Parasite dispersal influences introgression rate

Jorge Doña[1, 2], Andrew D. Sweet[2, 3], and Kevin P. Johnson[2]

[1]AllGenetics & Biology SL, Edificio CICA, Campus de Elviña s/n, 15008 A Coruña, Spain.

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

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

Dispersal is a central process in biology with implications at multiple scales of organization[1,2,3,4]. Organisms vary in their dispersal abilities, and these differences can have important biological consequences, such as impacting the likelihood of hybridization events[5]. 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)[6], 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.

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[1,2,3,4], including the reproduction of individuals, the composition of populations and communities, and the geographical distribution of species[1,7].

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[5]. For example, organisms with lower dispersal abilities tend to have smaller distributional ranges and populations that are genetically more structured[5,8,9].

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[10,11]. Similarly, dispersal differences explain more than 30% of the variation in the width of hybrid zones across animals[12]. However, overall the factors influencing hybridization events are poorly known[13], 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[6,8,14,15,16]. Both of these two lineages of feather lice occur across the diversity of pigeons and doves and have the same basic life history and diet, but they significantly differ in their dispersal ability[17,18,19]. 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[17,18,19]. Indeed, this additional dispersal mechanism profoundly influences their degree of population structure and cophylogenetic history[8,14,16,20]. In addition, wing lice have a higher rate of host-switching[6,14,15] (i.e., successful colonization of new host species) and of straggling[21] (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 species. We predicted that wing lice, which have higher dispersal abilities and thus higher odds of encountering individuals of a different louse species on the same host, should show more extensive evidence of introgression (Fig. 1).

We used two different approaches to quantify the differences in introgression between louse genera. First, in individual louse genomes, we quantified the genomic contributions from different closely related louse species of the same genus[22]. Second, we quantified introgression at the species level, accounting for incomplete lineage sorting (ILS) by inferring phylogenetic networks using a maximum pseudo-likelihood framework[23,24,25].

Both approaches revealed highly concordant results; higher levels of introgression among species of wing lice compared to body lice. In particular, using a read-mapping based method, the genomic signature of introgression was significantly higher in wing louse species than in body louse species (GLM with the mean values of the simulations; $F = 21.0705$, $df = 69$, $P = 2.367 \times 10^{-5}$; Fig. 2, Supplementary Table S1, Figs. S1-S12).

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).

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[26]. Notably, recent studies have found that straggling and host switching are relatively common processes in host-symbiont systems[27,28,29,30]. Our study suggests that in a straggling/host-switching scenario, hybridization can provide further variation with important eco-evolutionary consequences[31]. 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[13,32]. Further research is needed to understand the factors shaping the frequency of hybridization and how these factors influence eco-evolutionary dynamics.

Methods

Data

We studied whole genome data from 71 louse individuals belonging to five and seven taxa of *Columbicola* and *Physconelloides*, respectively (Supplementary Table S2). Data were available from previous studies[16,33,34] and represent all described New World ground-dove wing and body louse species, most host species in this group, and sampling across multiple biogeographic areas within species[16] (Supplementary Table S2). Illumina genome sequence data pre-processing included several steps[16]. 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[16,33,34,35]. 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[33] (raw data: SRR3161922). This data set was assembled de novo[35] using orthologous protein-coding genes from the human body louse genome (*Pediculus humanus humanus*[36]) 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[37]. 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[23,24,25], we first aligned each nuclear locus in MAFFT[38](--auto) and removed columns with only ambiguous sequences ("N"). Then, we estimated gene trees in RAxML v8.1.3[39] 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[40].

Quantifying introgression

We used two different approaches to quantify differences in the extent of introgression between the two louse genera. First, we used sppIDer[22] 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 loci. 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, using a maximum pseudo-likelihood framework with PhyloNet 3.6.1[23,24,25]. We trimmed the unrooted gene trees to the same individuals used as reference taxa in sppIDer, and performed ten 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

iteration using the *glm* function of the "stats" R package[41]. The extent of introgression for each louse genus was the dependent variable, the genus identity was the independent variable, and we accounted for the introgression differences between louse species including louse identity as a fixed factor. We confirmed assumptions underlying GLMs by testing the normality of regression residuals for normality against a Q-Q plot. We also considered the possibility that some of the reads mapping to other species were technical contaminations, i.e., due to index-swapping[42,43,44,45]. To account for possible contaminants, we wrote a simulation in R that randomly subtracted 9% from the mean coverage value of a particular sample (i.e., we subtracted a random proportion of the mean coverage value for each species until reaching 9 %). We ran 100 iterations of the simulation and ran a GLM for each iteration (Table S1). Finally, we used the χ^2 test to compare the number of species in pairwise comparisons of each genus with the number of reticulations found in each optimal phylogenetic network.

