

1

2 **Supplementary Information for**

3 Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the
4 Hawaiian Islands

5

6 Jae Young Choi^{a,*}, Xiaoguang Dai^b, Ornob Alam^a, Julie Z. Peng^c, Priyesh Rughani^b, Scott
7 Hickey^d, Eoghan Harrington^b, Sissel Juul^b, Julien Ayroles^c, Michael Purugganan^a, and Elizabeth
8 A. Stacy^{e,*}

9

10 ^aCenter for Genomics and Systems Biology, Department of Biology, New York University, New
11 York, New York, USA

12 ^bOxford Nanopore Technologies Inc., New York, New York, USA

13 ^cDepartment of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey,
14 USA

15 ^dOxford Nanopore Technologies Inc., San Francisco, California, USA

16 ^eSchool of Life Sciences, University of Nevada Las Vegas, Las Vegas, Nevada, USA

17

18 ***Corresponding authors:** JYC (jyc387@nyu.edu) and EAS (elizabeth.stacy@unlv.edu)

19 **This PDF file includes:**

20 Supplementary Text

21 Figures S1 to S19

22 Tables S1 to S5

23 References

24 **Supplementary information text**

25

26 **Plant material for the reference genome.** About 25 g of young leaves were collected from a
27 mature individual of *M. polymorpha* var. *incana* (NG4) maintained in a coldframe at the
28 University of Hawaii Hilo. NG4 was produced through a controlled-cross between two trees of
29 this variety occurring along Kuliouou Trail, Oahu (approximate location: 21.3160, -157.7296).
30 Leaf material was collected in small batches, wrapped in aluminum foil, and submerged in liquid
31 N₂ within one minute of collection. The bundled samples were then kept at -80C for 72 hours
32 and shipped overnight on dry ice to Oxford Nanopore Technologies, New York, NY and stored
33 again at -80C. The entire tree was covered in a black plastic bag (dark-treated) for 24 hours prior
34 to collection.

35

36 **Nanopore sequencing-based whole-genome-, RNA-, and Pore-C sequencing.** Using the
37 Qiagen DNeasy Plant Mini Kit, DNA was extracted from 2 g of collected leaf tissue. Separately,
38 total RNA was extracted from 1 g of collected leaf tissue using Thermofisher's PureLink RNA
39 Mini Kit. Full-length cDNA was synthesized from 50 ng of total RNA using the Oxford
40 Nanopore Technologies PCS109 kit, followed by 14 rounds of PCR amplification using the
41 primer mixture from the Oxford Nanopore Technologies EXP-PCA001 kit. A detailed protocol
42 outlining the Pore-C method can be found as supplemental text
43 (<https://doi.org/10.5281/zenodo.4264399>).

44 A sequencing library was prepared using the Oxford Nanopore Technologies standard
45 ligation sequencing kit SQK-LSK109. Sequencing was conducted on a GridION X5 sequencer
46 for 72 hours, and the raw data were base-called by Oxford Nanopore Technologies basecaller
47 Guppy (available on <https://community.nanoporetech.com/>) ver. 3.2.8 for the genomic DNA and
48 Pore-C DNA, and ver. 3.2.10 for the cDNA in the high-accuracy mode.

49

50 **Nanopore sequence-based reference genome assembly.** The FASTQ files from the whole-
51 genome sequencing data were filtered for high-quality long reads. We used the program `filtlong`
52 (<https://github.com/rrwick/Filtlong>) with parameters `--min_length 10000 --min_mean_q 85 --`
53 `min_window_q 70`, which selects for reads longer than 10 kbp, average Q-score greater than 8.2,

54 and a minimum sliding window Q-score of 5.2. The filtered nanopore reads were then assembled
55 with the genome assembler flye (1).

56

57 **Assembly contig scaffolding with Pore-C sequencing data.** The Pore-C data analysis was
58 conducted using the Pore-C workflow developed by Oxford Nanopore Technologies
59 (<https://github.com/nanoporetech/Pore-C-Snakemake>), which uses the snakemake workflow
60 engine (2). Briefly, the workflow first aligns the nanopore Pore-C chromosome contact sequence
61 reads to the unscaffolded *Metrosideros* genome assembly using bwa-sw ver. 0.7.17-r1188 (3)
62 with parameters -b 5 -q 2 -r 1 -T 15 -z 10. Compared to conventional Hi-C data, Pore-C contains
63 an enrichment of higher order contacts (4), and to process the multi-contact nanopore reads we
64 used Pore-C tools (<https://github.com/nanoporetech/pore-c>) also developed by Oxford Nanopore
65 Technologies. The alignment BAM file was processed with Pore-C tool to filter spurious
66 alignments, detect ligation junctions, and assign fragments that originated from the same
67 chromosomal contacts. Pore-C tools converted the alignment BAM file to a chromosome contact
68 pairs format (https://github.com/4dn-dcic/pairix/blob/master/pairs_format_specification.md) for
69 compatibility with the conventional downstream chromosome contact based analysis methods.
70 The pairs file was converted to a hic file format using the Juicer ver. 1.14.08 tools (5) to use as
71 the input data for the Juicebox assembly tools (6). The Juicebox tools were used to scaffold the
72 draft *Metrosideros* assembly, and we followed established guidelines
73 (<https://github.com/theaidenlab/Genome-Assembly-Cookbook>) to manually construct the
74 chromosome-scale scaffolds using the Pore-C based chromosome contact frequency information.
75 We assigned chromosome numbers to the superscaffold through synteny with the *Eucalyptus*
76 *grandis* genome assembly (7). Synteny between the *Metrosideros* superscaffold and *Eucalyptus*
77 chromosomes was determined by aligning the assemblies to each other and visualizing the
78 alignment through the program D-GENIES (8).

79

80 **Genome annotation.** The cDNA library that was sequenced on the nanopore sequencer was
81 used to annotate the coding sequence regions for the *M. polymorpha* genome assembly. Initially,
82 we used Pychopper ver. 2.3.1 (<https://github.com/nanoporetech/pychopper>) to trim primers,
83 identify full-length cDNA sequences, and orient the sequence to the correct strand. A total of
84 12,298,201 (70.2%) reads were classified by Pychopper and were used for downstream analysis.

85 The long reads were aligned to the reference genome using minimap2 ver. 2.17-r941 (9) with
86 options -ax splice -uf -k14. The alignment file was then used by stringtie2 ver. 2.1.3b (10), which
87 is optimized for *de novo* transcriptome assembly using long-read sequencing and a reference
88 genome. We used the MAKER program (11) for gene annotation using the workflow outlined on
89 the website <https://gist.github.com/darencard/bb1001ac1532dd4225b030cf0cd61ce2>. The
90 transcriptome assembly from stringtie2 was used as EST evidence in MAKER, and the protein
91 sequences from the previous *M. polymorpha* assembly (12), *E. grandis* assembly (7), and *A.*
92 *thaliana* (TAIR10) were used for a protein homology search in MAKER. After an initial round
93 of MAKER annotation the gene models were used by SNAP (13) and Augustus (14) to create
94 gene model training datasets specifically for our *M. polymorpha* genome assembly. The training
95 dataset was used for a second round of MAKER gene annotation.

96 We identified the repetitive regions of the *M. polymorpha* reference genome, first using
97 Repeatmodeler ver. 1.0.10 (<http://www.repeatmasker.org/RepeatModeler/>) for the *de novo*
98 identification repeat sequences in the reference genome, and then using Repeatmasker ver. 4.1.0
99 (<http://www.repeatmasker.org/RepeatMasker/>) to identify the genomic locations of the repetitive
100 sequences in our reference genome.

101

102 ***Metrosideros* population sequencing.** For the population genomic sampling, we collected young
103 leaf tissue from 9-11 adults from each of eight taxa on the focal island of Oahu and fewer adults
104 of two taxa on Kauai. Collected leaf tissue was kept cool and stored at -80C within 48 hours of
105 collection. Leaf material from the three outgroup samples was silica-dried in the field and stored
106 in a dessicator jar. DNA was extracted from both frozen and dried leaf samples using the
107 Macherey-Nagel NucleoSpin Plant II Mini kit. We used a Tn5 transposase-based method to
108 prepare the whole-genome sequencing library. Mosaic End adaptor A and B (Tn5ME-A:
109 TCGTCGGCAGCGTCAGATGTGTAT AAGAGACAG; Tn5ME-B:
110 GTCTCGTGGGCTCGGAGATGTGTATAAAGAGACAG) was annealed with Rev (Tn5ME-
111 Rev: /5Phos/CTGTCTCTTATACACATCT) by mixing 10uL (100uM) of each oligonucleotide
112 with 80 uL of reassociation buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA) in BioRad
113 thermocycler with the following program: 95°C for 10 min, 90°C for 1 min, and decrease
114 temperature by 1°C/cycle for 60 cycles, held for 1 min at each temperature. Pre-charge of Tn5
115 with adapters was carried out in solution by mixing 22.5 uL of 100 ng/uL Tn5 (Tn5 protein was

116 produced following the protocol described by (15)), 76.5 uL reassociation buffer/glycerol (1:1),
117 and 4.5 uL of equal molar of annealed adaptor 1 (A-Rev) and annealed adapter 2 (B-Rev). The
118 reaction was then incubated at 37°C for 30 min. The annealed adapters bind to Tn5 transposase
119 to form the transposome complex.

120 Genomic DNA was tagged by mixing with 1 uL of the above assembled Tn5
121 transposome, 4 uL of 5 X TAPS buffer (50 mM TAPS-NaOH pH 8.5 [Alfa aesar # J63268], 25
122 mM MgCl₂, 50% v/dimethylformamide [ThermoFisher #20673], pH 8.5 at 25°C) and water to a
123 total volume of 20 uL and incubated at 55°C for 7 min. The transposome fragments and attaches
124 adapters to gDNA. The reaction was completed by adding 5 uL of 0.2% SDS (Promega,
125 #V6551) to each reaction followed by incubation at 55°C for 7 min to inactivate and release the
126 Tn5. To enrich the DNA fragments that have adapter molecules on both ends we attached an
127 index to the library: 2µl of the stopped tagmentation, 1µl i5 index primer (1uM), 1µl i7 index
128 primer (1µM), 10µl of OneTaq HS Quick-Load 2x master mix (NEB #M0486L), and 6µl of
129 water were combined to make a 20-µl final reaction. The reaction was heated at 68°C for 3min
130 and 95°C for 30 sec, then thermocycled 12 times at 95°C for 10 sec, 55°C for 30 sec, and 68°C
131 for 30 sec, followed by a final extension of 5 min at 68°C.

132 We then pooled the libraries together, taking 5uL from individual libraries. The pooled
133 library was cleaned and size-selected using Agencourt AMPure XP beads (Beckman Coulter,
134 #A63881) at a 0.8:1 (beads: DNA) ratio. The final library was quantified with a Qubit high-
135 sensitivity DNA kit (Invitrogen Q32854) and examined on an Agilent 2100 Bioanalyzer high-
136 sensitivity DNA chip (Agilent p/n# 2938-85004) to observe the library size distribution. The
137 sequencing library was loaded on a NovaSeq 6000 S1 flow cell and sequenced under a 2×150-bp
138 conformation at the Genomics Core Facility within the Lewis-Sigler Institute for Integrative
139 Genomics at Princeton University.

140

141 **Calling genome-wide polymorphisms.** Raw sequencing reads were downloaded from our
142 previous study (16) and combined with the newly generated sequencing data from the current
143 study. From the sequence read archive (SRA) website we downloaded FASTQs with SRR
144 identifiers SRR8943660 to SRR8943653. The sequencing reads were adapter-trimmed and
145 quality-controlled using BBTools (<https://jgi.doe.gov/data-and-tools/bbtools/>) bbdduk program

146 version 37.66 with option: minlen = 25 qtrim = rl trimq = 10 ktrim = r k = 25 mink = 11 hdist = 1
147 tpe tbo.

148 Sequencing reads were then aligned to the scaffolded *M. polymorpha* reference genome
149 generated from this study using bwa-mem. PCR duplicate reads were removed using picard
150 version 2.9.0 (<http://broadinstitute.github.io/picard/>). Genome-wide read coverage statistics were
151 calculated using GATK version 3.8–0 (<https://software.broadinstitute.org/gatk/>).

152 We used the GATK HaplotypeCaller engine to call variant sites from the BAM alignment
153 file for each sample. The option –ERC GVCF was used to output the variants in the gVCF
154 format, and the gVCFs of each sample were merged together to allow a multi-sample joint
155 genotype procedure using GATK GenotypeGVCFs engine. Using standard GATK best-practice
156 hard-filter guidelines, the VariantFiltration engine was used to filter out low-quality
157 polymorphisms. In addition, we removed SNPs that were within 5 bp of an INDEL and
158 polymorphic sites that had less than 80% of individuals with a genotype call.

159

160 **Population relationship analysis.** The alignment BAM files were used to analyze the
161 population relationships between samples. We used ANGSD version 0.929 (17) and ngsTools
162 (18) to analyze the genotype likelihoods of each sample and infer the population relationships
163 using a probabilistic framework. We only analyzed potential variant sites where more than 80%
164 of the individuals had a genotype, while enforcing a total sequencing coverage filter such that
165 included sites had a minimum of 1/3 the average total sequencing depth (734×) and a maximum
166 of three times the average total sequencing depth (6,613×). To minimize the effect of linkage on
167 inferences on population relationships, polymorphic sites were randomly pruned using a 10-kbp
168 sliding window with a minimum distance of 5 kbp between random sites.

169 NGSadmix (19) was used to estimate the admixture proportions (K) for each individual.
170 For each of K = 3 to 15, the analysis was repeated 100 times and the run with the highest log-
171 likelihood was chosen. Principal component analysis was also conducted using genotype
172 likelihoods. Genotype posterior probabilities were calculated using ANGSD and used by the
173 program ngsCovar (18) to conduct the principal component analysis.

174 Phylogenetic relationships were investigated using the hard-called genotypes from
175 GATK and extracting four-fold degenerate sites using a Python script (available at
176 <https://github.com/tvkent/Degeneracy>). A maximum-likelihood phylogenetic tree was

177 reconstructed through RAxML ver 8.2.12 (20) using the $-f a$ algorithm (*i.e.* the rapid bootstrap
178 analysis and search for best-scoring maximum-likelihood tree). 100 bootstrap replicates were
179 generated to obtain confidence in the tree topology. We also conducted Bayesian phylogenetic
180 analysis using the SNAPP package as part of the program BEAST ver. 2.6.3 (21, 22). Because
181 SNAPP is a resource-intensive method, for each taxon/population we randomly selected four
182 individuals and selected for biallelic polymorphic sites. Sites were randomly pruned in 10-kbp
183 windows to generate unlinked biallelic sites, which are required for SNAPP. Using BEAST, an
184 MCMC analysis was run for 3,000,000 generations with sampling every 1,000 generations. At
185 the end, 30% of the chains were discarded as burn-in.

186

187 **Investigating reticulate evolutionary history.** The genotype call dataset was used to examine
188 the reticulate evolutionary history of *Metrosideros*. We conducted the ABBA-BABA D test (23,
189 24) using the R package admixr (25), which is based on the ADMIXTOOLS suite (26). Variants
190 were polarized using the high-coverage Fiji sample *M. vitiensis* as the outgroup genome. The
191 topologies used were based on the maximum-likelihood tree, the SNAPP-based tree, and the
192 TWISST analysis. Specifically, across the island we assumed the topology [(Hawaii, Oahu),
193 Kauai], which was supported by PCA, maximum-likelihood tree, and SNAPP based tree. Within
194 islands, only Oahu had topologies that differed between the maximum-likelihood and SNAPP
195 trees. Within Oahu, we examined the TWISST result and chose the topologies that were most
196 frequent to represent relationships (*i.e.* pubescent group = [(I,C),(F,R)]; glabrous group =
197 [(B,L),(M,T)].

