses, and 8 failures to diverge (Figure 7.1, Supplementary Figure 7.4). This least-costly solution was significantly lower than with randomized associations (observed cost = 27, mean randomized cost = 35.42, P = 0.03). Ground-doves and their body lice also had overall congruent phylogenies, but there were more cospeciation events (3) and significant individual associations (2) than in the wing louse system.

Comparisons of cophylogenetic analyses in ground-dove wing and body lice produced varied results. When only considering host species present in both data sets (the wing louse data set includes two more host species than the body louse data set), PACo residual values between wing and body louse links were not significantly different (Mann-Whitney U = 57, P = 0.847; Table 7.1), but they were positively correlated (ρ = 0.71, P = 0.019; Figure 7.2). Notably, lice from *Metriopelia* had low residual values in both wing and body lice, whereas lice from *Claravis pretiosa* had high residual values in both groups of lice. Lower residuals indicate a greater contribution to phylogenetic congruence. Conversely, PF1 (ρ = 0.45, P = 0.17; Supplementary Figure 7.5A) and PF2 (ρ = 0.59, P = 0.057; Supplementary Figure 7.5B) values were not correlated between wing and body louse links. However, body louse links had significantly higher PF1 and PF2 values than wing louse links (PF1: Mann-Whitney U = 12, P = 0.001; PF2: Mann-Whitney U = 11, P-value = 0.001; Table 7.1). Contrary to PACo residuals, higher PF1 and

PF2 values indicate a greater contribution to overall congruence between host and parasite phylogenies.

Comparative genetic distances between wing and body lice

Uncorrected distance values from the 1,007 nuclear genes were higher for *P. emersoni* and *P. robbinsi* (body lice; median = 0.007) than for *C. drowni* and *C. gymnopeliae* (wing lice; median = 0.005) (Mann-Whitney U = 1537100, P < 0.001) (Figure 7.3A). The body lice also had a higher proportion (0.008) of differences across all genes compared to the wing lice (0.006) (χ^2 = 249.49, P < 0.001). The opposite pattern was true of the mitochondrial distances: wing lice (median = 0.173) had higher distances than body lice (median = 0.146) (Mann-Whitney U = 0, P < 0.001; Figure 7.3B).

Population genomic patterns

STRUCTURE and DAPC analyses for the wing louse *C. passerinae* indicated K = 2 as the optimal number of clusters. The patterns of these two subsets correspond with the two OTUs recovered from ABGD and bGMYC (Supplementary Figure 7.6). Further STRUCTURE analysis on the two OTUs recovered an optimal K = 2 for *C. passerinae* 1 and K = 5 for *C. passerinae* 2. However, neither of these results suggested significant patterns of structure within these two OTUs (Supplementary Figure 7.7). DAPC indicated there are two clusters within *C. passerinae* 1 (Supplementary Figure 7.8). These patterns roughly correspond to biogeographic areas, which is further highlighted in the PCA (Figure 7.4A). DAPC did not recover any structure in *C. passerinae* 2, and although the PCA indicated there is some structure in the group, there are no clear patterns according to host species or biogeography as in other ground-dove louse taxa (Figure 7.4B). By comparison, the body louse species *P. eurysema* had an optimal K = 3, but

with more structure at higher values of K. Within the widespread OTU *P. eurysema* 3, some lice clustered according to host species and others according to biogeography (Figure 7.4C).

In assessments of heterozygosity in wing and body lice, estimates of raw heterozygosity (ratio of heterozygous sites to total sites) and θ (estimated from mlRho) were very comparable (Supplementary Table 7.3-7.4). Therefore, comparative tests using the two metrics gave similar results, and here we will report the results from the θ metric. Overall, wing lice had higher θ values than body lice (Mann-Whitney $U = 275.5$, $P < 0.001$; inset of Figure 7.5). However, this pattern is driven by differences between specialist lice (i.e., lice that are only associated with a single host species). Wing and body louse specialists have significantly different θ (Mann-Whitney $U = 0$, $P < 0.001$), whereas wing and body louse generalists have θ values that are not significantly different (Mann-Whitney $U = 235$, $P = 0.220$; Figure 7.5). Separately, wing louse specialists had higher θ than wing louse generalists (Mann-Whitney $U = 20$, $P < 0.001$), whereas the opposite pattern was apparent in body lice; body louse generalists had higher θ than body louse specialists (Mann-Whitney $U = 211.5$, $P = 0.004$).

DISCUSSION

Phylogenomic and population genomic comparisons of "ecological replicate" lice from ground-doves indicate that dispersal is a major force in shaping both micro- and macroevolutionary patterns in these parasites, providing a clear link between ecological mechanism and evolutionary patterns. As we predicted, wing lice, which have higher dispersal capability by using phoresis, showed less evidence of cospeciation with ground-doves and lower host-specificity than did body lice, which do not use phoresis. Wing lice also did not show any obvious population structure according to host species, whereas some body lice did show this

pattern. Finally, the most widespread wing louse species (*C. passerinae*) overall had less population structure than the most widespread body louse species (*P. eurysema*). Taken together, these findings suggest that parasite dispersal can be the primary driving force in host-parasite coevolution. Lice that cannot use hippoboscid flies for dispersal (i.e., body lice) can become isolated on a particular host species, leading to rapid population divergence and ultimately cospeciation with the host (Clayton and Johnson, 2003; Harbison et al., 2011).

Nevertheless, dispersal alone cannot account for all cophylogenetic patterns. In ground-doves, dispersal and host diversification interact to shape host-parasite interactions. Both wing and body louse phylogenies were significantly congruent with the host phylogeny, and both had at least one reconstructed cospeciation event. Although the lice can disperse to other host species – especially wing lice – both types of lice still have a strong association with the hosts and display some patterns of host-specificity. Lice cannot survive for long off the host, and they spend their entire lifecycle on the host (Marshall, 1981; Tompkins and Clayton, 1999). In theory, this type of host-parasite relationship should result in at least some phylogenetic congruence (Fahrenholtz, 1916; Eichler, 1948), which is what we observe here. It is also noteworthy that measures of congruence for individual host-parasite associations were positively correlated between wing and body lice (Figure 7.2). This further suggests host species are a key factor in promoting phylogenetic congruence, regardless of ecological differences between the two types of lice. Of course, neither louse system shows perfect phylogenetic congruence with hosts, and there is considerable variation between the two groups of lice. However, lice associated with *Metriopelia* ground-doves stand out as having consistent patterns of congruence. Both wing and body lice have a cospeciation event with these birds, and both are host-specific. Ecological barriers – notably the geographical and altitudinal differences among different species of

Metriopelia and between *Metriopelia* and lowland ground-dove species – could inhibit host switching and over time lead to phylogenetic congruence in both types of lice.

We also predicted that for shared divergence events, wing lice may exhibit delayed divergence, as compared to body lice, because they can more readily disperse among host species during the early stages of their divergence. Because they cannot effectively use hippoboscid flies for dispersal, body louse populations are expected to become isolated rapidly among diverging host lineages, whereas wing lice would be better able to retain some level of gene flow after an initial split. We therefore expected to see a higher genetic divergence between body louse species pairs compared to a pair of wing louse species that diverged with the same pair of host species. In comparison between the shared cospeciation event among *Metriopelia* doves and wing and body lice (Figure 7.1), this is exactly the pattern we see in the nuclear genes, with body lice showing more genetic divergence than wing lice across all loci. Surprisingly, the mitochondrial data show the opposite pattern; wing lice have a higher divergence between the same pair of host taxa than do body lice. This pattern could be related to the different architectures of the mitochondrial genomes. Body lice have a single mitochondrial chromosome, whereas wing lice likely have several mitochondrial "mini-chromosomes" (Covacin et al., 2006; Cameron et al., 2011). This uncommon architecture in wing louse mitochondrial genomes might enable the lice to withstand higher mutation rates in the mitochondria (S. Cameron, pers. comm.). Alternatively, the mini-chromosomes may actually cause increases in mutation rates, because of increased speed or frequency of replication. It could also be that the mutation rate differs for nuclear loci in the opposite direction, but there would be no known mechanism for this. In either case, comparisons of relative nuclear and mitochondrial divergence rates across different groups of lice appears to be a potentially rich field for further investigation.

Like host speciation, biogeography also plays an important role in shaping the phylogenetic and population divergence outcomes between lice and ground-doves, although it appears to affect wing and body lice differently. Wing lice show biogeographic structure at the phylogenetic level, but do not exhibit a similar structure within species. Conversely, body lice do not have biogeographic structure at the phylogenetic level, but they do within a widespread species (see also Sweet et al, 2018), a pattern that suggests body louse population structure can be shaped by biogeography, but the lineages sort according to host species over time. The underlying mechanism driving these patterns could still be dispersal differences (Weckstein, 2004). Because wing lice can more readily switch among sympatric host species, speciation may thus be driven by geographic events rather than host speciation. However, it is also possible that host and/or hippoboscid fly dispersal is responsible for the biogeographic patterns. Further phylogeographic analysis of lice, flies, and doves is needed to rigorously address these hypotheses.

Finally, dispersal also appears to have consequences at the population level, particularly as it relates to genetic diversity and inbreeding. Measures of heterozygosity in wing and body lice generally reflect the difference in dispersal ability. Overall, wing lice had higher levels of heterozygosity than body lice, which suggests the wing lice are more outbred. The ability of wing lice to disperse between different host individuals provides an opportunity for multiple populations to maintain gene flow. Because body lice have more limited opportunities for dispersal, they can become isolated on a host population or individual, thus leading to lower heterozygosity and more inbred louse populations (Nadler, 1995).

However, when examining patterns in heterozygosity in more detail, it is apparent that the overall pattern is driven by differences in heterozygosity for host specialists (i.e., species of

lice associated with a single host species). Whereas wing louse specialists had much higher heterozygosity than body louse specialists, heterozygosity levels of wing and body louse generalists were not significantly different. Not all body lice are strictly host-specific (i.e., there are some body louse species associated with multiple host species), and these body louse species appear to have gene flow on the same magnitude as some wing louse species, suggesting that body lice can disperse through modes other than phoresis. Many of the hosts parasitized by generalist body lice have overlapping geographic ranges and form mixed foraging flocks, so it is possible body lice are transferred through host contact or proximity (e.g., direct contact, shared dust baths, shared nest sites) (Clayton, 1991; Clayton and Tompkins, 1994; Clayton et al., 2016). Because wing and body lice generalists have similar levels of heterozygosity, perhaps interspecific dispersal via host contact occurs with similar frequency in both types of lice.

A more puzzling result is the higher heterozygosity of specialist wing lice compared to generalist wing lice (Barrett et al., 2008). This pattern may exist because wing lice are more able to disperse among individuals of a single host species than they are among multiple host species. In cases where multiple host species co-occur, it may be that overall dispersal rate in wing lice is lower than in cases where a single host species occurs alone. Another similar possibility is related to the number of new louse infrapopulations founded by phoresis versus those founded by parent-offspring transmission. Infestation prevalence of lice on doves (i.e., the fraction of host individuals with parasites) is often much less than 50%, suggesting there are many opportunities for founding of new infrapopulations (i.e., establishment of a louse population on an individual bird that previously did not have lice) (Price et al., 2003; Clayton et al., 2004). In the case of host specialists, a high proportion of the new infrapopulations would be founded by direct contact between male and female birds or through parent-offspring transmission at the nest. These

transmission (dispersal) events would often involve greater numbers of founding individual lice than would those founder events initiated through phoresis, which typically involve a very small number of lice (Harbison et al., 2011). For host generalist wing lice, there would likely be many founder events originating from phoresis, resulting in more population bottlenecks and thus leading to lower heterozygosity compared to host specialists.

Additional evidence for these explanations comes from examining the population genetic variation of the lice on *Metriopelia* doves, the most geographically isolated of all of the ground doves. All *Metriopelia* doves live at high elevations (usually >2,000 m.) in the Andes and are well-separated from closely related lowland ground-dove species (Gibbs et al., 2003). Some *Metriopelia* dove species are also separated from one another, by either geographical or altitudinal differences (e.g., some species are at higher elevations). The lice on these birds have large differences in estimates of θ, with wing lice having much larger values than body lice. It could be that without other host species in close proximity, wing lice avoid inbreeding by having a relatively high dispersal rate among conspecific host individuals, as compared to a mixed flock situation where some dispersal is within and some is between host species.

Variation in host population size could explain the differences in heterozygosity between wing and body louse specialists, if the pattern of specialization varies between the two groups (Hesse and Buckling, 2016). If wing louse specialists are associated with dove species with high population size, whereas body lice are associated with host with relatively small populations, then higher heterozygosity could be maintained in wing lice as compared to body lice. However, some wing and body louse specialists share two host species (*Metriopelia melanoptera* and *M. ceciliae*) in common, so differences in host population size could not be a factor in this case. Overall heterozygosity could also be a reflection of louse population sizes (Nei et al., 1975).

Wing lice are often more prevalent and have higher abundance than body lice, which could explain the observed patterns of heterozygosity (Harbison et al., 2011; Clayton et al., 2016). However, wing and body louse generalists have similar estimates of heterozygosity, and any relative differences in population size should have a similar effect on the heterozygosity of those taxa. It also seems plausible that generalist lice should have higher heterozygosity than specialist lice just because they could have higher overall population sizes by occurring on more host species. Indeed, in body lice, generalists have significantly higher heterozygosity than specialists. However, wing lice show the opposite pattern: specialists have significantly higher heterozygosity than generalists, so overall population size does not appear to be the most likely explanation for the variation in heterozygosity among specialist and generalist wing and body lice.

In conclusion, this study highlights the importance of an ecological process (dispersal) in shaping host-parasite micro- and macro-evolutionary patterns. By comparing two lineages of parasites that have different dispersal abilities but are associated with the same group of hosts, we were able to demonstrate that the lice with greater dispersal ability (wing lice) had less cophylogenetic congruence and population structure than the more dispersal-limited lice (body lice). We also show that other measures of population and evolutionary patterns, heterozygosity and genetic divergence, provide evidence for the importance of these dispersal differences in shaping the outcomes of these interactions. However, we also show that entirely independent factors, in particular biogeography and host diversification, can influence the pattern in these ground-dove lice. Thus, more generally, host-parasite systems are likely shaped by multiple factors, but in this system and others like it, dispersal ability can be a major predictor.

FIGURES

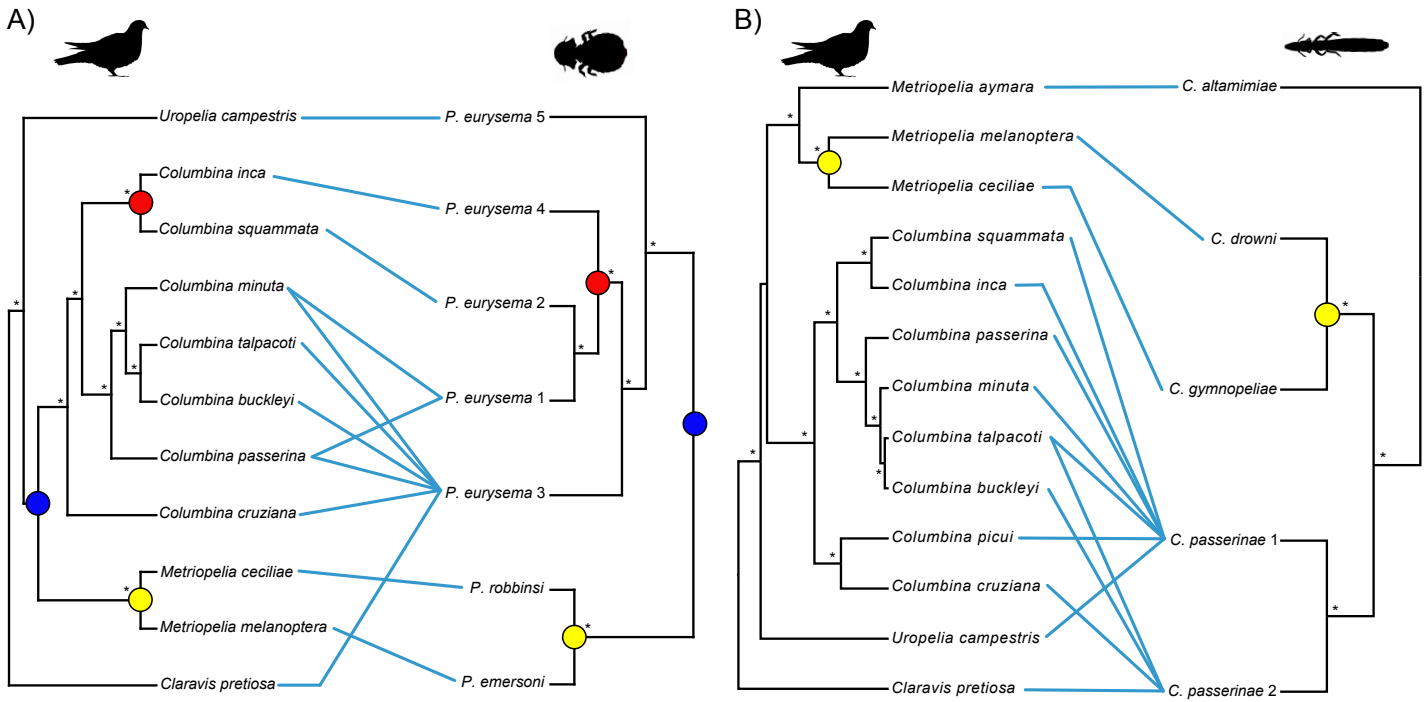


Figure 7.1. Tanglegrams comparing the phylogeny of New World ground-doves to the phylogenies of their A) body lice (*Physconelloides*) and B) wing lice (*Columbicola*). The ground-dove phylogeny is to the left and the louse phylogenies are to the right. The wing louse phylogeny is from this study, the body louse phylogeny is adapted from Sweet et al. (2018), and the ground-dove phylogeny is adapted from Sweet and Johnson (2015). Blue lines between the phylogenies indicate associated taxa. Asterisks indicate well-supported branches (>75% bootstrap support). In the louse phylogenies, all relationships have 100% bootstrap support. Circles over nodes indicate cospeciation events recovered from Jane4 reconciliation analyses.

