**Generating genomic data sets**

**De novo assembly**

After demultiplexing we trimmed reads to the same 85 base pairs length using trimmomatic (Bolger et al., 2014). To identify the optimal parameters for assembling loci de novo in STACKS, we followed the parameter-testing pipeline established by Rochette and Catchen (2017) and Paris et al. (2017). We ran the STACKS multiple times varying the parameters on each step. We ranged the ‘*m’* parameter from 3 to 10 (m3 – m10), the ‘*M’* parameter from 3 to 5 (M3–M5), the ‘*n*’parameter from 2 to 7 (n2–n7) and ‘*r*’ was kept constant at 0.80, with the rest of the parameters on default setting. ‘*r*’ = 0.80 signifies that a locus must be present in a minimum of 80% of individuals.

To examine the output from each de novo assembly run, we used vcftools to create statistics for each run (Danecek et al., 2011). These statistics were generated and examined to select the optimal parameters for our data set. The statistics used were variant quality, variant mean depth, variant missingness, minor allele frequency and were then observed in R. The results generated very similar outputs and we chose the parameter combination -*m* 5 -*M* 3 -*n* 3 for the. After parameter selection, we used SNPfiltR to streamline and automate the process of choosing appropriate filtering parameters for our dataset (DeRaad, 2022; Knaus and Grunwald, 2017). Briefly, we filtered for genotype depth and quality, allele balance, maximum depth, and missing data per sample and per SNP. Two different datasets were used to included different samples for different purposes. Lists of samples and full parameter optimization and filtering details are available in supplementary methods for each data set (organize sample list and link to R code and visualizations detailing filtering pipeline).

**Phylogenetic analysis**

We constructed a maximum likelihood (ML) tree with IQ-TREE 2.2.0 (Minh et al. 2020) and used ModelFinder (Kalyaanamoorthy et al. 2017) to select a model for Akaike Information Criterion (AIC). Branch support was evaluated with 1000 ultrafast bootstrap (UFBoot) (Hoang et al. 2018) and 1000 Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) (Guindon et al. 2010) replications. We report the selected model and branch support values in the results section.

**Population structure and assignment**

We used three different approaches to identify and describe genomic structure. First, we used PCA for an initial exploratory screening of data in R (cite adegenet package). We ran PCA analysis to assess population structure and the find the number of clusters (k). Since PCA analysis makes no assumptions about sampled and ancestral populations, we ran a tool in R, sNMF for estimates of individual admixture coefficients (Frichot et al., 2014), implemented in the r package lea 2.2.0 (Frichot & François, 2015). To infer the best-fit number of populations (k) and construct assignment plots with admixture coefficients. We set the parameter alpha at 10 and performed 50 replicates for each value of k tested. Due to the unstable nature of estimating k (Kalinowski, 2011) and given uneven sampling, we include and interpret values of k greater than the one inferred by the cross-entropy criterion. The third approach for calculating genetic structure was a phenotype-locality level pairwise Fst sampling (Nei & Chesser, 1983). Fst was calculated using a custom script as a wrapper for vcftools, which calculates

Weir and Cockerham’s (1984) estimator for Fst using a “ratio of averages” approach to test for gene flow and then were visualized using R plotting functions. PCA, sNMF and Fst scripts are available in supplementary methods (link pipeline).

**Testing for gene flow**

We used treemix v1.13 (Pickrell & Pritchard, 2012), to estimate genetic relationships and admixture between Aciurina groups utilizing the covariance of allele frequencies among populations. This tool estimates a maximum likelihood (ML) species tree and identifies species pairs exhibiting closer relationships than can be accounted for by the species tree structure. These identified species pairs are considered potential candidates for migration events. We set the Opaca samples as root, then applied the TreeMix model while considering linkage disequilibrium by organizing sites into blocks of 1,000 single-nucleotide polymorphisms (-k 1,000). Standard errors (-SE) was used to assess the reliability of the tree topology. Following the construction of a maximum-likelihood tree, migration edges (−m) were introduced until we observed an increase in likelihood of <1. To test for introgression between *A. trixa* and *A. bigloviae* to further examine the mitonuclear discordance, we also performed supervised runs of treemix by adding a priori migration edges using the -cor\_mig and -climb flags. Specifically, we manually tested # of edges: Cotton and Smooth groups SPECIFY. This eas to test the hypothesis of introgression to explain phylogenetic discrepancy between COI and nuclear analyses. The resulting maximum-likelihood trees were visualized using the built-in TreeMix R script plotting functions.

We conducted ABBA/BABA tests for introgression (Durand et al., 2011) within location-phenotype populations to assess instances of non-sister introgression or deviations from a strictly bifurcating evolutionary history. Specifically, we computed the D statistic, an integral part of ABBA/BABA tests requiring four taxa (P1, P2, P3, and P4). In this setup, P1 and P2 are considered sisters, with P3 as a sister to that group, and for P4 serving as an outgroup we used the opaca samples.