198 The genome-wide topological relationships were examined using the method TWISST
199 (27). The genotype call dataset was imputed and phased using the program Beagle ver. 5.0 (28,
200 29). The phased genotype dataset was used to conduct a sliding window-based estimation of the
201 local phylogenetic relationships. Using the `raxml_sliding_windows.py` script from the
202 `genomics_general` package
203 (https://github.com/simonhmartin/genomics_general/tree/master/phylo), phylogenetic trees were
204 reconstructed in windows of 200 polymorphic sites. The option ‘complete’ within the TWISST
205 program was used to calculate the exact weighting of each local window.

206

207 **Demographic modeling.** We used the methods $\delta a\delta i$ (30), G-PhoCS (31), and MSMC (32, 33) to
208 infer the demographic history of Hawaiian *Metrosideros*. For all analyses we used the genotype
209 call dataset.

210 For $\delta a\delta i$ analysis, we initially randomly thinned the dataset picking a SNP every 10 kbp
211 using PLINK ver. 2.0 (34). The site frequency spectrum was estimated using the easySFS.py
212 (<https://github.com/isaacovercast/easySFS>) script while using Fiji *M. vitiensis* as the outgroup
213 genome and polarizing the polymorphic sites. The easySFS.py script was also used to project
214 down the sample size to maximize the number of sites analyzed. The unfolded site frequency
215 spectrum data were used as input for $\delta a\delta i$, and we fit 20 demographic models (Supplemental
216 Figure 9). We optimized the model parameter estimates using the Nelder-Mead method by
217 randomly perturbing the parameter values for four rounds. The parameter estimates were
218 perturbed threefold, twofold, twofold, and onefold in incremental rounds. Each round the
219 perturbation was conducted for 10, 20, 30, and 40 replicates. Demography parameters were
220 extracted from the round with the highest likelihood. Demographic models were compared using
221 Akaike Information Criteria (AIC) values. The $\delta a\delta i$ analysis scripts were based on the study by
222 (35).

223 To prepare our dataset for G-PhoCS analysis we first partitioned our reference genome
224 into 1-kbp loci and determined those that are close to neutrality. Neutral loci were determined by
225 selecting loci that were 5 kb away from a genic sequence, 500 bp away from a repetitive DNA
226 sequence, and at least 10 kbp away from each other. Since G-PhoCS is designed to analyze the
227 variation within a single genome, we selected a single individual with high genome coverage to
228 represent each island. Selected samples included: H207 from Hawaii Island (population G_{HI}),
229 X83 from Molokai (population G_M), O385 from Oahu (taxon M), and K283 from Kauai
230 (population G_K). The *M. vitiensis* sample from Fiji was used as the outgroup. We used G-PhoCS
231 ver. 1.2.3 and ran every demographic model five times to check for convergence in the
232 demographic parameter estimates. Each MCMC run had 1,000,000 iterations, and the initial
233 500,000 iterations were discarded as burn-in. Priors were modeled using a gamma distribution
234 ($\alpha = 1$ and $\beta = 10,000$ for population size and divergence time; $\alpha = 0.002$ and $\beta = 0.00001$ for
235 migration rates). Different demographic models involved fitting migration bands between two
236 terminal lineages. After the MCMC run was complete, the program Tracer version 1.6
237 (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to estimate the 95% highest posterior density

238 for each demography parameter. The G-PhoCS-estimated divergence time τ is scaled according
239 to the mutation rate (μ). To convert divergence time to absolute divergence time T (in years) we
240 used the following equation:

$$T = \frac{\tau \times g}{\mu}$$

241 where g represents the generation time.

242 MSMC2 was used to estimate the past changes in effective population sizes and the
243 divergence times between individuals. From each taxon/population we chose a single
244 representative individual for the MSMC2 analysis as follows: O310 (taxon B), O65 (taxon C),
245 O194 (taxon L), H271 (taxon N), O464 (taxon R), O145 (taxon T), plus the individuals used in
246 the G-PhoCS analysis. We used the alignment BAM file for each individual and the mpileup
247 function of samtools ver. 1.3.1 (36) to detect sites that had a minimum base score of 30, a
248 mapping quality score of 30, and the coefficient to downgrade mapping qualities for excessive
249 mismatches at 50. The resulting text pileup output was used by bcftools ver. 1.3.1 to call variant
250 sites but excluding INDELs and limiting the calls to biallelic SNPs. The bamCaller.py script that
251 was provided by the MSMC suite (<https://github.com/stschiff/msmc-tools>) was then used to
252 produce the per-chromosome masks and VCF files for each individual. A genome-wide mask file
253 was also created for each individual by using the SNPable workflow
254 (<http://lh3lh3.users.sourceforge.net/snpable.shtml>) on each superscaffold/chromosome assembly
255 and then converting it to BED format using the makeMappabilityMask.py script within the
256 MSMC suite. Phasing was done by using the output from the beagle analysis. The input files for
257 MSMC2 were generated using the generate_multihetsep.py script, which is also part of the
258 MSMC suite. To estimate changes in effective population size we examined the two haplotypes
259 of each individual, while cross-coalescence rates were estimated from four haplotypes from two
260 individuals. The combineCrossCoal.py script from the MSMC suite was used to produce the
261 outputs for plotting.

262

263 **Simulations.** Simulations were conducted using the forward-time Wright-Fisher model-based
264 simulator SLiM (37). All of the SLiM code used is available on GitHub
265 (https://github.com/ornobalam/metrosideros_simulations). In all described simulations, the

266 starting population was allowed to evolve for a burn-in period of 50,000 non-overlapping
267 generations to generate the ancestral population.

268 For observing tree topologies, the basic model consisted of an ancestral population
269 (ABCD) with a fixed population size of N that split after 50,000 non-overlapping generations
270 into 2 daughter populations AB and CD, each with a fixed size of N . An outgroup population O
271 splits off from ABCD in the first generation with a fixed size of N . AB and CD were then
272 allowed to evolve for a further 15,500 and 5,000 non-overlapping generations before splitting
273 into daughter populations A and B, and C and D, respectively, each with a fixed size of N . A, B,
274 C, and D were then allowed to evolve for 150,000 generations since the origin of the ancestral
275 population. The split times for A, B, C, and D were chosen based on estimated split times for the
276 four sampled glabrous *Metrosideros* taxa from Oahu, using the cross-coalescence rate estimates
277 from MSMC2. We simulated a 1-Mbp chromosome with a recombination rate ranging between
278 10^{-9} and 10^{-7} , and mutation rate of 10^{-8} , both per base pair and per generation. The chromosome
279 comprised 30% genic regions with a 1:4 ratio of neutral to deleterious mutations, and 70% non-
280 coding regions with only neutral mutations.

281 We explored three population sizes (N) of the ancestor (ABCD): (1) a large $N = 100,000$
282 (2) a medium $N = 50,000$ and (3) a small $N = 10,000$. Ancient migration was modeled with a
283 rate, m_1 , between AB and CD, and recent migration was modeled with a rate, m_2 , among A, B,
284 C, and D (see Fig. S6 for visual representation of simulation model). Migration was exclusively
285 either ancient or recent in each simulation. For each N we simulated two different migration rates
286 of 0.1 or 0.001 for m_1 or m_2 . At the end of each simulation, 10 individuals were sampled and
287 outputted from populations A, B, C, D, and O, for analysis with TWISST.

288

289 **Population genomic analysis.** We used the gVCFs that were called from the previous step to
290 create a VCF file that had genotype calls for all sites including the non-variant positions. The
291 GATK GenotypeGVCFs engine was used with the option `-includeNonVariantSites`. We
292 analyzed a population VCF that had both variant and non-variant sites in order to obtain the
293 correct number of sites to be used as the denominator for the population genetic statistics we
294 were calculating. We used the `genomics_general` package
295 (https://github.com/simonhmartin/genomics_general) to calculate θ , D_{xy} , and F_{ST} in 10-kbp
296 windows, sliding the window by 5 kbp. For each window we imposed a quality filter only

297 analyzing sites that had a minimum quality score of 30 and a minimum depth of 5×. Windows
298 that had more than 30% of the sites with a genotype call after filtering were chosen for
299 downstream analysis.

300 Evidence of admixture in localized windows (38) was investigated using the
301 `abbababawindows.py` script from the `genomics_general` package to calculate the f_{dM} statistics. To
302 search for genomic outliers of differentiation, the F_{ST} values were Z-transformed (zF_{ST}), and
303 genomic windows with $zF_{ST} \geq 4$ were considered outliers (39).

304 The strength of evidence of selective sweeps was estimated using the ω statistic (40). We
305 used the program OmegaPlus (41), which is specifically designed to estimate the ω statistics for
306 genome-wide SNP datasets. We set the grid size of OmegaPlus so that the ω statistics would be
307 estimated at 10-kbp windows for each super-scaffold/chromosome.

308 Allele state and count per polymorphic site was determined using PLINK and used for
309 analyzing the evolutionary origin of differentiation outliers.

310

311 **Gene Ontology Enrichment.** Coding sequences of each gene model were assigned a
312 computationally predicted function and gene ontology using the eggNOG pipeline (42). We
313 required an ontology to have more than two gene group members for further consideration. Gene
314 ontology enrichment was tested through a hypergeometric test.

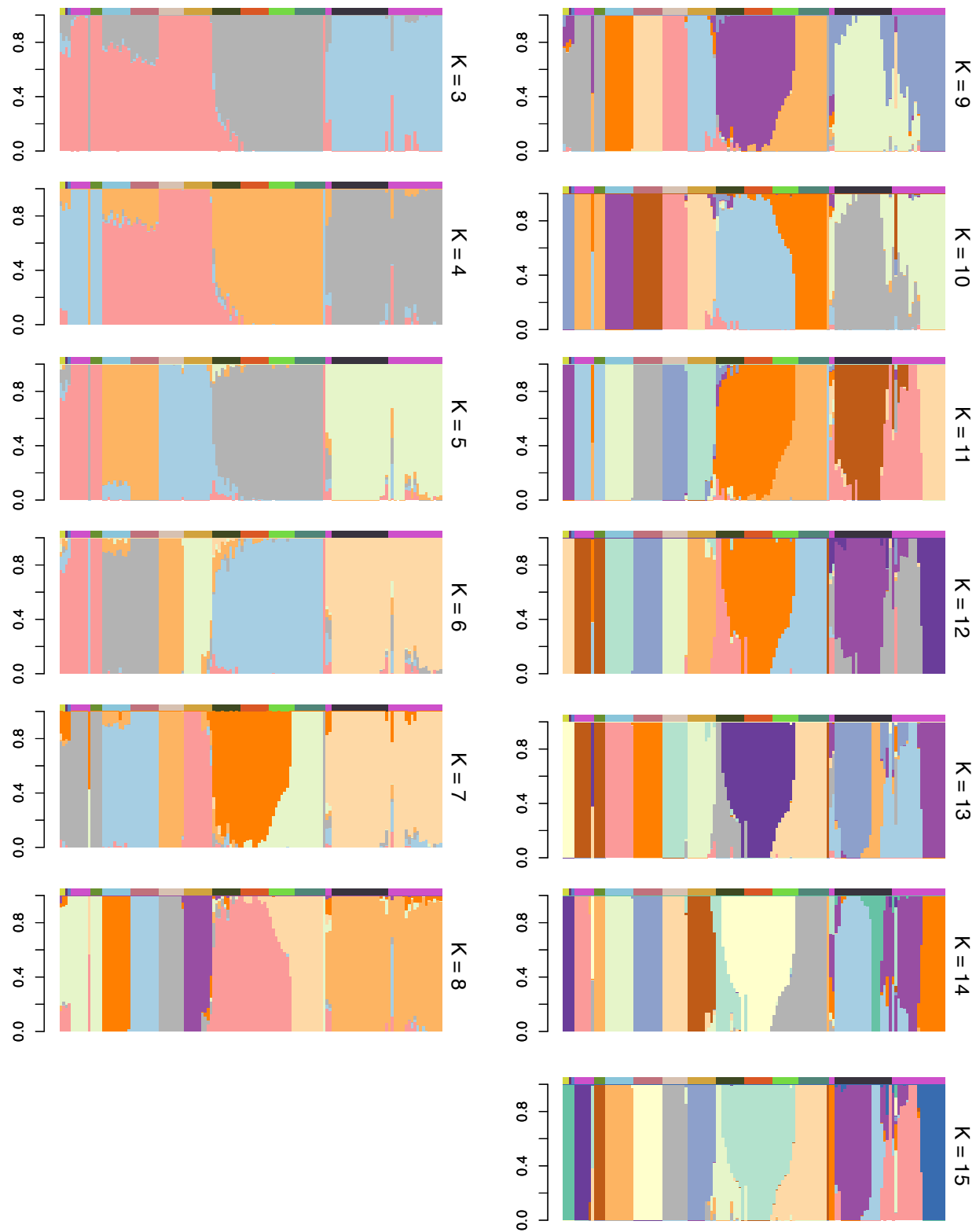


Fig. S1. Ancestry proportions for K=3 to K=15 across all Hawaiian archipelago samples.

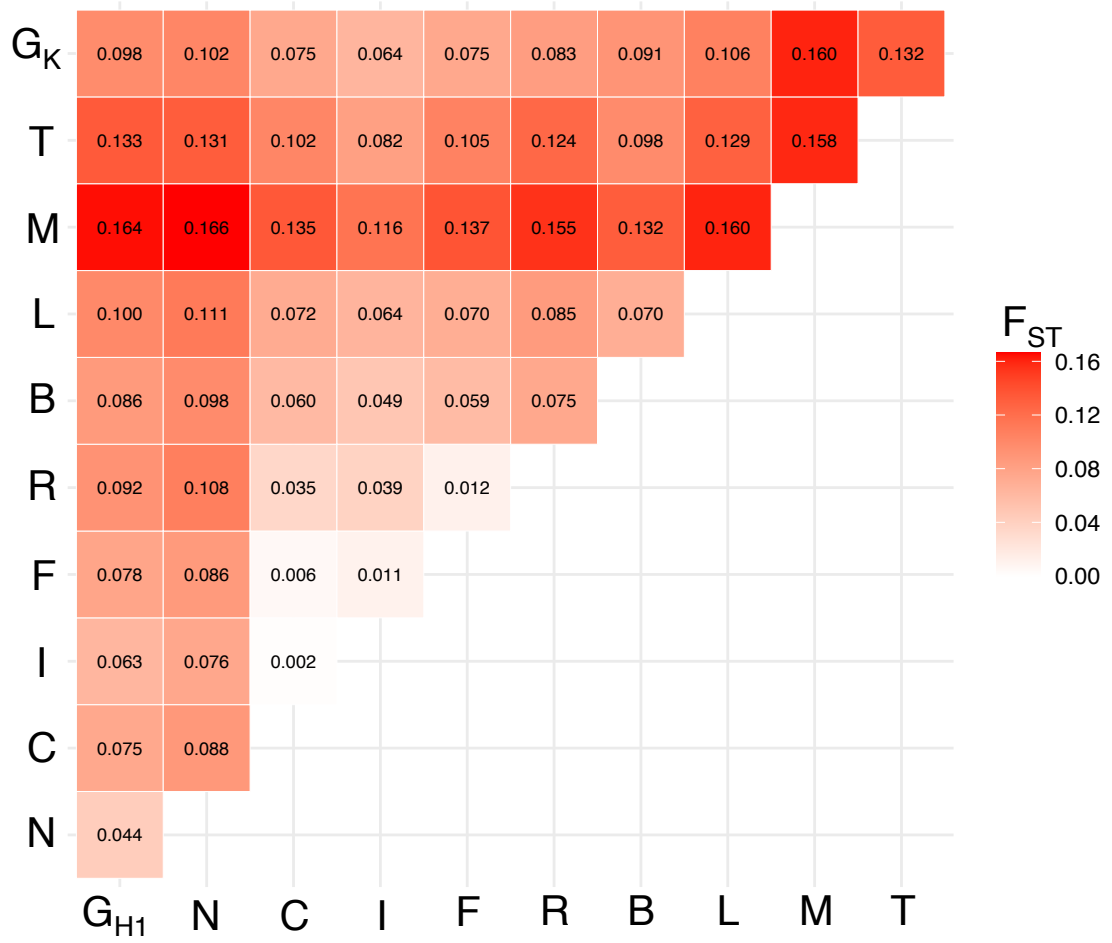


Fig. S2. Pairwise average F_{ST} between taxa/populations.

