

Dear Editorial Office,

Please accept our submitted manuscript, entitled "Reconstructing the history and biological consequences of a plant invasion on the Galápagos islands" for consideration as a Research Report in the *Proceedings of the National Academy of Sciences*. My co-authors have approved this submission and we declare no conflicts of interest.

Invasive species are a growing threat to global biodiversity, yet the genetic and demographic factors allowing successful invasions is poorly understood. Although strong genetic bottlenecks might be expected to constrain the ability of colonizers to adapt to novel habitats, cases of successful invasions are common, suggesting that other diversity-maintaining processes may be at play. In this manuscript, we investigate these processes in the case of a recent invasion on the Galápagos Islands. Using genome sequence data from 174 novel island collections and a panel of 132 continental accessions, we quantify the contributions of demographic bottlenecks, single versus multiple introductions, and post-invasion introgression to patterns of genetic diversity in populations of the invasive *Solanum pimpinellifolium* and the Galapagos endemic *Solanum cheesmaniae*, two wild tomato species. We find that the majority of *S. pimpinellifolium* originated in central Ecuador, are the product of a recent human-mediated invasion, and are actively hybridizing with *S. cheesmaniae*. Furthermore, we characterize the detailed genomic ancestry of introgression in two admixed *S. pimpinellifolium* populations and show that the recent appearance of a novel orange-fruited phenotype in these invasive populations can be explained by introgression from *S. cheesmaniae* at two distinct and well-characterized carotenoid biosynthesis loci. Our findings provide critical data on the evolutionary importance of hybridization during colonization, including the role of introgression in shaping invasion outcomes.

We believe this manuscript will be of broad interest to the *PNAS* readership. Our study uses state-of-the-art population genomic approaches to reconstruct a tangled history of invasion, expansion, and gene flow in an invasive wild tomato. In doing so, we describe an exciting and novel case of phenotypic convergence for fruit color, driven by the exchange of carotenoid loci between endemic and invasive populations. Aside from a select few high-impact case studies, the role of introgression in shaping evolutionary trajectories in the wild (including convergent and adaptive evolution) is not well understood. Our study will add to this important literature, and the broad range of biologists engaging with it, including researchers working on speciation and hybridization, the genetics of important functional trait variation, dynamics of introduction and invasion, and conservation genetics. By assessing these processes in crop wild relatives (*S. pimpinellifolium* and *S. cheesmaniae* are both close relatives of domesticated tomato), the reach of our findings is extended from genetic and ecological to plant breeding disciplines. Finally, our focal invasion is occurring on the Galápagos Islands—impacting species that were first collected by Darwin himself—so provides a new example of how the isolation and unique diversity of this archipelago provides a cradle for natural experiments that enrich our understanding of evolutionary biology. For all of these reasons, we believe our manuscript will fit well at *PNAS*, which has a history of highlighting analyses of many diverse aspects of the natural world, using cutting-edge tools to examine pressing contemporary questions.

Thank you for your time and consideration,

Matthew JS Gibson
gibsomat@indiana.edu

1 **Reconstructing the history and biological consequences of a plant invasion on**
2 **the Galápagos islands**

5 Matthew J.S. Gibson^{1*}, María de Lourdes Torres^{2,3}, Yaniv Brandvain⁴, & Leonie C.
6 Moyle¹

9 ¹Department of Biology, Indiana University, Bloomington, Indiana, 47405

11 ²Universidad San Francisco de Quito (USFQ). Colegio de Ciencias Biológicas y
12 Ambientales, Laboratorio de Biotecnología Vegetal. Campus Cumbayá. Quito, Ecuador.

14 ³Galapagos Science Center, Universidad San Francisco de Quito and University of
15 North Carolina at Chapel Hill, San Cristobal, Galapagos, Ecuador

17 ⁴Department of Plant Biology, University of Minnesota-Twin Cities, St. Paul, Minnesota,
18 55455

21 ORCID Identifiers:

23 Matthew J.S. Gibson: 0000-0001-7855-1628
24 María de Lourdes Torres: 0000-0001-7207-4568
25 Leonie C. Moyle: 0000-0003-4960-8001

28 **Corresponding Author:** Matthew J.S. Gibson, gibsomat@indiana.edu

30 **Classification:** Biological Sciences, Evolution

32 **Keywords:** Galápagos, colonization, hybridization, introgression, convergence

35 Author contributions: MJSG, MLT, & LCM designed the research; MJSG performed field
36 work; MJSG performed molecular and bioinformatic analyses; YB designed the
37 introgression HMM; MJSG implemented the HMM. MJSG wrote the manuscript, with
38 input from YB, MLT, and LCM.

40 **Abstract**

41
42 The introduction of non-native species into new habitats is one of the foremost risks to
43 global biodiversity. Here, we evaluate a recent invasion of wild tomato (*Solanum*
44 *pimpinellifolium*) onto the Galápagos islands from a population genomic perspective,
45 using a large panel of novel collections from the archipelago as well as historical
46 accessions from mainland Ecuador and Peru. We infer a recent invasion of *S.
47 pimpinellifolium* on the islands, largely the result of a single event from central Ecuador
48 which, despite its recency, has rapidly spread onto several islands in the Galápagos. By
49 reconstructing patterns of local ancestry throughout the genomes of invasive plants, we
50 uncover evidence for recent hybridization and introgression between *S. pimpinellifolium*
51 and the closely related endemic species *Solanum cheesmaniae*. Two large introgressed
52 regions overlap with known fruit color loci involved in carotenoid biosynthesis. Instead of
53 red fruits, admixed individuals with endemic haplotypes at these loci have orange fruit
54 colors that are typically characteristic of the endemic species. We therefore infer that
55 introgression explains the observed trait convergence. Moreover, we infer roles for two
56 independent loci in driving this pattern, and a likely history of selection favoring the
57 repeated phenotypic transition from red to orange fruits. Together, our data reconstruct a
58 complex history of invasion, expansion, and gene flow among wild tomatoes on the
59 Galápagos islands. These findings provide critical data on the evolutionary importance of
60 hybridization during colonization and its role in influencing conservation outcomes.

61
62
63
64

65 **Significance Statement**

66
67 The isolation and unique diversity of the Galápagos Islands provide numerous natural
68 experiments that have enriched our understanding of evolutionary biology. Here we use
69 population genomic sequencing to reconstruct the timing, path, and consequences of a
70 biological invasion by wild tomato onto the Galápagos. We infer that invasive
71 populations originated from a recent human-mediated migration event from central
72 Ecuador. Our data also indicate that invasive populations are hybridizing with endemic
73 populations, and that this has led to some invasive individuals adopting both fruit color
74 genes and the fruit color characteristic of the endemic island species. Our results
75 demonstrate how hybridization can shape patterns of trait evolution over very short time
76 scales, and characterize genetic factors underlying invasive success.

77
78
79

80 **Introduction**

81 The success of colonizing species depends on complex interactions between
82 local environments and the availability of relevant genetic variation. Introduction events
83 are often associated with strong genetic bottlenecks (Kolbe et al., 2004; Colautti et al.,
84 2005; Golani et al., 2007) and reduced effective population sizes, features which may
85 constrain the ability of colonizers to adapt to novel environments and compete with
86 native biota (Lande, 1988; Lee, 2002). This suggests that biological invasions should
87 rarely follow from introductions (Queller, 2000; Kolbe et al., 2004), yet successful
88 invasions are nonetheless pervasive (Kolbe et al., 2004; Allendorf & Lundquist, 2003;
89 Estoup et al., 2016; Comeault et al., 2020).

90 Several factors could be involved in this success. Despite intense bottlenecks,
91 diversity could be maintained by other means, including multiple independent
92 introductions (Facon et al., 2008; Kolbe et al., 2004) or via hybridization with
93 congenerics present in the new habitat (Ellstrand & Schierenbeck, 2000; Lavergne &
94 Molofsky, 2007; Reatini and Vision, 2020; Stepien et al., 2005). Of these mechanisms,
95 hybridization might be particularly important for facilitating invasion into island habitats.
96 Hybridization among native and introduced taxa is common on islands (Carlquist, 1974),
97 potentially because of limitations on geographic extent, the abundance of generalist
98 pollinators (Olesen et al., 2002), and/or frequent anthropogenic disturbance (Bertolo et
99 al., 2012; Lin et al., 2013; Long et al., 2014; Cao et al., 2014). In addition, while the
100 geographic isolation of insular habitats makes them hot spots for species endemism,
101 only a small subset of continental taxa are successful in colonizing remote islands. The
102 resulting incomplete trophic networks provide abundant ecological opportunities for
103 invaders, including reproductive interactions between closely related species. Given
104 these potentially complex contributing factors, describing the occurrence and
105 consequences of invasion is critical for understanding both the dynamics of colonizing
106 populations and for predicting conservation outcomes.

107 In this study, we investigate the contributions of demographic bottlenecks, single
108 versus multiple introductions, and post-invasion hybridization, to patterns of genomic
109 variation in populations of invasive and endemic tomato species on the Galápagos
110 Islands. Two yellow/orange fruited tomato species are considered endemic to the
111 islands: *Solanum cheesmaniae* (L. Riley) Fosberg [CHS] and *Solanum galapagense*
112 S.C. Darwin and Peralta [GAL] (SI Appendix, section S1). Two red-fruited invasive
113 species from continental Ecuador and Peru are now also documented on the
114 archipelago: *Solanum pimpinellifolium* L. [PIM] and *Solanum lycopersicum* L. [LYC]—
115 the domesticated tomato. Domesticated LYC was almost certainly introduced for
116 agriculture (Rick, 1963). Wild species PIM was likely also introduced by early human
117 colonizers (Darwin et al., 2003; Darwin, 2009), however the timing and source of this
118 introduction is not known. Recent field surveys indicate substantially increased
119 abundance of the invasive species while abundance of the two endemic species has
120 markedly declined over the past two decades (Darwin, 2009; Nuez et al., 2004; Gibson
121 et al., 2020), suggesting that recent demographic shifts may pose an extinction threat to
122 the endemic species. Several factors also indicate a high potential for hybridization
123 between native and invasive species, including overlapping habitats (Darwin, 2009;
124 Nuez et al., 2004), similar flower morphologies (Darwin et al., 2003; Darwin, 2009; Rick
125 et al., 1977; Vosters et al., 2014), and shared pollinators (Darwin, 2009). Moreover, all

126 four species are closely related—having diverged less than 500 kya (Pease et al.,
127 2016)—and all can be crossed to produce hybrids in the greenhouse (Rick, 1956; Rick
128 & Bowman, 1961; Rick & Fobes, 1975).

129 Using genomic sequencing data from 174 plants (representing all four species)
130 from the largest islands of San Cristobal, Santa Cruz, and Isabela, and a panel of 132
131 mainland PIM accessions from across the entire native range on continental South
132 America, we (*i*) infer the timing, source, and number of invasions by PIM onto the
133 Galápagos and (*ii*) evaluate evidence for post-colonization gene flow between the four
134 tomato taxa, and its evolutionary consequences. We find that the majority of PIM
135 originated in central Ecuador, are the product of a recent invasion, and are actively
136 hybridizing with an endemic relative. By characterizing fine-scale local ancestry, we find
137 that the emergence of novel orange-fruited plants—which resemble the endemic
138 species in color—in two invasive populations can be explained by endemic
139 introgression at distinct carotenoid loci with known phenotypic effects specifically on fruit
140 color. Our findings reconstruct a recent path of invasion via Ecuador, provide evidence
141 for ongoing interspecific gene flow, and suggest a history of natural selection favoring
142 orange fruits in the island habitat.

143
144

145 **Results**

146

147 **Sequencing and collections**

148

149 Sequence data were drawn from 306 individual samples. We performed double-digest
150 RAD (ddRAD) sequencing (using *PstI* and *EcoRI* enzymes) of 174 wild collected
151 individuals from 13 populations of endemic and invasive tomatoes from three islands in
152 the Galápagos archipelago: San Cristobal, Santa Cruz, and Isabela (**Figure 1; Table 1;**
153 *SI Appendix, Fig. S5 and Table S1*). We complemented these data with ddRAD reads
154 from 132 mainland PIM (*SI Appendix, Fig. S1 and Table S2*), previously sequenced in
155 Gibson & Moyle (2020) using the same enzymes. We recovered 18573 high quality
156 RAD loci, each sequenced to an average of 61.4X (s.d. = 35X) in 80% of all 306
157 samples (*SI Appendix, Table S3 and Table S4*). Average insert size was 192 bp (s.d. =
158 51.7) after adapter and quality trimming. After filtering for depth (> 8 reads), 11297
159 SNPs were retained. After filtering for LD ($r^2 < 0.7$), 5767 SNPs were retained. Refer to
160 *SI Appendix, Table S5* for a summary of each filtering step and the analyses for which
161 each dataset was used.

162

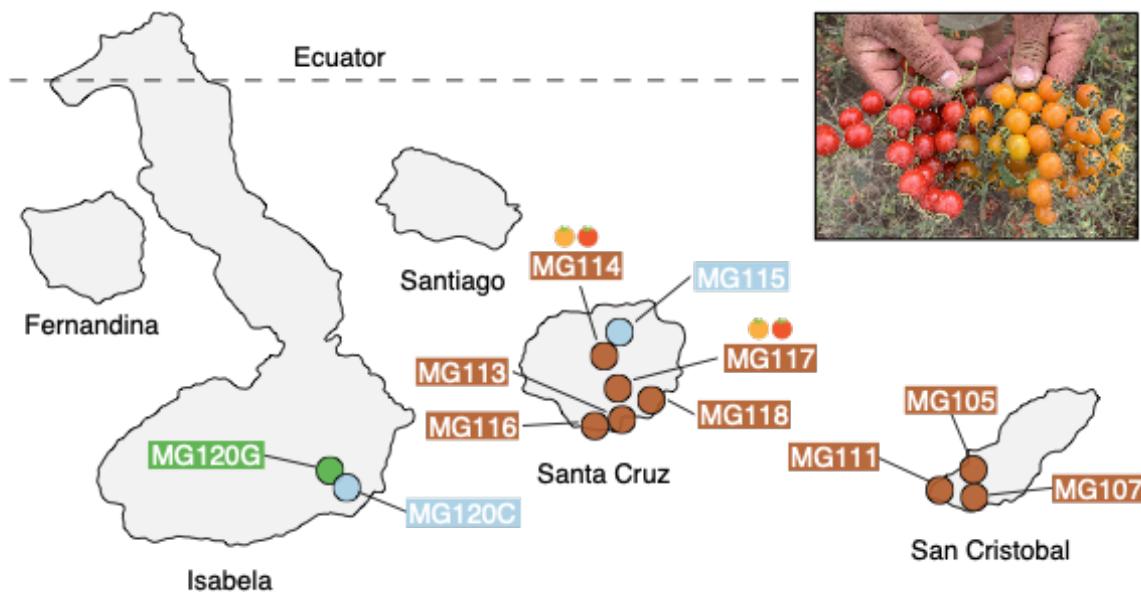


Figure 1: Geographic distribution of focal sampling sites on the Galápagos Islands. *Inset:* Photograph of polymorphic (red/orange) PIM fruits representative of populations MG114 and MG117. For simplicity, LYC populations as well as sampling sites with < 8 individuals are not included here. Refer to Table S1 for a full list of collection localities and sample sizes.

163

164

165

166

167

168

169 **Genetic data support an Ecuadorian origin for most invasive populations**

170
 171 Using our ddRAD sequencing data for Galápagos and continental PIM, we analyzed
 172 population genetic signatures of colonization and characterized the origin and path of
 173 invasion into the archipelago. Nucleotide diversity (π in 100kb overlapping windows;
 174 **Figure 2C**) was reduced on average 6.6-fold in island populations relative to mainland
 175 accessions (**Table 1**), a pattern consistent with population genetic expectations
 176 following colonization.

177 Genetic variation in the native (mainland) range of PIM is highly geographically
 178 structured (Gibson & Moyle, 2020; *SI Appendix, Fig. S1*), allowing us to infer a putative
 179 origin of PIM lineages invasive on the Galápagos. To do so, we estimated genome-wide
 180 patterns of relatedness between invasive and mainland individuals using several
 181 methods. A rate-smoothed maximum likelihood tree constructed in *Treemix* (Pickrell &
 182 Pritchard, 2012) identified Galápagos PIM as monophyletic, and clearly separated
 183 island and non-island clades (**Figure 2B**). In general, pairwise sequence divergence
 184 was lower in Galápagos-Ecuador comparisons (average $d_{xy} = 2.3 \times 10^3$) than between
 185 Galápagos-Peru comparisons (average $d_{xy} = 3.6 \times 10^3$; **Figure 2C**), and samples
 186 showing low genome-wide divergence were clustered in central Ecuador (similar
 187 patterns were observed using F_{ST} ; *SI Appendix, Table S10*).
 188
 189

190 **Table 1:** Diversity statistics for focal population samples (H = observed heterozygosity; π = genome wide
 191 nucleotide diversity).

Taxa	Population	Island	Endemic	Segregating Sites	H	π
PIM	Peru	N	N	32820	0.00025	0.00094
	Ecuador	N	N	21773	0.00033	0.00130
	MG105	Y	N	1520	0.00011	0.00009
	MG107	Y	N	4434	0.0002	0.00034
	MG111	Y	N	2562	0.00012	0.00010
	MG113	Y	N	3763	0.00015	0.00029
	MG114	Y	N	4730	0.00016	0.00023
	MG116	Y	N	4929	0.00024	0.00024
	MG117	Y	N	4776	0.00016	0.00028
CHS	MG115	Y	Y	8540	0.00023	0.00045
	MG120C	Y	Y	854	0.0001	0.00008
GAL	MG120G	Y	Y	3282	0.00013	0.00014
GALxCHS	MG120GC	Y	Y	2857	0.00023	0.00026
LYC	MG125	Y	N	7219	0.00035	0.00063
	MG126	Y	N	4567	0.00018	0.00052

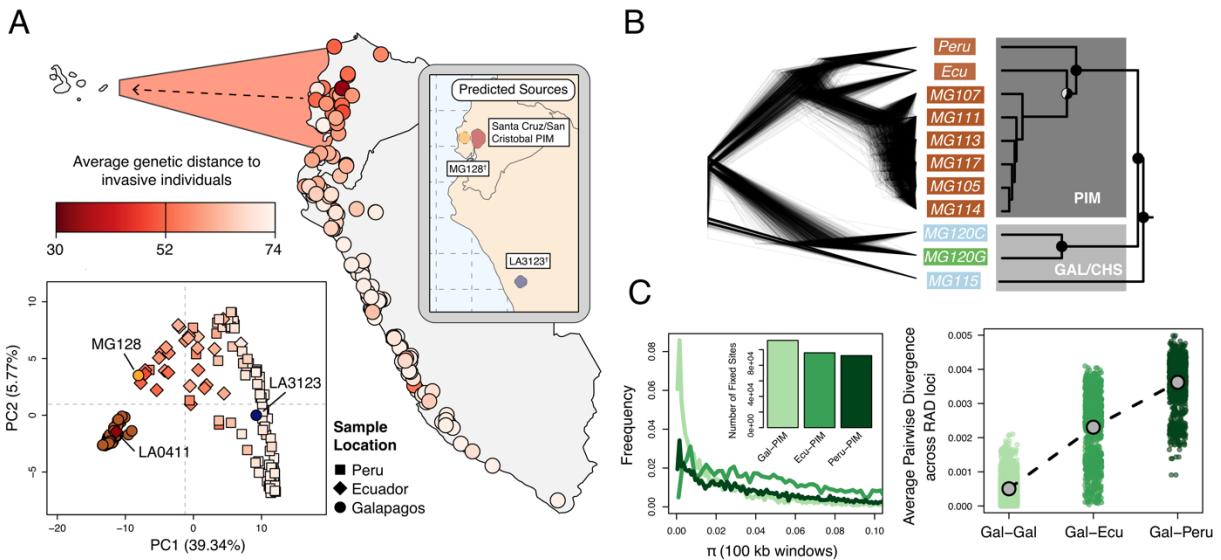


Figure 2: Galápagos PIM is the result of a recent invasion from Ecuador. (A) *Map*: average genetic distance between Galápagos-PIM collections and each of the 140 mainland accessions. *Plot*: multi-locus PCA. Squares, diamonds, and circles indicate Peruvian, Ecuadorian, and Galápagos collections, respectively. *Inset*: Predicted continental origins for Galápagos PIM collections. Colors are same as shown in the multi-locus PCA (†Exact locations vary substantially between runs. Results from a single run are shown). (B) Maximum likelihood relationships among focal populations calculated with *Treemix* (allowing no migration). *Left*: inferred trees of 1000 resampled datasets (500 SNPs, with replacement). *Right*: consensus topology. All trees were rate smoothed ($\lambda = 1$). (C) Diversity and divergence metrics. *Left*: nucleotide diversity (π) calculated for Galápagos-PIM, Ecuador-PIM, and Peru PIM in overlapping 100kb windows. Invariant windows ($\pi = 0$) are truncated and are instead shown in the inset bar plot. *Right*: average pairwise sequence divergence for three PIM comparisons: Gal x Gal, Gal x Ecu, and Gal x Peru. Each point represents a comparison between individuals, averaged over all loci.

198
199
200 To investigate potential source localities for invasive populations at a finer scale,
201 we implemented the software *Locator* (Battey et al., 2020) which uses a machine
202 learning algorithm to predict sample origins from genotype data. *Locator* predictions
203 indicated 2 to 3 source regions for Galápagos PIM, although the exact locations varied
204 across runs and depended on which island PIM collections were considered (*SI*
205 *Appendix*, *Fig. S2*). Santa Cruz and San Cristobal PIM collections were predicted to
206 have originated in central Ecuador; this result was generally consistent across runs, with
207 the consensus being an origin near Los Ríos and Guayas provinces in southcentral
208 Ecuador (*SI Appendix*, *Fig. S2*). Interestingly, we also infer that one mainland accession
209 represents a back migration from the Galápagos to Los Ríos (LA0411; *SI Appendix*,
210 *section S2*), further highlighting the high degree of connectivity between this region and
211 the islands. In contrast, the remaining two samples, LA3123 (a historical collection from
212 Santa Cruz sampled in 1991) and MG128-1 (newly sampled on Isabela), were predicted
213 to have originated in alternative locations, with most runs supporting a Peruvian origin
214 for LA3123 and an Ecuadorian origin for MG128-1 (*SI Appendix*, *Fig. S2*). The exact
215 origin locations for these samples varied substantially across runs. In general, *Locator*

216 predictions were consistent with the pattern of low pairwise sequence divergence
217 between Galápagos PIM and central Ecuadorian samples, pointing to Ecuador, and
218 perhaps central Ecuador in particular, as the source of the majority of invasive PIM
219 populations on the Galápagos.

220 Together our data support 2-3 independent introductions of PIM onto the
221 archipelago, each with variable consequences for current invasive populations: (i) a
222 minor event from Peru [LA3123], (ii) a minor event from Ecuador [MG128-1], and (iii) a
223 major event from central Ecuador that is responsible for nearly all sampled populations.
224
225

226 **Demographic reconstruction supports a recent colonization by PIM on the 227 Galápagos**

228

229 We used the allele frequency spectrum to model the demographic history of invasive
230 populations. In particular, we evaluated two demographic models using $\delta\alpha\delta i$
231 (Gutenkunst et al., 2010): (i) a neutral model of constant population size and (ii) an
232 instantaneous population bottleneck model. Since this species is self-fertile (that is, it
233 lacks genetic self-incompatibility that is present in some wild tomato species; Rick et al.,
234 1977) we simultaneously inferred the inbreeding coefficient (F ; Blischak et al., 2020).
235 The bottleneck model thus included 5 parameters: bottleneck population size (N_B), final
236 population size (N_F), timing of the bottleneck (T_B), timing of recovery (T_F), and the
237 inbreeding coefficient. To limit potential confounding effects due to population structure
238 within PIM, we estimated the folded site frequency spectrum (SFS) of a single
239 population (MG114, which was the most deeply sampled of our PIM populations). We
240 also masked regions of inferred introgression (as detected by our HMM, see below) as
241 these can spuriously inflate rare variants and thus bias the inference of a bottleneck and
242 subsequent parameter estimation (*SI Appendix, section S6*). In our masked dataset, we
243 observed an excess of rare variants (genome wide Tajima's $D = -0.48$) more consistent
244 with a bottleneck model ($RSS_{Bottle} = 1.39$; $\ln(L) = -18.93$) than a neutral model ($RSS_{Neutral}$
245 = 15.12; $\ln(L) = -25.89$).

246 We used the best-fit bottleneck model to estimate the timing of the introduction,
247 performing the $\delta\alpha\delta i$ optimization procedure described in Portik et al. (2017). We inferred
248 a recent bottleneck occurring 284.1 generations in the past (**Table 2**). During the
249 bottleneck, N_E (effective population size) was reduced by 27% relative to the ancestral
250 reference (N_{Ref}). Variance in all parameter distributions as estimated by a nonparametric
251 bootstrap was high (*SI Appendix, Fig. S6*), and median estimates deviated slightly from
252 the optimized estimates (**Table 2**). In all cases, the optimized parameter estimate fell
253 within the bootstrapped 95% CIs.

254 For comparison, we also model the history of an endemic CHS population
255 (MG115) using the same framework as above. A bottleneck model again fit the data
256 better ($\ln(L) = -49.91$) than a neutral model ($\ln(L) = -73.51$). We estimate the timing of
257 the bottleneck at 3,747.39 generations in the past, approximately 13X higher than the
258 estimate for PIM. Similarly, N_E during the bottleneck was reduced 34% relative to the
259 ancestral population and subsequently increased following the bottleneck (increase of
260 1.7-fold relative to N_{Ref}).
261

262
263
264
265
266
267

Table 2: Demographic model estimates for PIM population MG114 inferred using $\delta\alpha\delta i$. 95% CI values were obtained from 2,000 bootstrap replicates of the SFS. Each estimate is shown in rescaled units (rescaled by N_{Ref} for N_B and N_F , and by $2N_{Ref}$ for T_B and T_F).

Parameter	Optimum	Bootstrap Median	95% CI
N_B	184.69	286.16	4.49 - 1483.26
N_F	2730.97	977.28	319.12 – 5333.40
T_B	43.12	846.90	91.88 – 7871.41
T_F	240.98	839.13	22.28 – 13591.11
F	0.24	0.02	$2.18 \times 10^{-10} – 0.54$

268
269
270

271 Admixture analyses support the occurrence of inter- and intraspecific gene flow

272
273
274
275
276
277
278
279
280
281
282
283
284
285

The close evolutionary relationship of PIM, CHS, and GAL, their similar floral morphologies, and the presence of only a single major pollinator on the islands (*Xylocarpus darwini*; McMullen, 1999), indicate the potential for interspecific gene flow between tomato species may be high. Key morphological observations also suggest that these species may be exchanging genes (Darwin, 2009). In particular, we have previously described a novel fruit color polymorphism in two Santa Cruz PIM populations (MG114 & MG117; Gibson et al., 2020), where approximately 40% of individuals have orange instead of their ancestrally red fruits. Orange fruits are very rare in mainland PIM (TGRC passport data; www.tgrc.ucdavis.edu) but are diagnostic of the two endemic Galápagos species. Accordingly, we used multiple population genomic methods to investigate evidence of hybridization and introgression in the genomes of island plants, paying special attention to patterns of admixture in the polymorphic PIM populations.

