Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers

Sean Stankowski[1,2,§], Madeline A. Chase[1,3,§], Allison M. Fuiten[1,4], Murillo F. Rodrigues[1], Peter L. Ralph[1], and Matthew A. Streisfeld[1*].

[1]Institute of Ecology and Evolution, 335 Pacific Hall 5289, University of Oregon, Eugene, OR 97405, USA
[2]Present address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
[3]Present address: Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden
[4]Present address: Department of Dermatology, Oregon Heath and Science University, L468R Portland, OR, USA
[§]These authors contributed equally to this work

*Corresponding author: mstreis@uoregon.edu

ORCID ids:
SS: https://orcid.org/0000-0003-0472-9299; MAC: https://orcid.org/0000-0002-7916-3560; AMF: https://orcid.org/0000-0003-3926-5455; MFR: https://orcid.org/0000-0001-7508-1384 PLR: https://orcid.org/0000-0002-9459-6866; MAS: https://orcid.org/0000-0002-2660-8642

Keywords: Indirect selection; introgression; linked selection *Mimulus*; speciation

Short title: Rapid genomic landscape evolution during a radiation of monkeyflowers

Abstract—Speciation genomic studies aim to interpret patterns of genome-wide variation in light of the processes that give rise to new species. However, interpreting the genomic 'landscape' of speciation is difficult, because many evolutionary processes can impact levels of variation. Facilitated by the first chromosome-level assembly for the group, we use whole-genome sequencing and simulations to shed light on the processes that have shaped the genomic landscape during a recent radiation of monkeyflowers. After inferring the phylogenetic relationships among the nine taxa in this radiation, we show that highly similar diversity (π) and differentiation (F_{ST}) landscapes have emerged across the group. Variation in these landscapes was strongly predicted by the local density of functional elements and the recombination rate, suggesting that the landscapes have been shaped by widespread natural selection. Using the varying divergence times between pairs of taxa, we show that the correlations between F_{ST} and genome features arose almost immediately after a population split and have become stronger over time. Simulations of genomic landscape evolution suggest that background selection (i.e., selection against deleterious mutations) alone is too subtle to generate the observed patterns, but scenarios that involve positive selection and genetic incompatibilities are plausible alternative explanations. Finally, tests for introgression among these taxa reveal

widespread evidence of heterogeneous selection against gene flow during this radiation. Thus, combined with existing evidence for adaptation in this system, we conclude that the correlation in F_{ST} among these taxa informs us about the genomic basis of adaptation and speciation in this system.

Author summary—What can patterns of genome-wide variation tell us about the speciation process? The answer to this question depends upon our ability to infer the evolutionary processes underlying these patterns. This, however, is difficult, because many processes can leave similar footprints, but some have nothing to do with speciation *per se*. For example, many studies have found highly heterogeneous levels of genetic differentiation when comparing the genomes of emerging species. These patterns are often referred to as differentiation 'landscapes' because they appear as a rugged topography of 'peaks' and 'valleys' as one scans across the genome. It has often been argued that selection against deleterious mutations, a process referred to as background selection, is primarily responsible for shaping differentiation landscapes early in speciation. If this hypothesis is correct, then it is unlikely that patterns of differentiation will reveal much about the genomic basis of speciation. However, using genome sequences from nine emerging species of monkeyflower coupled with simulations of genomic divergence, we show that it is unlikely that background selection is the primary architect of these landscapes. Rather, differentiation landscapes have probably been shaped by adaptation and gene flow, which are processes that are central to our understanding of speciation. Therefore, our work has important implications for our understanding of what patterns of differentiation can tell us about the genetic basis of adaptation and speciation.

Introduction

The primary goal of speciation genomics is to interpret patterns of genome-wide variation in light of the ecological and evolutionary processes that contribute to the origin of new species (Ravinet et al. 2017, Wolf and Ellegren 2017, Campbell et al. 2018). Advances in DNA sequencing now allow us to capture patterns of genome-wide variation from organisms across the tree of life, but inferring the processes underlying these patterns remains a formidable challenge (Ravinet et al. 2017). This is because speciation is highly complex, involving a range of factors and processes that shape genomes through time and across different spatial and ecological settings (Abbott et al. 2013).

The difficulty of inferring process from pattern is illustrated by recent efforts to characterize the genomic basis of reproductive isolation using patterns of genome-wide variation (Ravinet et al. 2017, Wolf and Ellegren 2017, Campbell et al. 2018). Numerous studies have revealed highly heterogeneous differentiation 'landscapes' between pairs of taxa at different stages in the speciation process (Turner et al. 2005, Hohenlohe et al. 2010, Ellegren et al. 2012, Martin et al. 2013, Renaut et al. 2013, Poelstra et al. 2014, Soria-Carrasco et al. 2014, Lamichhaney et al. 2015, Malinsky et al. 2015). This pattern, which is characterized by peaks and valleys of relative differentiation (i.e., F_{ST}) across the genome, was initially thought to provide insight into the genomic architecture of porous species boundaries (Wu 2001). Peaks in the differentiation landscape were interpreted as genomic regions containing loci underlying reproductive barriers, while valleys were thought to reflect regions that were homogenized by ongoing gene flow (Turner et al.

2005, Nosil et al. 2009, Feder et al. 2012). However, as the field of speciation genomics has matured, it has become clear that heterogeneous differentiation landscapes can be influenced by factors that have nothing to do with speciation *per se*.

For example, it is now clear that levels of genome-wide differentiation (F_{ST}) can be influenced by the genomic distribution of intrinsic properties, including the recombination rate and the local density of functional sites (Cruickshank and Hahn 2014). This is because these properties affect the way that natural selection impacts levels of variation across the genome. First, regions enriched for functional sequence are more likely to be subject to selection, because they provide a larger target size for mutations with fitness effects. Second, when positive or negative directional selection acts on these mutations, it can indirectly reduce levels of genetic variation at statistically associated sites (Maynard-Smith and Haigh 1974, Charlesworth et al. 1993, Hudson and Kaplan 1995, Charlesworth 1998, Gillespie 2000, Coop and Ralph 2012, Cruickshank and Hahn 2014). Because these indirect effects of selection are mediated by linkage disequilibrium (LD) between selected and neutral sites, stronger reductions in diversity (and corresponding increases in F_{ST}) are expected in genomic regions with low recombination, as this is where LD breaks down most gradually (Baird 2015). This is also why these indirect effects of selection are referred to as 'linked selection', even though physical linkage is not actually the cause. In this paper, we refer to the indirect effects of selection, as 'indirect selection' for short.

Although genomic features were not initially expected to play a major role in shaping patterns of between species variation (Kimura 1968, Ohta 1973) (but see (Charlesworth 1998, Gillespie 2000)), recent empirical studies indicate that they can have a strong impact on the topography of the differentiation landscape (Hahn 2008, Cruickshank and Hahn 2014, Burri 2017a, b, Wolf and Ellegren 2017). The most compelling evidence comes from studies that compare not one pair of taxa, but several closely related taxa that share a similar distribution of genomic elements due to their recent common ancestry (Burri 2017a, b, Wolf and Ellegren 2017). For example, Burri et al. (Burri et al. 2015) examined multiple episodes of divergence in *Ficedula* flycatchers and found strikingly similar differentiation landscapes between distinct pairs of taxa, presumably due to a shared pattern of indirect selection across the genome. Highly correlated differentiation landscapes have also been found in other groups of taxa, including sunflowers (Renaut et al. 2013), *Heliconius* butterflies (Kronforst et al. 2013, Martin et al. 2013), Darwin's finches (Han et al. 2017), and other birds (Van Doren et al. 2017, Delmore et al. 2018).

But what, if anything, do these parallel signatures of selection reveal about the genomic basis of adaptation and speciation? Part of the answer lies in understanding which forms of selection cause these patterns. In a recent article, Burri (Burri 2017b) argued that selection against deleterious mutations (*i.e.*, background selection, or BGS) is primarily responsible for the evolution of differentiation landscapes that are correlated both among taxa and with the distribution of intrinsic features. This argument was based on two premises: first, deleterious mutations are far more common than beneficial ones, so BGS has a greater opportunity to generate a genome-wide correlation between levels of variation and the distributions of intrinsic properties; second, although all functional elements are potential targets of BGS, the same loci are not expected to be repeatedly involved in adaptation or speciation across multiple taxa (Burri 2017b). Although this

interpretation implies that correlated differentiation landscapes are unlikely to inform us about adaptation or speciation directly, it has been argued that this shared pattern of differentiation can be used to control for the effects of BGS when attempting to identify regions of the genome that have been affected by positive selection (Burri 2017b).

Even though it is often argued that correlated differentiation landscapes reflect the action of recurrent background selection, it stands to reason that they also could arise as a direct consequence of adaptation and speciation. For example, when adaptation occurs from standing variation, the rate of adaptation is not limited by the mutation rate, meaning that heterogeneous differentiation landscapes can evolve rapidly (Barrett et al. 2008, Bassham et al. 2018). Moreover, if adaptation is highly polygenic, positive selection will inevitably impact most of the genome (Rockman 2012). This could cause levels of differentiation to become correlated with the distribution of intrinsic genomic features, even across multiple taxa that are adapting to different environments. Also, correlated differentiation landscapes may arise across multiple closely related taxa due to a common basis of reproductive isolation among taxa. This is because incomplete isolating barriers generate a heterogeneous pattern of selection against gene flow across the genome. The loci underlying these porous barriers, including genetic incompatibilities and locally adapted alleles, are expected to accumulate in gene rich regions and will have stronger barrier effects in genomic regions where the recombination rate is low (Schumer et al. 2018, Martin et al. 2019). Thus, speciation also may result in genome-wide correlations between levels of differentiation and genome features, especially if reproductive isolation is highly polygenic.

Facilitated by a new chromosome-level genome assembly, genetic map, and annotation, we combine analyses of whole-genome sequencing with simulations to understand how different processes have contributed to the evolution of correlated genomic landscapes during a recent radiation. The bush monkeyflower radiation consists of eight taxa of *Mimulus aurantiacus* distributed mainly throughout California. (Fig. 1; (Chase et al. 2017). Together with their sister species *M. clevelandii,* they span a range of divergence times over the past approximately one million years. The plants inhabit a range of environments, including temperate coastal regions, mountain ranges, semi-arid habitats, and offshore islands (Thompson 2005). Crossing experiments have shown that all of these taxa are at least partially inter-fertile (McMinn 1951), and many hybridize in nature in narrow regions where their distributions overlap (Streisfeld and Kohn 2005, 2007, Sobel and Streisfeld 2015, Stankowski et al. 2015, 2017). Despite their close evolutionary relationships and opportunities for gene flow, these taxa show striking phenotypic differentiation (Chase et al. 2017). The most conspicuous trait differences are associated with their flowers, which show heritable variation in color, size, shape, and the placement of the reproductive organs (Streisfeld and Kohn 2005, Stankowski et al. 2015). In his seminal works on plant speciation, Grant (Grant 1981, 1993b, a) postulated that these floral trait differences were due to pollinator-mediated selection by different avian and insect pollinators. Detailed studies in one pair of taxa support this hypothesis and have shown that pollinator-mediated selection can generate strong premating reproductive isolation in the face of extensive gene flow (Streisfeld and Kohn 2005, 2007, Sobel and Streisfeld 2015, Stankowski et al. 2015, 2017).

After inferring the phylogenetic relationships among these taxa, we show that highly similar diversity (π) and differentiation (F_{ST}) landscapes have emerged across the

group. Variation in these landscapes was strongly predicted by the local density of functional elements and the recombination rate, suggesting that they have been shaped by widespread selection. Using the varying divergence times between pairs of taxa, we show that the correlations between F_{ST} and genome features arose almost immediately after a population split and have become stronger over the course of time. Simulations of genomic landscape evolution suggest that background selection (i.e., selection against deleterious mutations) alone is too subtle to generate the observed patterns, but scenarios that involve positive selection and genetic incompatibilities are plausible alternative explanations. Finally, tests for introgression among these taxa reveal widespread evidence of heterogeneous selection against gene flow during this radiation. We discuss the implications of these results for our general understanding of genomic landscape evolution, particularly in light of recent efforts to reveal the genomic basis of adaptation and speciation.

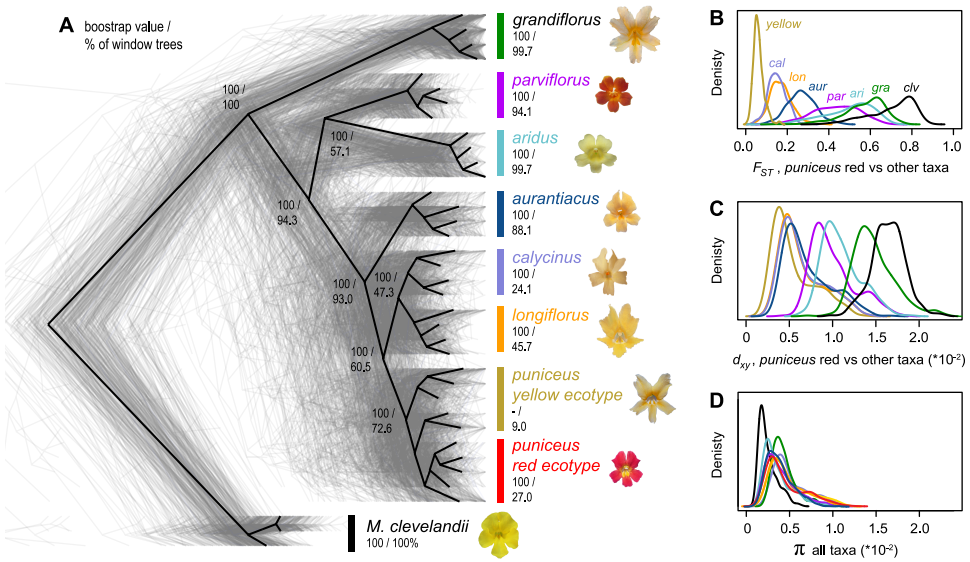

Figure 1. Evolutionary relationships and patterns of genome-wide variation across the radiation. A) The black tree was constructed from a concatenated alignment of genome-wide SNPs and is rooted using *M. clevelandii*. The 387 gray trees were constructed from 500 kb genomic windows. The first number associated with each node or taxon is the bootstrap support for that clade in the whole genome tree, and the second number is the percentage of window-based trees in which that clade is present. B) Levels of differentiation (F_{ST}), C) divergence (d_{xy}), and D) diversity (π) within and among taxa based on the same 500 kb windows. For simplicity, F_{ST} and d_{xy} are shown only for comparisons with the red ecotype of subspecies *puniceus*. See Fig. S7 for the distributions of F_{ST} and d_{xy} across all pairs of taxa.

Results and Discussion

A chromosome-level genome assembly, map, and annotation for the bush monkeyflower

To facilitate the analysis of genome-wide variation in this group, we constructed the first chromosome-level reference genome for the bush monkeyflower using a combination of long-read Single Molecule Real Time (SMRT) sequencing (PacBio), overlapping and mate-pair short-reads (Illumina), and a high-density genetic map (7,589

217 segregating markers across 10 linkage groups; Fig. S1; Table S1). Contig building and
218 scaffolding yielded 1,547 scaffolds, with an N50 size of 1.6 Mbp, and a total length of
219 207 Mbp. The high-density map allowed us to anchor and orient 94% of the assembled
220 genome onto 10 linkage groups, which is the number of chromosomes inferred from
221 karyotypic analyses in all subspecies of *M. aurantiacus* and *M. clevelandii* (Vickery
222 1995). Analysis of assembly completeness based on conserved gene space (Simao et al.
223 2015) revealed that 93% of 1440 universal single copy orthologous genes were
224 completely assembled, with a further 2% partially assembled (Table S2). Subsequent
225 annotation yielded 23,018 predicted genes.
226
227 *Phylogenetic relationships among taxa and patterns of discordance across the genome*
228 To infer phylogenetic relationships among the taxa in this radiation, we sequenced
229 37 whole genomes from the seven subspecies and two ecotypes of *Mimulus aurantiacus*
230 (n = 4-5 per taxon) and its sister taxon *M. clevelandii* (n = 3) (Fig. S2; Table S3). Close
231 sequence similarity allowed us to align reads from all samples to the reference assembly
232 with high confidence (average 91.7% reads aligned; Table S3). After mapping, we
233 identified 13.2 million variable sites that were used in subsequent analyses (average
234 sequencing depth of 21x per individual, Table S3). Relationships were then inferred
235 among the nine taxa from a concatenated alignment of genome-wide SNPs using
236 maximum-likelihood (ML) phylogenetic analysis in *RAxML* (Stamatakis 2014).
237 The tree topology obtained from this analysis (Fig. 1) confirmed the same
238 phylogenetic relationships as previous analyses based on reduced-representation
239 sequencing and five different methods of phylogenetic reconstruction (Stankowski and
240 Streisfeld 2015, Chase et al. 2017), and was supported by patterns of clustering from
241 principal components analysis (Fig. S3). Individuals of each of the seven subspecies
242 formed monophyletic groups with 100% bootstrap support (Fig. 1). Relationships within
243 subspecies *puniceus* were more complex, as the red ecotype formed a monophyletic sub-
244 clade within the paraphyletic yellow ecotype. This is consistent with the recent origin of
245 red flowers from a yellow-flowered ancestor (Stankowski and Streisfeld 2015).
246 Although the whole genome phylogeny provides a well-supported summary of the
247 relationships among these taxa, concatenated phylogenies can obscure phylogenetic
248 discordance in more defined genomic regions (Pease et al. 2016) (Fig. 1A). To test for
249 fine-scale phylogenetic discordance, we next constructed ML phylogenies for 500 kb and
250 100 kb genomic windows. We then calculated a 'concordance score' for each tree by
251 computing the correlation between the distance matrix generated from each window-
252 based tree and the whole-genome tree, with a stronger correlation indicating that two
253 trees have a more similar topology.
254 At the 500 kb scale, only 22 (6%) trees showed the same taxon branching order as
255 the whole-genome tree. However, concordance scores tended to be very high for all of
256 the trees (mean = 0.964, s.d. 0.039; min 0.719), suggesting that variation in the topologies
257 was due primarily to minor differences in branching order. This was confirmed by
258 quantifying how often each node in the genome-wide tree was recovered in the set of
259 window-based trees (Fig. 1). Specifically, we found that differences in branching order
260 were associated with the most recent splits, which included pairs of closely related taxa.
261 For example, even though subspecies *puniceus* was monophyletic in the majority of trees
262 (72.6%), individuals from the red ecotype only formed a monophyletic group in 27% of

263 the trees. Similarly, the closely related subspecies *longiflorus* and *calycinus* were
264 monophyletic in only 45.7% and 24.1% of trees, respectively. Higher discordance among
265 these closely related, geographically proximate taxa is likely due to two factors. First,
266 only a short time has passed since they shared a common ancestor, meaning that ancestral
267 polymorphisms have had little time to sort among lineages. Second, ongoing gene flow
268 between some taxa may have opposed sorting, prolonging the retention of ancestral
269 variants among them.
270 Next, we examined how the level of phylogenetic discordance varied across the
271 bush monkeyflower genome. If the discordance was due to the stochastic effects of
272 neutral processes, variation in tree concordance scores should be distributed randomly
273 across the genome (Hudson 1990). To test this prediction, we plotted the tree
274 concordance scores across the 10 linkage groups (Fig. 2A; Fig. S4 for results from 100 kb
275 windows and Fig. S5 for plots along each chromosome). Rather than being randomly
276 distributed, trees with lower concordance scores tended to cluster in relatively narrow
277 regions of all 10 chromosomes (Fig. 2A; autocorrelation analysis permutation tests p =
278 0.001 - 0.023; Fig. S6). This non-random pattern indicates that the rate of sorting varies
279 along chromosomes, which could be due to variation in the strength of indirect selection
280 across the bush monkeyflower genome.
281
282 *Correlated patterns of genome-wide variation across the radiation*
283 To gain deeper insight into the evolutionary processes that have shaped patterns
284 of genome-wide variation during this radiation, we used three summary statistics to
285 quantify patterns of diversity (π), divergence (d_{xy}), and differentiation (F_{ST}) within and
286 between these taxa (the same 500 kb and 100 kb sliding windows as above). π, or
287 nucleotide diversity, was used to quantify the level of genetic variation within each taxon.
288 It is defined as the average number of nucleotide differences per site between two
289 sequences drawn from the population (Nei and Li 1979). Second, we calculated d_{xy} as a
290 measure of sequence divergence between all 36 pairs of taxa. In principle, π and d_{xy} are
291 the same measure, except that the former is calculated between sequences within a taxon
292 and the latter between pairs of taxa; they have the same units and both are proportional to
293 the average coalescence time of the sequences being compared multiplied by mutation
294 rate. Third, we calculated F_{ST} between each pair of taxa. Unlike d_{xy}, which is a measure
295 of divergence between sequences drawn from different taxa, F_{ST} measures differentiation
296 between the taxa relative to the total diversity in the sample. Between samples x and y,
297 F_{ST} is roughly equal to $1 - \pi/d_{xy}$, where π is the mean diversity in the two taxa (Slatkin
298 1991). F_{ST} is strongly influenced by current levels of within-taxon diversity, while d_{xy} is
299 strongly influenced by the level of ancestral variation when divergence times are small.
300 The use of the term 'divergence' to describe both d_{xy} and F_{ST} has caused some confusion
301 in the literature, leading to alternative naming schemes. Here, we follow a recently
302 proposed convention, referring to d_{xy} as 'divergence' and F_{ST} as 'differentiation' (Ravinet
303 et al. 2017).
304 The variation in F_{ST} among all 36 pairs of taxa highlights the continuous nature
305 of differentiation across the group (Fig. 1B; Fig. S7), with mean window-based estimates
306 ranging from 0.06 (red vs. yellow ecotypes of *puniceus*) to more than 0.70. Distributions
307 of divergence (d_{xy}) show a similar pattern (Fig. 1C), with mean values ranging from
308 0.54% (red vs. yellow ecotypes) to 1.6% (yellow ecotype vs. *M. clevelandii*). Broad

distributions of window-based estimates indicate high variability in levels of differentiation and divergence among genomic regions (Fig. 1B & 1C). Window-based estimates of nucleotide diversity also varied markedly (π; Fig. 1D), ranging from 0.09% to 1.26%, even though mean estimates were very similar among the ingroup taxa (0.37% to 0.53%) and were only slightly lower in *M. clevelandii* (0.26%).

As with tree concordance, variation in these summary statistics was non-randomly distributed across broad regions of each chromosome ($p < 0.005$; Fig. 2; Fig. S4 Fig. S5; Fig. S6). To account for the magnitude of variation in these statistics across all nine taxa (for π) or among the 36 pairs of taxa (for d_{xy} and F_{ST}), we normalized the window-based estimates using *Z*-transformation and plotted them across the genome (Fig. S5). Visual inspection of these data revealed that patterns of genome-wide variation in each statistic were qualitatively similar in all comparisons. We therefore used principal components analysis to quantify their similarity and extracted a single variable (PC1) that summarized the common pattern (Fig. S5).