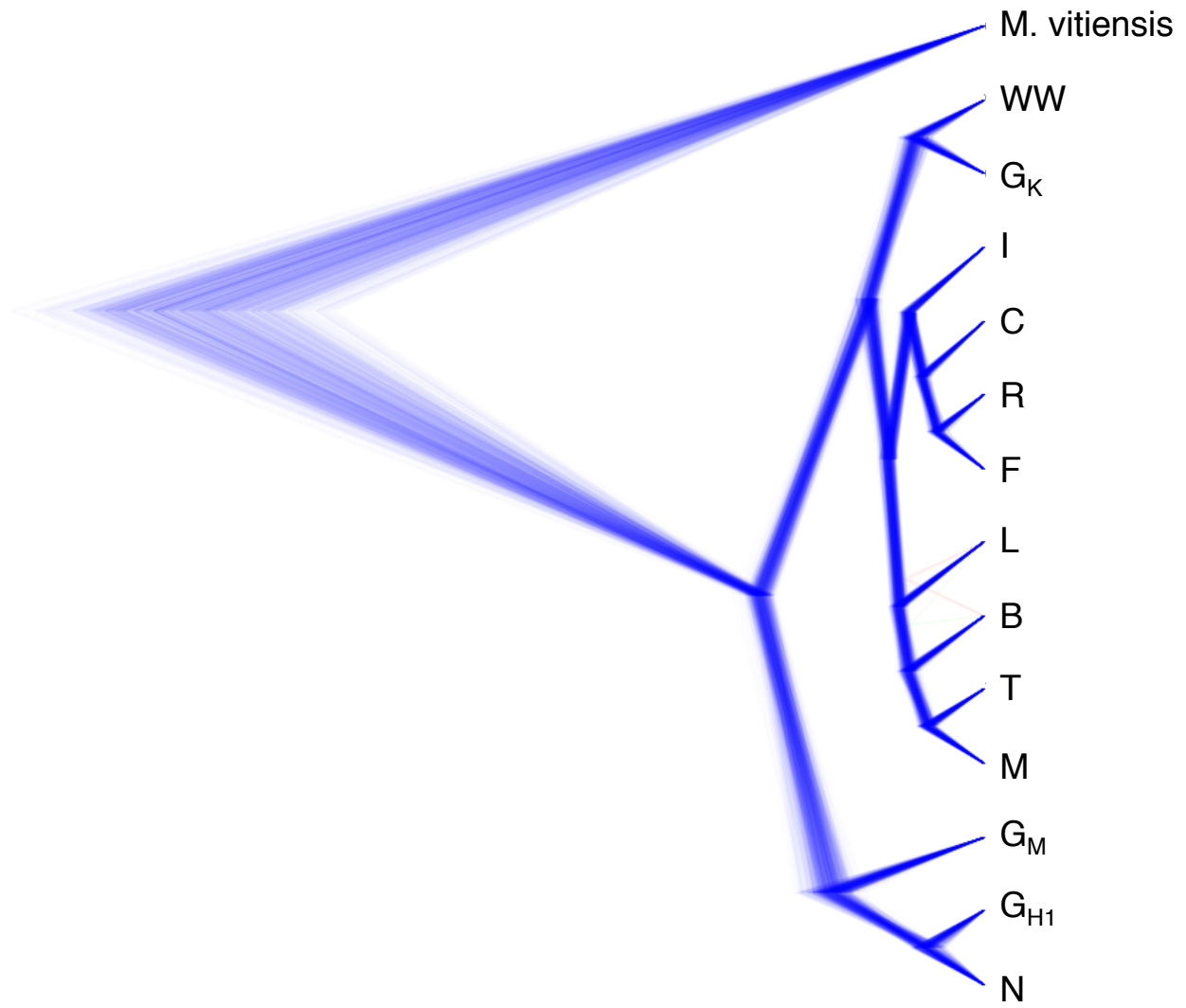


Fig. S3. A cloudogram generated from SNAPP-reconstructed population topology. The different colors represent different species topologies. Topologies with the highest clade credibility are shown in blue, while red represents minor trees with an alternative topology.

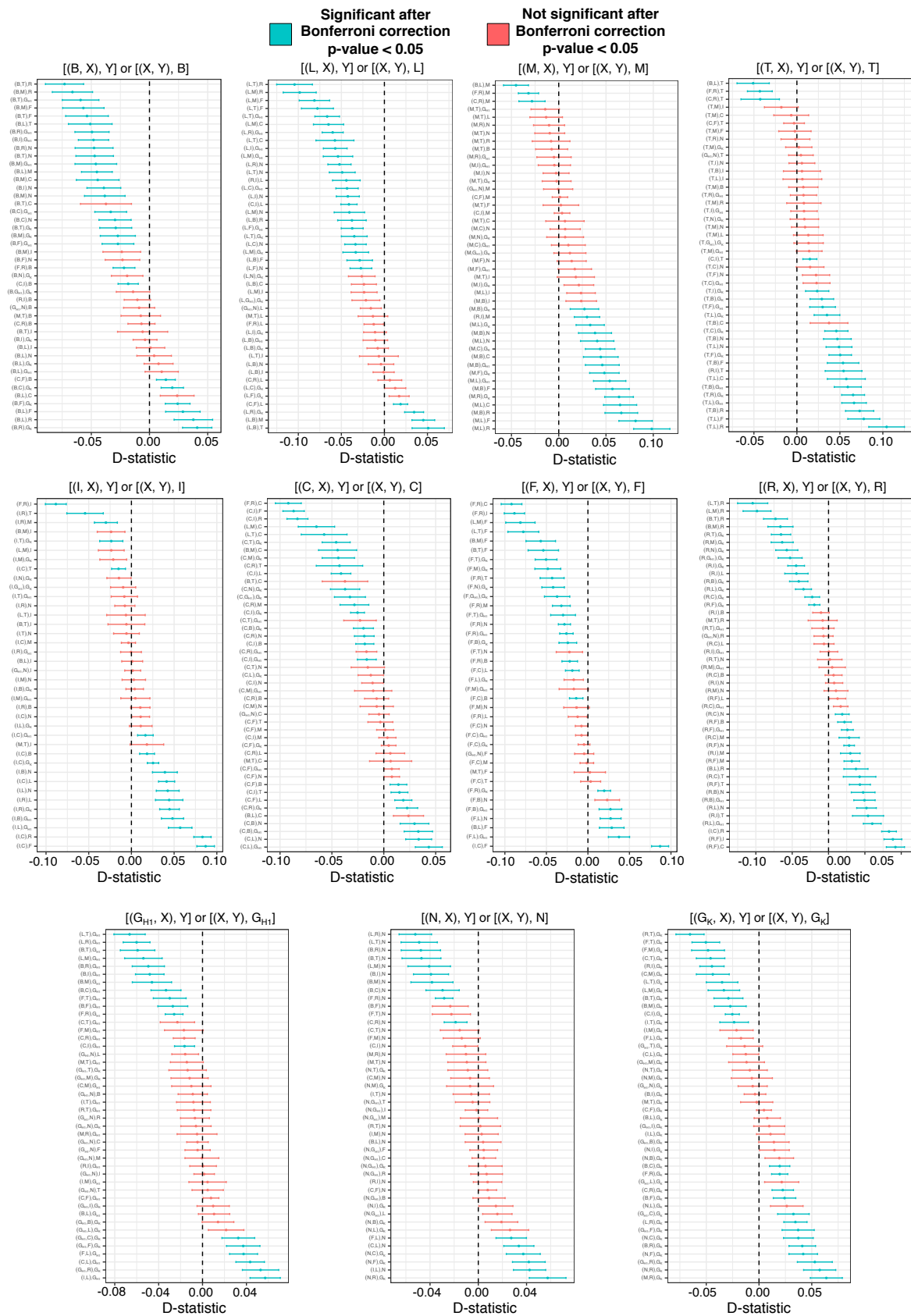
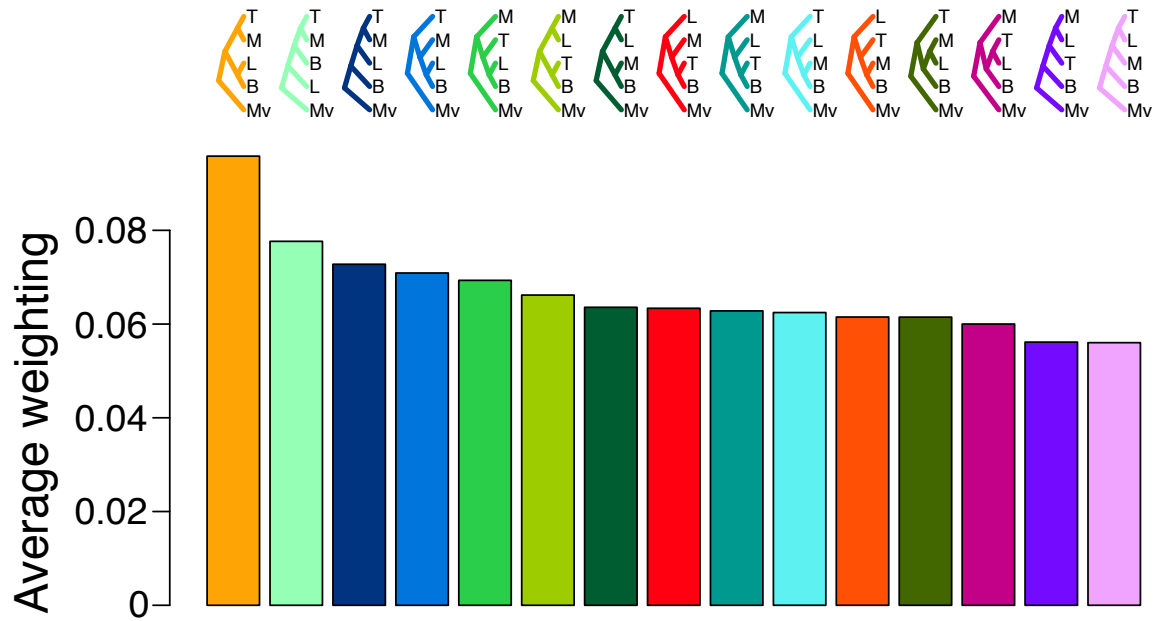


Fig. S4. ABBA-BABA D-test statistics for all taxon/population trio combinations.

Glabrous group topology



Pubescent group topology

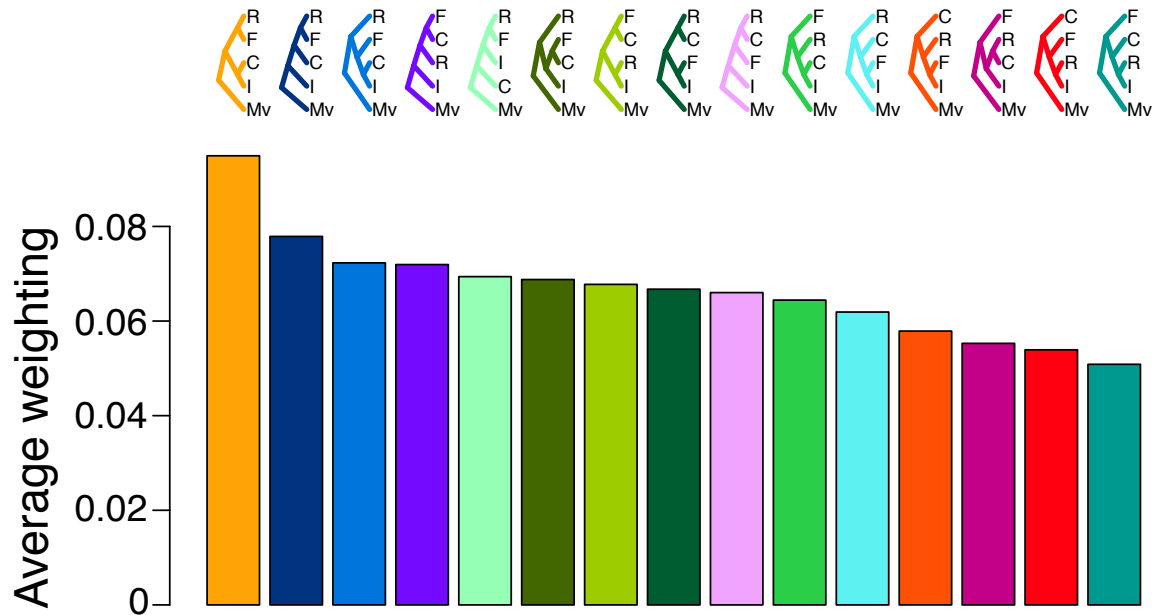


Fig. S5. TWISST-estimated topology weightings for the 15 possible topologies for a rooted 4-taxon tree. Shown are topology weights for glabrous (top) and pubescent (bottom) *Metrosideros* taxa on Oahu.