286
287
288
289
290
291
292
293

We first examined evidence for recent (early generation) hybrids by evaluating genome-wide signatures in *fastStructure* (Raj et al., 2014) and *NewHybrids* (Anderson & Thompson, 2002). Interestingly, we find no evidence of early generation CHSxPIM hybrids in either of the polymorphic PIM populations MG114 and MG117 (**Figure 3A**). However, these analyses did detect variable levels of CHS x PIM admixture at the nearby site MG115 (**Figure 3A**), a pattern which is also reflected in principle component-space (*SI Appendix, Fig. S8*). Using *NewHybrids*, we classified 4/6 of these admixed plants as first- or second-generation hybrids (*SI Appendix, Table S6 and S8*).

294
295
296
297
298
299
300
301

To complement the above analyses, we employed *Treemix* (Pickrell & Pritchard, 2012). The most likely topology inferred by *Treemix* implies three separate admixture events between PIM and CHS: two cases of CHS → PIM admixture and one case of PIM → CHS admixture on Santa Cruz (**Figure 3C**). This analysis therefore indicates repeated gene flow between PIM and CHS, although how many distinct events were involved is difficult to infer given the high genomic similarity of island PIM populations and recency of the invasion (*SI Appendix, Fig. S3 and S4*). As independent support for a history of gene flow, we calculated the four taxa D statistic of Durand et al. (2011)

302 using *Solanum pennellii* (LA3778) as an outgroup and treating PIM population MG114
 303 and CHS (MG115) as P2 and P3, respectively. We found that D was 0.818 (s.d. =
 304 0.028; bootstrapped $P < 0.02$), indicating an excess of ABBA sites and strong
 305 evidence for admixture between island PIM and CHS. D was also significant when other
 306 invasive PIM populations—Santa Cruz population MG117 or San Cristobal populations
 307 MG107 and MG105—were used as P2, indicating that the detected admixture likely
 308 predates the dispersal and differentiation between Santa Cruz and San Cristobal
 309 invasive PIM. This is consistent with inferences in the *Treemix* graph, in which
 310 admixture events between PIM and CHS involve internal branches that subtend current
 311 San Cristobal and Santa Cruz PIM populations (**Figure 3C**).

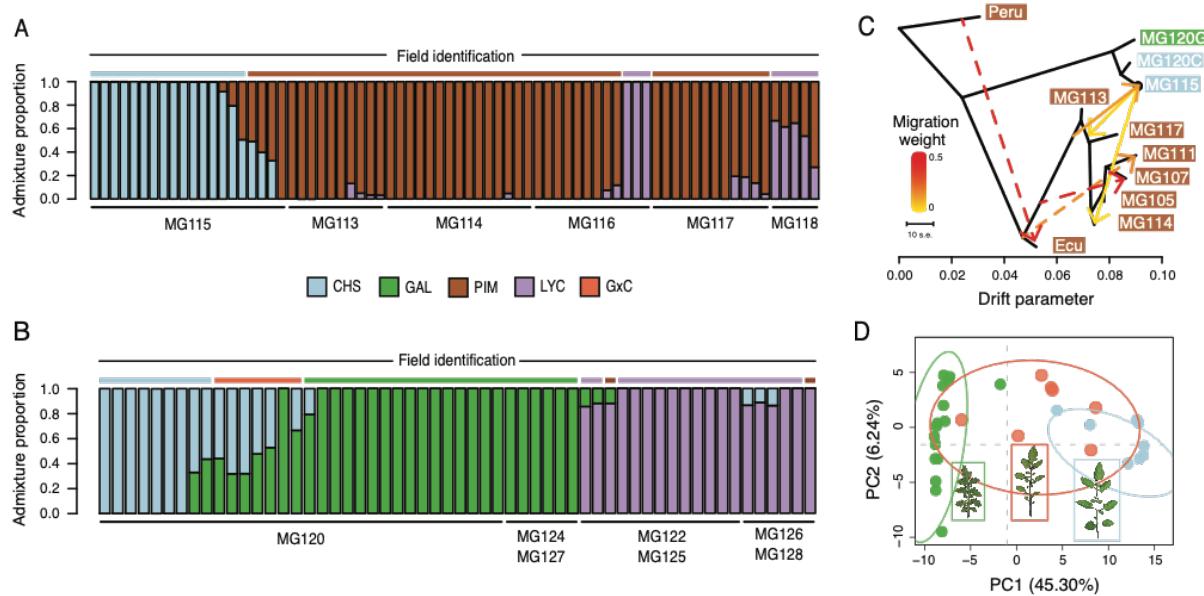


Figure 3: Patterns of population genetic structure and admixture on Santa Cruz and Isabela. (A) *fastStructure* inference for all Santa Cruz samples ($N = 74$). $K = 3$. (B) *fastStructure* inference for all Isabela samples ($N = 57$). $K = 3$. (C) *Treemix* analysis summary ($m = 6$; $\ln[L] = 395.08$). (D) Principle components analysis for samples at site MG120, a hybrid zone between CHS and GAL.

312 Interestingly, *fastStructure* and *Treemix* produced conflicting results regarding
 313 evidence for admixture in the polymorphic PIM populations; *Treemix* appears to support
 314 this while *fastStructure* does not. To evaluate whether this was due to differences in the
 315 detection of more subtle—and potentially older—signals of introgression, we
 316 implemented a local ancestry assignment algorithm using a Hidden Markov Model
 317 (HMM) to probe for evidence of introgression at a finer scale. Doing so, we found
 318 evidence for bidirectional gene flow between CHS and PIM (**Figure 4**; *SI Appendix*, Fig.
 319 S13), with inferred introgression being more common in the CHS → PIM direction and
 320 in PIM populations that were polymorphic for fruit color. Our HMM detected clear
 321 evidence for CHS ancestry within the polymorphic PIM populations MG114 and MG117,
 322
 323
 324

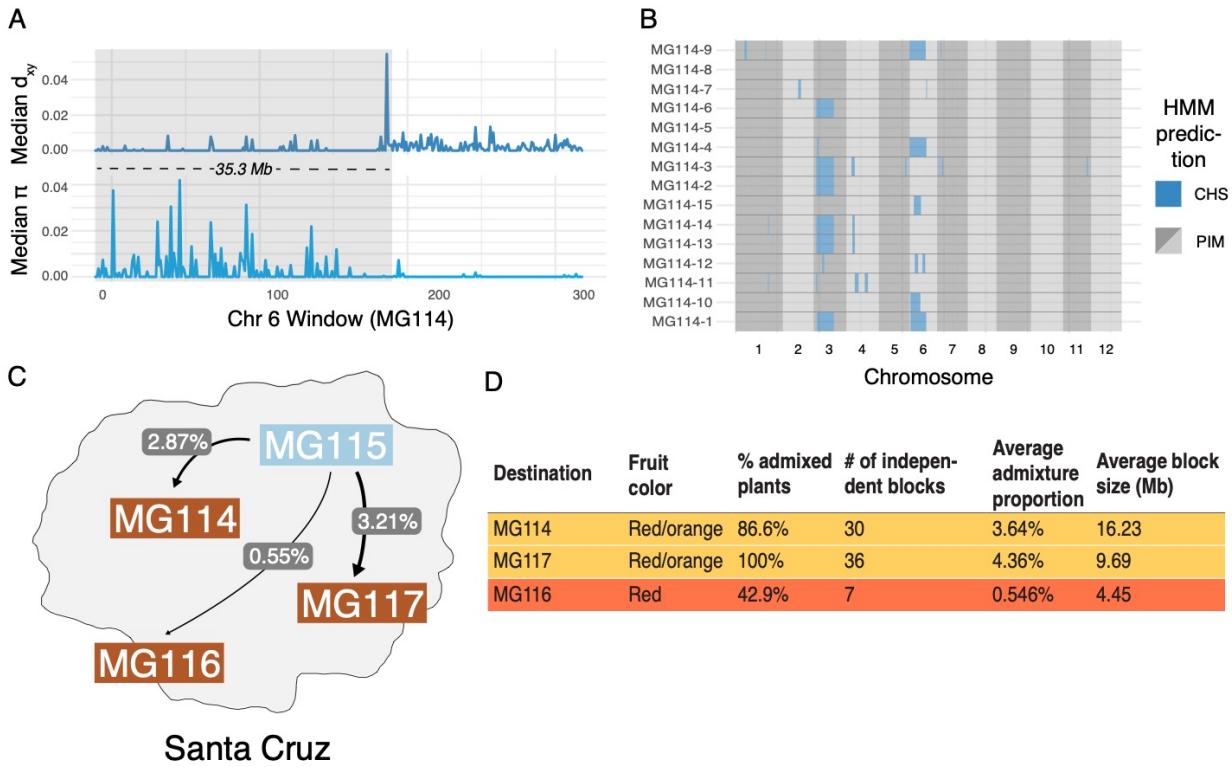


Figure 4: Local ancestry assignment using HMM characterizes a history of endemic x invasive introgression. (A) Patterns of diversity and divergence along chromosome 6 for an MG114 individual. The region of recent coalescence (low divergence; high diversity) with CHS is annotated in gray. This 20.2 kb block segregates at 20% in MG114. (B) Genome-wide HMM predictions for all individuals in MG114. The x-axis is ordered by chromosome and y-axis is ordered by individual. Two large CHS haplotypes segregate at high frequency on chromosomes 3 (40%) and 6 (20%). (C) Visual summary of admixture proportions from CHS into three PIM populations. (D) Summary of HMM assignment for each PIM population. Populations displaying variation in fruit color (MG114 & MG117) have more CHS ancestry than those which are fixed for the ancestral red state (MG116).

reflecting admixture from CHS → PIM. In contrast, inferred admixture in nonpolymorphic PIM (e.g., MG116) was much more restricted. For MG114, CHS ancestry blocks were large (average = 16,235 kb), of varying size ($sd = 139.43$ kb), and composed on average 3.64% of the genomes of any given MG114 plant (Figure 4C, 4D). Shared ancestry between MG114 and MG115 was dominated by two large CHS haplotypes segregating at moderate to high frequencies on chromosome three (40%; mean size = 51.35 Mb) and chromosome six (20%; mean size = 35.3 Mb; Figure 4B). The genomic distribution of CHS ancestry blocks in different individuals indicates they are not independent. For example, on chromosome 3 all but one individual (5/6) carrying the CHS haplotype had identical breakpoints, consistent with them being derived from the same hybridization (and subsequent recombination) event and/or the individuals being closely related (SI Appendix, Table S10). In MG117, CHS ancestry made up 4.36% ($sd = 2.56\%$) of any given MG117 genome and average block size was 9,687.5 kb (Figure 4C, 4D; SI Appendix, Table S12). As with MG114, a large CHS haplotype on

340 chromosome six occurs at a frequency of 0.42. This block varied substantially in size in
341 each individual, and all were discontinuous across the chromosome (i.e., there is an
342 implied double crossover event). Further, two individuals were heterozygous for
343 ancestry at the downstream portion of the haplotype (*SI Appendix*, Fig. S10, S11, S12).
344 In comparison to MG114 and MG117, MG116 showed little evidence for shared
345 ancestry. While 3/7 individuals had inferred signals of CHS ancestry, these blocks were
346 generally small compared to MG114 and MG117 (average = 4,450 kb; **Figure 5**) and
347 made up a substantially smaller fraction of the total genome (average admixture
348 proportion = 0.55%; **Figure 4D**). No large CHS haplotypes were segregating in MG116,
349 unlike those observed in MG114 and MG117 (*SI Appendix*, Table S13).

350 The size of detected ancestry blocks contains information regarding the timing of
351 gene flow, because it depends on the number of recombination events (generations)
352 that have occurred since an initial hybridization event. We can broadly estimate the age
353 of these haplotypes using a simple logarithmic relationship (*SI Appendix*, section S5;
354 Lynch & Walsh, 1998). In MG114 and MG117, age estimates are within the range of 4-
355 12 generations (e.g., the large chromosome 3 and 6 CHS haplotypes in MG114 are
356 estimated at 4.23 and 4.74 generations, respectively). In addition to placing boundaries
357 on when the initial hybridization took place in the past (>4 and up to 12 generations),
358 there is close agreement between age estimates in MG114 and MG117, suggesting that
359 these instances of CHS introgression might have been derived from the same
360 admixture event (*SI Appendix*, section S5).

361 Relative to the patterns of gene flow from CHS into polymorphic PIM described
362 above, gene flow in the opposite direction (PIM → CHS) was more restricted. The
363 average proportion of PIM ancestry within CHS individuals (at MG115) was 1.62% (s.d.
364 = 2.72%; *SI Appendix*, Table S13). These results point to a potential bias in the direction
365 of gene flow, with more exchange occurring from CHS into PIM than from PIM into
366 CHS.

367 In addition to inferred introgression between PIM and CHS, we also found evidence for
368 hybridization/introgression involving the other two taxa: GAL and LYC. In particular, we
369 uncovered a recent history of hybridization between CHS and GAL on Isabela—at *la*
370 *Laguna de Manzanilla* (**Figure 3B & 3D**). These two species have been reported as co-
371 occurring at this site since 2000, and hybridization has previously been hypothesized
372 based on allozyme and morphological analyses (Darwin, 2009; Gibson et al., 2020);
373 **Figure 3D**; *inset images*). Our *fastStructure* and PCA analyses of individuals at this site
374 clearly identify 9 samples as admixed (**Figure 3B**), mostly corresponding with
375 morphological classifications of intermediacy (**Figure 3D**). Of the 9 admixed plants, 4
376 appear to be first generation backcrosses, whereas 4 are F₂ (*SI Appendix*, Table S7).
377 We also identified putative cases of CHS/GAL x LYC admixture in two populations on
378 Isabela (**Figure 3B**), although some of these signals might be the product of unmodeled
379 genetic substructure (*SI Appendix*, section S3). On Santa Cruz, low levels of LYC
380 (domesticated tomato) ancestry were also detected within some PIM populations,
381 including a potential hybrid PIM x LYC population MG118 that was predicted to be
382 entirely first-generation hybrids.

383

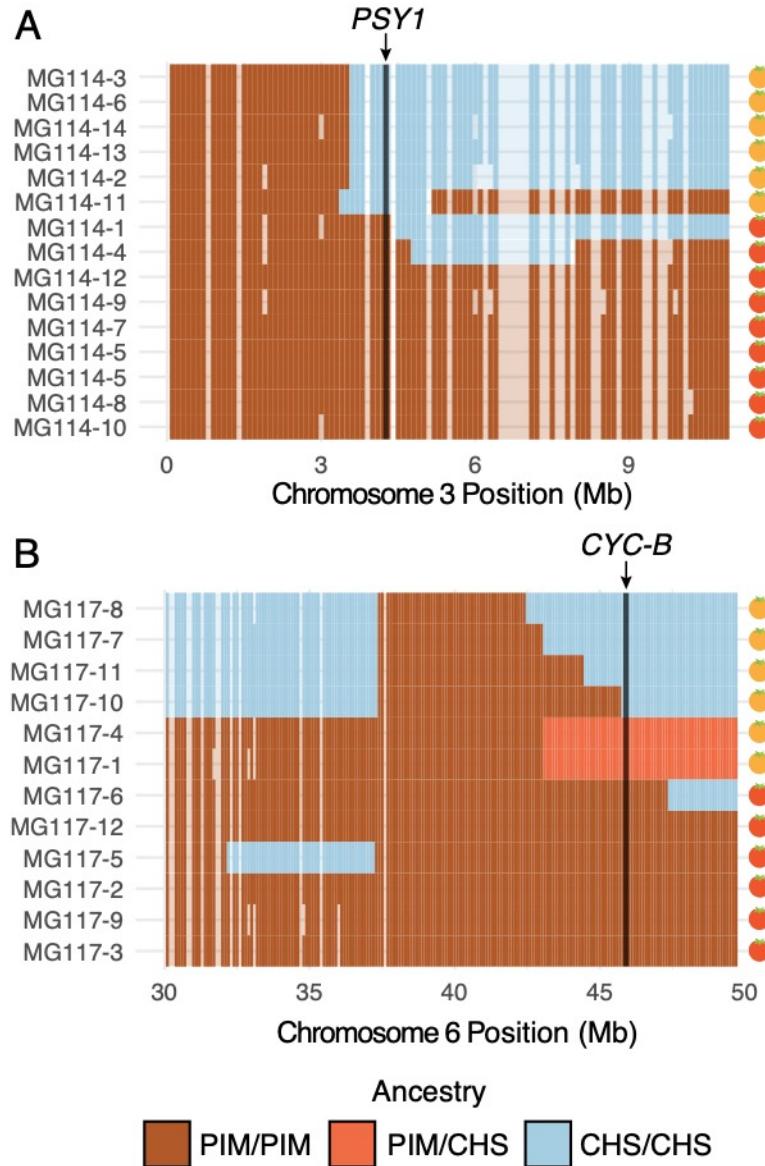


Figure 5: Patterns of local ancestry across focal chromosome regions of MG114 and MG117, enlarged to show variation in introgression block break points at color pathway genes. (A) CHS ancestry at carotenoid biosynthesis gene *PSY1* on chromosome 3 correlates with observed fruit color variation in MG114. (B) CHS ancestry at carotenoid biosynthesis gene *CYC-B* on chromosome 6 correlates with fruit color variation in MG117. Each cell represents 100kb. Empty cells indicate windows with no sequence data. Empty cells are ghost shaded with each ancestry color based on neighboring assignments.

384 **An introgressed origin for orange fruits in PIM**

385
 386 Because introgression from CHS into PIM was most evident in PIM populations that
 387 were polymorphic for fruit color (MG114 and MG117), we took an admixture mapping
 388 approach to investigate whether introgression influences fruit color variation in these
 389 populations. Specifically, we examined the association between local ancestry across
 390 the genome and observed fruit color phenotypes, paying special attention to genomic
 391 locations of 8 known genes involved in carotenoid biosynthesis (*SI Appendix, Table*

392 S14, S15; Paran & van der Knaap, 2007). For MG117, we found that the presence of
393 orange fruits correlated perfectly with CHS ancestry at only one gene: CYC-B on
394 chromosome 6 (**Figure 5B**; *SI Appendix, Table S15*). This association is significant
395 based on a χ^2 test of independence ($\chi^2 = 8.33$; df = 1; P = 0.0039). CYC-B is a lycopene
396 beta cyclase and the specific locus known to underlie the lighter (orange) colored fruits
397 observed in the endemic species (Stommel & Haynes, 1994).

398 In contrast, the similarly-sized chromosome 6 haplotype in MG114 (**Figure 4B**)
399 does not include the CHS allele at the CYC-B locus. Instead, in MG114 we find that the
400 presence of orange fruits was solely predicted by CHS ancestry at PSY1 on
401 chromosome 3, the first enzyme in the carotenoid fruit color pathway (*SI Appendix,*
402 *Table S14*; $\chi^2 = 11.12$; df = 1; P = 0.0009). Although the role of PSY1 in the coloration
403 of endemic fruits has not yet been studied, loss of function mutations in PSY1 have
404 been described in LYC and these produce orange fruit color (Fray and Grierson, 1993).
405 To further investigate the association of PSY1 with endemic fruit pigmentation, we used
406 previously published RNAseq data (Pease et al. 2016) to examine variation in PSY1
407 across the entire wild tomato clade. PSY1 is expressed at detectable levels in both
408 endemic species and has a highly conserved coding sequence (1203/1239 shared
409 sites), whose exceptions include a single non-synonymous substitution (62R → W)
410 unique to the endemic clade (*SI Appendix, Fig. S14*). This substitution lies outside of the
411 trans-isoprenyl diphosphate synthase protein domain associated with the enzyme's
412 function, but within the transit peptide signal sequence (residues 1-70), although the
413 specific functional importance of this variant remains to be assessed. Regardless,
414 based on our observed associations between carotenoid biosynthesis loci and fruit color
415 variation, our data support two separate mechanisms underlying the emergence of
416 orange fruits in Galápagos PIM, both of which were likely derived via introgression from
417 CHS.

418
419
420

421 **Discussion**

422
423 Biological invasions are one of the foremost threats to global biodiversity, yet we still
424 have a poor understanding of the processes that contribute to invasive colonization
425 success. Here, we studied patterns of genome-wide ancestry and relatedness between
426 endemic, invasive, and continental wild tomato populations in order to reconstruct the
427 history and consequences of a recent biological invasion on the Galápagos. Invasive
428 populations of *S. pimpinellifolium* (PIM) had low levels of genetic diversity and an
429 excess of rare alleles, and we inferred 2-3 recent introductions onto the archipelago
430 from Ecuador and Peru. As a consequence of this invasion, we uncovered evidence for
431 recent and ongoing gene flow between PIM and the congeneric endemic species *S.*
432 *cheesmaniae* (CHS). Local ancestry at two key carotenoid loci further supported an
433 introgressed (CHS) origin for orange fruits in at least two invasive populations.
434 Together, our results reconstruct the history of invasion, and infer that the source of
435 convergent phenotypic evolution in the invasive populations is introgression of important
436 functional alleles from endemic relatives.

437
438 **Genomic data reconstruct the demographic history of invasion onto the**
439 **Galápagos Islands**

440 Our analyses identify 3 independent introduction events, yet only a single event from
441 Ecuador comprises 98% of sampled invasives on the archipelago. The other two
442 introductions—each represented by single plant collections—either did not produce
443 large invasive populations or they result from much more recent introductions which
444 have not yet established broadly. Indeed, our PCA (**Figure 2A**) supports the idea that
445 they are the product of more recent introduction events, as these two collections are
446 more closely related to the mainland PIM populations than plants derived from the
447 primary introduction.

448 For the primary introduction, although we observed variance in source region
449 predictions (*SI Appendix, Fig. S2*), the consensus prediction supports a southcentral
450 origin near Guayas or Los Ríos provinces. Intriguingly, historical data on human
451 migration and trade on the islands also point to this as a likely region for the source of
452 invasive PIM. First, although Ecuadorian colonizers of the Galápagos originate from
453 across the country, one of the earliest and largest bursts of migration coincided with the
454 Tungurahua province earthquake in 1949 (Toral-Granda et al., 2017). This province is
455 geographically central and close to Los Ríos. Second, the vast majority of all trade
456 between Galápagos and the continent occurred—and continues to occur—from
457 Guayaquil, the second largest city in Ecuador (Lundh, 2004; Toral-Granda et al., 2017).
458 The surrounding agriculture regions, which includes Los Ríos, would be the most
459 proximate sources for raw product shipments to the islands. This historical context
460 provides additional support for our genetic inference of a majority southcentral origin for
461 invasive populations.

462 Our inferences also clearly implicate humans as the source of PIM introduction.
463 Our demographic reconstruction points to a recent bottleneck and expansion of PIM on
464 the archipelago (**Table 2**), much more recent than our estimate for CHS. Similarly, our
465 inference that LA0411 (a mainland Ecuador accession) is the product of back migration

467 from the Galápagos underscores the recent and likely substantial human influence on
468 the movement of PIM. We conclude that PIM is most likely the result of a recent,
469 human-mediated expansion on the archipelago. Human introduced species represent
470 upwards of 70% of all alien plant species on the Galápagos (Quiroga, 2018), and PIM
471 has similarly been hypothesized to be the product of a human introduction; however the
472 timing and mode of its introduction—including the role of humans—was not previously
473 known.

474

475 **Hybridization as a consequence of invasion onto the Galápagos**

476

477 One key evolutionary consequence of PIM's introduction onto the Galápagos that
478 emerges from our analyses is its hybridization with endemic congeneric species—
479 primarily CHS. Hybridization has been hypothesized as a mechanism for promoting
480 invasive colonization success, as it could help overcome the adaptive limits that might
481 otherwise be imposed by genetic bottlenecks during the colonization process [the so-
482 called “genetic paradox” of invasion (Allendorf & Lundquist, 2003; Sakai et al., 2001)].
483 These bottlenecks can be especially severe during introductions onto islands (Kolbe et
484 al., 2004; Colautti et al., 2005; Golani et al., 2007). In addition, several factors indicate
485 the high potential for gene flow specifically between the four studied species (CHS,
486 GAL, PIM, LYC), including their very close evolutionary relationships (all are members
487 of the red-fruited *Esculentum* subclade within the wild tomatoes; Pease et al., 2016),
488 and their incomplete reproductive barriers (Rick, 1956; Rick & Bowman, 1961; Rick &
489 Fobes, 1975). Nonetheless, previous analyses based on handfuls of loci provided
490 conflicting evidence for and against the occurrence of gene flow between species
491 presents on the island (Nuez et al., 2004; Darwin, 2009).

492 Our data provide clear evidence for recent hybridization and introgression
493 between all four tomato taxa on the archipelago. Although our focus here is primarily on
494 CHS and PIM, we also find evidence for recent hybridization and/or introgression
495 between CHS and PIM (Santa Cruz), PIM and LYC (Santa Cruz), CHS and GAL
496 (Isabela), and, to a lesser extent, CHS and LYC (Isabela). These patterns suggest that
497 hybridization—both with congeneric endemics (CHS and GAL) and invasives (LYC)—
498 could serve as a source of adaptive genetic variation in invasive PIM.

499 The most prominent signal of gene flow is between PIM and CHS (**Figure 3**;
500 **Figure 4**; **Figure 5**), including clear evidence for both early generation (F1 and F2)
501 hybrid offspring as well as older introgression 4-12 generations in the past. Our results
502 indicate (i) that CHS ancestry is maintained in some PIM populations beyond initial
503 hybridization and (ii) that gene flow is ongoing.

504 The potential consequences of secondary genetic contact are numerous (Wolf et
505 al., 2001; Todesco et al., 2016). While we do not have direct data on relative fitness of
506 hybrids, the persistence of later generation CHSxPIM hybrids indicates they are not
507 immediately selected against. Indeed, the genomes of most admixed PIM (MG114 and
508 MG117) are consistent with a history of secondary contact and gene flow characterized
509 not by strong hybrid incompatibility, but a less restricted exchange of alleles between
510 species. Furthermore, the nonrandom distribution of CHS ancestry throughout admixed
511 PIM suggests that it may be selectively maintained in certain regions of the genome.
512 Instead of observing a heterogeneous set of CHS alleles in the backcrossed genome of

513 PIM, we find that CHS ancestry is enriched on chromosomes 3 and 6, and absent in
514 much of the rest of the genome, in both MG114 and MG117 (**Figure 4B**; *SI Appendix*,
515 *Fig. S12*). Moreover, our local ancestry predictions provide evidence that these
516 introgressed regions may contain key genes responsible for the emergence of orange
517 fruit color in MG114 and MG117.