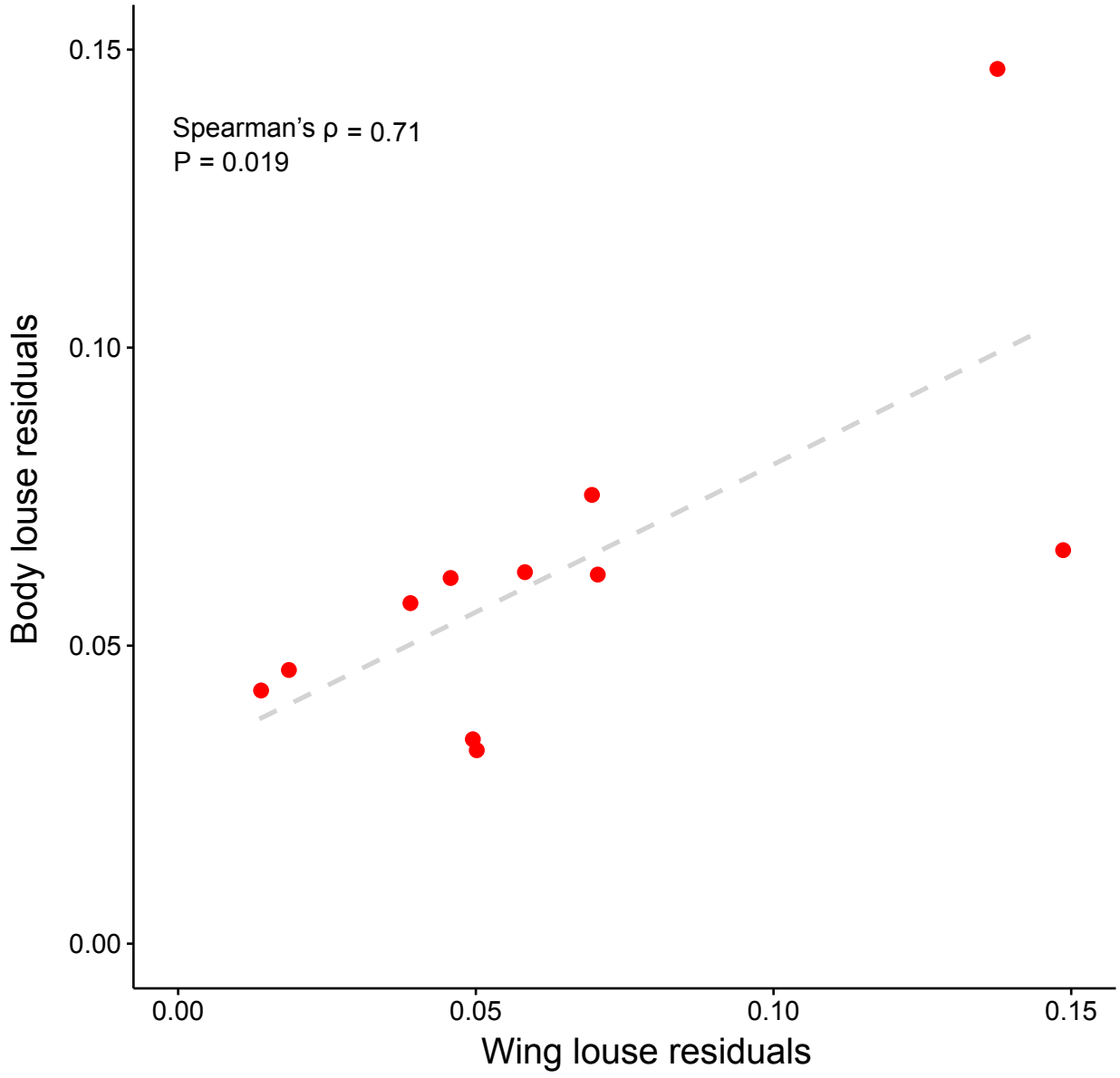


Figure 7.2. Correlation of New World ground-dove wing and body louse residuals from a PACo analysis. Red points indicate host species. For hosts with multiple wing or body louse associations, the points represent mean residual values. A regression line is provided to indicate trend.

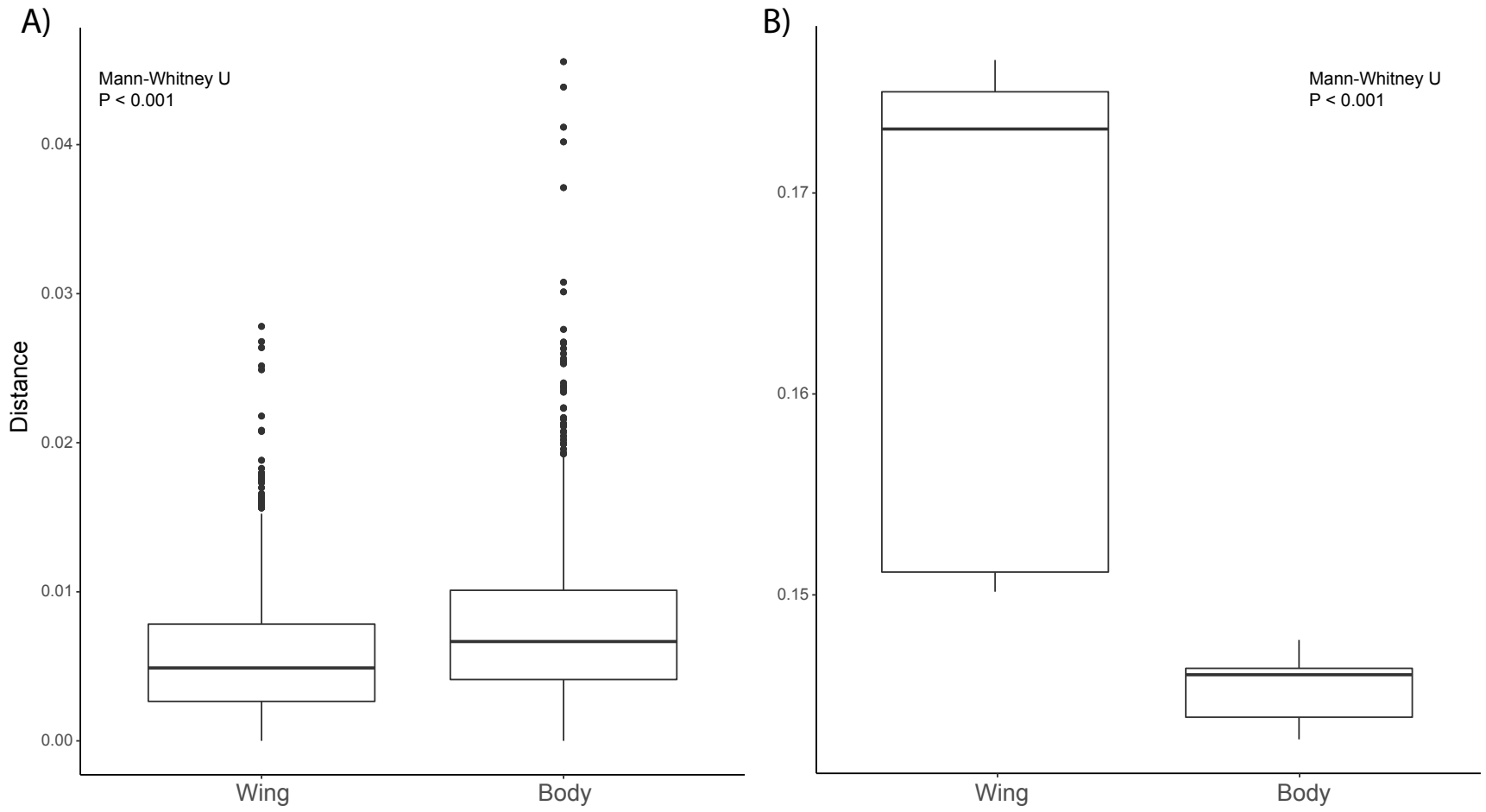


Figure 7.3. Box plots of uncorrected genetic distances between lice from *Metriopelia melanoptera* and lice from *M. ceciliae*. For both wing and body lice, the distribution of nuclear distances (A) are comparisons of 1,006 genes from two individuals. The mitochondrial distances (B) were calculated from 5 genes comparing 3 wing louse and 4 body louse individuals.

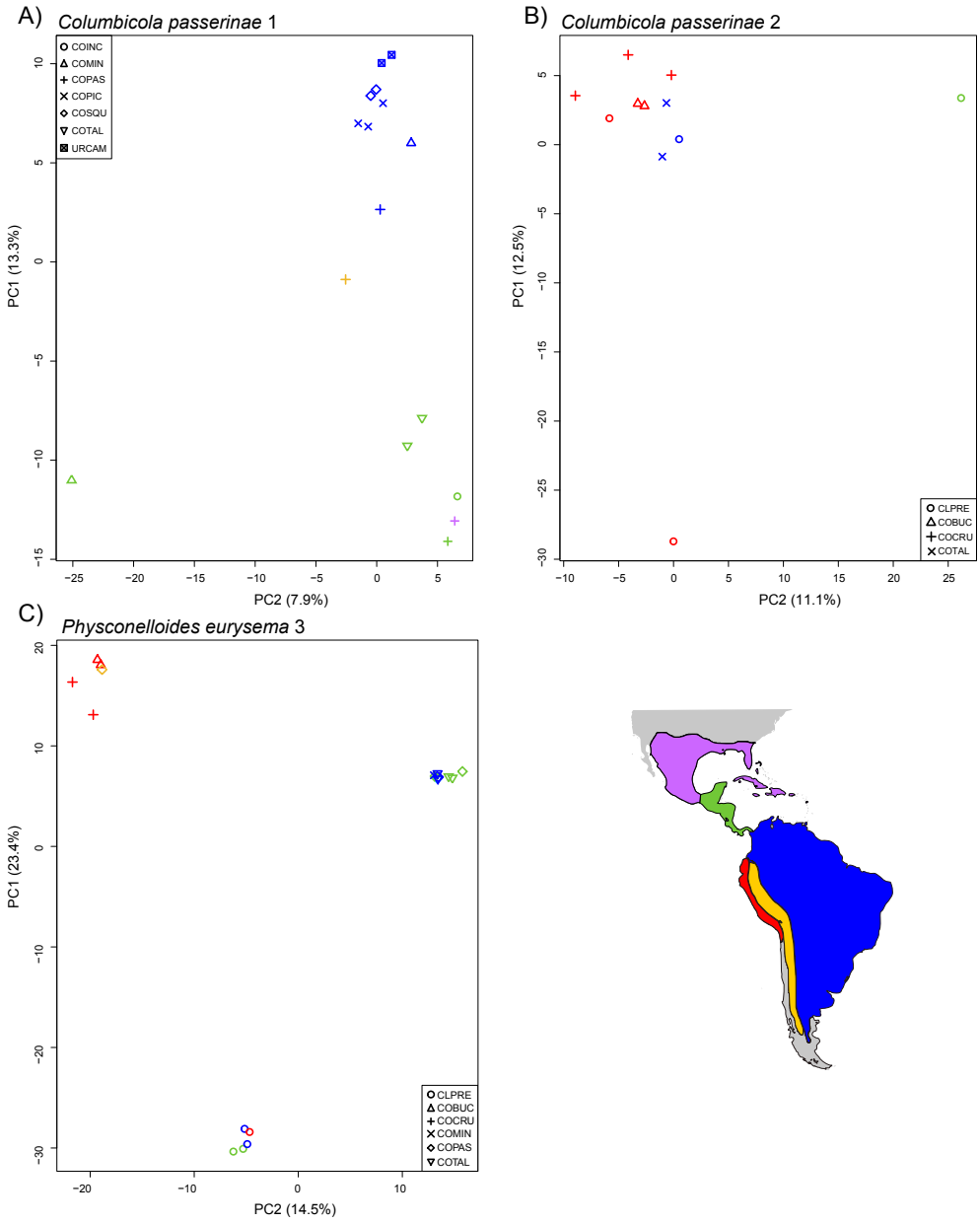


Figure 7.4. PCA plots from SNP data of the ground-dove louse taxa A) *Columbicola passerinae* 1, B) *C. passerinae* 2, and C) *Physconelloides eurysema* 3 (adapted from Sweet et al., 2018). The points represent individual lice; they are colored according to biogeographic regions (see inset map) and shaped according to host species. The host species key shows the first two letters of the genus and first three letters of the species (e.g., Coinc = *Columbina inca*).

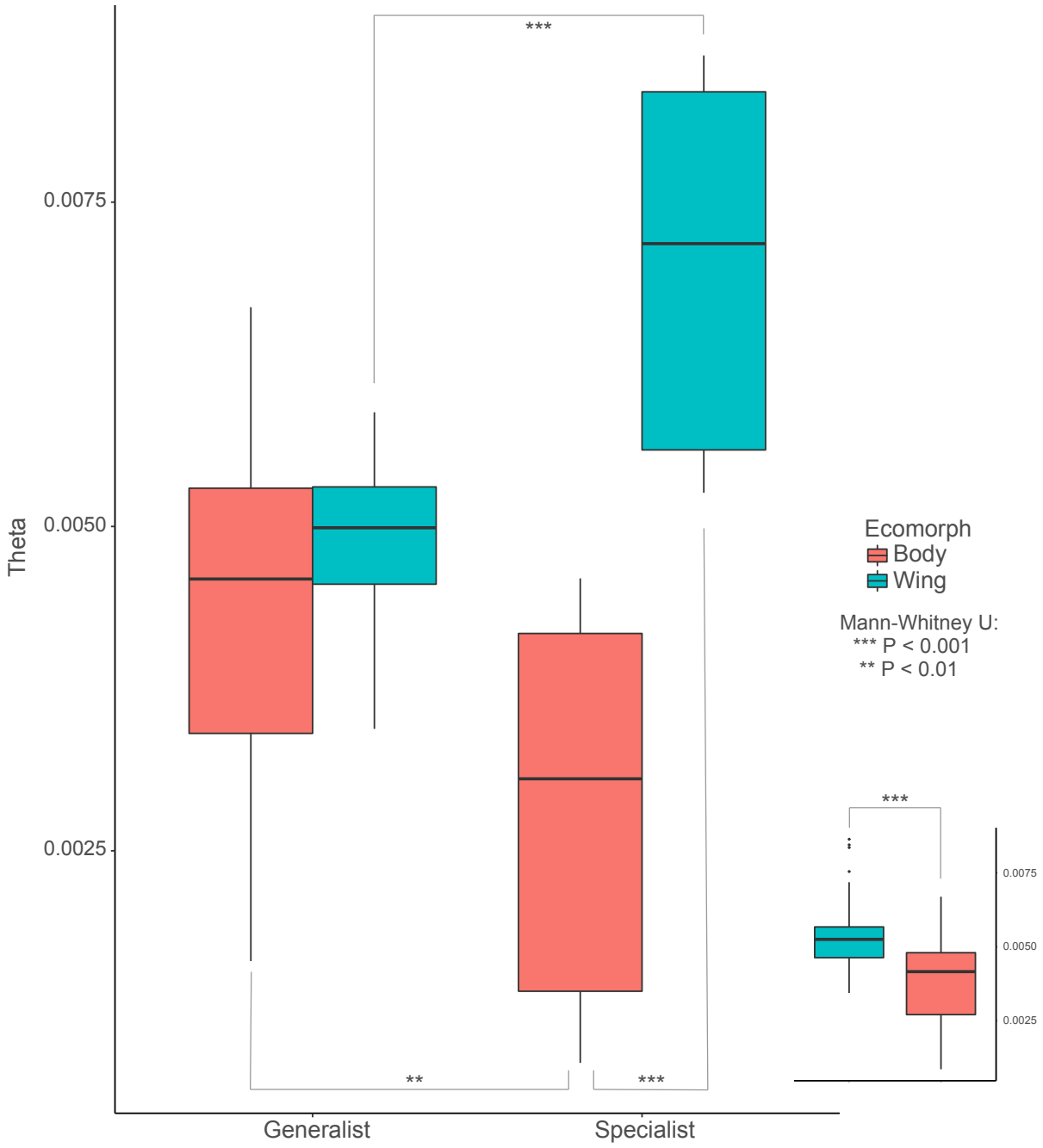


Figure 7.5. Box plots of θ estimated from individual New World ground-dove wing and body louse genomes. The inset plot (bottom right) shows the overall values of θ for wing (blue) and body (red) lice. The main figure shows the values of generalist (associated with multiple host species) and specialist (associated with a single host species) lice. Significantly different distributions are indicated with asterisks.

TABLES

Table 7.1. Comparison of cophylogenetic measures for New World ground-dove wing and body lice. Included are the residuals from PACo, and the ParaFitLink1 and ParaFitLink2 statistics from ParaFit. Average values are reported for host species with multiple wing or body louse associations.

Host	Wing residuals	Body residuals	Wing PF1	Wing PF2	Body PF1	Body PF2
Claravis pretiosa	0.138	0.147	-1.33E-06	-4.05E-04	-1.39E-05	-4.37E-03
Columbina buckleyi	0.070	0.062	1.17E-06	3.57E-04	1.34E-05	4.21E-03
Columbina cruziana	0.069	0.075	3.21E-07	9.78E-05	3.43E-06	1.07E-03
Columbina inca	0.050	0.032	1.57E-06	4.78E-04	8.11E-06	2.54E-03
Columbina minuta	0.046	0.061	1.81E-06	5.51E-04	1.13E-05	3.54E-03
Columbina passerina	0.039	0.057	1.61E-06	4.90E-04	9.92E-06	3.11E-03
Columbina squammata	0.049	0.034	1.56E-06	4.75E-04	8.46E-06	2.65E-03
Columbina talpacoti	0.058	0.062	1.49E-06	4.54E-04	1.35E-05	4.23E-03
Metriopelia ceciliae	0.014	0.042	3.49E-06	1.06E-03	2.80E-05	8.77E-03
Metriopelia melanoptera	0.019	0.046	2.91E-06	8.88E-04	2.58E-05	8.07E-03
Uropelia campestris	0.149	0.066	-2.04E-06	-6.22E-04	1.85E-05	5.79E-03

LITERATURE CITED

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

Allen J.M., Boyd B., Nguyen N.P., Vachaspati P., Warnow T., Huang D., Grady P.G., Bell K.C., Cronk Q.C., Mugisha L., Pittendrigh B.R., Leonardi M.S., Reed D.L., Johnson K.P. 2017. Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology* 66: 786-798.

Balbuena J.A., Míguez-Lozano R., Blasco-Costa I. 2013. PACo: A novel Procrustes application to cophylogenetic analysis. *PLoS ONE* 8: e61048.

Banks J.C., Paterson A.M. 2005. Multi-host parasite species in cophylogenetic studies. *International Journal for Parasitology* 35: 741-746.

Barrett L.G., Thrall P.H., Burdon J.J., Linde C.C. 2008. Life history determines genetic structure and evolutionary potential of host--parasite interactions. *Trends in Ecology & Evolution* 23: 678–685.

Bell K.C., Calhoun K.L., Hoberg E.P., Demboski J.R., Cook J.A. 2016. Temporal and spatial mosaics: Deep host association and shallow geographic drivers shape genetic structure in a widespread pinworm, *Rauschtineria eutamii* (Nematoda: Oxyuridae). *Biological Journal of the Linnean Society* 119: 397–413.

Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* 57: 289–300.

Bordes F., Morand S. 2009. Coevolution between multiple helminth infestations and basal immune investment in mammals: Cumulative effects of polyparasitism? *Parasitology Research* 106: 33-37.

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

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

Brooks D.R., McLennan D.A. 1991. Phylogeny, ecology, and behavior: A research program in comparative biology. The University of Chicago Press, Chicago.

Cameron S.L., Yoshizawa K., Mizukoshi A., Whiting M.F., Johnson K.P. 2011. Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). *BMC Genomics* 12: 394.

Clayton D.H., Lee P.L.M., Tompkins D.M., Brodie E.D. 1999. Reciprocal natural selection on host-parasite phenotypes. *The American Naturalist* 154: 261–270.

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

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

Clayton D.H., Tompkins D.M. 1994. Ectoparasite virulence is linked to mode of transmission. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 256: 211–217.

Clayton D.H. 1991. Coevolution of avian grooming and ectoparasite avoidance. In: J.E. Loye and M. Zuk, eds. Bird-parasite interactions: ecology, evolution, and behavior. Oxford University Press, Oxford.

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

Clayton D., Bush S., Johnson K. 2004. Ecology of congruence: Past meets present. *Systematic Biology* 53: 165–173.

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.

Couch A.B. 1962. Phoretic mallophagans from hippoboscids of mourning doves *Zenaidura macroura*. *The Journal of Parasitology* 48: 497.

Covacin C., Shao R., Cameron S., Barker S.C. 2006. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). *Insect Molecular Biology* 15: 63–68.

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

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

Cruaud A., Rasplus J.-Y. 2016. Testing cospeciation through large-scale cophylogenetic studies. *Current Opinion in Insect Science* 18: 53–59.