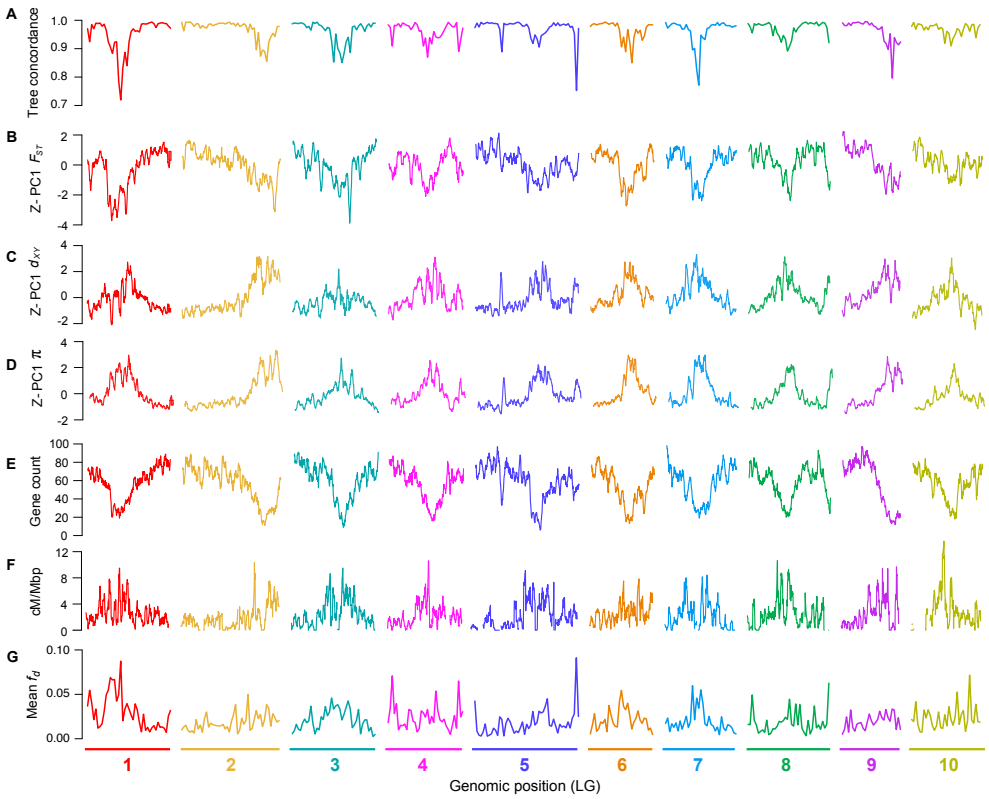

Figure 2. Common genomic landscapes mirror variation in the local properties of the genome. A) Tree concordance scores for 500 kb non-overlapping genomic windows plotted across the 10 bush monkeyflower chromosomes. B – D) Plots of the first principal component (PC1) for F_{ST}, d_{xy}, and π in overlapping 500 kb windows (step size = 50 kb). PC1 explains 66%, 70%, and 85% of the variation in F_{ST}, d_{xy} and π, respectively, and is Z-transformed such that above average values have positive values and below average values have negative values. E – F) Gene count and recombination rate (cM/Mbp) in overlapping 500 kb windows. G), Mean f_d (admixture proportion) in 500 kb non-overlapping genomic windows. See Fig. S4 for the same plot made for 100 kb windows.

These analyses confirmed that patterns of genome-wide variation were highly correlated across this group of taxa. Indeed, PC1 explained 65.9% of the variation in F_{ST} across the 36 pairwise comparisons. Further, all comparisons loaded positively onto PC1 (mean loading = 0.78 s.d. 0.18; Table S5 for all loadings), indicating that peaks and troughs of F_{ST} tended to occur in the same genomic regions across all comparisons. Patterns of genome-wide divergence (d_{xy}) and diversity (π) also were highly correlated across comparisons, with PC1 explaining 69.5% and 84.7% of the variation among the window-based estimates, respectively. Again, all taxa (for π) and taxon comparisons (for d_{xy}) loaded positively onto the first principal component (mean loading for d_{xy} = 0.78 s.d. 0.18; for π, 0.91 s.d. 0.07). PC1 therefore provides a summary of the original landscapes and is effectively the same as taking the mean window-based scores for each statistic (r^2 between PC1 and mean scores > 0.995 for all three statistics).

Common genomic landscapes have been shaped by heterogeneous indirect selection

Observing highly similar genomic landscapes suggests that a common pattern of heterogeneous selection has shaped variation in all nine taxa. Indeed, if a region experiences recurrent indirect selection across the phylogenetic tree, then it should show lower diversity (π) within species and lower divergence (d_{xy}) between species, because d_{xy} is influenced by levels of diversity in the common ancestor (Charlesworth et al. 1993, Charlesworth 1998, Burri 2017b). In agreement with this prediction, we observed a strong positive correlation between PC1 d_{xy} and PC1 π (r = 0.84), indicating that regions of the genome with lower diversity tended to be less diverged between these taxa (and thus had lower ancestral diversity) (Fig. 3, Fig. S8 for scatterplots and Fig. S9 for results at 100 kb scale). Regions with reduced diversity also tended to show higher levels of differentiation (F_{ST}) (r = -0.84) and tree concordance (r = -0.69). These relationships are also predicted by models of recurrent indirect selection, because local reductions in diversity decrease the amount of genetic variation in future generations, similar to a local reduction in N_e. As a result, ancestral variants in impacted regions sort more rapidly than under neutrality (Pease and Hahn 2013).

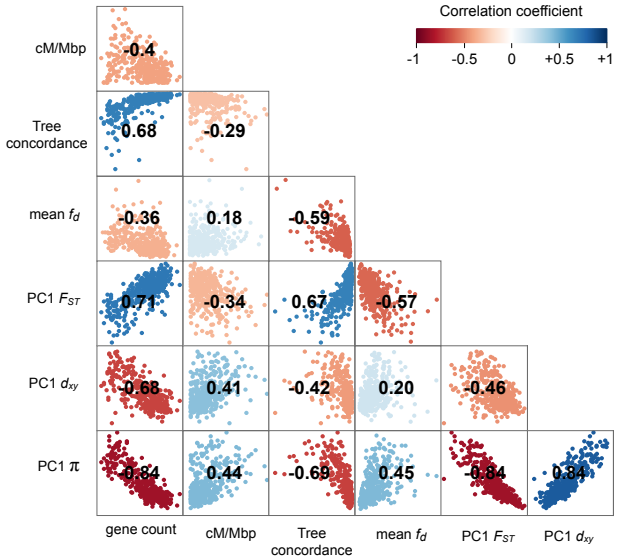

Figure 3. Correlations between measures of diversity and intrinsic features reveal the impact of heterogeneous indirect selection. Matrix of pairwise correlations between PC1 F_{ST}, PC1 d_{xy}, PC1 π, tree concordance, mean f_d, gene density, and recombination rate, all estimated in 500 kb non-overlapping windows. The heat map indicates the strength of the correlation and its sign. All correlations are statistically significant at $p < 0.001$. (See Fig. S8 for a more detailed correlation matrix and Fig. S9 for the correlation matrix from 100 kb windows.

406 In models of recurrent indirect selection, variation in its impact on associated sites
407 is determined by the distribution of genomic features, including the local density of
408 functional elements and the recombination rate (Maynard-Smith and Haigh 1974,
409 Charlesworth et al. 1993). To test these theoretical predictions, we used our genome
410 annotation and genetic map to calculate the number of protein coding genes and the
411 average recombination rate (cM/Mbp) in each 500 kb window (Fig. 2E-F; Fig. S4; Fig.
412 S5). Regions of the genome with more functional elements tended to have a lower
413 recombination rate (r = -0.40; p < 0.0001), leading to large variation in the predicted
414 strength of indirect selection among regions. Consistent with the theoretical predictions
415 outlined above, we found strong correlations between PC1 π and gene count (r = -0.84; p
416 < 0.0001) and PC1 π and recombination rate (r = 0.44; p < 0.0001; Fig. 3; Figs. S8 &
417 S9), indicating that diversity is indeed lower in regions where selection is predicted to
418 have stronger indirect effects. However, we did not observe a significant interactive
419 effect of gene count and recombination rate on diversity (p = 0.057). This may be
420 because the distribution of these features is correlated, making it difficult to tease apart
421 their relative impacts.
422 Taken together, our results indicate that a common pattern of indirect selection
423 has caused correlated genomic landscapes to evolve across this radiation. As predicted by
424 theory (Maynard-Smith and Haigh 1974, Charlesworth et al. 1993, Langley et al. 2012)
425 (Hudson and Kaplan 1995) and observed in diverse taxa (Pease and Hahn 2013, Burri et
426 al. 2015, Corbett-Detig et al. 2015), genome-wide variation in the strength of indirect
427 selection is caused by the heterogeneous distributions of intrinsic genome features,
428 namely the density of functional elements and the local recombination rate.
429
430 *A known adaptive locus shows a strong deviation from the common pattern of*
431 *differentiation*
432 Because widespread signatures of indirect selection are often assumed to reflect
433 the impact of heterogeneous background selection (BGS) across the genome, several
434 authors have proposed that the correlation in F_{ST} across multiple pairs of taxa can be used
435 as a baseline for detecting genomic regions that have been affected by positive selection
436 (Berner and Salzburger 2015, Burri 2017b). More specifically, loci that have contributed
437 to adaptation or speciation should be detectable as a positive deviation from the common
438 pattern of differentiation (i.e., PC1), which is considered to reflect processes that are
439 unrelated to ecologically relevant positive selection (Burri 2017b).
440 We have a unique opportunity to test the effectiveness of this comparative
441 genomic approach by examining patterns of differentiation around a locus that is known
442 to contribute to adaptation and speciation in subspecies *puniceus*. Using a candidate gene
443 approach, Streisfeld et al. (Streisfeld et al. 2013) showed that the shift from yellow to red
444 flowers in *puniceus* was caused by a *cis*-regulatory mutation in the R2R3-MYB
445 transcription factor *MaMyb2*. Patterns of haplotype variation within the gene indicate that
446 the red allele was subject to strong positive selection and rapidly swept to fixation in
447 what is now the geographic range of the red ecotype (Stankowski and Streisfeld 2015).
448 To test for a linage-specific signature of differentiation at this locus, we examined
449 patterns of Z-F_{ST} in a relatively narrow region (~3 Mbp) surrounding *MaMyb2*. (Fig. 4;
450 Fig. S10). At the 500 kb scale, the comparison between the red and yellow ecotypes
451 shows a sharp peak centered near the flower color locus. The peak is strongly elevated

above PC1 Z-F_{ST} (3.33 standard deviations), indicating that differentiation is indeed accentuated relative to the level observed between other taxon pairs. Further, other comparisons, including those with the red ecotype, do not show accentuated differentiation in this region, indicating that the signal is specific to this taxon pair. The peak is even more pronounced in the analysis at the 100 kb scale, rising 9.5 standard deviations above PC1 Z-F_{ST}. At this scale, the signature of positive selection is visible in other comparisons that include the red ecotype, though the signal is less pronounced than in the comparison with the yellow ecotype (Fig. S10).

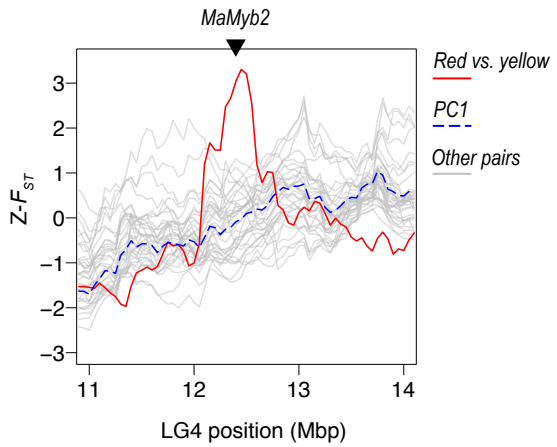

Figure 4. A large-effect adaptive locus shows a lineage-specific signature of positive selection. Plots of Z-transformed F_{ST} across the genome, estimated in 500 kb sliding windows (step size 50 kb). The red line shows values between the red and yellow ecotypes of subspecies *puniceus*, while the gray lines show the values of all other comparisons. The dashed blue line shows the first PC calculated across all of the comparisons. The triangle marks the position of the gene *MaMyb2*. A *cis*-regulatory mutation that is tightly linked to this gene is responsible for the shift from yellow to red flowers. See Fig. S10 for the same plot made for 100 kb windows.

Correlated differentiation landscapes emerge rapidly following a population split

Although correlated landscapes may evolve due to the indirect effects of widespread background selection, it is also conceivable that they could evolve due to other evolutionary processes. To gain insight into the role that background selection may have played in shaping these patterns, we next tested several hypotheses about how BGS is expected to shape the genomic landscape over time. To do this, we used the level of genetic distance between each pair of taxa as a proxy for their divergence time and constructed a temporal picture of genomic landscape evolution that spans the first million years following a population split.

In a verbal model, Burri (Burri 2017b) made several predictions about how correlations between measures of variation and genome features should be impacted by BGS following a population split. First, the correlations between diversity (π) and intrinsic features should be relatively consistent over time, as heterogeneous BGS will continue to constrain patterns of diversity in each daughter population after the split. Second, π and d_{xy} should remain highly correlated with one another, because d_{xy} is largely influenced by ancestral diversity at these short timescales. Third, levels of genome-wide differentiation (F_{ST}) initially should be low and vary stochastically across the genome, due mainly to the sampling effect that accompanies a split. Therefore, F_{ST} should not be correlated with the distribution of genomic properties early in divergence. However, as time passes, and the indirect effects of ancestral and lineage-specific BGS accumulate, levels of differentiation should gradually become more correlated with underlying genome features.

To test these hypotheses, we computed correlations between relevant window-based measures of variation and genomic features between all 36 pairs of taxa. We then plotted these correlations on corresponding measures of between-taxon d_a, which ranged from very low (0.05%) between the parapatric red and yellow ecotypes of subspecies *punicieus*, up to 1.3% in allopatric comparisons that included *M. clevelandii*. We used d_a (d_{xy} minus mean π), because it corrects for sequence variation present in the common ancestor, making it a better proxy of recent divergence time than d_{xy}.

This analysis revealed very clear temporal signatures of genomic landscape evolution, some of which were consistent with the above predictions (Fig. 5; See Fig. S11 for results with d_{xy} as a proxy for time). First, the presence of strong correlations between π (mean of the two taxa) and the distribution of genomic features barely changed with increasing divergence time between populations (Fig. 5A & 5B; Table S6). This is consistent with the prediction that diversity landscapes are inherited from the ancestor and are then maintained by the indirect effects of selection in each daughter population following the split. Second, the relationships between F_{ST} and gene count, F_{ST} and recombination rate, and F_{ST} and π all become stronger as d_a increases (Fig. 5C & 5D; Table S6). Thus, as predicted, the differentiation landscape increasingly reflects the distribution of intrinsic features as divergence time increases.

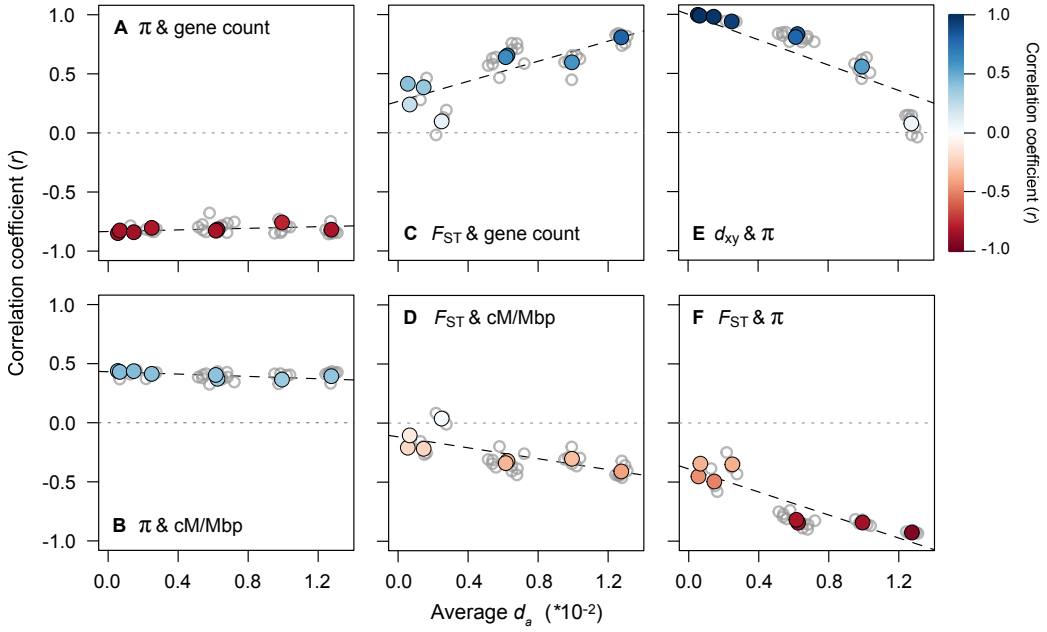

Figure 5. The range of divergence times reveals static and dynamic signatures of recurrent indirect selection. Correlations between variables (500 kb windows) for all 36 taxonomic comparisons (gray dots) plotted against the average d_a as a measure of divergence time. The left panels show how the relationships between π (each window averaged across a pair of taxa) and (A) gene count and (B) recombination rate vary with increasing divergence time. The middle panels (C & D) show the same relationships, but with F_{ST}. The right panels show the relationships between (E) d_{xy} and π and (F) F_{ST} and π. The regressions (dashed lines) in each plot are fitted to the eight independent contrasts (colored points) obtained using a phylogenetic correction. The color gradient shows the strength of the correlation.

However, two of the observed patterns differed markedly from existing predictions (Burri 2017b). First, π and d_{xy} did not remain highly correlated over this relatively short timeframe. Although the correlation is almost perfect at the earliest time points ($r > 0.99$), it decays rapidly and is not significantly different from zero in the most diverged taxon pairs (Fig. 5E). In addition, the correlations between F_{ST} and proxies of the strength of indirect selection are much higher than expected at these early divergence times. For example, for the red and yellow ecotypes, the correlations between F_{ST} and gene density ($r = 0.415$), recombination rate ($r = -0.208$), and mean π ($r = -0.452$) are substantial and highly significant ($p < 0.0001$) at a very early stage of divergence. The strength of these correlations also increases rapidly, with the diversity and differentiation landscapes almost perfectly mirroring one another in the most divergent comparisons (Fig. 5F; $r = -0.94$; Fig. 6; Fig. S12). As our results only partially match the above predictions, it is possible that other forces aside from BGS are responsible for the rapid emergence of differentiation landscapes in this system. Therefore, we next considered the potential for other evolutionary processes to generate these patterns.

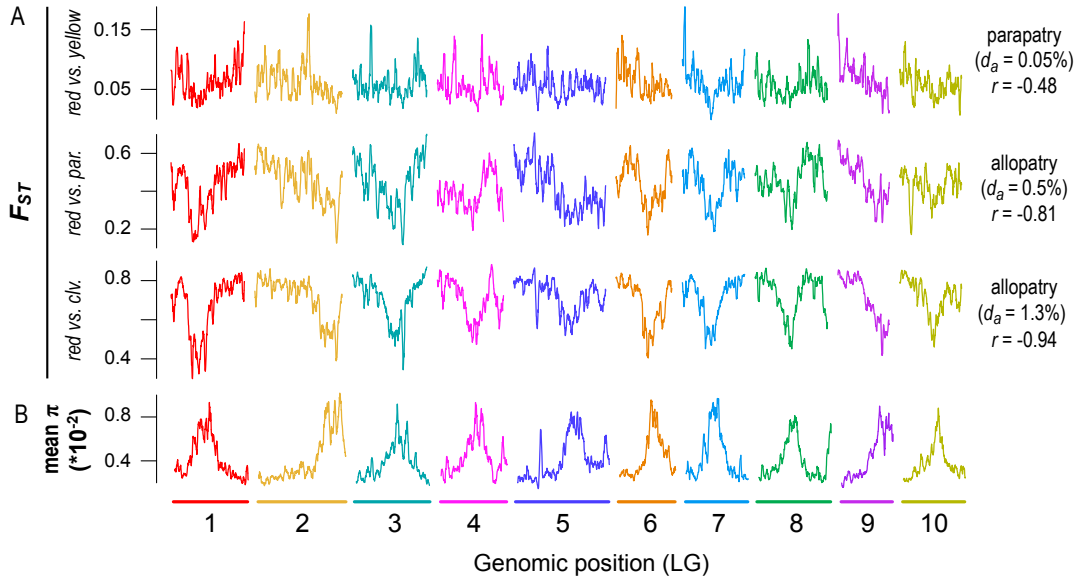

Figure 6. Emergence of a heterogeneous differentiation landscape across one million years of divergence. A) Plots of F_{ST} (500 kb windows) across the genome for pairs of taxa at early (red vs. yellow), intermediate (red vs. *parviflorus*), and late stages (red vs. *M. clevelandii*) of divergence. B) Average nucleotide diversity (for the red ecotype of subspecies *puniceus*, yellow ecotype of subspecies *puniceus*, subspecies *parviflorus*, and *M. clevelandii*) across the genome in 500 kb windows. The geographic distribution (parapatric or allopatric), sequence divergence ($d_a \times 10^{-2}$), and correlation between F_{ST} and mean π are provided next to each taxon pair.

Simulations suggest that adaptation has played an important role in genomic landscape evolution

To assess the plausibility of different modes of selection for generating the observed genomic landscapes and temporal patterns, we used individual-based simulations in SLiM (Haller et al. 2019, Haller and Messer 2019). Our basic model

557 consisted of an ancestral population (N = 10,000) that evolved for 10N generations and
558 then split into two daughter populations that diverged for a further 10N generations. Each
559 individual carried a 21 Mbp chromosome (similar to the physical size of a bush
560 monkeyflower chromosome), partitioned into three equally sized regions. In the central
561 third, all mutations were neutral, while some mutations in the two distal ends could affect
562 fitness, depending on the scenario. While simple, this partitioning scheme generates
563 broad-scale variation in the strength of indirect selection across the chromosome and
564 roughly approximates the distributions of genomic features in this system (e.g., LG 7 in
565 Fig. 2).
566 We implemented six modifications to this basic model: (*i*) *Neutral divergence*: no
567 mutations affect fitness; (*ii*) *Background selection* (BGS): non-neutral mutations are
568 deleterious; (*iii*) *Bateson-Dobzhansky-Muller incompatibility* (BDMI): As *ii*, except that
569 after the split, a fraction of non-neutral mutations is deleterious in one population and
570 neutral in the other (randomly chosen). (*iv*) *Positive selection*: non-neutral mutations have
571 positive fitness effects. (*v*) *BGS and positive selection*: *ii* and *iv combined*; (*vi*) *Local*
572 *adaptation:* As *iv*, but after the split some non-neutral mutations are beneficial in one
573 population and neutral in the other. Migration between populations was allowed only in
574 scenarios *iii* and *vi*. See Methods for more details and Table S7 for parameter values. To
575 summarize the results, measures of variation (π, d_{xy}, and F_{ST}) were calculated in 500 kb
576 regions at 10 time points for each simulation.
577 This broad range of scenarios generated quite different patterns of variation, both
578 through time and across the chromosome (Fig. 7; Fig. S13). As expected, the neutral
579 model did not produce heterogeneous genomic landscapes. Background selection
580 produced a diversity landscape characterized by higher diversity in the unconstrained
581 central third of the chromosome (Fig 7; Fig S13). Consistent with the predictions of Burri
582 (2017a), π and d_{xy} remained highly correlated over the course of the simulations.
583 However, even with a high proportion of deleterious mutations (10%) and substantial
584 fitness effects (1%), the variation in π and d_{xy} was modest, as was variation in the
585 differentiation landscape (Fig. 7).
586 Under some conditions, the BDM incompatibility model produced patterns of
587 variation that were more similar to our empirical findings than the BGS model (Fig. 7,
588 S13). In all simulations, patterns of diversity and divergence were relatively homogenous
589 at the time of the split. However, selection against gene flow induced by incompatibility
590 loci on the chromosome ends caused patterns of variation to become highly structured
591 over time. Indeed, higher π, lower d_{xy}, and lower F_{ST} were observed in the neutrally
592 evolving chromosome center due to the homogenizing effects of higher gene flow in this
593 region. As in our empirical data, the correlation between π and d_{xy} decayed over the
594 course of some simulations, and even became negative depending on rates of gene flow
595 and selection. Although we simulated this scenario starting with undifferentiated
596 populations (i.e., primary contact), widespread BDMIs should produce the same patterns
597 for populations that diverged in allopatry and came back into secondary contact. This is
598 because gene flow will have a stronger homogenizing effect in regions of the genome that
599 are not associated with incompatibilities (Durrett et al. 2000).
600 The three models with positive selection (Positive, BGS & Positive, and Local
601 adaptation) all produced highly heterogeneous genomic landscapes across the range of
602 parameter values explored (Fig. 7; Fig. S13). In all cases, positive selection in the

ancestor created a highly heterogeneous diversity landscape that was inherited by each daughter population and was maintained for the duration of the simulation. As in our data, π and d_{xy} were perfectly correlated at early divergence times, but the correlation decayed rapidly and even became negative in scenarios with stronger selection. Unlike in the BDMI scenario, where the correlation between π and d_{xy} decayed due to higher gene flow in the chromosome center, it broke down in these simulations because positive selection caused d_{xy} to increase more rapidly in the chromosome ends (Fig. S13). Highly heterogeneous differentiation landscapes, characterized by lower F_{ST} in the chromosome center, emerged rapidly in all three scenarios. This pattern appeared irrespective of the parameter values examined and was not heavily influenced by other factors, like gene flow, the inclusion of BGS, or whether adaptation occurred from standing variation. However, some of these factors may have a larger influence in simulations with different non-neutral mutation rates and selection strengths.