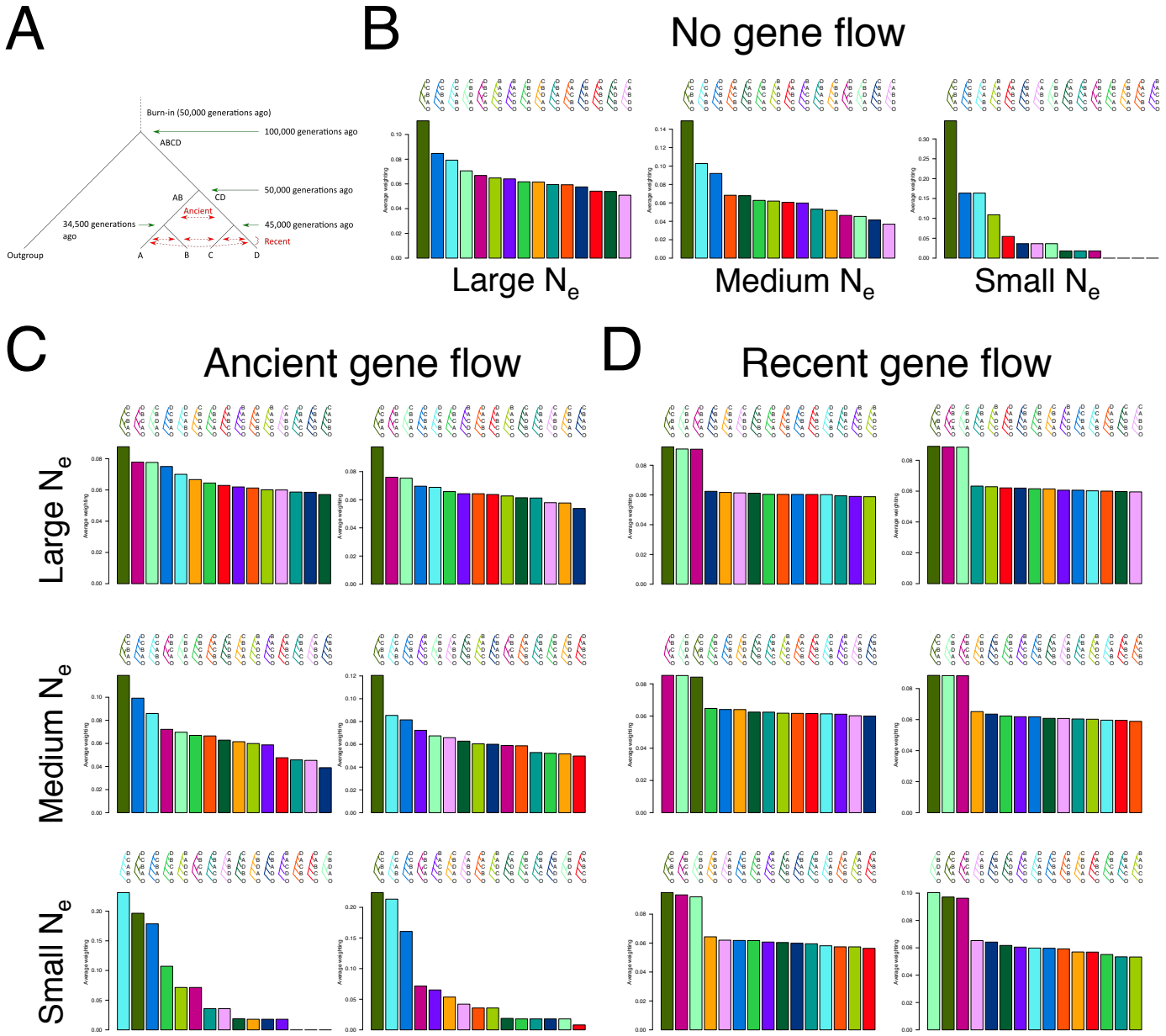


Fig. S6. Simulations examining the effects of population size (N_e) and gene flow on the topology weights. (A) The simulation scenario that is based on the history of the four glabrous taxa from Oahu. Topology weights for scenarios with no gene flow (B), ancient gene flow (C), and recent gene flow (D).

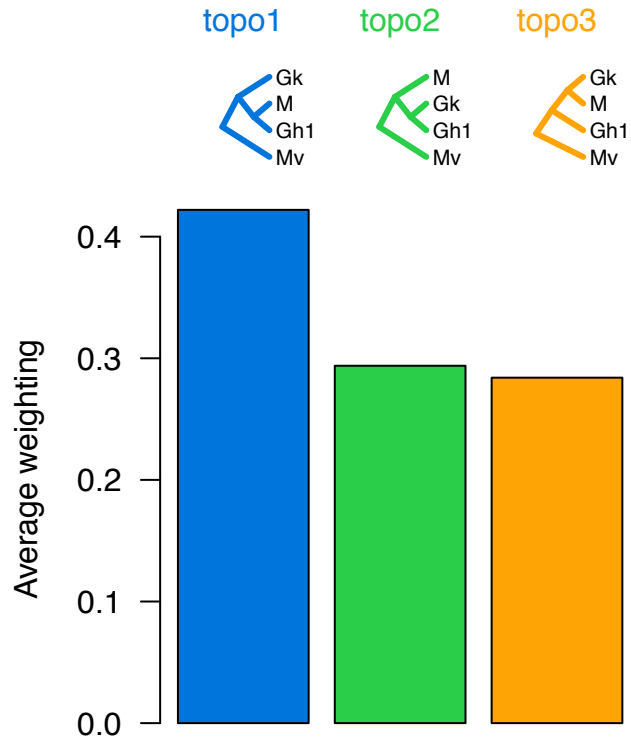


Fig. S7. TWISST topology weights involving taxa/populations G_{H1} , M , and G_K . Topologies were built using *M. vitiensis* (M_v) as the outgroup.

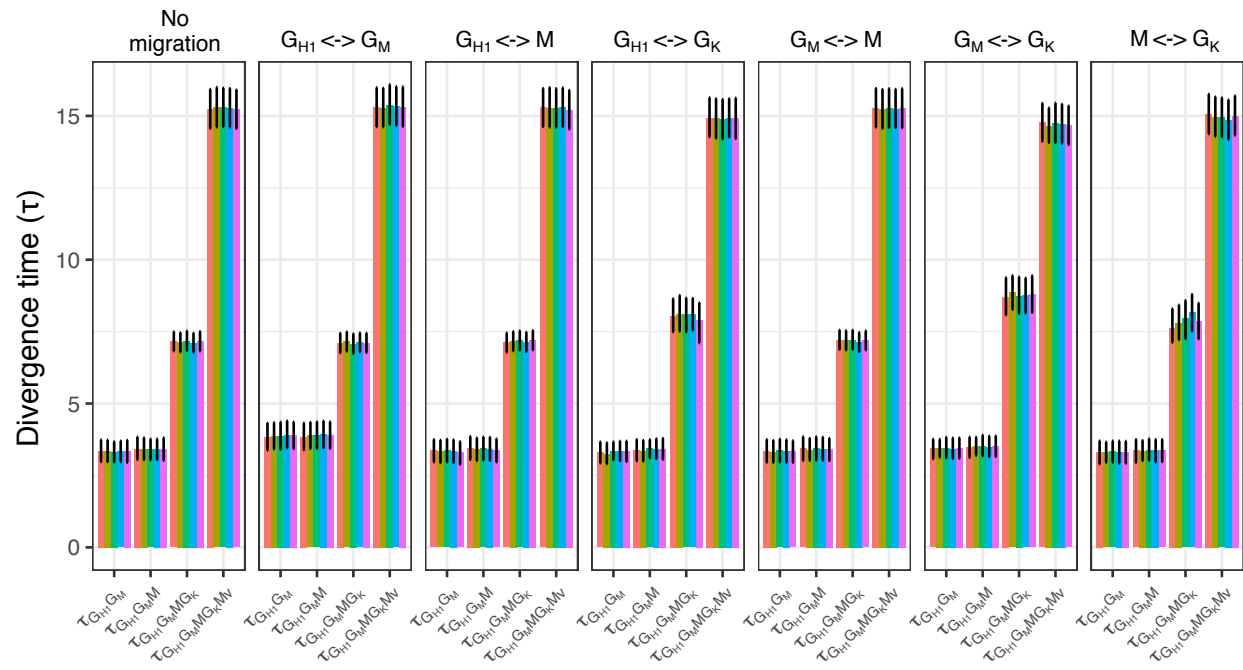


Fig. S8. G-PhoCS-based divergence time estimates for different migration models.

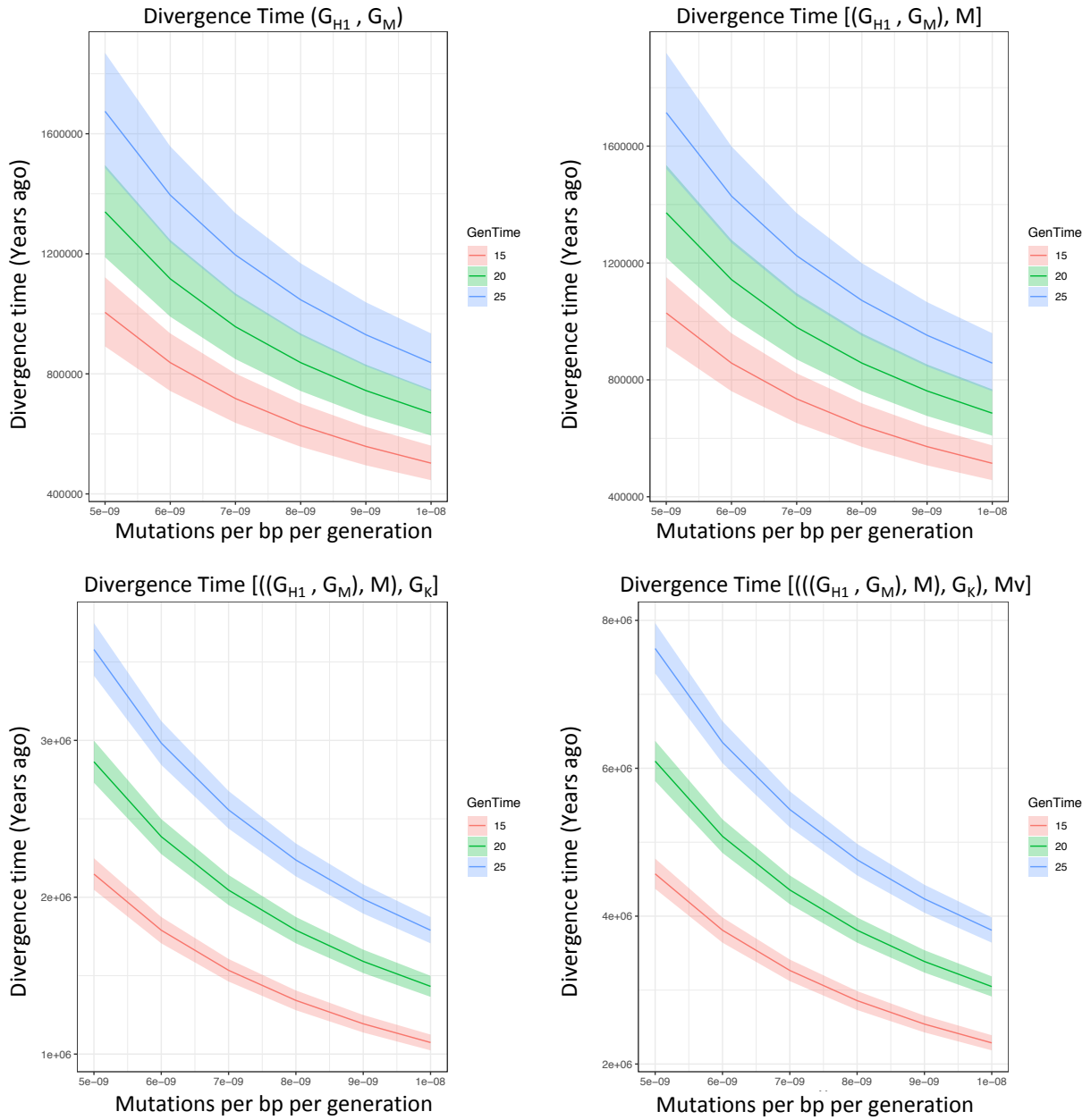


Fig. S9. G-PhoCS-based divergence time estimates assuming different mutation rates and generation times.

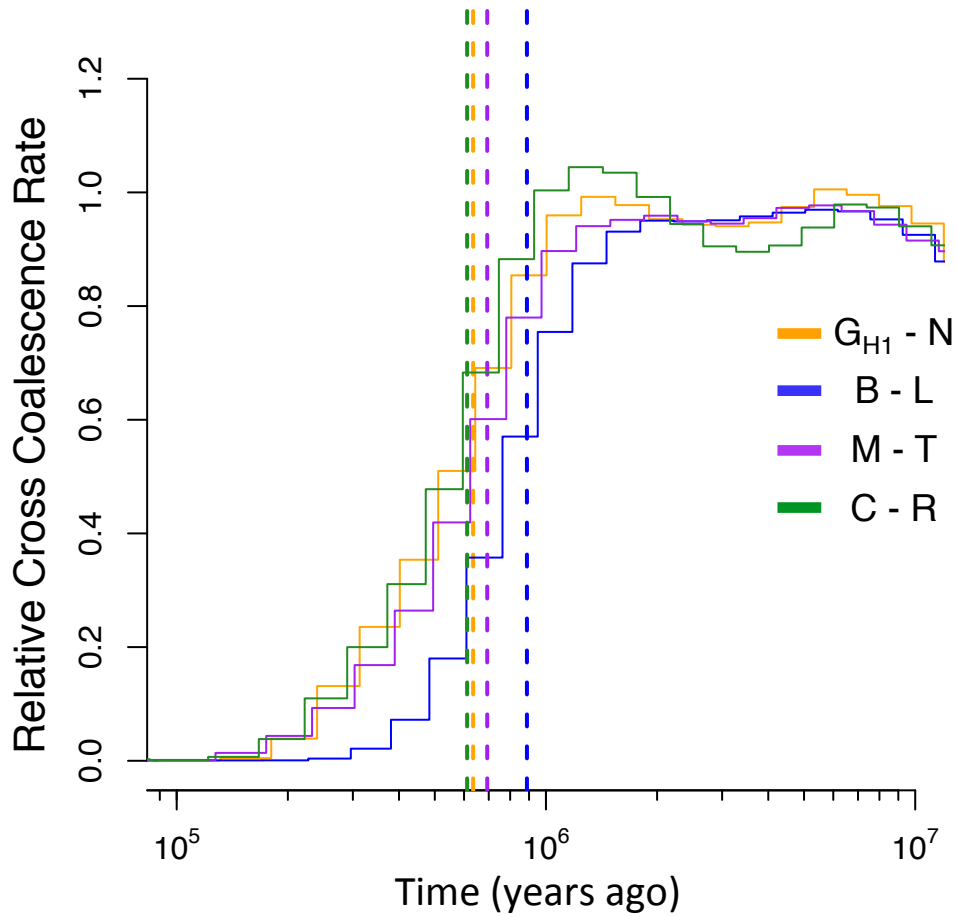


Fig. S10. The relative cross-coalescence rates for the four phylogenetic sister pairs. A cross-coalescence rate of 0.5 was assumed to reflect the divergence time between the paired taxa. Divergence times are indicated with dotted lines and shown in the key. Relative times were converted to absolute times assuming a mutation rate of $7e-9$ mutations per base pair per generation and a 20-year generation time.

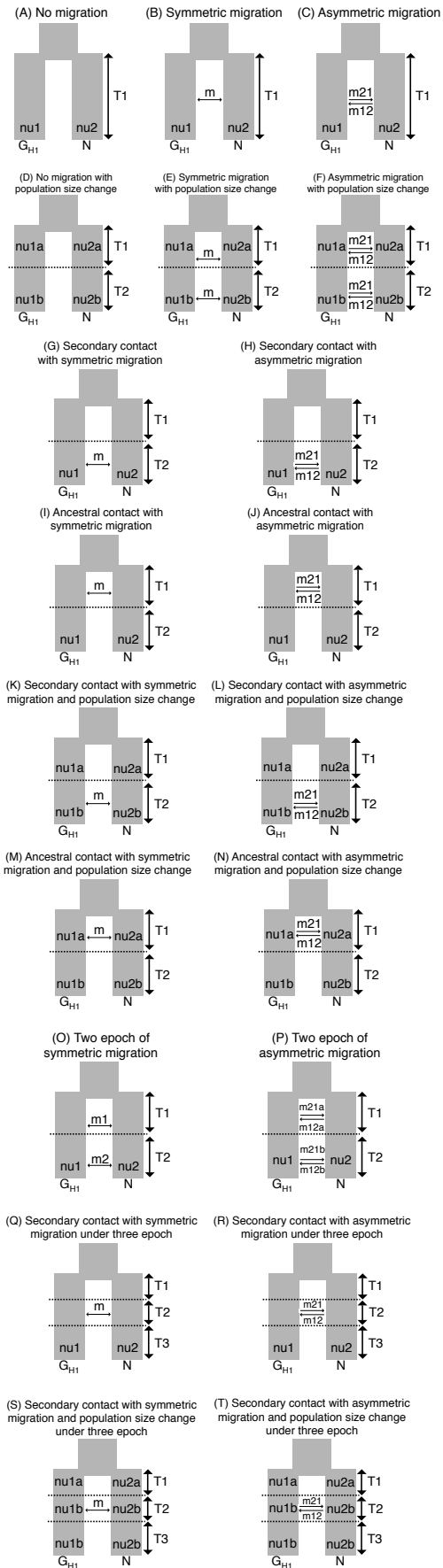


Fig. S11. The 20 different δadi scenarios that were modeled.