We used the f4-ratio statistic to assess genomic evidence for interspecific gene exchange. We calculated the f4-ratio for all combinations of trios of species on the filtered VCF files using the software Dsuite (v.0.2 r20). The f4-ratio statistic estimates the admixture proportion, that is, the proportion of the genome affected by gene flow. The results presented in this study (Fig. Dsuite) are based on the ‘tree’ output of the Dsuite function Dtrios, with each trio arranged according to the species tree on the basis of the maximum-likelihood topology found using IQ-Tree.

The ABBA/BABA test examines alleles present exclusively in P3 and in either P1 (ABBA pattern) or P2 (BABA pattern). Both patterns could arise from incomplete lineage sorting or gene flow between P3 and the other population. Significantly more ABBA patterns than BABA patterns suggest gene flow between P2 and P3 (and vice versa). The D statistic, generated by this test, represents the ratio of ABBA patterns minus BABA patterns divided by the sum of ABBA and BABA patterns (Durand et al., 2011).

D ranges from zero (no introgression) to one (complete introgression). In our study, we focused solely on gene flow between P2 and P3 for clarity. The magnitude of D is proportionate to the extent of gene flow between P3 and the other population, although it may be underestimated if there is gene flow between P1 and P2 (Durand et al., 2011). We executed these tests in Bash scripting, utilizing scripts developed by Ethan Gyllenhaal (https://github.com/ethangyllenhaal), along with custom bootstrapping functions.

**Results**

* **Stacks**

After cleaning the first data set consisted in all New Mexico samples with the Reno ones included in a filtered SNP data sets containing 33 and original samples, filtered from 22,412 to 14,102 variants with 12.42% overall missing data. The second dataset included all New Mexico samples without the three Reno ones. The initial 30 samples were retained and contained 26,337 sites filtered down to 15,706 with 13,706% missing data.

* **Phylogenetics**

Mitochondrial and nuclear data showed conflicting relationships across the *A. trixa* and *A. bigloviae* species/groups.

* **PCA, sNMF and Fst**

Our results indicated all different gall phenotypes, smooth, cotton and Notata are three independent genetic clusters regardless of location.

Our sMNF analyses show structure across the study system and that all Aciurina species included are distinct contradicting? mitochondrial DNA. Optimal K was 3 but because k optimization is often conservative and can be unreliable (Kalinowski, 2011), we included analyses from a range of k (2–5). As k is increased (k=4 and K=5), populations start parsing geographically within phenotype. For example, at k = 3, all three species are separated, but at k = 4 cotton populations from northern and southern New Mexico start to be separated and at k = 5, this happens with the smooth populations as well.

\*\*\*For Fst values, in the other diagonal we can put mitochondrial distances.

**Figure SNMF.** No evidence of recent hybridization in nuclear genomes between *A. trixa* and *A. bigloviae* individuals. INCLUDE MAP and PCA IN THIS FIGURE? SNMF: Population assignment plots generated in sNMF of each individual. \*\*\*Include (1) Variance explained across different values of m (likelihood values) and (2) delta m spike supporting the optimal migrations at m = 1.

**Figure Optimal K.**

**Figure Fst.** Fixation index statistics between clusters are shown in the matrix. Warmer colors show a higher Fst value indicating more differentiation between groups whereas cooler colors indicate less differentiation/more proximity.

* **Treemix**

Ran models with 0 to 3 migration edges (Figure Treemix).

For migration edges, a weight close to 1 indicates strong gene flow or migration between the populations. A weight close to 0 suggests limited or no gene flow. Migration edges ranged from 0.018 to 0.067. Two out of three p-values were below 0.05 in model 3 (3 migration events). Weight suggests that ~2% and ~7% of the genome is shared between the groups. (Treemix table)

Suggested gene flow between the RGG-S to RGG-C + TES-C, AND MAL-N + TES-N to RGG-C.

treemix produced a topology concordant with phylogeny IQ-TREE.

BUT edges coming from nodes are maybe just noise

**Figure Treemix.** Species trees estimated in TreeMix. 1. TreeMix species tree with no migration edges permitted. 2-6. Inferred species tree with edges added from 1 to 5; Best-inferred species tree with highest likelihood score was 4, this edge p-value (p= 0.53). Scale bar shows 10 times the average standard error of the estimated entries in the sample covariance matrix. Migration edges were colored according to percent ancestry received from the donor population.

* **Dsuite**

**Figure Dsuite.** F-branch (fb(C)) statistics across our dataset highlight excess allele-sharing between tips in the tree (which represent individuals grouped by locality and gall phenotype; horizontally arranged at the top of the figure) and each other tip (solid line) and node (dotted line) in the phylogenetic tree (vertically arranged on the left of the figure), compared to its sister branch. The associated locality and gall phenotype of each tree tip is indicated by the symbol and colour (as in Fig. X). The redness of each cell in the matrix indicates the degree of excess allele-sharing between each tree tip (C) and each tip or node (b) with significant instances of excess allele-sharing, where the Z-score was >4.41 (equivalent to the Bonferroni multiple-testing corrected P-value of 0.01), are highlighted with a dot. Grey shading indicates tests which cannot be carried out due to the topology of the tree.