References

1 Clobert J, Danchin E, Dhondt AA, Nichols J D. *Dispersal*. Oxford Univ. Press, 2001.

2 Nathan R. The challenges of studying dispersal. *Trends in Ecology & Evolution* 2001; **16**: 481–483.

3 Matthysen E. Multicausality of dispersal: a review. In: *Dispersal Ecology and Evolution*. Oxford University Press, 2012, pp 3–18.

4 Barton NH. The genetic consequences of dispersal. In: *Animal Dispersal*. Springer Netherlands, 1992, pp 37–59.

161 5 Bohonak AJ. Dispersal Gene Flow, and Population Structure. *The Quarterly Review of Biology*
162 1999; **74**: 21–45.

163 6 Clayton DH, Bush SE, Johnson KP. *Coevolution of Life on Hosts*. University of Chicago Press,
164 2016 doi:10.7208/chicago/9780226302300.001.0001.

165 7 Clobert J, Baguette M, Benton TG, Bullock JM (eds.). *Dispersal Ecology and Evolution*.
166 Oxford University Press, 2012 doi:10.1093/acprof:oso/9780199608898.001.0001.

167 8 DiBlasi E, Johnson KP, Stringham SA, Hansen AN, Beach AB, Clayton DH *et al.* Phoretic
168 dispersal influences parasite population genetic structure. *Molecular Ecology* 2018; **27**:
169 2770–2779.

170 9 Dawson MN, Hays CG, Grosberg RK, Raimondi PT. Dispersal potential and population
171 genetic structure in the marine intertidal of the eastern North Pacific. *Ecological*
172 *Monographs* 2014; **84**: 435–456.

173 10 Nussberger B, Currat M, Quilodran CS, Ponta N, Keller LF. Range expansion as an
174 explanation for introgression in European wildcats. *Biological Conservation* 2018; **218**:
175 49–56.

176 11 Currat M, Ruedi M, Petit RJ, Excoffier L. The hidden side of invasions: massive introgression
177 by local genes. *Evolution* 2008; **62**: 1908–20.

178 12 McEntee JP, Burleigh JG, Singhal S. Dispersal predicts hybrid zone widths across animal
179 diversity: Implications for species borders under incomplete reproductive isolation.
180 *BioRxiv* 2018. doi:10.1101/472506.

13 Arnold ML. *Divergence with Genetic Exchange*. Oxford University Press, 2015 doi:10.1093/acprof:oso/9780198726029.001.0001.

14 Clayton DH, Johnson KP. Linking coevolutionary history to ecological process: doves and lice. *Evolution* 2003; **57**: 2335–41.

15 Johnson KP, Clayton DH. Untangling coevolutionary history. *Syst Biol* 2004; **53**: 92–4.

16 Sweet AD, Johnson KP. The role of parasite dispersal in shaping a host-parasite system at multiple evolutionary scales. *Molecular Ecology* 2018. doi:10.1111/mec.14937.

17 Harbison CW, Bush SE, Malenke JR, Clayton DH. Comparative transmission dynamics of competing parasite species. *Ecology* 2008; **89**: 3186–3194.

18 Harbison CW, Jacobsen MV, Clayton DH. A hitchhiker's guide to parasite transmission: The phoretic behaviour of feather lice. *International Journal for Parasitology* 2009; **39**: 569–575.

19 Bartlow AW, Villa SM, Thompson MW, Bush SE. Walk or ride? Phoretic behaviour of amblyceran and ischnoceran lice. *International Journal for Parasitology* 2016; **46**: 221–227.

20 Sweet AD, Chesser RT, Johnson KP. Comparative cophylogenetics of Australian phabine pigeons and doves (Aves: Columbidae) and their feather lice (Insecta: Phthiraptera). *International Journal for Parasitology* 2017; **47**: 347–356.

21 Whiteman NK, Santiago-Alarcon D, Johnson KP, Parker PG. Differences in straggling rates between two genera of dove lice (Insecta: Phthiraptera) reinforce population genetic and cophylogenetic patterns. *International Journal for Parasitology* 2004; **34**: 1113–1119.

202 22 Langdon QK, Peris D, Kyle B, Hittinger CT. sppIDer: A Species Identification Tool to
203 Investigate Hybrid Genomes with High-Throughput Sequencing. *Mol Biol Evol* 2018; **35**:
204 2835–2849.

205 23 Than C, Ruths D, Nakhleh L. PhyloNet: a software package for analyzing and reconstructing
206 reticulate evolutionary relationships. *BMC Bioinformatics* 2008; **9**: 322.

207 24 Wen D, Yu Y, Zhu J, Nakhleh L. Inferring Phylogenetic Networks Using PhyloNet. *Syst Biol*
208 2018; **67**: 735–740.