518

519 **Orange fruit color in island PIM was derived via introgression from CHS**

520

521 A key finding of our analyses is that introgression is likely the source of
522 phenotypic convergence on orange fruits that is observed in invasive Santa Cruz PIM.
523 Orange/yellow fruit color is diagnostic for the endemic species (CHS is typically pale
524 yellow; GAL is typically orange) but extremely rare in PIM. At the genetic level,
525 convergence could be based on three potential sources of variation: ancestrally
526 segregating variation, introgression, or via a *de novo* transition. Of these, ancestral
527 variation is the least likely: The very few described examples of orange fruits among
528 continental PIM are all located in Peru (e.g. Sifres et al. 2007), and none have been
529 reported in the inferred geographic region of origin of this invasion (TGRC passport
530 data; www.tgrc.ucdavis.edu). One goal here was therefore to distinguish between
531 introgressed and novel mutation as the source of phenotypic convergence. We did so
532 by mapping the landscape of introgression throughout the genomes of invasive PIM
533 plants, and evaluating its association with observed fruit color variation, and with loci
534 known to underlie this trait in *Solanum* (Paran & van der Knaap, 2007). With these data
535 we inferred a unique scenario in which phenotypic transitions to orange fruits in two
536 different invasive PIM populations were each derived from introgression at a distinct
537 carotenoid locus: *CYC-B* or *PSY1*.

538 Our data in conjunction with existing experimental evidence indicate that *CYC-B*
539 is the causative locus for orange fruits in MG117. Interestingly, *CYC-B* mutants were
540 first identified as natural allelic variation in the endemic species CHS (Rick, 1956).
541 Introgression of the CHS *beta* allele at *CYC-B* into LYC causes the accumulation of β-
542 carotene in ripening tissues and the production of orange fruits (Stommel & Haynes,
543 1994). Orange fruits segregate as a single dominant gene, and genotypic variation at
544 this locus explains a large fraction of fruit color variation in experimental crosses (Rick,
545 1956; Stommel & Haynes, 1994). Our data show a clear association between CHS
546 ancestry and orange fruit color at this locus (**Figure 5**)—including the observation that
547 individuals heterozygous for ancestry display the dominant phenotype—so we infer that
548 introgression of CHS *CYC-B* into PIM has the same large, dominant effect on fruit color
549 in admixed individuals of this wild species.

550 Unlike *CYC-B*, *PSY1* was first identified in the spontaneous fruit color mutant
551 *yellow-flesh* in LYC (accession LA2997; Fray and Grierson, 1993), and its role has not
552 been directly evaluated in CHS or GAL. Recessive *r* mutants at *yellow-flesh* carry a
553 truncated version of *PSY1* that is unable to convert precursor into phytoene. The
554 resulting fruits accumulate almost no carotenoids and the yellow skin pigmentation is
555 driven primarily by the accumulation of the flavonoid chalconaringenin (Fray and
556 Grierson, 1993). Using previously published RNA-seq data (Pease et al., 2016), we
557 confirmed the expression of *PSY1* in both endemic species and did not detect any
558 truncation or premature stop mutations. Rather, we identified a single non-synonymous

559 substitution (62R→W) within the transit peptide signal domain, found in both endemic
560 species. Disruption of the transit signal sequence may prevent localization to the
561 chloroplast and thus result in a nonfunctional enzyme, although determining the exact
562 role *PSY1* has in endemic—and by extension PIM—fruit coloration would require future
563 functional confirmation. Regardless, *CYC-B* and *PSY1* in invasive orange-fruited PIM
564 are unequivocally derived from CHS, and current functional knowledge of both loci
565 indicate their effects on fruit color could entirely explain observed phenotypic variation in
566 orange-fruited PIM.

567 Finally, the ubiquitous lighter (orange and yellow) fruits of the two endemic
568 species, the appearance of convergence towards endemic-like fruit colors in invasive
569 PIM, and the likely independent recruitment of endemic fruit color alleles at *PSY1* and
570 *CYC-B* in MG114 and MG117, together suggest intriguing evidence that lighter fruits
571 may have a specific selective advantage on the islands. The potential environmental
572 basis of this selection is unknown, however differences in fruit dispersal—including
573 disperser color preference(s) and/or fruit color appärence—on the islands vs the
574 continental mainland, could be a likely mechanism. Alternatively, at least in the case of
575 *PSY1* which likely involves either a full or partial loss of function, orange pigmentation
576 could arise due to relaxed selection, if it is more costly to produce red fruits and they
577 have no specific advantage in island environments. Future field experiments and fitness
578 measurements will help to distinguish among these selective hypotheses.
579

580 **Conclusions**

581 Overall, our results reconstruct a complex and recent history of invasion by wild tomato
582 onto the Galápagos Islands, and highlight the potential importance of gene flow during
583 colonization. Our results also add to an emerging phenotypic convergence literature by
584 highlighting how admixture brought on by anthropogenic change can drive convergence
585 over very short time scales. While the adaptive benefit of orange fruits remains to be
586 evaluated, our finding of two separate molecular mechanisms underlying orange
587 coloration each derived from CHS is highly suggestive that lighter fruit pigmentation is
588 favored in the island environment. This study underscores how the long history of
589 research on the Galápagos Islands continues to enrich our understanding of
590 evolutionary processes in the natural world.
591

592

593

594

595 Materials and methods

596
597 **Population sampling and genotyping.** We sampled leaf tissue from 13 wild
598 populations of invasive and endemic tomato taxa on the three largest islands of the
599 Galápagos archipelago: San Cristobal, Santa Cruz, and Isabela (**Figure 1; SI Appendix,**
600 *Table S1*). Leaf tissue was dried in silica and DNA was extracted using Qiagen Plant
601 Mini Kits (Qiagen, Valencia, Calif., USA). Two double-digest restriction site associated
602 DNA sequencing (ddRAD) libraries were prepared using *PstI* and *EcoRI* enzymes by
603 the Indiana University Center for Genomics and Bioinformatics. Libraries were
604 sequenced across two Illumina NextSeq flowcells (150 bp, paired-end, mind-output).
605 Raw reads were filtered for quality, trimmed of adapter sequence and low-quality bases
606 using *fastp* (Chen et al., 2018), and demultiplexed by individual using the
607 *process_radtags* program in *Stacks* (ver. 2; Catchen et al., 2013). Reads were mapped
608 to the *S. lycopersicum* reference genome version SL3.0 using BWA (Li & Durbin, 2009).
609 Bam files of 140 continental accessions representing the full species range of PIM (*SI*
610 *Appendix, Fig. S1 and Table S2*) were jointly reanalyzed with the new samples in
611 *Stacks*. Mapped reads were assembled and variants were called with the *Stacks*
612 *ref_map* pipeline. Genotype calls made with fewer than 8 reads were removed and
613 subsequently we retained only sites having data for at least 80% of all 306 individuals.
614 For all analyses except diversity/divergence calculations (π , Tajima's D, d_{xy} , F_{ST}) and
615 *Treemix*, we pruned sites in high LD ($r^2 > 0.7$) using *bcftools*. All scripts are available at
616 <http://github.com/gibsonmatt/galtom>.

617
618 **Nucleotide diversity and divergence estimates.** Within-population diversity and
619 divergence estimates across the genome were calculated using the *Stacks* program
620 *populations*. Windowed π (**Figure 1C**) was extracted from *Stacks* output. For pairwise
621 comparisons between the islands, Ecuador, and Peru (right panel **Figure 1C**), we
622 calculate genome-wide pairwise divergence directly from the assembled RAD loci
623 (samples.fa file) using a custom Python script (<http://github.com/gibsonmatt/galtom>). For
624 each pairwise comparison between samples, we count the total number of sequence
625 differences and total number of sites for which both samples have data. This choice
626 allows us to conveniently model patterns of diversity between diploid samples in our
627 introgression HMM using a binomial. We also calculated the average genetic distance
628 from each accession to all Galápagos PIM across polymorphic sites in the R package
629 *adegenet* (Jombart & Ahmed, 2008; **Figure 1A**).
630

631 **Phylogenetic reconstruction.** We inferred a maximum likelihood tree of population
632 relationships (**Figure 1B**) using *Treemix* with no specified migration. The *Treemix* input
633 file was generated from a VCF using a custom Python script
634 (<http://github.com/gibsonmatt/galtom>). We expected abundant phylogenetic discordance
635 both within Galápagos PIM (given its recent divergence from Ecuador) and within the
636 red-fruited tomato clade in general (Pease et al., 2016). To this end, we generated
637 1,000 replicate datasets by sampling (with replacement) 500 SNPs from the full dataset.
638 These trees were visualized using *densitree* as implemented in the R package
639 *phangorn* (Schliep, 2011). Both the set of replicate trees and the consensus tree were
640 rate smoothed using *r8s* ($\lambda = 1$), using the *chronopl* function available in the R package

641 *ape* (Paradis & Schliep, 2019). In addition to *Treemix*, we also inferred a maximum
642 likelihood phylogeny of individual sample relationships using *RAxML* (*SI Appendix, Fig.*
643 S3). We subset our dataset to one individual per population for Galápagos collections
644 and to 15–20 samples per geographic region (Peru and Ecuador) for mainland
645 accessions. We ran *RAxML* using the GTRCAT approximation of the general time
646 reversible model of substitution allowing for rate heterogeneity. 25 alternative runs from
647 distinct maximum parsimony trees were performed, from which we selected the best
648 single tree.
649

650 **Demographic inference.** We modeled bottleneck demographic histories from the site
651 frequency spectrum (SFS) using $\delta\alpha\delta i$ (Gutenkunst et al., 2009), calculating the folded
652 SFS using *easySFS* (<https://github.com/isaacovercast/easySFS>). For population
653 MG114, we down sampled from 30 to 14 chromosomes, striking a balance between the
654 number of segregating sites and levels of missing data in frequency bins. We chose this
655 population because the larger number of samples ($2N = 15$) relative to other PIM
656 populations afforded more statistical power. Prior to calculating the SFS for MG114, we
657 removed regions inferred as introgressed by our HMM, as we found that a large fraction
658 of singleton and doubleton sites in MG114 were shared with CHS (*SI Appendix, section*
659 S6). These sites would be spuriously interpreted as *de novo* mutations derived in PIM
660 post-colonization, thereby biasing our parameter estimates. See *SI Appendix, section*
661 S6 for further discussion of our filtering scheme and its effects on $\delta\alpha\delta i$ estimates. For
662 population MG115, we down sampled to 24 chromosomes. For each population, two
663 models were evaluated: (1) a neutral model of no population size change and (2) an
664 instantaneous bottleneck model. The parameters of the bottleneck model are described
665 in the results. We use an extensive three-step optimization pipeline described in Portik
666 et al. (2017) to explore the demographic parameter space of the PIM bottleneck model.
667 For each successive round, we increase the number of replicate runs (10, 200, and
668 500) and decrease the amount of parameter perturbation (3, 2, and 1-fold). The
669 parameters from the replicate with the highest likelihood in each round are taken as the
670 starting parameters in the next round. To estimate confidence intervals for our
671 parameters, we performed a nonparametric bootstrapping procedure (assuming
672 independence among sites) by sampling randomly from the observed SFS 2,000 times.
673 Since we filter for LD, we consider this method to be appropriate.
674

Time in $\delta\alpha\delta i$ is represented in units of $2N_{ref}$ generations. To convert from these
675 coalescent time values to numbers of generations, we estimate N_{ref} as $\theta/4\mu L$, where θ is
676 the population scaled mutation parameter (estimated by $\delta\alpha\delta i$; 99.95 for MG114), μ is
677 the per-generation mutation rate (assumed here to be 1×10^{-8} muts/bp/generation), and
678 L is the length of queried sequence (5,806,952 bp; estimated from the data as the total
679 number of bases where at least one sample in the focal population had data).
680 Bottleneck and final effective population sizes were similarly converted from relative to
681 absolute values using N_{ref} .
682

683 **Inferring gene flow between contemporary island populations.** We used several
684 independent methods to characterize genetic structure in our dataset. First, we applied
685 *Treemix* (Pickrell & Pritchard, 2012) to assess evidence for broad signatures of
686 admixture among populations. Since we were interested in understanding the history of

687 PIM, we subset our full SNP dataset to exclude LYC populations since the exact
688 taxonomic status of these samples are unclear and do not help address questions of
689 gene flow between PIM and endemic species. Furthermore, we remove PIM
690 populations with fewer than 8 samples. We ran *Treemix* with several values for m (0-8),
691 the migration parameter which determines how many reticulation branches are allowed.
692 Based on the likelihoods provided by *Treemix*, we determined that a migration
693 parameter of 6 was most appropriate. Increasing m led us to infer more intraspecific
694 migration within PIM, but no additional interspecific events could be inferred and the
695 increase in data likelihood was marginal (*SI Appendix, Fig. S7 and Table S9*).

696 We also employed model-based (*fastStructure* ([Raj et al., 2014]) and non-
697 model-based (multi-locus PCA) methods to evaluate genetic structure. We ran both
698 methods using the LD-filtered dataset of 5,767 SNPs, and on each island separately.
699 For each island, we ran *fastStructure* for values of k between 1 and 7. We then chose
700 an appropriate model complexity using the *chooseK.py* script supplied with
701 *fastStructure*. After a value for k was chosen, we evaluated the stability of the ancestry
702 assignments across 5-10 separate runs. The multi-locus PCA was implemented in the R
703 package *adegenet*.

704 For select populations with evidence for admixture, we ran *NewHybrids*
705 (Anderson & Thompson, 2002) to identify any early generation hybrid individuals (F_1 , F_2 ,
706 and backcrosses). *NewHybrids* was ran using the same LD-filtered dataset as in
707 *fastStructure* and PCA (script for converting from vcf to *NewHybrids* format is provided
708 at <http://github.com/gibsonmatt/galtom>). For each *NewHybrids* run, we specify three
709 groups: a parent population A group, a parent population B group, and an admixed
710 group. For the CHSxPIM interaction on Santa Cruz, parent populations were defined as
711 non-admixed MG115 and MG113 individuals. For the CHSxGAL interaction on Isabela,
712 parent populations were MG120C and MG120G. For each population we ran the
713 Markov chain for at least 6,000 iterations. Individual assignments were not sensitive to
714 the choice of priors. Lastly, four-sample D-statistics (Durand et al. 2011) were
715 calculated using the *compD* software package
<https://github.com/stevemussmann/Comp-D>), using *S. pennellii* (LA3778) as an
717 outgroup species.

718 **Introgression analysis.** We implemented a hidden Markov model (HMM) to identify
719 fine-scale genomic signatures of introgression. Although RAD sequencing is not optimal
720 for genome scanning due to lower marker density, recent signatures of introgression
721 should be large and detectable. Nonetheless, we acknowledge the fact that our
722 sequencing methods may not allow for full characterization of the landscape of
723 introgression. In our analysis we leveraged the fact that, in regions of recent
724 introgression, genetic diversity within the destination population (π) should be elevated
725 relative to diversity between the source and destination populations (d_{XY}). In other
726 words, regions of recent interspecific coalescence should resemble the introgressing
727 species more than the destination species. We calculate these divergence/diversity
728 estimates directly from assembled RAD loci (samples.fa file from Stacks).

729 Our HMM featured three hidden states: (A) CHS ancestry, (B) PIM ancestry, and
730 (C) heterozygous CHS/PIM ancestry, for which we used π , d_{XY} , and the mean of π and
731 d_{XY} (m) in nonoverlapping 100kb windows, respectively, to calculate emission
732

733 probabilities. We used three binomial models to obtain these probabilities (see *SI*
734 *Appendix, section S4* for a detailed description of the model). For each chromosome
735 and for each focal comparison, we found the most likely hidden state path for a given
736 sequence of π and d_{XY} using the Viterbi algorithm, controlling for underflow by operating
737 in log-space. Because of the coarse scale (100kb windows) used in ancestry
738 assignment, our HMM struggled to identify smaller genetic signals potentially consistent
739 with introgression. *SI Appendix, Fig. S9* provides a per-window look at how ancestry
740 assignment correlated with patterns of diversity and divergence in MG114. Patterns of
741 diversity/divergence vary substantially even between adjacent windows and thus it is
742 likely that we are not capturing sub-100kb signals of introgression. Nonetheless, the
743 large size of the blocks we do identify allows us to be confident they are the result of
744 introgression rather than ILS.
745
746

747 **Acknowledgements**

748

749 We thank the Galápagos Science Center staff on San Cristóbal for logistic and
750 permitting support, and the Galápagos National Park for assistance locating and
751 sampling endemic populations. On-site field support was provided by Marcelo Loyola
752 and Genaro Garcia. This work was supported by a United States National Science
753 Foundation award (IOS 1127059) to LCM and the Indiana University Brackenridge
754 award to MJSG. All field collections were made with appropriate permits and prior
755 authorization by the Galápagos National Park and Ecuadorian Ministry of Environment
756 (Permits PC-40-18 and PC-72-19).

757

758

759

760

761

762 **References**

763

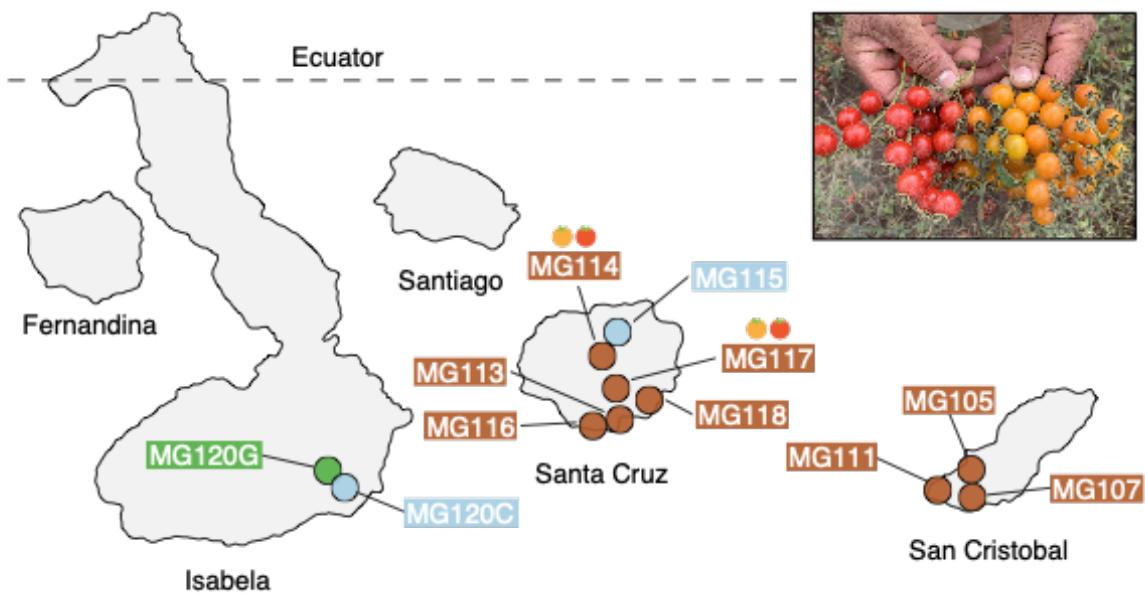
- 764 **Allendorf, F. W., & Lundquist, L. L. (2003).** Introduction: Population Biology,
765 Evolution, and Control of Invasive Species. *Conservation Biology*, 17(1), 24–30.
- 766 **Anderson, E. C., & Thompson, E. A. (2002).** A model-based method for identifying
767 species hybrids using multilocus genetic data. *Genetics*, 160(3), 1217–1229.
- 768 **Battey, C. J., Ralph, P. L., & Kern, A. D. (2020).** Predicting geographic location from
769 genetic variation with deep neural networks. *eLife*, 9.
770 <https://doi.org/10.7554/eLife.54507>
- 771 **Bertolo, L. S., Lima, G. T. N. P., & Santos, R. F. (2012).** Identifying change trajectories
772 and evolutive phases on coastal landscapes. Case study: São Sebastião Island,
773 Brazil. *Landscape and Urban Planning*, 106(1), 115–123.
- 774 **Blischak, P. D., Barker, M. S., & Gutenkunst, R. N. (2020).** Inferring the Demographic
775 History of Inbred Species from Genome-Wide SNP Frequency Data. *Molecular
776 Biology and Evolution*, 37(7), 2124–2136.
- 777 **Cao, W., Li, R., Chi, X., Chen, N., Chen, J., Zhang, H., & Zhang, F. (2017).** Island
778 urbanization and its ecological consequences: A case study in the Zhoushan
779 Island, East China. *Ecological Indicators*, 76, 1–14.
- 780 **Carlquist, S. J. (1974).** Island biology. Columbia University Press, NY.
- 781 **Colautti, R. I., Manca, M., Viljanen, M., Ketelaars, H. A. M., Bürgi, H., Macisaac, H.**
782 **J., & Heath, D. D. (2005).** Invasion genetics of the Eurasian spiny waterflea:
783 evidence for bottlenecks and gene flow using microsatellites. *Molecular Ecology*,
784 14(7), 1869–1879.
- 785 **Comeault, A. A., Wang, J., Tittes, S., Isbell, K., Ingle, S., Hurlbert, A. H., & Matute,**
786 **D. R. (2020).** Genetic Diversity and Thermal Performance in Invasive and Native
787 Populations of African Fig Flies. *Molecular Biology and Evolution*, 37(7), 1893–
788 1906.
- 789 **Darwin, S. C. (2009).** The systematics and genetics of tomatoes on the Galápagos
790 Islands (Solanum, Solanaceae) [Doctoral, UCL (University College London)].
791 <https://discovery.ucl.ac.uk/id/eprint/18994/>
- 792 **Darwin, S. C., Knapp, S., & Peralta, I. E. (2003).** Taxonomy of tomatoes in the
793 Galápagos Islands: Native and introduced species of Solanum section
794 Lycopersicon (Solanaceae). *Systematics and Biodiversity*, 1(1), 29–53.
- 795 **Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011).** Testing for ancient
796 admixture between closely related populations. *Molecular Biology and Evolution*,
797 28(8), 2239–2252.
- 798 **Ellstrand, N. C., & Schierenbeck, K. A. (2000).** Hybridization as a stimulus for the
799 evolution of invasiveness in plants? *Proceedings of the National Academy of
800 Sciences of the United States of America*, 97(13), 7043–7050.
- 801 **Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016).** Is
802 There a Genetic Paradox of Biological Invasion? *Annual Review of Ecology, Evolution,
803 and Systematics*, 47(1), 51–72.
- 804 **Facon, B., Pointier, J.-P., Jarne, P., Sarda, V., & David, P. (2008).** High genetic
805 variance in life-history strategies within invasive populations by way of multiple
806 introductions. *Current Biology*, 18(5), 363–367.

- 807 **Fray, R. G., & Grierson, D. (1993).** Identification and genetic analysis of normal and
808 mutant phytoene synthase genes of tomato by sequencing, complementation and
809 co-suppression. In Plant Molecular Biology (Vol. 22, Issue 4, pp. 589–602).
810 **Gibson, M. J. S., de Lourdes Torres, M., & Moyle, L. C. (2020).** Local extirpation is
811 pervasive among historical populations of Galapagos endemic tomatoes.
812 Evolutionary Ecology, 1–19.
813 **Gibson, M. J. S., & Moyle, L. C. (2020).** Regional differences in the abiotic
814 environment contribute to genomic divergence within a wild tomato species.
815 Molecular Ecology, 29(12), 2204–2217.
816 **Golani, D., Azzurro, E., Corsini-Foka, M., Falautano, M., Andaloro, F., & Bernardi,
817 G. (2007).** Genetic bottlenecks and successful biological invasions: the case of a
818 recent Lessepsian migrant. Biology Letters, 3(5), 541–545.
819 **Gutenkunst, R., Hernandez, R., Williamson, S., & Bustamante, C. (2010).** Diffusion
820 Approximations for Demographic Inference: DaDi. Nature Precedings.
821 **Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos,
822 J. B. (2004).** Genetic variation increases during biological invasion by a Cuban
823 lizard. Nature, 431(7005), 177–181.
824 **Lande, R. (1988).** Genetics and demography in biological conservation. Science,
825 241(4872), 1455–1460.
826 **Lavergne, S., & Molofsky, J. (2007).** Increased genetic variation and evolutionary
827 potential drive the success of an invasive grass. Proceedings of the National
828 Academy of Sciences of the United States of America, 104(10), 3883–3888.
829 **Lee, C. E. (2002).** Evolutionary genetics of invasive species. Trends in Ecology &
830 Evolution, 17(8), 386–391.
831 **Lin, T., Xue, X., Shi, L., & Gao, L. (2013).** Urban spatial expansion and its impacts on
832 island ecosystem services and landscape pattern: A case study of the island city
833 of Xiamen, Southeast China. Ocean & Coastal Management, 81, 90–96.
834 **Long, H., Liu, Y., Hou, X., Li, T., & Li, Y. (2014).** Effects of land use transitions due to
835 rapid urbanization on ecosystem services: Implications for urban planning in the
836 new developing area of China. Habitat International, 44, 536–544.
837 **Lundh JP. (2004).** *Galápagos: A Brief History*. Unpublished manuscript, Human and
838 Cartographic History of the Galápagos Islands. Oslo, Norway.
839 <http://www.galapagos.to/TEXTS/LUNDH1-1.php>
840 **Lynch, M., Walsh, B., & Others. (1998).** Genetics and analysis of quantitative traits
841 (Vol. 1). Sinauer Sunderland, MA.
842 **McMullen, C. K. (1999).** Flowering Plants of the Galápagos. Cornell University Press,
843 NY.
844 **Nuez, F., Prohens, J., & Blanca, J. M. (2004).** Relationships, origin, and diversity of
845 Galapagos tomatoes: implications for the conservation of natural populations.
846 American Journal of Botany, 91(1), 86–99.
847 **Olesen, J. M., Eskildsen, L. I., & Venkatasamy, S. (2002).** Invasion of pollination
848 networks on oceanic islands: importance of invader complexes and endemic
849 super generalists. Diversity & Distributions, 8(3), 181–192.
850 **Paran, I., & van der Knaap, E. (2007).** Genetic and molecular regulation of fruit and
851 plant domestication traits in tomato and pepper. Journal of Experimental Botany,
852 58(14), 3841–3852.

