Title: Parasite dispersal influences introgression rate: the case of dove wing and body lice

Authors: Jorge Doña[1]*, Andrew D. Sweet[1, 2], and Kevin P. Johnson[1]*

Affiliations:

[1]Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, 1816 S. Oak St., Champaign, Illinois 61820, USA.

[2]Department of Entomology, Purdue University, 901 W. State St., West Lafayette, Indiana 47907, USA.

*Correspondence to: jorged@illinois.edu, kpjohnso@illinois.edu.

Running title: Dispersal influences introgression rate.

Author contributions: J.D., and K.P.J. conceived the study. J.D., A.D.S., and K.P.J. designed the study. A.D.S. collected the data. J.D. and A.D.S. analysed the data. K.P.J. obtained financial support for the project. J.D. wrote the manuscript and all authors contributed to editing the manuscript.

Acknowledgments: This study was supported by the US National Science Foundation (DEB-1239788 and DEB-1342604 to K.P.J). We declare that we have no conflict of interest.

Data availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Supplementary Materials: Supplementary data are available at Figshare (DOI: 10.6084/m9.figshare.9176204).

Abstract: Organisms vary in their dispersal abilities, and these differences can have important biological consequences, such as impacting the likelihood of hybridization events. However, the factors shaping the frequency of hybridization are still poorly understood, and therefore how dispersal ability affects the opportunities for hybridization is still unknown. Here, using the ecological replicate system of dove wing and body lice (Insecta: Phthiraptera), we show that species with higher dispersal abilities exhibited increased genomic signatures of introgression. Specifically, we found a higher proportion of introgressed genomic reads and more reticulated phylogenetic networks in wing lice, the louse group with higher dispersal abilities. Our results illustrate how differences in dispersal ability can drive differences in the extent of introgression through hybridization. The results from this study represent an important step for understanding the factors driving hybridization. We expect our approach will stimulate future studies on the ecological factors shaping hybridization to further understand this important process.

Introduction

Dispersal is the permanent movement of organisms away from their place of origin. It is a fundamental process in biology with significant implications at multiple scales of organization (Barton 1992; Clobert et al. 2001; Nathan 2001; Matthysen 2012), including the reproduction of individuals, the composition of populations and communities, and the geographical distribution of species (Clobert et al. 2001, 2012).

Organisms differ in their dispersal abilities, and these differences have an impact on their biology, such as on the distributional range of a species or gene flow between populations (Bohonak 1999). For example, organisms with lower dispersal abilities tend to have smaller distributional ranges and populations that are genetically more structured (Bohonak 1999; Dawson et al. 2014; DiBlasi et al. 2018).

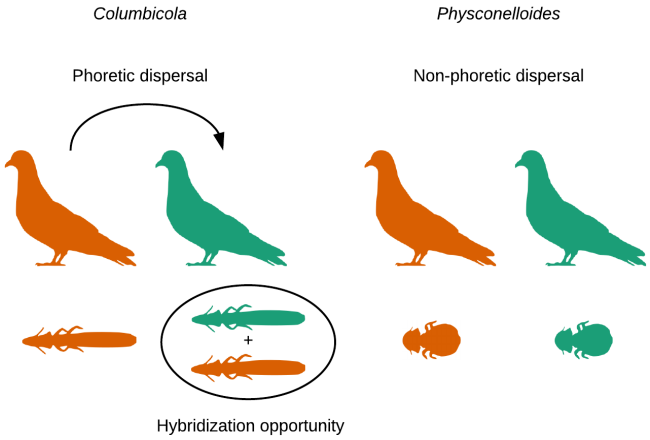
Dispersal ability might also affect the opportunities for hybridization between species because the rates at which individuals encounter different species are likely to be higher in organisms with higher dispersal capabilities. Indeed, recent evidence supports this prediction by demonstrating that range expansion is associated with the extent of introgression (Currat et al. 2008; Nussberger et al. 2018). Similarly, dispersal differences explain more than 30% of the variation in the width of hybrid zones across animals (McEntee et al. 2018). However, overall the factors influencing hybridization events are poorly known (Randler 2006; Arnold 2015; Taylor and Larson 2019), and, in particular, the influence of dispersal ability on the rate of hybridization remains understudied.

