

# Phylogenomic Data Reveal Widespread Introgression Across the Range of an Alpine and Arctic Specialist

Erik R. Funk<sup>1\*</sup>, Garth M. Spellman<sup>2</sup>, Kevin Winker<sup>3</sup>, Jack J. Withrow<sup>3</sup>, Kristen C. Ruegg<sup>4</sup>, Erika Zavaleta<sup>5</sup>, Scott A. Taylor<sup>1</sup>

<sup>1</sup>*Department of Ecology and Evolutionary Biology, University of Colorado, Boulder CO 80301, USA*

<sup>2</sup>*Department of Zoology, Denver Museum of Nature and Science, Denver, CO 80205, USA*

<sup>3</sup>*University of Alaska Museum, University of Alaska, Fairbanks, AK 99709, USA*

<sup>4</sup>*Department of Biology, Colorado State University, Fort Collins, CO 80523, USA*

<sup>5</sup>*Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA*

\*Corresponding author: [erik.funk@colorado.edu](mailto:erik.funk@colorado.edu)

## SUPPLEMENTAL TEXT

### *SM 1.1: Phylogenetic Methods*

We constructed phylogenetic trees using both quartets, and maximum likelihood approaches for all datasets. Quartet trees were constructed using the program SVDquartets, implemented the PAUP\* v4.0a (Swofford 2003), and maximum likelihood trees were constructed using RAxML v.8.2.12 (Stamatakis 2014) and run in the CIPRES Science Gateway V. 3.3 (Miller, Pfeiffer, and Schwartz 2010).

SVDquartets .— Input nexus files for SVDquartets analyses were generated from VCF tables for all datasets using the *vcf2phylip.py* script with the “-n” flag

(<https://github.com/edgardomortiz/vcf2phylip/blob/master/vcf2phylip.py>). Analyses were run by examining 500,000 quartets, and 200 bootstrap replicates. Resulting trees were used to generate a 50% majority rule consensus tree, with all compatible groupings included.

**RAxML .**— Input phylip files for RAxML were generated from VCF tables using custom python scripts ([https://github.com/erikrfunk/genomics\\_tools](https://github.com/erikrfunk/genomics_tools)). Consensus sequences were constructed using `generate-consensus.py` with the flags “--format” set to “phylip” and “--variants” set to “True”. This generated a phylip file containing only variant sites for every individual. In the case of heterozygous genotypes, the alternative allele was used.

RAxML was run using a Felsenstein ascertainment bias correction to account for the absence of non-variant sites, under the GTRCAT model. We allowed RAxML to halt bootstrapping automatically using the autoMRE criterion. Bootstrap support values were drawn on the best ML tree on a local computer using the *raxmlHPC-PTHREADS-SSE3* command using the “-z” flag.

### *SM 1.2: Additional Phylogenetic Results and Taxonomic Implications*

Across both quartet and maximum likelihood phylogenetic reconstruction methods, and all datasets, individuals from the Pribilof islands (*L. t. umbrina*) and from the Aleutian islands (*L. t. griseonucha*) were recovered as fully supported, reciprocally monophyletic groups (Fig 1d) that together, were sister to all remaining North American individuals. Along with conStruct (Fig. 1c) and PCA (Fig.3), these results strongly support the individuals from the Alaska islands as a distinct lineage. Additionally, given the recovery of individuals from each of the two island systems (Pribilof and Aleutian) as separate clades, and as clusters with slightly different admixture proportions in conStruct models with high values of  $K$  ( $K > 4$ ), these taxa may themselves form distinct lineages, but should be explored further with more complete, island-wide sampling.

The remaining gray-crowned rosy-finches did not form a clade; however, with the exception of one individual from California, the three remaining subspecies were recovered as two separate clades in all SVDquartets analyses. The first clade contained all individuals of the two widespread gray-crowned rosy-finch subspecies (*L. t. tephrocotis*, and *L. t. littoralis*). Within this clade, in the 0p0.05maf SVDquartets phylogeny, eight of the nine gray-cheeked subspecies (*L. t. littoralis*) formed a clade, while all brown-cheeked subspecies (*L. t. tephrocotis*) and the one remaining gray-cheeked individual formed a second clade. While these results suggest some divergence between these two widespread subspecies (possibly resulting from assortative mating by cheek color or drift), the relationships of individuals from these two taxa were unsupported in all RAxML analyses. As a result, it is unclear how supported the divergence of these two subspecies are. Due to the sister relationship of these taxa, there is no valid topology to test for introgression between these taxa using ABBA-BABA tests, so it is also uncertain the extent to which these subspecies are exchanging genetic information. Broadly, our results support the two subspecies together forming a distinct lineage, but no further divergence among them.