209 25 Yu Y, Nakhleh L. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC*
210 *Genomics* 2015; **16 Suppl 10**: S10.

211 26 Detwiler JT, Criscione CD. An infectious topic in reticulate evolution: introgression and
212 hybridization in animal parasites. *Genes (Basel)* 2010; **1**: 102–23.

213 27 Doña J, Serrano D, Mironov S, Montesinos-Navarro A, Jovani R. Unexpected bird-feather
214 mite associations revealed by DNA metabarcoding uncovers a dynamic ecoevolutionary
215 scenario. *Molecular Ecology* 2018. doi:10.1111/mec.14968.

216 28 Bourguignon T, Lo N, Dietrich C, Šobotník J, Sidek S, Roisin Y *et al.* Rampant Host
217 Switching Shaped the Termite Gut Microbiome. *Current Biology* 2018; **28**: 649–654.e2.

218 29 Vienne DM de, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T.
219 Cospeciation vs host-shift speciation: methods for testing evidence from natural
220 associations and relation to coevolution. *New Phytologist* 2013; **198**: 347–385.

30 Nylin S, Agosta S, Bensch S, Boeger WA, Braga MP, Brooks DR *et al.* Embracing Colonizations: A New Paradigm for Species Association Dynamics. *Trends in Ecology & Evolution* 2018; **33**: 4–14.

31 Barton NH. The consequences of an introgression event. *Molecular Ecology* 2018; **27**: 4973–4975.

32 Folk RA, Soltis PS, Soltis DE, Guralnick R. New prospects in the detection and comparative analysis of hybridization in the tree of life. *Am J Bot* 2018; **105**: 364–375.

33 Boyd BM, Allen JM, Nguyen N, Sweet AD, Warnow T, Shapiro MD *et al.* Phylogenomics using Target-restricted Assembly Resolves Intra-generic Relationships of Parasitic Lice (Phthiraptera: Columbicola). *Systematic Biology* 2017; syx027.

34 Sweet AD, Boyd BM, Allen JM, Villa SM, Valim MP, Rivera-Parra JL *et al.* Integrating phylogenomic and population genomic patterns in avian lice provides a more complete picture of parasite evolution. *Evolution* 2017; **72**: 95–112.

35 Allen JM, Huang DI, Cronk QC, Johnson KP. aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 2015; **16**. doi:10.1186/s12859-015-0515-2.

36 Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM *et al.* Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A* 2010; **107**: 12168–73.

37 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2.. *Nat Methods* 2012; **9**: 357–9.

242 38 Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast
243 Fourier transform. *Nucleic Acids Research* 2002; **30**: 3059–3066.

244 39 Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
245 thousands of taxa and mixed models. *Bioinformatics* 2006; **22**: 2688–2690.

246 40 Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things).
247 *Methods in Ecology and Evolution* 2011; **3**: 217–223.

248 41 Team RC, others. R: A language and environment for statistical computing. 2013.

249 42 Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H. Dont make a
250 mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing
251 studies?. *Fungal Ecology* 2012; **5**: 747–749.

252 43 Schnell IB, Bohmann K, Gilbert MT. Tag jumps illuminated–reducing sequence-to-sample
253 misidentifications in metabarcoding studies. *Mol Ecol Resour* 2015; **15**: 1289–303.

254 44 Esling P, Lejzerowicz F, Pawlowski J. Accurate multiplexing and filtering for high-
255 throughput amplicon-sequencing. *Nucleic Acids Res* 2015; **43**: 2513–24.

256 45 Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E *et al.* Index Switching Causes
257 Spreading-Of-Signal Among Multiplexed Samples In Illumina HiSeq 4000 DNA
258 Sequencing. *BioRxiv* 2017. doi:10.1101/125724.

259

260 **Additional information**

261 Table S1, S2 and, Figures S1-S12 are embedded into the supplementary_material.html file.

Acknowledgements

This study was supported by NSF DEB-1239788 and DEB-1342604 to KPJ.

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.