Comparisons of the effect of dispersal on hybridization should ideally hold constant most factors other than dispersal. The ecological replicate system of wing and body lice (Insecta: Phthiraptera) of pigeons and doves (Aves: Columbidae) has proven to be an ideal system for

comparing the impact of dispersal differences on other aspects of biology, such as population structure and codivergence (Clayton and Johnson 2003; Johnson and Clayton 2004; Clayton et al. 2015; DiBlasi et al. 2018; Sweet and Johnson 2018). Specifically, this is an excellent system in which to assess the effect of differences in dispersal capabilities on levels of introgression because both of these two lineages of feather lice: 1) co-occur across the diversity of pigeons and doves, 2) present highly comparable temporal patterns of diversification; specifically, cophylogenetic analyses and bird time-calibrated trees indicate that both lineages originated on the common ancestor of *Metropelia* doves (11.3-14.9 mya) and also share a cospeciation event which occurred within the *Metriopelia* genus (5.2-7.4 mya; Sweet and Johnson 2015, 2018), 3) have the same basic life history and diet (Clayton et al. 2015; Sweet and Johnson 2015, 2018), but 4) they significantly differ in their dispersal ability (Harbison et al. 2008, 2009; Bartlow et al. 2016). Both wing and body lice disperse vertically between parents and offspring in the nest. However, wing lice can also attach to and hitchhike on hippoboscid flies to disperse "phoretically" between host individuals or host species (Harbison et al. 2008, 2009; Bartlow et al. 2016). Indeed, this hitch-hiking dispersal mechanism profoundly influences their degree of population structure and cophylogenetic history (Clayton and Johnson 2003; Sweet et al. 2017b; DiBlasi et al. 2018; Sweet and Johnson 2018). In addition, wing lice have a higher rate of host-switching (Clayton and Johnson 2003; Clayton et al. 2015; Sweet et al. 2017b) (i.e., successful colonization of new host species) and of straggling (Whiteman et al. 2004) (i.e., dispersal to new host species without reproduction on that new host).

To compare differences in the extent of introgression between wing and body lice, we used whole-genome data from 71 louse individuals belonging to five taxa of wing lice (*Columbicola*) and seven taxa of body lice (*Physconelloides*) occurring across the same host

82 species and have highly comparable patterns of diversification (Sweet and Johnson 2015, 2018).
83 We predicted that wing lice, which have higher dispersal abilities and thus higher odds of
84 encountering individuals of a different louse species on the same host, should show more
85 extensive evidence of introgression (Fig. 1).



86

87 **Figure 1**. Diagram depicting the ecological replicate system and the hypothesis of this study. Wing lice (*Columbicola*) have
88 higher dispersal abilities than body lice (*Physconelloides*), and thus higher odds of encountering individuals of a different louse
89 sp.

90 ## Materials and Methods

91 ### Data

92 We studied whole genome data from 71 louse individuals belonging to five and seven taxa of
93 *Columbicola* and *Physconelloides*, respectively (Supplementary Table S2). Data were available
94 from previous studies (Boyd BM, et al. 2017; Sweet et al. 2017a; Sweet and Johnson 2018) and
95 represent all described New World ground-dove wing and body louse species, most host species
96 in this group, including samples across multiple biogeographic areas within species (Sweet and
97 Johnson 2018; Supplementary Table S2). Illumina genome sequence data pre-processing