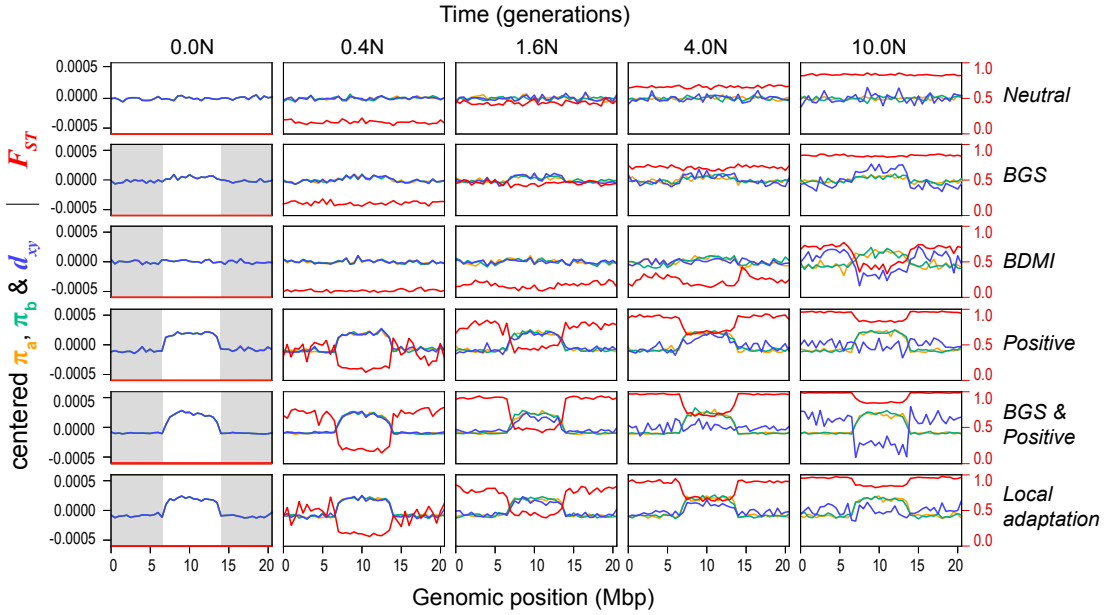

Figure 7. Genomic landscapes simulated under different divergence histories. Each row of plots shows patterns of within- and between-population variation (π, d_{xy}, and F_{ST}) across the chromosome (500 kb windows) at five time points (N generations, where N = 10000) during one of scenarios The selection parameter (Ns, where s = Ns/N), proportion of deleterious (-) and positive mutations (+), and number of migrants per generation (Nm; 0 unless stated) for these simulations are as follows: (*i*) neutral divergence (no selection), (*ii*) background selection (-Ns = 100; -prop = 0.01), (*iii*) Bateson-Dobzhansky-Muller incompatibilities (BDMI) (-Ns = 100, -prop = 0.05, Nm = 0.1), (*iv*) positive selection (+Ns = 100, +prop = 0.01), (*v*) background selection and positive selection (-Ns = 100, -prop = 0.01; +Ns = 100, +prop = 0.005), and (*vi*) local adaptation (+Ns = 100, +prop = 0.001, Nm = 0.1). The gray boxes in the first column show the areas of the chromosome that are experiencing selection, while the white central area evolves neutrally. Note that π (in populations a and b) and d_{xy} have been mean centered so they can be viewed on the same scale. Un-centered values and additional simulations with different parameter combinations and more time points can be found in Fig. S13.

634 Overall, the results of these simulations suggest that background selection (BGS)
635 is not primarily responsible for heterogeneous genomic landscapes. Although our
636 simulations should be interpreted cautiously, because we have not thoroughly explored
637 parameter space associated with each model, the results of other recent simulation studies
638 also support this conclusion. For example, Matthey-Doret & Whitlock (Matthey-Doret
639 and Whitlock 2018) simulated population divergence with BGS under different scenarios
640 with simulations using parameters estimated from humans and stickleback. They found
641 that BGS was unable to generate heterogeneous differentiation landscapes over short
642 timescales, with or without gene flow. Similarly, Rettelbach *et al*. (Rettelbach et al. 2019)
643 simulated the evolution of diversity landscapes using empirical estimates of
644 recombination rate and functional densities from the collared flycatcher genome. They
645 found that BGS was able to generate modest variation in levels of diversity across
646 chromosomes, but they concluded that it was not sufficient to explain the pronounced
647 dips observed in empirical studies of flycatcher genomes (e.g., (Burri et al. 2015)). Both
648 of these studies suggest that other processes (probably positive selection) are responsible
649 for generating the observed patterns (Matthey-Doret and Whitlock 2018, Rettelbach et al.
650 2019).
651 Our simulations suggest that divergence histories involving positive selection
652 and/or incompatibilities are plausible explanations for heterogeneous genomic landscapes
653 that emerge rapidly after a population split. However, in order to account for the presence
654 of strong correlations between measures of variation and genomic features across
655 multiple comparisons, these scenarios assume that the genomic basis of adaptation and/or
656 reproductive isolation is diffuse, shared across multiple taxa, and evolves rapidly.
657 Although this was once considered unlikely, recent studies suggest that these conditions
658 may be common. For example, adaptation and speciation are now thought to be highly
659 polygenic (Rockman 2012, Jiggins and Martin 2017, Martin et al. 2019), and recent
660 evidence suggests that adaptation draws heavily on ancestral mutations, rather than those
661 arising independently in multiple, related descendent lineages (Barrett et al. 2008, Nelson
662 and Cresko 2018, Marques et al. 2019). Similarly, the interaction between widespread
663 hybrid incompatibilities (intrinsic or extrinsic) and intrinsic genomic features is thought
664 to have caused similar patterns of genome-wide variation to evolve between different
665 pairs of hybridizing taxa (Martin et al. 2019).
666 Given the rapid and extensive trait diversification that has occurred during the
667 bush monkeyflower radiation (Stankowski et al. 2015, Chase et al. 2017, Sobel et al.
668 2019), positive selection has probably played an important role in shaping patterns of
669 genome-wide variation across the group. This includes the striking divergence of floral
670 traits (Fig. 1A), which is thought to be due to divergent pollinator-mediated selection
671 across southern California (Grant 1981, 1993b, a). In the taxa that have been studied in
672 detail, crossing experiments have shown that most of these traits are polygenic, and there
673 is evidence that they are targets of selection (Sobel and Streisfeld 2015, Stankowski et al.
674 2015). Intrinsic genetic incompatibilities may also influence patterns of variation between
675 some of the taxa, but they are unlikely to explain the correlated landscapes that have
676 evolved across the whole group, as most of these taxa show little or no evidence for
677 intrinsic incompatibilities between them (McMinn 1951, Sobel and Streisfeld 2015).
678 However, local adaptation also can generate a porous isolating barrier (Wu 2001), so it is
679 likely that gene flow has contributed to genomic landscape evolution among some taxa,

680 particularly those that are geographically proximate and/or hybridize in areas where their
681 distributions overlap.
682
683 *Radiation-wide evidence for selection against gene flow*
684 To determine if gene flow has contributed to the evolution of these correlated
685 genomic landscapes, we first tested for evidence of introgression among these taxa using
686 *D*-statistics (Green et al. 2010). Patterson's *D* measures asymmetry between the numbers
687 of sites with ABBA and BABA patterns (where A and B are ancestral and derived alleles,
688 respectively) across three in-group taxa and an out-group (in our case, *M. clevelandii*)
689 that have the relationship (((P1, P2), P3) O). A significant excess of either pattern gives a
690 non-zero value of *D*, which is taken as evidence that gene flow has occurred between P3
691 and one of the in-group taxa (Green et al. 2010). Overall, these tests provide evidence
692 that gene flow has occurred during this radiation. All 48 of the appropriate four-taxon
693 combinations yielded significant non-zero values of Patterson's *D* ($p < 0.0001$; Table S7;
694 Fig. S14). The admixture proportions (*f*) for each test indicate that an average of 2.5% of
695 the genome has been transferred between pairs of in-group taxa, but the proportion varies
696 among the tests, ranging from less than 1% to more than 10% (Fig. S4.
697 Although gene flow initially involves the exchange of whole genomes between
698 populations, selection against gene flow reduces effective migration (m_e) in regions of the
699 genome that are associated with barrier loci (Jiggins and Martin 2017, Ravinet et al.
700 2017). Therefore, porous isolating barriers should result in a heterogeneous pattern of
701 introgression across the genome (Barton and Hewitt 1985, Wu 2001). To test this
702 hypothesis, we calculated the f_d statistic, a version of the admixture proportion modified
703 for application to genomic windows (Martin et al. 2015). As predicted, estimates of f_d,
704 varied markedly among genomic regions (Fig. 2; Fig. S14). Most windows showed levels
705 of f_d at or near zero, indicating that foreign alleles have been purged from these regions
706 by selection. However, a proportion of windows showed substantial admixture
707 proportions. In some tests, values of f_d exceeded 0.5, indicating that more than half of the
708 sites in some windows have been shared between taxa.
709 Because gene flow opposes differentiation, we would expect to see lower F_{ST} in
710 regions with a higher proportion of introgressed variants. Consistent with this prediction,
711 we observed a substantial negative correlation ($r = -0.57$; $p < 0.0001$) between mean f_d
712 (for each window, averaged over the 48 tests) and PC1 F_{ST}. This result is not driven by a
713 limited number of the four-taxon tests, as 44 of the 48 comparisons showed the same
714 significant negative relationship (Table S15). Also consistent with models of widespread
715 selection against gene flow, regions of the genome with higher f_d scores tended to have
716 higher diversity ($r = 0.45$; $p < 0.0001$), lower tree concordance scores ($r = -0.59$; $p <$
717 0.0001), fewer functional genes ($r = -0.36$; $p < 0.0001$), and a higher recombination rate
718 ($r = -0.18$; $p < 0.0001$).
719 Overall, these results indicate that widespread gene flow has played a key role in
720 the formation of the genomic landscape in this system. In addition to reductions in
721 diversity and increased differentiation owing to selection against gene flow, the
722 persistence of introgressed variants has probably resulted in higher levels of diversity in
723 regions with fewer genes and higher recombination rate. The four-taxon tests show that
724 the impact of gene flow is widespread across the radiation, though some caution should
725 be exercised when interpreting the specific pattern of individual introgression events.

726 Given that there is the potential for gene flow between so many taxa and ancestral
727 lineages, it is difficult to infer the source and timing of individual admixture events. For
728 example, rather than reflecting recent gene flow between all pairs of taxa, some
729 introgression events may have occurred deeper in history, and their consequences
730 inherited by multiple taxa. Although more sophisticated methods will be needed to model
731 gene flow across this group, these results clearly show that it has contributed to the rapid
732 evolution of correlated genomic landscapes during this radiation.
733
734 *Conclusions and implications for understanding genomic landscape evolution and the*
735 *basis of adaptation and speciation*
736 Facilitated by a new chromosome-level genome assembly, we have shed light on
737 the causes of correlated genomic landscapes across a radiation of monkeyflowers.
738 Adaptive divergence and gene flow are hallmarks of rapid radiations (Schluter 2000), and
739 our data suggest that the indirect effects of selection resulting from these processes have
740 contributed to a common pattern of differentiation among these taxa. Our 'time-course'
741 analysis shows that the common landscape emerged rapidly after populations split and
742 has become more correlated with the distribution of genomic features as divergence time
743 has increased. Although background selection may play some role, its effects are
744 probably too subtle to have made a strong contribution to the correlations during over the
745 timeframe of this radiation.
746 In addition to enhancing our understanding of the processes that have shaped the
747 genomic landscape during this radiation, our study contributes toward a more general
748 understanding of the role that natural selection plays in shaping genome-wide variation.
749 In line with the findings of other recent studies, our results indicate that little, if any, of
750 the genome evolves free of the effects of natural selection (Hahn 2008, Corbett-Detig et
751 al. 2015). Moreover, our 'time-course' analysis shows that between-taxon signatures of
752 selection can emerge very rapidly after a population split and can be substantial, even
753 between populations at the early stages of divergence. Overall, these results suggest that
754 patterns of between-population variation cannot be understood without taking the effects
755 of natural selection into account.
756 An important point arising from this work is that multiple divergence histories can
757 generate heterogeneous differentiation landscapes that are correlated both among taxa
758 and with the distribution of intrinsic genomic properties. When divergence is recent,
759 possible explanations include polygenic adaptation across multiple populations and
760 porous barriers to gene flow arising from divergent ecological selection and/or intrinsic
761 incompatibilities. Although it has often been assumed that recurrent background selection
762 is the primary cause of correlated landscapes, it is important to remember that all forms
763 of selection can indirectly impact levels of variation at associated sites. In fact, our
764 simulations show that alternative explanations may be more likely when divergence times
765 are short and there is opportunity for gene flow among taxa. Thus, we advocate for a
766 more nuanced approach when interpreting correlated differentiation landscapes, rather
767 than assuming that they are caused by a single evolutionary process.
768 Finally, our results have important practical implications for studies attempting to
769 identify the genomic basis of adaptation and speciation from patterns of genome-wide
770 variation. Indeed, detecting these loci is a major goal of adaptation and speciation studies,
771 and genome scans are now commonly used to identify promising candidate regions

(Ravinet et al. 2017). In an effort to correct for the potentially confounding effects of background selection (BGS), it has been proposed that the correlation in F_{ST} among multiple population pairs can be used as a baseline for outlier detection (Berner and Salzburger 2015, Burri 2017b). A core premise of this method is that correlated differentiation landscapes are caused primarily by BGS, so removing this signature should expose the loci relevant to adaptation and/or speciation. Although this approach may be successful in identifying large-effect loci targeted by lineage-specific positive selection, as we illustrated for the *MaMyb2* locus, we caution against treating the common pattern of differentiation simply as background noise. Rather than parsing out the effects of BGS, studies that use this approach may actually be discarding the genomic signature of polygenic adaptation and speciation.

Data Accessibility

Raw sequencing reads used for the genome assembly, linkage map construction, and genome resequencing and are available on the Short-Read Archive (SRA) under the bioproject ID xxx. The genetic map, annotation, VCF file, and analysis scripts have been deposited on DRYAD. The reference assembly is available at mimubase.org. Code used to run the simulations is available on Github at: https://github.com/mufernando/mimulus_sims.

Acknowledgments

We thank William A. Cresko, Thomas Nelson, Roger K. Butlin, Anja M. Westram, Mark Ravinet, Sarah. P. Otto, and Michael C. Whitlock for stimulating discussion. John Willis performed the Illumina mate-pair library preps used for the genome assembly. Julian Catchen, Clay Small, Susan Bassham, Mark Rausher and Janna Fierst provided technical or analytical support or advice. Doug Turnbull and Maggie Weitzman conducted the Illumina sequencing at the University of Oregon Genomics Core facility. Thomas Nelson, Martin Garlovsky, Roger K. Butlin, Anja M. Westram and Jeff Ross-Ibarra provided comments on an earlier version of this manuscript. We thank Reto Burri, Stuart J.E. Baird, and two anonymous reviewers for providing insightful comments that greatly improved the paper. We also thank people who contributed to insightful discussions at the 2017 Speciation Genomics Meeting, Cambridge UK, the 2019 Gordon Speciation Conference, Ventura CA, USA, and on Twitter. Funding was provided by the National Science Foundation grant DEB-1258199 to MS and the Sloan Foundation to PR.

Materials and Methods

Genome Assembly

We used a combination of short-read Illumina and long-read Single Molecule, Real Time (SMRT) sequencing to assemble the genome of a single individual from the red ecotype of *M. aurantiacus* subspecies *puniceus* (population UCSD; Table S1). Genomic DNA was isolated from leaf tissue using either ZR plant/seed DNA miniprep kits (Zymo Research) or GeneJet Plant Genomic DNA purification kits (Thermo Fisher). Illumina libraries were generated following the *Allpaths-LG* assembly pipeline (Gnerre et al. 2011), which included a single fragment library with average 180 bp insert size and three mate pair libraries (average insert sizes: 3.5-5 kb, 5-7 kb, and 7-13 kb). Libraries

818 were sequenced on the Illumina HiSeq 2500 using paired-end 100 bp reads. An initial
819 scaffold-level assembly was performed with *Allpaths-LG* using default parameters and
820 the *haploidify* function enabled. This assembly yielded 11,123 contigs (N50 = 40.5 kb)
821 and 2,299 scaffolds (N50 = 1,310 kb), for a total assembly size of 193.3 Mbp. Long-read
822 sequencing was performed from the same individual using 12 SMRT cells sequenced on
823 the Pacific Biosystems RS II machine at Duke University. We obtained a total of 6.4 Gbp
824 of sequence, which corresponds to ~21× coverage of the genome. The PacBio reads were
825 used to re-scaffold the *Allpaths-LG* scaffolds using *Opera-LG* (Gao et al. 2016). This
826 reduced the number of scaffolds to 1,547 (N50 = 1,578 kb).
827 We then manually improved the scaffold containing the flower color gene
828 *MaMyb2* (Streisfeld et al. 2013). We first aligned this scaffold to a previously published
829 draft sequence assembly from this same individual (Stankowski et al. 2017), which was
830 generated using Illumina short-reads and the *Velvet* assembler (Zerbino and Birney
831 2008). We used long range PCR and cloning to generate Sanger sequences across three
832 regions within 20 kb of *MaMyb2* that did not assemble well. Genomic DNA was
833 amplified using Phusion high fidelity polymerase (NEB). PCR products were cloned into
834 the pCR2.1 TOPO-TA vector (Life Technologies), and purified plasmids were sequenced
835 with Sanger technology. Resulting sequences were aligned to the scaffold containing
836 *MaMyb2*, and new PCR primers were designed to sequence internal fragments until the
837 entire insert was sequenced. Using this approach, we sequenced a total of 9,824 bp across
838 the three regions. The reference sequence in the assembly was corrected manually to
839 match the Sanger data.
840 Finally, we gap filled the assembly using the PacBio data and the program
841 *PBJelly* (English et al. 2012). Resulting scaffolds were assembled into pseudomolecules
842 using *Chromonomer* (http://catchenlab.life.illinois.edu/chromonomer/), according to the
843 online manual. This software anchored and oriented scaffolds based on the order of
844 markers in a high-density linkage map (see below) and made corrections to scaffolds
845 when differences occurred between the genetic and physical positions of markers in the
846 map. A final round of gap filling with *PBJelly* was performed to fill any gaps that were
847 created by broken scaffolds in *Chromonomer*. To assess the completeness of the gene
848 space in the assembly, we used both the BUSCO and CEGMA pipelines to estimate the
849 proportion of 956 single copy plant genes (BUSCO) or 248 core eukaryotic genes
850 (CEGMA) that were completely or partially assembled (Parra et al. 2007, Simao et al.
851 2015). The proportion of these genes present in an assembly has been shown to be
852 correlated with the total proportion of assembled gene space, and thus serves as a good
853 predictor of assembly completeness.
854
855 *Construction of high-density linkage map*
856 We generated an outbred F_2 mapping population by crossing two F_1 individuals,
857 each the product of crosses between different greenhouse-raised red and yellow ecotype
858 plants collected from one red ecotype and one yellow ecotype population (populations
859 UCSD and LO, respectively; Table S1). We then used restriction-site associated DNA
860 sequencing (RADseq) to genotype F_1 and F_2 individuals. DNA was extracted from leaf
861 material using Zymo ZR plant/seed DNA miniprep kits, and RAD library preparation
862 followed the protocol outlined in Sobel and Streisfeld (Sobel and Streisfeld 2015).

863 Libraries were sequenced on the Illumina HiSeq 2000 platform using single-end 100 bp
864 reads at the Genomics Core Facility, University of Oregon.
865 Reads were filtered based on quality, and errors in the barcode sequence or RAD
866 site were corrected using the *process_radtags* script in *Stacks v. 1.35* (Catchen et al.
867 2011, Catchen et al. 2013). Loci were created using the *denovo_map.pl* function of
868 *Stacks*, with three identical raw reads required to create a stack, two mismatches allowed
869 between loci for an individual, and two mismatches allowed when processing the catalog.
870 Single nucleotide polymorphisms (SNPs) were determined and genotypes called using a
871 maximum-likelihood (ML) statistical model implemented in *Stacks* and a stringent χ^2
872 significance level of 0.01 to distinguish between heterozygotes and homozygotes. We
873 then used the *genotypes* program implemented in *Stacks* to identify a set of 9,029
874 mappable markers. We specified a 'CP' cross design (F_1 individuals coded as the
875 parents), requiring that a marker was present in at least 85% of progeny at a minimum
876 depth of 12 reads per individual, and we allowed automated corrections to be made to the
877 data.
878 Linkage map construction was performed using *Lep-MAP2* (Rastas et al. 2016).
879 The data were filtered using the *Filtering* module to include only individuals with less
880 than 15% missing data and excluded markers that showed evidence for extreme
881 segregation distortion (χ^2 test, $P < 0.01$). To assign markers to linkage groups, we used
882 the *SeparateChromosomes* module with a logarithm of odds (LOD) score limit of 20 and
883 no minimum size for linkage groups (LG). This assigned 7,217 markers to 10 linkage
884 groups, which matches the number of chromosomes in *M. aurantiacus*. The *JoinSingles*
885 module was executed again with a LOD limit of 10 to join an additional 877 ungrouped
886 markers to the 10 previously formed LGs. Fifty-seven singles that were not joined at this
887 stage were discarded from the dataset. Initial marker orders were determined using sex-
888 averaged and sex-specific recombination rates using the *OrderMarkers* module. For each
889 LG, we conducted 10 independent runs using the Kosambi mapping function
890 (*useKosambi=1*), with the dataset split into seven pseudo-families to take advantage of
891 parallel processing. When multiple markers had identical genotypes, only the duplicate
892 marker with the least missing data was used in marker ordering. We retained the marker
893 order from the run with the best likelihood. After removing markers with an error rate >
894 0.05, the ML order was re-evaluated using the *evaluateOrder* flag. The map contained
895 8,094 informative loci from 269 F2 individuals, with an average of 3.5% ± SD 3.86
896 missing data per individual.
897 After the integration of our assembly and genetic map using the *Chromonomer*
898 software (Amores et al. 2014), we made corrections to the map order based on the
899 physical position of markers within assembled scaffolds. Using the output of
900 *Chromonomer*, we identified markers that were out of order in the map compared to their
901 local assembly order and aligned these markers to the assembly from *Chromonomer*
902 using *Bowtie2* v. 2.2.5 (Langmead and Salzberg 2012) with the *very_sensitive* settings to
903 obtain their physical order. We then re-estimated the map using the *evaluateOrder* flag in
904 *Lep-MAP2* as described above, but with the marker order constrained to the physical
905 order (*improveOrder=0*) and with all duplicate markers included in the analysis
906 (*removeDuplicates=0*). After initial map construction, we removed 17 markers with an
907 estimated error rate greater than 5% and estimated the map one last time using the same
908 settings. The final map contained 7,589 markers across the 10 linkage groups.