- 853 **Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016).** Phylogenomics
854 Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLoS
855 Biology, 14(2), e1002379.
- 856 **Pickrell, J. K., & Pritchard, J. K. (2012).** Inference of population splits and mixtures
857 from genome-wide allele frequency data. PLoS Genetics, 8(11), e1002967.
- 858 **Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M.,**
859 **Rödel, M.-O., Blackburn, D. C., & Fujita, M. K. (2017).** Evaluating mechanisms
860 of diversification in a Guineo-Congolian tropical forest frog using demographic
861 model selection. Molecular Ecology, 26(19), 5245–5263.
- 862 **Queller, D. C. (2000).** Pax Argentinica. Nature, 405(6786), 519–520.
- 863 **Quiroga D. (2018).** Introduced species and the threats to the Galapagos. In: de
864 Lourdes Torres M, Mena CF (eds) Understanding invasive species on the
865 galapagos, 1st edn. Springer, Berlin, pp 13–2
- 866 **Raj, A., Stephens, M., & Pritchard, J. K. (2014).** fastSTRUCTURE: variational
867 inference of population structure in large SNP data sets. Genetics, 197(2), 573–
868 589.
- 869 **Reatini, B., & Vision, T. J. (2020).** Genetic architecture influences when and how
870 hybridization contributes to colonization. Evolution, 74(8), 1590-1602.
- 871 **Rick, C. M. (1956).** Genetic and Systematic Studies on Accessions of *Lycopersicon*
872 from the Galapagos Islands. American Journal of Botany, 43(9), 687–696.
- 873 **Rick, C. M. (1963).** Biosystematic studies on Galapagos tomatoes. Occasional Papers
874 of the California Academy of Sciences, 44, 59–77.
- 875 **Rick, C. M., & Bowman, R. I. (1961).** Galapagos tomatoes and tortoises. Evolution;
876 International Journal of Organic Evolution, 407–417.
- 877 **Rick, C. M., & Fobes, J. F. (1975).** Allozymes Of Galapagos Tomatoes: Polymorphism,
878 Geographic Distribution, And Affinities. Evolution; International Journal of Organic
879 Evolution, 29(3), 443–457.
- 880 **Rick, C. M., Fobes, J. F., & Holle, M. (1977).** Genetic variation in *Lycopersicon*
881 *pimpinellifolium*: Evidence of evolutionary change in mating systems. Plant
882 Systematics and Evolution, 127, 139-170.
- 883 **Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A.,**
884 **Baughman, S., Cabin, R. J., Cohen, J. E., Ellstrand, N. C., McCauley, D. E.,**
885 **O'Neil, P., Parker, I. M., Thompson, J. N., & Weller, S. G. (2001).** The
886 Population Biology of Invasive Species. Annual Review of Ecology and
887 Systematics, 32(1), 305–332.
- 888 **Sifres, A., Picó, B., Blanca, J. M., De Frutos, R., Nuez, F. (2007).** Genetic structure of
889 *Lycopersicon pimpinellifolium* (Solanaceae) populations collected after the ENSO
890 event of 1997-1998. Genetic Resources and Crop Evolution, 54, 359-377.
- 891 **Stepien, C. A., Brown, J. E., Neilson, M. E., & Tumeo, M. A. (2005).** Genetic diversity
892 of invasive species in the Great Lakes versus their Eurasian source populations:
893 insights for risk analysis. Risk Analysis: An Official Publication of the Society for
894 Risk Analysis, 25(4), 1043–1060.
- 895 **Stommel, J. R., & Haynes, K. G. (1994).** Inheritance of Beta Carotene Content in the
896 Wild Tomato Species *Lycopersicon cheesmanii*. In Journal of Heredity (Vol. 85,
897 Issue 5, pp. 401–404).

- 898 **Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner,**
- 899 S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H.
- 900 (2016). Hybridization and extinction. *Evolutionary Applications*, 9(7), 892–908
- 901 **Toral-Granda, M. V., Verónica Toral-Granda, M., Causton, C. E., Jäger, H.,**
- 902 **Trueman, M., Izurieta, J. C., Araujo, E., Cruz, M., Zander, K. K., Izurieta, A., &**
- 903 **Garnett, S. T. (2017). Alien species pathways to the Galapagos Islands, Ecuador.**
- 904 **PLoS One**, 12(9), e0184379.
- 905 **Vosters, S. L., Jewell, C. P., Sherman, N. A., Einterz, F., Blackman, B. K., & Moyle,**
- 906 **L. C. (2014). The timing of molecular and morphological changes underlying**
- 907 **reproductive transitions in wild tomatoes (Solanum sect. Lycopersicon). Molecular**
- 908 **Ecology**, 23(8), 1965–1978.
- 909 **Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the Risk of**
- 910 **Extinction through Hybridization. Conservation Biology: The Journal of the Society**
- 911 **for Conservation Biology**, 15(4), 1039–1053.
- 912
- 913

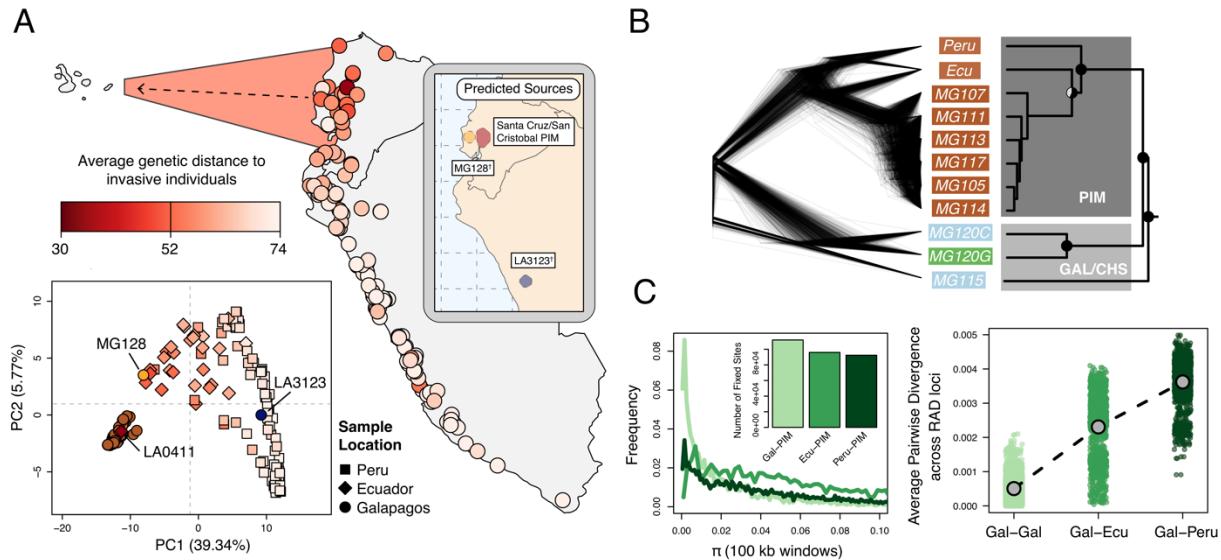
914



915
916
917
918
919
920
921

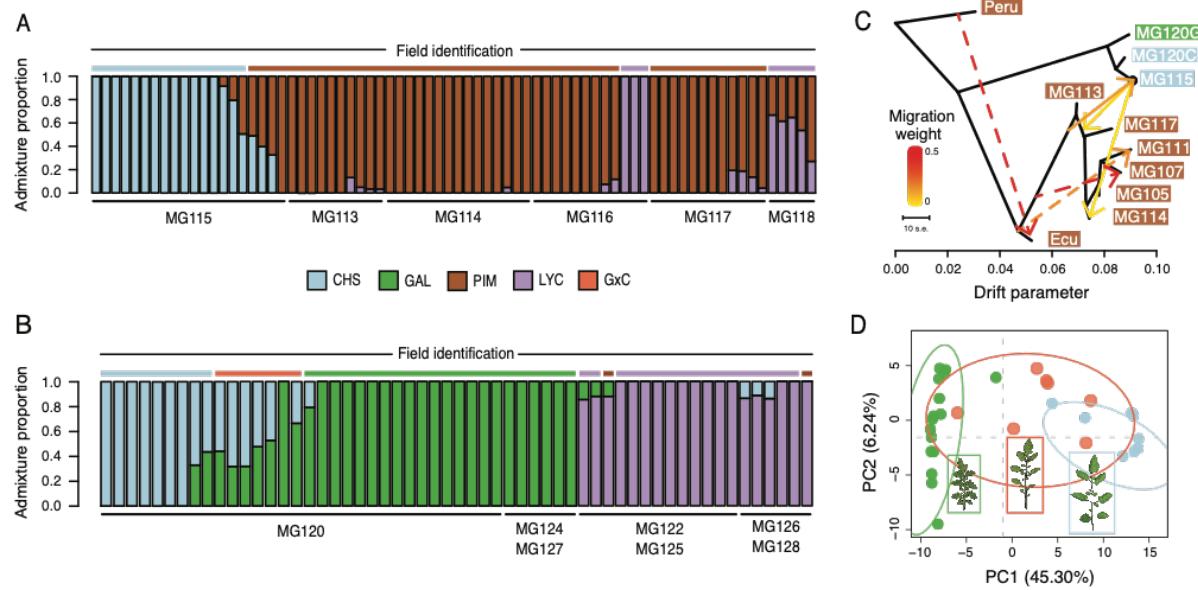
Figure 1: Geographic distribution of focal sampling sites on the Galápagos Islands. *Inset:* Photograph of polymorphic (red/orange) PIM fruits representative of populations MG114 and MG117. For simplicity, LYC populations as well as sampling sites with < 8 individuals are not included here. Refer to Table S1 for a full list of collection localities and sample sizes.

922
923



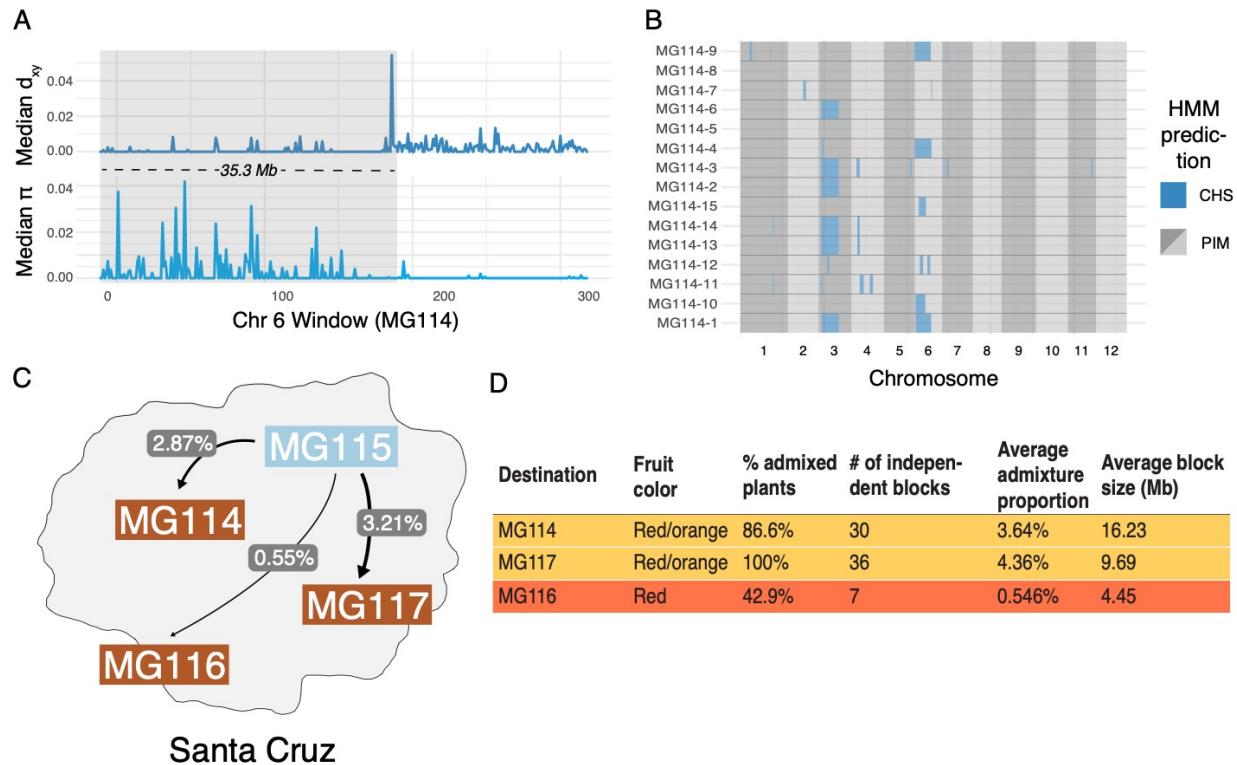
924
925
926 **Figure 2:** Galápagos PIM is the result of a recent invasion from Ecuador. (A) *Map*:
927 average genetic distance between Galápagos-PIM collections and each of the 140
928 mainland accessions. *Plot*: multi-locus PCA. Squares, diamonds, and circles indicate
929 Peruvian, Ecuadorian, and Galápagos collections, respectively. *Inset*: Predicted
930 continental origins for Galápagos PIM collections. Colors are same as shown in the
931 multi-locus PCA ([†]Exact locations vary substantially between runs. Results from a single
932 run are shown). (B) Maximum likelihood relationships among focal populations
933 calculated with *Treemix* (allowing no migration). *Left*: inferred trees of 1000 resampled
934 datasets (500 SNPs, with replacement). *Right*: consensus topology. All trees were rate
935 smoothed ($\lambda = 1$). (C) Diversity and divergence metrics. *Left*: nucleotide diversity (π)
936 calculated for Galápagos-PIM, Ecuador-PIM, and Peru PIM in overlapping 100kb
937 windows. Invariant windows ($\pi = 0$) are truncated and are instead shown in the inset
938 bar plot. *Right*: average pairwise sequence divergence for three PIM comparisons: Gal
939 x Gal, Gal x Ecu, and Gal x Peru. Each point represents a comparison between
940 individuals, averaged over all loci.

941
942
943
944
945
946
947
948
949
950
951

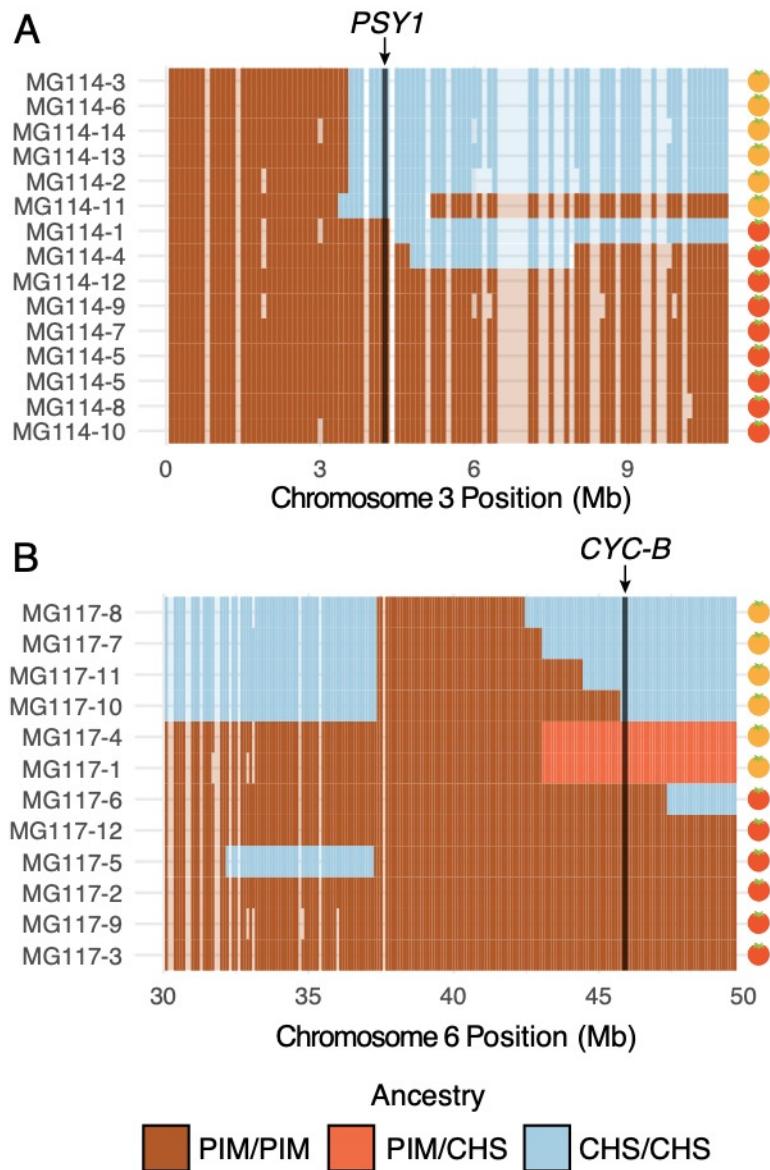


953
954
955
956
957
958
959
960

Figure 3: Patterns of population genetic structure and admixture on Santa Cruz and Isabela. (A) *fastStructure* inference for all Santa Cruz samples ($N = 74$). $K = 3$. (B) *fastStructure* inference for all Isabela samples ($N = 57$). $K = 3$. (C) *Treemix* analysis summary ($m = 6$; $\ln[L] = 395.08$). (D) Principle components analysis for samples at site MG120, a hybrid zone between CHS and GAL.



961
962 **Figure 4:** Local ancestry assignment using HMM characterizes a history of endemic x
963 invasive introgression. (A) Patterns of diversity and divergence along chromosome 6 for
964 an MG114 individual. The region of recent coalescence (low divergence; high diversity)
965 with CHS is annotated in gray. This 20.2 kb block segregates at 20% in MG114. (B)
966 Genome-wide HMM predictions for all individuals in MG114. The x-axis is ordered by
967 chromosome and y-axis is ordered by individual. Two large CHS haplotypes segregate
968 at high frequency on chromosomes 3 (40%) and 6 (20%). (C) Visual summary of
969 admixture proportions from CHS into three PIM populations. (D) Summary of HMM
970 assignment for each PIM population. Populations displaying variation in fruit color
971 (MG114 & MG117) have more CHS ancestry than those which are fixed for the ancestral
972 red state (MG116).
973



974
975
976
977
978
979
980
981
982

Figure 5: Patterns of local ancestry across focal chromosome regions of MG114 and MG117, enlarged to show variation in introgression block break points at color pathway genes. (A) CHS ancestry at carotenoid biosynthesis gene *PSY1* on chromosome 3 correlates with observed fruit color variation in MG114. (B) CHS ancestry at carotenoid biosynthesis gene *CYC-B* on chromosome 6 correlates with fruit color variation in MG117. Each cell represents 100kb. Empty cells indicate windows with no sequence data. Empty cells are ghost shaded with each ancestry color based on neighboring assignments.

983
984
985
986

Table 1: Diversity statistics for focal population samples (H = observed heterozygosity; π = genome wide nucleotide diversity).

Taxa	Population	Island	Endemic	Segregating Sites	H	π
PIM	Peru	N	N	32820	0.00025	0.00094
	Ecuador	N	N		0.00033	0.00130
	MG105	Y	N		0.00011	0.00009
	MG107	Y	N		0.0002	0.00034
	MG111	Y	N		0.00012	0.00010
	MG113	Y	N		0.00015	0.00029
	MG114	Y	N		0.00016	0.00023
	MG116	Y	N		0.00024	0.00024
	MG117	Y	N		0.00016	0.00028
CHS	MG115	Y	Y	8540	0.00023	0.00045
	MG120C	Y	Y	854	0.0001	0.00008
GAL	MG120G	Y	Y	3282	0.00013	0.00014
GALxCHS	MG120GC	Y	Y	2857	0.00023	0.00026
LYC	MG125	Y	N	7219	0.00035	0.00063
	MG126	Y	N	4567	0.00018	0.00052

987
988
989
990

991
992
993
994
995

Table 2: Demographic model estimates for PIM population MG114 inferred using $\delta\alpha\delta i$.
95% CI values were obtained from 2,000 bootstrap replicates of the SFS. Each estimate
is shown in rescaled units (rescaled by N_{Ref} for N_B and N_F ; and by $2N_{Ref}$ for T_B and T_F).

Parameter	Optimum	Bootstrap Median	95% CI
N_B	184.69	286.16	4.49 - 1483.26
N_F	2730.97	977.28	319.12 – 5333.40
T_B	43.12	846.90	91.88 – 7871.41
T_F	240.98	839.13	22.28 – 13591.11
F	0.24	0.02	$2.18 \times 10^{-10} – 0.54$

996
997



Supplementary Information for

Reconstructing the history and biological consequences of a plant invasion on the Galápagos islands

Matthew J.S. Gibson, María de Lourdes Torres, Yaniv Brandvain, & Leonie C. Moyle

Corresponding Author: Matthew JS Gibson

Email: gibsomat@indiana.edu

This PDF file includes:

Supplementary text

Figures S1 to S14

Tables S1 to S16

SI References

Supplementary Text

S1. Field collections and taxonomic treatments

S1.1. Collections and DNA extraction. We visited San Cristóbal, Santa Cruz, and Isabela—the three most populated islands in the archipelago. These islands contain 40 documented TGRC locations (TGRC passport data) and these sites, as well as those described in Darwin (2009) and Nuez et al. (2004), were searched during our expeditions. We looked for plants within a 200m radius of each previously documented collection site, and we also searched for additional undocumented populations. Populations were sampled across linear transects and detailed photographs of leaf, fruit, and flower morphology were taken of all sampled individuals against color and length standards. The latitude and longitude of each collection location was logged using a smartphone. Each site and the identity of species found within are described in **Table S1**. For extracting DNA, 3-5 leaves were sampled from each plant and immediately dried in silica gel. Tissue disruption was performed with mini-pestles and DNA for each sample was extracted using Qiagen Plant Mini kits (Qiagen, Valencia, Calif., USA) at the Galápagos Science Center (San Cristobal, Galápagos, Ecuador). Our sequencing pipeline is described in the main text.

S1.2. Taxonomy. Population identity was determined based on the taxonomic treatments described in Darwin et al. (2003), with the exception of LYC, which we separated into two forms—LYC (domesticated tomato) and LYC var. *cerasiforme* (CER; cherry tomato; Rick, 1956)—to maintain consistency with Nuez et al (2004). Ripe fruit color was the primary trait used in species identification, with the exception of individuals designated as orange-fruited *S. pimpinellifolium*, based on all other characters. We additionally measured nine leaf traits [leaf length, leaf width, terminal leaflet length, leaflet count, interjected leaflet count, petiole length, internode length, leaflet length, and leaflet width previously identified as diagnostic by Darwin (2003)].

S2. LA0411 and our inference of back migration

S2.1. Original collection notes for LA0411. Charles Rick, 1956: "Pichilingue. 200 m. L. *pimpinellifolium* or L. *esculentum* var. *cerasiforme*? Weedy vigorous form growing as weed in corn field N of station buildings. Population of 15 plants found here. 5 had ripe fruit. Long internodes, leaves more elaborate than ? pimp. Few long hairs at growing point like Santa Cruz *pimpinellifolium*. Alternation all = 3 or +/- 3. Flowers small, stigmas all exserted 1 mm or more. Population seems to be entirely uniform. Inflorescence all 1st order, racemose. Fruits large for pimp. = 1.5 cm, red. No vectors seen. Long tailed thick billed black birds seen near patch. These said to eat wild tomatoes near Guayaquil cement factory. Odor = *esculentum*. Anthocyanin = dark".

S2.2. Evidence for back migration. Based on patterns of genomic relatedness to Galápagos and surrounding PIM accessions, we infer that Ecuadorian accession LA0411 (collected in 1956 by Charles Rick) is the product of a back migration from the Galápagos to mainland Ecuador. This accession showed a particularly strong resemblance to Galápagos PIM (average d_{xy} = 0.0006) and was also divergent from neighboring mainland Los Ríos accessions (**Figure S4**). This accession was indeed noted as morphologically similar to Galápagos PIM tomato collections when sampled on

mainland Ecuador in 1956 (TGRC passport data; <http://tgrc.ucdavis.edu>; S2.1, above). This inference has two implications for understanding the history of PIM and other invasive species on the islands. First, it sets an upper bound on the timing of the initial introduction of PIM to the Galápagos as no later than 1956, as it must have already been established there prior to a back migration event. Second, it highlights the substantial connectivity between this region (the putative source of invasive PIM) and the Galápagos in general.

This finding does not affect our analyses of invasive population origins, including our inference that most invasive PIM have an Ecuadorian source. Invasive PIM have high genetic similarity with many accessions in this region of Ecuador (**Figure 2A**). Furthermore, running *Locator* without accession LA0411 produces results identical to those described above.

S3. Further details on evidence for gene flow

We use several statistical methods for detecting gene flow between invasive and endemic populations. Our primary focus was on characterizing admixture between CHS and PIM on Santa Cruz (described in text), however several additional patterns are also worth presenting. These are discussed below.

S3.1. Treemix. In addition to the two cases of interspecific PIM X CHS admixture, three cases of intraspecific admixture within PIM were inferred using *Treemix*: one from Peru PIM into Ecuador PIM, one from the ancestral Galápagos branch to MG107, and one from Ecuador PIM to MG111. It is difficult to interpret the factors responsible for these inferred events, although they may not all reflect true admixture cases. The edge between Ecuador and Peru PIM is most likely a byproduct of our distinction between these groups; in reality the relatedness among mainland samples reflects a pattern of IBD across latitude. The edges leading to MG107 and MG111 could suggest the occurrence of additional minor introduction events nested within the major Ecuador event. Such a scenario might be plausible given the known substantial trade between the Galápagos and central mainland Ecuador, and the occurrence of at least one reintroduction event. However, we have no independent support for such additional events. Increasing m in *Treemix* had the effect of inferring additional intraspecific migration within PIM, but did not infer any additional between species admixture (see **Figure S6** for additional *Treemix* run summaries) and the increase in data likelihood was marginal (**Table S8**). This indicates that the precise parameter choices in these analyses do not change the number of introgression events inferred between invasive PIM and the endemic island species, beyond those that described in the main text.

S3.2. *S. lycopersicum*. In addition to MG118 (population of F1 PIM x LYC plants) described in text, other weaker signals of LYC ancestry in PIM (e.g., in MG113, 114, 116, & 117; **Figure 4A**) may also be the result of post-colonization admixture, however they may also represent latent (unmodeled) population structure within PIM and/or reflect the hybrid ancestry of LYC var. *cerasiforme* (Ranc et al., 2008). Increasing K from 3 to 4 in *fastStructure* swapped the minor LYC ancestry fractions seen in MG113, MG116, and MG117 for a fourth ancestry class. However, the minor LYC component seen in MG114 as well as the 50/50 LYCxPIM MG118 population retained their original LYC classifications. It is difficult to interpret this given the complex

and unresolved history of LYC var. *cerasiforme* (Ranc et al., 2008), however it may indicate (i) the presence of two or more distinct LYC lineages on the Galápagos or (ii) a contribution by *S. galapagense* (GAL; not sampled on Santa Cruz but historically present). The former scenario may be expected if multiple varieties of domesticated seed have been used for cultivation on the archipelago.

Separate from the clear case of CHS x GAL admixture on Isabela at *la laguna de manzanilla* (discussed in the main text), additional signals of admixture between CHS/GAL (MG120) x LYC (MG122 and MG126) were also detected (**Figure 4B**). In MG122, increasing K from 3 to 4 clearly separates the three inferred GAL x LYC individuals into a separate non-admixed sub cluster, suggesting that the inferred GAL ancestry in these samples likely reflects latent substructure within LYC. In contrast, even when increasing K from 4 to 5, the three inferred CHS x LYC samples in MG126 retain their minority CHS component, consistent with a real signal of shared ancestry between CHS and LYC for these individuals. Whether this is the result of gene flow post-colonization of LYC on the islands or a reflection of breeding history (CHS germplasm has been used as a source for certain beneficial crop traits; Rick, 1967; Stommel and Haynes, 1994) is unclear.