included several steps (Sweet and Johnson 2018). First, we discarded duplicate read pairs using the fastqSplitDups script (https://github.com/McIntyre-Lab/mcscriptand https://github.com/McIntyre-Lab/mclib). We then eliminated the Illumina sequencing adapters with Fastx_clipper v0.014 from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Also, we removed the first 5 nt from the 5' ends of reads using Fastx_trimmer v0.014 and trimmed bases from the 3' ends of reads until reaching a base with a phred score ≥28 using Fastq_quality_trimmer v0.014. Finally, we removed any reads less than 75 nt and analyzed the cleaned libraries with Fastqc v0.11.5 to check for additional errors. We assembled nuclear loci in aTRAM following previous studies (Allen et al. 2015; Boyd BM, et al. 2017; Sweet and Johnson 2018). In particular, we mapped modest coverage (25-60X), 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 a previous study (Boyd BM, et al. 2017) (raw data: SRR3161922). This data set was assembled de novo (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). For body lice, we obtained nuclear data using the same pipeline and software parameters, except that we used 1,095 loci from *P. emersoni* as the reference for mapping. To generate the input ultrametric gene trees for Phylonet v3.6.8 (Than et al. 2008; Yu and Nakhleh 2015; Wen et al. 2018), we first aligned each nuclear locus in MAFFT (Katoh 2002)(--auto) and removed columns with only ambiguous sequences ("N"). Then, we estimated gene trees in RAxML v8.1.3 (Stamatakis 2006) with a GTR + Γ substitution model for

each gene alignment. Finally, we made trees ultrametric using the nnls method in the force.ultrametric function within the "phytools" R package (Revell 2011).

Quantifying introgression

We used two different approaches to quantify differences in the extent of introgression between the two louse genera. We employed methods suitable to both detect introgression between species and individuals from the same species (i.e., we did not employ methods aimed to detect differences at the population level, e.g., TreeMix; Pickrell and Pritchard 2012). First, we used sppIDer (Langdon et al. 2018) to quantify the genomic contributions of different louse species in an individual louse genome. We built our reference for each genus using all the nuclear loci from a single individual per species. For the reference, we selected those individuals for which we assembled the highest number of genes for each genus. Finally, we estimated the extent of introgression as the sum of the mean coverages of reads mapped from all the species excluding the focal louse species, divided by the mean coverage of the focal louse species. Second, we quantified introgression at the species level, while accounting for ILS (i.e., reticulations in this method can be attributed to hybridization events), using a maximum pseudo-likelihood framework with PhyloNet 3.6.1 (Than et al. 2008; Yu and Nakhleh 2015; Wen et al. 2018). We trimmed the unrooted gene trees to the same individuals used as reference taxa in sppIDer, and performed eleven independent analyses with a differing maximum number of reticulation nodes (i.e., from zero to ten). We conducted ten runs per analysis. We then selected the optimal network for each genus based on AIC values.

Analyses

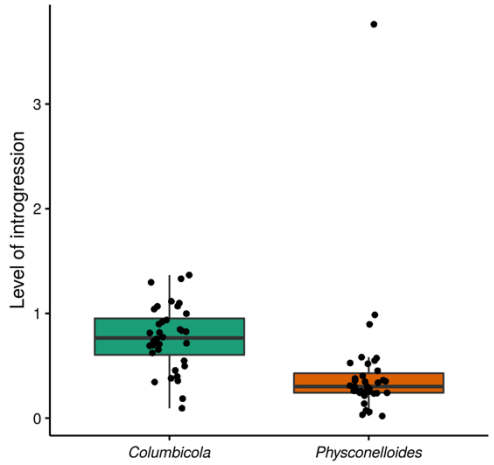
We compared the sppIDer results using generalized linear models (GLMs). We used a Gaussian distribution of errors and an identity link function. We performed one GLM for each simulation