909
910 *Genome annotation*
911 Prior to genome annotation, the assembly was soft-masked for repetitive elements
912 and areas of low complexity with *RepeatMasker* (RepeatMasker Open-4.0) using a
913 custom *Mimulus aurantiacus* library created by *RepeatModeler* (RepeatModeler Open-
914 1.0), Repbase repeat libraries (Jurka et al. 2005), and a list of known transposable
915 elements provided by *MAKER* (Holt and Yandell 2011). In total, 30.99% of the genome
916 assembly was masked by *RepeatMasker*. Repetitive elements were annotated with
917 *RepeatModeler*. Hidden Markov Models for gene prediction were generated by *SNAP*
918 (Korf 2004) and *Augustus* (Stanke and Waack 2003) and were trained iteratively to the
919 assembly using *MAKER*, as described by Cantarel et al. (Cantarel et al. 2008). Training
920 was performed on the 14.5 Mbp sequence from LG9. Evidence used by *MAKER* for
921 annotation included protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Solanum*
922 *lycopersicum*, *Solanum tuberosum*, *Daucus carota*, *Vitis vinifera* (all downloaded from
923 EnsemblPlants on 9 August 2016), *Salvia miltiorrhiza* (downloaded from Herbal
924 Medicine Omics Database on 9 August 2016), *Mimulus guttatus v. 2* (downloaded from
925 JGI Genome Portal on 9 August 2016), and all Uniprot/swissprot proteins (downloaded
926 on 18 August 2016) (Goodstein et al. 2012, Nordberg et al. 2013, Kersey et al. 2016)
927 (Herbal Medicine Omics Database; Uniprot). We filtered the annotations with *MAKER* to
928 include: 1) only evidence-based information that also contained assembled protein
929 support, and 2) those *ab initio* gene predictions that did not overlap with the evidence-
930 based annotations and that contained protein family domains, as detected with
931 InterProScan (Quevillon et al. 2005).
932
933 *Genome re-sequencing and variant calling*
934 We collected leaf tissue from each taxon (Collection locations available in Table
935 S3 and Fig. S2) and extracted DNA from dried tissue using the Zymo Plant/Seed
936 MiniPrep DNA kit following the manufacturer's instructions. We prepared sequencing
937 libraries using the Kapa Biosystems HyperPrep kit, and libraries with an insert size
938 between 400-600 bp were sequenced on the Illumina HiSeq 4000 using paired-end 150
939 bp reads at the Genomics Core Facility, University of Oregon.
940 We filtered raw reads using the *process_shortreads* script in *Stacks* v1.46 to
941 remove reads with uncalled bases or poor quality scores. We then aligned the retained
942 reads to the reference assembly using the BWA-MEM algorithm in *BWA* v0.7.15 (Li
943 2013). An average of 91.7% of reads aligned (range: 82.6-96.0%), and the average
944 sequencing depth was 21x (range: 15.16 – 30.86x). We then marked PCR duplicates
945 using *Picard* (http://broadinstitute.github.io/picard). We performed an initial run of
946 variant calling using the UnifiedGenotyper tool in *GATK* v3.8 (McKenna et al. 2010) and
947 filtered the data to remove variants with a mapping quality < 50, a quality depth < 4, and
948 a Fisher Strand score > 50. We then used these variants to perform base quality score
949 recalibration for each individual, before performing another run of the UnifiedGenotyper
950 to call final variants. After the second run of variant calling, we removed variants with a
951 mapping quality < 40, a quality depth < 2, and a Fisher Strand score > 60. The final
952 dataset contained 13,233,829 SNPs across the nine taxa. Finally, we ran
953 UnifiedGenotyper with the EMIT_ALL_SITES option to output all variant and invariant
954 genotyped sites.

955
956 *Phylogenetic and principal component analyses*
957 We used *RAxML* v8 to reconstruct the evolutionary relationships among the nine
958 taxa by concatenating variant sites from across the genome. To investigate patterns of
959 phylogenetic discordance across the genome, we also built trees from windows across the
960 genome. We phased SNPs using *BEAGLE* v4.1 (Browning and Browning 2007), using a
961 window size of 100 kb and an overlap of 10 kb. We incorporated information on
962 recombination rate from the genetic map and did not impute missing genotypes. After
963 phasing, we used *MVFtools* (https://www.github.com/jbpease/mvftools) to run *RAxML*
964 from 100 kb and 500 kb nonoverlapping windows, with the *M. clevelandii* samples set as
965 the outgroups. We then visualized the window trees in *DensiTree* v2.01 (Bouckaert
966 2010).
967 To assess concordance between the window-based trees and the whole-genome
968 tree, we converted trees to distance matrices using the *Ape* package in R (Paradis et al.
969 2004). We then calculated the Pearson's correlation coefficient between the distance
970 matrix from each window and the whole-genome tree, with a stronger correlation
971 indicating higher concordance with the whole-genome tree. We used one-dimensional
972 autocorrelation analysis to determine if variation in tree concordance was randomly
973 distributed across the genome. This involved estimating the autocorrelation between
974 genomic position and tree concordance for each LG with a lag size of 2 Mbp. The
975 significance of the observed value for each LG was determined from a null distribution of
976 autocorrelation coefficients estimated from 1000 random permutations of the genome-
977 wide data.
978 We also conducted a Principal Components Analysis (PCA) based on all variant
979 sites from across the entire genome using *Plink* v. 1.90 (Chang et al. 2015). Initially, we
980 ran the PCA with all 37 samples, but we consecutively re-ran the analysis by removing
981 different taxa in order to assess clustering patterns among more closely related samples.
982
983 *Population genomic analyses*
984 To examine how genome-wide patterns of diversity, differentiation, and
985 divergence varied among taxa, we calculated nucleotide diversity (π), between-taxon
986 differentiation (F_{ST}), and between-taxon divergence (d_{xy}), respectively, in non-
987 overlapping and overlapping 100 kb (step size = 10kb) and 500 kb (step size = 50 kb)
988 windows using custom Python scripts downloaded from
989 https://github.com/simonhmartin/genomics_general. We calculated measures of
990 differentiation and divergence across all 36 pairwise comparisons among the nine taxa,
991 and diversity was estimated separately for each taxon. These scripts estimated π and d_{xy}
992 by dividing the number of sequence differences within a window (either within or
993 between taxa) by the total number of sites in that window. To account for missing data,
994 the script counted the number of differences between each sample, divided by the total
995 number of variant sites that were genotyped within those samples, and then averaged
996 across all pairs of samples. To provide an unbiased estimate of diversity and divergence,
997 we incorporated invariant sites into the calculation by dividing the number of pairwise
998 differences (within and between taxa, respectively) by the total number of genotyped
999 sites (variant and invariant) within a window. F_{ST} was calculated following the measure
1000 of K_{ST} (Hudson et al. 1992), equation 9), but was modified to incorporate missing data

using the same approach as π and d_{xy}. We filtered the data separately for each taxonomic comparison, so that each site was required to be genotyped in at least three individuals for comparisons within the *M. aurantiacus* complex or at least two individuals for each comparison involving *M. clevelandii.*

We summarized the variation in each statistic across comparisons using a Principal Components Analysis (PCA), with taxon or taxon pair as the variables. Thus, across each window, the first principal component of π, F_{ST}, and d_{xy} provided multivariate measures that explained the greatest covariance in the data. We used a one-dimensional autocorrelation analysis and permutation test to determine if the genome-wide patterns of PC1 π, F_{ST}, and d_{xy} departed from a random expectation, as described above for tree concordance (see section 'phylogenetic analyses').

To examine the relationships among PC1 diversity, differentiation, and divergence, we estimated Pearson's correlation coefficient among all three statistics across genomic windows. Further, we estimated correlations among these three statistics and tree concordance, gene density, and recombination rate. Recombination rate was estimated by comparing the genetic and physical distance (in cM/Mbp) between all pairs of adjacent markers on each LG from the genetic linkage map described above. We removed the top 5% of recombination rates, as these represented unrealistically high values of recombination. A minimum of three estimates was required to obtain an average recombination rate estimate within each window. Gene density was calculated from the number of predicted genes in each window, as determined from the annotation described above. Genes spanning two windows were counted in both.

To determine how the correlations among the statistics (diversity, differentiation, divergence, recombination rate, gene count) changed with increasing divergence time, we examined the correlation coefficient among all pairs of statistics individually for each of the 36 pairwise comparisons. Because diversity was measured within taxa, we calculated the mean value of π between each pair of taxa. Also, because many of the pairwise comparisons are non-independent, we applied the phylogenetic correction outlined by (Felsenstein 1985, Coyne and Orr 1989) to produce a statistically independent set of data points for this analysis.

We calculated the divergence time between *M. clevelandii* and *M. aurantiacus*, based on a corrected estimate of sequence divergence (d_a) between individuals of *M. clevelandii* and all subspecies of *M. aurantiacus* combined. We then converted this value into a divergence time T (in generations) using the equation: $T = d_a/(2\mu)$, where μ is the mutation rate, 1.5×10^{-8} (Koch et al. 2001, Brandvain et al. 2014). This value was then converted into years by multiplying by a generation time of two years.

Simulations

To assess the plausibility of different scenarios in producing heterogeneous genomic landscapes, we implemented forward-time Wright-Fisher simulations using SLiM (Haller et al. 2019, Haller and Messer 2019). The basic model consisted of an ancestral population with a fixed population size of N=10,000 that split after 10N non-overlapping generations into two daughter populations, each with a fixed size of N. These populations were then allowed to evolve for a further 10N non-overlapping generations. We simulated a 21 Mbp chromosome with a recombination rate of 2×10^{-8} and a mutation rate of 10^{-8}, both per base pair and per generation.

1047 We explored the following six modifications of this basic model. (*i*) *Neutral*
1048 *evolution*: mutations did not impact fitness; (*ii*) *Background selection* (BGS): mutations
1049 in the ancestor and daughter populations were neutral in the middle third of the
1050 chromosome but can be deleterious in the chromosome ends; (*iii*) *Bateson-Dobzhansky-*
1051 *Muller incompatibility* (BDMI): Neutral and deleterious mutations occurred in the
1052 ancestor. After the split, we allowed migration between the two daughter populations. To
1053 simulate BDM incompatibilities, a fraction of the selected mutations was deleterious in
1054 each of the populations and neutral in the other. (*iv*) *Positive selection*: mutations in the
1055 ancestor and daughter population were neutral in the middle third of the chromosome but
1056 could be beneficial in the remaining regions. (*v*) *Local adaptation:* mutations in the
1057 ancestor and daughter population were neutral in the middle third of the chromosome but
1058 could be beneficial in the remainder. After the split, we allowed migration between the
1059 daughter populations. To simulate local adaptation, a fraction of the selected mutations
1060 was beneficial in each population and neutral in the other. Unlike the BDMI simulation,
1061 we allowed local adaptation from standing variation, so some of the mutations that were
1062 neutral in the ancestral population became beneficial after the split. *BGS and positive*
1063 *selection*: mutations in the ancestor and daughter population were neutral in the middle
1064 third of the chromosome but could be either deleterious or beneficial in the tails.
1065 Each scenario was simulated with varying proportions of beneficial/deleterious to
1066 neutral mutations, and varying mean selective coefficients and migration rates, where
1067 applicable (Table S7). For each combination of parameters, we ran five replicate
1068 simulations. As described in Kelleher et al. (Kelleher et al. 2018), to improve simulation
1069 speed we recorded genealogies and non-neutral mutations in a tree sequence and added
1070 neutral mutations afterwards with *msprime* (Kelleher et al. 2016). We used *scikit-allel*
1071 (v1.1.8) (Miles and Harding 2017) to calculate π, F_{ST}, and d_{xy} in 500 kb regions. All of
1072 the SLiM code used is available on GitHub
1073 (https://github.com/mufernando/mimulus_sims).
1074
1075 *Tests for genome-wide admixture*
1076 We tested for introgression and quantified levels of admixture by calculating
1077 Patterson's *D* and the admixture proportion (Green et al. 2010, Durand et al. 2011).
1078 Patterson's *D* takes four taxa with the relationship (((P1, P2), P3) O) and looks for an
1079 excess of either ABBA or BABA sites, where A and B represent the ancestral and
1080 derived alleles respectively. We calculated Patterson's *D* for all possible groups of three
1081 ingroup taxa based on the set of relationships inferred from the genome-wide
1082 (concatenated) dataset. *M. clevelandii* was used as the outgroup for all tests. The *D*
1083 statistic was calculated from allele frequency data for biallelic sites, with the ancestral
1084 allele called as the allele fixed within *M. clevelandii.* Therefore, only sites fixed within
1085 *M. clevelandii* were included in the analysis. We converted genotype data into allele
1086 frequency data using a Python script download from
1087 https://github.com/simonhmartin/genomics. Patterson's *D* was calculated from the
1088 proportion of ABBA and BABA sites in R using a custom script. To evaluate the
1089 significance of each test, we performed block jack-knifing with a block size of 500 kb.
1090 We then calculated the genome-wide admixture proportion, *f*, for all tests with a
1091 significant value of Patterson's *D*. This test compares the excess of ABBA sites shared
1092 between P2 and P3 to the expected excess of ABBA sites between two completely

admixed populations. The expected excess is calculated by splitting the samples from P3 into two different groups (P3a and P3b) and calculating the excess of ABBA to BABA sites with one group taking the place of P2 and the other the place of P3.

For tests with significant values of Patterson's *D*, we then estimated levels of admixture in each 500 kb genomic window. Although Patterson's *D* is a powerful test for detecting introgression genome-wide, it is not suited for estimating admixture in defined genomic regions (Martin et al. 2015). Thus, we calculated the four-taxon statistic f_d, which is a modified version of the admixture proportion developed for estimating local admixture (Martin et al. 2015). We calculated f_d in 500 kb non-overlapping windows, using Python scripts (https://github.com/simonhmartin/genomics_general). Because f_d is designed to detect an excess of ABBA sites (*i.e.,* gene flow from P3 to P2), we switched the order of P1 and P2 in tests where *D* was negative (excess of BABA) in the genome-wide tests for introgression. Note that this does not affect the set of relationships in these comparisons. We summarized genome-wide variation in f_d across the different tests by taking the average value of f_d within each window and the maximum value of f_d from the 48 different tests in each window, and estimated the Pearson's correlation coefficient between these values other statistics.

Figure captions

Figure 1. Evolutionary relationships and patterns of genome-wide variation across the radiation. A) The black tree was constructed from a concatenated alignment of genome-wide SNPs and is rooted using *M. clevelandii*. The 387 gray trees were constructed from 500 kb genomic windows. The first number associated with each node or taxon is the bootstrap support for that clade in the whole genome tree, and the second number is the percentage of window-based trees in which that clade is present. B) Levels of differentiation (F_{ST}), C) divergence (d_{xy}), and D) diversity (π) within and among taxa based on the same 500 kb windows. For simplicity, F_{ST} and d_{xy} are shown only for comparisons with the red ecotype of subspecies *puniceus*. See Fig. S7 for the distributions of F_{ST} and d_{xy} across all pairs of taxa.

Figure 2. Common genomic landscapes mirror variation in the local properties of the genome. A) Tree concordance scores for 500 kb non-overlapping genomic windows plotted across the 10 bush monkeyflower chromosomes. B – D) Plots of the first principal component (PC1) for F_{ST}, d_{xy}, and π in overlapping 500 kb windows (step size = 50 kb). PC1 explains 66%, 70%, and 85% of the variation in F_{ST}, d_{xy} and π, respectively, and is Z-transformed such that above average values have positive values and below average values have negative values. E – F) Gene count and recombination rate (cM/Mbp) in overlapping 500 kb windows. G), Mean f_d (admixture proportion) in 500 kb non-overlapping genomic windows. See Fig. S4 for the same plot made for 100 kb windows.

Figure 3. Correlations between measures of diversity and intrinsic features reveal the impact of heterogeneous indirect selection. Matrix of pairwise correlations between PC1 F_{ST}, PC1 d_{xy}, PC1 π, tree concordance, mean f_d, gene density, and recombination rate, all estimated in 500 kb non-overlapping windows. The heat map indicates the strength of the correlation and its sign. All correlations are statistically significant at $p <$

0.001. (See Fig. S8 for a more detailed correlation matrix and Fig. S9 for the correlation matrix from 100 kb windows.

Figure 4. A large-effect adaptive locus shows a lineage-specific signature of positive selection. Plots of Z-transformed F_{ST} across the genome, estimated in 500 kb sliding windows (step size 50 kb). The red line shows values between the red and yellow ecotypes of subspecies *puniceus*, while the gray lines show the values of all other comparisons. The dashed blue line shows the first PC calculated across all of the comparisons. The triangle marks the position of the gene *MaMyb2*. A *cis*-regulatory mutation that is tightly linked to this gene is responsible for the shift from yellow to red flowers. See Fig. S10 for the same plot made for 100 kb windows.

Figure 5. The range of divergence times reveals static and dynamic signatures of recurrent indirect selection. Correlations between variables (500 kb windows) for all 36 taxonomic comparisons (gray dots) plotted against the average d_a as a measure of divergence time. The left panels show how the relationships between π (each window averaged across a pair of taxa) and (A) gene count and (B) recombination rate vary with increasing divergence time. The middle panels (C & D) show the same relationships, but with F_{ST}. The right panels show the relationships between (E) d_{xy} and π and (F) F_{ST} and π. The regressions (dashed lines) in each plot are fitted to the eight independent contrasts (colored points) obtained using a phylogenetic correction. The color gradient shows the strength of the correlation.

Figure 6. Emergence of a heterogeneous differentiation landscape across one million years of divergence. A) Plots of F_{ST} (500 kb windows) across the genome for pairs of taxa at early (red vs. yellow), intermediate (red vs. *parviflorus*), and late stages (red vs. *M. clevelandii*) of divergence. B) Average nucleotide diversity (for the red ecotype of subspecies *puniceus*, yellow ecotype of subspecies *puniceus*, subspecies *parviflorus*, and *M. clevelandii*) across the genome in 500 kb windows. The geographic distribution (parapatric or allopatric), sequence divergence ($d_a \times 10^{-2}$), and correlation between F_{ST} and mean π are provided next to each taxon pair.

Figure 7. Genomic landscapes simulated under different divergence histories. Each row of plots shows patterns of within- and between-population variation (π, d_{xy}, and F_{ST}) across the chromosome (500 kb windows) at five time points (N generations, where N = 10000) during one of scenarios The selection parameter (Ns, where s = Ns/N), proportion of deleterious (-) and positive mutations (+), and number of migrants per generation (Nm; 0 unless stated) for these simulations are as follows: (*i*) neutral divergence (no selection), (*ii*) background selection (-Ns = 100; -prop = 0.01), (*iii*) Bateson-Dobzhansky-Muller incompatibilities (BDMI) (-Ns = 100, -prop = 0.05, Nm = 0.1), (*iv*) positive selection (+Ns = 100, +prop = 0.01), (*v*) background selection and positive selection (-Ns = 100, -prop = 0.01; +Ns = 100, +prop = 0.005), and (*vi*) local adaptation (+Ns = 100, +prop = 0.001, Nm = 0.1). The gray boxes in the first column show the areas of the chromosome that are experiencing selection, while the white central area evolves neutrally. Note that π (in populations a and b) and d_{xy} have been mean centered so they can be viewed on the

same scale. Un-centered values and additional simulations with different parameter combinations and more time points can be found in Fig. S13.

List of Supplementary tables

Table S1. Summary of the genetic linkage map constructed using an F_2 intercross between the red and yellow ecotypes of subspecies *puniceu*s.

Table S2. Analysis of gene space completeness in the *M. aurantiacus* genome using CEGMA and BUSCO.

Table S3. Sample information for the 37 sequenced individuals.

Table S4. Loadings for principal component 1 calculated across all 36 pairwise comparisons (for F_{ST} and d_{xy}) or all nine taxa (for π).

Table S5. Details for the linear regressions presented in Figure 5 of the main text.

Table S6. Parameters used to simulate genomic landscape evolution.

Table S7. Genome-wide estimates of admixture across the bush monkeyflower radiation.

Table S8. Correlation coefficients between the window-based admixture proportion (f_d) and PC1 F_{ST}.

List of Supplementary figures

Fig. S1. Map distance (cM/Mbp) vs. physical distance across the 10 linkage groups.

Fig. S2. Geographic distribution of sampling locations for each sample sequenced in this study.

Fig. S3. Genome-wide Principal Components Analysis (PCA).

Fig. S4. Common genomic landscapes mirror variation in the local properties of the genome.

Fig. S5. Patterns of variation plotted across each bush monkeyflower linkage group.

Fig. S6. Patterns of variation are non-randomly distributed across the genome.

Fig. S7. Patterns of differentiation and divergence for all 36 pairs of taxa.

Fig. S8. Bivariate plots among measures of variation and genomic features across 500 kb windows.

Fig. S9. Bivariate plots among measures of variation and genomic features.

Fig. S10. A large-effect adaptive locus shows a lineage-specific signature of positive selection.

Fig. S11. Emergence of a heterogeneous differentiation landscape across one million years of divergence.

Fig. S12. Negative correlation between nucleotide diversity and differentiation becomes stronger with increasing divergence time.

Fig. S13. Genomic landscapes simulated under different divergence histories.

Fig. S14. Evidence for widespread gene flow across the bush monkeyflower radiation.

Fig. S15. Variation in levels of admixture across the genome.