S4. Local ancestry assignment with Hidden Markov Models

Our HMM was implemented in R and Python (code available at <http://github.com/gibsonmatt/galtom>). We used binned pairwise sequence divergence in 100kb nonoverlapping windows to define emission probabilities and defined three hidden states (homozygous CHS ancestry, homozygous PIM ancestry, and heterozygous CHS/PIM). We use our HMM to identify regions of recent coalescence between each population.

For a focal individual and in each window we calculated emission probabilities using three binomial models:

$$P_{CHS} = \binom{d_1}{s} \pi^s (1 - \pi)^{d_1-s}$$

$$P_{PIM} = \binom{d_2}{s} d_{XY}^s (1 - d_{XY})^{d_2-s}$$

$$P_{HET} = \binom{d_3}{s} m^s (1 - m)^{d_3-s}$$

where s is the number of sites in the window and d_1 , d_2 , and d_3 are the median number of differences to MG114, MG115, and the mean of d_1 and d_2 , respectively. Transition probabilities (for the CHS into PIM model) were defined as follows:

$$P[CHS - CHS] = ([1 - r] + r) + a$$

$$P[PIM - PIM] = ([1 - r] + r) + (1 - a)$$

$$P[CHS - PIM] = P[PIM - CHS] = P[HET - CHS] = P[HET - PIM] = r \times a$$

where r and a can be interpreted as proportional to the per-window recombination rate and admixture proportion, respectively. We scaled r by a factor t , which can be interpreted as proportional to the time since admixture. Increasing t will cause the HMM to be more likely to switch between states since recombination will have had more time to break up introgressed blocks. We found that our HMM is relatively insensitive to chosen transition probabilities. For all populations we chose to use 0.002, 0.48, and 10 for r , a , and t , respectively, as these produced consistent annotations that agreed with patterns of observed nucleotide diversity. Each focal individual in a population was analyzed separately against each of the individuals in the potential donor population. For example, to analyze introgression from CHS → PIM, divergence to all MG114 and MG115 individuals is calculated and used to define emissions.

S5. On the timing of introgression

The size of introgression blocks contains information about their age. Over time, recombination will break up contiguous blocks into smaller tracts, resulting in a mosaic of ancestry throughout the genome. We can roughly estimate the age of any given block if we assume that an initial hybridization event was followed by subsequent backcrossing to non-admixed parental individuals. Using a simple logarithmic relationship, the expected time t required to arrive at a proportion p of the donor genome after repeated backcrossing is estimated as:

$$\log\left(\frac{1}{p}\right)/\log(2)$$

via Lynch & Walsh (1998). As block size gets smaller, the expected time since hybridization increases.

The distribution of introgression tract sizes and the concordance of break points between blocks within and between populations is indicative of non-independence among blocks. Such a pattern makes any detailed assessment of the timing of these events difficult, and our broad estimate makes several simplifying assumptions. The occurrence of selfing and/or small population sizes may also affect the relationship between block length and age of introgression. Nonetheless, it is clear that many of the observed patterns of CHS ancestry throughout the genomes of MG114 and MG117 plants were derived from very recent (and likely shared) events. We further investigated the degree to which these histories of introgression were shared by comparing the ancestry of each population window-by-window. Specifically, we called windows of shared CHS ancestry in MG114 and MG117 as those where at least one individual in each population was assigned as CHS by our HMM. Across the genome, 250 windows were inferred to be of CHS ancestry in both populations, representing 26.8% of all CHS ancestry in MG114 and 35.4% of all CHS ancestry in MG117. These patterns imply a complex and partially shared history of admixture in MG114 and MG117. This inference is consistent with several of our other analyses, including *Treemix* (**Figure 4**) and D-statistics which also pointed to a shared basis to detected patterns of admixture. At the resolution which we are able to assign local ancestry, our data also firmly indicate that

endemic variation is maintained within PIM beyond the first or second generation of hybridization.

S6. The impact of recent introgression on demographic inferences with $\delta\alpha\delta i$

A recent history of introgression from CHS/GAL into invasive PIM populations could bias $\delta\alpha\delta i$ parameter estimates towards older dates by introducing low frequency alleles that would incorrectly be interpreted as *de novo* mutations derived post-expansion. To examine this possibility, we determined whether rare alleles in MG114 were private (i.e., informative of the time since bottleneck) or shared with CHS (e.g., from introgression). We found that a large portion (52%) of singleton and doubleton SNPs in MG114 were shared with CHS (MG115), suggesting that introgressed variants could substantially affect our parameter estimates of the occurrence and timing of demographic changes within invasive PIM. For this reason, prior to model fitting we removed SNPs within all genomic regions where at least one MG114 individual had evidence for CHS ancestry based on our HMM. Any region inferred to be of CHS ancestry in any individual was removed from the dataset for all individuals. This resulted in 365 sites being filtered from the SFS relative to the original dataset. The results of model fitting using the masked data are reported in the main text while the unmasked dataset results are reported in **Table S16**.

As expected, θ was higher in the unmasked dataset including introgression (108.094) than in the masked dataset (99.95). Accordingly, N_{ref} was also higher in the unmasked dataset (465.36) compared to the masked (430.32). The total time since the bottleneck was inferred to be substantially older using the unmasked SFS (a difference of 1894.45 generations). N_F (recovery population size) and F (inbreeding) were also heavily affected by filtering of introgression regions. The optimal N_F and F were 335.06 and 0.0008, respectively, in the unmasked dataset and 2730.97 and 0.24, respectively in the masked dataset. These results confirmed our expectation, based on allele sharing, that including introgressed genomic locations substantially inflated $\delta\alpha\delta i$ parameter estimates.

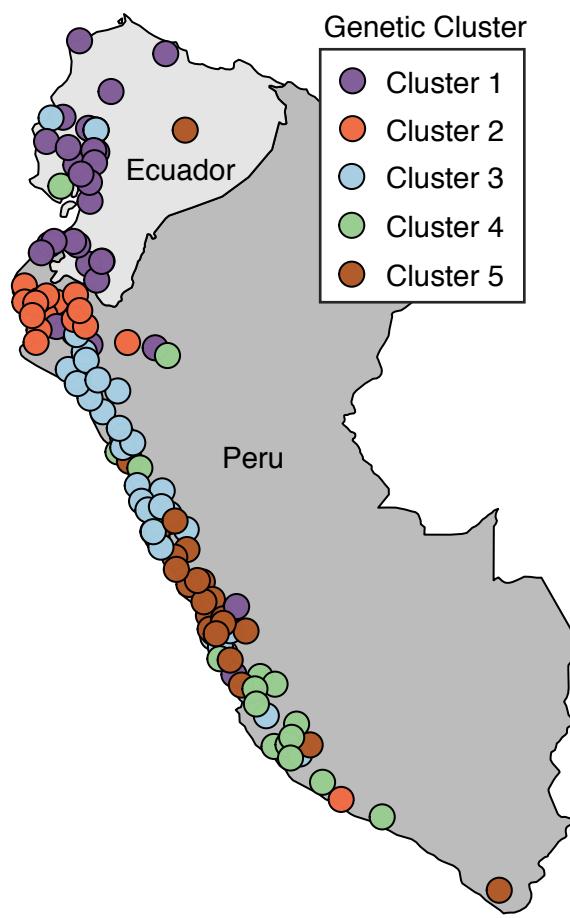


Figure S1: Map of mainland collection sites, colored by genetic ancestry cluster as determined in Gibson & Moyle (2020).

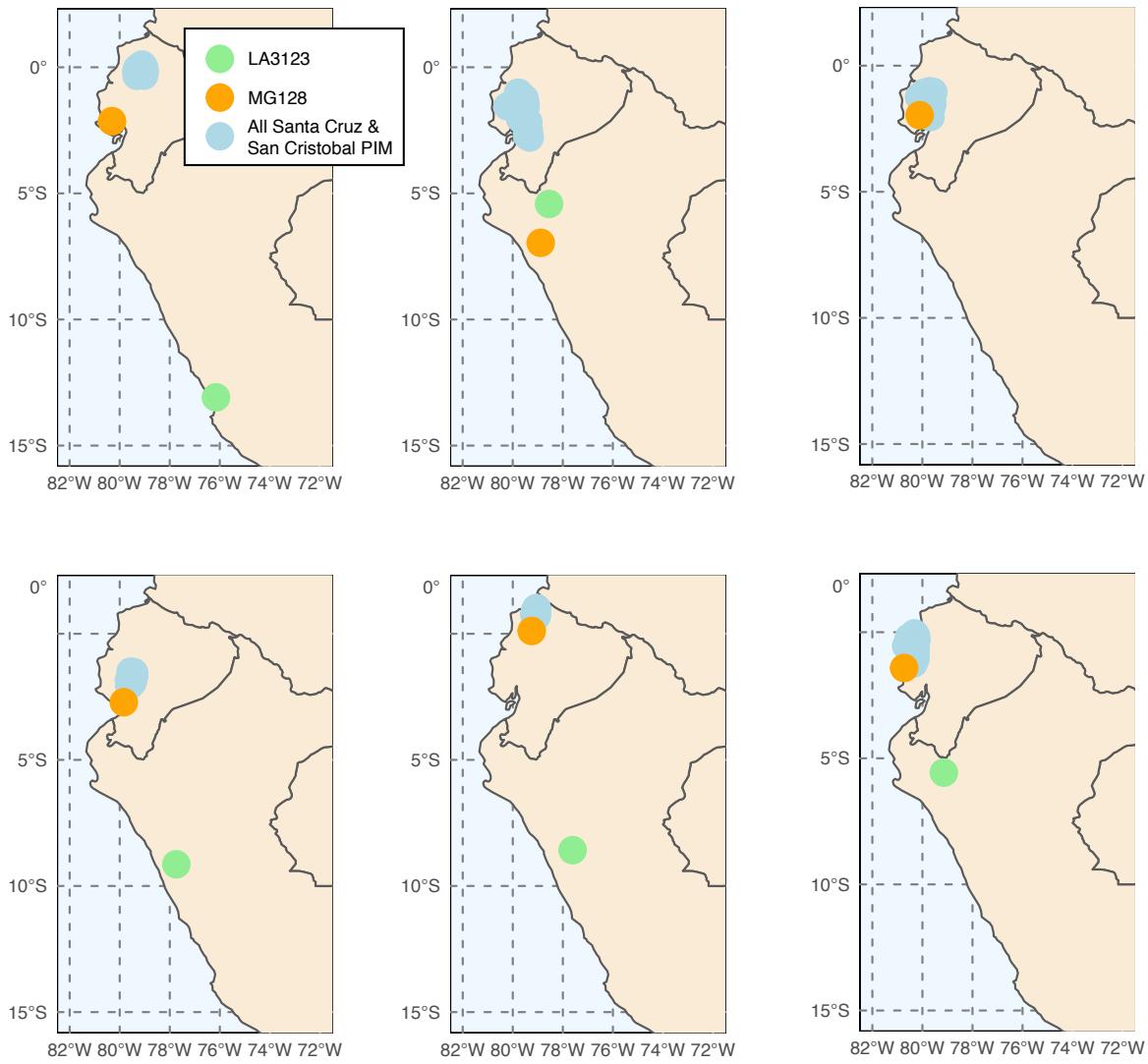


Figure S2: Six runs of *Locator* (Battey et al., 2019) generally support a three-invasion scenario. Exact source localities for MG128-1 and LA3123 varied substantially across run.

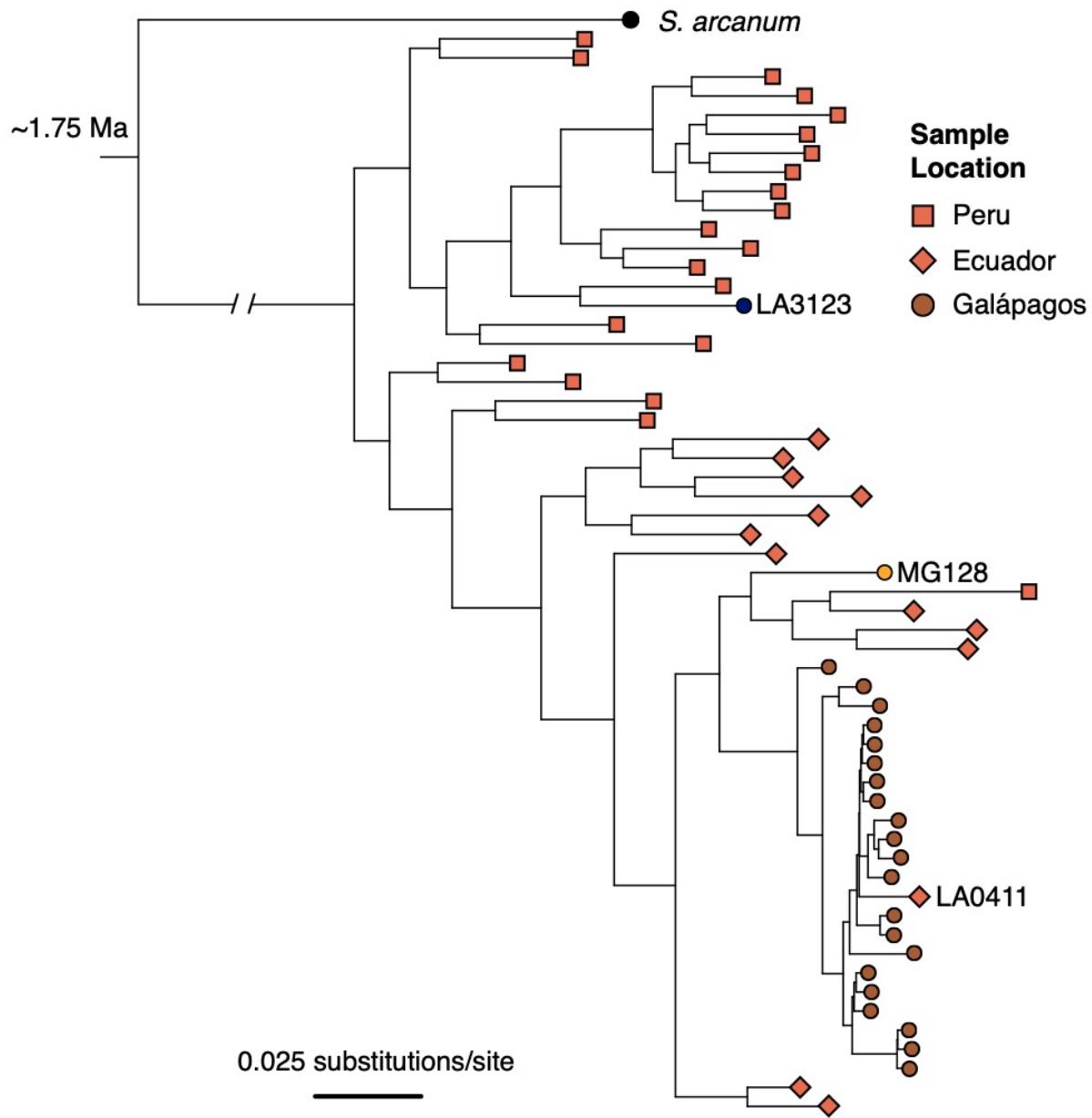


Figure S3: ML tree of individual samples inferred with RAxML, using data concatenated across all RAD loci. Samples were subset by population (for Galápagos collections, 1-2 individuals/population) and by geographic region (for mainland accessions; 20 individuals from Peru, 14 individuals from Ecuador) to limit redundancy and increase computation speed.

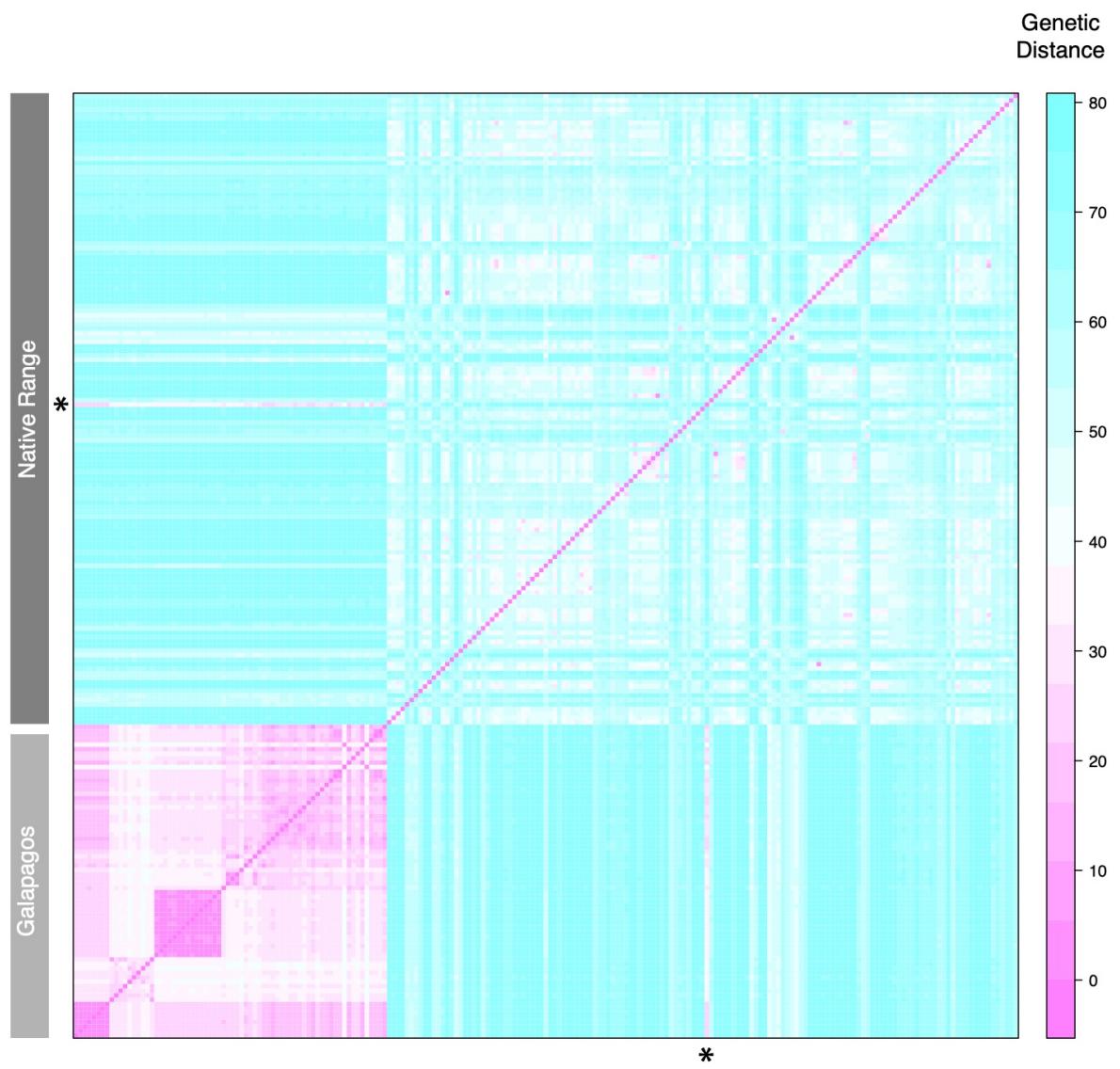


Figure S4: Pairwise genetic distances at all polymorphic loci. Row indicated with an asterisk is LA0411, a sample putatively reintroduced to mainland Ecuador from Galápagos.

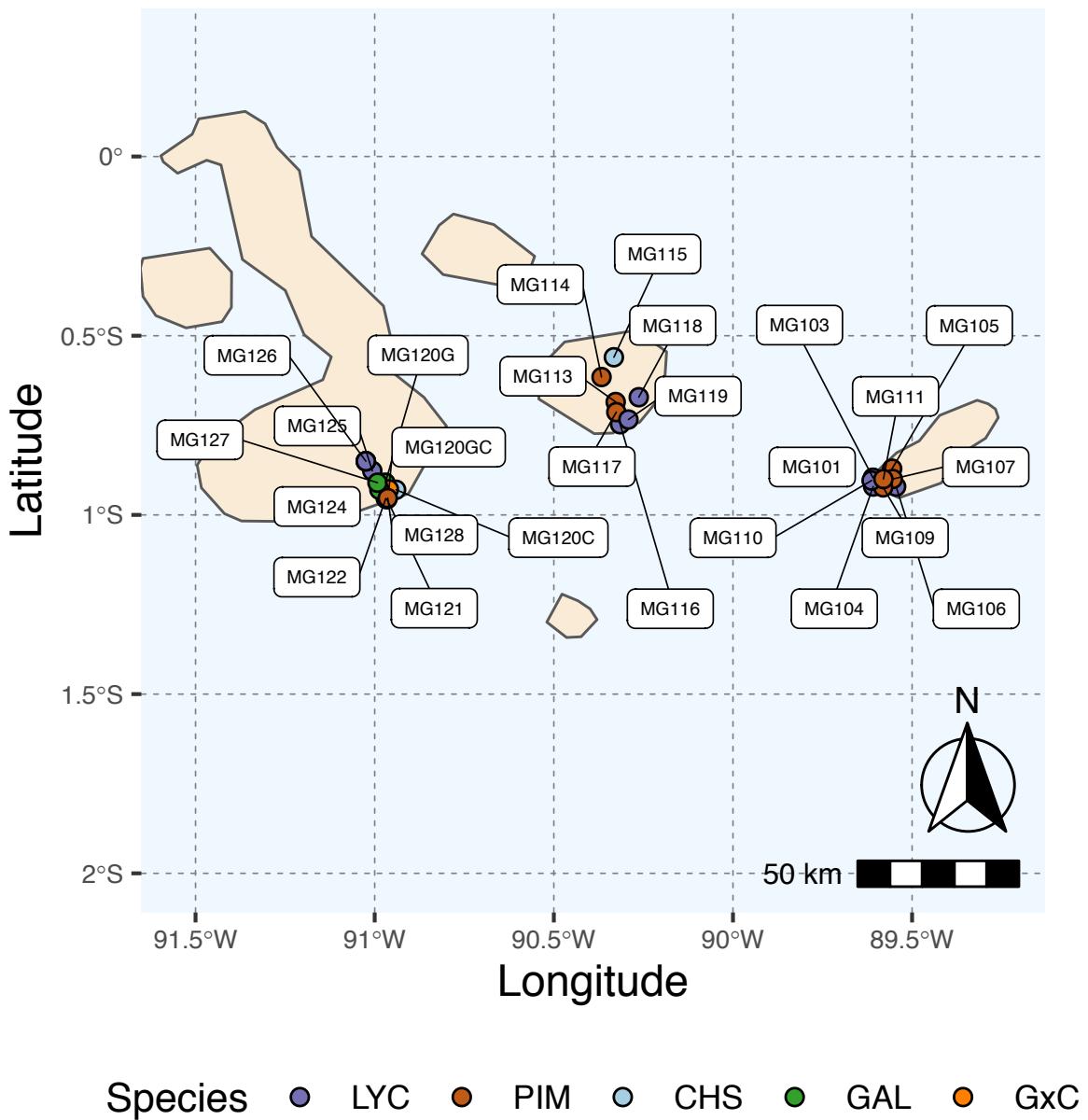


Figure S5: Map of all island collection locations. Refer to Table S1 for details.

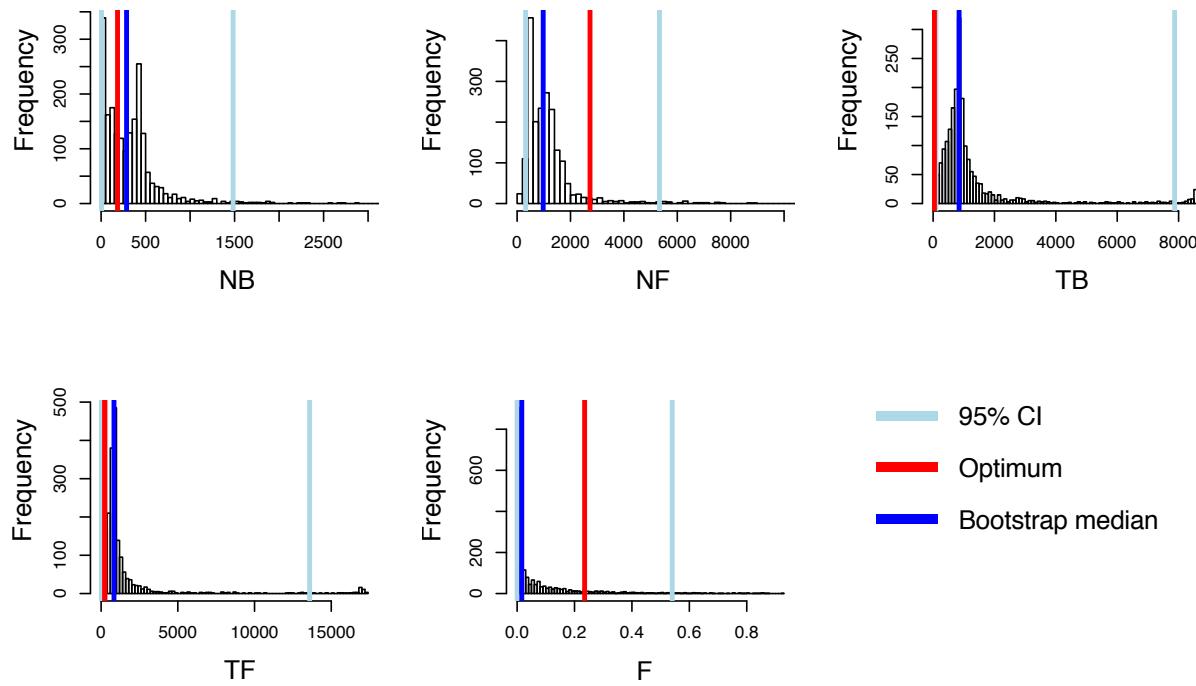


Figure S6: Histograms of bootstrapped parameter estimates from *dadi* for PIM population MG114, using the introgression-masked site frequency spectrum. Light blue bars indicate 2.5% and 97.5% quantiles. Red bars indicate the optimum value inferred. Dark blue bars indicate the bootstrapped median value.