143 iteration using the glm function of the "stats" R package (R Core Team 2013). The extent of
144 introgression for each louse genus was the dependent variable, the genus identity was the
145 independent variable, and we accounted for the introgression differences between louse species
146 including louse identity as a fixed factor. We confirmed assumptions underlying GLMs by
147 testing the normality of regression residuals for normality against a Q-Q plot. We also
148 considered the possibility that some of the reads mapping to other species were technical
149 contaminations, i.e., due to index-swapping (Carlsen et al. 2012; Esling et al. 2015; Schnell et al.
150 2015; Sinha et al. 2017). Previous studies have found that the misassignment of reads generally
151 ranges from 1 to 9% (Carlsen et al. 2012; Esling et al. 2015; Schnell et al. 2015; Sinha et al.
152 2017). Thus, to account for possible contaminants, we wrote a simulation in R that randomly
153 subtracted 9% (Carlsen et al. 2012; Esling et al. 2015; Schnell et al. 2015; Sinha et al. 2017)
154 from the mean coverage value of a particular sample (i.e., we subtracted a random proportion of
155 the mean coverage value for each sample until reaching 9 %). We ran 100 iterations of the
156 simulation and ran a GLM for each iteration (Table S1). Finally, we used the χ^2 test to compare
157 the number of species in pairwise comparisons of each genus with the number of reticulations
158 found in each optimal phylogenetic network.

159 **Results**

160 We used two different approaches to quantify the overall (i.e., ancient plus recent) differences in
161 introgression between louse genera. First, in individual louse genomes, we quantified the
162 genomic contributions from different closely related louse species of the same genus (Langdon et
163 al. 2018). Second, we quantified introgression at the species level, while accounting for
164 incomplete lineage sorting (ILS), by inferring phylogenetic networks using a maximum pseudo-
165 likelihood framework (Than et al. 2008; Yu and Nakhleh 2015; Wen et al. 2018).

166 Both approaches revealed highly concordant results: higher levels of introgression among
167 species of wing lice compared to body lice. In particular, using a read-mapping based method,
168 the genomic signature of introgression was significantly higher in wing louse species than in
169 body louse species (GLM with the mean values of the simulations; F = 21.0705, df = 69, *P* =
170 2.367 x 10^{-5}; Fig. 2, Supplementary Table S1, Figs. S1-S12). Contrary to this effect, one body
171 louse individual (included in the GLMs) exhibited the highest level of introgression (Fig. 2,
172 Figs. S1-S12). However, the other individual from the same taxon, inhabiting the same host
173 species and collected in the same geographic region, did not show these elevated levels of
174 introgression (Supplementary Table S2, Figs. S1-S12).



175

Figure 2. Boxplot showing the differences in levels of introgression between wing (green) and body (orange) lice. Level of introgression represents the sum of the mean coverage of reads mapped from all the species excluding the focal louse species, divided by the mean coverage of the focal louse species (see Methods). Black dots represent individual samples (horizontally jittered).

178

179

Secondly, in a phylogenetic network framework, the optimal networks of wing lice were more reticulated than those of body lice even though the number of taxa included in the networks was lower (seven reticulations in *Columbicola* vs. four in *Physconelloides*, Fig. 3). Accordingly, the number of reticulations given the number of potential combinations was significantly higher (χ^2= 3.8132; df=1; *P*= 0.03). Also, the specific lineages involved in the reticulations were generally congruent with signatures of introgression from the read-mapping based approach (Fig. S1-S12).