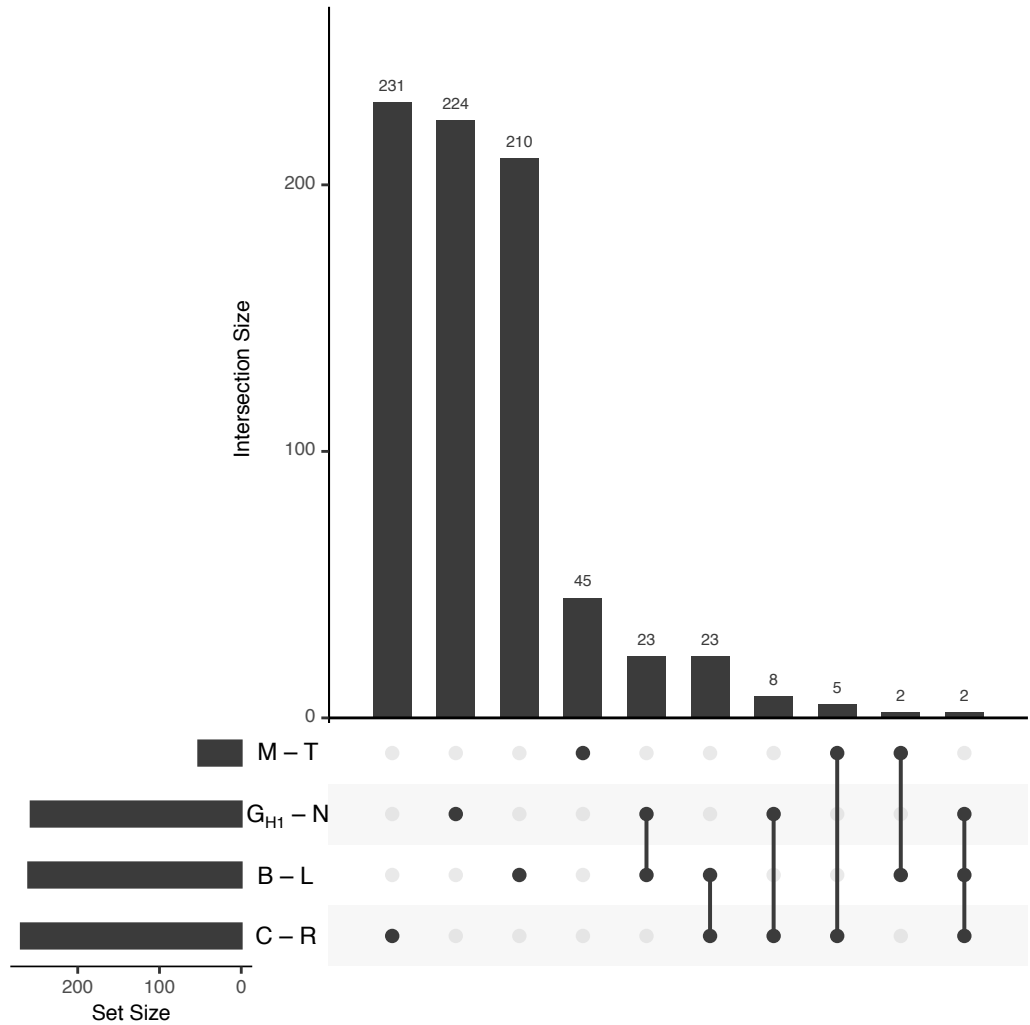


Fig. S12. The number of overlapping zFST outlier positions.

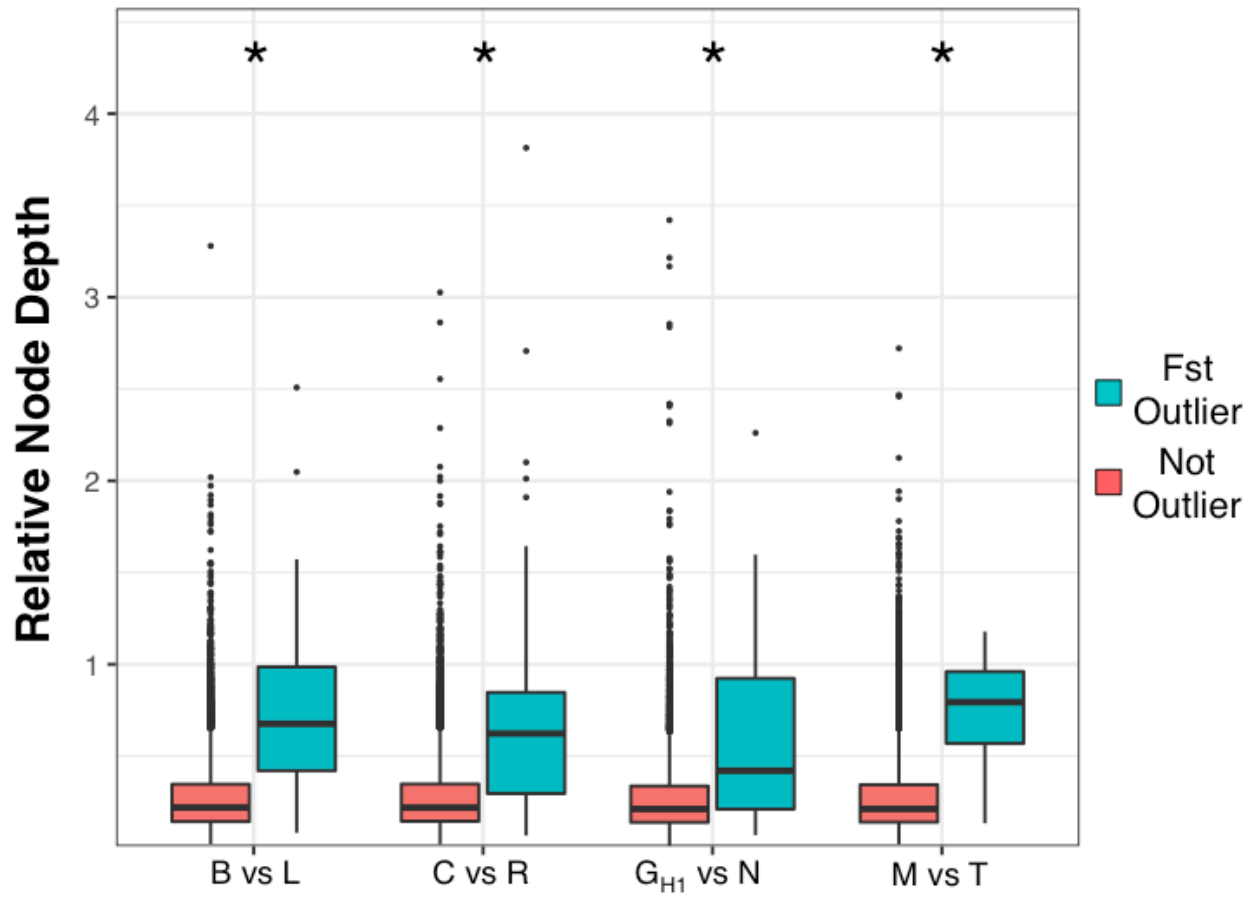


Fig. S13. Relative node depth (RND) values (calculated in 10-kbp window) for differentiation (F_{ST}) outliers identified for each sister pair. Red boxes are statistics from the genomic background, and green boxes are statistics from the differentiation outlier regions. * indicate significant difference with Mann-Whitney U test after Bonferroni correction; $p < 0.05$.

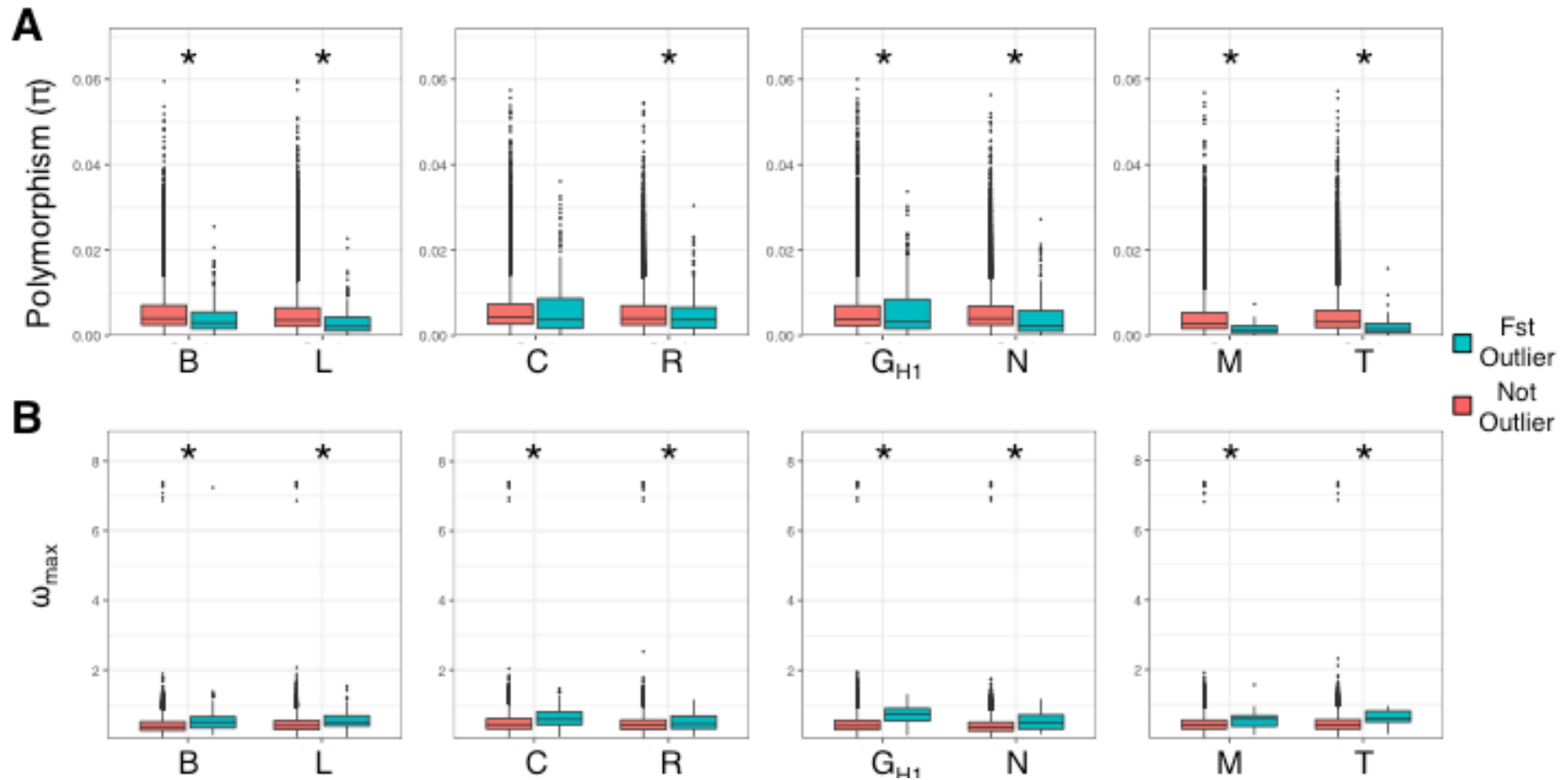


Fig. S14. Evidence of selection (calculated in 10-kbp window) in differentiation outlier regions. Red boxes are statistics from the genomic background, and green boxes are statistics from the differentiation outlier regions. (A) Levels of polymorphism in the differentiation outlier regions. (B) Selective sweep statistics (ω_{\max}) in the differentiation outlier regions. * indicate significant difference with Mann-Whitney U test after Bonferroni correction; $p < 0.05$.

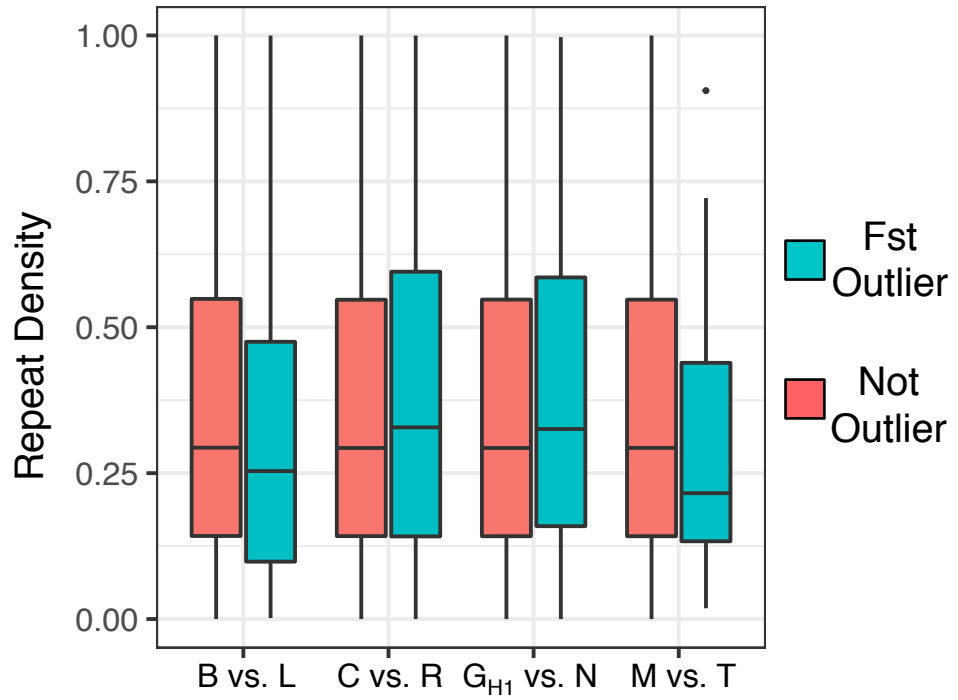


Fig. S15. Repeat sequence density in differentiation outlier regions compared to genome-wide 10-kbp windows. Red boxes are statistics from the genomic background, and green boxes are statistics from the differentiation outlier regions.