Cutter A.D. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution* 69: 1172–1185.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin, R., 1000 Genomes Project Analysis Groups. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

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

Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* 6: 361–375.

Demastes J.W., Spradling T.A., Hafner M.S., Spies G.R., Hafner D.J., Light J.E. 2012. Cophylogeny on a Fine Scale: *Geomydoecus* chewing lice and their pocket gopher hosts, *Pappogeomys bulleri*. *Journal of Parasitology* 98: 262–270.

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

Dybdahl M.F., Lively C.M. 1996. The Geography of coevolution : Comparative population structures for a snail and its trematode parasite. *Evolution*. 50: 2264-2275.

Earl D.A., VonHoldt B.M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Ehrlich P.R., Raven P.H. 1964. Butterflies and plants: A study in coevolution. *Evolution*. 18: 586-608.

Eichler W.D. 1948. Some rules in ectoparasitism. *The Annals and Magazine of Natural History* 1: 588–598.

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

Fahrenholz H. 1913. Ectoparasiten und abstammungslehre. *Zoologischer Anzeiger* 41: 371–374.

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

Hackett-Jones E., Cobbold C., White A. Coexistence of multiple parasitoids on a single host due to differences in parasitoid phenology. 2009. *Theoretical Ecology* 2: 19-31.

Hafner M.S., Page R.D. 1995. Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 349: 77–83.

Hafner M.S., Sudman P.D., Villablanca F.X., Spradling T.A., Demastes J.W., Nadler S.A. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265:1087–1090.

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

Harbison C.W., Clayton D.H. 2011. Community interactions govern host-switching with implications for host-parasite coevolutionary history. *Proceedings of the National Academy of Sciences* 108: 9525–9529.

Harbison C.W., Jacobsen M. V., Clayton D. H. 2009. A hitchhiker's guide to parasite transmission: The phoretic behaviour of feather lice. *International Journal for Parasitology* 39: 569–575.

Haubold B., Pfaffelhuber P., Lynch M. 2010. mlRho - a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Molecular Ecology* 19: 277–284.

Hesse E., Buckling A. 2016. Host population bottlenecks drive parasite extinction during antagonistic coevolution. *Evolution* 70: 235–240.

Hoberg E.P., Brooks D.R. 2008. A macroevolutionary mosaic: Episodic host-switching, geographical colonization and diversification in complex host-parasite systems. *Journal of Biogeography* 35: 1533–1550.

Hutchinson M.C., Cagua E.F., Balbuena J.A., Stouffer D.B., Poisot T. 2017. paco: implementing Procrustean Approach to Cophylogeny in R. *Methods in Ecology and Evolution* 8: 932-940.

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

Jackson A.P., Machado C.A, Robbins N., Herre E.A. 2008. Multi-locus phylogenetic analysis of neotropical figs does not support co-speciation with the pollinators: The importance of systematic scale in fig/wasp cophylogenetic studies. *Symbiosis* 45:16.

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

Johnson K.P., Reed D.L., Hammond Parker S.L., Kim D., Clayton D.H. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for *Columbicola* (Insecta: Phthiraptera). *Molecular Phylogenetics and Evolution* 45: 506–518.

Johnson K.P., Williams B.L., Drown D.M., Adams R.J., Clayton D.H. 2002. The population genetics of host specificity: genetic differentiation in dove lice (Insecta: Phthiraptera). *Molecular Ecology* 11: 25–38.

Johnson K.P., Weckstein J.D., Meyer M.J., Clayton D.H. 2011. There and back again: Switching between host orders by avian body lice (Ischnocera: Goniodidae). *Biological Journal of the Linnean Society* 102: 614–625.

Johnson K.P., Clayton D. 2004. Untangling coevolutionary history. *Systematic Biology* 53: 92–94.

Johnson, K. P., Shreve S. M., Smith V. S. 2012. Repeated adaptive divergence in microhabitat specialization in avian feather lice. *BMC Biology* 10: 52.

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

Keirans J.E. 1975. A review of the phoretic relationship between Mallophaga (Phthiraptera: Insecta) and Hippoboscidae (Diptera: Insecta). *Journal of Medical Entomology* 12: 71–76.

Kirkness E.F., Haas B.J., Sun W., Braig H.R., Perotti M.A., Clark J.M., Lee H., Robertson H.M., Kennedy R.C., Gerlach D., Kriventseva E.V, Elsik G., Graur D., Hill C.A., Veenstra J.A., Tubío J.M.C., Ribeiro J.M.C., Rozas J., Johnston J.S., Reese J.T., Popadic A., Raoult D., Reed D.L., Tomoyasu Y., Mittapalli O., Margam V.M., Li H., Meyer J.M., Johnson R.M., Romero-Severson J., Vanzee J.P., Alvarez-ponce D., Vieira F.G., Guirao-rico S., Anzola J.M., Yoon S., Strycharz J.P., Unger M.F., Christley S., Lobo N.F., Seufferheld M.J., Wang N., Dasch A., Struchiner C.J., Madey G., Hannick L.I., Bidwell S., Joardar V., Caler E., Shao R., Barker S.C., Cameron S., Bruggner R.V, Regier A., Johnson J., Viswanathan L., Utterback R., Sutton G.G., Lawson D., Robert M., Venter J.C., Strausberg R.L., Berenbaum M.R., Collins F.H., Zdobnov E.M., Pittendrigh B.R. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences.* 107:12168-12173

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

Lanfear R., Calcott B., Kainer D., Mayer C., Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14: 82.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

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

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

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

Light J.E., Hafner M.S. 2007. Cophylogeny and disparate rates of evolution in sympatric lineages of chewing lice on pocket gophers. *Molecular Phylogenetics and Evolution* 45: 997-1013.

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

Maa T.C. 1969. A revised checklist and concise host index of Hippoboscidae (Diptera). *Pacific Insects Monograph*. 20: 261-299.

Marussich W.A., Machado C.A. 2007. Host-specificity and coevolution among pollinating and nonpollinating New World fig wasps. *Molecular Ecology* 16: 1925–1946.

Mccoy K.D., Boulinier T., Tirard C., Michalakis Y. 2001. Host specificity of a generalist parasite: Genetic evidence of sympatric host races in the seabird tick *Ixodes uriae. Journal of Evolutionary Biology* 14: 395–405.

McCoy K.D., Boulinier T., Chardine J.W., Danchin E., Michalakis Y. 1999. Dispersal and Distribution of the Tick *Ixodes uriae* within and among seabird host populations: The need for a population genetic approach. *The Journal of Parasitology* 85: 196-202.

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

McCoy K.D., Boulinier T., Tirard C., Michalakis Y. 2003. Host-dependent genetic structure of parasite populations: Differential dispersal of seabird tick host races. *Evolution* 57: 288–296.

Miller M.A., Pfeiffer W., Schwartz T. 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., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: i44-i52.

Moran N.A., Baumann P. 2000. Bacterial endosymbionts in animals. *Current Opinions in Microbiology*. 3: 270-275.

Nadler S. 1995. Microevolution and the genetic structure of parasite populations. *The Journal of Parasitology* 81: 395–403.

Nei M., Maruyama T., Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29: 1–10.

Nelson B.C., Murray M.D. 1971. The distribution of Mallophaga on the domestic pigeon (*Columba livia*). *International Journal for Parasitology* 1: 21–29.

Nieberding C.M., Durette-Desset M.-C., Vanderpoorten A., Casanova J.C., Ribas A., Deffontaine V., Feliu C., Morand S., Libois R., Michaux J.R. 2008. Geography and host biogeography matter for understanding the phylogeography of a parasite. *Molecular Phylogenetics and Evolution* 47: 538-554.

Page R.D.M. 1993. Parasites, phylogeny and cospeciation. *International Journal for Parasitology* 23: 499–506.

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

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.

Paterson A.M., Wallis G.P., Wallis L.J., Gray R.D. 2000. Seabird and louse coevolution: Complex histories revealed by 12S rRNA sequences and reconciliation analyses. *Systematic Biology* 49: 383-399.

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

Poulin R. 2007. Are there general laws in parasite ecology? *Parasitology* 134: 763-776.

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

Price P.W. 1980. Evolutionary biology of parasites. Princeton University Press.

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

Puillandre N., Lambert A., Brouillet S., Achaz G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21: 1864–1877.

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

Ricklefs R.E., Fallon S.M., Bermingham E. 2004. Evolutionary relationships, cospeciation, and host switching in avian malaria parasites. *Systematics Biology*. 53: 111-119.

Rosenberg N. A. 2004. DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.

Rothschild M., Clay T. 1952. Fleas, flukes, and cuckoos: A study of bird parasites. 3rd ed. Collins, London.

Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.

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

Sorenson M.D., Balakrishnan C.N., Payne R.B. 2004. Clade-limited colonization in brood parasitic finches (*Vidua* spp.). *Systematic Biology* 53: 140-153.

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

Stefka J., Hoeck P.E., Keller L.F., Smith V.S. 2011. A hitchhikers guide to the Galapagos: Co-phylogeography of Galapagos mockingbirds and their parasites. *BMC Evolutionary Biology* 11: 284.

Sweet A. D., Boyd B. M., Johnson K. P. 2016. Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. *Biological Journal of the Linnean Society* 118: 813–828.

Sweet A.D., Johnson K.P. 2016. Cophylogenetic analysis of New World ground-doves (Aves: Columbidae) and their parasitic wing lice (Insecta: Phthiraptera: *Columbicola*). *Molecular Phylogenetics and Evolution* 103: 122–132.

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

Tompkins D.M., Clayton D.H. 1999. Host resources govern the specificity of swiftlet lice: Size matters. *Journal of Animal Ecology* 68: 489–500.

Vaidya G., Lohman D.J., Meier R. 2011. SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27: 171–180.

Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., Del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K.V, Altshuler D., Gabriel S., DePristo M.A. 2013. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 43: 11.10.1-33.

Vinarski M.V., Korallo N.P., Krasnov B.R., Shenbrot G.I., Poulin R. 2007. Decay of similarity of gamasid mite assemblages parasitic on Palaearctic small mammals: Geographic distance, host-species composition or environment. *Journal of Biogeography* 34: 1691–1700.

Vogwill T., Fenton A., Brockhurst M.A. 2008. The impact of parasite dispersal on antagonistic host-parasite coevolution. *Journal of Evolutionary Biology* 21: 1252–1258.

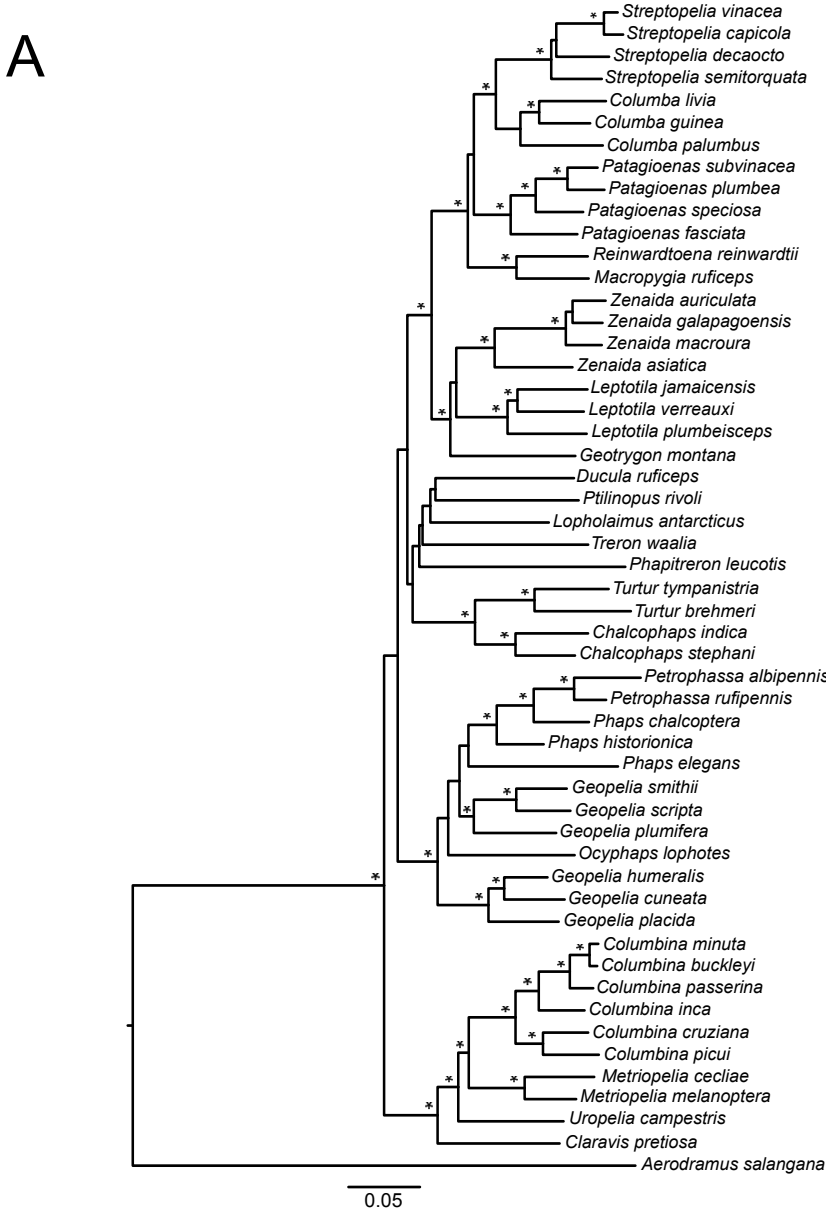
Weckstein J.D. 2004. Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology* 53: 154–164.

Weiblen G. D., Bush G. L. 2002. Speciation in fig pollinators and parasites. *Molecular Ecology* 11:1573–1578. Wiley Online Library.

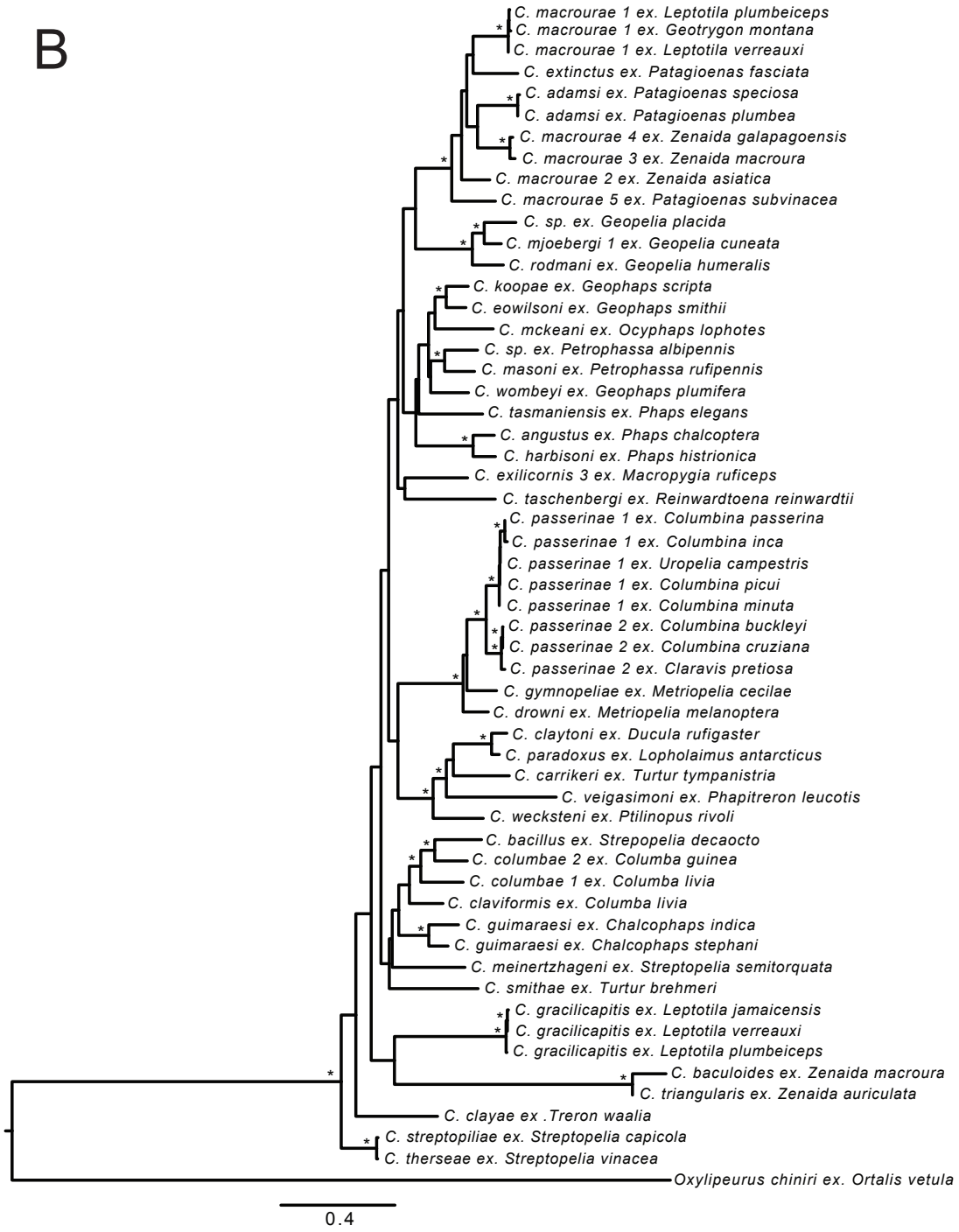
Whiteman N.K., Kimball R.T., Parker P.G. 2007. Co-phylogeography and comparative population genetics of the threatened Galápagos hawk and three ectoparasite species: Ecology shapes population histories within parasite communities. *Molecular Ecology* 16: 4759–4773.