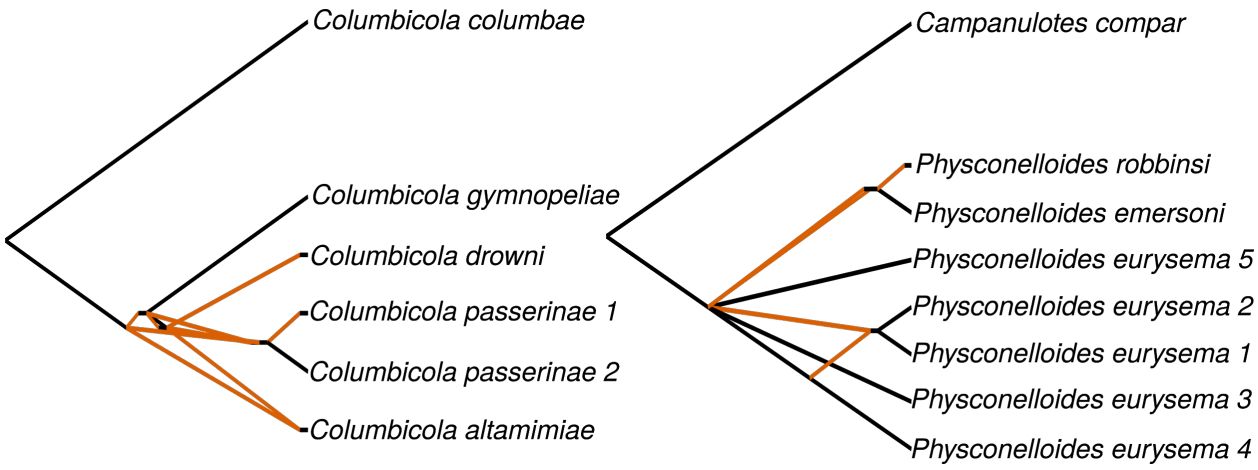


Figure 3. Optimal phylogenetic networks of feather lice genera. Orange branches depict reticulations. From left to right, *Columbicola* (seven reticulations) and *Physconelloides* (four reticulations) networks (See Methods).

Discussion

Taken together, evidence from wing and body louse genomes suggests that differences in dispersal ability drive differences in the extent of introgression in this system of ecological replicate parasites. This work is among the first studies of introgression in a host-symbiont system (Detwiler and Criscione 2010). Notably, recent studies have found that straggling and host-switching are relatively common processes in host-symbiont systems (De Vienne et al. 2013; Bourguignon et al. 2018; Nylin et al. 2018; Doña et al. 2019). Our study suggests that in a straggling/host-switching scenario, hybridization can provide further variation with important

eco-evolutionary consequences (Barton 2018). Indeed, we may have found a potential recent hybridization event (i.e., the *Physconelloides* individual showing the highest level of introgression), though this requires further study to rule out methodological issues (e.g., wet-lab contamination). Overall, the results from this study represent a significant step towards understanding the factors driving hybridization, because most previous studies focus on the presence/absence of hybridization and the evolutionary consequences of hybridization events (Arnold 2015; Folk et al. 2018; Taylor and Larson 2019). Further research is needed to understand the factors shaping the frequency of hybridization and to ascertain their influence across different scales (e.g., from ancient to recent hybridization events).

References:

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

Arnold, M. L. 2015. Divergence with genetic exchange. OUP Oxford, Oxford, UK.

Bartlow, A. W., S. M. Villa, M. W. Thompson, and S. E. Bush. 2016. Walk or ride? Phoretic behaviour of amblyceran and ischnoceran lice. Int J Parasitol 46:221–227.

Barton, N. H. 2018. The consequences of an introgression event. Mol Ecol 27:4973–4975.

Barton, N. H. 1992. The genetic consequences of dispersal. In: Animal Dispersal. Springer, Netherlands.

Bohonak, A. J. 1999. Dispersal Gene Flow, and Population Structure. Q Rev Biol 74:21–45.