**References**

**Figures**

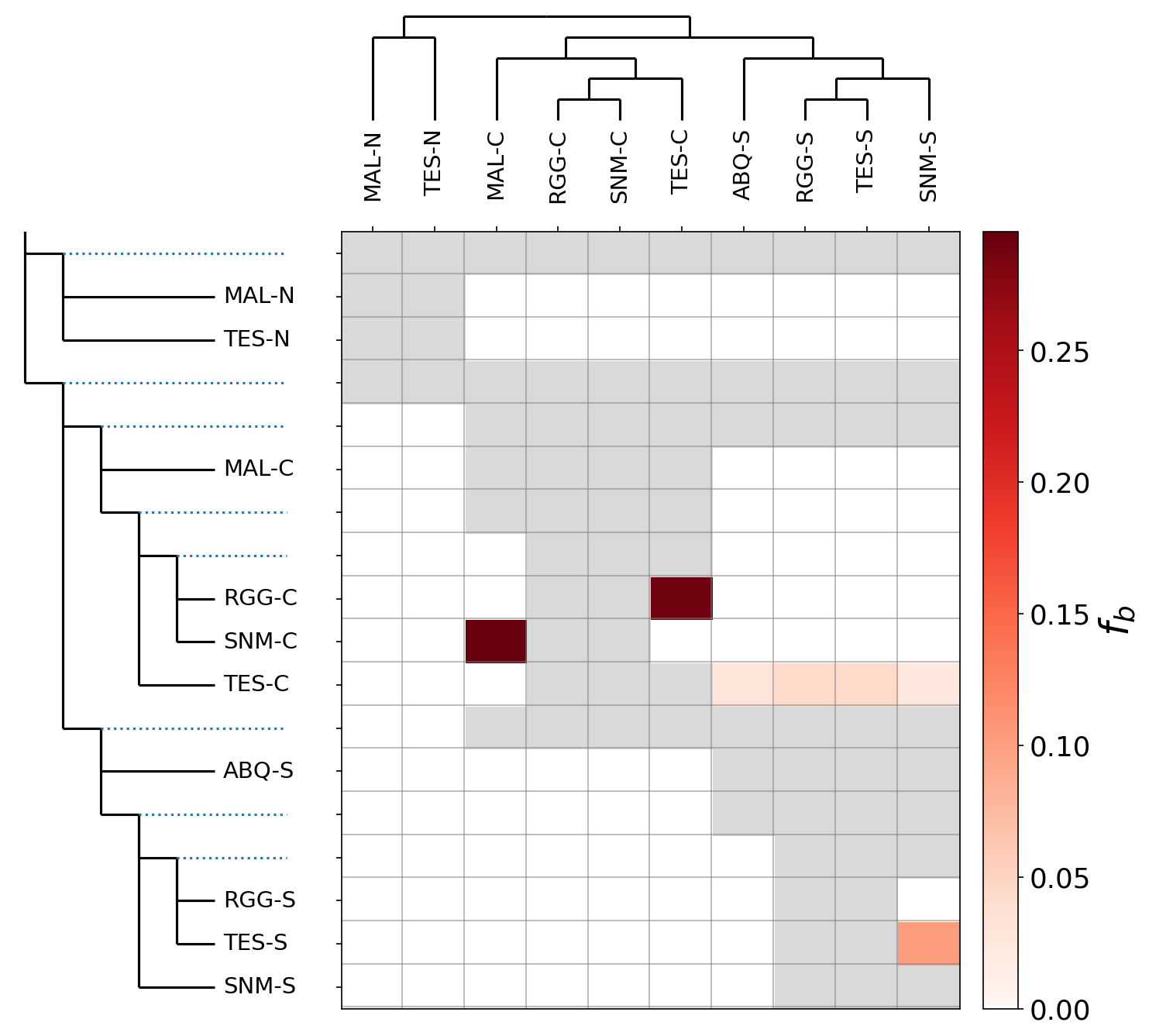
****







****

****

****

**Tables**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Weight** | **Jackknife** | **Weight Jackknife** | **P-Value** | **Subtree\_Origin** | **Subtree\_Destination** |
| 0.0587654 | 0.0605928 | 0.0406347 | 0.0679598 | (RGG-C:0,TES-C:0):0.0021851 | SLV-O:0.0929742 |
| 0.0670482 | 0.0670237 | 0.00963541 | 1.75E-12 | RGG-S:0 | (RGG-C:0,TES-C:0):0.0021851 |
| 0.0188397 | 0.0190513 | 0.00618688 | 0.00103741 | (MAL-N:0,TES-N:0):0.0660543 | RGG-C:0 |