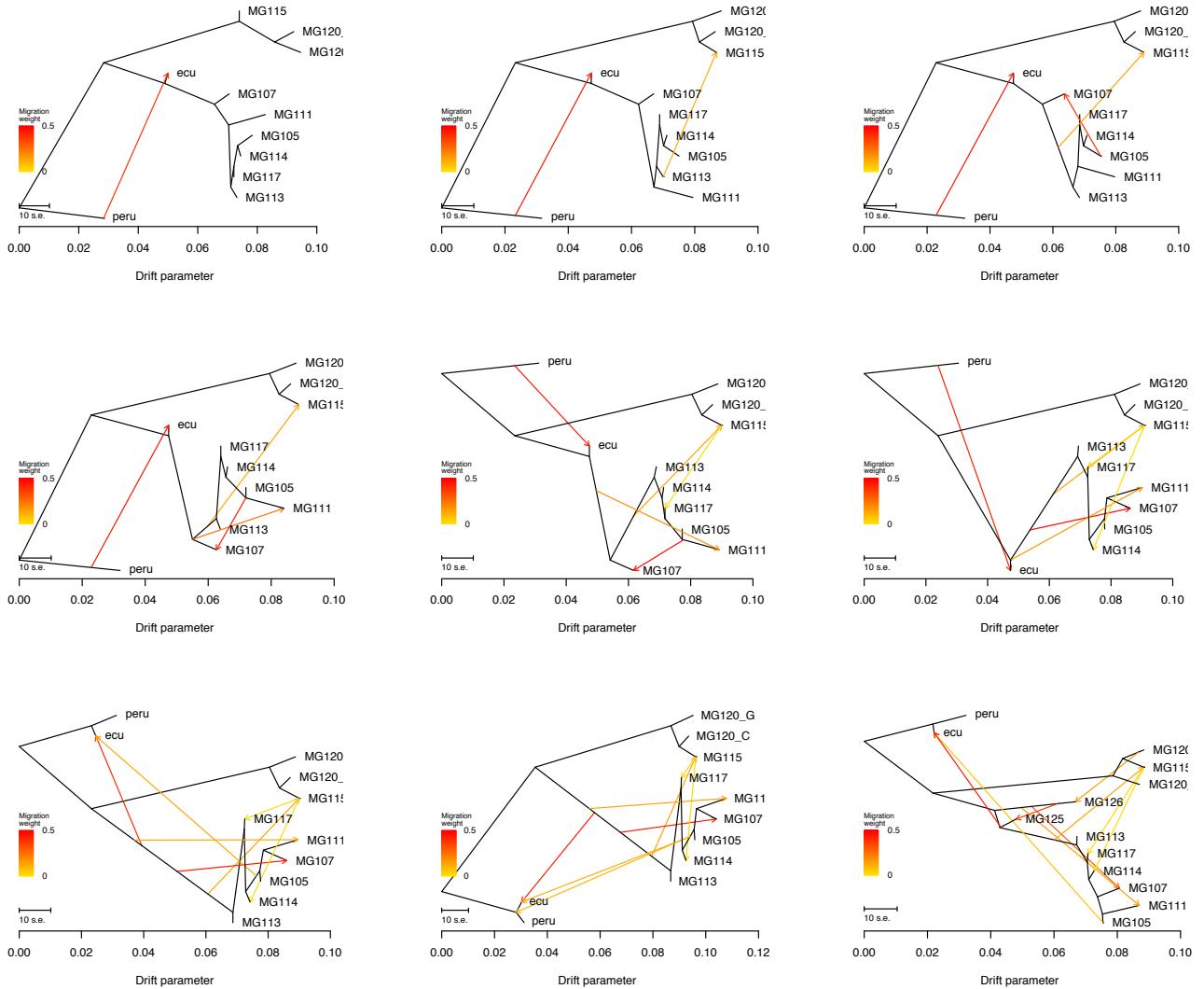


Figure S7: Treemix summary figures for all tested values for m (migration events; 1-8).

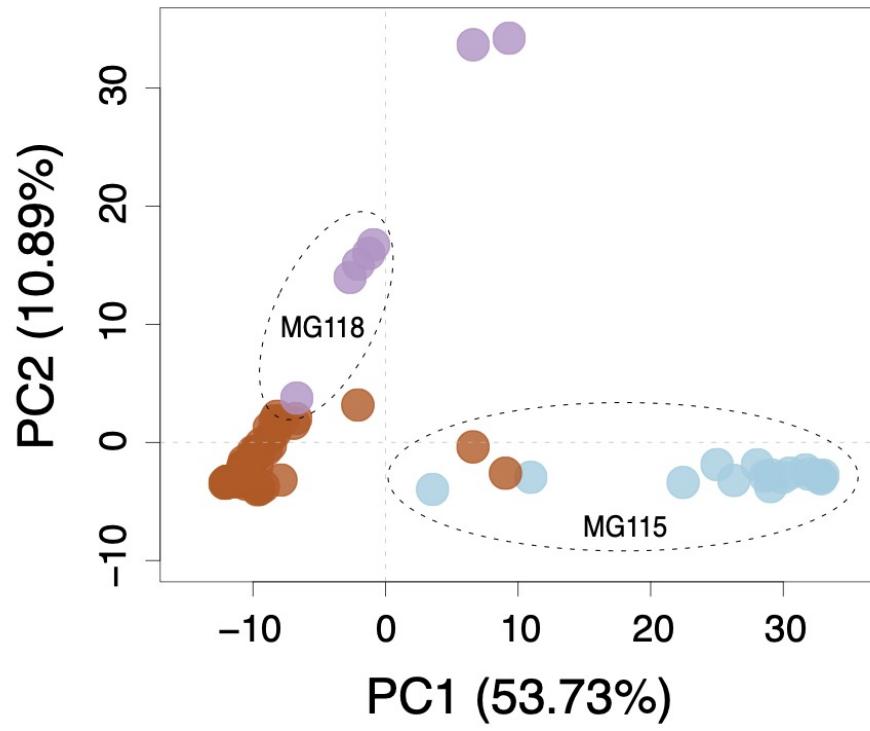
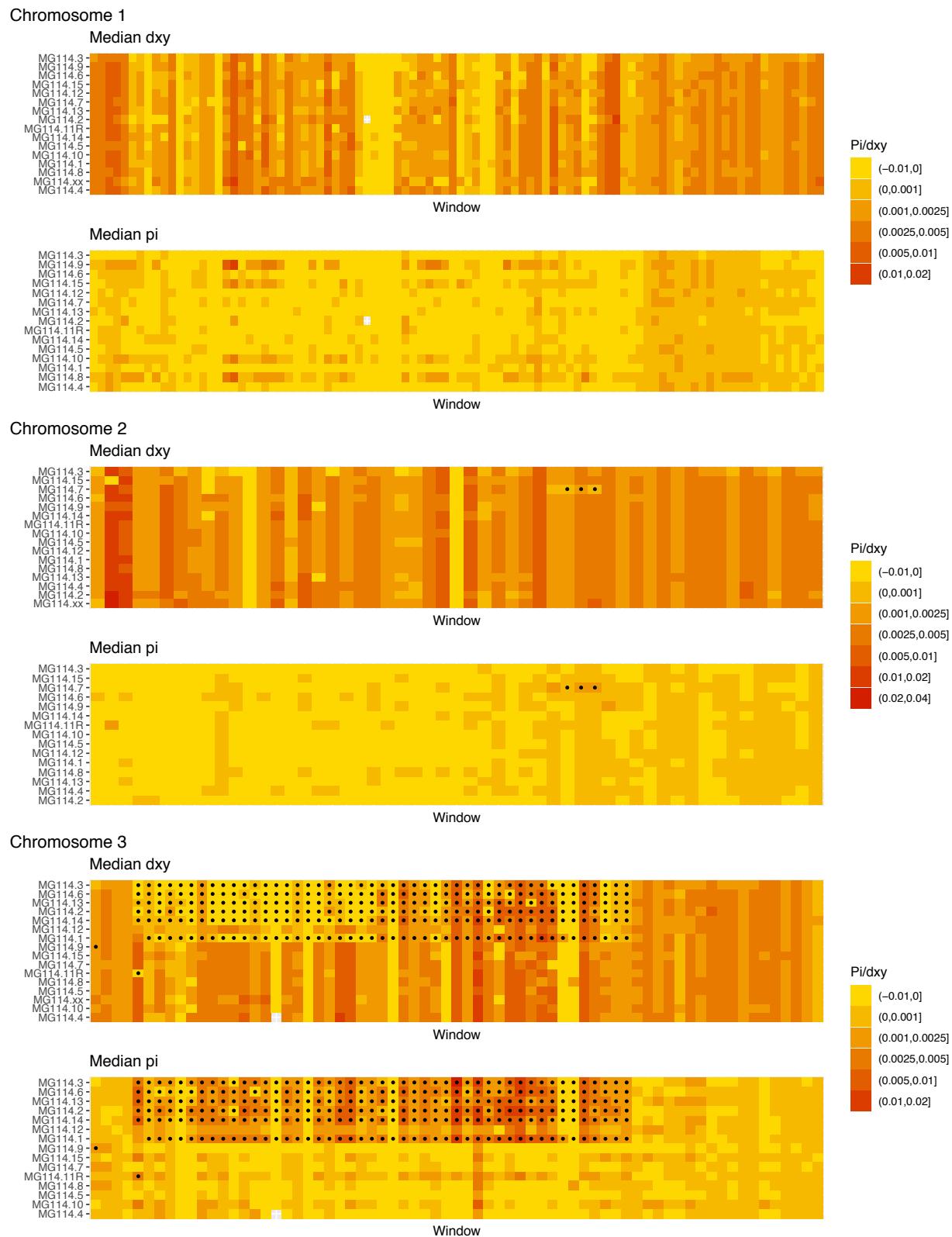
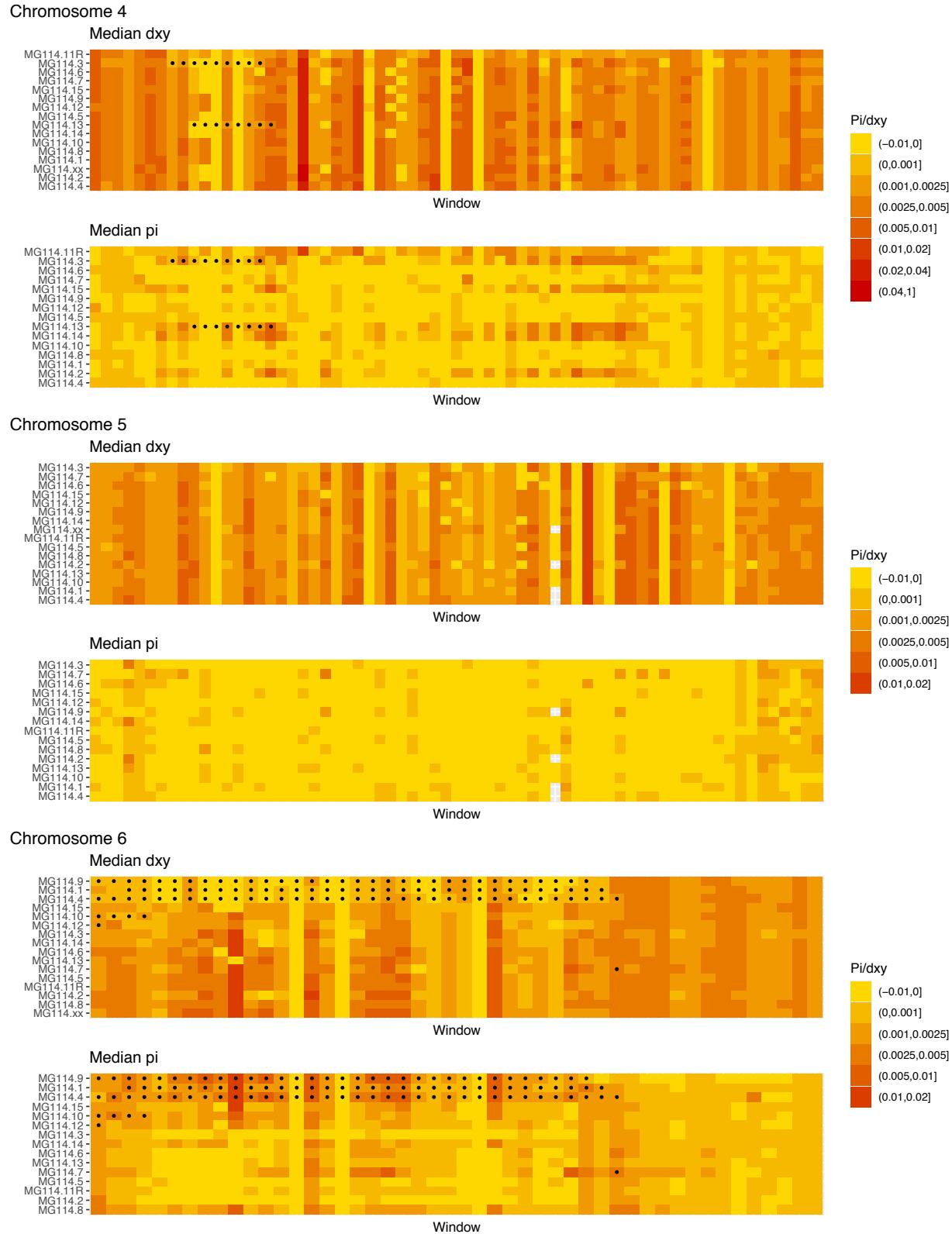


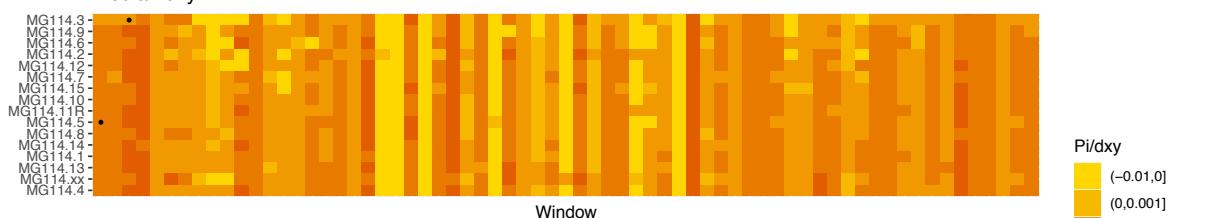
Figure S8: Multilocus PCA for Santa Cruz collections. Brown, purple, and blue points correspond to PIM, LYC, and CHS, respectively.



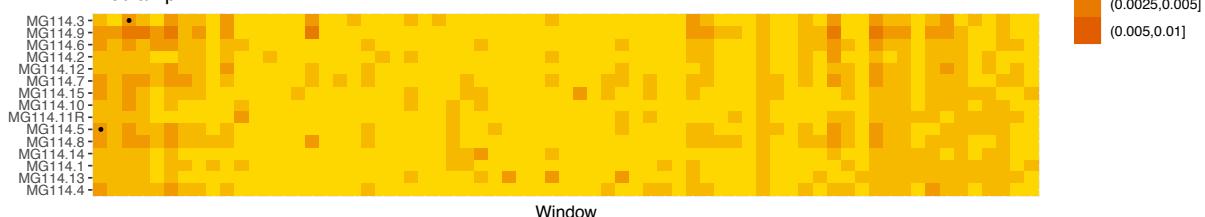


Chromosome 7

Median dxy

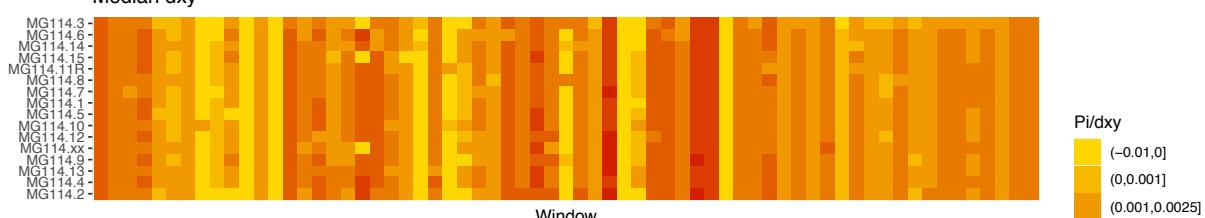


Median pi

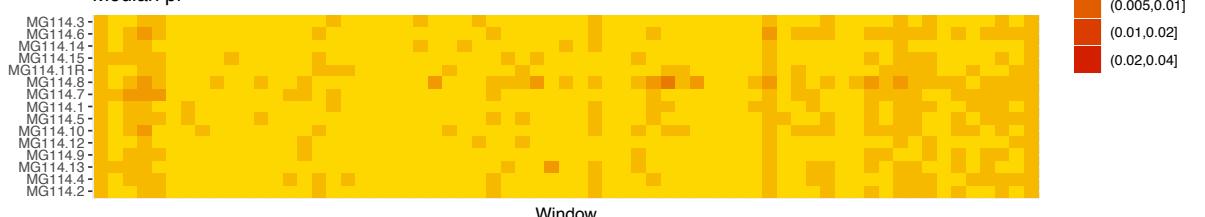


Chromosome 8

Median dxy

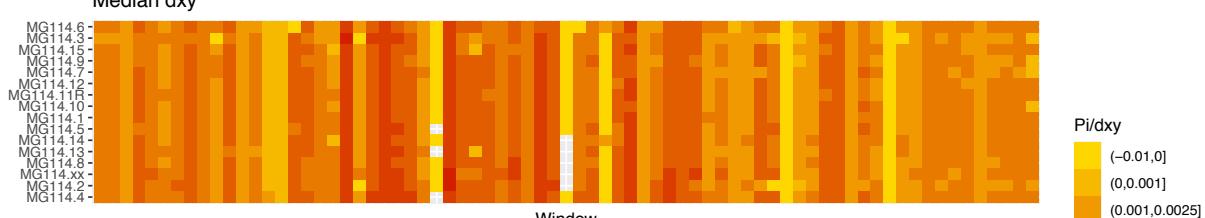


Median pi

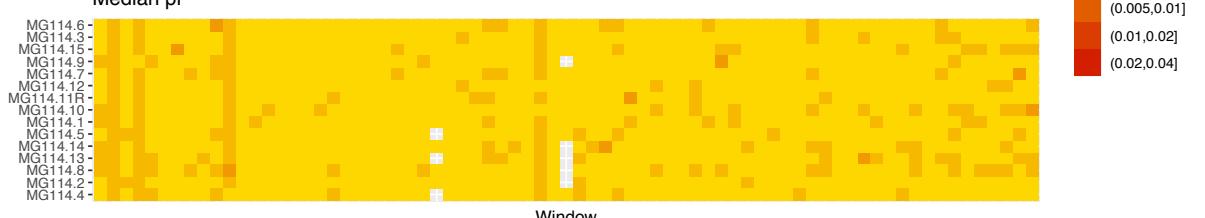


Chromosome 9

Median dxy



Median pi



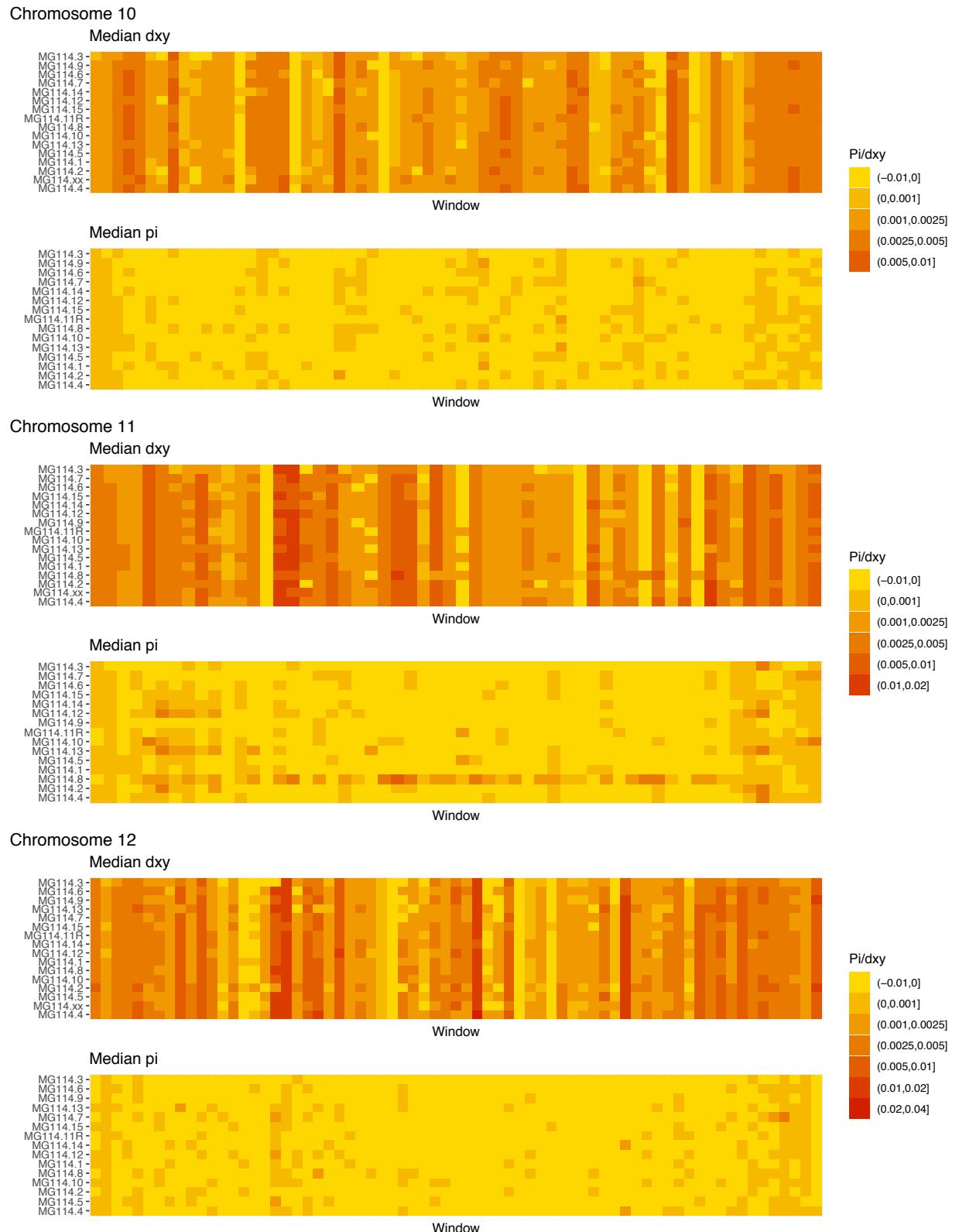


Figure S9: Diversity and divergence across the genomes of MG114 individuals. Each cell represents 1 Mb and colors correspond to median π or dxy (to CHS population MG115). Black dots show regions of CHS ancestry predicted by the HMM. HMM

predictions were done in 100kb windows. Annotations here reflect the consensus prediction of all windows within each 1 Mb cell.

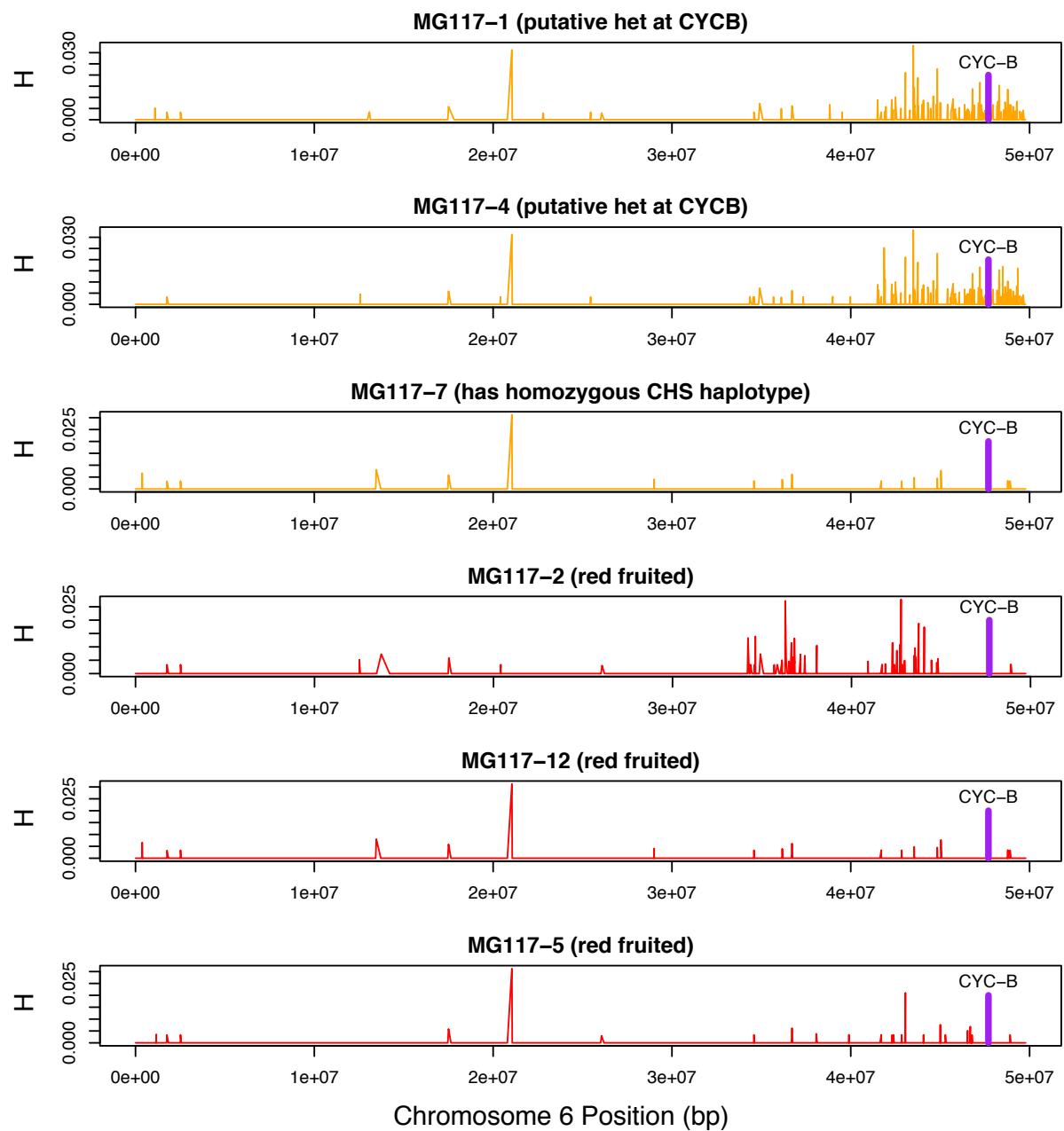


Figure S10: Heterozygosity along chromosome 6 of population MG117. CHS ancestry at CYC-B (dashed line) in MG117-1 and MG117-4 is heterozygous, as shown by elevated heterozygosity estimates at that location in these individuals.