References

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. W. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Most, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig,

1229 A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura,
1230 R. Vainola, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation.
1231 Journal of Evolutionary Biology **26**:229-246.
1232 Amores, A., J. Catchen, I. Nanda, W. Warren, R. Walter, M. Schartl, and J. H.
1233 Postlethwait. 2014. A RAD-tag genetic map for the platyfish (Xiphophorus
1234 maculatus) reveals mechanisms of karyotype evolution among teleost fish.
1235 Genetics **197**:625-641.
1236 Baird, S. J. E. 2015. Exploring linkage disequilibrium. Mol Ecol Resour **15**:1017-1019.
1237 Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major
1238 armor gene in threespine stickleback. Science **322**:255-257.
1239 Barton, N. H. and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annual Review of
1240 Ecology and Systematics **16**:113-148.
1241 Bassham, S., J. Catchen, E. Lescak, F. A. von Hippel, and W. A. Cresko. 2018. Repeated
1242 Selection of Alternatively Adapted Haplotypes Creates Sweeping Genomic
1243 Remodeling in Stickleback. Genetics **209**:921-939.
1244 Berner, D. and W. Salzburger. 2015. The genomics of organismal diversification
1245 illuminated by adaptive radiations. Trends in Genetics **31**:491-499.
1246 Bouckaert, R. R. 2010. DensiTree: making sense of sets of phylogenetic trees.
1247 Bioinformatics **26**:1372-1373.
1248 Brandvain, Y., A. M. Kenney, L. Flagel, G. Coop, and A. L. Sweigart. 2014. Speciation
1249 and Introgression between Mimulus nasutus and Mimulus guttatus. Plos Genetics
1250 **10**.
1251 Browning, S. R. and B. L. Browning. 2007. Rapid and accurate haplotype phasing and
1252 missing-data inference for whole-genome association studies by use of localized
1253 haplotype clustering. American Journal of Human Genetics **81**:1084-1097.
1254 Burri, R. 2017a. Dissecting differentiation landscapes: a linked selection's perspective.
1255 Journal of Evolutionary Biology **30**:1501-1505.
1256 Burri, R. 2017b. Interpreting differentiation landscapes in the light of long-term linked
1257 selection. Evolution Letters **1**:118-131.
1258 Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit,
1259 S. Bures, L. Z. Garamszegi, S. Hogner, J. Moreno, A. Qvarnstrom, M. Ruzic, S.
1260 A. Saether, G. P. Saetre, J. Torok, and H. Ellegren. 2015. Linked selection and
1261 recombination rate variation drive the evolution of the genomic landscape of
1262 differentiation across the speciation continuum of Ficedula flycatchers. Genome
1263 Research **25**:1656-1665.
1264 Campbell, C. R., J. W. Poelstra, and A. D. Yoder. 2018. What is Speciation Genomics?
1265 The roles of ecology, gene flow, and genomic architecture in the formation of
1266 species. Biological Journal of the Linnean Society **124**:561-583.
1267 Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S.
1268 Alvarado, and M. Yandell. 2008. MAKER: An easy-to-use annotation pipeline
1269 designed for emerging model organism genomes. Genome Research **18**:188-196.
1270 Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks:
1271 an analysis tool set for population genomics. Molecular Ecology **22**:3124-3140.
1272 Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011.
1273 Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences.
1274 G3-Genes Genomes Genetics **1**:171-182.

1275 Chang, C. C., C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee.
1276 2015. Second-generation PLINK: rising to the challenge of larger and richer
1277 datasets. Gigascience **4**.
1278 Charlesworth, B. 1998. Measures of divergence between populations and the effect of
1279 forces that reduce variability. Molecular Biology and Evolution **15**:538-543.
1280 Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious
1281 mutations on neutral molecular variation. Genetics **134**:1289-1303.
1282 Chase, M. A., S. Stankowski, and M. A. Streisfeld. 2017. Genomewide variation provides
1283 insight into evolutionary relationships in a monkeyflower species complex
1284 (Mimulus sect. Diplacus). American Journal of Botany **104**:1510-1521.
1285 Coop, G. and P. Ralph. 2012. Patterns of neutral diversity under general models of
1286 selective sweeps. Genetics **192**:205-224.
1287 Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. Natural selection constrains
1288 neutral diversity across a wide range of species. Plos Biology **13**:e1002112.
1289 Coyne, J. A. and H. A. Orr. 1989. Patterns of Speciation in Drosophila. Evolution
1290 **43**:362-381.
1291 Cruickshank, T. E. and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of
1292 speciation are due to reduced diversity, not reduced gene flow. Molecular
1293 Ecology **23**:3133-3157.
1294 Delmore, K. E., J. S. L. Ramos, B. M. Van Doren, M. Lundberg, S. Bensch, D. E. Irwin,
1295 and M. Liedvogel. 2018. Comparative analysis examining patterns of genomic
1296 differentiation across multiple episodes of population divergence in birds.
1297 Evolution Letters **2**:76-87.
1298 Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient
1299 admixture between closely related populations. Molecular Biology and Evolution
1300 **28**:2239-2252.
1301 Durrett, R., L. Buttel, and R. Harrison. 2000. Spatial models for hybrid zones. Heredity
1302 (Edinb) **84 (Pt 1)**:9-19.
1303 Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner,
1304 H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W.
1305 Wolf. 2012. The genomic landscape of species divergence in Ficedula flycatchers.
1306 Nature **491**:756-760.
1307 English, A. C., S. Richards, Y. Han, M. Wang, V. Vee, J. Qu, X. Qin, D. M. Muzny, J. G.
1308 Reid, K. C. Worley, and R. A. Gibbs. 2012. Mind the gap: upgrading genomes
1309 with Pacific Biosciences RS long-read sequencing technology. Plos One
1310 **7**:e47768.
1311 Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow.
1312 Trends in Genetics **28**:342-350.
1313 Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist
1314 **125**:1-15.
1315 Gao, S., D. Bertrand, B. K. Chia, and N. Nagarajan. 2016. OPERA-LG: efficient and
1316 exact scaffolding of large, repeat-rich eukaryotic genomes with performance
1317 guarantees. Genome Biol **17**:102.
1318 Gillespie, J. H. 2000. Genetic drift in an infinite population: The pseudohitchhiking
1319 model. Genetics **155**:909-919.

1320 Gnerre, S., I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T.
1321 Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R.
1322 Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, and D. B.
1323 Jaffe. 2011. High-quality draft assemblies of mammalian genomes from
1324 massively parallel sequence data. Proc Natl Acad Sci U S A **108**:1513-1518.
1325 Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W.
1326 Dirks, U. Hellsten, N. Putnam, and D. S. Rokhsar. 2012. Phytozome: a
1327 comparative platform for green plant genomics. Nucleic Acids Res **40**:D1178-
1328 1186.
1329 Grant, V. 1981. Plant Speciation. 2 edition. Columbia University Press, New York.
1330 Grant, V. 1993a. Effects of hybridization and selection on floral isolation. Proc Natl Acad
1331 Sci U S A **90**:990-993.
1332 Grant, V. 1993b. Origin of floral isolation between ornithophilous and sphingophilous
1333 plant species. Proc Natl Acad Sci U S A **90**:7729-7733.
1334 Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson,
1335 H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D.
1336 Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M.
1337 Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M.
1338 Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J.
1339 Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V.
1340 B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A.
1341 Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C.
1342 Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, and S. Paabo.
1343 2010. A draft sequence of the Neandertal genome. Science **328**:710-722.
1344 Hahn, M. W. 2008. Toward a selection theory of molecular evolution. Evolution **62**:255-
1345 265.
1346 Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph. 2019. Tree-
1347 sequence recording in SLiM opens new horizons for forward-time simulation of
1348 whole genomes. Mol Ecol Resour **19**:552-566.
1349 Haller, B. C. and P. W. Messer. 2019. SLiM 3: Forward Genetic Simulations Beyond the
1350 Wright-Fisher Model. Molecular Biology and Evolution **36**:632-637.
1351 Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster.
1352 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the
1353 genomic landscape of divergence among Darwin's finches. Genome Research
1354 **27**:1004-1015.
1355 Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko.
1356 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback
1357 using Sequenced RAD Tags. Plos Genetics **6**.
1358 Holt, C. and M. Yandell. 2011. MAKER2: an annotation pipeline and genome-database
1359 management tool for second-generation genome projects. BMC Bioinformatics
1360 **12**:491.
1361 Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxford Surveys in
1362 Evolutionary Biology **7**:44.
1363 Hudson, R. R., D. D. Boos, and N. L. Kaplan. 1992. A Statistical Test for Detecting
1364 Geographic Subdivision. Molecular Biology and Evolution **9**:138-151.

1365 Hudson, R. R. and N. L. Kaplan. 1995. Deleterious Background Selection with
1366 Recombination. Genetics **141**:1605-1617.
1367 Jiggins, C. D. and S. H. Martin. 2017. Glittering gold and the quest for Isla de Muerta.
1368 Journal of Evolutionary Biology **30**:1509-1511.
1369 Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, and J. Walichiewicz.
1370 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet
1371 Genome Res **110**:462-467.
1372 Kelleher, J., A. M. Etheridge, and G. McVean. 2016. Efficient Coalescent Simulation and
1373 Genealogical Analysis for Large Sample Sizes. Plos Computational Biology **12**.
1374 Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph. 2018. Efficient pedigree
1375 recording for fast population genetics simulation. Plos Computational Biology **14**.
1376 Kersey, P. J., J. E. Allen, I. Armean, S. Boddu, B. J. Bolt, D. Carvalho-Silva, M.
1377 Christensen, P. Davis, L. J. Falin, C. Grabmueller, J. Humphrey, A. Kerhornou, J.
1378 Khobova, N. K. Aranganathan, N. Langridge, E. Lowy, M. D. McDowall, U.
1379 Maheswari, M. Nuhn, C. K. Ong, B. Overduin, M. Paulini, H. Pedro, E. Perry, G.
1380 Spudich, E. Tapanari, B. Walts, G. Williams, M. Tello-Ruiz, J. Stein, S. Wei, D.
1381 Ware, D. M. Bolser, K. L. Howe, E. Kulesha, D. Lawson, G. Maslen, and D. M.
1382 Staines. 2016. Ensembl Genomes 2016: more genomes, more complexity. Nucleic
1383 Acids Res **44**:D574-580.
1384 Kimura, M. 1968. Evolutionary rate at the molecular level. Nature **217**:624-626.
1385 Koch, M., B. Haubold, and T. Mitchell-Olds. 2001. Molecular systematics of
1386 Brassicaceae: evidence from plastidic matK and nuclear Chs sequences. American
1387 Journal of Botany **88**:534-544.
1388 Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics **5**:59.
1389 Kronforst, M. R., M. E. B. Hansen, N. G. Crawford, J. R. Gallant, W. Zhang, R. J.
1390 Kulathinal, D. D. Kapan, and S. P. Mullen. 2013. Hybridization Reveals the
1391 Evolving Genomic Architecture of Speciation. Cell Reports **5**:666-677.
1392 Lamichhaney, S., J. Berglund, M. S. Almen, K. Maqbool, M. Grabherr, A. Martinez-
1393 Barrio, M. Promerova, C. J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R.
1394 Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and
1395 their beaks revealed by genome sequencing. Nature **518**:371-375.
1396 Langley, C. H., K. Stevens, C. Cardeno, Y. C. Lee, D. R. Schrider, J. E. Pool, S. A.
1397 Langley, C. Suarez, R. B. Corbett-Detig, B. Kolaczkowski, S. Fang, P. M. Nista,
1398 A. K. Holloway, A. D. Kern, C. N. Dewey, Y. S. Song, M. W. Hahn, and D. J.
1399 Begun. 2012. Genomic variation in natural populations of Drosophila
1400 melanogaster. Genetics **192**:533-598.
1401 Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat
1402 Methods **9**:357-359.
1403 Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
1404 MEM. arXiv **1303.3997v1**.
1405 Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A.
1406 Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of
1407 speciation separate cichlid ecomorphs in an East African crater lake. Science
1408 **350**:1493-1498.
1409 Marques, D. A., J. I. Meier, and O. Seehausen. 2019. A Combinatorial View on
1410 Speciation and Adaptive Radiation. Trends in Ecology & Evolution.

1411 Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson,
1412 M. Blaxter, A. Manica, J. Mallet, and C. D. Jiggins. 2013. Genome-wide evidence
1413 for speciation with gene flow in Heliconius butterflies. Genome Research
1414 **23**:1817-1828.
1415 Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the use of ABBA-BABA
1416 statistics to locate introgressed loci. Molecular Biology and Evolution **32**:244-
1417 257.
1418 Martin, S. H., J. W. Davey, C. Salazar, and C. D. Jiggins. 2019. Recombination rate
1419 variation shapes barriers to introgression across butterfly genomes. Plos Biology
1420 **17**.
1421 Matthey-Doret, R. and M. C. Whitlock. 2018. Background selection and the statistics of
1422 population differentiation: consequences for detecting local adaptation. BiorXiv.
1423 Maynard-Smith, J. and J. Haigh. 1974. Hitch-Hiking Effect of a Favorable Gene.
1424 Genetics Research **23**:23-35.
1425 McMinn, H. E. 1951. Studies in the genus *Diplacus*, Scrophulariaceae. Madrono **11**:33-
1426 128.
1427 Miles, A. and N. Harding. 2017. cggh/scikit-allel: v1.1.8.
1428 Nei, M. and W. H. Li. 1979. Mathematical-Model for Studying Genetic-Variation in
1429 Terms of Restriction Endonucleases. Proc Natl Acad Sci U S A **76**:5269-5273.
1430 Nelson, T. C. and W. A. Cresko. 2018. Ancient genomic variation underlies repeated
1431 ecological adaptation in young stickleback populations. Evolution Letters **2**:9-21.
1432 Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I.
1433 V. Grigoriev, and I. Dubchak. 2013. The genome portal of the Department of
1434 Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res **42**:D26-D31.
1435 Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and
1436 heterogeneous genomic divergence. Molecular Ecology **18**:375-402.
1437 Ohta, T. 1973. Slightly Deleterious Mutant Substitutions in Evolution. Nature **246**:96-98.
1438 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and
1439 Evolution in R language. Bioinformatics **20**:289-290.
1440 Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core
1441 genes in eukaryotic genornes. Bioinformatics **23**:1061-1067.
1442 Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics Reveals
1443 Three Sources of Adaptive Variation during a Rapid Radiation. Plos Biology **14**.
1444 Pease, J. B. and M. W. Hahn. 2013. More Accurate Phylogenies Inferred from Low-
1445 Recombination Regions in the Presence of Incomplete Lineage Sorting. Evolution
1446 **67**:2376-2384.
1447 Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Muller, V. Baglione, P.
1448 Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic
1449 landscape underlying phenotypic integrity in the face of gene flow in crows.
1450 Science **344**:1410-1414.
1451 Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez.
1452 2005. InterProScan: protein domains identifier. Nucleic Acids Res **33**:W116-
1453 W120.
1454 Rastas, P., F. C. F. Calboli, B. C. Guo, T. Shikano, and J. Merila. 2016. Construction of
1455 Ultradense Linkage Maps with Lep-MAP2: Stickleback F-2 Recombinant Crosses
1456 as an Example. Genome Biol Evol **8**:78-93.

1457 Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlovic, M. A. F. Noor,
1458 B. Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of
1459 speciation: a road map for finding barriers to gene flow. Journal of Evolutionary
1460 Biology **30**:1450-1477.
1461 Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, J.
1462 M. Burke, and L. H. Rieseberg. 2013. Genomic islands of divergence are not
1463 affected by geography of speciation in sunflowers. Nature Communications **4**.
1464 Rettelbach, A., A. Nater, and H. Ellegren. 2019. How Linked Selection Shapes the
1465 Diversity Landscape in Ficedula Flycatchers. Genetics **212**:277-285.
1466 Rockman, M. V. 2012. The Qtn Program and the Alleles That Matter for Evolution: All
1467 That's Gold Does Not Glitter. Evolution **66**:1-17.
1468 Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford University Press, Oxford.
1469 Schumer, M., C. L. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S.
1470 Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2018. Natural
1471 selection interacts with recombination to shape the evolution of hybrid genomes.
1472 Science **360**:656-659.
1473 Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov.
1474 2015. BUSCO: assessing genome assembly and annotation completeness with
1475 single-copy orthologs. Bioinformatics **31**:3210-3212.
1476 Slatkin, M. 1991. Inbreeding Coefficients and Coalescence Times. Genetical Research
1477 **58**:167-175.
1478 Sobel, J. M., S. Stankowski, and M. A. Streisfeld. 2019. Variation in ecophysiological
1479 traits might contribute to ecogeographic isolation and divergence between
1480 parapatric ecotypes of Mimulus aurantiacus. J Evol Biol.
1481 Sobel, J. M. and M. A. Streisfeld. 2015. Strong premating reproductive isolation drives
1482 incipient speciation in Mimulus aurantiacus. Evolution **69**:447-461.
1483 Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.
1484 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J.
1485 Crespi, and P. Nosil. 2014. Stick Insect Genomes Reveal Natural Selection's Role
1486 in Parallel Speciation. Science **344**:738-742.
1487 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
1488 analysis of large phylogenies. Bioinformatics **30**:1312-1313.
1489 Stanke, M. and S. Waack. 2003. Gene prediction with a hidden Markov model and a new
1490 intron submodel. Bioinformatics **19**:Ii215-Ii225.
1491 Stankowski, S., J. M. Sobel, and M. A. Streisfeld. 2015. The geography of divergence
1492 with gene flow facilitates multitrait adaptation and the evolution of pollinator
1493 isolation in Mimulus aurantiacus. Evolution **69**:3054-3068.
1494 Stankowski, S., J. M. Sobel, and M. A. Streisfeld. 2017. Geographic cline analysis as a
1495 tool for studying genome-wide variation: a case study of pollinator-mediated
1496 divergence in a monkeyflower. Molecular Ecology **26**:107-122.
1497 Stankowski, S. and M. A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive
1498 divergence in a recent radiation of monkeyflowers. Proceedings of the Royal
1499 Society B-Biological Sciences **282**:154-162.
1500 Streisfeld, M. A. and J. R. Kohn. 2005. Contrasting patterns of floral and molecular
1501 variation across a cline in Mimulus aurantiacus. Evolution **59**:2548-2559.

1502 Streisfeld, M. A. and J. R. Kohn. 2007. Environment and pollinator-mediated selection
1503 on parapatric floral races of Mimulus aurantiacus. Journal of Evolutionary
1504 Biology **20**:122-132.
1505 Streisfeld, M. A., W. N. Young, and J. M. Sobel. 2013. Divergent Selection Drives
1506 Genetic Differentiation in an R2R3-MYB Transcription Factor That Contributes
1507 to Incipient Speciation in Mimulus aurantiacus. Plos Genetics **9**.
1508 Thompson, D. M. 2005. Systematics of Mimulus subgenus Schizoplacus
1509 (Scrophulariaceae). Systematic Botany Monographs **75**:1-213.
1510 Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in
1511 Anopheles gambiae. Plos Biology **3**:1572-1578.
1512 Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel.
1513 2017. Correlated patterns of genetic diversity and differentiation across an avian
1514 family. Molecular Ecology **26**:3982-3997.
1515 Vickery, R. K. 1995. Speciation by Aneuploidy and Polyploidy in Mimulus
1516 (Scrophulariaceae). Great Basin Naturalist **55**:174-176.
1517 Wolf, J. B. and H. Ellegren. 2017. Making sense of genomic islands of differentiation in
1518 light of speciation. Nature Reviews Genetics **18**:87-100.
1519 Wu, C. I. 2001. The genic view of the process of speciation. Journal of Evolutionary
1520 Biology **14**:851-865.
1521 Zerbino, D. R. and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly
1522 using de Bruijn graphs. Genome Research **18**:821-829.
1523
1524

Supplementary information for:

Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers

Sean Stankowski[1,2,§], Madeline A. Chase[1,3,§], Allison M. Fuiten[1], Murillo F. Rodrigues[1], Peter L. Ralph[1], and Matthew A. Streisfeld[1*].

[1]Institute of Ecology and Evolution, 335 Pacific Hall 5289, University of Oregon, Eugene, OR 97405, USA
[2]Present address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
[3]Present address: Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden
[§]These authors contributed equally to this work

*Corresponding author: mstreis@uoregon.edu

ORCID ids:
SS: https://orcid.org/0000-0003-0472-9299; MAC: https://orcid.org/0000-0002-7916-3560; AMF: https://orcid.org/0000-0003-3926-5455; MFR: https://orcid.org/0000-0001-7508-1384 PLR: https://orcid.org/0000-0002-9459-6866; MAS: https://orcid.org/0000-0002-2660-8642

Supplementary tables

Table S1. Summary of the genetic linkage map constructed using an F2 intercross between the red and yellow ecotypes of subspecies *puniceu*s. The table includes map length in cM for each linkage group (LG), the number of markers associated with each LG, the number of unique map positions, and the average genetic distance in cM between each unique map position. Standard deviations are given in parentheses.

LG	Map length (cM)	Number of markers	Unique map positions	Avg. genetic dist. between unique markers in cM (sd)
1	93.9	969	335	0.28 (0.33)
2	71.37	893	253	0.28 (0.88)
3	76.3	912	256	0.30 (0.65)
4	70.2	851	257	0.28 (0.41)
5	78.7	778	295	0.27 (0.29)
6	69.1	741	247	0.28 (0.88)
7	59.8	738	234	0.26 (0.33)
8	65.6	674	246	0.27 (0.41)
9	68.6	623	182	0.37 (0.58)
10	71.1	410	150	0.48 (0.96)
Avg.	72.74 (8.69)	758.90 (155.75)	245.50 (49.04)	0.31 (0.06)

Table S2. Analysis of gene space completeness in the *M. aurantiacus* genome using CEGMA and BUSCO. The number and percent of core genes found in the final assembly are shown for each analysis (CEGMA, *n* = 248; BUSCO, *n* = 1440).

Analysis	# Genes	% Found in Assembly
CEGMA Complete	233	93.95
CEGMA Partial	244	98.39
BUSCO total complete (duplicated)	1340 (61)	93 (4.2)
BUSCO Fragmented	29	2.0
BUSCO Missing	71	5.0

Table S3. Sample information for the 37 sequenced individuals. Includes their taxon identity, sampling location, percent read alignment, and average sequencing depth.

Sample	Taxon	Latitude	Longitude	% Reads aligned	Seq. Depth
159_83	ssp. *aridus*	32.6630	-116.2230	91.7	21.12
159_84	ssp. *aridus*	32.6630	-116.2230	89.3	21.98
195_1	ssp. *aridus*	32.6300	-116.1429	92.6	20.20
T84	ssp. *aridus*	32.6526	-116.2449	87.2	21.75
T102	ssp. *aurantiacus*	39.0424	-122.7727	94.9	23.74
T104	ssp. *aurantiacus*	39.2045	-123.7646	94.6	25.09
T50	ssp. *aurantiacus*	35.9865	-121.4928	88.3	24.36
T92	ssp. *aurantiacus*	37.8459	-120.6110	94.0	15.16
T144	ssp. *calycinus*	34.1929	-117.2784	93.2	26.00
T150	ssp. *calycinus*	33.8564	-116.8481	94.7	24.02
T90	ssp. *calycinus*	35.5918	-118.5052	91.3	19.97
T91	ssp. *calycinus*	35.3172	-118.5871	95.5	27.91
T101	ssp. *grandiflorus*	39.5536	-121.4301	92.0	16.05
T61	ssp. *grandiflorus*	39.5590	-120.8243	91.6	17.31
T96	ssp. *grandiflorus*	39.0122	-120.7552	92.0	28.21
T99	ssp. *grandiflorus*	39.4376	-121.0599	91.4	23.84
DPR	ssp. *longiflorus*	33.7459	-117.4485	96.0	26.88
SS	ssp. *longiflorus*	34.2722	-118.6100	94.2	30.86
T33	ssp. *longiflorus*	34.3438	-118.5099	94.6	18.87
T8	ssp. *longiflorus*	34.1347	-118.6452	82.6	25.11
KK168	ssp. *parviflorus*	34.0180	-119.6730	91.8	23.66
KK161	ssp. *parviflorus*	34.0180	-119.6730	92.0	19.11
KK180	ssp. *parviflorus*	34.0180	-119.6730	92.4	18.18
KK182	ssp. *parviflorus*	34.0193	-119.6802	91.3	19.46
ELF	ssp. *puniceus*, red	33.0860	-117.1453	93.0	18.20
JMC	ssp. *puniceus*, red	32.7373	-116.9541	93.8	19.06
LH	ssp. *puniceus*, red	33.0609	-117.1188	87.1	19.77
MT	ssp. *puniceus*, red	32.8210	-117.0618	93.7	20.85
UCSD	ssp. *puniceus*, red	32.8894	-117.2362	87.0	18.23
BCRD	ssp. *puniceus*, yellow	32.9496	-116.6380	94.6	20.85
INJ	ssp. *puniceus*, yellow	33.0979	-116.6643	93.1	18.83
LO	ssp. *puniceus*, yellow	32.6767	-116.3312	93.4	18.04
PCT	ssp. *puniceus*, yellow	32.7326	-116.4698	92.3	19.68
POTR	ssp. *puniceus*, yellow	32.6038	-116.6339	90.5	19.27
CLV_GH	*M. clevelandii*	33.1589	-116.8122	92.3	21.31
CLV_11	*M. clevelandii*	33.3391	-116.9325	84.4	15.52
CLV_4	*M. clevelandii*	33.3391	-116.9325	89.3	17.31

Table S4. Loadings for principal component 1 calculated across all 36 pairwise comparisons (for F_{ST} and d_{xy}) or all nine taxa (for π).