219 Bourguignon, T., N. Lo, C. Dietrich, J. Šobotník, S. Sidek, Y. Roisin, A. Brune, and T. A. Evans.
220 2018. Rampant host switching shaped the termite gut microbiome. Curr Biol 28:649–654.
221 Boyd BM, Allen JM, Nguyen N, Sweet AD, Warnow T, Shapiro MD, and et al. 2017.
222 Phylogenomics using Target-restricted Assembly Resolves Intra-generic Relationships of
223 Parasitic Lice (Phthiraptera: Columbicola). Syst Biol 66:896–911.
224 Carlsen, T., A. B. Aas, D. Lindner, T. Vrålstad, T. Schumacher, and H. Kauserud. 2012. Dont
225 make a mista(g)ke: is tag switching an overlooked source of error in amplicon
226 pyrosequencing studies? Fungal Ecol 5:747–749.
227 Clayton, D. H., S. E. Bush, and K. P. Johnson. 2015. Coevolution of life on hosts: integrating
228 ecology and history. University of Chicago Press, Chicago, IL.
229 Clayton, D. H., and K. P. Johnson. 2003. Linking coevolutionary history to ecological process:
230 doves and lice. Evolution 57:2335–2341.
231 Clobert, J., M. Baguette, T. G. Benton, and J. M. Bullock (eds). 2012. Dispersal Ecology and
232 Evolution. Oxford University Press.
233 Clobert, J., E. Danchin, A. A. Dhondt, and N. J. D. 2001. Dispersal. Oxford Univ. Press.
234 Currat, M., M. Ruedi, R. J. Petit, and L. Excoffier. 2008. The hidden side of invasions: massive
235 introgression by local genes. Evolution 62:1908–1920.
236 Dawson, M. N., C. G. Hays, R. K. Grosberg, and P. T. Raimondi. 2014. Dispersal potential and
237 population genetic structure in the marine intertidal of the eastern North Pacific. Ecol
238 Monogr 84:435–456.
239 De Vienne, D. M., G. Refrégier, M. López-Villavicencio, A. Tellier, M. E. Hood, and T. Giraud.
240 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural
241 associations and relation to coevolution. New Phytol 198:347–385.

242 Detwiler, J. T., and C. D. Criscione. 2010. An infectious topic in reticulate evolution:
243 introgression and hybridization in animal parasites. Genes (Basel) 1:102–23.
244 DiBlasi, E., K. P. Johnson, S. A. Stringham, A. N. Hansen, A. B. Beach, and D. H. Clayton.
245 2018. Phoretic dispersal influences parasite population genetic structure. Mol Ecol
246 27:2770–2779.
247 Doña, J., D. Serrano, S. Mironov, A. Montesinos-Navarro, and R. Jovani. 2019. Unexpected
248 bird–feather mite associations revealed by DNA metabarcoding uncovers a dynamic
249 ecoevolutionary scenario. Mol Ecol 28:379–390.
250 Esling, P., F. Lejzerowicz, and J. Pawlowski. 2015. Accurate multiplexing and filtering for high-
251 throughput amplicon-sequencing. Nucleic Acids Res 43:2513–24.
252 Folk, R. A., P. S. Soltis, D. E. Soltis, and R. Guralnick. 2018. New prospects in the detection and
253 comparative analysis of hybridization in the tree of life. Am J Bot 105:364–375.
254 Harbison, C. W., S. E. Bush, J. R. Malenke, and D. H. Clayton. 2008. Comparative transmission
255 dynamics of competing parasite species. Ecology 89:3186–3194.
256 Harbison, C. W., M. V. Jacobsen, and D. H. Clayton. 2009. A hitchhiker's guide to parasite
257 transmission: The phoretic behaviour of feather lice. Int J Parasitol 39:569–575.
258 Johnson, K. P., and D. H. Clayton. 2004. Untangling coevolutionary history. Syst Biol 53:92–4.
259 Katoh, K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast
260 Fourier transform. Nucleic Acids Res 30:3059–3066.
261 Kirkness, E. F., B. J. Haas, W. Sun, H. R. Braig, M. A. Perotti, and J. M. Clark. 2010. Genome
262 sequences of the human body louse and its primary endosymbiont provide insights into
263 the permanent parasitic lifestyle. Proc Natl Acad Sci U S A 107:12168–73.