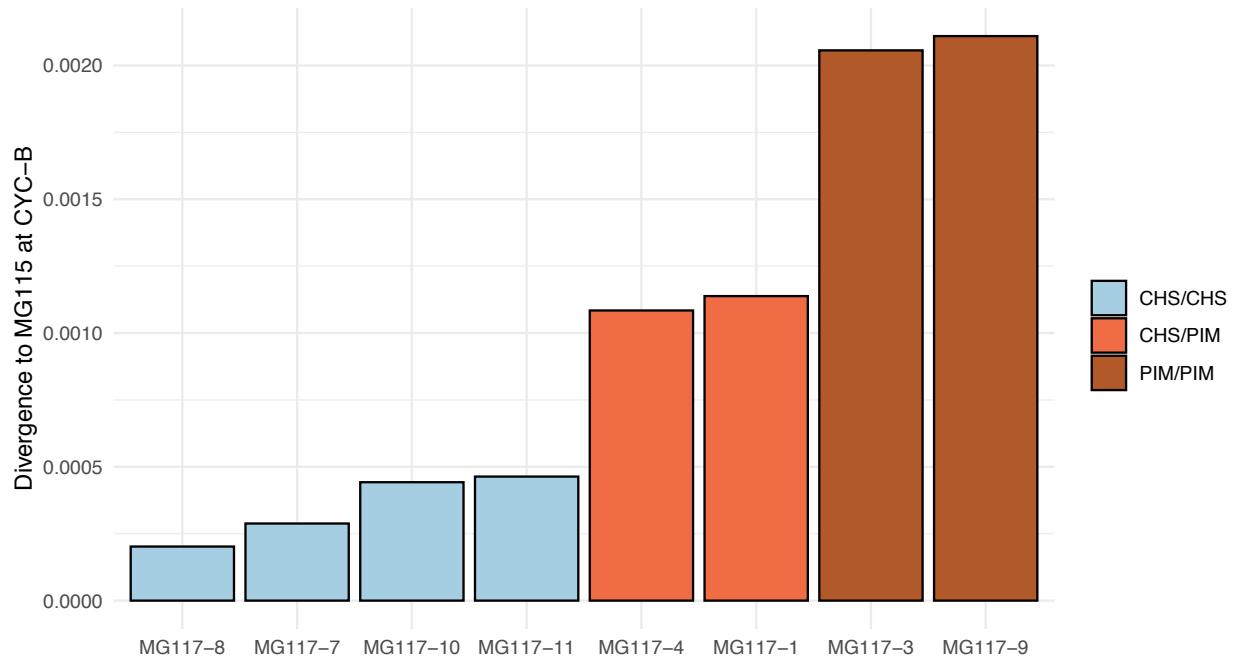


Figure S11: Median divergence estimates between MG117 individuals and MG115 at CYC-B. The HMM correctly classifies intermediately diverged regions as heterozygous.

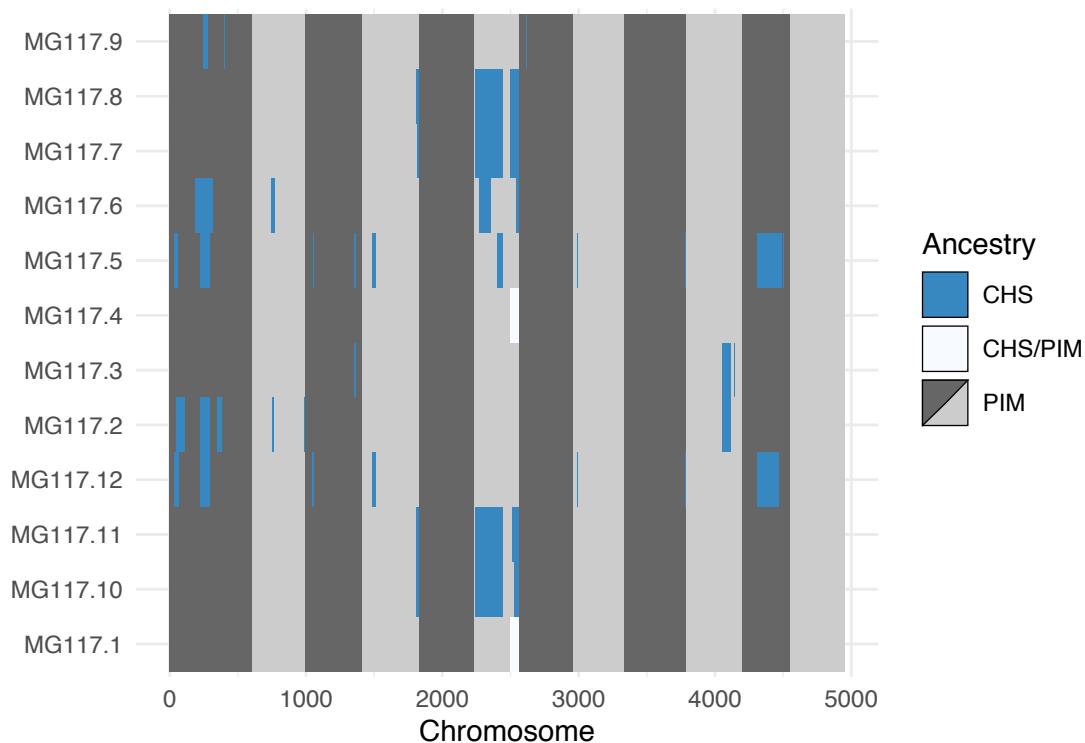


Figure S12: Genome-wide local ancestry in population MG117 as inferred by the HMM.

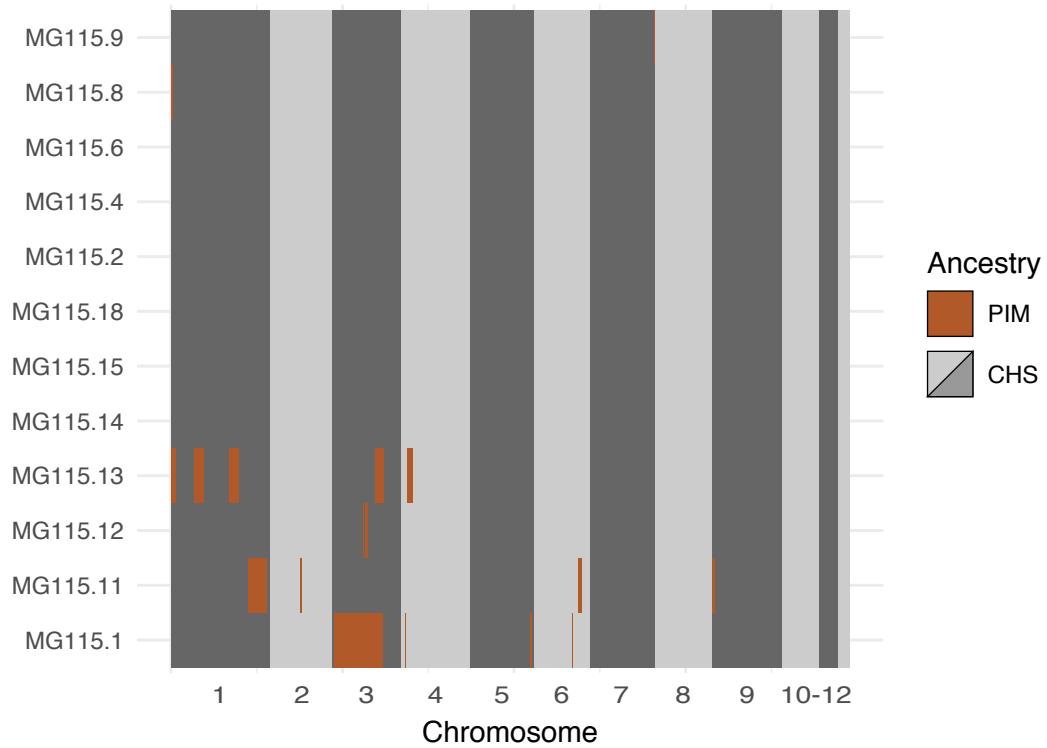


Figure S13: Local ancestry assignment throughout the genomes of CHS plants from population MG115 (Santa Cruz), using MG114 as the PIM reference population. These data represent the opposite direction of gene flow (PIM → CHS) from that shown in Figure 5B (CHS → PIM). Admixture in the PIM → CHS direction was markedly lower than CHS → PIM.

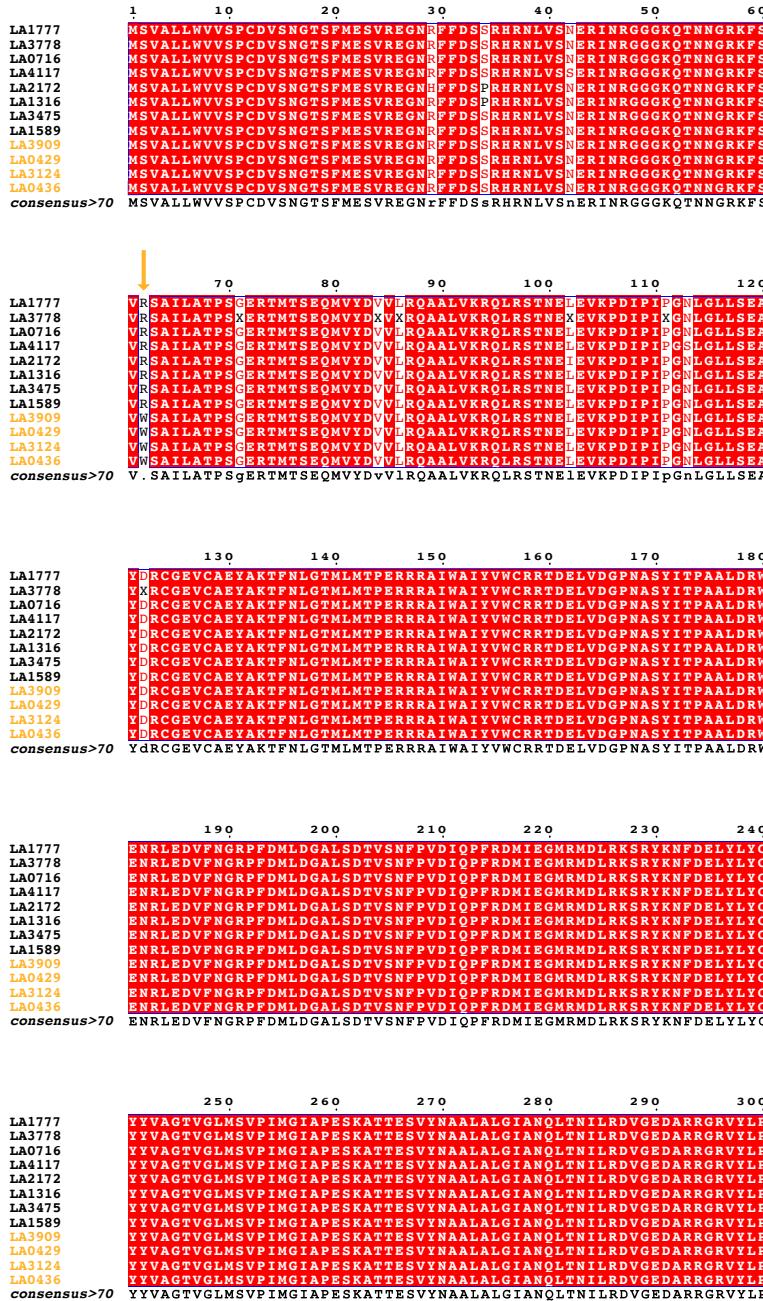


Figure S14: Coding sequence alignment of PSY1 for 9 wild tomato species (12 accessions). Endemic accessions are colored orange. The nonsynonymous substitution which defines the endemic clade is indicated with an orange arrow. Data were obtained from Pease et al. (2016).

Table S1: Full list of population collection sites and their geographic coordinates.

Island	Population ID	Species	Elevation (m)	Lat	Lon	Description
San Cristobal	MG105	PIM	319	00° 54' 28.76" S	89° 33' 20.37" W	El Progresso
	MG107	PIM	319	00° 53' 58.99" S	89° 33' 13.66" W	Soledad
	MG111	PIM	8	00° 54' 04.09" S	89° 36' 36.28" W	Puerto Baquerizo
Santa Cruz	MG113	PIM	259	00° 41' 05.56" S	90° 19' 38.02" W	Bellavista
	MG114	PIM	571	00° 36' 54.89" S	90° 22' 00.64" W	Mina Roja
	MG116	PIM/LYC	10	00° 44' 49.02" S	90° 18' 53.42" W	Puerto Ayora
	MG117	PIM	129	00° 42' 47.28" S	90° 19' 31.81" W	Thomas Berlanga
	MG118	PIMxLYC	264	00° 40' 17.50" S	90° 15' 46.42" W	Cascajo
	MG115	CHS	251	00° 33' 36.23" S	90° 19' 56.65" W	Santa Cruz Hwy
	MG120G	GAL	9	00° 55' 45.81" S	90° 58' 44.25" W	Laguna Manzanilla
Isabela	MG120C	CHS	9	00° 55' 45.81" S	90° 58' 44.25" W	Laguna Manzanilla Road to
	MG124	GAL	12	00° 55' 47.27" S	90° 59' 12.43" W	Santa Tomás Road to
	MG127	GAL	16	00° 55' 23.01" S	90° 59' 20.93" W	Santa Tomás
	MG122	LYC	6	00° 57' 15.21" S	90° 58' 08.41" W	Puerto Villamil
	MG125	LYC	138	00° 52' 37.26" S	91° 00' 23.73" W	Puerto Villamil
	MG126	LYC	332	00° 51' 00.49" S	91° 01' 28.39" W	Santa Tomás
	MG128	PIM	7	00° 57' 11.17" S	90° 57' 48.62" W	Puerto Villamil

Table S2: List of mainland collection sites

Accession	Longitude	Latitude			
LA0121	-78.99556	-8.18444	LA 1586	-78.73	-8.36
LA0122	-78.72	-8.01	LA 1589	-78.74	-8.39
LA0373	-	-9.936944	LA 1598	-78.19	-9.93
LA0375	78.218333	-	LA 1600	-78.04	-10.28
LA0375	-78.21	-9.94	LA 1601	-77.68	-10.67
LA0418	-79.93333	-1.86667	LA 1602	-76.98333	-11.78333
LA 0420	-79.65	-1.06	LA 1603	-77	-11.53333
LA 0442	-78.25917	-9.48167			
LA 0443	-79.48333	-1.1	LA 1604	-77.08333	-11.48333
LA 0753	-79.03333	-8.11667	LA 1605	-76.41667	-13.08333
LA 0722	-76.88333	-12.26667	LA 1611	-76.61667	-12.6
LA 1248	-79.35	-3.98333	LA 1614	-77.36667	-11.11667
LA 1256	-79.61667	-2.66667	LA 1615	-80.63333	-5.23333
LA 1280	-76.71667	-12.03333	LA 1635	-75.06667	-14.63333
LA 1301	-75.91667	-13.73333	LA 1636	-76.08333	-13.45
LA 1381	-79.96667	-5.56667	LA 1637	-75.83333	-13.36667
LA 1416	-	-0.266667	LA 1651	-76.94611	-12.08528
LA 1416	79.333333	-	LA 1659	-77.85861	-9.54667
LA 1428	-	-0.816667	LA 1660	-77.96667	-9.53333
LA 1469	80.216667	-	LA 1661	-76.5	-12.76667
LA 1469	-79.79	-5.86	LA 1670	-70.51667	-17.83333
LA 1470	-79.73333	-6.06667	LA 1683	-81.11	-4.87
LA 1472	-76.25	-13.31667	LA 1684	-80.15	-5.1
LA 1478	-80.08333	-5.21667	LA 1685	-80.6975	-4.88667
LA 1520	-77.11889	-11.04528	LA 1686	-80.62	-5.07
LA 1547	-77.93333	0.58333	LA 1687	-80.62	-5.07
LA 1561	-77.41667	-11.1	LA 1688	-80.375	-4.88333
LA 1562	-76.81667	-12.13333	LA 1689	-80.6175	-5.17639
LA 1572	-76.78333	-11.96667	LA 1690	-80.6175	-5.17639
LA 1576	-76.86667	-12.16667	LA 1697	-76.98333	-12.05
LA 1579	-79.87	-6.59	LA 1728	-75.95	-13.4
LA 1584	-79.79	-6.37	LA 1729	-75.64056	-13.29694
			LA 1921	-75.12722	-14.31194
			LA 1923	-75.28333	-14.66667

LA 1924	-75.21417	-14.62889	PI 365918	-	-2.066667
LA 1933	-74.44583	-15.45639	PI 365958	-74.95	-14.83333
LA 1936	-74.0325	-15.83361	PI 365960	-78.32944	-9.46028
LA 1950	-73.25806	-16.38861	PI 365961	-78.68056	-8.53639
LA 1987	-78.83	-8.43	PI 365962	-78.88806	-8.08611
LA 2069	-	-4.050556	PI 365964	-78.35	-9.18333
LA 2093	-79.88806	-3.62806	PI 365965	-78.15	-9.04167
LA 2102	-79.4675	-4.40167	PI 365966	-77.9	-9.55833
LA 2181	-78.78306	-5.77583	PI 379021	-78.66667	-7.98333
LA 2186	-78.16667	-5.89167	PI 379023	-76.65	-11.85833
LA 2401	-78.22778	-9.50833	PI 379024	-76.65389	-11.88944
LA 2412	-76.71667	-12.48333	PI 379025	-75.15833	-14.68333
PI 251318	-	-5.260833	PI 379026	-80.05056	-5.2525
PI 251319	79.964167	-	PI 379059	-80.1	-1.48333
PI 251320	-79.48333	-1.1	PI 390519	-79.9833	-3.55
PI 251321	-79.48333	-1.1	PI 390688	-77.633	-10.333
LA 2544	-80.26667	-2.33333	PI 390689	-77.733	-10.5
LA 2576	-79.33333	-7.3	PI 390690	-77.5	-1.1
LA 2578	-77.89167	-9.55	PI 390691	-77.766	-10.633
LA 2578	-77.925	-9.525	PI 390695	-79.6166	-7
LA 2649	-80.16667	-5.575	PI 390696	-79.6166	-7
LA 2725	-75.86944	-13.71833	PI 390699	-79.9	-6.68
LA 2832	-75.13333	-14.46667	PI 390700	-79.7	-6.1666
LA 2839	-78.06667	-5.925	PI 390701	-79.91666	-5.6
LA 2852	-80.48333	-0.83333	PI 390705	-80.25	-5
LA 2854	-80.58333	-1.35	PI 390706	-80.75	-5.5
LA 2857	-90.96667	-0.95	PI 390707	-	-5.5333
LA 2915	-79.74528	-5.98472	PI 390707	80.816667	-
LA 2933	-80.5625	-1.44167	PI 390716	-	-
LA 2983	-76.76278	-12.53194	PI 390718	81.066666	4.8833333
LA 3123	-90.38306	-0.62333	PI 390721	-79.96666	-5.266666
PI 365909	-79.15	-0.25	PI 390722	-79.71666	-5.416667
PI 365910	-79.85	0.86667	PI 390728	-80.75	-4.85
PI 365911	-79.96667	-3.45833	PI 390728	-80.51666	-3.616666
PI 365912	-79.36	-3.99	PI 390748	-	-3.55
PI 365914	-79.46667	-1.58333	PI 407533	80.416666	-
PI 365915	-79.53333	-1.55	PI 407534	-76.71	-11.95
PI 365916	-79.51667	-1.81667	PI 407535	-76.83333	-11.98333
PI 365917	-79.63	-2.25	PI 407537	-76.76667	-12.08333
				-79.87	-6.59

PI 407538	-79.89	-6.6
PI 407544	-79	-8.10833
PI 407550	-78.26	-9.49
PI 407551	-78.47	-9.27
PI 407552	-78.56667	-8.925
PI 407553	-78.34	-9.47
PI 407554	-	-
	78.126389	10.067222
PI 503517	-80.18333	-5.16667
PI 503519	-79.98333	-5.175
PI 503521	-80.58333	-4.75
PI 503522	-80.825	-4.90833
PI 503523	-80.625	-5.21667
PI 503524	-80.7	-3.8

Table S3: Stacks ref_map assembly summary

Statistic	Value
Bam records	444,288,642
Primary alignments retained	688,775,808
Records per sample	2,128,095 (17,586-13,836,425)
Assembled/genotyped loci	269,892
Effective per-sample coverage	61.4x (sd = 35.0x)
Mean sites per locus	200.3

Table S4: Stacks per-sample assembly summary

Sample	Bam records	Fraction kept	# Loci	Mean coverage
0121-101	860527	0.974	17034	33
0122-101	1630820	0.972	25079	48.589
0373-102	3329913	0.972	23171	103.581
0375-101	2126539	0.966	32746	54.023
0418-101	2143602	0.978	32138	55.606
0420-101	1719200	0.979	29998	46.226
0442-101	4607456	0.975	46057	102.695
0443-101	2497478	0.973	35378	61.342
0753-101	1107015	0.97	24562	32.976
0772-101	2065633	0.972	27580	57.748
1248-101	2745125	0.981	30936	73.403
1256-101	3632845	0.978	35536	90.728
1280-101	3120897	0.977	30504	83.85
1301-101	2155029	0.978	31797	56.615
1316-1	2454676	0.938	24096	81.476
1316-6	2622282	0.939	26819	82.11
1381-101	2114032	0.974	33684	53.253
1416-101	313218	0.98	13820	13.808
1428-101	2700077	0.982	35638	68.555
1469-101	1457202	0.973	29690	38.898
1470-101	3475525	0.975	35756	86.178
1472-101	12103090	0.971	45961	291.096
1478-101	989959	0.978	19611	34.414
1520-101	4548266	0.976	29953	125.193
1547-101	874210	0.981	23120	27.471
1561-101	2172059	0.978	27342	61.321
1562-101	2288002	0.976	27199	64.904
1572-101	4330309	0.978	35422	110.974
1576-101	4227007	0.968	38571	106.326
1579-101	2739722	0.974	30155	73.004
1584-101	3938382	0.973	37777	95.768
1586-101	3544559	0.972	37577	85.758
1589-12	823915	0.972	17903	30.026
1598-101	4697846	0.977	33154	120.776
1600-101	2096174	0.972	30537	56.59
1601-101	301039	0.963	11510	15.188

1602-101	2736356	0.972	35178	67.97
1603-101	5257707	0.977	37743	130.705
1604-101	2492691	0.946	28257	67.72
1605-101	3183592	0.98	37344	76.445
1611-101	2652137	0.972	30520	72.3
1614-101	2035135	0.978	26826	57.963
1615-101	2245827	0.979	28566	63.405
1635-101	5004334	0.968	41560	119.504
1636-101	1900561	0.976	30965	49.755
1637-101	1664781	0.978	25995	49.517
1651-101	266205	0.98	14669	11.852
1659-101	5447399	0.979	41153	129.37
1660-101	1923801	0.968	26594	56.287
1661-101	2289988	0.979	31529	60.589
1670-101	3133778	0.978	31472	83.181
1683-101	3318632	0.969	35756	81.26
1684-101	3191539	0.975	22414	105.04
1685-101	2731893	0.977	31859	71.467
1686-101	3440511	0.977	35442	85.595
1687-101	4310773	0.979	38098	103.703
1688-101	4800347	0.978	40430	114.266
1689-101	1127857	0.98	21617	36.639
1690-101	2555961	0.976	31373	69.644
1697-101	5430669	0.978	41428	128.326
1728-101	5326058	0.961	21262	186.408
1729-101	5544196	0.98	43872	129.781
1777-27	1692290	0.93	19922	62.233
1921-101	2499661	0.971	31127	67.084
1923-101	3376990	0.974	33440	86.66
1924-101	3200102	0.977	33993	82.095
1933-101	2357305	0.978	32867	61.884
1936-101	13836425	0.975	41433	332.854
1950-101	4561743	0.975	36384	113.567
1987-101	290857	0.968	12644	13.926
2069-101	3869190	0.978	36932	93.812
2093-101	2700416	0.98	37549	66.224
2102-101	2219077	0.974	34680	55.023
2172-15	980024	0.938	19945	34.482
2172-2	2497911	0.943	28183	75.962
2181-101	4189761	0.974	39622	100.017

2186-101	2449823	0.979	32297	64.939
2401-101	2238459	0.972	34308	55.845
2412-101	555326	0.961	16565	21.261
251318-101	2181450	0.977	29419	59.609
251319-101	977057	0.977	23040	30.38
251320-101	3661717	0.974	40335	85.945
251321-101	2735553	0.978	27966	77.639
2544-101	4029122	0.977	36496	100.314
2576-101	2840786	0.971	32821	73.02
2578-101	1676259	0.974	28117	46.642
2649-101	420490	0.964	14742	17.394
2650-101	229441	0.981	9556	13.814
2725-101	4544274	0.978	37961	112.358
2832-101	3536608	0.973	38029	86.212
2839-101	4515964	0.969	36132	116.373
2852-101	1051515	0.973	22886	32.64
2854-101	1761127	0.979	26764	50.474
2857-101	2620451	0.98	37788	63.593
2915-101	3738875	0.976	34888	95.234
2933-101	73476	0.954	6306	6.1
2983-101	4009330	0.978	34339	104.643
3123-101	1426303	0.971	29396	37.907
365909-101	1356728	0.98	27700	37.888
365910-101	3475812	0.982	35251	87.426
365911-101	17586	0.94	3173	2.728
365912-101	2690411	0.98	28908	75.043
365914-101	2912501	0.978	39856	68.799
365915-101	2507336	0.98	36527	61.872
365916-101	1756143	0.978	31283	46.149
365917-101	1106543	0.98	22047	35.6
365918-101	3367378	0.979	42736	76.855
365958-101	4835275	0.975	37191	118.94
365960-101	2534529	0.977	30820	67.633
365961-101	2291525	0.978	29046	64.786
365962-101	3773949	0.975	36760	92.937
365964-101	3201558	0.976	36215	79.441
365965-101	3722221	0.978	28182	104.862
365966-101	1822739	0.979	25301	55.208
3778-5	2004880	0.925	21279	74.21
3778-9	1646601	0.927	20385	63.181

379021-101	1778550	0.977	28344	49.177
379023-101	2888964	0.977	34130	73.825
379024-101	1718608	0.973	25672	50.495
379025-101	4022951	0.978	34063	103.335
379026-101	4340700	0.98	38470	104.786
379059-101	2421589	0.981	31746	64.709
390519-101	1666679	0.977	29865	44.378
390688-101	1752668	0.978	30987	46.157
390689-101	4512953	0.977	38270	110.085
390690-101	2646066	0.976	10851	150.443
390691-101	3381952	0.976	36319	83.565
390695-101	423532	0.976	14109	18.452
390696-101	3692047	0.976	36228	90.892
390699-101	2130632	0.975	31947	55.986
390700-101	3748237	0.975	34294	95.713
390701-101	2286853	0.976	35498	56.239
390705-101	2541073	0.976	38360	60.362
390706-101	5667961	0.979	42438	132.915
390707-101	2150065	0.977	32006	55.456
390716-101	649115	0.978	23213	20.37
390718-101	2263338	0.976	34473	55.94
390721-101	1990569	0.978	32194	52.027
390722-101	2426132	0.978	30980	64.961
390728-101	4222460	0.978	36856	105.316
390748-102	1619440	0.978	28684	44.355
407533-101	1313574	0.973	29425	36.275
407534-101	3650109	0.976	36673	88.856
407535-103	2199289	0.976	33381	56.649
407537-101	3609456	0.975	32089	94.558
407538-101	2500804	0.972	33091	64.03
407544-101	1986374	0.972	31221	51.989
407550-101	4712208	0.978	36252	118.624
407551-101	2952304	0.969	33357	75.274
407552-101	2026851	0.978	25959	59.894
407553-101	3324274	0.977	33136	86.48
407554-101	1600623	0.974	30346	42.51
4117A-103	2270802	0.925	21943	82.778
4117A-31	2189807	0.927	24388	73.222
503517-101	41587	0.977	6038	3.769
503519-101	2150172	0.976	30221	57.298