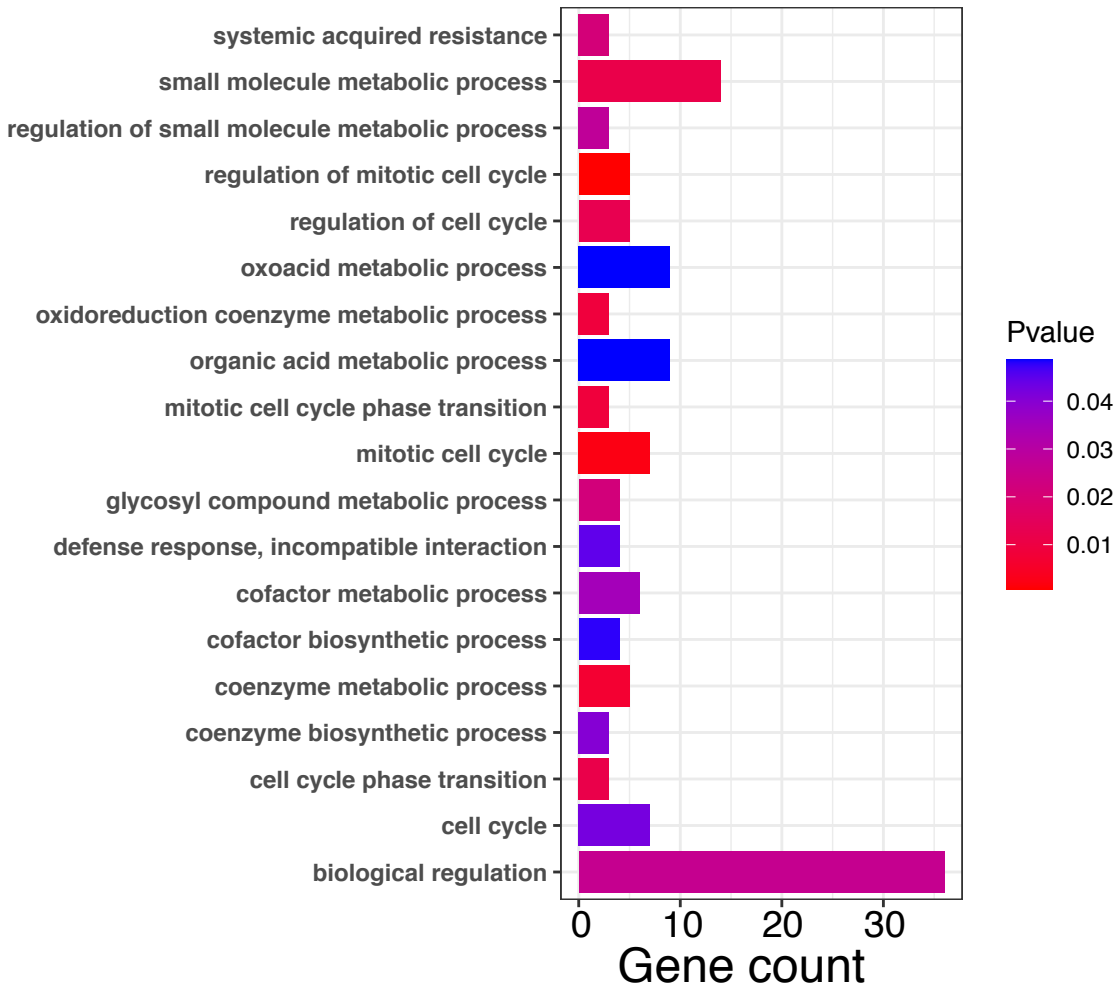


Fig. S16. Significantly enriched gene ontology terms for genes overlapping genomic regions of differentiation outliers.

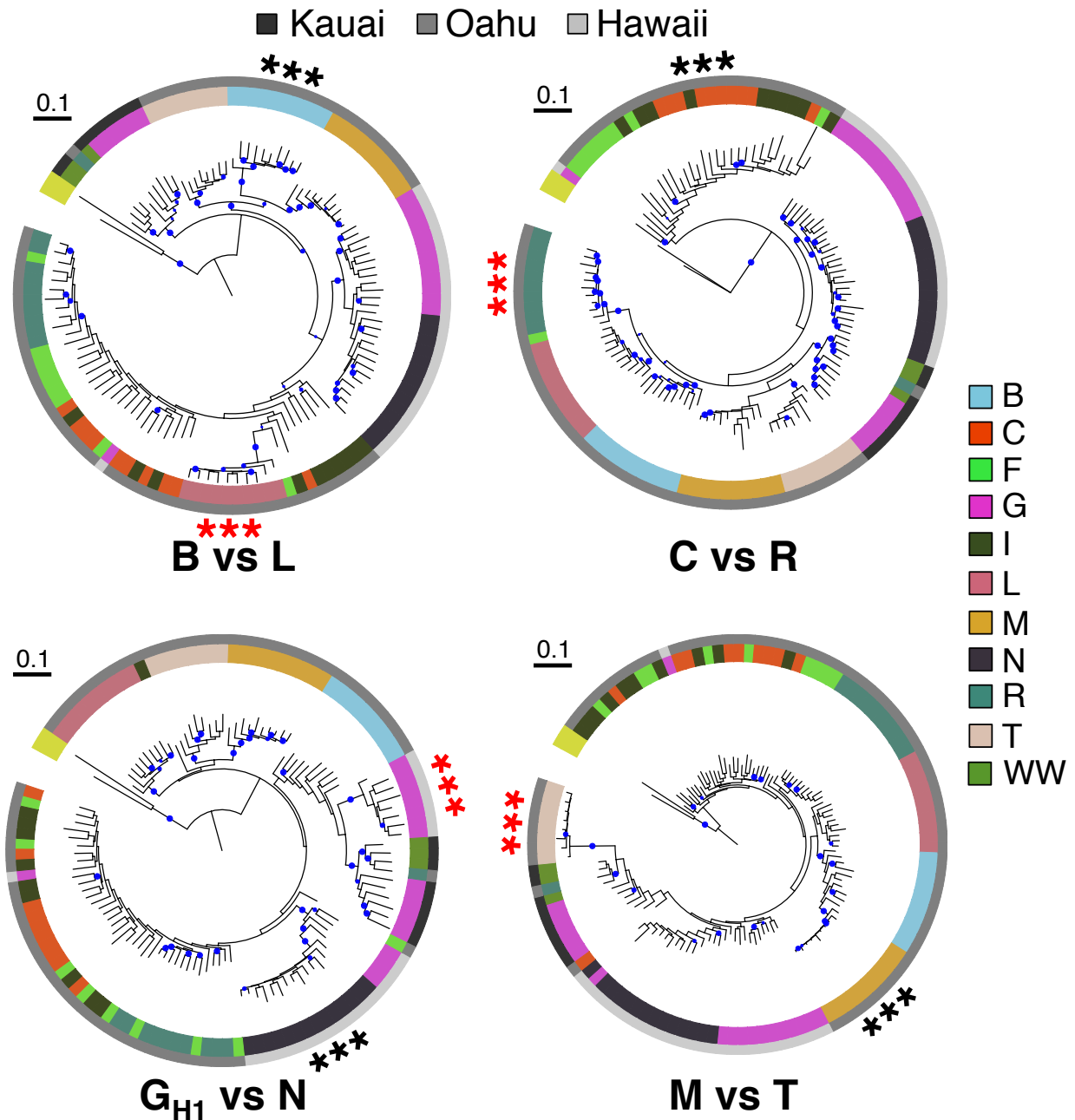


Fig. S17. Maximum-likelihood phylogenetic tree of the differentiation (F_{ST}) outlier regions identified for each sister pair in Fig. 3B. For each phylogeny the topological position of the sister pair that was examined is indicated with star and the topologically incongruent taxa from the pair is indicated with red star. Nodes with greater than 95% bootstrap support are indicated with blue circles.

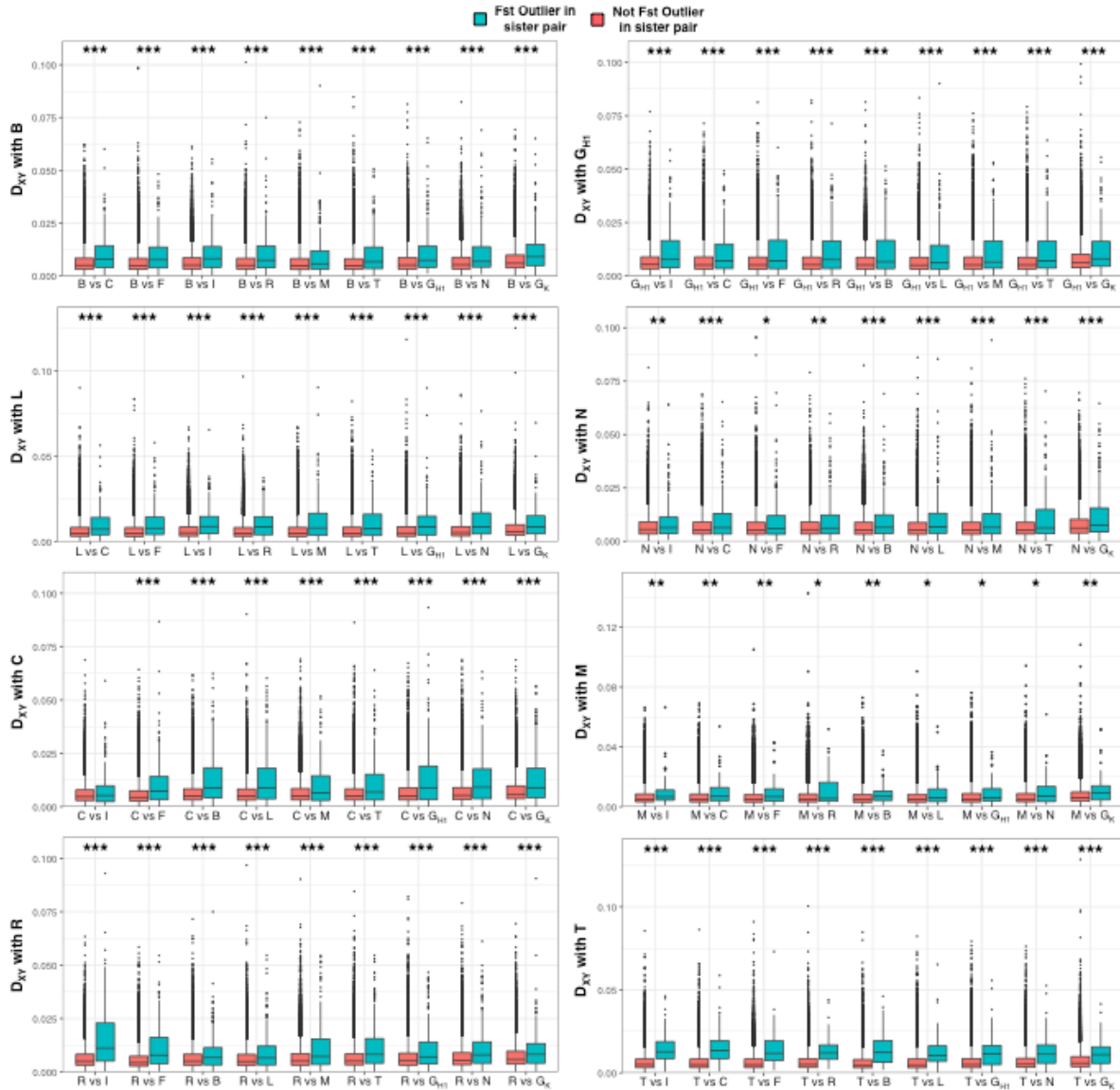


Fig. S18. Levels of D_{xy} between a taxon of the sister pair and all other Hawaiian *Metrosideros* taxa sampled. Red boxes are statistics from the genomic background, and green boxes are statistics from the differentiation outlier regions identified in a sister pair. Significant differences are shown as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$ following Mann Whitney U test with Bonferroni correction.

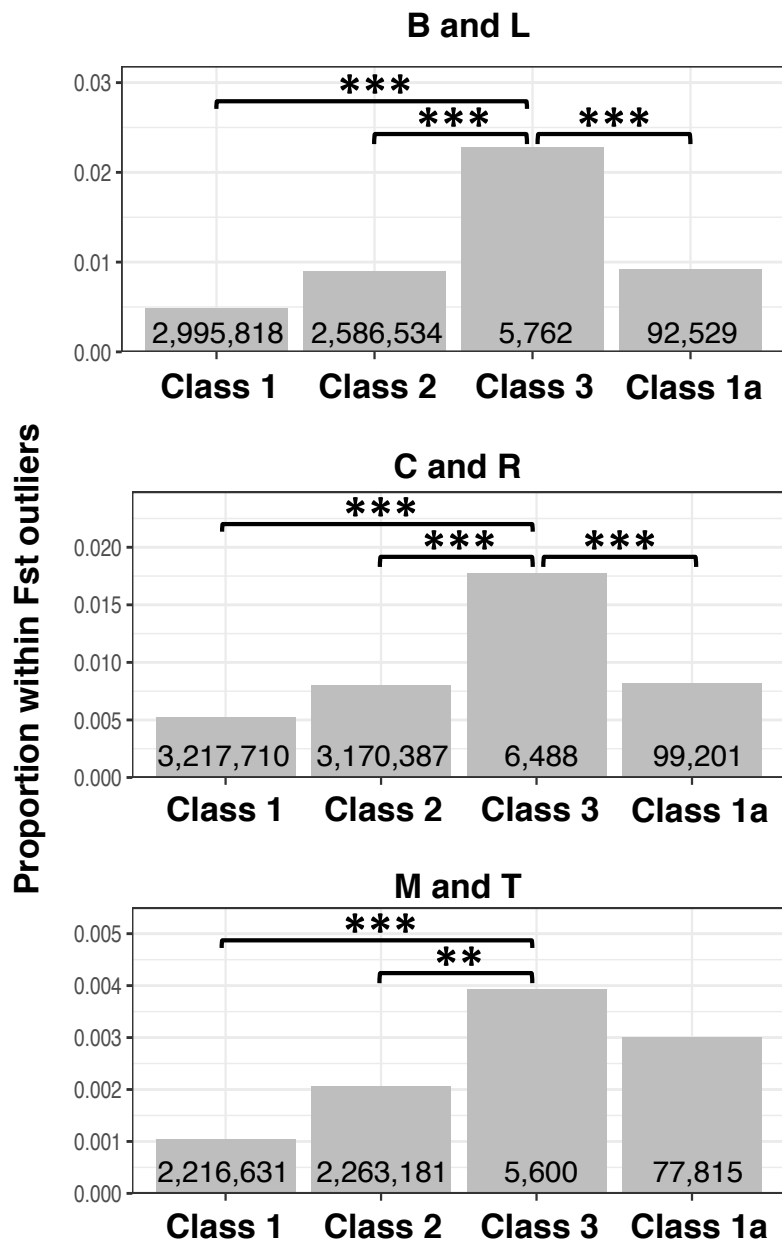


Fig. S19. Classification of sites polymorphic within a sister pair [B-L (top), C-R (middle), M-T (bottom)] into four categories according to the distribution of allele states on other islands (see text and Fig. 6 for visual representation of the four categories). Bars show the proportion within each category of sites that are polymorphic within the differentiation outlier regions of the focal sister pair. Numbers within the bars indicate the total number of SNPs within each category from the focal sister pair. ** indicate $p < 0.01$ and *** indicate $p < 0.001$ after Fisher's exact test.

Table S1. Sequencing statistics from the nanopore sequencing of the whole genome.

Flowcell	Number of reads	Median read length	Read length N50	Median quality score (QS)	Total bases
Raw					
FAL75935	1,821,430	2,929	13,392	11.1	11,244,606,349
FAL76650	1,909,759	2,895	13,081	11.1	11,574,820,609
FAN08536	483,618	22,785	29,641	11.3	11,013,147,889
Total	4,214,807	3,433	18,983	11.1	33,832,574,847
Filtered					
FAL75935	294,849	16,568	19,761	12.5	5,458,999,683
FAL76650	301,870	16,519	19,679	12.4	5,574,321,352
FAN08536	282,078	25,613	29,484	12.6	7,694,845,247
Total	878,797	19,185	23,559	12.5	18,728,166,282

Table S2. Aggregated Pore-C sequencing and number of contact results.

Reads				Read contact order						Pairwise Contacts	
Num of reads	Read bp	Num of reads > 1 contacts	Read bp > 1 contacts	<1	2	3	4	5	6>	Count	Contacts per Gbp sequenced
2,562,228	5,255,145,653	1,523,754	3,406,299,103	1,038,474	615,983	426,185	232,250	119,043	130,293	7,780,611	1,480,570

Table S3. Sequencing statistics from the nanopore sequencing of the cDNA library.