Data and materials availability

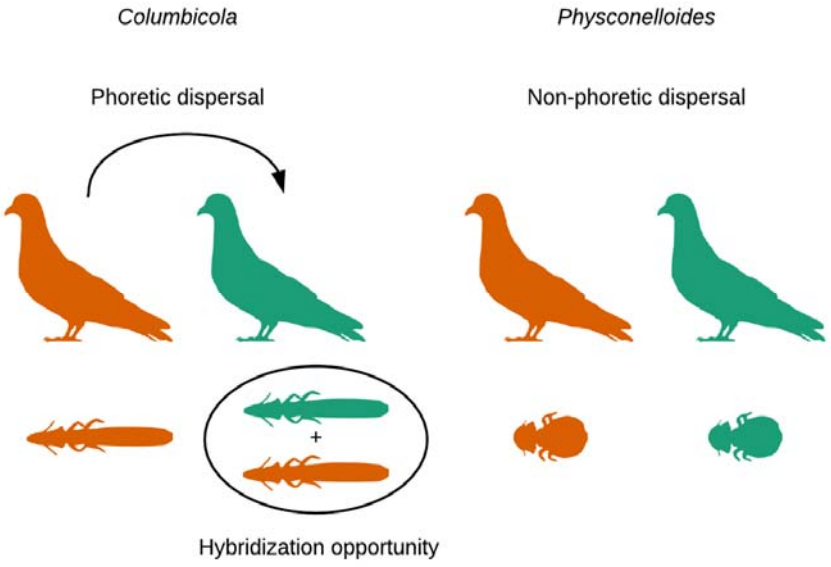
All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Competing interests

The authors declare that they have no competing interests.

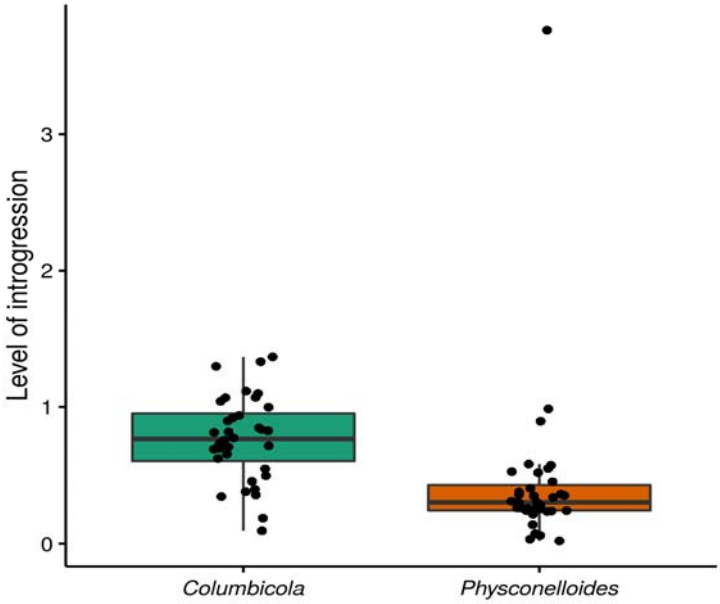
Correspondence and requests for materials should be addressed to J.D (jorged@illinois.edu) or K.P.J. (kpjohnso@illinois.edu).

Figure 1. Diagram depicting the ecological replicate system and the hypothesis of this study. Wing lice (*Columbicola*) have higher dispersal abilities than body lice (*Physconelloides*), and thus higher odds of encountering individuals of a different louse species on the same host. Thus, wing lice are predicted to show higher levels of introgression compared to body lice.



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

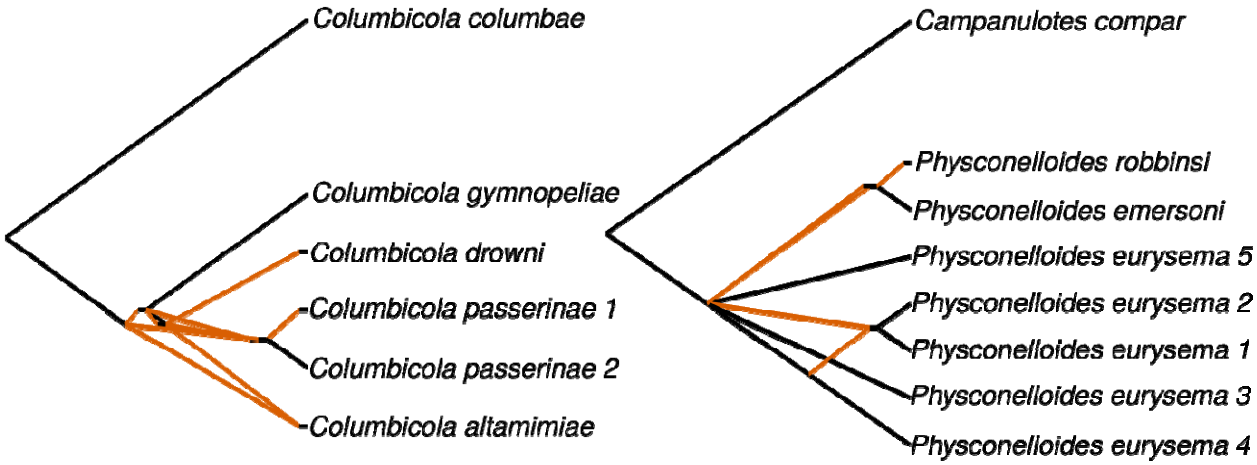
294



295

296

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



300