503521-101	6484748	0.979	45894	147.141
503522-101	2816838	0.975	35615	70.381
503523-101	2419634	0.973	31376	63.588
503524-101	1960823	0.982	30638	53.171
MG101-1	725195	0.928	18335	25.537
MG103-1	977295	0.96	18436	35.374
MG103-2	872069	0.941	19843	29.093
MG104-1	1057938	0.958	11116	58.649
MG105-1R	802728	0.958	20614	25.944
MG105-2	1251346	0.955	22039	38.886
MG105-3	913195	0.957	14281	37.933
MG105-4	419832	0.941	14741	16.625
MG105-5	1097887	0.95	18938	37.605
MG105-6	620727	0.928	15240	24.131
MG105-7	1065864	0.94	22244	32.856
MG105-8	1669040	0.946	21116	53.846
MG106-1	1048647	0.935	21094	33.885
MG106-2	2067945	0.951	24944	61.98
MG106-3	1508324	0.919	25277	42.628
MG107-1	1020866	0.949	20784	33.154
MG107-10	1371695	0.95	23053	41.631
MG107-2	1013620	0.954	19659	34.407
MG107-3	1570291	0.943	23608	46.753
MG107-4	1191986	0.947	22672	36.376
MG107-5	1233479	0.949	23085	37.537
MG107-6	1400747	0.951	24138	41.631
MG107-7	1509667	0.943	25277	42.924
MG107-8	817534	0.944	22834	24.648
MG107-9	1247038	0.948	23293	37.524
MG109-1	487019	0.934	16752	17.768
MG109-2	1366389	0.922	23161	40.268
MG110-1	1495047	0.944	19625	50.782
MG111-1	1184637	0.951	22352	36.789
MG111-10	1062784	0.923	19524	35.003
MG111-11	971504	0.938	18317	34.064
MG111-12	941272	0.932	16856	34.805
MG111-13	1611403	0.9	10274	95.883
MG111-14	652155	0.932	13871	27.868
MG111-15	1149410	0.921	20600	37.079
MG111-2	1319828	0.956	21336	42.519

MG111-3	650475	0.938	19195	21.732
MG111-4	1467335	0.951	24297	43.762
MG111-5	975986	0.939	18848	33.374
MG111-6	840759	0.934	17313	30.492
MG111-7	1203553	0.948	18638	42.454
MG111-8	1160326	0.941	17337	42.655
MG111-9	1297324	0.929	20208	42.701
MG113-1	571968	0.772	8039	33.615
MG113-10	1244608	0.938	24388	35.874
MG113-2	653093	0.944	15521	25.727
MG113-3R	300687	0.943	13535	12.897
MG113-4R	843535	0.937	18422	28.902
MG113-5	575110	0.935	16909	20.961
MG113-6R	2593997	0.944	25708	74.595
MG113-7	850539	0.943	18776	28.988
MG113-8	2269982	0.952	23278	68.572
MG113-9	1278430	0.935	22422	38.587
MG114-1	1611257	0.943	25513	45.471
MG114-10	1796885	0.95	25910	51.116
MG114-11R	1613402	0.944	24758	46.904
MG114-12	1507046	0.935	21076	47.935
MG114-13	1342628	0.94	19642	44.421
MG114-14	1317485	0.947	19916	43.755
MG114-15	1591814	0.944	21411	50.567
MG114-2	1300896	0.944	12767	60.66
MG114-3	1723271	0.944	11520	89.564
MG114-4	1603043	0.943	22800	48.25
MG114-5	1228445	0.936	20553	39.003
MG114-6	762486	0.937	17455	27.083
MG114-7	1141877	0.942	16898	41.686
MG114-8	1205965	0.947	24169	35.389
MG114-9	1141136	0.955	15419	45.178
MG115-1	1477117	0.938	23093	44.483
MG115-10	973572	0.944	20389	32.138
MG115-11	1231265	0.941	24113	36.387
MG115-12	1567784	0.942	24255	46.785
MG115-13	2586243	0.963	21691	85.989
MG115-14	1941036	0.957	27676	53.886
MG115-15	2145739	0.961	30369	57.138
MG115-16	2070350	0.961	28224	57.151

MG115-17R	1726535	0.952	24777	50.676
MG115-18	1979915	0.963	24131	60.335
MG115-19	2262119	0.97	17555	85.434
MG115-2	1106841	0.926	24626	31.365
MG115-20	2150328	0.961	28516	58.657
MG115-3	1533423	0.938	24524	44.596
MG115-4	1800051	0.948	23123	54.73
MG115-5	1345188	0.932	23577	39.921
MG115-6	922730	0.931	18308	32.439
MG115-7	1383340	0.938	23713	40.832
MG115-8	685980	0.942	17987	24.383
MG115-9	856138	0.943	22300	26.311
MG116-1	2620413	0.955	27302	73.4
MG116-10	1474196	0.959	23858	44.53
MG116-11	1157341	0.959	24159	34.504
MG116-12	1533917	0.964	24423	46.12
MG116-2	1364474	0.945	23098	40.972
MG116-3	1327137	0.954	20642	44.534
MG116-4	1621290	0.941	23231	48.327
MG116-5	1527120	0.967	24202	45.724
MG116-6	1905794	0.96	27432	52.444
MG116-7	2153685	0.968	23252	67.593
MG116-8	1782314	0.965	24634	52.398
MG116-9	790971	0.957	18616	27.702
MG117-1	2809514	0.961	28967	77.132
MG117-10	1502902	0.958	26639	42.204
MG117-11	1923532	0.95	28226	52.276
MG117-12	1211097	0.959	22834	37.889
MG117-2	3476301	0.96	30638	92.095
MG117-3	1632177	0.964	24771	48.427
MG117-4	2307685	0.956	26581	65.874
MG117-5	1610157	0.954	24949	46.59
MG117-6	1206813	0.951	21453	38.663
MG117-7	1801874	0.946	26204	51.04
MG117-8	1715525	0.953	24281	51.474
MG117-9	2179544	0.958	26274	62.452
MG118-1	1103683	0.941	19515	37.384
MG118-2	1269249	0.955	22429	40.491
MG118-3	1445480	0.956	21846	46.781
MG118-4	1301649	0.954	23723	39.913

MG119-1	1354403	0.945	23381	41.215
MG120-1	1317924	0.953	22436	41.413
MG120-10	1063381	0.952	17837	38.961
MG120-11-1	1041049	0.943	18670	36.437
MG120-11-2	1009798	0.97	20749	33.949
MG120-13	982881	0.959	20783	31.933
MG120-15	1270578	0.964	23248	39.201
MG120-17	2247542	0.967	23367	70.732
MG120-18	2030229	0.967	24802	61.665
MG120-19	2262015	0.973	13697	106.883
MG120-2	886664	0.944	19009	30.448
MG120-20	2718318	0.968	18435	104.637
MG120-21	2427974	0.96	19247	85.343
MG120-22	1318297	0.923	16698	49.39
MG120-23	1374128	0.97	12017	72.572
MG120-24	2036867	0.966	19389	72.398
MG120-25R	1643420	0.968	21039	54.345
MG120-26	1708398	0.97	19285	60.138
MG120-27	1396758	0.959	19094	49.861
MG120-28	2766644	0.967	21790	91.26
MG120-29	2761259	0.965	24255	85.876
MG120-3	825272	0.903	18024	28.046
MG120-30	2487753	0.956	21644	80.888
MG120-31	2521148	0.952	15268	111.736
MG120-32	3240975	0.966	25778	95.885
MG120-33	2757052	0.964	25710	81.585
MG120-34	1868699	0.963	24014	57.217
MG120-35	1630391	0.961	12558	85.032
MG120-4	1228342	0.949	18728	43.217
MG120-5	935242	0.946	13198	42.036
MG120-6	1011413	0.933	18383	35.368
MG120-7	1012083	0.941	18590	35.337
MG120-8	1593234	0.948	20843	53.045
MG120-9	1572342	0.948	19438	54.613
MG121-1	1945828	0.973	19266	69.698
MG122-1	2686292	0.976	12638	144.7
MG122-2	4586924	0.978	23482	150.272
MG124-1R	1017508	0.968	17320	37.994
MG124-2	955443	0.954	20026	31.856
MG124-3R	2370949	0.969	24067	72.399

MG124-4R	907607	0.971	19503	31.376
MG125-1	2535112	0.964	24049	76.8
MG125-10	2150726	0.964	27549	60.85
MG125-2	3231953	0.973	25269	98.059
MG125-3	3279297	0.974	23299	105.228
MG125-4	4015968	0.958	25993	118.148
MG125-5	1402070	0.957	23665	42.37
MG125-6	1569825	0.958	25790	45.052
MG125-7	2363601	0.963	27347	66.782
MG125-8	2594420	0.962	28513	72.106
MG125-9	1980450	0.961	26345	57.123
MG126-1	1505836	0.963	25147	45.062
MG126-2	1631100	0.973	14470	71.207
MG126-3	1551656	0.966	19077	54.742
MG126-4	1382566	0.958	23114	43.06
MG126-5	1190730	0.93	24328	34.349
MG127-1	707240	0.963	19487	24.347
MG127-2	1420116	0.965	23195	44.543
MG128-1	1454426	0.965	24074	44.481

Table S5: Summary of sequence and genotype filters

Filter #	Filter applied	Script calls	# of SNPs	Used for
Initial Stacks calls	Internal Stacks quality filters	Stacks ref_map	-	
1	Remove stacks present in less than 80% of samples	Stacks populations -R 0.8	49,580	Diversity and divergence ^{*†} ; introgression HMM ^{*†} ; <i>Treemix</i> [†]
2	Remove sites supported by less than 8 reads	vcftools --minDP --max-missing-count	11,297	-
3	Remove individuals with > 60% missing data	vcftools --remove-indv	11,297	D-stats
4	Remove sites in high LD ($r^2 > 0.7$)	bcftools +prune	5,767	<i>fastStructure</i> ; <i>NewHybrids</i> ; <i>RAxML</i> ; $\delta\alpha$

^{*} Performed on Stacks assembled haplotypes, not SNPs.

[†] Individuals with missing data also removed

Table S6: Full likelihoods for *NewHybrids* classifications of CHSxPIM admixture on Santa Cruz. MCMC was run for 6,000 steps. The most likely classification for each individual is shown in bold.

Individual	Pop	CHS	PIM	F1	F2	BC_CHS	BC_PIM
MG113-10	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-2	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-3R	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-4R	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-5	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-6R	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-7	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-8	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG113-9	PIM	1.00	0.00	0.00	0.00	0.00	0.00
MG115-10	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-11	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-12	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-13	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-14	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-15	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-16	ADMX	0.00	1.00	0.00	0.00	0.00	0.00
MG115-17R	ADMX	0.00	1.00	0.00	0.00	0.00	0.00
MG115-18	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-19	CHS	1.00	0.00	0.00	0.00	0.00	0.00
MG115-1	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-20	ADMX	0.00	0.00	0.00	1.00	0.00	0.00
MG115-2	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-3	ADMX	0.00	0.00	0.00	1.00	0.00	0.00
MG115-4	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-5	ADMX	0.00	0.00	1.00	0.00	0.00	0.00
MG115-6	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-7	ADMX	0.00	0.00	1.00	0.00	0.00	0.00
MG115-8	CHS	0.00	1.00	0.00	0.00	0.00	0.00
MG115-9	CHS	0.00	1.00	0.00	0.00	0.00	0.00

Table S7: Full likelihoods for *NewHybrids* classifications of CHSxGAL admixture on Isabela. MCMC was run for 6,000 steps. The most likely classification for each individual is shown in bold.

Individual	Pop	CHS	PIM	F1	F2	BC_CHS	BC_PIM
MG120-13	ADMX	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
MG120-15	ADMX	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-19	ADMX	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-23	ADMX	0.00000	0.00000	0.00000	0.00017	0.00000	0.99982
MG120-28	ADMX	0.00000	0.00000	0.07191	0.04213	0.00000	0.88596
MG120-33	ADMX	0.00000	0.00000	0.00000	0.99986	0.00000	0.00014
MG120-34	ADMX	0.00000	0.00000	0.00003	0.99995	0.00002	0.00000
MG120-11-2	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-17	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-18	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-20	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-22	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-24	CHS	0.01289	0.00000	0.00000	0.75656	0.00000	0.23055
MG120-27	CHS	0.00000	0.00000	0.19540	0.00655	0.00000	0.79804
MG120-29	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-31	CHS	0.99983	0.00000	0.00000	0.00017	0.00000	0.00000
MG120-10	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-11-1	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-21	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-25R	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-26	PIM	0.00000	0.00000	0.00000	0.00843	0.99157	0.00000
MG120-2	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-1	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-30	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-32	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-3	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-4	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-5	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-6	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-7	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-8	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000
MG120-9	PIM	0.00000	0.99983	0.00000	0.00017	0.00000	0.00000

Table S8: *NewHybrids* MCMC summary. Only the most likely genotype category classifications are shown. Refer to Table S5 for alternative class probabilities.

Population A	Population B	Type	Classification			
			Pure	F₁	F₂	BC
MG115	MG113	CHS-PIM	25	2	2	0
MG120C	MG120G	CHS-GAL	24	0	4	4

Table S9: Likelihoods for different *Treemix* runs

m	ln(L)
0	157.79953
1	350.8362
2	374.0593
3	381.47799
4	389.98895
5	393.66234
6	395.08137
7	396.37992
8	396.8767

Table S10: Mean Fst between focal island populations and mainland groups.

	Peru	Ecu
MG105	0.351069	0.193202
MG106	0.335434	0.25599
MG107	0.338202	0.172407
MG111	0.357811	0.21111
MG113	0.329355	0.169055
MG114	0.348367	0.188794
MG115	0.362342	0.298977
MG117	0.342618	0.182225
MG120_G	0.387008	0.347495
MG120_C	0.412477	0.370005
MG120_GC	0.380804	0.329513
MG125	0.308738	0.130834
MG126	0.33319	0.190727

Table S11: Summary of inferred introgression blocks for population MG114 (polymorphic). Shaded groups indicate blocks that are likely the same age/the result of the same hybridization event based on break points.

Block	Size (Mb)	Chr	Start (Mb)	Stop (Mb)	Count	Individuals	Type
114_1A	1.3	1	80	81.3	1	MG114-11	CHS/CHS
114_1B	1.2	1	80.1	81.3	1	MG114-14	CHS/CHS
114_1C	0.30	1	85.20	85.50	1	MG114-13	CHS/CHS
114_1D	6.8	1	16.0	22.8	1	MG114-9	CHS/CHS
114_1E	0.7	1	76.6	77.3	1	MG114-9	CHS/CHS
114_2A	3.60	2	36.9	40.1	1	MG114-7	CHS/CHS
114_3A	5.0	3	23.0	28.0	1	MG114-12	CHS/CHS
114_3B	3.1	3	4.8	7.9	1	MG114-4	CHS/CHS
114_3C	1.70	3	3.54	5.10	1	MG114-11	CHS/CHS
114_3D*	50.9	3	4.50	55.4	1	MG114-1	CHS/CHS
114_3E*	51.8	3	3.6	55.4	5	MG114-13, MG114-2, MG114-6, MG114-3, MG114-14	CHS/CHS
114_4A	8.70	4	8.40	17.10	1	MG114-13	CHS/CHS
114_4B	9.2	4	7.20	16.4	1	MG114-3	CHS/CHS
114_4C	8.0	4	8.4	16.4	1	MG114-14	CHS/CHS
114_4D	9.3	4	16.9	26.2	1	MG114-11	CHS/CHS
114_4E	4.8	4	43.0	48.8	1	MG114-11	CHS/CHS
114_5A	2.2	5	60.3	62.5	1	MG114-3	CHS/CHS
114_6A†	34.70	6	0.00	35.70	1	MG114-9	CHS/CHS
114_6B†	36.70	6	0.00	36.70	1	MG114-4	CHS/CHS
114_6C†	34.50	6	1.50	36.00	1	MG114-1	CHS/CHS
114_6D	4.30	6	1.1	25.7	1	MG114-10	CHS/CHS
114_6E	7.9	6	12.6	20.5	1	MG114-12	CHS/CHS
114_6F	5.4	6	29.7	35.1	1	MG114-12	CHS/CHS
114_6G	0.90	6	36.50	37.40	1	MG114-7	CHS/CHS
114_6H	19.7	6	7.0	26.7	1	MG114-15	CHS/CHS
114_7A	7.9	7	12.6	20.5	1	MG114-3	CHS/CHS
114_7B	5.4	7	29.7	35.1	1	MG114-3	CHS/CHS
114_7C	0.80	7	2.30	3.10	1	MG114-3	CHS/CHS
114_9A	0.60	9	72.3	72.9	1	MG114-10	CHS/CHS
114_11A	1.9	11	50.9	52.8	1	MG114-3	CHS/CHS

*CHS chromosome 3 haplotype; †CHS chromosome 6 haplotype;

Table S12: Summary of inferred introgression blocks for population MG117 (polymorphic). Shaded groups indicate blocks that are likely the same age/the result of the same hybridization event based on break points.

Block	Size (Mb)	Chr	Start (Mb)	Stop (Mb)	Count	Individuals	Type
117_1A	6.6	1	51.4	58.0	1	MG117-9	CHS/CHS
117_1B	0.7	1	76.6	77.3	1	MG117-9	CHS/CHS
117_1C	5.8	1	4.1	9.9	1	MG117-12	CHS/CHS
117_1D	4.2	1	4.1	8.3	1	MG117-5	CHS/CHS
117_1E	17	1	44	61	3	MG117-12, MG117-2, MG117-5	CHS/CHS
117_1F	10.1	1	5.7	15.8	1	MG117-2	CHS/CHS
117_1G	5.7	1	69.3	75.0	1	MG117-2	CHS/CHS
117_1H	33.8	1	31	64.8	1	MG117-6	CHS/CHS
117_2A	2.0	2	29.2	31.2	1	MG117-2	CHS/CHS
117_2B	3.9	2	28.5	32.4	1	MG117-6	CHS/CHS
117_2C	0.3	2	55.6	55.9	1	MG117-2	CHS/CHS
117_3A	1.6	3	66.2	67.8	2	MG117-3, MG117-5	CHS/CHS
117_3B	3.2	3	7.2	10.4	1	MG117-12	CHS/CHS
117_3B	2.3	3	8.1	10.4	1	MG117-5	CHS/CHS
117_4A*	2.2	4	64.4	66.6	1	MG117-8	CHS/CHS
117_4B*	2.9	4	63.7	66.6	2	MG117-11, MG117-10	CHS/CHS
117_4C*	1.9	4	64.7	66.6	1	MG117-7	CHS/CHS
117_4D	8	4	8.4	16.4	2	MG117-12, MG117-5	CHS/CHS
117_6A†	36.3	6	1.1	37.4	4	MG117-8, MG117-11, MG117-7, MG117-10	CHS/CHS
117_6B†	7.3	6	42.5	49.8	1	MG117-8	CHS/CHS
117_6C†	5.3	6	44.5	49.8	1	MG117-11	CHS/CHS
117_6D†	5.4	6	44.4	49.8	1	MG117-7	CHS/CHS
117_6E†	4.0	6	45.8	49.8	1	MG117-10	CHS/CHS
117_6F	6.7	6	43.1	49.8	2	MG117-4, MG117-1	CHS/PIM
117_6G	2.4	6	47.4	49.8	1	MG117-6	CHS/CHS
117_6H	5.1	6	32.2	37.3	1	MG117-5	CHS/CHS
117_6I	21.3	6	4.3	25.7	1	MG117-6	CHS/CHS
117_7A	1.2	7	5.2	6.4	1	MG117-9	CHS/CHS
117_8A	0.3	8	3.1	3.4	2	MG117-5, MG117-12	CHS/CHS
117_9A	0.6	9	72.3	72.9	2	MG117-5, MG117-5	CHS/CHS
117_10A	11.8	10	44.3	56.1	2	MG117-3, MG117-2	CHS/CHS
117_10B	1.3	10	59.2	60.5	1	MG117-3	CHS/CHS
117_11A	0.5	11	56.1	56.6	1	MG117-3	CHS/CHS

117_11B	30.9	11	14.8	45.7	1	MG117-12	CHS/CHS
117_11C	34.8	11	14.8	49.6	1	MG117-5	CHS/CHS
117_11D	0.6	11	50.9	51.5	1	MG117-5	CHS/CHS

*CHS chromosome 4 haplotype; †CHS chromosome 6 haplotype

Table S13: Summary of inferred introgression blocks for population MG116
(fixed, red-fruited)

Block	Size (Mb)	Chr	Start (Mb)	Stop (Mb)	Count	Individuals	Type
116_1A	4.6	1	4.8	9.4	1	MG116-9	CHS/CHS
116_2A	4.9	1	14.7	19.6	1	MG116-9	CHS/CHS
116_2A	20.4	2	0.4	20.8	1	MG116-4	CHS/CHS
116_2B	2.1	2	29.1	31.2	1	MG116-4	CHS/CHS
116_2C	1	2	39.9	40.9	1	MG116-4	CHS/CHS
116_5A	0.9	5	3.6	4.5	1	MG116-6	CHS/CHS
116_6A	1.1	6	45.6	46.7	1	MG116-4	CHS/CHS
116_9A	0.6	9	72.3	72.9	1	MG116-6	CHS/CHS

Table S14: Population MG114 ancestry at carotenoid biosynthesis loci, as inferred by our HMM. The genomic location of each locus was determined based on the *Solanum lycopersicum* reference build SL3.0 and iTAG3.0 annotation. The association between ancestry and fruit color at PSY1 is significant based on a χ^2 test of independence ($\chi^2 = 11.123$; df = 1; P = 0.00085). The association at LCY-B is not statistically significant ($\chi^2 = 2.934$; df = 1; P = 0.08673).

Gene	ZDS	PSY1*	PDS	LCY-B	CYC-B	CCS	CRTISO	LCY-E
Chromosome Position (bp)	1 88514122	3 4325334	3 70499494	4 11946753	6 45897927	8 63274891	10 62681526	12 2284379
Fruit color	Ancestry							
MG114-1	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-2	Orange	PIM	CHS	PIM	PIM	PIM	PIM	PIM
MG114-3	Orange	PIM	CHS	PIM	PIM	PIM	PIM	PIM
MG114-4	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-5	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-6	Orange	PIM	CHS	PIM	PIM	PIM	PIM	PIM
MG114-7	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-8	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-9	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-10	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-11	Orange	PIM	CHS	PIM	PIM	PIM	PIM	PIM
MG114-12	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG114-13	Orange	PIM	CHS	PIM	PIM	PIM	PIM	PIM
MG114-14	Orange	PIM	CHS	PIM	PIM	PIM	PIM	PIM
MG114-15	Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM

Table S15: Population MG117 ancestry at carotenoid biosynthesis loci, as inferred by our HMM. The genomic location of each locus was determined based on the *Solanum lycopersicum* reference build SL3.0 and iTAG3.0 annotation. The association between ancestry and fruit color at CYC-B is significant based on a χ^2 test of independence ($\chi^2 = 8.333$; df = 1; P = 0.00389). The association at LCY-B is not statistically significant ($\chi^2 = 0.6$; df = 1; P = 0.4386).

Gene Chromosome Position (bp)	ZDS	PSY1	PDS	LCY-B	CYC-B*	CCS	CRTISO	LCY-E
	1 88514122	3 4325334	3 70499494	4 11946753	6 45897927	8 63274891	10 62681526	12 2284379
Fruit Color	Ancestry							
MG117-1 Orange	PIM	PIM	PIM	PIM	CHS	PIM	PIM	PIM
MG117-2 Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG117-3 Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG117-4 Orange	PIM	PIM	PIM	PIM	CHS	PIM	PIM	PIM
MG117-5 Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG117-6 Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG117-7 Orange	PIM	PIM	PIM	PIM	CHS	PIM	PIM	PIM
MG117-8 Orange	PIM	PIM	PIM	PIM	CHS	PIM	PIM	PIM
MG117-9 Red	PIM	PIM	PIM	PIM	PIM	PIM	PIM	PIM
MG117-10 Orange	PIM	PIM	PIM	PIM	CHS	PIM	PIM	PIM
MG117-11 Orange	PIM	PIM	PIM	PIM	CHS	PIM	PIM	PIM
MG117-12 Red	PIM	PIM	PIM	CHS	PIM	PIM	PIM	PIM

Table S16: Demographic model estimates for PIM population MG114 inferred using $\delta\alpha\delta i$ on the unmasked dataset (including introgressed regions). 95% CI values were obtained from 2,000 bootstrap replicates of the SFS. Each estimate is shown in rescaled units (rescaled by N_{Ref} for N_B and N_F ; and by $2N_{Ref}$ for T_B and T_F).

Parameter	Optimum	Bootstrap Median	95% CI
N_B	115.37	314.33	4.82 - 1444.46
N_F	335.06	991.97	345.55 – 6644.64
T_B	2125.97	928.21	109.74 – 8509.09
T_F	52.59	917.73	35.77 – 16116.06
F	0.0008	0.0134	$3.48 \times 10^{-9} – 0.53$

SI References

- Darwin, S. C. (2009).** The systematics and genetics of tomatoes on the Galápagos Islands (*Solanum*, Solanaceae) [Doctoral, UCL (University College London)]. <https://discovery.ucl.ac.uk/id/eprint/18994/>
- Darwin, S. C., Knapp, S., & Peralta, I. E. (2003).** Taxonomy of tomatoes in the Galápagos Islands: Native and introduced species of *Solanum* section *Lycopersicon* (Solanaceae). *Systematics and Biodiversity*, 1(1), 29–53.
- Lynch, M., Walsh, B., & Others. (1998).** Genetics and analysis of quantitative traits (Vol. 1). Sinauer Sunderland, MA.
- Nuez, F., Prohens, J., & Blanca, J. M. (2004).** Relationships, origin, and diversity of Galapagos tomatoes: implications for the conservation of natural populations. *American Journal of Botany*, 91(1), 86–99.
- Rick, C. M. (1956).** Genetic and Systematic Studies on Accessions of *Lycopersicon* from the Galapagos Islands. *American Journal of Botany*, 43(9), 687–696.
- Rick, C. M. (1967).** Fruit and pedicel characters derived from Galápagos tomatoes. *Economic Botany*, 21(2), 171.
- Stommel, J. R., & Haynes, K. G. (1994).** Inheritance of Beta Carotene Content in the Wild Tomato Species *Lycopersicon cheesmanii*. In *Journal of Heredity* (Vol. 85, Issue 5, pp. 401–404).
<https://doi.org/10.1093/oxfordjournals.jhered.a111486>