APPENDIX A: SUPPLEMENTARY FIGURES

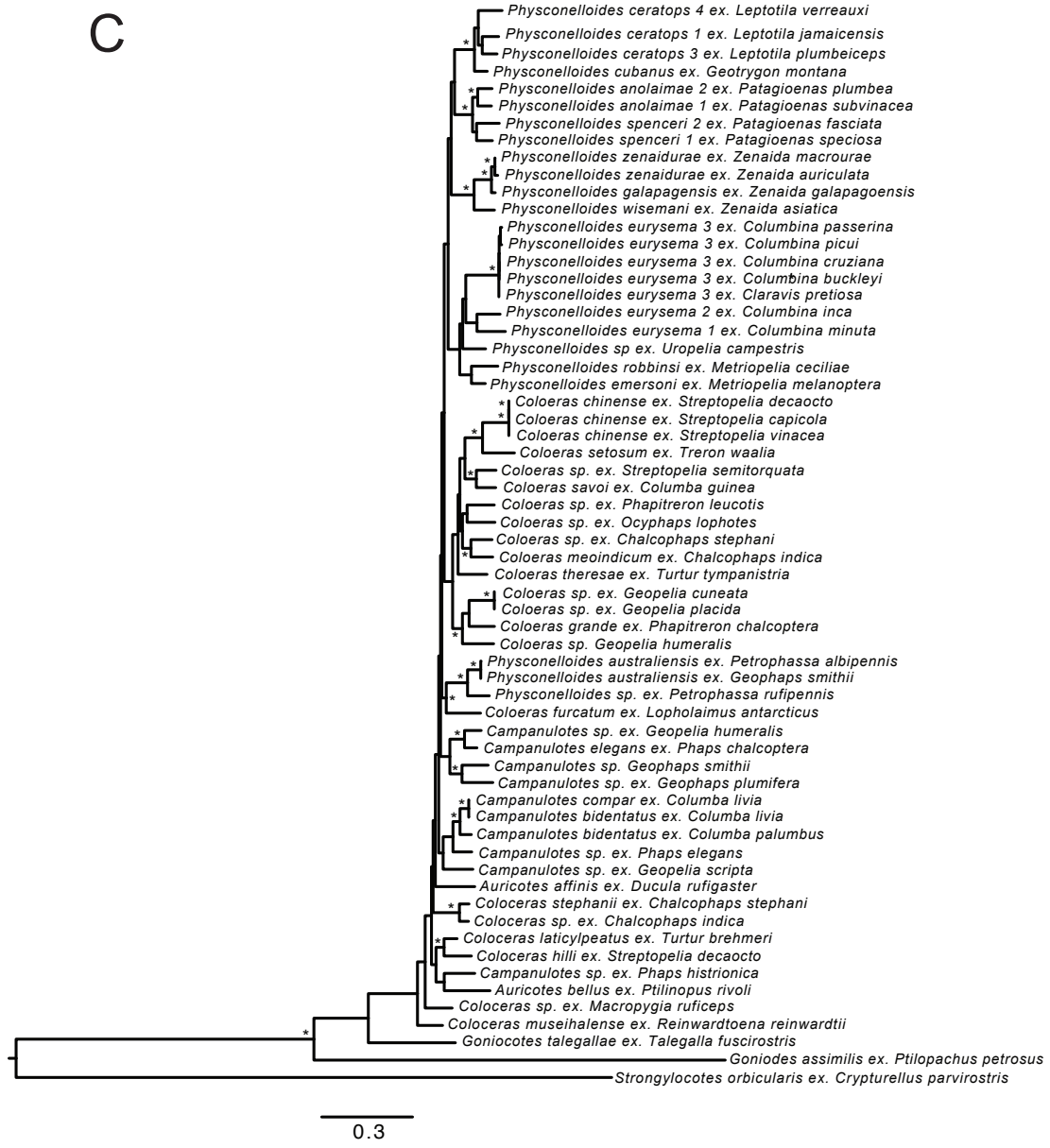
Supplementary Figures for Chapter 2



Supplementary Figure 2.1. Maximum likelihood phylogeny of A) doves, B) dove wing lice, and C) dove body lice. All phylogenies were estimated from 500 boostrap (BS) replicates in RAxML v7.0.4 (Stamatakis, 2006). Asterisks (*) indicate BS values ≥70. Scale bars indicate nucleotide substitutions per site.

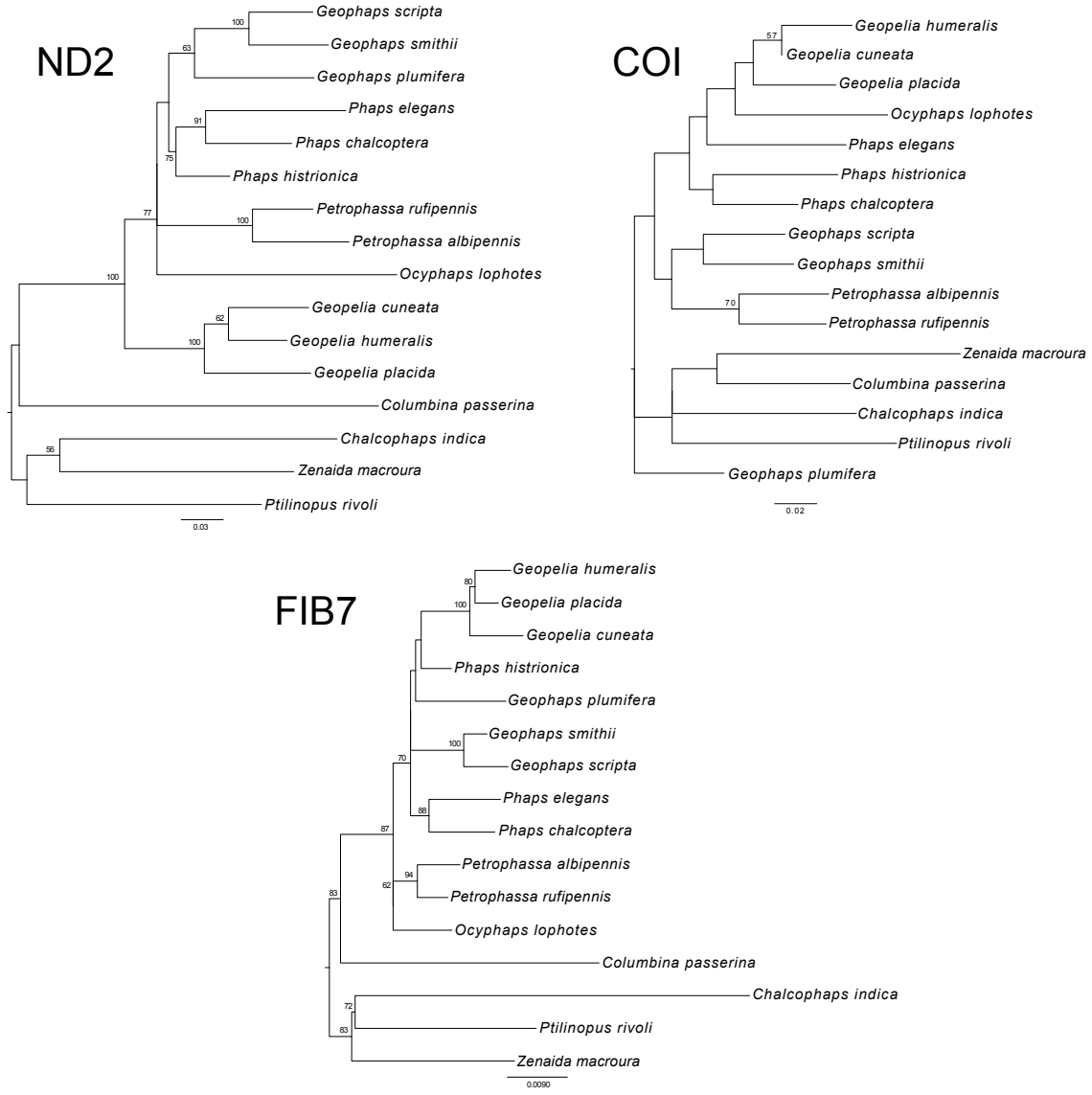


Supplementary Figure 2.1. Continued

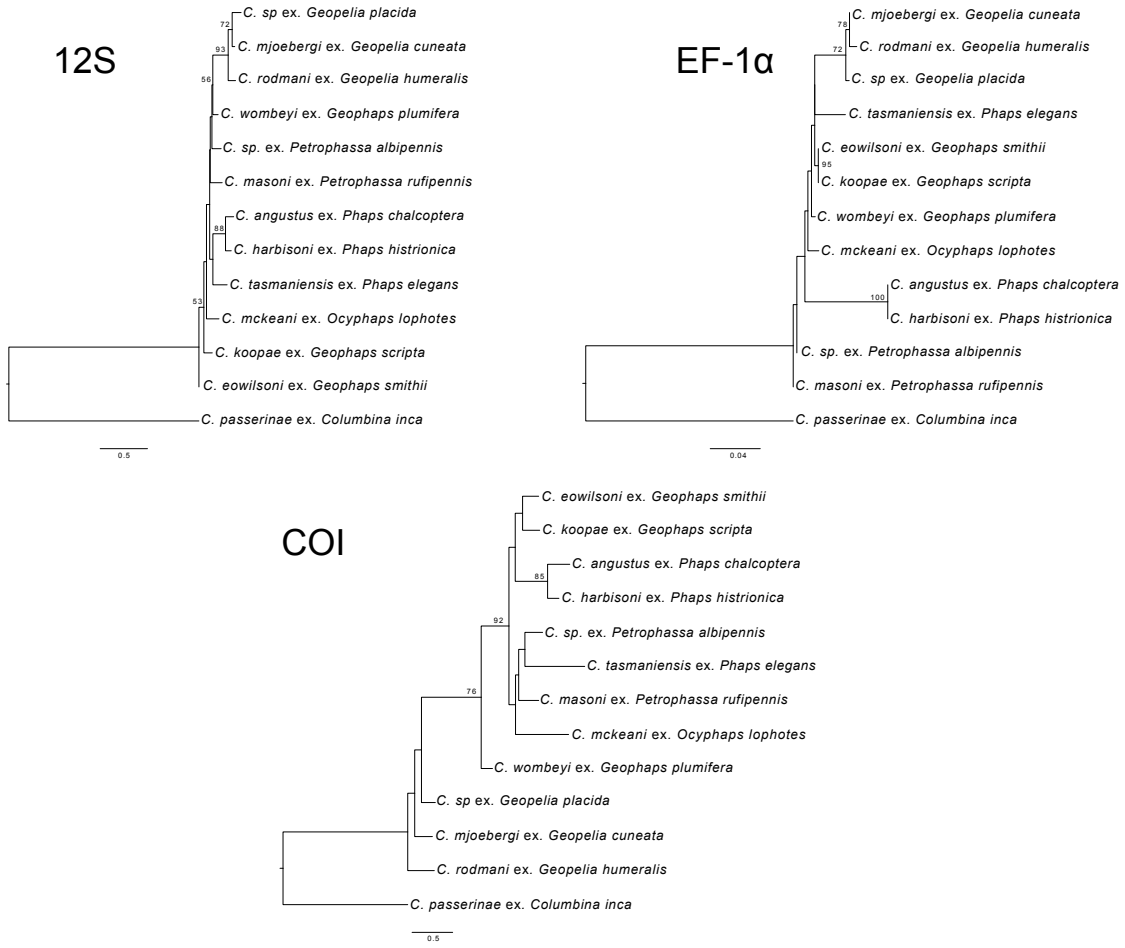


Supplementary Figure 2.1. Continued.

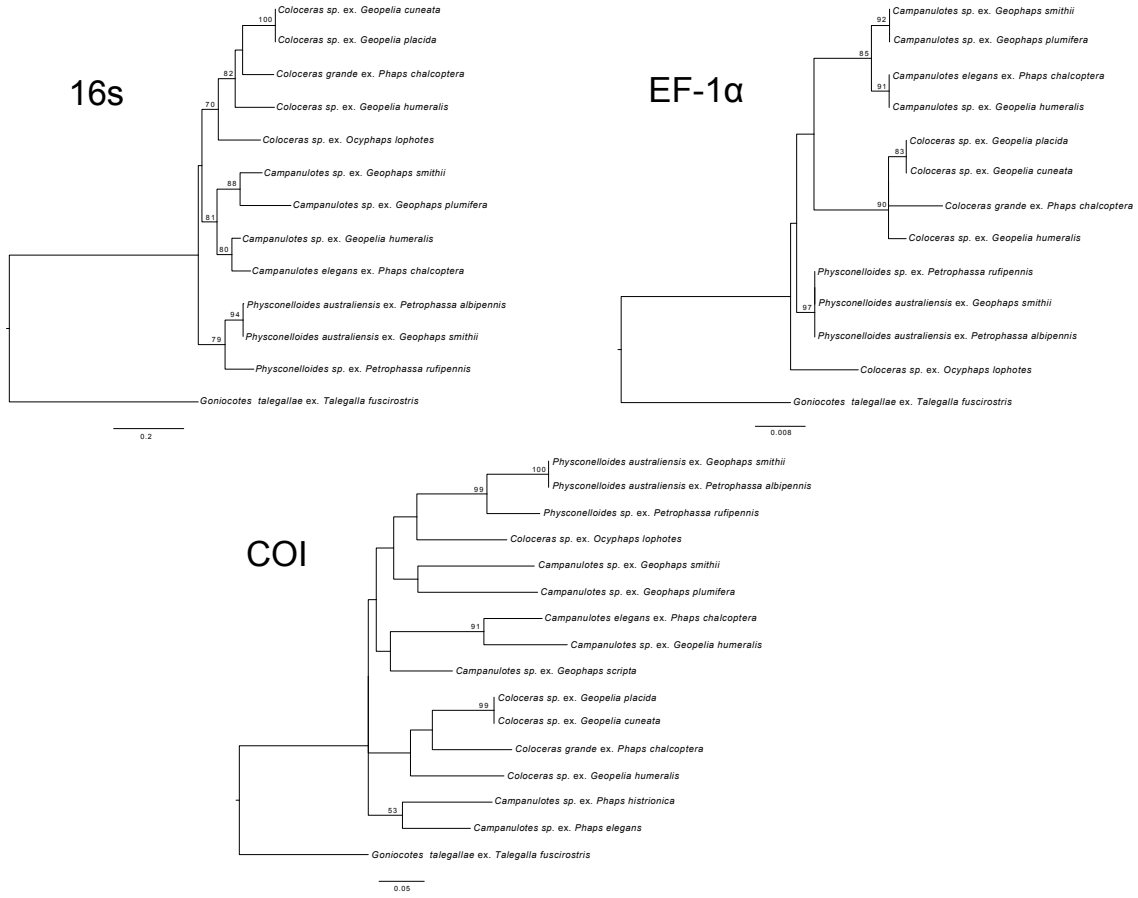
Supplementary Figures for Chapter 3



Supplementary Figure 3.1. Gene trees for phabine pigeons and doves estimated in RAxML for NADH dehydrogenase subunit 2 gene (ND2), mitochondrial cytochrome oxidase subunit 1 gene (*Cox*1), and gene for nuclear beta-fibrinogen intron 7 (FIB7). Bootstrap values ≥50 are indicated before nodes. Branch lengths are in nucleotide substitutions per site.

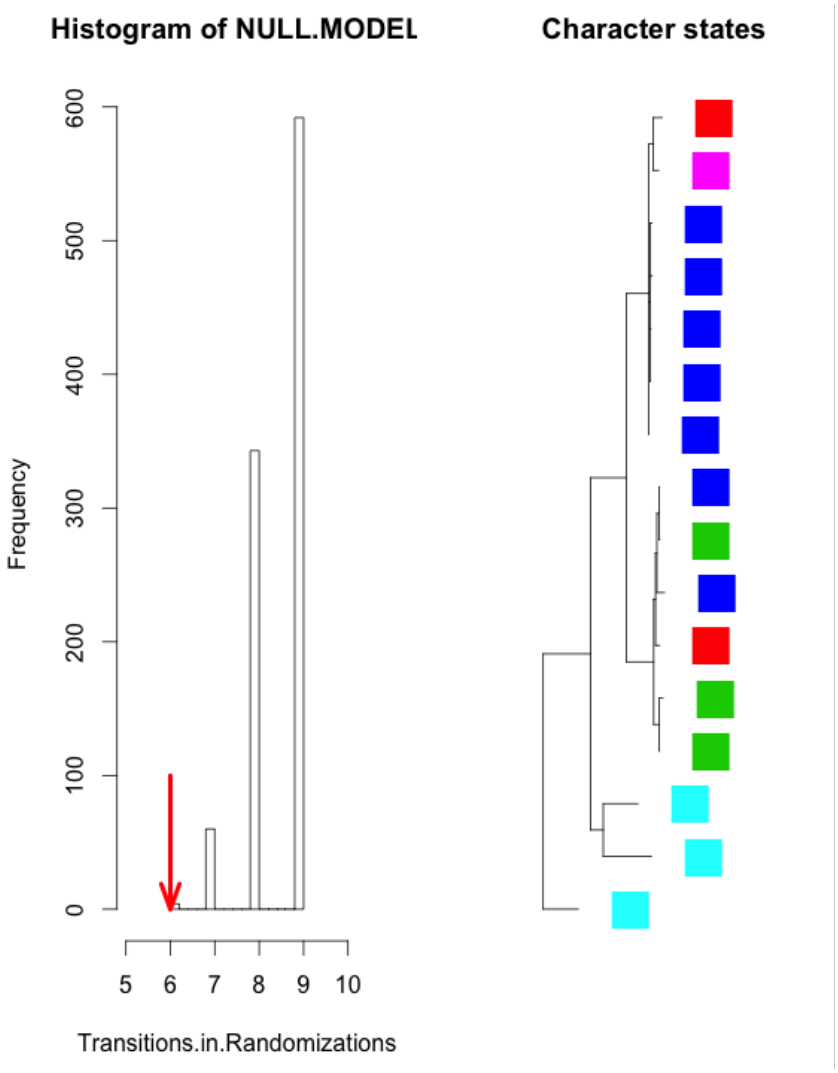


Supplementary Figure 3.2. Gene trees for phabine wing lice estimated in RAxML for 12S rRNA gene (12S), gene for elongation factor 1α (EF-1α), and mitochondrial cytochrome oxidase subunit 1 gene (*Cox*1). Bootstrap values ≥50 are indicated before nodes. Branch lengths are in nucleotide substitutions per site.



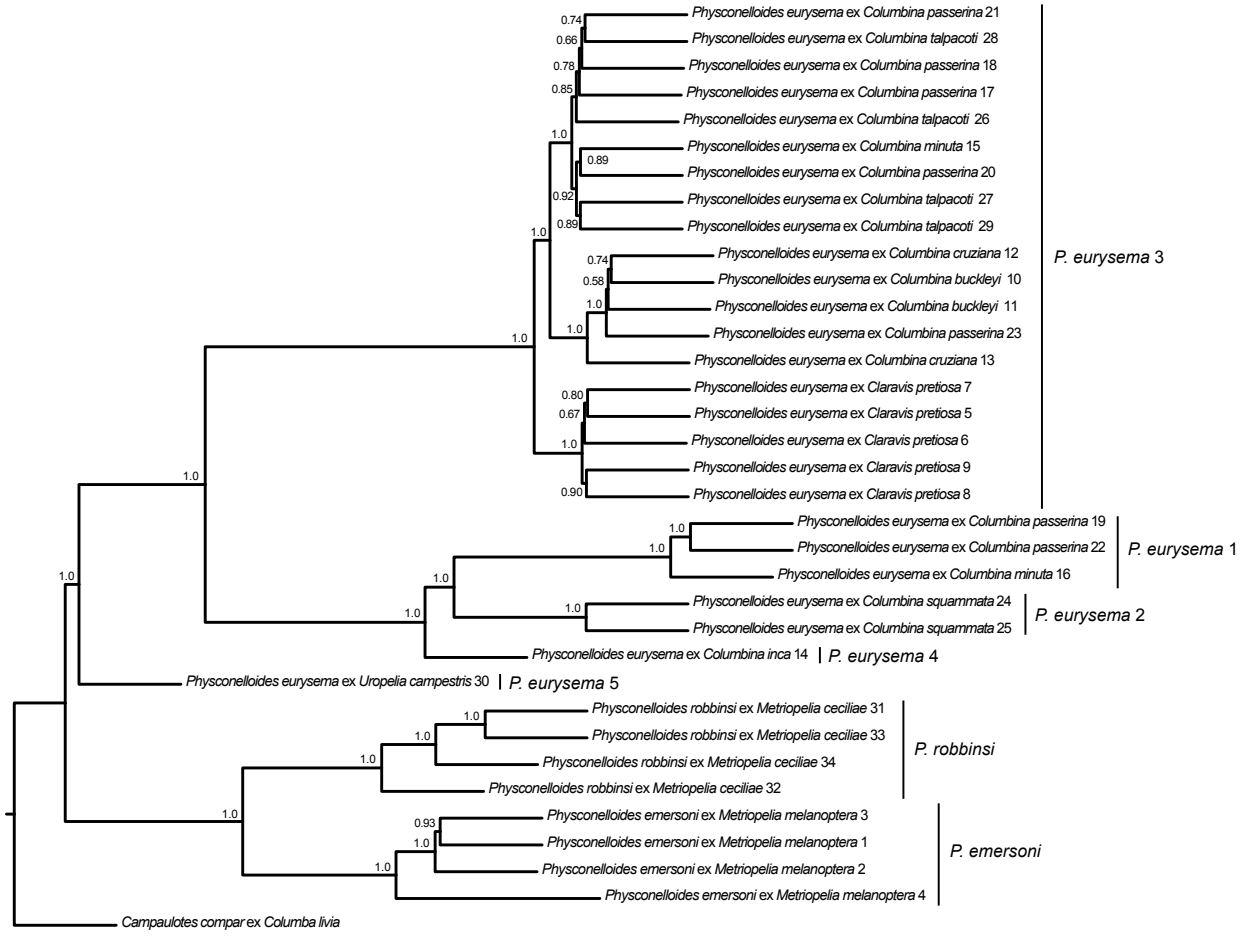
Supplementary Figure 3.3. Gene trees for phabine body lice estimated in RAxML for 16S rRNA gene (16S), gene for elongation factor 1α (EF-1α), and mitochondrial cytochrome oxidase subunit 1 gene (*Cox*1). Bootstrap values ≥50 are indicated before nodes. Branch lengths are in nucleotide substitutions per site.