Comparison	F_{ST}	d_{xy} PC1
Aur_Ari	0.85707	0.95575
Gra_Ari	0.89606	0.86182
Cal_Ari	0.88738	0.95423
CLV_Ari	0.8969	0.39278
Par_Ari	0.85342	0.94563
Lon_Ari	0.9046	0.96057
Aur_Gra	0.81191	0.85563
Aur_Cal	0.32407	0.90376
Aur_CLV	0.90725	0.46698
Aur_Par	0.79434	0.94778
Lon_Aur	0.40962	0.90381
Ari_Y	0.90097	0.94262
Aur_Y	0.451	0.90512
Gra_Y	0.91203	0.90305
Cal_Y	0.49255	0.88634
CLV_Y	0.90576	0.54264
Par_Y	0.86511	0.94318
Lon_Y	0.59046	0.88521
Cal_Gra	0.89043	0.88319
CLV_Gra	0.91181	0.32045
Par_Gra	0.89579	0.86769
Lon_Gra	0.88549	0.89344
Cal_Par	0.83543	0.94124
Lon_Cal	0.3786	0.87337
Cal_CLV	0.90548	0.52607
Par_CLV	0.91991	0.45233
Lon_CLV	0.90738	0.51482
Lon_Par	0.84197	0.94507
R_Ari	0.90783	0.94574
R_Aur	0.51788	0.89861
R_Y	0.51423	0.84798
R_Gra	0.91936	0.90695
R_Cal	0.62051	0.88316
R_CLV	0.90758	0.55149
R_Par	0.87615	0.93825
R_Lon	0.66912	0.88264

Table S4 continued.

Taxon	π PC1
Ari	0.9251
Aur	0.86508
Cal	0.96734
CLV	0.92017
Gra	0.77181
Lon	0.95166
Par	0.91044
R	0.97282
Y	0.97255

Table S5. Details for the linear regressions presented in Figure 5 of the main text.

Variables	Pearson's r	Regression equation	*p*
d_{xy} & π	-0.93	y = -60.5*x* + 1.33	< 0.001
π & gene count	0.59	y = 3.9*x* - 0.85	0.130
π & cM/mbp	-0.79	y = -5.4*x* + 0.46	0.020
F_{ST} & π	-0.89	y = -53.3*x* - 0.13	0.003
F_{ST} & gene count	0.81	y = 46.5*x* + 0.04	0.016
F_{ST} & cM/mbp	-0.73	y = -25.5*x* + 0.01	0.041

Table S6. Parameters used in SLiM to simulate genomic landscape evolution. (*N*=10,000).

Class	Mean Ns of deleterious mutations	Percent of deleterious mutations	Mean Ns of beneficial mutations	Percent of beneficial mutations	Migration rate (Nm)
Neutral	-	-	-	-	-
BGS	-10	5%	-	-	-
	-10	10%	-	-	-
	-100	5%	-	-	-
	-100	10%	-	-	-
BDMI	-10	5%	-	-	0.1
	-10	10%	-	-	1
	-10	5%	-	-	0.1
	-10	10%	-	-	1
	-100	5%	-	-	0.1
	-100	10%	-	-	1
	-100	5%	-	-	0.1
	-100	10%	-	-	1
Positive	-	-	100	0.1%	-
	-	-	100	0.5%	-
Local adaptation	-	-	100	0.1%	0.1
	-	-	100	0.1%	1
	-	-	100	0.5%	0.1
	-	-	100	0.5%	1
BGS & positive	-10	5%	100	0.1%	-
	-10	5%	100	0.5%	-
	-10	10%	100	0.1%	-
	-10	10%	100	0.5%	-
	-100	5%	100	0.1%	-
	-100	5%	100	0.5%	-
	-100	10%	100	0.1%	-
	-100	10%	100	0.5%	-

Table S7. Genome-wide estimates of admixture across the bush monkeyflower radiation. Measures of Patterson's *D* and the admixture proportion, *f*, are given for all 48 possible pairs of four-taxon tests, with *M. clevelandii* as the outgroup for each test. All values of Patterson's *D* are statistically significant (P < 0.0001) based on a block jackknife approach.

P1	P2	P3	Patterson's D	Admixture proportion
Aur	Cal	Ari	0.0808	0.0209
Lon	Cal	Ari	0.0233	0.0055
Aur	Lon	Ari	0.0633	0.0155
Aur	R	Ari	0.1360	0.0389
Cal	R	Ari	0.0651	0.0184
Lon	R	Ari	0.0883	0.0238
Aur	Y	Ari	0.1595	0.0474
Cal	Y	Ari	0.0949	0.0270
Lon	Y	Ari	0.1177	0.0324
R	Y	Ari	0.0371	0.0088
R	Cal	Aur	0.0831	0.0903
Y	Cal	Aur	0.0976	0.1034
R	Lon	Aur	0.0866	0.0909
Y	Lon	Aur	0.1009	0.1041
Y	R	Aur	0.0188	0.0145
Ari	Par	Gra	0.1053	0.0243
Ari	Aur	Gra	0.1821	0.0477
Cal	Aur	Gra	0.0648	0.0096
Lon	Aur	Gra	0.0936	0.0141
Y	Aur	Gra	0.1392	0.0218
R	Aur	Gra	0.1471	0.0222
Par	Aur	Gra	0.1072	0.0240
Ari	Cal	Gra	0.1523	0.0384
Lon	Cal	Gra	0.0344	0.0045
Y	Cal	Gra	0.0886	0.0123
R	Cal	Gra	0.0943	0.0126
Par	Cal	Gra	0.0665	0.0145
Ari	Lon	Gra	0.1371	0.0341
Par	Lon	Gra	0.0478	0.0101
Y	Lon	Gra	0.0589	0.0078
R	Lon	Gra	0.0650	0.0082
Ari	R	Gra	0.1072	0.0261
Par	R	Gra	0.0099	0.0019
Ari	Y	Gra	0.1103	0.0265

Par	Y	Gra	0.0127	0.0023
R	Y	Gra	0.0059	0.0004
R	Y	Cal	0.0149	0.0438
Aur	Cal	Par	0.0415	0.0182
R	Cal	Par	0.0104	0.0055
Y	Cal	Par	0.0176	0.0077
Aur	Lon	Par	0.0456	0.0207
Cal	Lon	Par	0.0049	0.0025
R	Lon	Par	0.0151	0.0080
Y	Lon	Par	0.0222	0.0102
Aur	R	Par	0.0311	0.0128
Y	R	Par	0.0090	0.0022
Aur	Y	Par	0.0240	0.0106
R	Y	Lon	0.0001	-0.0018

Table S8. Correlation coefficients between the window-based admixture proportion (f_d) and PC1 F_{ST}. Pearson's r was calculated between PC1 F_{ST} and f_d for each of the 48 four-taxon tests, measured in 500kb non-overlapping windows. *M. clevelandii* is the outgroup for each test.

P1	P2	P3	Pearson's r	*p*
Aur	Y	Par	-0.0682	0.1809
R	Lon	Aur	-0.1645	0.0012
Y	Cal	Aur	-0.1388	0.0062
Ari	Cal	Gra	-0.3383	8.19E-12
Y	Aur	Gra	-0.3126	3.23E-10
Cal	R	Ari	-0.4135	2.06E-17
Cal	Lon	Par	-0.1024	0.0441
R	Y	Gra	-0.2732	4.75E-08
Cal	Aur	Gra	-0.2313	4.28E-06
Y	Cal	Gra	-0.2318	4.08E-06
Y	Lon	Par	-0.2484	7.51E-07
Aur	Y	Ari	-0.5850	6.54E-37
R	Lon	Gra	-0.2797	2.18E-08
Y	Lon	Aur	-0.1870	0.0002
Par	Aur	Gra	-0.3334	1.67E-11
Ari	Aur	Gra	-0.3344	1.45E-11
Y	R	Par	-0.2681	8.58E-08
R	Y	Cal	-0.0093	0.8546
R	Aur	Gra	-0.3389	7.40E-12
Ari	Y	Gra	-0.2782	2.61E-08
Ari	Lon	Gra	-0.3538	7.45E-13
Aur	Lon	Ari	-0.3056	8.24E-10
Aur	R	Par	-0.0897	0.0779
Par	Y	Gra	-0.2413	1.56E-06
Aur	Cal	Ari	-0.3456	2.68E-12
R	Lon	Par	-0.2026	5.96E-05
Lon	Cal	Ari	-0.2312	4.29E-06
Y	Lon	Gra	-0.2408	1.65E-06
Par	Cal	Gra	-0.3394	6.91E-12
R	Y	Lon	0.0104	0.8381
Lon	Y	Ari	-0.4775	1.94E-23
Ari	Par	Gra	-0.2272	6.35E-06
Lon	R	Ari	-0.4144	1.71E-17
Aur	R	Ari	-0.5208	2.77E-28
Aur	Lon	Par	-0.1015	0.0460

Ari	R	Gra	-0.2636	1.42E-07
Lon	Aur	Gra	-0.1697	0.0008
Cal	Y	Ari	-0.4707	9.75E-23
Par	Lon	Gra	-0.3104	4.33E-10
Lon	Cal	Gra	-0.1296	0.0107
R	Y	Ari	-0.3998	2.75E-16
Aur	Cal	Par	-0.1359	0.0074
R	Cal	Aur	-0.0881	0.0836
R	Cal	Gra	-0.2463	9.31E-07
Y	R	Aur	-0.1112	0.0287
R	Cal	Par	-0.1981	8.75E-05
Par	R	Gra	-0.2533	4.45E-07
Y	Cal	Par	-0.2735	4.58E-08

Supplementary figures

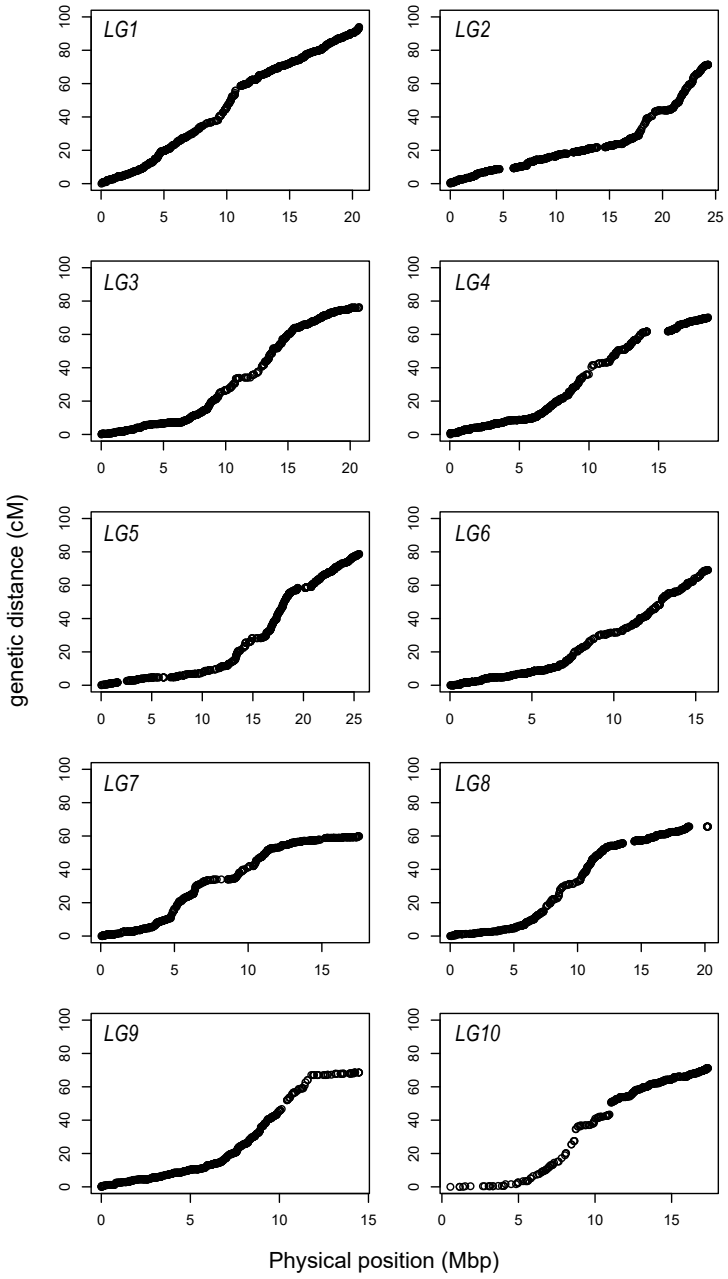

Figure. S1. Map distance (cM/Mbp) vs. physical distance across the 10 linkage groups. Recombination for each marker is estimated relative to the start of the linkage group and plotted at its physical location on each chromosome in the reference assembly.

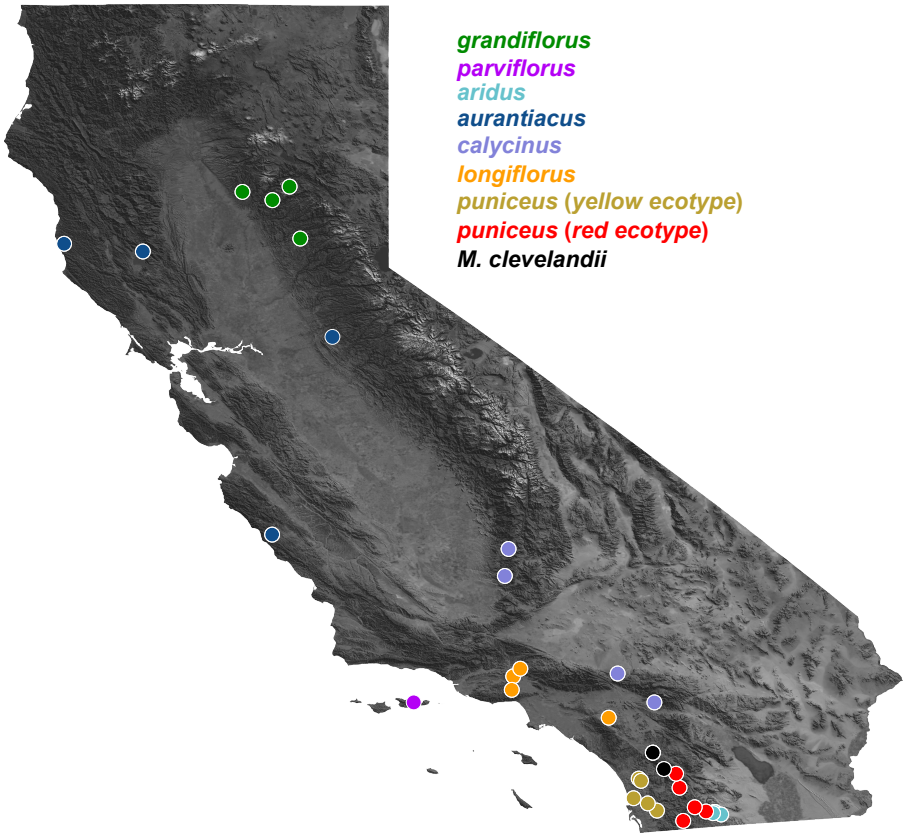

Figure S2. Geographic distribution of sampling locations for each sample sequenced in this study. Detailed position information for each population can be found in Table S3.

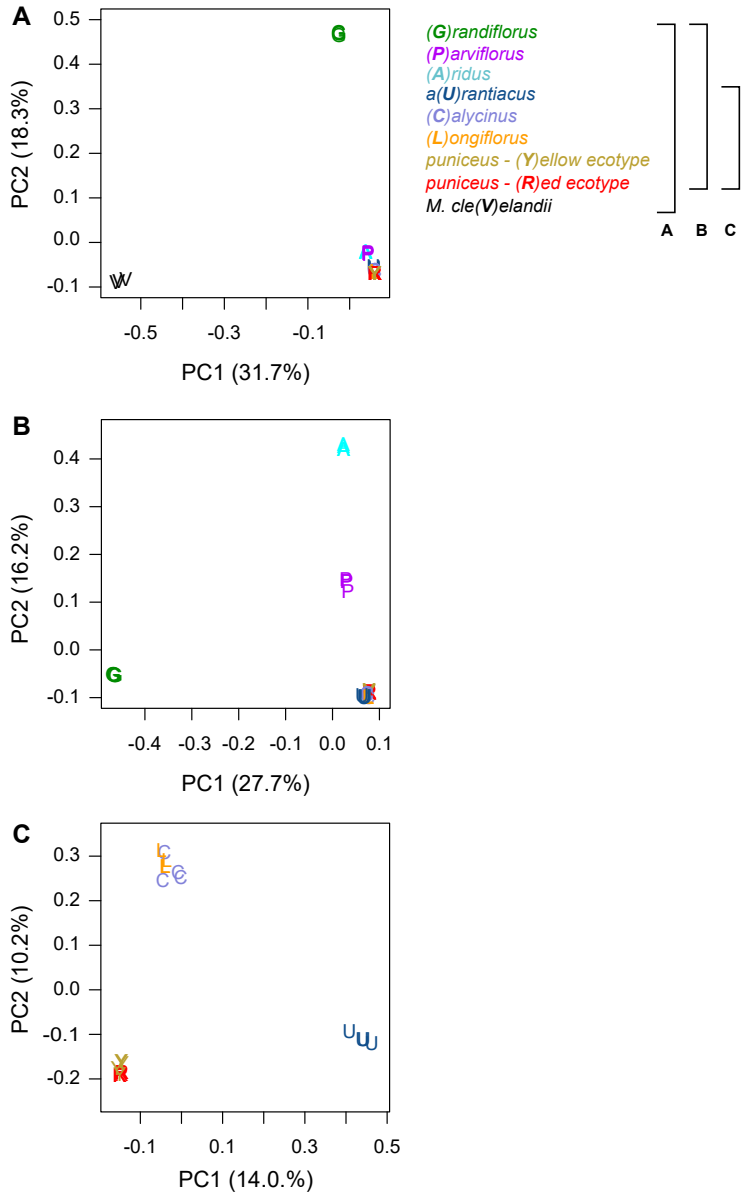

Figure S3. Genome-wide Principal Components Analysis (PCA). Each plot is a separate PCA performed using different sets of taxa. The legend to the right describes the set of taxa included in each analysis, with the capital letter in parentheses and the color representing the specific taxon. A) All taxa; B) all subspecies of *M. aurantiacus*, but excluding *M. clevelandii*; C) only subspecies *aurantiacus*, *longiflorus*, *calycinus*, and the red and yellow ecotypes of subspecies *puniceus*. The percent variation explained by each principal component is given in parentheses.

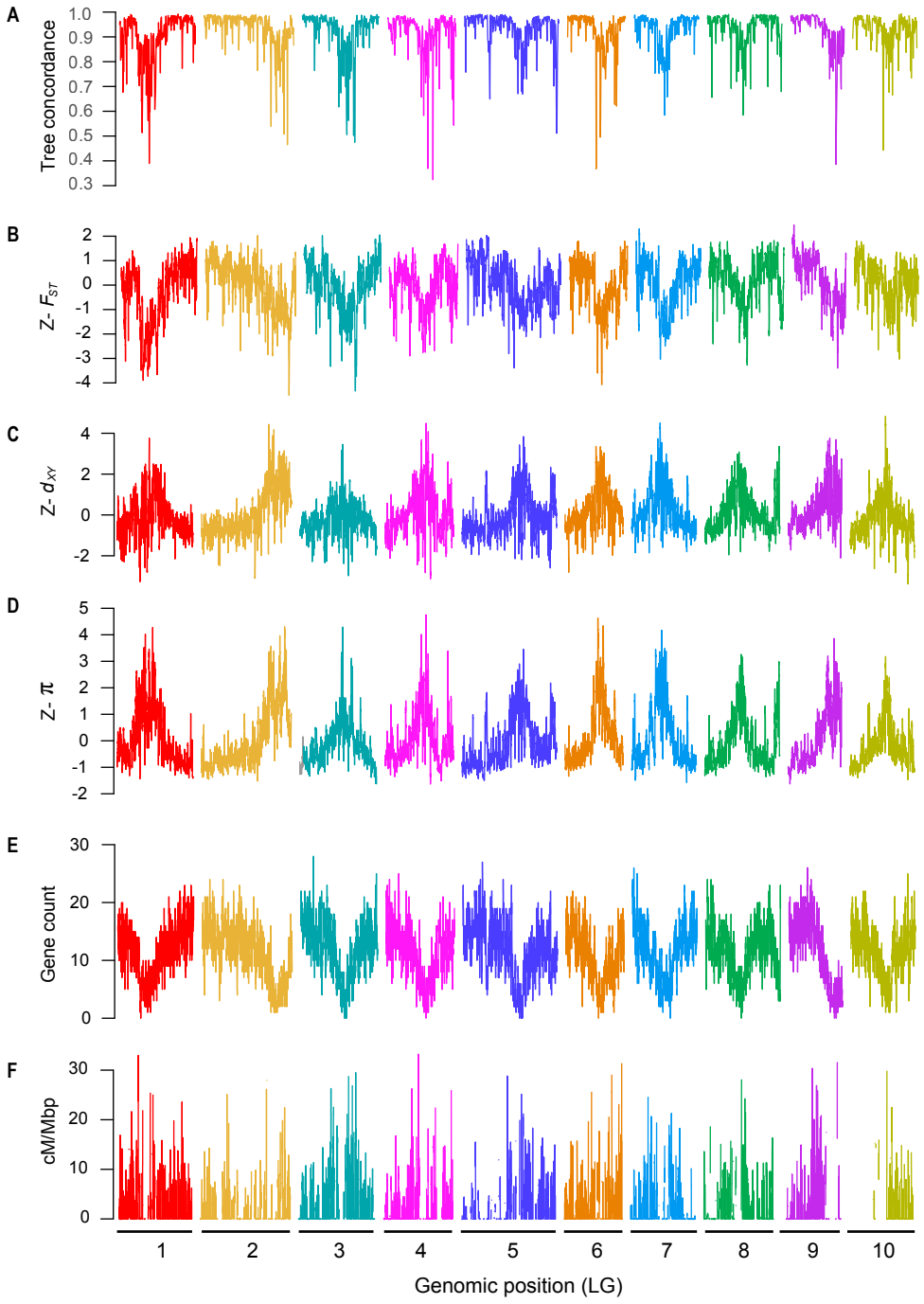

Figure S4. Common patterns of genome-wide variation mirror variation in the local properties of the genome. Plots are the same as in Fig. 2 of the main text, but for 100 kb windows (step size 10 kbp). A) Tree concordance; B-D) Z-transformed PC1 for F_{ST} , d_{xy} and π, respectively; E) gene count; and F) recombination rate (cM/Mbp) are plotted across the 10 monkeyflower LGs.