The second remaining gray-crowned rosy-finch clade consists of four of the five individuals from the California population (*L. t. dawsoni*). The fifth individual was consistently recovered as more closely related to the two widespread subspecies than to the remaining individuals from California. While additional sampling of individuals from this population will be necessary to fully evaluate the distinctiveness of this lineage, it is possible the fifth individual may be admixed, or a vagrant from a more northern population that is unrepresentative of the rest of this population. Support of this hypothesis with additional samples would suggest the California population is distinct, not only geographically, but evolutionary as well.

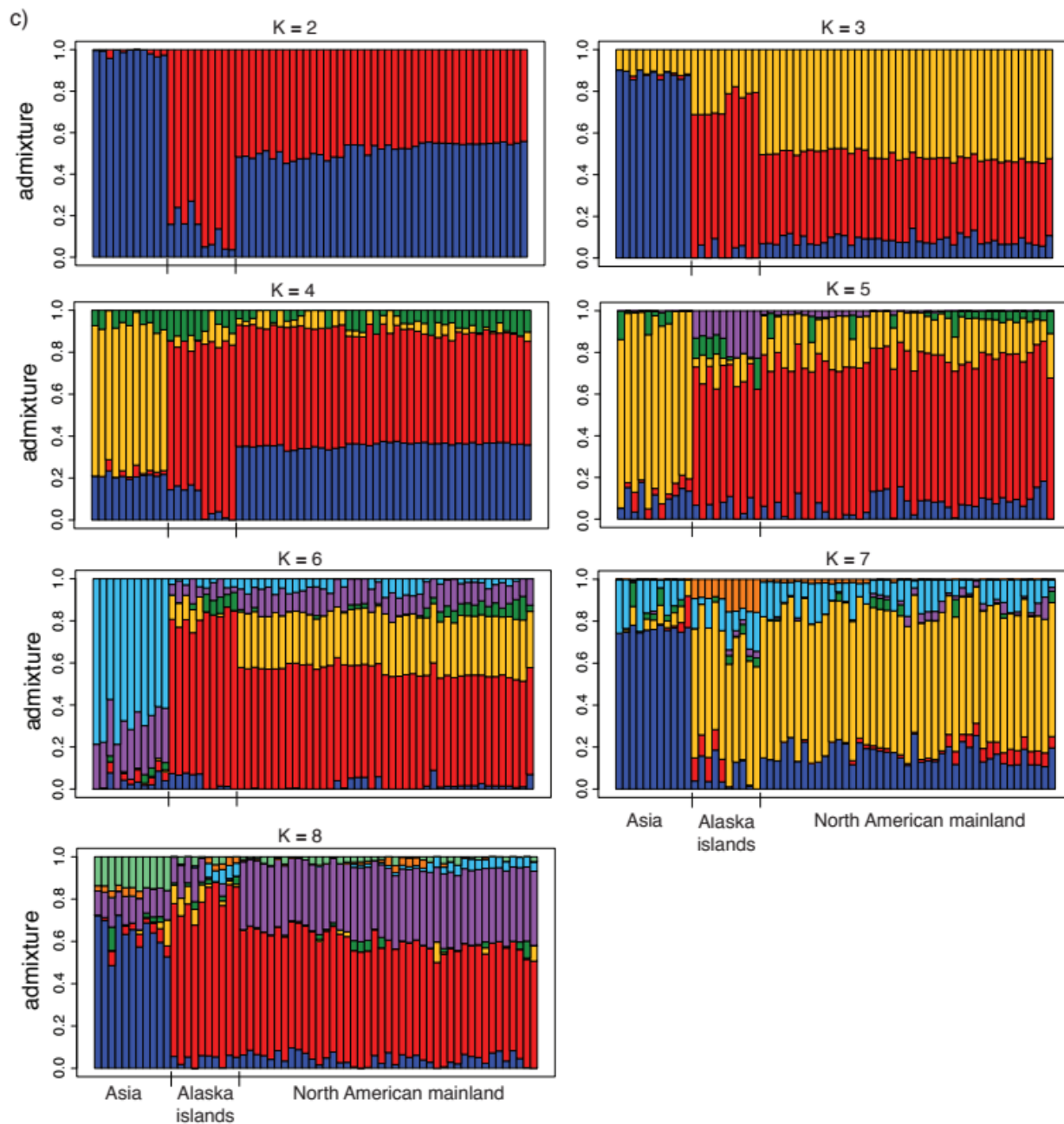
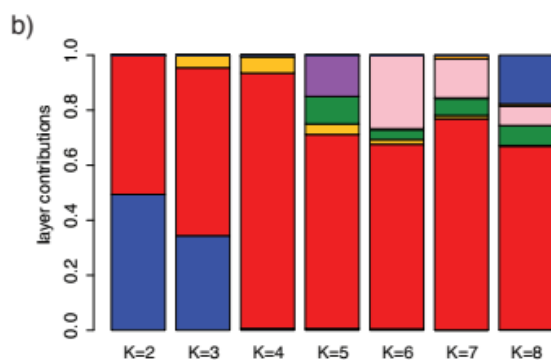
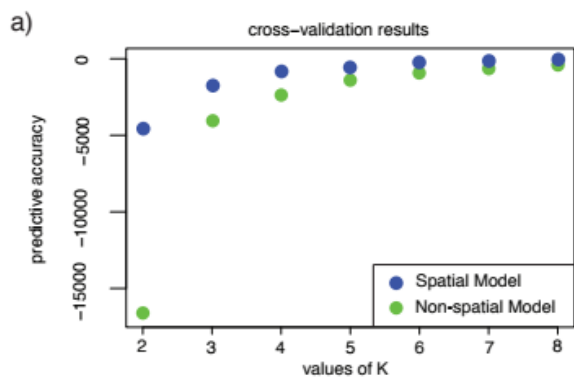
The remaining clades include two sister clades that correspond to the black rosy-finch, and the brown-capped rosy-finch. These clades were supported by all phylogenetic analyses