Supplementary Figures for Chapter 5

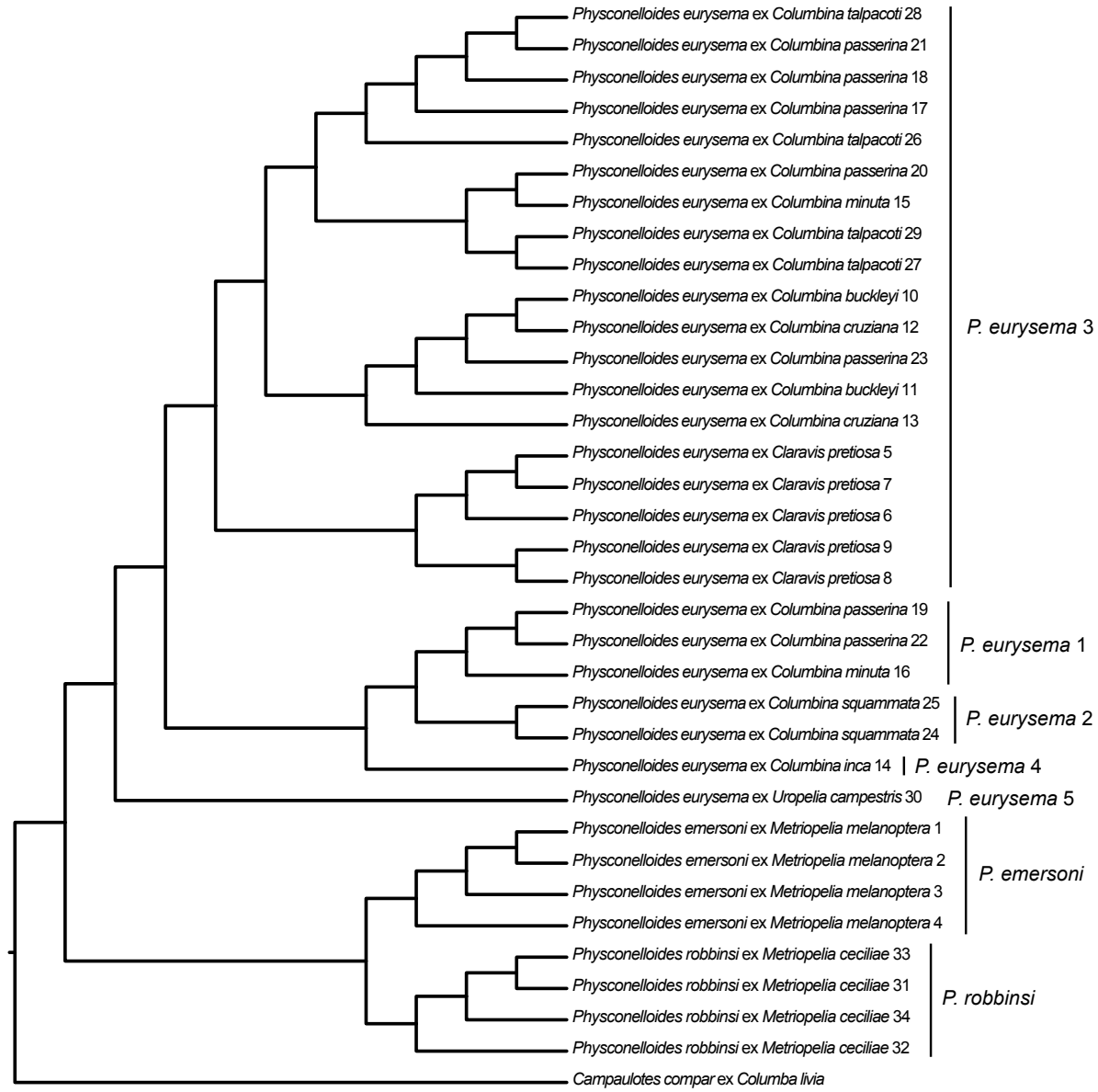


Supplementary Figure 5.1. Results from the Maddison-Slatkin procedure testing for significant biogeographic structure within the ground-dove *Columbicola* phylogeny. Character states on the phylogeny are colored according to host biogeography. The histogram indicates the number of character state transitions from 999 randomizations across the *Columbicola* phylogeny. The observed number of transitions is indicated with a red arrow.

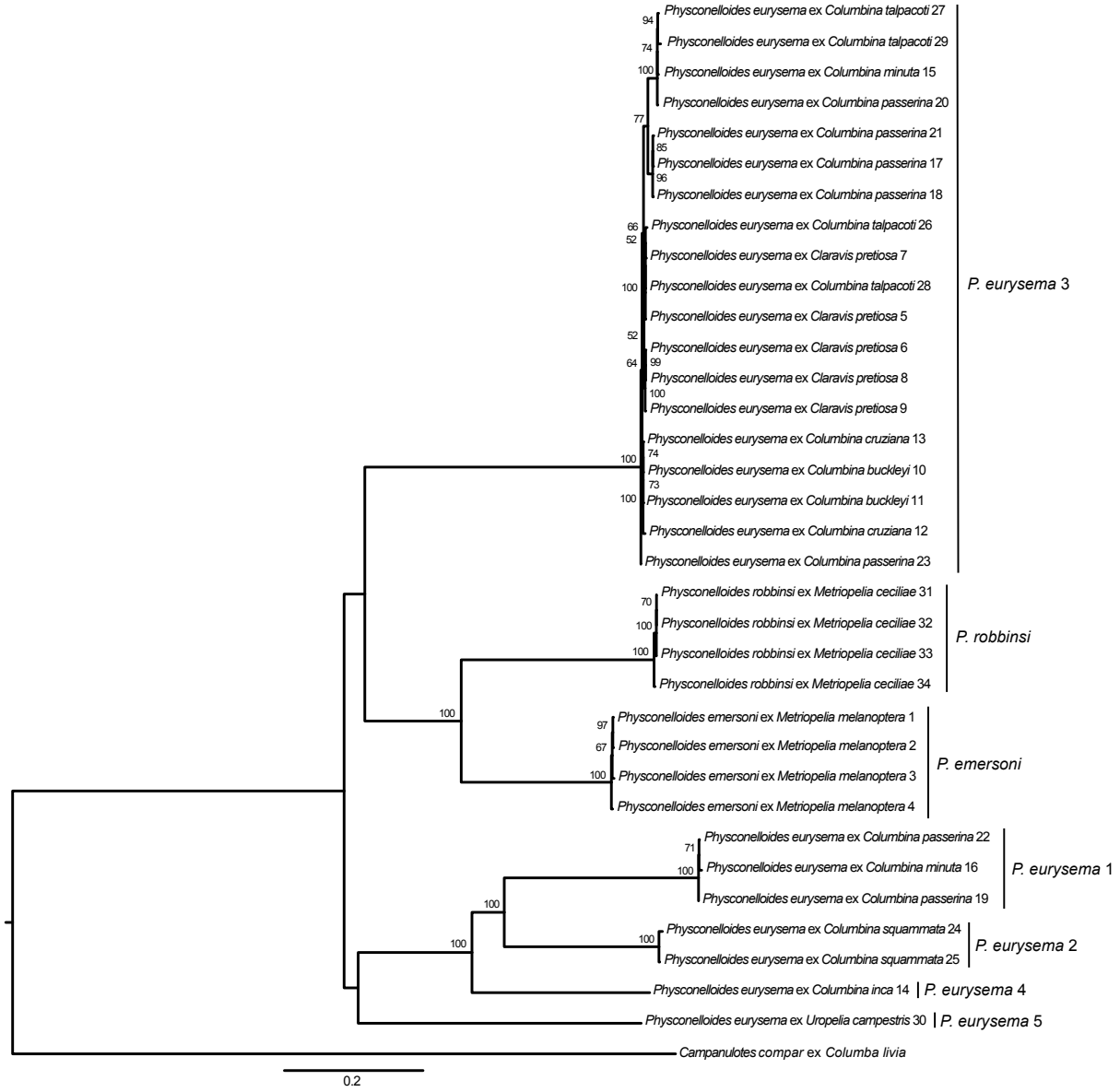
Supplementary Figures for Chapter 6



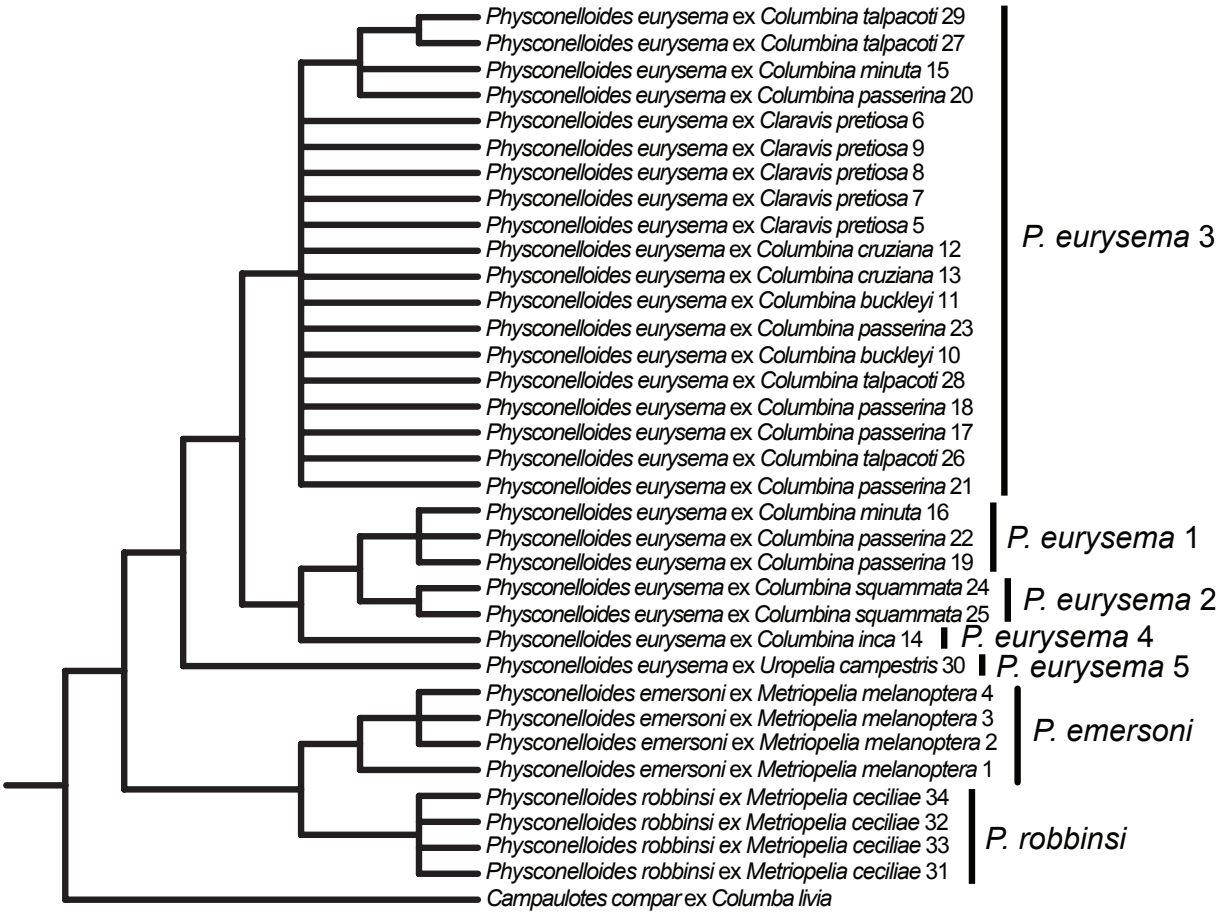
Supplementary Figure 6.1. ASTRAL phylogeny from gene trees of body lice from small New World ground-doves. Vertical lines to the right of the phylogeny indicate taxa recovered from OTU analyses. Local posterior probabilities are indicated at each node.



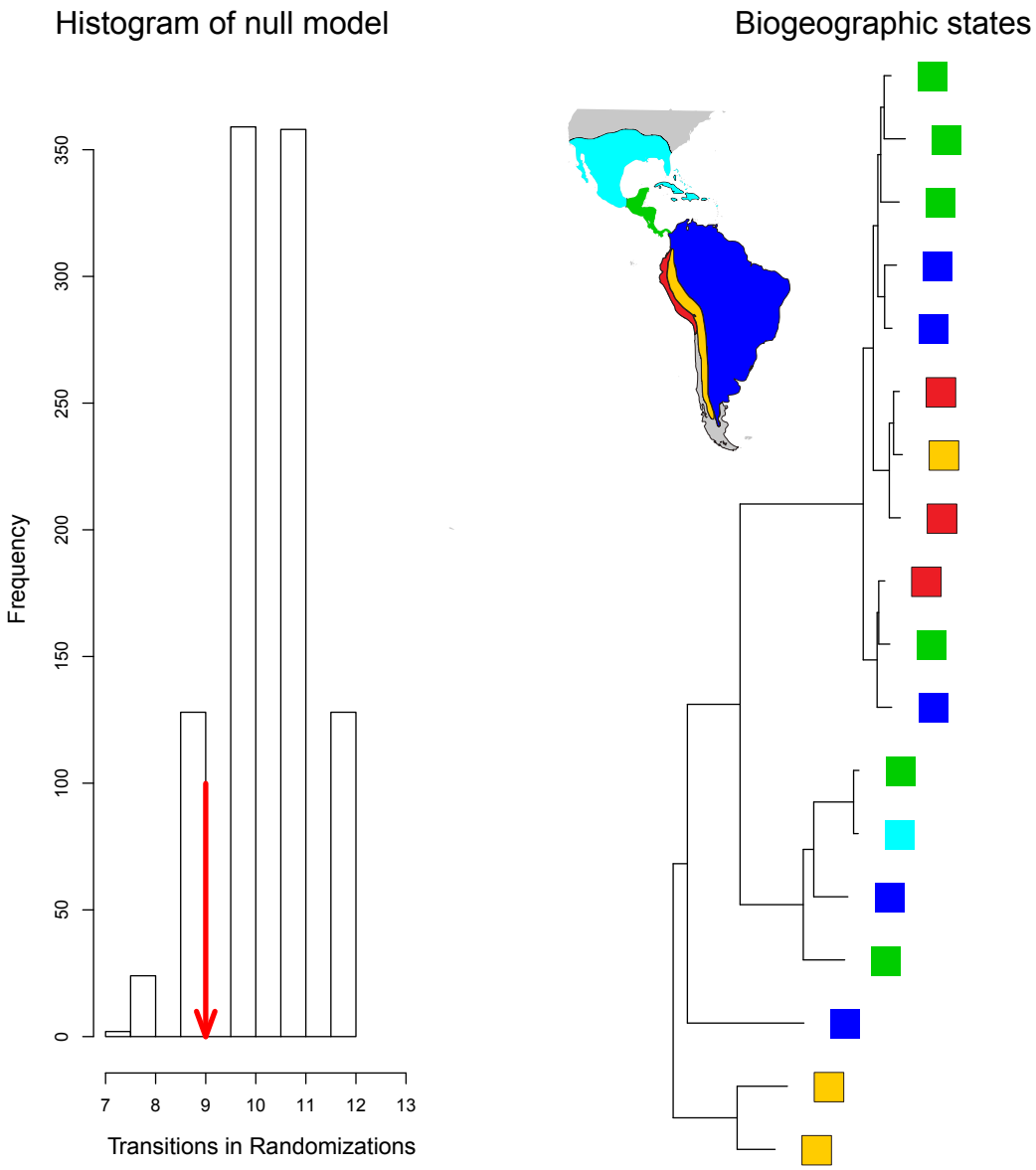
Supplementary Figure 6.2. ASTRID cladogram from gene trees of body lice from small New World ground-doves. Vertical lines to the right of the phylogeny indicate taxa recovered from OTU analyses.



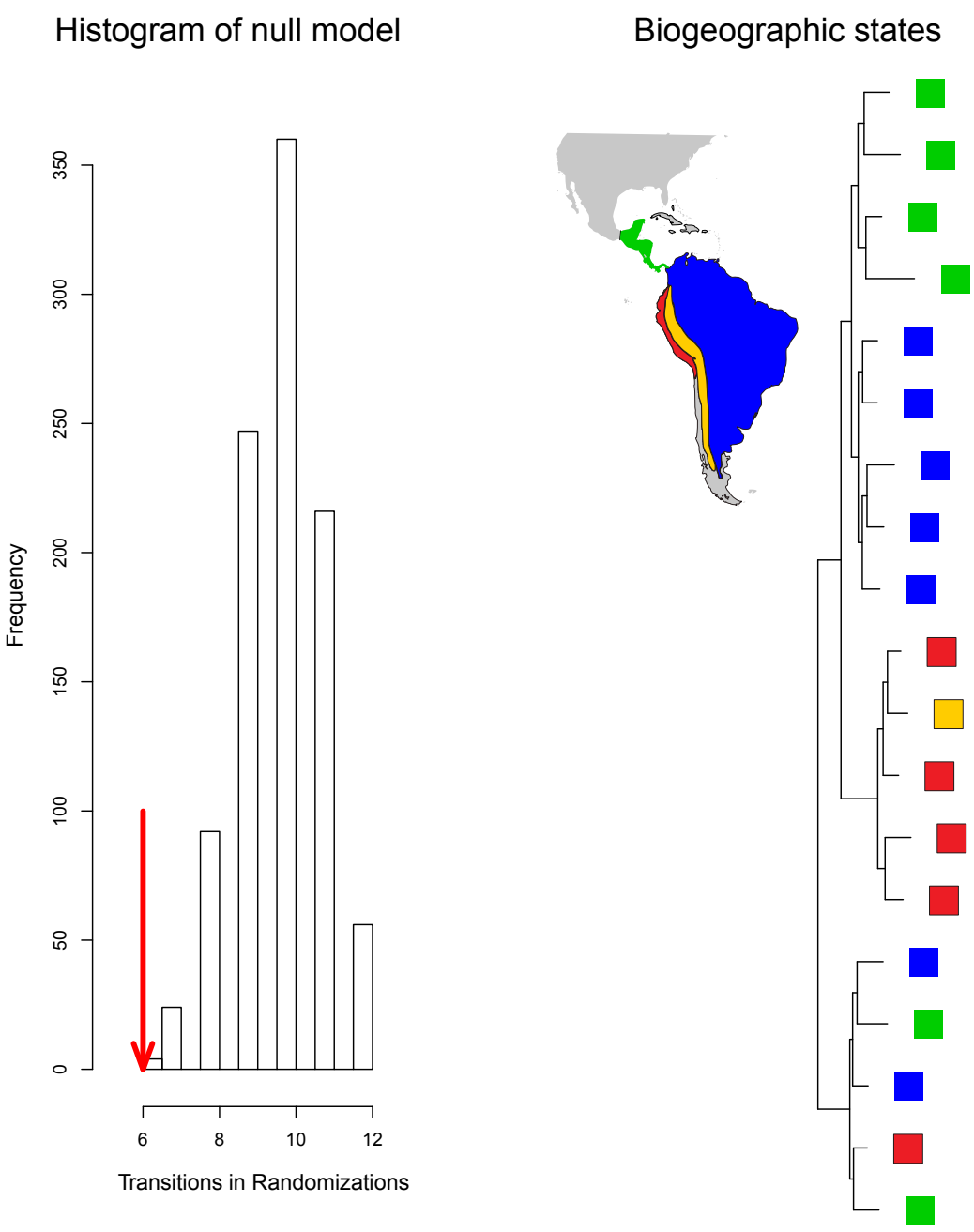
Supplementary Figure 6.3**.** Maximum likelihood phylogeny from mitochondrial sequence data of body lice from small New World ground-doves. Bootstrap support values >50% are indicated at each node. Scale bar indicates nucleotide substitutions per site. Vertical lines to the right of the phylogeny indicate taxa recovered from OTU analyses.