264 Langdon, Q. K., D. Peris, B. Kyle, and C. T. Hittinger. 2018. sppIDer: A Species Identification
265 Tool to Investigate Hybrid Genomes with High-Throughput Sequencing. Mol Biol Evol
266 35:2835–2849.
267 Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat
268 Methods 9:357–9.
269 Matthysen, E. 2012. Multicausality of dispersal: a review. In: Dispersal Ecology and Evolution.
270 Oxford University Press.
271 McEntee, J. P., J. G. Burleigh, and S. Singhal. 2018. Dispersal predicts hybrid zone widths
272 across animal diversity: Implications for species borders under incomplete reproductive
273 isolation. bioRxiv, doi: 10.1101/472506.
274 Nathan, R. 2001. The challenges of studying dispersal. Trends Ecol Evol 16:481–483.
275 Nussberger, B., M. Currat, C. S. Quilodran, N. Ponta, and L. F. Keller. 2018. Range expansion as
276 an explanation for introgression in European wildcats. Biol Conserv 218:49–56.
277 Nylin, S., S. Agosta, S. Bensch, W. A. Boeger, M. P. Braga, D. R. Brooks, M. L. Forister, P. A.
278 Hambäck, E. P. Hoberg, and T. Nyman. 2018. Embracing colonizations: a new paradigm
279 for species association dynamics. Trends Ecol Evol 33:4–14.
280 Pickrell, J. K., and J. K. Pritchard. 2012. Inference of Population Splits and Mixtures from
281 Genome-Wide Allele Frequency Data. PLoS Genetics 8.
282 R Core Team. 2013. R: A language and environment for statistical computing.
283 Randler, C. 2006. Behavioural and ecological correlates of natural hybridization in birds. Ibis
284 148:459–467.
285 Revell, L. 2011. phytools: an R package for phylogenetic comparative biology (and other
286 things). Methods Ecol Evol 3:217–223.

287 Schnell, I. B., K. Bohmann, and M. T. Gilbert. 2015. Tag jumps illuminated–reducing sequence-
288 to-sample misidentifications in metabarcoding studies. Mol Ecol Resour 15:1289–303.
289 Sinha, R., G. Stanley, G. S. Gulati, C. Ezran, K. J. Travaglini, and E. Wei. 2017. Index
290 Switching Causes Spreading-Of-Signal Among. bioRxiv, doi: 10.1101/125724.
291 Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
292 thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
293 Sweet, A. D., B. M. Boyd, J. M. Allen, S. M. Villa, M. P. Valim, and J. L. Rivera-Parra. 2017a.
294 Integrating phylogenomic and population genomic patterns in avian lice provides a more
295 complete picture of parasite evolution. Evolution 72:95–112.
296 Sweet, A. D., R. T. Chesser, and K. P. Johnson. 2017b. Comparative cophylogenetics of
297 Australian phabine pigeons and doves (Aves: Columbidae) and their feather lice (Insecta:
298 Phthiraptera). Int J Parasitol 47:347–356.
299 Sweet, A. D., and K. P. Johnson. 2015. Patterns of diversification in small New World ground
300 doves are consistent with major geologic events. Auk 132:300–312.
301 Sweet, A. D., and K. P. Johnson. 2018. The role of parasite dispersal in shaping a host–parasite
302 system at multiple evolutionary scales. Mol Ecol 27:5104–5119.
303 Taylor, S. A., and E. L. Larson. 2019. Insights from genomes into the evolutionary importance
304 and prevalence of hybridization in nature. Nat Ecol Evol 3:170.
305 Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: a software package for analyzing and
306 reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9:322.
307 Wen, D., Y. Yu, J. Zhu, and L. Nakhleh. 2018. Inferring Phylogenetic Networks Using
308 PhyloNet. Syst Biol 67:735–740.

309 Whiteman, N. K., D. Santiago-Alarcon, K. P. Johnson, and P. G. Parker. 2004. Differences in
310 straggling rates between two genera of dove lice (Insecta: Phthiraptera) reinforce
311 population genetic and cophylogenetic patterns. Int J Parasitol 34:1113–1119.
312 Yu, Y., and L. Nakhleh. 2015. A maximum pseudo-likelihood approach for phylogenetic
313 networks. BMC Genomics 16:Suppl 10: S10.
314
315