Sequencing statistics	
Number of reads	12,235,945
Total bases	12,251,510,900 bp
Median read length	847 bp
Read length N50	1,174 bp
Median quality score (QS)	11.1

Table S4. Population sample information.											
Sample ID	Code	TaxonName	Island	SourceLocation	Mean Coverage	Median Coverage	Percent Read Aligned	Sequenced By	Latitude	Longitude	Altitude
K177	G _k	M. polymorpha var. glaberrima	Kauai	Kuilau Trail, Kaimana, Kauai, HI	15.29	14	0.92	This_study			
K283	G _k	M. polymorpha var. glaberrima	Kauai	Iliau Loop Trail, Edge of Waimea Canyon, Kauai, HI	18.46	17	0.94	This_study			
K287	G _k	M. polymorpha var. glaberrima	Kauai	Iliau Loop Trail, Edge of Waimea Canyon, Kauai, HI	15.76	14	0.95	This_study			
K288	G _k	M. polymorpha var. glaberrima	Kauai	Iliau Loop Trail, Edge of Waimea Canyon, Kauai, HI	21	19	0.95	This_study			
K290	G _k	M. polymorpha var. glaberrima	Kauai	Iliau Loop Trail, Edge of Waimea Canyon, Kauai, HI	13.11	12	0.94	This_study			
K292	G _k	M. polymorpha var. glaberrima	Kauai	Iliau Loop Trail, Edge of Waimea Canyon, Kauai, HI	13.35	11	0.95	This_study			
K299	G _k	M. polymorpha var. glaberrima	Kauai	Iliau Loop Trail, Edge of Waimea Canyon, Kauai, HI	15.31	13	0.94	This_study			
K202	WW	M. waialealae var. waialealae	Kauai	Kahili Ridge, Kauai, HI	7.2	5	0.91	This_study			
K205	WW	M. waialealae var. waialealae	Kauai	Kahili Ridge, Kauai, HI	8.21	7	0.92	This_study			
K207	WW	M. waialealae var. waialealae	Kauai	Kahili Ridge, Kauai, HI	1.6	2	0.91	This_study			
K214	WW	M. waialealae var. waialealae	Kauai	Kahili Ridge, Kauai, HI	14.54	12	0.96	This_study			
O296	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	12.42	9	0.95	This_study	21.35	-157.79	953.00
O297	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	9.48	6	0.83	This_study	21.35	-157.79	953.00
O301	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	16.5	14	0.92	This_study	21.35	-157.79	940.00
O303	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	17.07	15	0.93	This_study	21.35	-157.79	924.00
O304	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	17.68	14	0.91	This_study	21.35	-157.79	929.00
O305	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	16.73	14	0.95	This_study	21.35	-157.79	902.00
O306	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	13.35	11	0.93	This_study	21.35	-157.79	896.00
O308	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	11.65	10	0.95	This_study	21.35	-157.79	882.00
O309	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	16.97	13	0.93	This_study	21.35	-157.79	877.00
O310	B	M. polymorpha race B	Oahu	Konahuanui, Oahu, HI	17.79	15	0.94	This_study	21.35	-157.79	871.00
O45	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	23.12	22	0.93	This_study	21.42	-157.86	598.00
O55	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	20.98	20	0.90	This_study	21.41	-157.87	534.00
O60	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	20.81	20	0.93	This_study	21.41	-157.87	499.00
O63	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	19.83	19	0.92	This_study	21.41	-157.87	536.00
O64	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	31.59	30	0.92	This_study	21.41	-157.87	536.00
O65	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	20.22	19	0.92	This_study	21.41	-157.87	537.00
O67	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	19.62	19	0.94	This_study	21.41	-157.87	547.00
O68	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	18.5	18	0.93	This_study	21.41	-157.87	537.00
O70	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	19.06	18	0.93	This_study	21.41	-157.87	527.00
O71	C	M. polymorpha race C	Oahu	'Aiea, Oahu, HI	21.72	20	0.92	This_study	21.41	-157.87	527.00
O119	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	9.95	9	0.90	This_study	21.35	-157.80	683.00
O122	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	11.04	10	0.90	This_study	21.35	-157.80	677.00
O128	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	8.55	8	0.89	This_study	21.35	-157.80	656.00
O131	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	14.01	11	0.89	This_study	21.35	-157.80	651.00
O138	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	14.8	14	0.94	This_study	21.35	-157.80	658.00
O141	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	12.96	12	0.94	This_study	21.35	-157.80	662.00
O142	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	13.01	12	0.92	This_study	21.35	-157.80	657.00
O149	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	15.14	13	0.95	This_study	21.35	-157.80	648.00
O168	F	M. polymorpha race F	Oahu	Konahuanui, Oahu, HI	14.97	13	0.94	This_study	21.35	-157.80	603.00
O72	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	20.25	19	0.94	This_study	21.41	-157.87	530.00
O75	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	20.82	20	0.90	This_study	21.41	-157.88	527.00
O76	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	19.39	18	0.93	This_study	21.41	-157.88	476.00
O77	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	20.15	20	0.93	This_study	21.41	-157.88	492.00
O78	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	21.55	20	0.93	This_study	21.41	-157.87	498.00
O88	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	20.72	17	0.92	This_study	21.41	-157.87	533.00
O89	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	24.98	22	0.93	This_study	21.41	-157.87	531.00
O90	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	17.75	16	0.90	This_study	21.41	-157.87	521.00
O91	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	11.48	10	0.85	This_study	21.41	-157.87	517.00
O92	I	M. polymorpha var. incana	Oahu	'Aiea, Oahu, HI	19.14	16	0.91	This_study	21.41	-157.87	517.00
O177	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	21.62	18	0.92	This_study	21.51	-158.14	1230.00
O180	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	13.63	11	0.93	This_study	21.51	-158.15	1224.00
O183	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	18.4	16	0.94	This_study	21.50	-158.15	1224.00
O189	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	13.96	11	0.93	This_study	21.50	-158.15	1214.00
O191	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	16.34	13	0.95	This_study	21.50	-158.15	1211.00
O194	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	19.09	17	0.92	This_study	21.50	-158.15	1200.00
O196	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	17.15	15	0.93	This_study	21.50	-158.15	1191.00
O199	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	14.17	12	0.95	This_study	21.50	-158.15	1202.00
O200	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	15.44	11	0.91	This_study	21.50	-158.15	1197.00
O355	L	M. polymorpha race L	Oahu	Mt. Ka'ala, Oahu, HI	12.74	11	0.86	This_study	21.50	-158.15	1203.00
NG154	M	M. macropus	Oahu	Lanihuli, Oahu, HI	14.45	12	0.95	This_study	21.35	-157.80	658.00
O135	M	M. macropus	Oahu	Konahuanui, Oahu, HI	12.65	10	0.96	This_study			
O344	M	M. macropus	Oahu	Konahuanui, Oahu, HI	14.57	12	0.85	This_study			
O367	M	M. macropus	Oahu	Konahuanui, Oahu, HI	16.59	14	0.97	This_study			
O368	M	M. macropus	Oahu	Konahuanui, Oahu, HI	12.75	9	0.96	This_study	21.35	-157.80	579.00
O369	M	M. macropus	Oahu	Konahuanui, Oahu, HI	17.3	13	0.95	This_study	21.35	-157.80	659.00
O370	M	M. macropus	Oahu	Konahuanui, Oahu, HI	13.03	11	0.95	This_study	21.35	-157.80	657.00
O371	M	M. macropus	Oahu	Konahuanui, Oahu, HI	8.85	7	0.96	This_study	21.35	-157.80	659.00
O373	M	M. macropus	Oahu	Konahuanui, Oahu, HI	8.43	6	0.96	This_study	21.35	-157.80	663.00
O385	M	M. macropus	Oahu	Konahuanui, Oahu, HI	17.14	14	0.94	This_study	21.35	-157.80	657.00
O462	R	M. rugosa	Oahu	Waikane, Oahu, HI	10.12	9	0.89	This_study			
O463	R	M. rugosa	Oahu	Waikane, Oahu, HI	11.9	10	0.86	This_study			
O464	R	M. rugosa	Oahu	Waikane, Oahu, HI	17.81	15	0.86	This_study			
O465	R	M. rugosa	Oahu	Waikane, Oahu, HI	10.52	9	0.77	This_study			
O466	R	M. rugosa	Oahu	Waikane, Oahu, HI	13.13	12	0.86	This_study			
O467	R	M. rugosa	Oahu	Waikane, Oahu, HI	15.48	13	0.87	This_study			
O468	R	M. rugosa	Oahu	Waikane, Oahu, HI	14.57	12	0.84	This_study			
O469	R	M. rugosa	Oahu	Waikane, Oahu, HI	26.93	24	0.91	This_study			
O470	R	M. rugosa	Oahu	Waikane, Oahu, HI	10.22	9	0.85	This_study			
O471	R	M. rugosa	Oahu	Waikane, Oahu, HI	9.52	8	0.82	This_study			
M. rugosa	R	M. rugosa	Oahu		24.75	22	0.95	Choi_et al	-	-	
M. tremuloid	T	M. tremuloides	Oahu		19.79	18	0.91	Choi_et al	-	-	
O143	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	12.89	12	0.91	This_study	21.35	-157.80	660.00
O145	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	19.17	17	0.95	This_study	21.35	-157.80	652.00
O146	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	16.62	14	0.95	This_study	21.35	-157.80	652.00
O148	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	14.31	13	0.93	This_study	21.35	-157.80	650.00
O151	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	13.75	12	0.94	This_study	21.35	-157.80	650.00
O154	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	2.45	3	0.94	This_study			
O157	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	18.1	16	0.94	This_study	21.35	-157.80	635.00
O166	T	M. tremuloides	Oahu	Konahuanui, Oahu, HI	14.81	11	0.94	This_study	21.35	-157.80	604.00
X36	G _M	M. polymorpha var. glaberrima	Molokai		24.63	23	0.95	Choi_et al	21.14	-156.93	
X83	G _M	M. polymorpha var. glaberrima	Molokai		13	12	0.95	Choi_et al	21.12	-156.92	
NG207	-	M. polymorpha var. glaberrima	Hawaii	Saddle Rd, Hawaii Island, HI	3.64	3	0.92	This_study			
E106	N	M. polymorpha var. newellii	Hawaii		20.38	19	0.95	Choi_et al	19.72	-155.19	
E110	N	M. polymorpha var. newellii	Hawaii		18.85	18	0.93	Choi_et al	19.72	-155.19	
E111	G _{H2}	M. polymorpha var. glaberrima	Hawaii		21.67	19	0.96	Choi_et al	19.72	-155.19	
E114	G _{H2}	M. polymorpha var. glaberrima	Hawaii		5.89	5	0.96	Choi_et al	19.72	-155.19	
E126	G _{H2}	M. polymorpha var. glaberrima	Hawaii		27.09	23	0.97	Choi_et al	19.72	-155.19	
E132	G _{H2}	M. polymorpha var. glaberrima	Hawaii		14.22	11	0.94	Choi_et al	19.71	-155.19	
E70	-	M. polymorpha var. glaberrima	Hawaii		18.96	15	0.95	Choi_et al	19.71	-155.24	
E72	-	M. polymorpha var. glaberrima	Hawaii		17.83	15	0.96	Choi_et al	19.71	-155.24	
E74	-	M. polymorpha var. glaberrima	Hawaii		35.43	36	0.95	Choi_et al	19.71	-155.24	
H193	G _{H1}	M. polymorpha var. glaberrima	Hawaii		25.85	24	0.91	Choi_et al	20.07	-155.67	
H195	-	M. polymorpha var. glaberrima	Hawaii		17.32	15	0.90	Choi_et al	20.07	-155.67	
H196	-	M. polymorpha var. glaberrima	Hawaii		0.04	1	0.96	Choi_et al	20.07	-155.67	
H197	G _{H1}	M. polymorpha var. glaberrima	Hawaii		46.72	43	0.96	Choi_et al	20.07	-155.67	
H198	G _{H1}	M. polymorpha var. glaberrima	Hawaii		22.92	21	0.93	Choi_et al	20.07	-155.67	
H202	G _{H1}	M. polymorpha var. glaberrima	Hawaii		22.7	21	0.94	Choi_et al	20.07	-155.67	
H204	G _{H1}	M. polymorpha var. glaberrima	Hawaii		10.87	9	0.95	Choi_et al	20.07	-155.67	
H206	G _{H1}	M. polymorpha var. glaberrima	Hawaii		11.25	9	0.92	Choi_et al	20.07	-155.67	
H207	G _{H1}	M. polymorpha var. glaberrima	Hawaii		18.87	16	0.96	Choi_et al	20.07	-155.67	
H212	G _{H1}	M. polymorpha var. glaberrima	Hawaii		15.26	14	0.93	Choi_et al	20.07	-155.67	
H235	-	M. polymorpha var. newellii	Hawaii		23.2	21	0.96	Choi_et al	19.68	-155.15	
H261	-	M. polymorpha var. newellii	Hawaii		8.82	7	0.96	Choi_et al	19.71	-155.23	
H269	N	M. polymorpha var. newellii	Hawaii		20.08	19	0.96	Choi_et al			