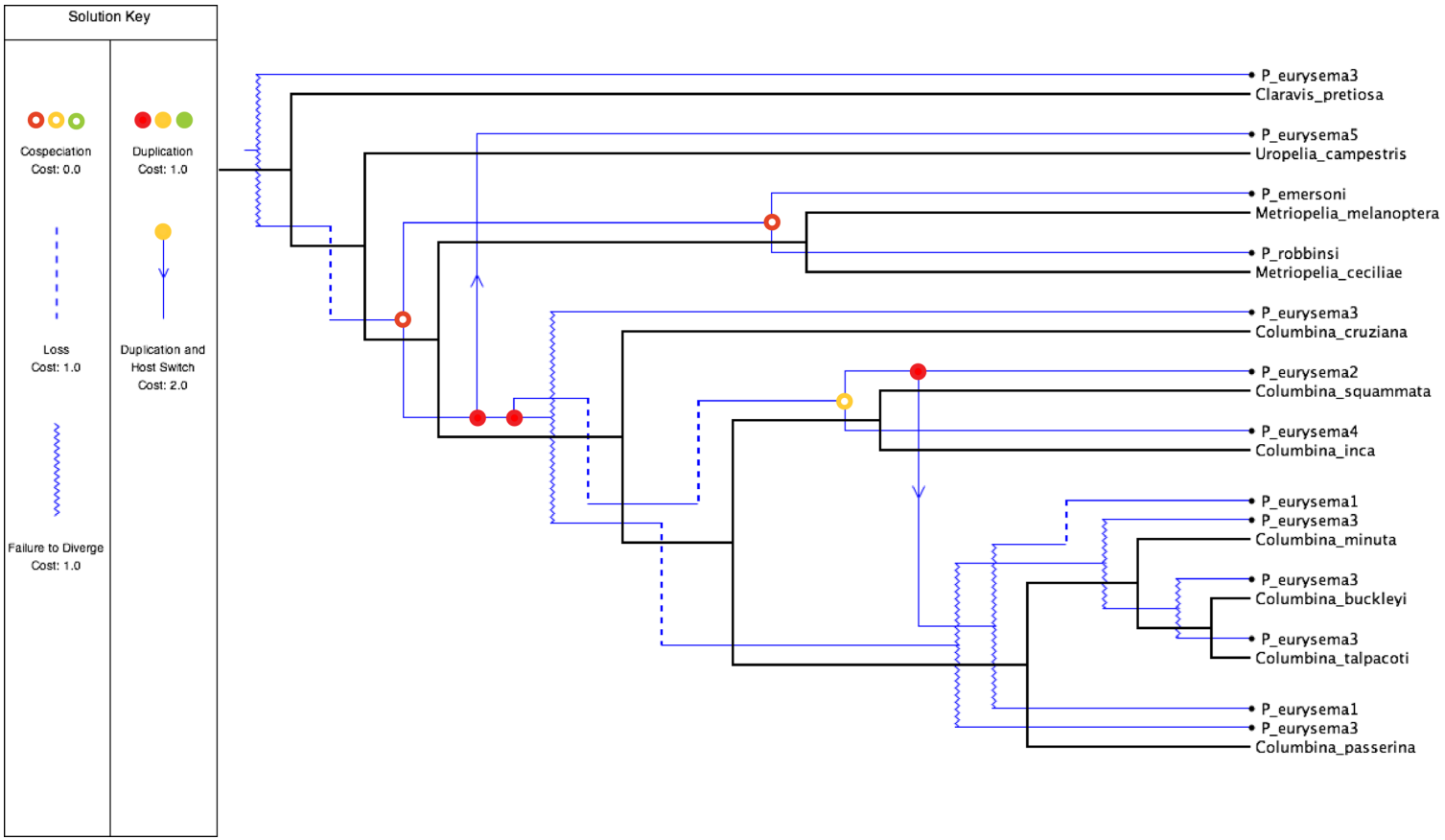
Supplementary Figure 6.4. Summary of phylogenetic relationships among *Physconelloides* lice from small New World ground-doves. This is a strict consensus tree of 50% majority-rule consensus phylogenies from the concatenated, gene tree, and mitochondrial data sets for the lice. Vertical lines to the right of the tip labels indicate the taxa recovered from OTU analyses.



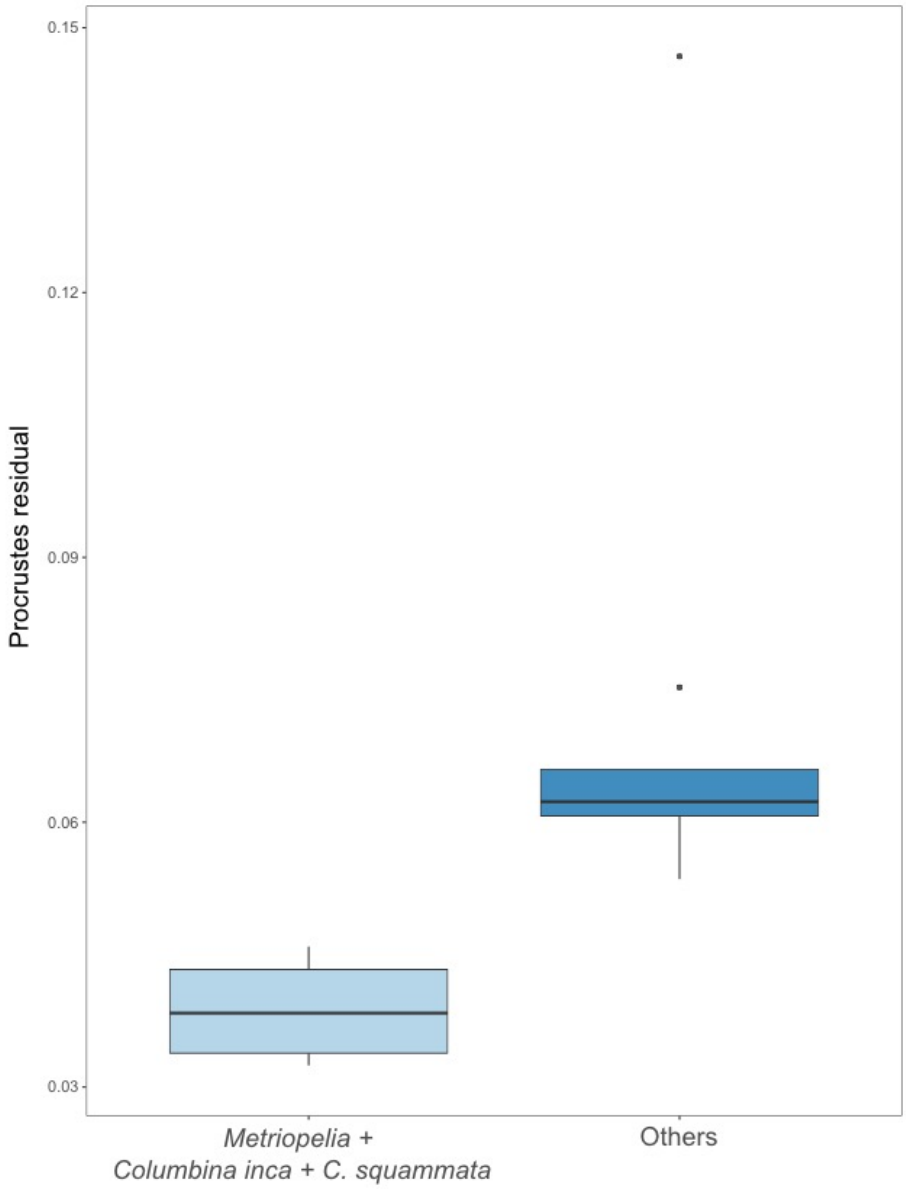
Supplementary Figure 6.5. Biogeographic states and Maddison-Slatkin randomization results of *Physconelloides* lice from small New World ground-doves. Tips are colored according to the map. The phylogeny is from the concatenated nuclear phylogeny, with identical tips collapsed. Individuals were considered identical if they were separated by short branch lengths, and were from the same host species and geographic region. The red line indicates the observed number of character state transitions.



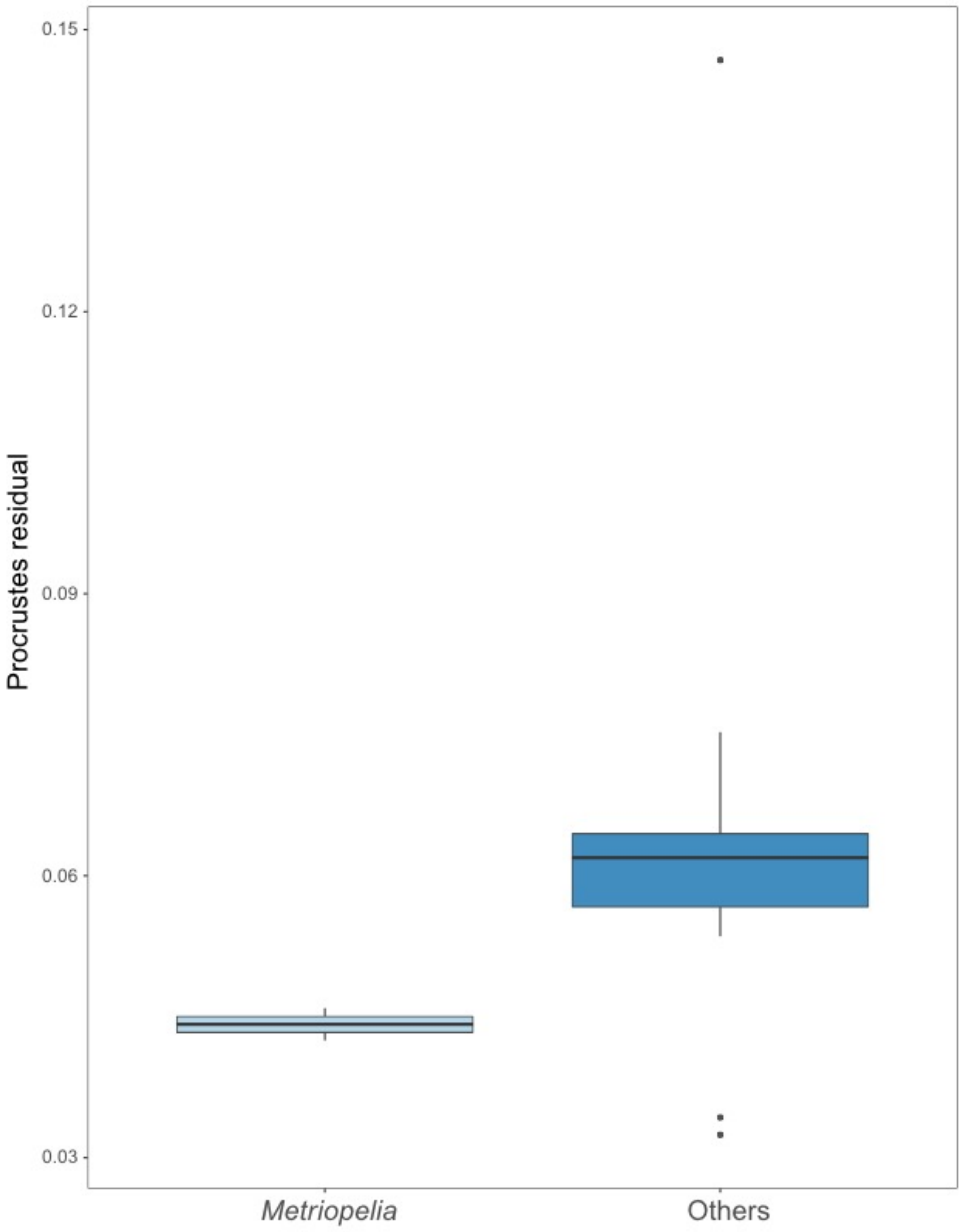
Supplementary Figure 6.6. Biogeographic states and Maddison-Slatkin randomization results of *Physconelloides eurysema* 3 lice from small New World ground-doves. Tips are colored according to the map. The phylogeny is from the concatenated nuclear phylogeny. The red line indicates the observed number of character state transitions.



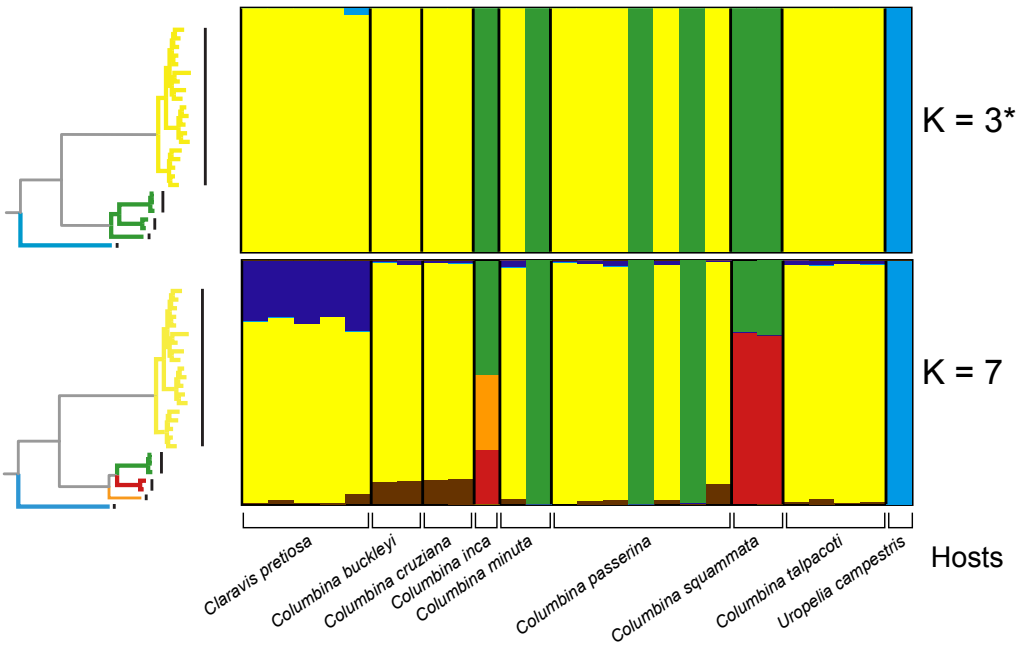
Supplementary Figure 6.7. Reconciliation of phylogenetic trees of small New World ground-doves and their parasitic body lice. Results are an optimal solution from Jane4, an event-based cophylogenetic method. The parasite tree is represented with blue lines, and the host tree with black lines. Coevolutionary "events" are represented on the figure as indicated in the Solution Key. Red circles indicate no equally optimal solution exits, whereas yellow circles indicate at least one other equally optimal solution exists.



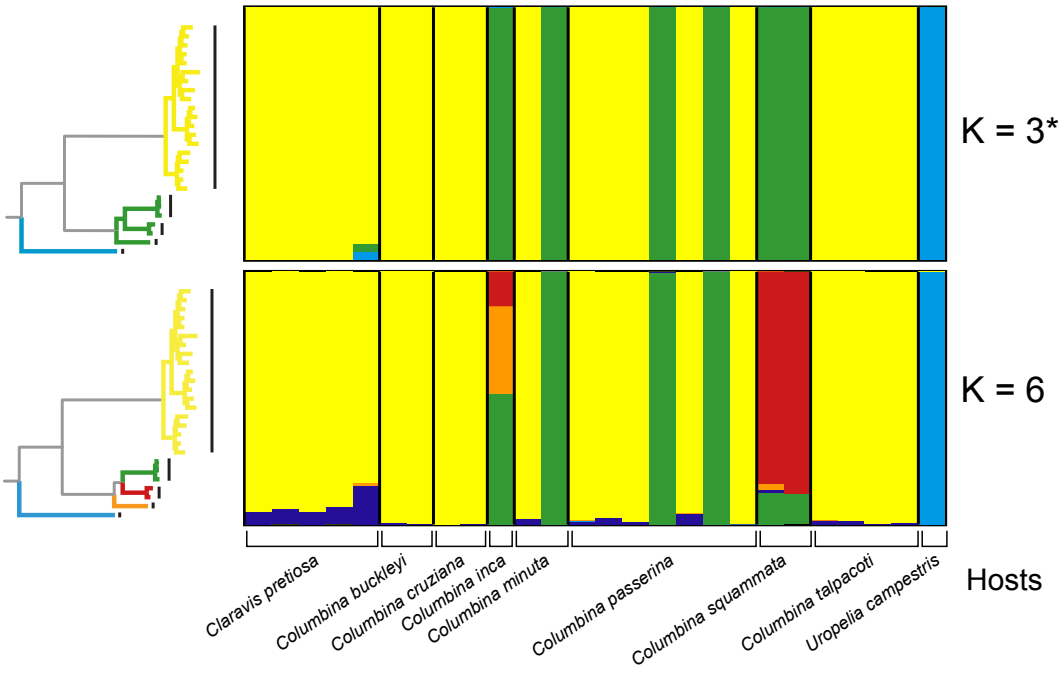
Supplementary Figure 6.8. Box-and-whisker plot of jackknifed Procrustes squared residuals from individual links between small New World ground-doves and their *Physconelloides* body lice. Lower residual values suggest a greater contribution to phylogenetic congruence. Links associated with cospeciation events recovered from event-based analysis are represented by the left box (light blue). All other links are represented by the left box (dark blue). The two sets of links are significantly different (t = -3.32, P = 0.008).