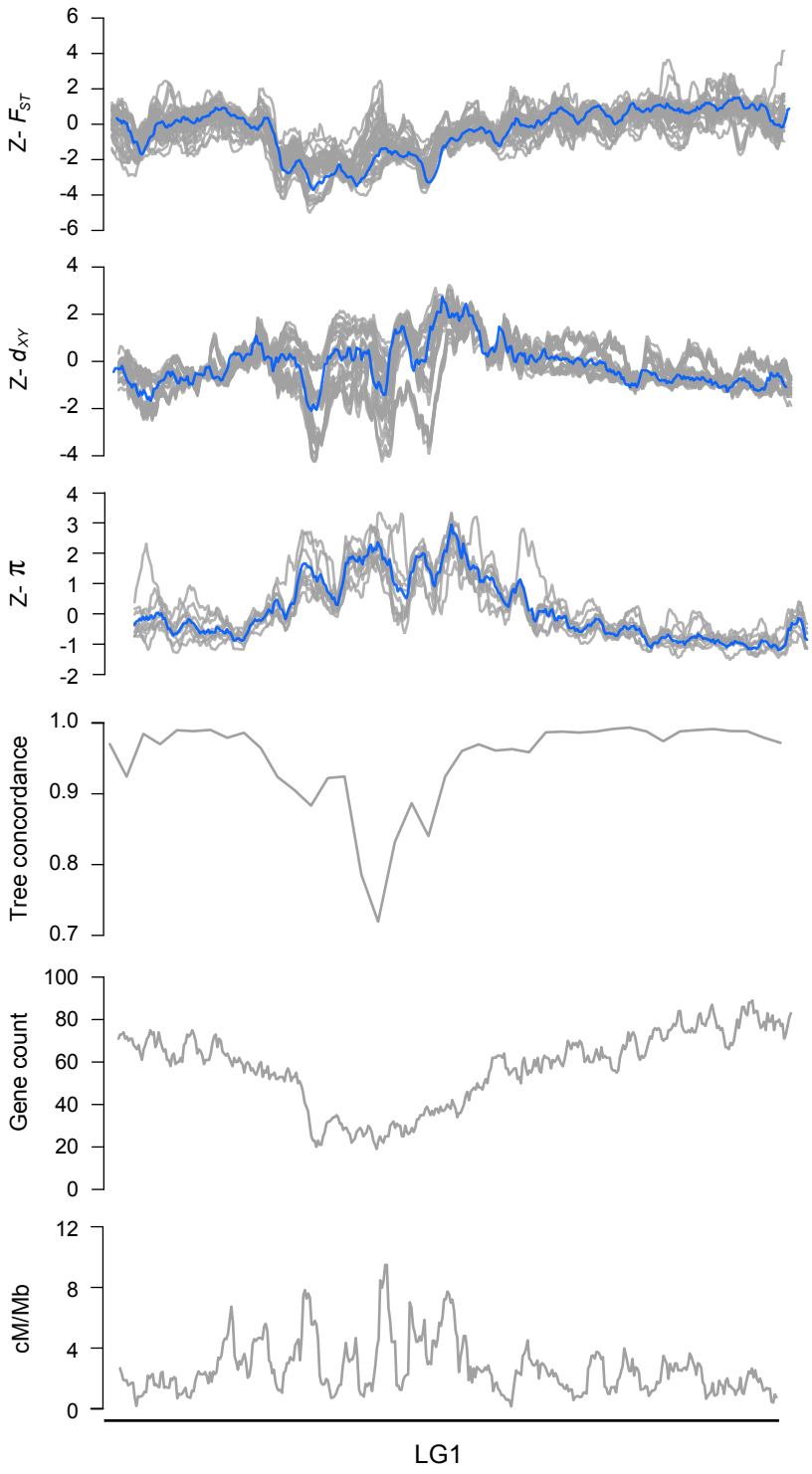

Figure S5. Patterns of variation plotted across each bush monkeyflower linkage group. Z-transformed F_{ST}, d_{xy}, and π in overlapping 500 kb windows (step size = 50 kbp). The gray lines are z-transformed scores for each of the 36 pairwise comparisons (F_{ST} and d_{xy}) or nine taxa (π), and the blue line is the z-transformed score for the first principal component (PC1). Estimates of tree concordance, gene count and recombination rate (cM/Mbp) are also shown.