Table S5. ω results for all 20 demography models for all 4 sister pairs.	Model	log-likelihood(AIC)	theta	nu1	nu2	nu3a	nu2a	nu1b	nu2b	m12	m21	m12a	m21a	m12b	m21b	T1	T2	T3	
B vs L																			
(K) secondary contact with symmetric migration and population size change	-1322.58	2659.36	3451.83	-	-	-	0.7786	0.5566	1.9384	1.9225	0.514	-	-	-	-	0.2281	0.2918	-	
(O) two epoch of symmetric migration	-1357.08	2726.16	3281.74	-	1.9189	1.3714	-	-	311.36	-	-	-	-	-	-	0.4732	0.3446	-	
(G) secondary contact with asymmetric migration	-1376.12	2764.24	2947.57	-	2.0169	1.5855	-	-	0.4709	0.5356	0.0759	-	-	-	0.7318	-	-	-	
(E) symmetric migration with population size change	-1385.23	2784.46	3325.77	-	-	-	1.4578	0.5669	1.9474	1.3091	0.5088	-	-	-	-	-	0.2638	0.354	-
(B) symmetric migration	-1417.07	2842.14	2959.52	-	2.0664	1.4611	-	-	-	-	0.4322	-	-	-	-	-	0.7063	-	-
(G) secondary contact with symmetric migration	-1418.45	2846.49	2909.33	-	1.9254	1.3774	-	-	-	-	0.5568	-	-	-	-	-	0.2994	0.534	-
(R) secondary contact with asymmetric migration under three epoch (isolation - migration - isolation)	-1419.6	2853.2	2784.57	-	2.043	1.7189	-	-	-	-	2.3693	-	-	-	-	-	0.5645	0.1166	3.5671
(F) asymmetric migration and population size change	-1422.79	2861.58	2451.21	-	-	-	1.2181	0.2806	2.2552	2.3538	0.5323	0.2851	-	-	-	-	0.1309	0.9174	-
(S) secondary contact with symmetric migration population size change under three epoch (isolation - migration - isolation)	-1435.59	2887.18	2483.93	-	-	-	19.4443	0.3535	2.1869	1.7371	0.4786	-	-	-	-	-	0.123	1.0174	0.0116
(C) asymmetric migration	-1440.41	2890.82	2811.63	-	2.1431	1.6868	-	-	-	-	0.3678	0.3945	-	-	-	-	0.7474	-	-
(T) secondary contact with asymmetric migration population size change under three epoch (isolation - migration - isolation)	-1601.04	3220.08	3212.22	-	-	-	1.8862	0.3564	2.9301	1.6835	0.2513	3.147	-	-	-	-	0.1902	0.146	0.145
(I) ancestral symmetric migration	-1665.25	3340.5	2632.89	-	2.0042	1.6651	-	-	-	-	0.9616	-	-	-	-	-	0.879	0.131	-
(P) two epoch of asymmetric migration	-1668.47	3352.94	1177.11	-	4.3647	3.0194	-	-	-	-	3.4212	-	-	-	-	-	4.2787	1.869	-
(J) ancestral asymmetric migration	-1683.79	3379.58	2973	-	-	-	-	-	-	-	0.9233	1.049	-	-	5.5715	0.464	0.2377	0.9208	0.0629
(N) ancestral asymmetric migration with population size change	-1863.83	3743.66	1653.34	-	-	-	1.8043	1.3067	5.0066	3.2053	1.5766	3.656	-	-	-	-	3.3099	0.5117	-
(M) ancestral symmetric migration with population size change	-1922.59	3859.18	4007.25	-	-	-	0.6065	0.4997	4.8627	3.1101	0.3963	-	-	-	-	-	0.1158	0.131	-
(D) no migration with population size change	-1935.76	3883.52	3959.14	-	-	-	0.6841	0.5264	2.7482	2.3279	-	-	-	-	-	-	0.088	0.1584	-
(Q) secondary contact with symmetric migration under three epoch (isolation - migration - isolation)	-1936.34	3884.68	2730.04	-	1.9826	1.1116	-	-	-	-	4.4816	-	-	-	-	-	0.9534	0.1019	0.1821
(A) no migration	-1995.35	3996.7	3715.68	-	1.9358	1.3804	-	-	-	-	-	-	-	-	-	-	0.3143	-	-
(L) secondary contact with asymmetric migration and population size change	-2270.84	4557.68	3843.12	-	-	-	0.3238	11.7112	1.8081	0.4678	0.2155	2.9631	-	-	-	-	1.0995	0.8536	-
G _{III} vs N																			
(O) two epoch of symmetric migration	-985.13	1982.26	3106.12	-	1.5946	1.3956	-	-	-	-	-	-	0.5827	-	-	-	0.3535	0.276	-
(R) secondary contact with asymmetric migration under three epoch (isolation - migration - isolation)	-988.35	1990.7	3145.91	-	2.045	1.1846	-	-	-	-	2.41	7.3335	-	-	-	-	0.4663	0.0422	5.4051
(G) secondary contact with symmetric migration	-1005.59	2021.75	16754	-	1.5555	-	-	-	-	0.848	-	-	-	-	-	-	0.0343	0.6724	-
(C) asymmetric migration	-1016.71	2043.82	2900.22	-	1.78	1.5327	-	-	-	-	0.6218	0.9316	-	-	-	-	0.6906	-	-
(K) secondary contact with symmetric migration and population size change	-1052.91	2119.42	2945.76	-	-	-	1.3458	0.3478	1.916	1.4179	0.9676	-	-	-	-	-	0.2502	0.5832	-
(H) secondary contact with asymmetric migration	-1058.24	2128.48	3077.18	-	1.2441	1.5301	-	-	-	-	1.581	0.7094	-	-	-	-	0.3014	0.4243	-
(E) symmetric migration with population size change	-1062.58	2139.16	1962.7	-	-	0.0579	3.0444	2.4837	1.9242	0.6876	-	-	-	-	-	-	1.3156	1.5752	-
(L) secondary contact with asymmetric migration and population size change	-1120.55	2257.1	1707.25	-	-	0.3682	0.2582	3.2357	1.7831	0.379	0.9169	-	-	-	-	-	1.7921	1.8851	-
(F) asymmetric migration and population size change	-1204.8	2425.6	7498.76	-	-	0.4041	0.066	0.9611	0.4363	0.249	6.0537	-	-	-	-	-	5.1216	0.1719	-
(T) secondary contact with asymmetric migration population size change under three epoch (isolation - migration - isolation)	-1205.72	2429.44	2093.11	-	-	0.0928	0.7276	2.0026	2.2502	1.2399	0.4926	-	-	-	-	-	0.452	1.4495	0.034
(M) ancestral symmetric migration with population size change	-1211.01	2436.02	3211.78	-	-	1.216	1.7234	1.5029	1.1368	1.488	-	-	-	-	-	-	0.5262	0.0634	-
(Q) secondary contact with symmetric migration under three epoch (isolation - migration - isolation)	-1266.08	2544.16	2205.43	-	2.6816	1.6849	-	-	-	-	3.8161	-	-	-	-	-	1.0274	0.128	0.1328
(I) ancestral symmetric migration	-1304.92	2619.84	2999.38	-	1.5383	1.4232	-	-	-	-	26.1851	-	-	-	-	-	5.5518	0.1697	-
(P) two epoch of asymmetric migration	-1320.69	2657.38	1942.64	-	1.2401	1.9424	-	-	-	-	23.047	-	-	-	-	-	0.8375	0.3446	-
(J) ancestral asymmetric migration	-1323.57	2659.14	2318.22	-	2.2324	1.4854	-	-	-	-	0.3962	4.1008	-	-	5.564	0.0645	2.1746	0.1079	0.6375
(S) secondary contact with symmetric migration population size change under three epoch (isolation - migration - isolation)	-1351.4	2718.8	672.5	-	-	2.9562	0.4088	5.4202	5.8626	0.4472	-	-	-	-	-	-	5.0116	5.1573	0.251
(B) symmetric migration	-1376.15	2760.3	1516.92	-	1.9357	2.6375	-	-	-	-	0.5183	-	-	-	-	-	2.651	-	-
(A) no migration	-1526.21	3058.42	3988.4	-	1.6743	1.3762	-	-	-	-	-	-	-	-	-	-	0.2008	-	-
(N) ancestral asymmetric migration with population size change	-1529.26	3074.52	3878.77	-	-	7.85	26.963	1.486	1.2973	0.4348	0.9963	-	-	-	-	-	0.0484	0.1748	-
(D) no migration with population size change	-1550.67	3113.34	3975.28	-	-	22.6936	1.3213	1.0187	1.3429	-	-	-	-	-	-	-	0.113	0.0963	-
C vs I																			
(O) two epoch of symmetric migration	-1091.03	2194.06	3024.92	-	0.8638	5.8947	-	-	-	-	-	-	0.2736	-	-	-	15.2904	-	-
(G) secondary contact with symmetric migration	-1119.13	2248.26	2915.47	-	1.2173	4.5389	-	-	-	-	11.4593	-	-	-	-	-	0.5115	0.0592	-
(E) symmetric migration with population size change	-1156.87	2327.74	8174.78	-	3.2848	6.4283	-	0.0859	1.8727	0.5935	1.7333	18.0019	-	-	-	-	0.5176	0.0729	-
(C) asymmetric migration	-1235.79	2487.58	2902.19	-	3.2848	4.6283	-	-	-	-	-	-	0.5042	1.004	1.4067	9.4359	0.3944	0.2555	-
(K) secondary contact with symmetric migration and population size change	-1253.84	2517.68	3025.92	-	3.1179	3.9395	-	-	-	-	0.2878	8.4204	-	-	-	-	0.5422	-	-
(H) secondary contact with asymmetric migration	-1255.89	2523.78	3026.62	-	2.3434	5.2936	-	-	-	-	1.8644	9.7638	-	-	-	-	0.0121	0.5353	-
(B) symmetric migration	-1274.02	2556.04	3136.09	-	1.4342	5.1103	-	-	-	-	8.4561	-	-	-	-	-	0.4495	-	-
(I) ancestral symmetric migration	-1306.38	2622.76	3144.11	-	1.6976	4.4919	-	-	-	-	10.5463	-	-	-	-	-	0.4093	0.0105	-
(F) asymmetric migration and population size change	-1337.48	2690.96	1771.62	-	-	0.1983	1.6391	1.9934	8.3463	7.033	1.9232	-	-	-	-	-	6.9754	0.7126	-
(J) ancestral asymmetric migration	-1468.33	2948.66	2638.51	-	1.8252	4.7451	-	-	-	-	29.6255	1.2755	-	-	-	-	0.6613	0.0232	-
(M) ancestral symmetric migration with population size change	-1506.49	3026.98	3377.72	-	-	1.5991	1.0153	6.2071	24.0289	25.5961	-	-	-	-	-	-	0.3761	0.0781	-
(K) secondary contact with symmetric migration and population size change	-1537.19	3088.38	1142.17	-	-	0.1781	0.1657	2.2247	23.1174	0.4043	-	-	-	-	-	-	2.4777	2.8818	-
(S) secondary contact with symmetric migration population size change under three epoch (isolation - migration - isolation)	-1566.83	3149.66	2134.6	-	0.6552	0.1955	-	1.835	7.9273	13.9999	-	-	-	-	-	-	0.1319	0.0756	0.0173
(Q) secondary contact with asymmetric migration under three epoch (isolation - migration - isolation)	-1707.18	3426.26	2301.44	-	1.6228	7.5548	-	-	-	-	5.367	-	-	-	-	-	0.3248	0.5493	0.0236
(T) secondary contact with asymmetric migration population size change under three epoch (isolation - migration - isolation)	-1716.21	3450.42	1232.09	-	-	5.4939	1.1765	3.9713	7.9346	13.9931	0.3071	-	-	-	-	-	8.3158	2.5922	0.0342
(N) ancestral asymmetric migration with population size change	-1876.62	3769.24	4168.87	-	-	3.8477	23.0938	3.2409	6.9858	0.5037	5.3401	-	-	-	-	-	0.0979	0.0287	-
(A) no migration	-1988.48	3982.96	4224.85	-	4.7308	16.0142	-	-	-	-	-	-	-	-	-	-	0.1138	-	-
(D) no migration with population size change	-2033.33	4078.66	4223.78	-	-	7.5318	14.3295	2.3013	2.6296	-	-	-	-	-	-	-	0.0968	0.0239	-
(R) secondary contact with asymmetric migration under three epoch (isolation - migration - isolation)	-2186.84	4387.68	3656.55	-	4.7998	8.865	-	-	-	-	0.5733	2.422	-	-	-	-	0.1294	0.0277	0.1471
(L) secondary contact with asymmetric migration and population size change	-2553.1	5122.2	8182.32	-	-	8.2458	2.0076	0.5881	2.4257	0.1099	22.9295	-	-	-	-	-	20.7734	17.8449	-
M vs T																			
(S) secondary contact with symmetric migration population size change under three epoch (isolation - migration - isolation)	-2076	4168	4199.58	-	-	0.3094	1.1757	2.9057	0.9393	1.8086	-	-	-	-	-	-	0.3692	0.0641	0.0619
(M) ancestral symmetric migration with population size change	-2215.89	4445.78	3822.47	-	-	0.5073	0.6306	15.2277	15.4825	0.7442	-	-	-	-	-	-	1.3466	0.0797	-
(F) asymmetric migration and population size change	-2274	4564	4274.99	-	-	0.272	0.8452	2.2173	0.8006	0.9317	0.4443	-	-	-	-	-	1.0315	0.1001	-
(H) secondary contact with asymmetric migration	-2449.35	4916.7	3685.32	-	0.831	1.193	-	-	-	-	0.7101	0.3337	-	-	-	-	0.4972	0.2582	-
(G) secondary contact with symmetric migration	-2546.86	5103.72	3635.68	-	0.9581	1.1164	-	-	-	-	0.538	-	-	-	-	-	0.5689	0.1793	-
(C) asymmetric migration	-2654.28	5318.56	4019.14	-	0.7266	0.9885	-	-	-	-	0.5459	0.337	-	-	-	-	0.7347	-	-
(B) symmetric migration	-2754.61	5517.22	3808.84	-	0.8437	0.9757	-	-	-	-	0.4692	-	-	-	-	-	0.8443	-	-
(K) secondary contact with symmetric migration and population size change	-2786.1	5586.2	3732.87	-	-	-	15.1116	1.4335	0.8697	1.11	0.3407	-	-	-	-	-	0.0573	0.6637	-
(O) two epoch of symmetric migration	-2835.26	5682.52	2933.45	-	1.0421	1.3858	-	-	-	-	-	0.7857	-	-	-	-	0.3907	1.0894	-
(P) two epoch of asymmetric migration	-2892.66	5801.32	1689.53	-	1.609	2.5379	-	-	-	-	-	-	22.807	2.6106	0.3055	0.1493	1.8543	2.2994	-

SI References

1. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540 (2019).
2. J. Köster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
3. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997v2 (2013).
4. N. Ulahannan, *et al.*, Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure. *bioRxiv*, 833590 (2019).
5. N. C. Durand, *et al.*, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *cells* **3**, 95–98 (2016).
6. N. C. Durand, *et al.*, Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).
7. A. A. Myburg, *et al.*, The genome of *Eucalyptus grandis*. *Nature* **510**, 356–362 (2014).
8. F. Cabanettes, C. Klopp, D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6** (2018).
9. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
10. S. Kovaka, *et al.*, Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**, 278 (2019).
11. B. L. Cantarel, *et al.*, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188–196 (2008).
12. A. Izuno, T. Wicker, M. Hatakeyama, D. Copetti, K. K. Shimizu, Updated Genome Assembly and Annotation for *Metrosideros polymorpha*, an Emerging Model Tree Species of Ecological Divergence. *G3: Genes, Genomes, Genetics* **9**, 3513–3520 (2019).
13. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
14. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
15. S. Picelli, *et al.*, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

16. J. Y. Choi, M. Purugganan, E. A. Stacy, Divergent Selection and Primary Gene Flow Shape Incipient Speciation of a Riparian Tree on Hawaii Island. *Mol Biol Evol* **37**, 695–710 (2020).
17. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics* **15**, 356 (2014).
18. M. Fumagalli, F. G. Vieira, T. Linderoth, R. Nielsen, ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–7 (2014).
19. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics* **195**, 693–702 (2013).
20. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
21. D. Bryant, R. Bouckaert, J. Felsenstein, N. A. Rosenberg, A. RoyChoudhury, Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution* **29**, 1917–1932 (2012).
22. R. Bouckaert, *et al.*, BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* **10**, e1003537 (2014).
23. R. E. Green, *et al.*, A draft sequence of the Neandertal genome. *Science* **328**, 710–22 (2010).
24. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution* **28**, 2239–2252 (2011).
25. M. Petr, B. Vernot, J. Kelso, admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).
26. N. Patterson, *et al.*, Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
27. S. H. Martin, S. M. Van Belleghem, Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* **206**, 429–438 (2017).
28. B. L. Browning, Y. Zhou, S. R. Browning, A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* **103**, 338–348 (2018).
29. S. R. Browning, B. L. Browning, Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am J Hum Genet* **81**, 1084–1097 (2007).
30. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).

31. I. Gronau, M. J. Hubisz, B. Gulko, C. G. Danko, A. Siepel, Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034 (2011).
32. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* **46**, 919–925 (2014).
33. S. Schiffels, K. Wang, “MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent” in *Statistical Population Genomics*, Methods in Molecular Biology., J. Y. Duthiel, Ed. (Springer US, 2020), pp. 147–166.
34. C. C. Chang, *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4** (2015).
35. D. M. Portik, *et al.*, Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology* **26**, 5245–5263 (2017).
36. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
37. B. C. Haller, P. W. Messer, SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol Biol Evol* **36**, 632–637 (2019).
38. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
39. F. Han, *et al.*, Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome research* **27**, 1004–1015 (2017).
40. Y. Kim, R. Nielsen, Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**, 1513–24 (2004).
41. N. Alachiotis, A. Stamatakis, P. Pavlidis, OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* **28**, 2274–2275 (2012).
42. J. Huerta-Cepas, *et al.*, Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).