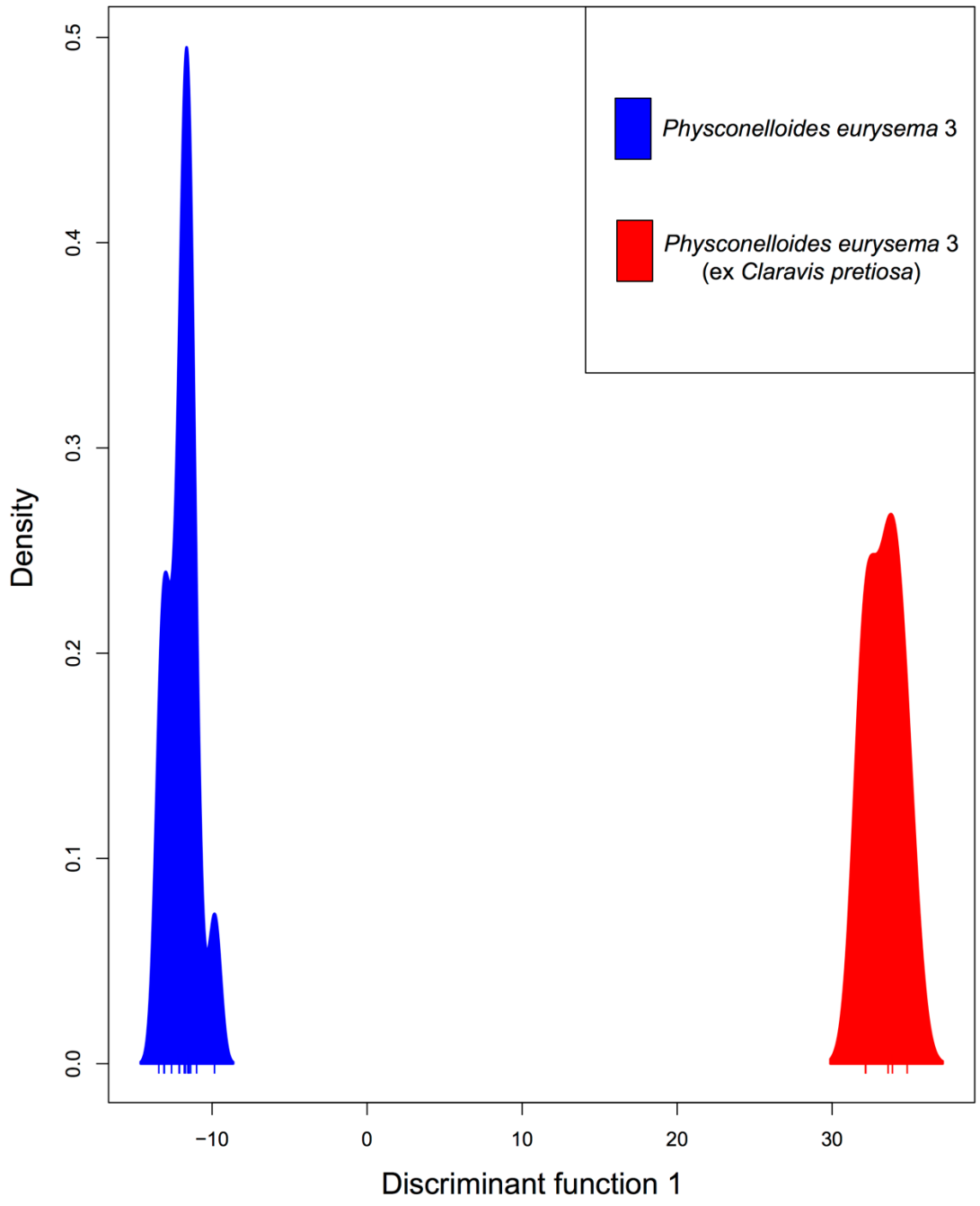
Supplementary Figure 6.9. Box-and-whisker plot of jackknifed Procrustes squared residuals from individual links between small New World ground-doves and their *Physconelloides* body lice. Lower residual values suggest a greater contribution to phylogenetic congruence. Links that had significant ParaFitLink1 statistics are represented by the left box (light blue). All other links are represented by the left box (dark blue). The two sets of links are significantly different (t = -2.27, P = 0.045).



Supplementary Figure 6.10. 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. 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.

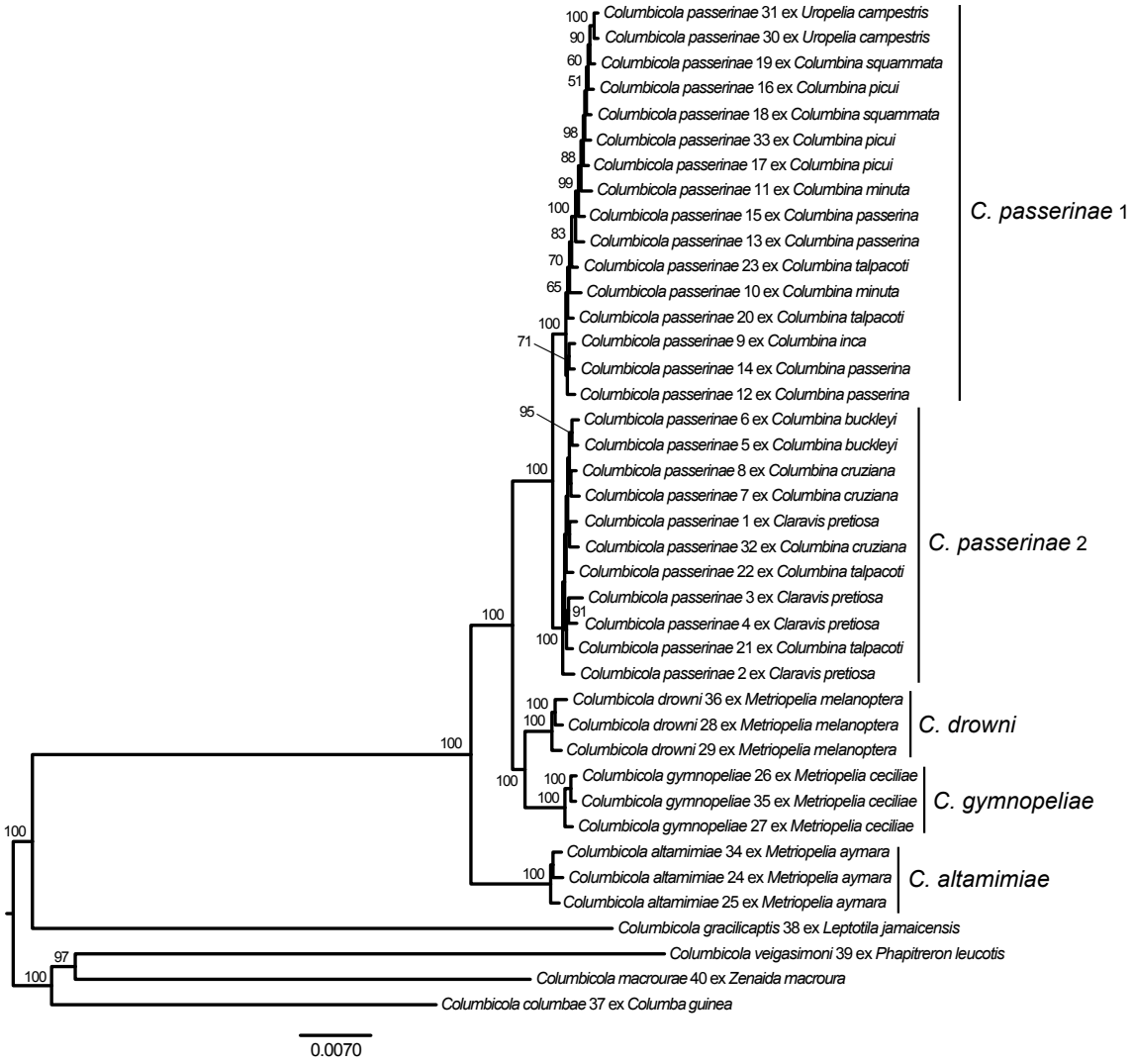


Supplementary Figure 6.11. STRUCTURE plot from 880 randomly sampled unlinked SNPs called for *Physconelloides* body lice from small New World ground-doves. Individual lice are grouped according to host species, and colored according to the likelihood of being in a particular cluster. Phylogenies to the left of the STRUCTURE plots are modified from the concatenated *Physconelloides* phylogeny, and are colored according to the clusters from the STRUCTURE plot. 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.

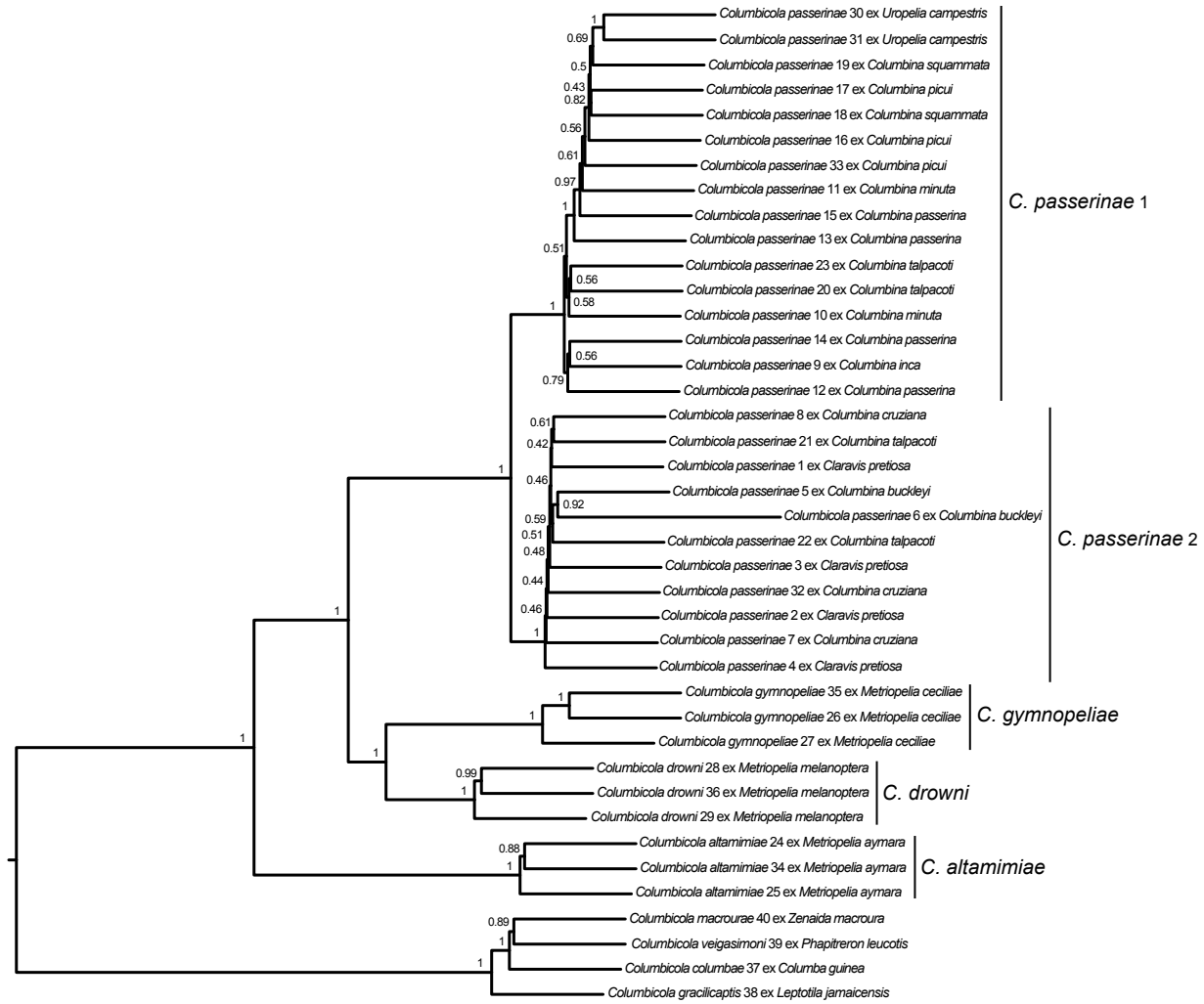


Supplementary Figure 6.12. DAPC density plot generated using SNPs from *Physconelloides eurysema* 3 (K = 2). Individual lice are indicated by vertical lines along the x-axis. Lice from *Claravis pretiosa* are red, and lice from all other host species are blue.

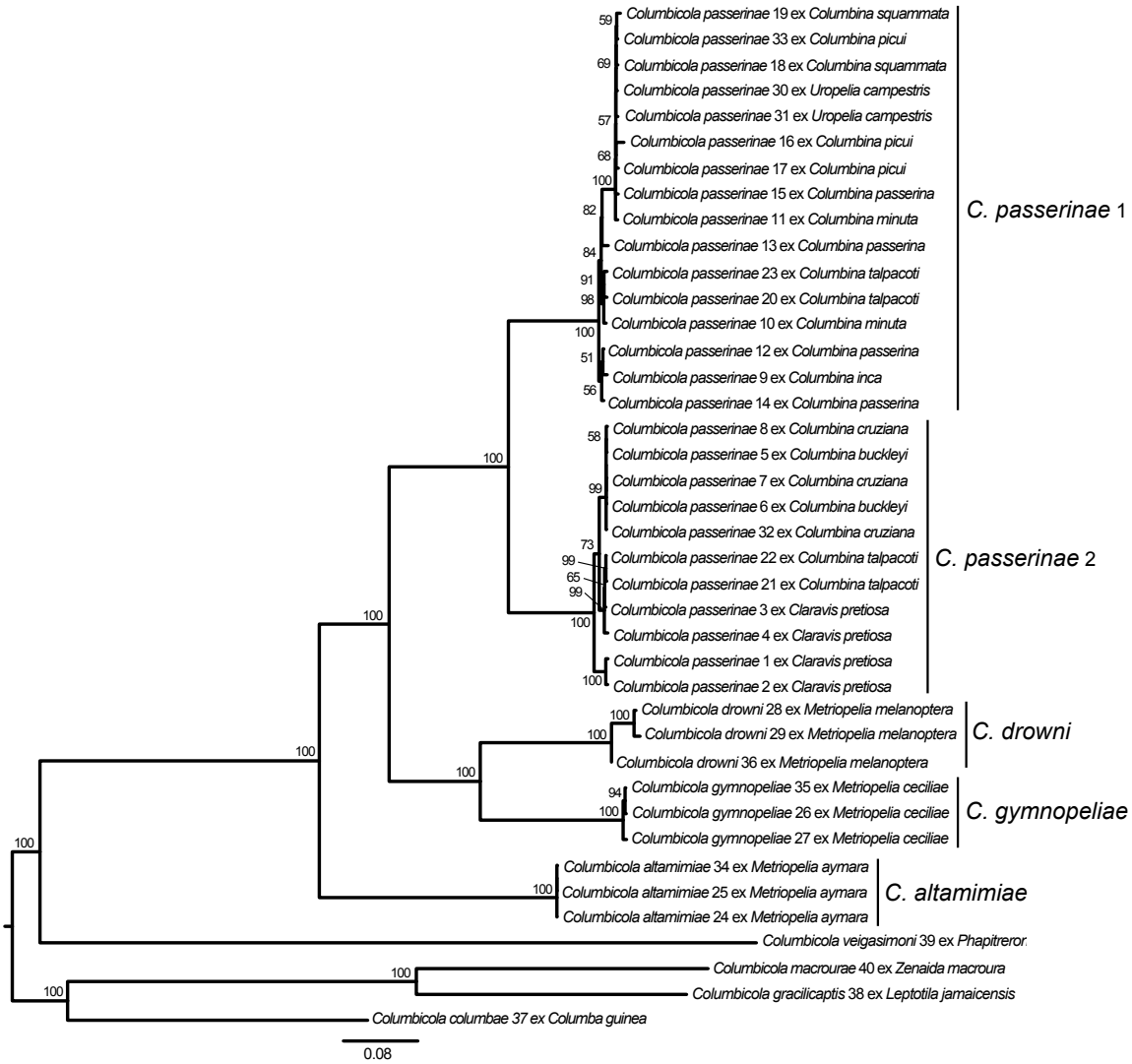
Supplementary Figures for Chapter 7



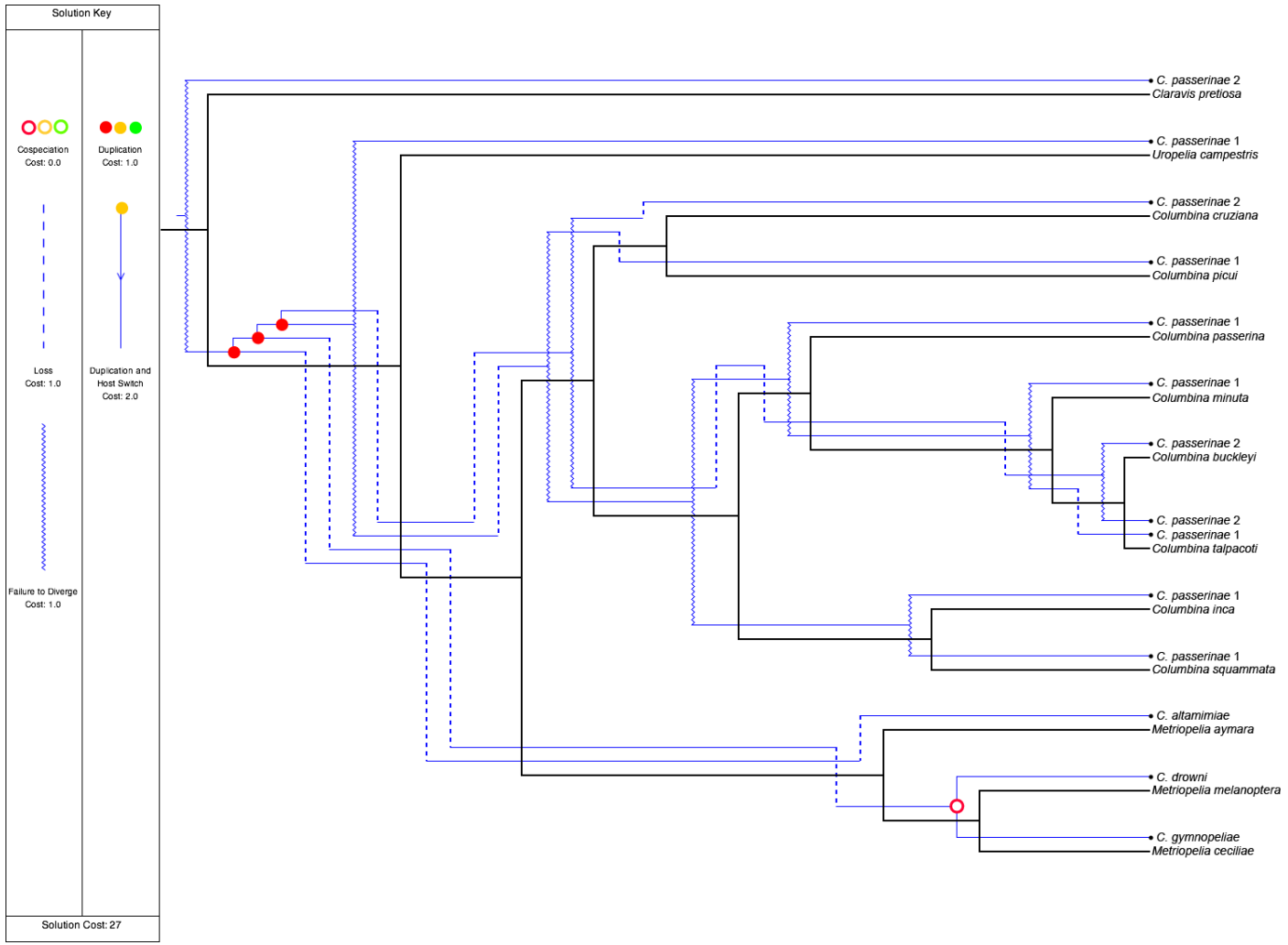
Supplementary Figure 7.1. Maximum likelihood phylogeny of New World ground-dove wing lice (*Columbicola*) estimated from a concatenated alignment of 1,039 nuclear genes. Bootstrap support values ≥50% are indicated at each node. Branch lengths represent nucleotide substitutions per site, as indicated by the scale bar. OTUs are indicated to the right of the tip labels. Numbers in each tip label refer to Supplementary Table 7.1.