(but see the main text description and Fig. S2 for variation in the relationships of these clades with respect to the California individuals). As described in the main text, one black rosy-finch (RF37, MSB:30447) varied in placement. This variation in placement appears to be likely due to sequence coverage, and is likely not reflective of a true lack of monophyly. Further, the information added by the inclusion of additional sequence data in the 0p datasets compared to the 75p datasets recover this individual as a member of the black rosy-finch clade. These results lend support for the conclusion that these taxa both form distinct, divergent evolutionary lineages.

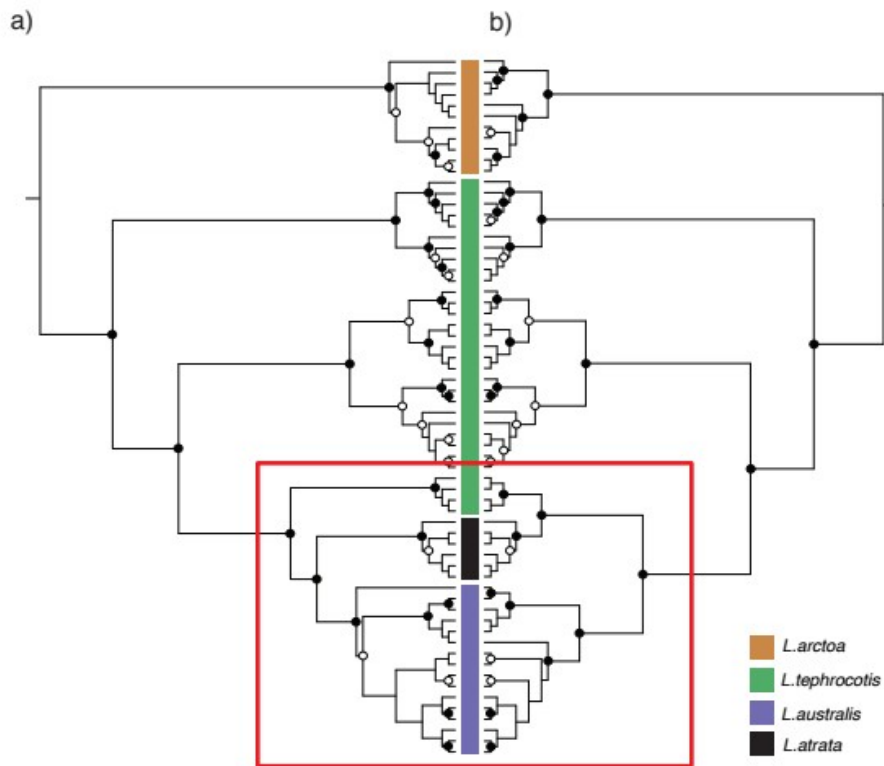
#### *Supplemental References*

- Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 – 8.
- Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.
- Swofford D.L. 2003. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods). Sunderland, MA: Sinauer Associates.