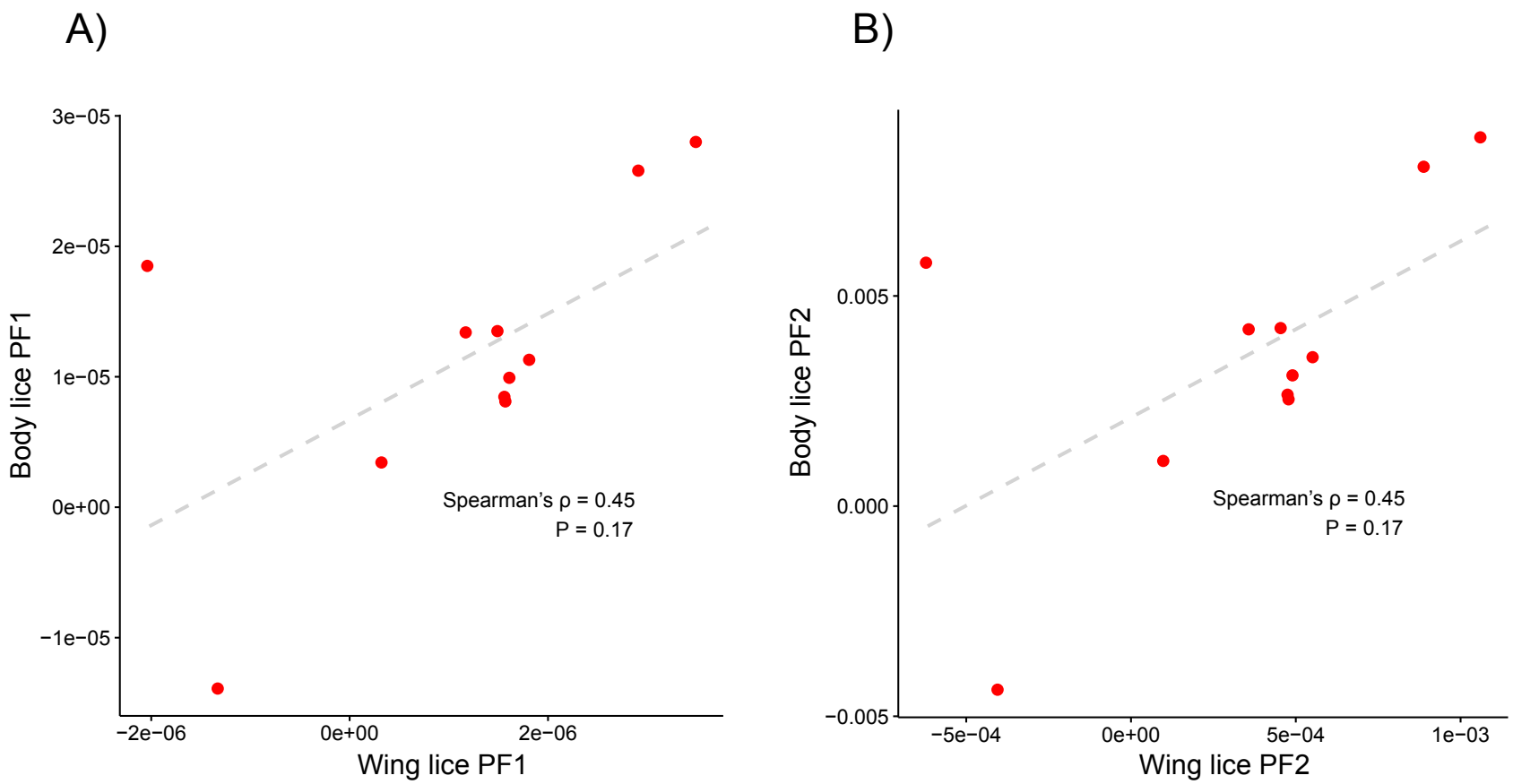
Supplementary Figure 7.2. Phylogeny of New World ground-dove lice (*Columbicola*) summarized from 1,039 gene trees using ASTRAL. Numbers at each node indicate local posterior probability. Internal branch lengths represent coalescent units, but tip branches are not meaningful. OTUs are indicated to the right of the tip labels. Numbers in each tip label refer to Supplementary Table 7.1.



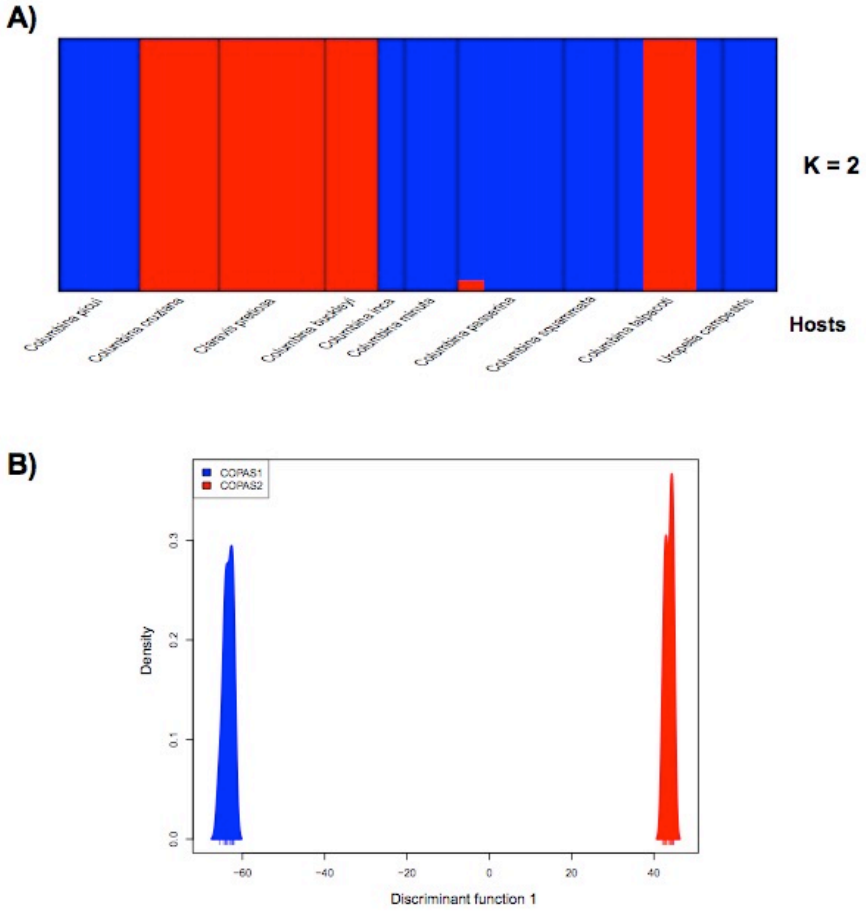
Supplementary Figure 7.3. Maximum likelihood phylogeny of New World ground-dove wing lice (*Columbicola*) estimated from a concatenated alignment of 7 mitochondrial genes. Bootstrap support values ≥50% are indicated at each node. Branch lengths represent nucleotide substitutions per site, as indicated by the scale bar. OTUs are indicated to the right of the tip labels. Numbers in each tip label refer to Supplementary Table 7.1.



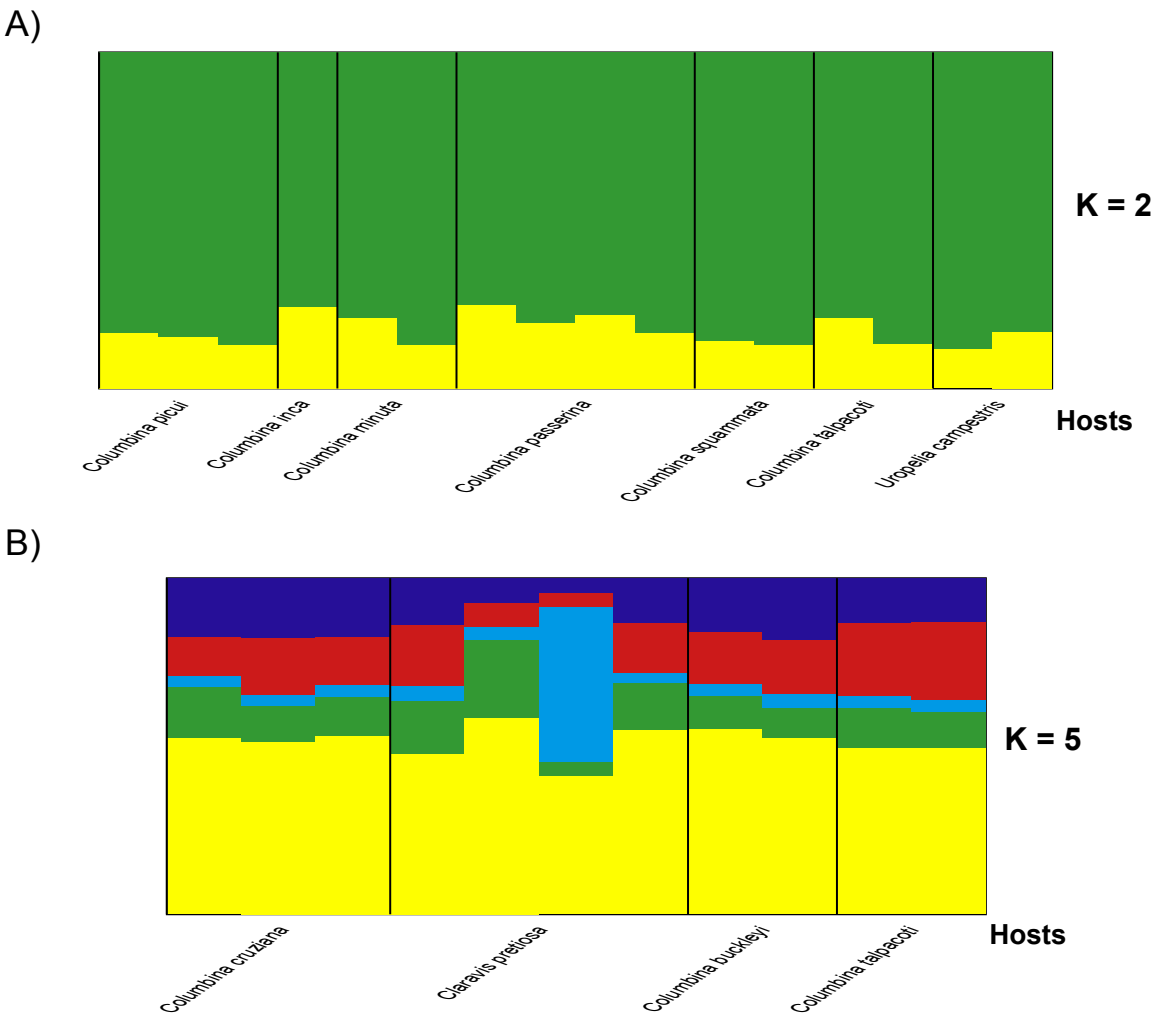
Supplementary Figure 7.4. Reconciliation of the New World ground-dove phylogeny with the phylogeny of their wing lice (*Columbicola*). Results are from an analysis with Jane4. The ground-dove phylogeny is represented with black lines, and the louse phylogeny with blue lines.



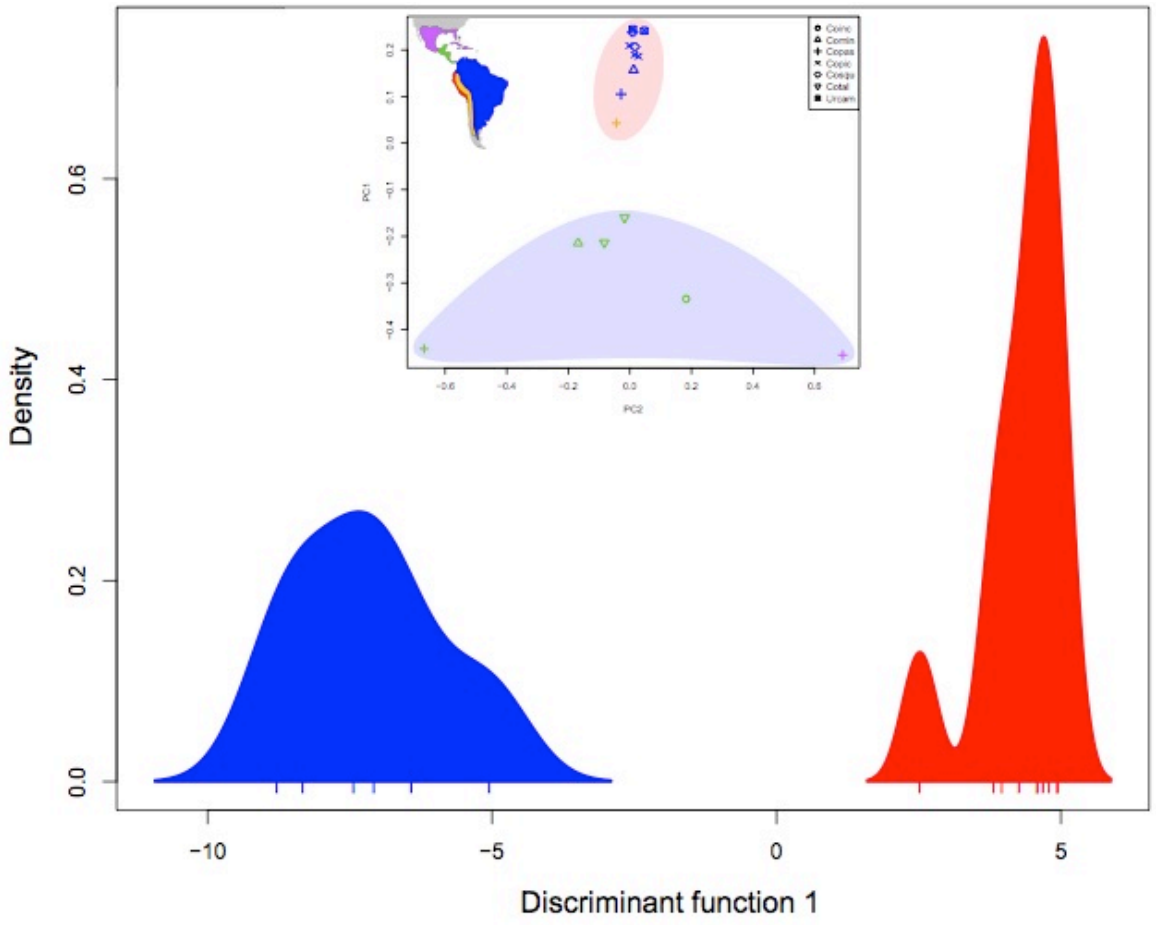
Supplementary Figure 7.5. Correlation of A) ParaFitLink1 (PF1) and B) ParaFitLink2 (PF2) values from a ParaFit analysis of New World ground-doves and their wing and body lice. Red dots indicate host species. For hosts with multiple wing or body louse associations, the points represent mean residual values. Regression lines are provided to indicate trends.



Supplementary Figure 7.6. A) STRUCTURE and B) DAPC plots for the ground-dove wing louse species *Columbicola passerinae*. Columns in the STRUCTURE plot and distributions in the DAPC plot are colored by OTU: blue = *C. passerinae* 1, red = *C. passerinae* 2. Columns in the STRUCTURE plot represent individual lice, with the host species indicated below the plot.



Supplementary Figure 7.7. STRUCTURE plot for the ground-dove wing louse taxa A) *Columbicola passerinae* 1 and B) *C. passerinae* 2, two potentially cryptic species. Columns in the STRUCTURE plot represent individual lice, with the host species indicated below the plots.



Supplementary Figure 7.8. DAPC plot for the ground-dove wing louse taxon *Columbicola passerinae* 1. Individual lice are indicated by red or blue lines along the x-axis and by individual points on the inset PCA plot. The two specific clusters identified in DAPC are indicated with corresponding shaded shapes on the PCA plot (blue and red). Points on the PCA plot are colored according to biogeographic regions and shaped according to host species.

APPENDIX B: SUPPLEMENTARY TABLES

Supplementary Tables for Chapter 2

Supplementary Table 2.1 (Supp_table2.1.xlsx). Sampling matrices for doves and their A) wing lice and B) body lice. Hosts are listed to the left and associated louse samples to the right. The number of known recorded wing and body louse species associated with each host taxon is also listed. GenBank accession numbers are listed for all loci used in this study. Accession numbers for novel sequences are listed in bold. Missing data are indicated by dashes (-). Collecting locality for each host sample is also listed.

Supplementary Tables for Chapter 3

Supplementary Table 3.1 (Supp_table3.1.xlsx). Sampling matrix of Australian phabine pigeons and doves and their wing and body lice. GenBank accession numbers are indicated for available sequence data.

Supplementary Table 3.2 (Supp_table3.2.xlsx). Uncorrected COI pairwise distances for phabine body lice.

Supplementary Table 3.3 (Supp_table3.3.xlsx). Uncorrected COI pairwise distances for phabine wing lice.

Supplementary Table 3.4 (Supp_table3.4.xlsx). ParaFit individual link test statistics and P-values for phabine pigeons and doves and their body lice with 50% majority-rule consensus trees.

Supplementary Table 3.5 (Supp_table3.5.xlsx). ParaFit individual link test statistics and P-values for phabine pigeons and doves and their wing lice (C. = Columbicola) with 50% majority-rule consensus.

Supplementary Tables for Chapter 4

Supplementary Table 4.1 (Supp_table4.1.xlsx). Uncorrected mitochondrial pairwise distances of small New World ground-doves and their outgroups taxa. Numbers at the end of each ingroup taxon name refer to the collection voucher numbers as indicated in Table 4.1. Outgroup taxa are listed at the bottom of the table (#s 27-36), and do not have voucher numbers listed (see Table 4.1 for voucher information).

Supplementary Tables for Chapter 5

Supplementary Table 5.1 (Supp_table5.1.xlsx). Uncorrected distances of ground-dove *Columbicola* based on CO1. Taxa are colored according to *Columbicola* species. Yellow = *C. drowni*, Green = *C. gymnopeliae,* Purple = *C. altamimiae*, Red = *C. passerinae* 2, Blue = *C. passerinae* 1, Brown = *C. columbae* (outgroup).

Supplementary Tables for Chapter 6

Supplementary Table 6.1 (Supp_table6.1.xlsx). 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.

Supplementary Tables for Chapter 7

Supplementary Table 7.1 (Supp_table7.1.xlsx). Specimen information, extraction and library preparation details, Illumina sequencing statistics, locus assembly, and raw sequence data deposition for *Columbicola* wing lice from New World ground-doves.

Supplementary Table 7.2 (Supp_table7.2.xlsx). Results from a ParaFit analysis of New World ground-doves and their wing lice. Each row is a host-parasite association, and includes ParaFitLink1 (PF1), ParaFitLink2 (PF2) results along with associated P-values generated from randomizations of the association matrix.

Supplementary Table 7.3 (Supp_table7.3.xlsx). mlRho estimates of θ for New World ground-dove wing and body lice. Average values of θ are listed along with the 95% confidence intervals. Numbers in the sample names refer to Supplementary Table 7.1 for wing lice, and Supplementary Table 6.1 for body lice. Several relevant characters of the lice and hosts are included in columns to the right.

Supplementary Table 7.4 (Supp_table7.4.xlsx). Raw heterozygosity (ratio of heterozygous sites to total sites) for New World ground-dove wing and body lice. Numbers in the sample names refer to Supplementary Table 7.1 for wing lice, and Supplementary Table 6.1 for body lice. Several relevant characters of the lice and hosts are included in columns to the right.