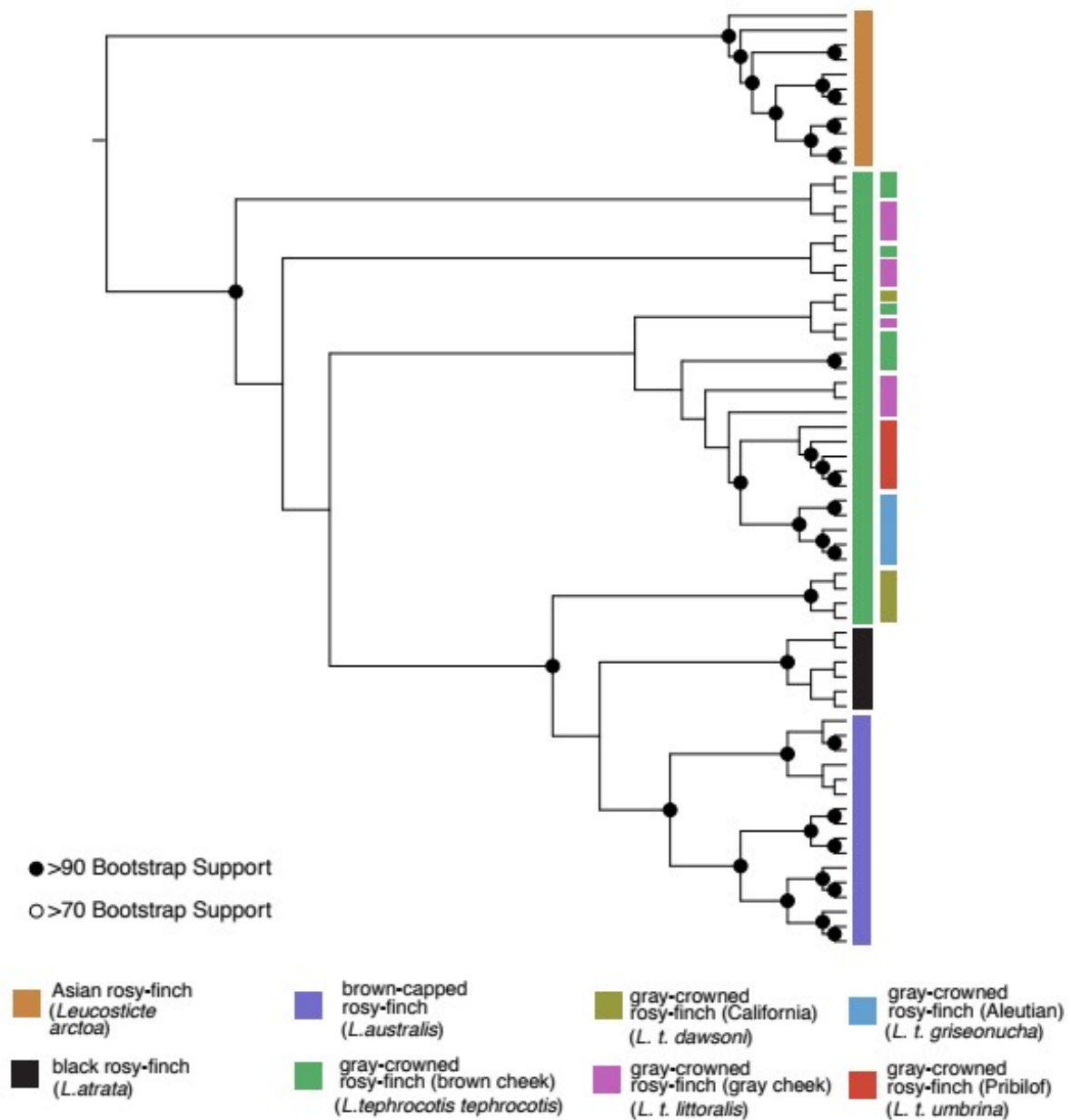
SUPPLEMENTAL FIGURES



**FIGURE S1.** Results from conStruct showing a) cross validation, b) Layer contributions, and c) admixture plots for all values of  $K$ . “Alaska island” cluster includes individuals from both the Aleutian and Pribilof islands.



**FIGURE S2.** Phylogenies generated using SVDquartets from total data matrices (0p) for both a) 0.05 and b) 0.1 minor allele frequency filter data sets. The two topologies conflict in the region highlighted by the red box. The only strongly supported topological conflict that relates to species level relationships is the placement of the clade containing four *L. tephrocotis* individuals from California, shown in a) as sister to a clade with both black rosy-finches and brown-capped rosy-finches, and in b) as sister to black rosy-finches. Open circles represent nodes with greater than 70% bootstrap support and closed circles represent nodes with greater than 90% bootstrap support.



**FIGURE S3.** Best ML tree and bootstrap support values generated using RAxML from the total data matrix (0p) and a 0.05 minor allele frequency. Color bars indicate taxonomic classification at both the specific (left column), and subspecific levels (right column). Subspecies for the Asian rosy-finch (*L. arctoa*) are not indicated. Open circles represent nodes with greater than 70% bootstrap support and closed circles represent nodes with greater than 90% bootstrap support.