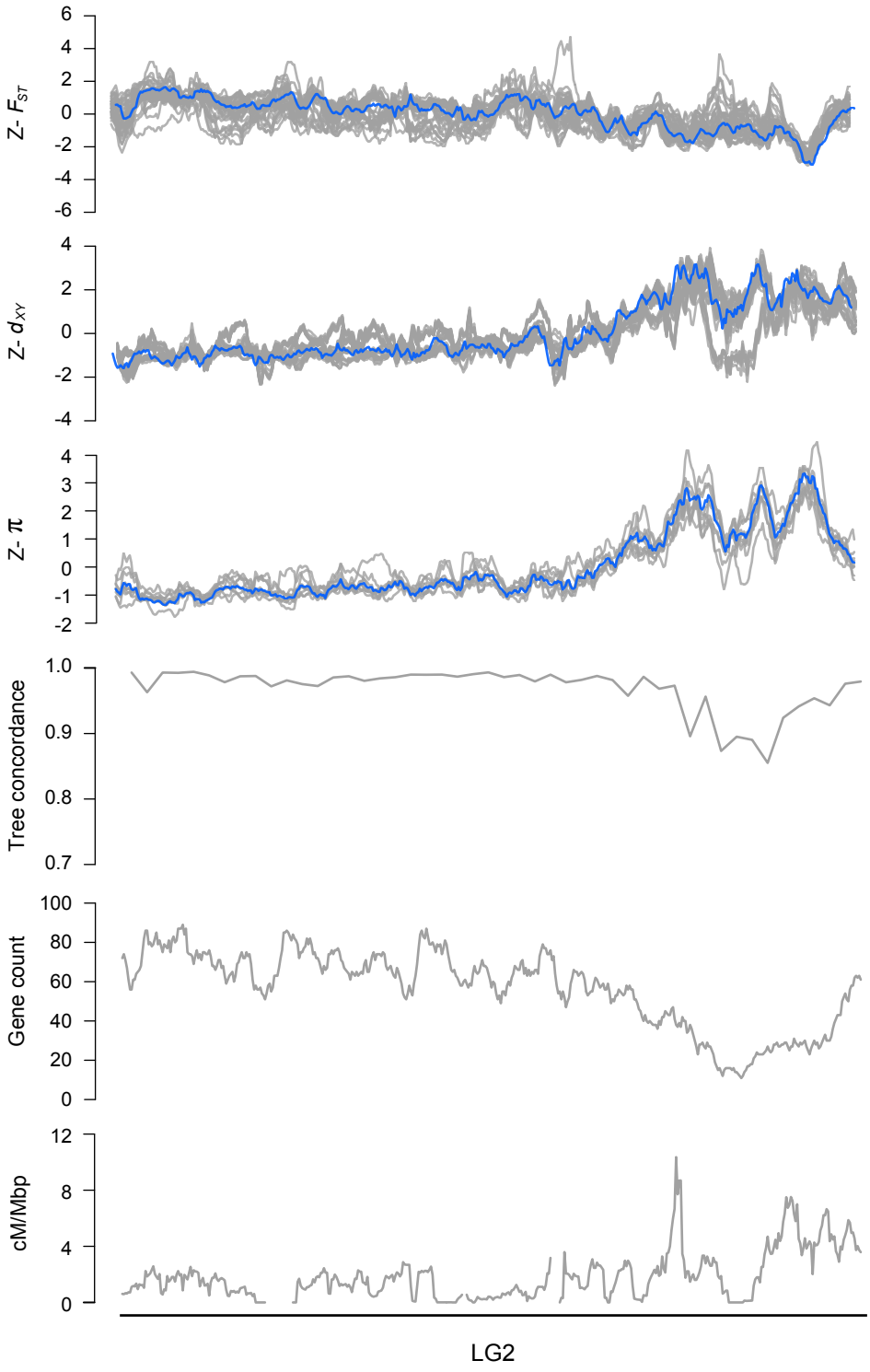

Figure S5. continued

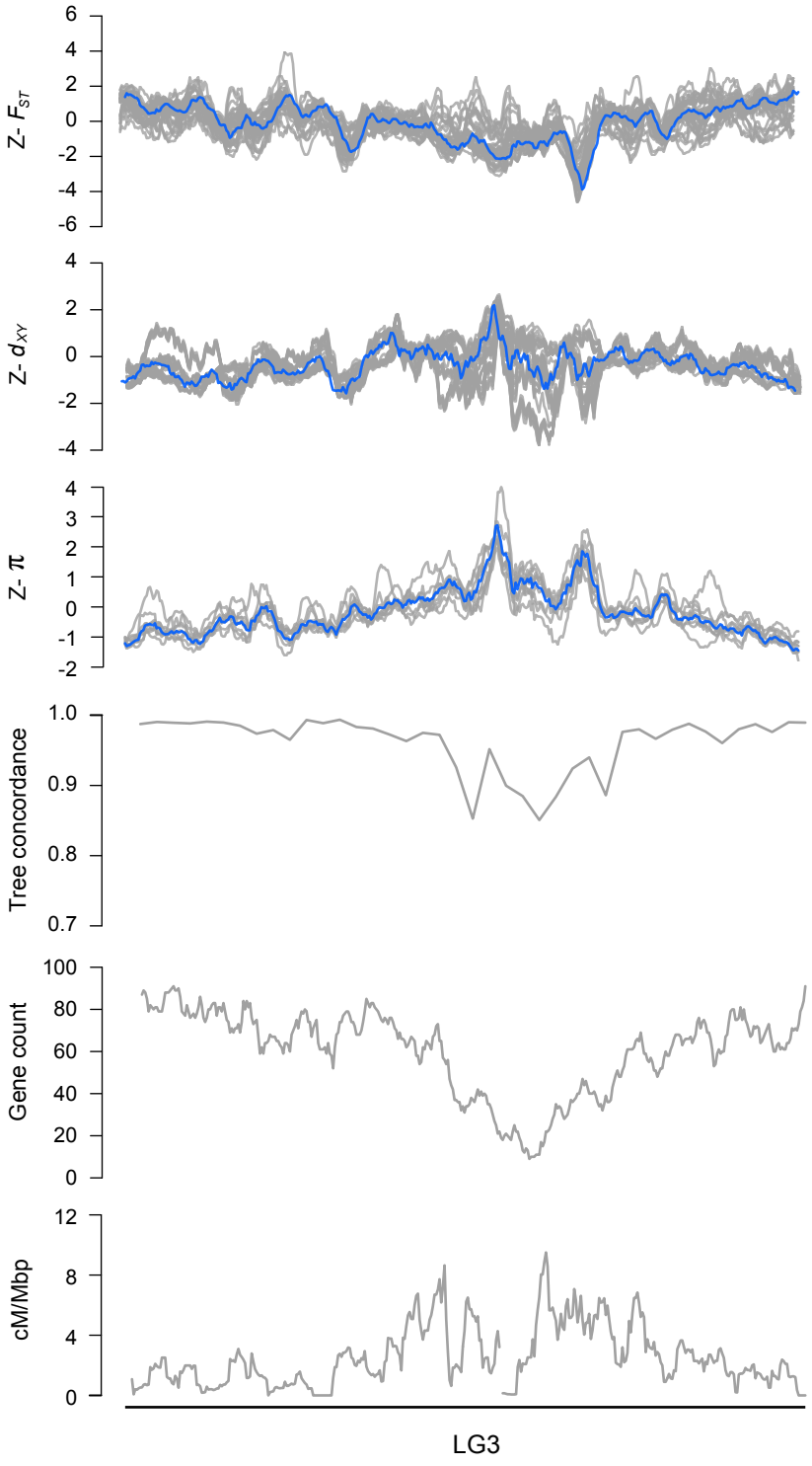

Figure S5. continued

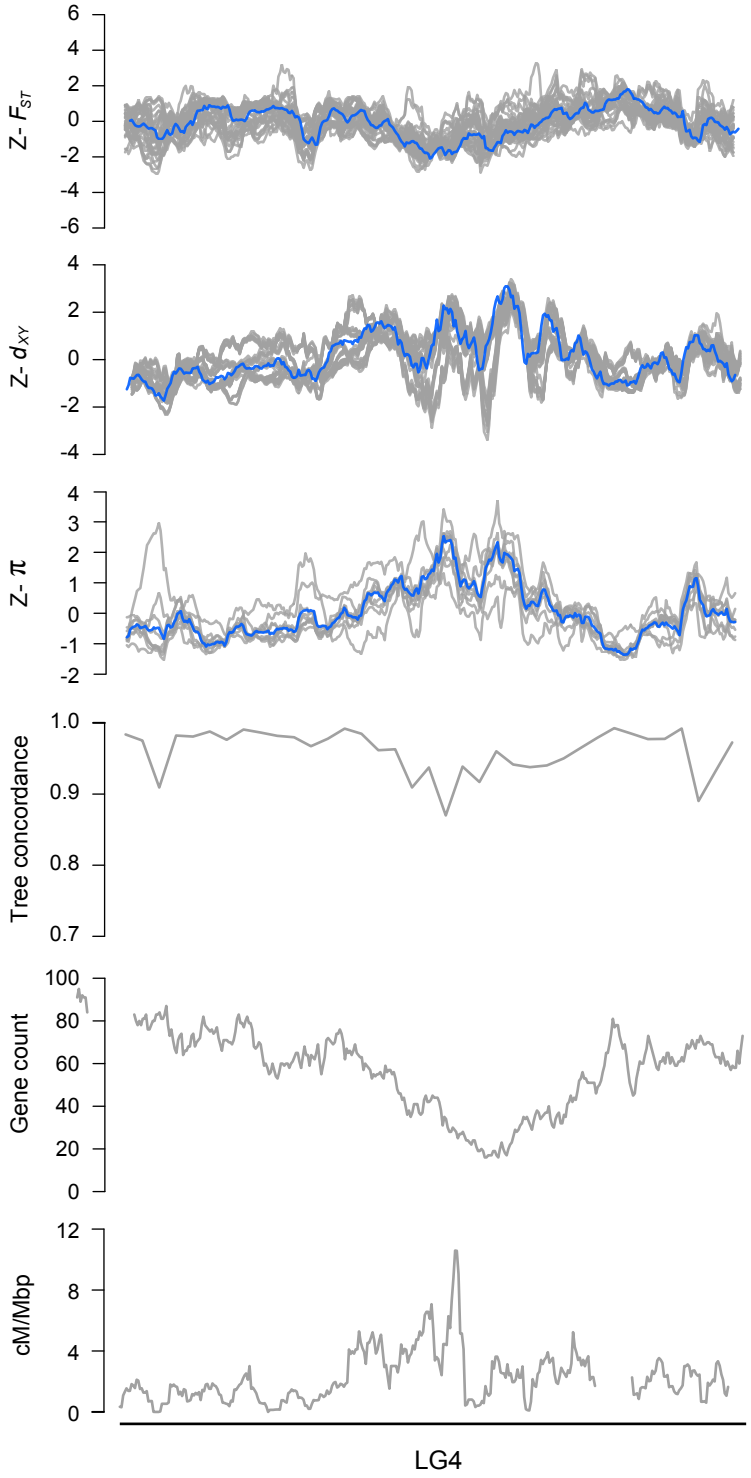

Figure S5. continued

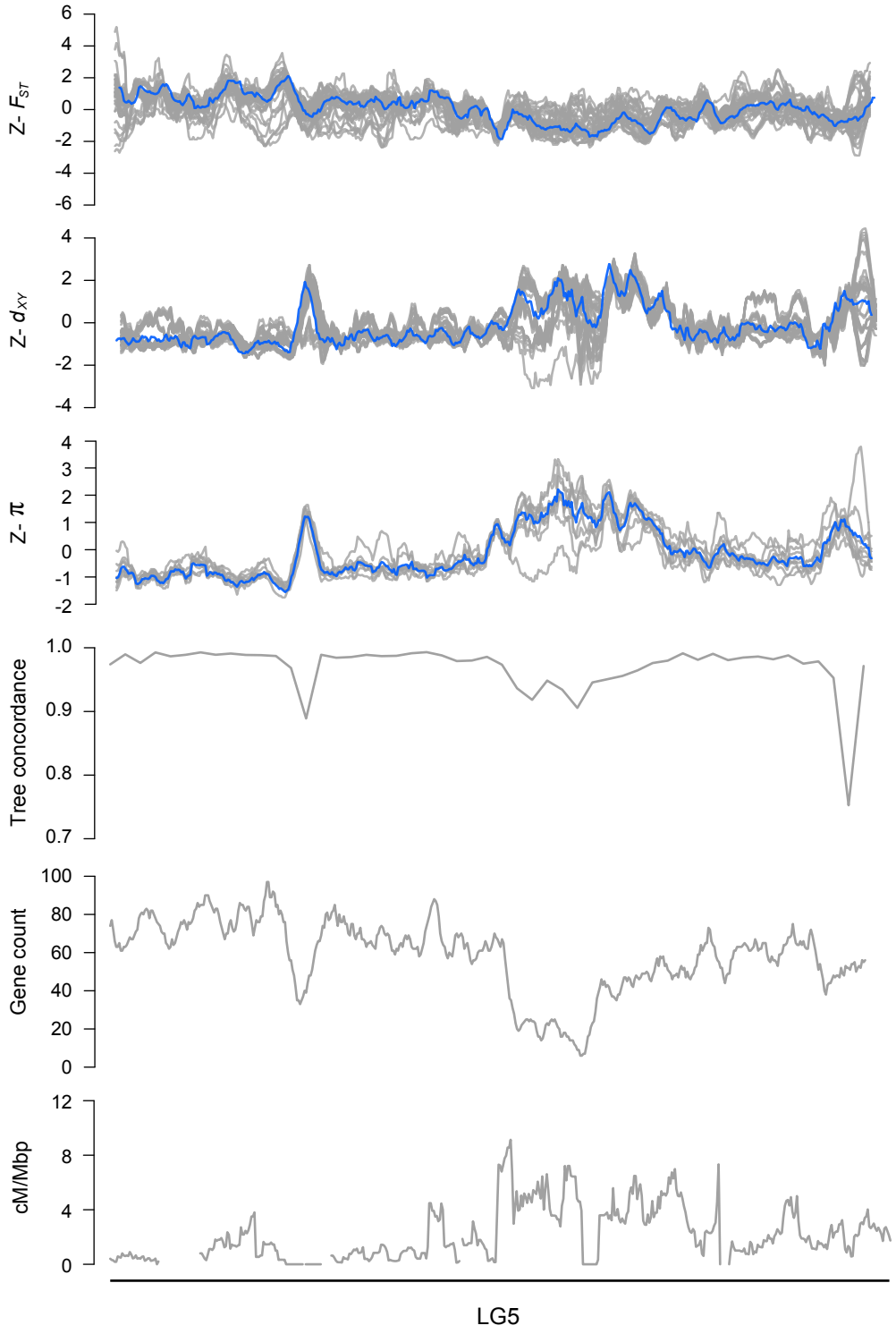

Figure S5. continued

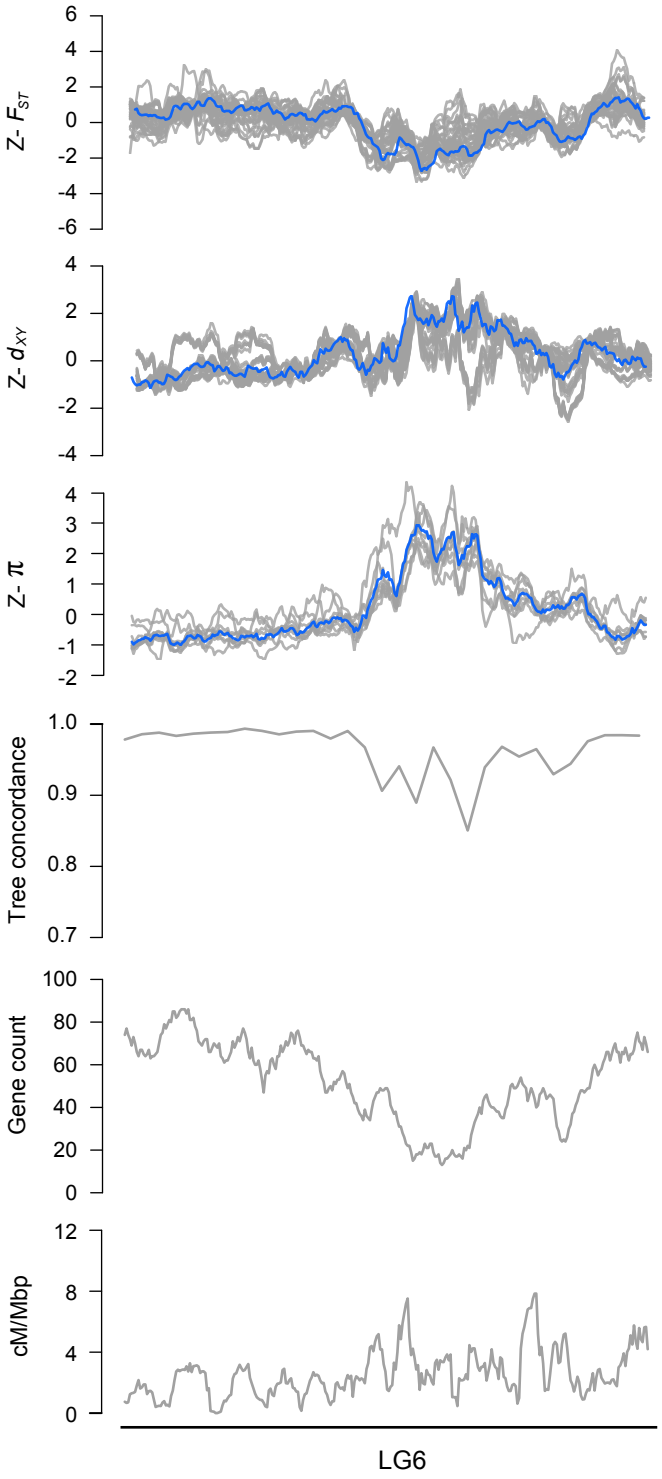

Figure S5. continued

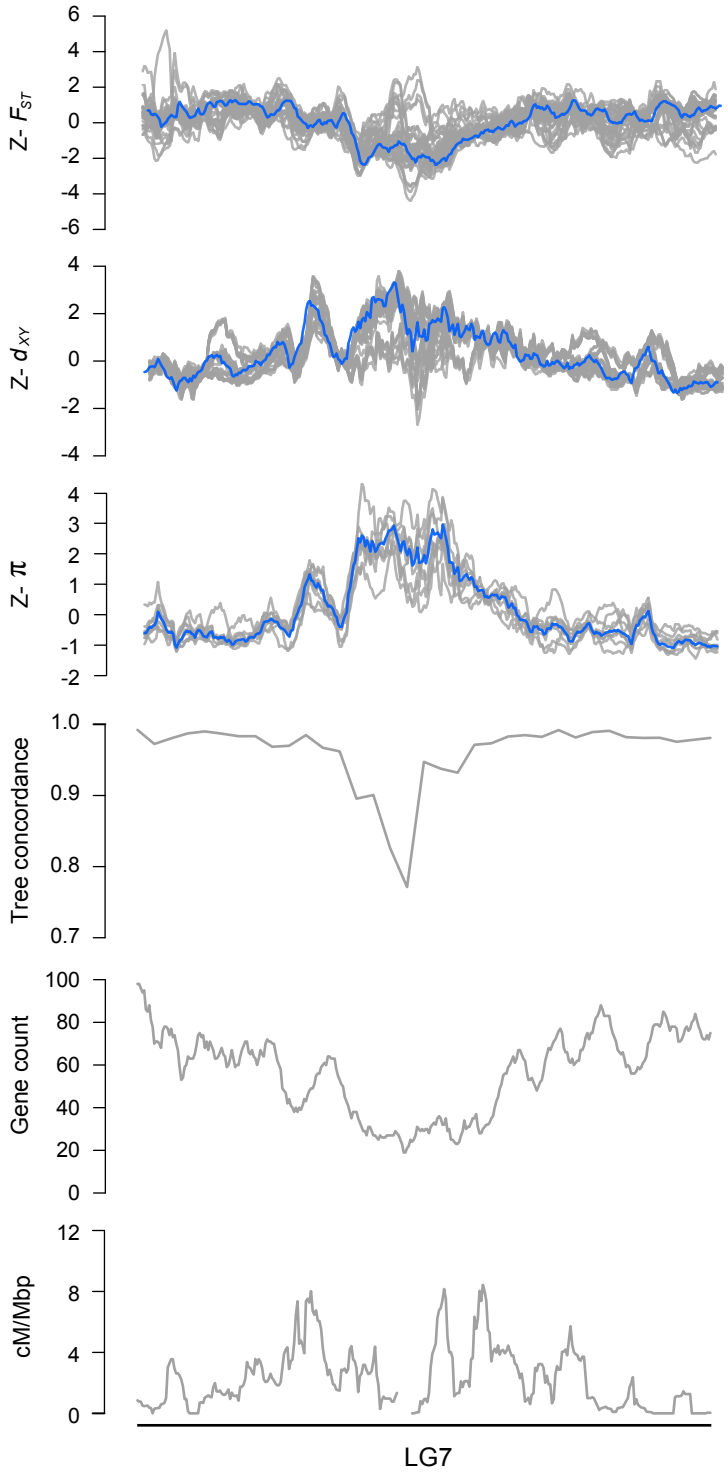

Figure S5. continued

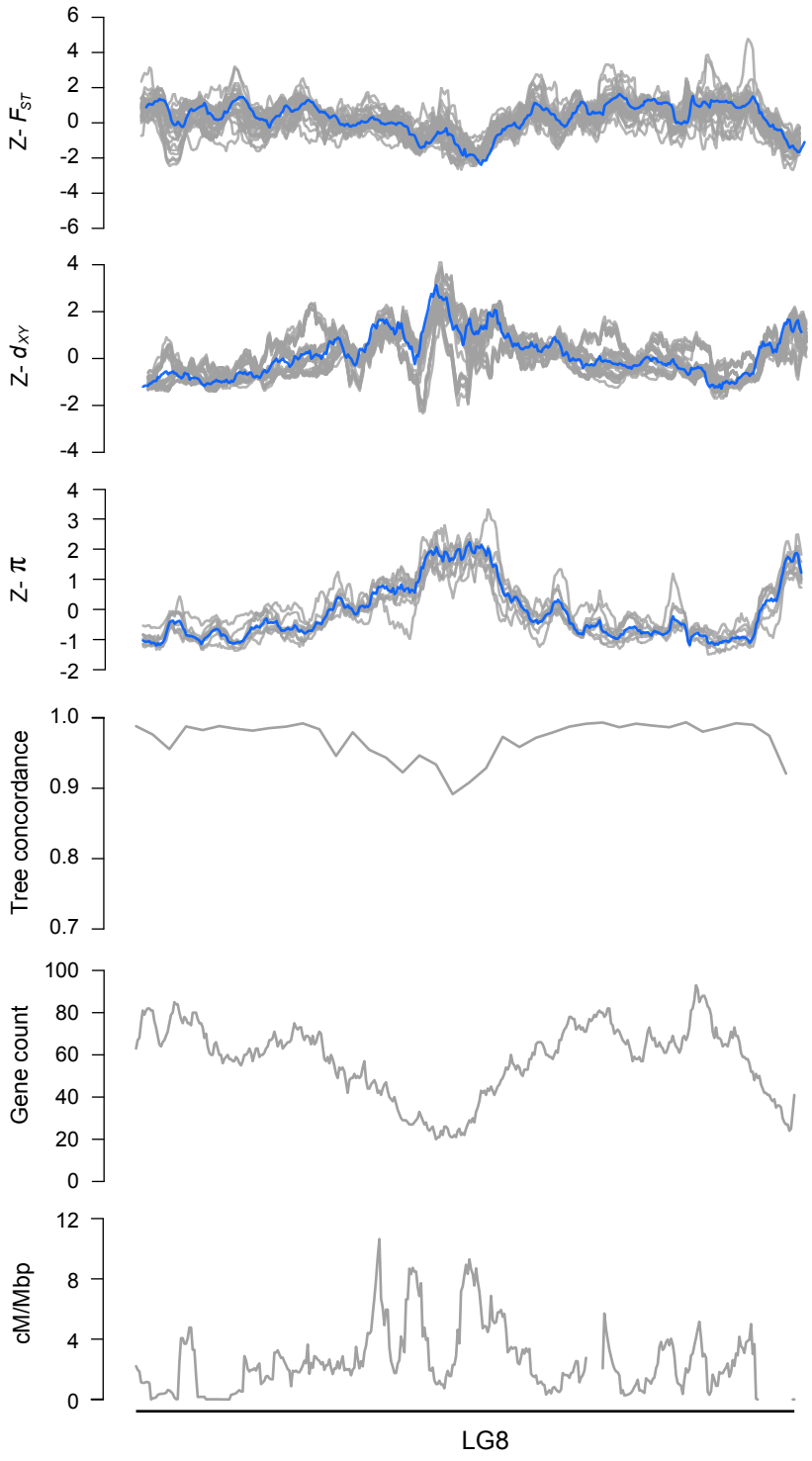

Figure S5. continued

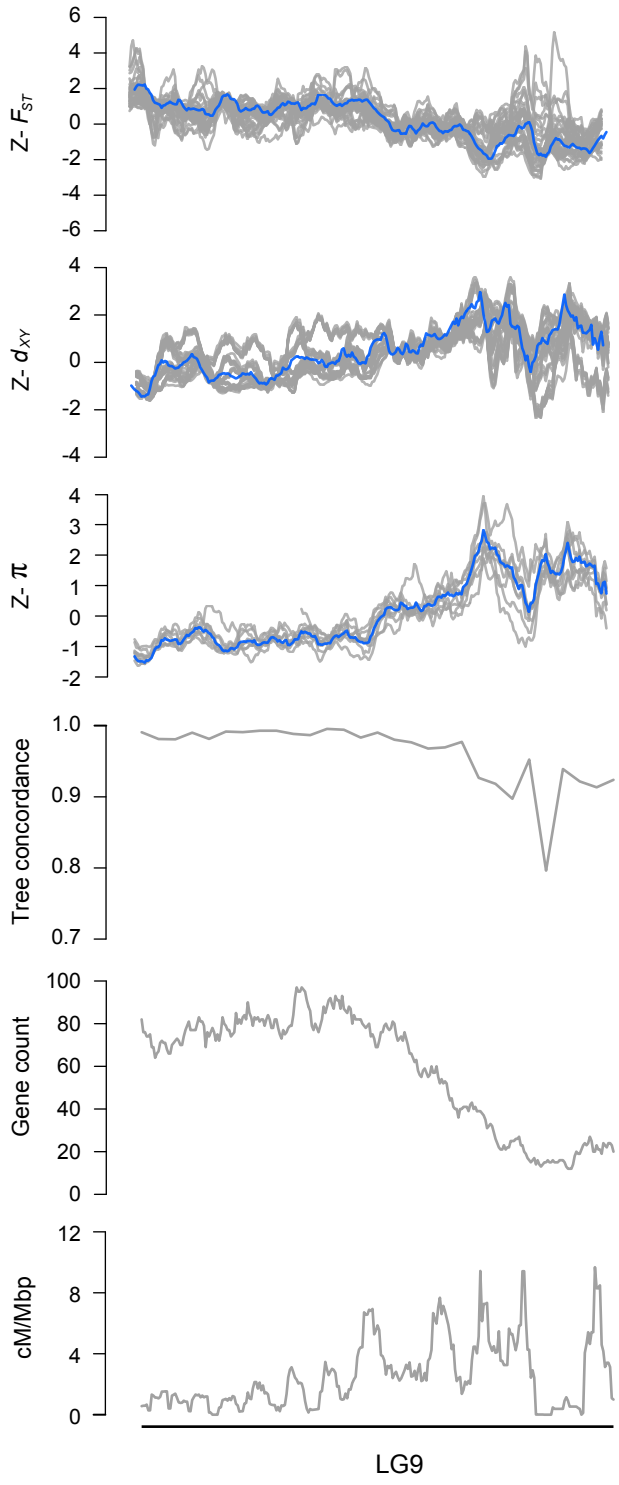

Figure S5. continued

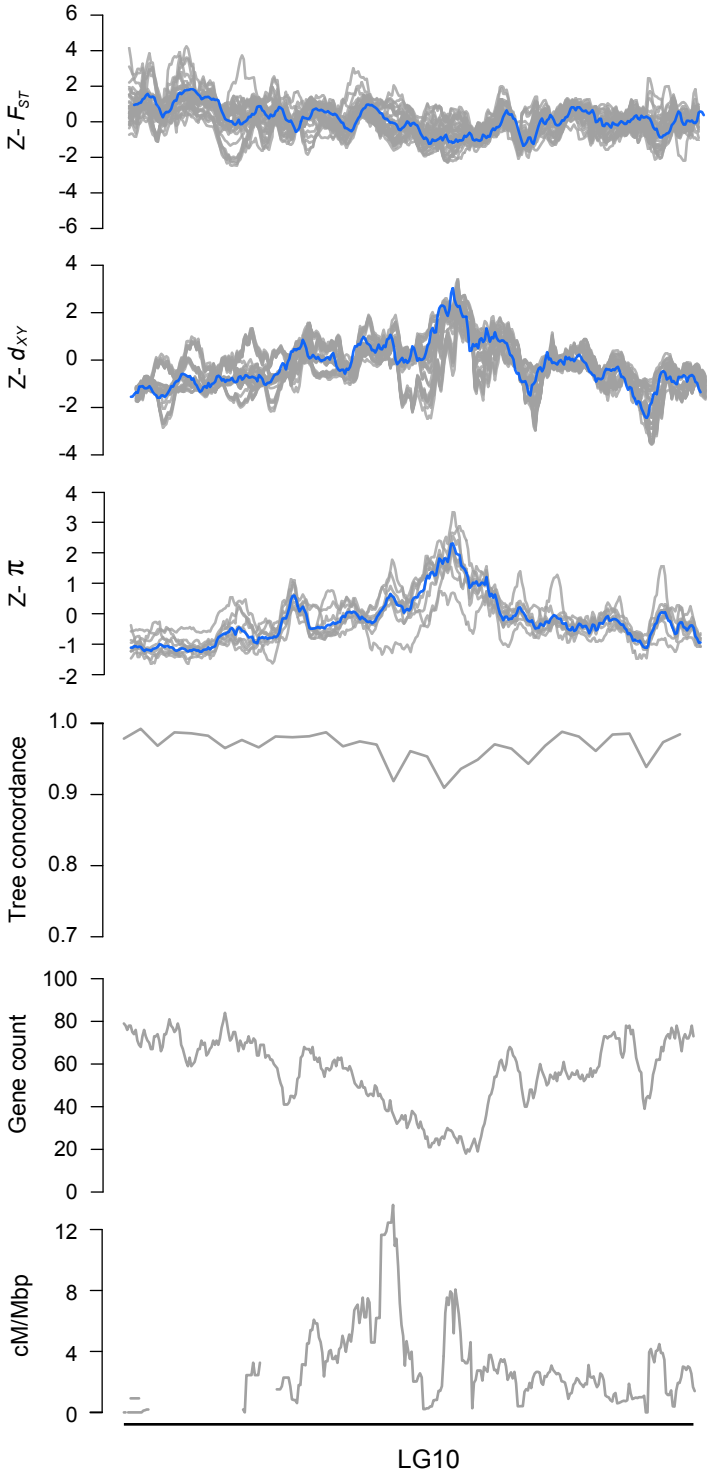

Figure S5. continued

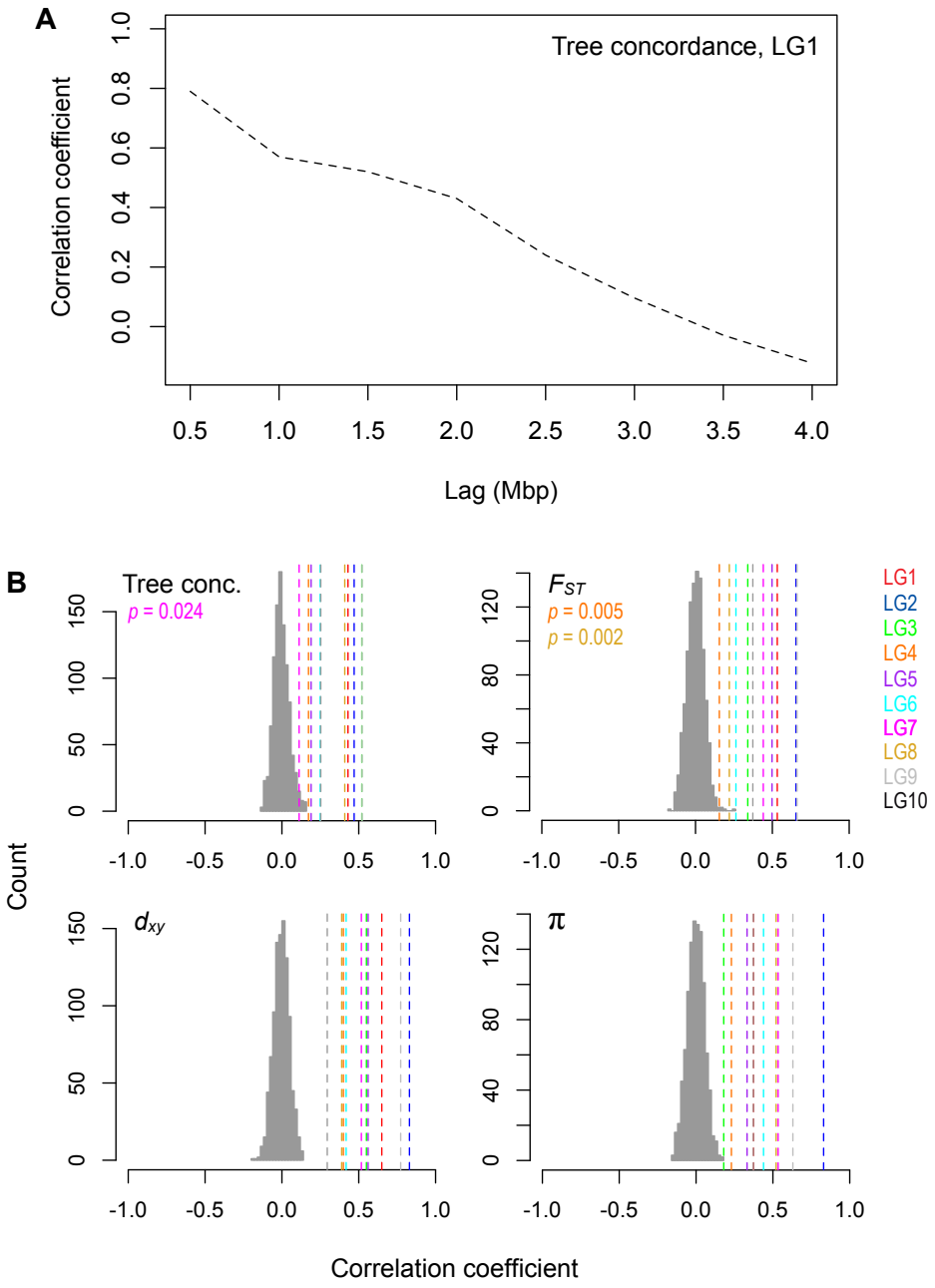

Figure S6. Patterns of variation are non-randomly distributed across the genome. A) Strong autocorrelation of tree concordance scores on LG1 over a Mbp scale. B) Levels of tree concordance, F_{ST}, d_{xy}, and π all show significant autocorrelation at the 2 Mbp scale. The dashed vertical lines show the observed autocorrelation coefficients for each LG with a 2 Mbp lag. The histogram shows the null distribution of autocorrelation coefficients (same lag) generated from 1000 random permutations of the genome-wide values. The observed data are significant at $p = 0.001$ unless stated otherwise.

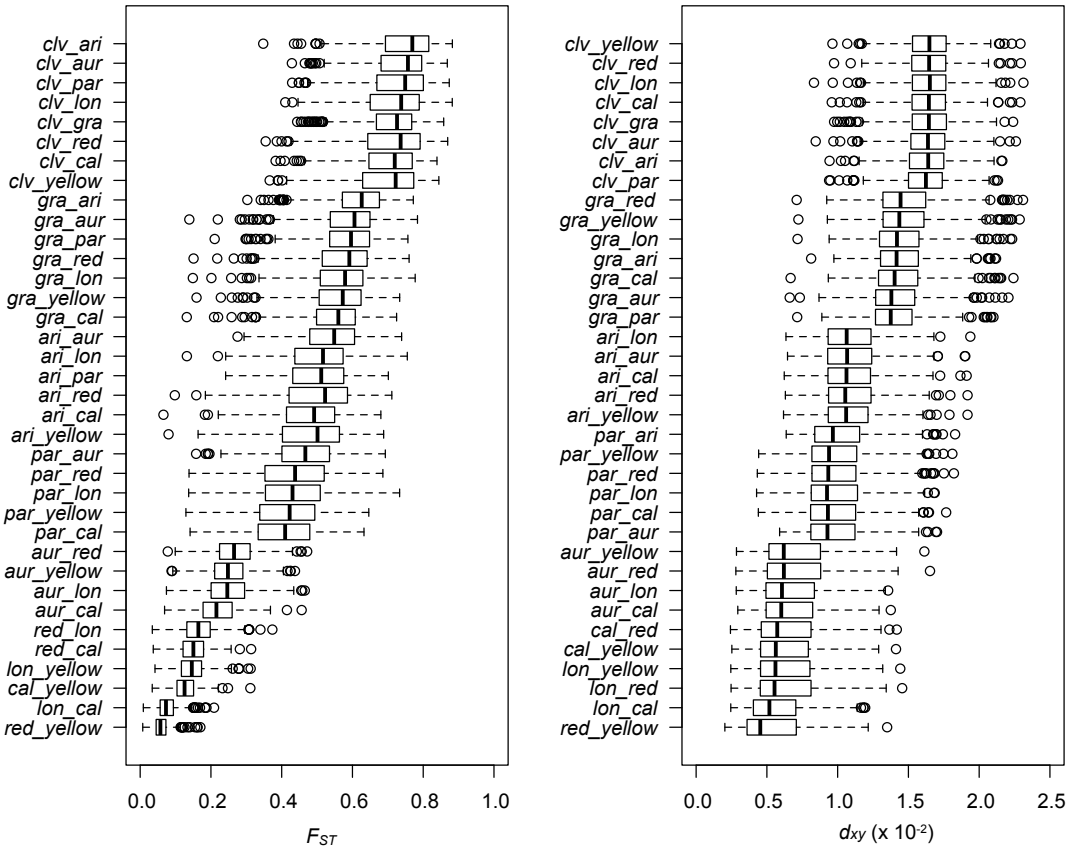

Figure S7. Patterns of differentiation and divergence for all 36 pairs of taxa. Box plots for each of the 36 pairwise taxonomic comparisons reveal the range of variation in F_{ST} and d_{xy} across the radiation. Moreover, the data show extensive variance among genomic windows within each comparison. Vertical black lines indicate the median, boxes represent the lower and upper quartiles, and whiskers extend to 1.5 times the interquartile range. Taxon abbreviations: *cal, calycinus; lon, longiflorus; aur, aurantiacus; par, parviflorus; ari, aridus; gra, grandiflorus; clv, M. clevelandii*.

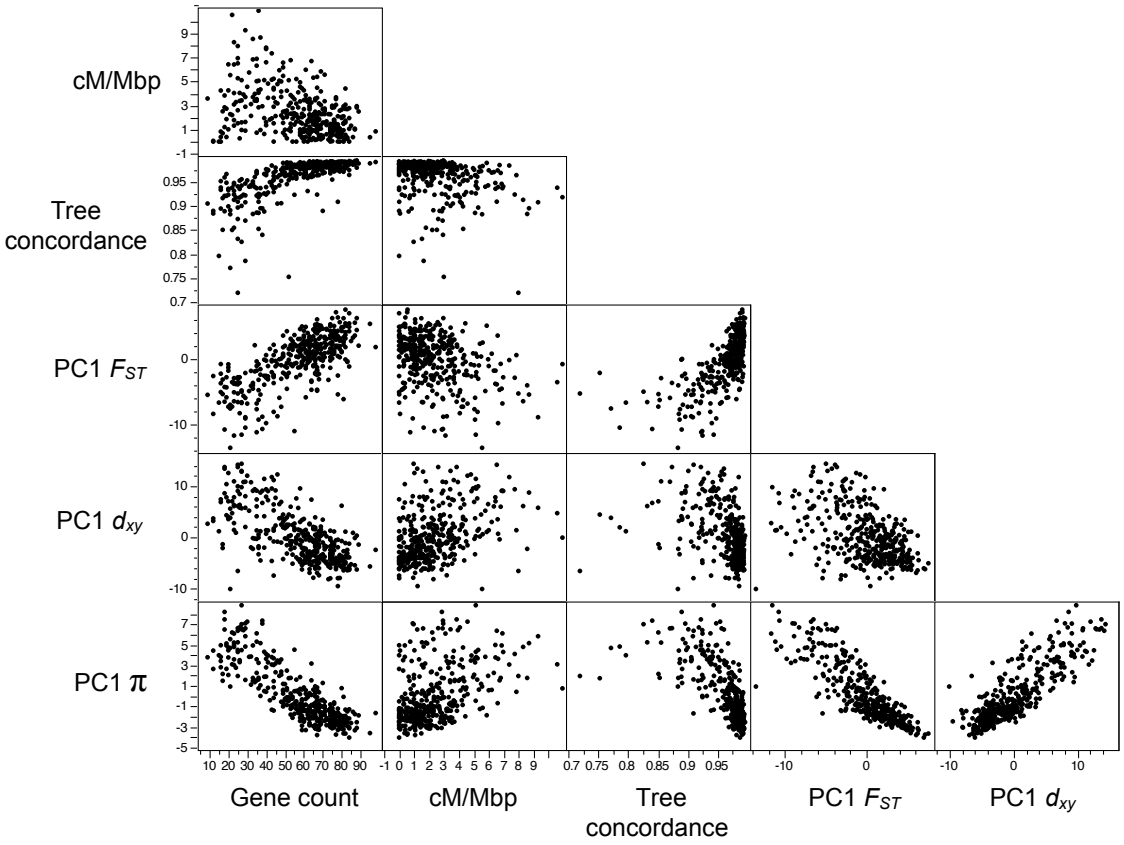

Figure S8. Bivariate plots among measures of variation and genomic features across 500 kb genomic windows. Note that this is the same as Figure 3 but with axes units. Also note that the axes are different across rows and columns of the matrix.

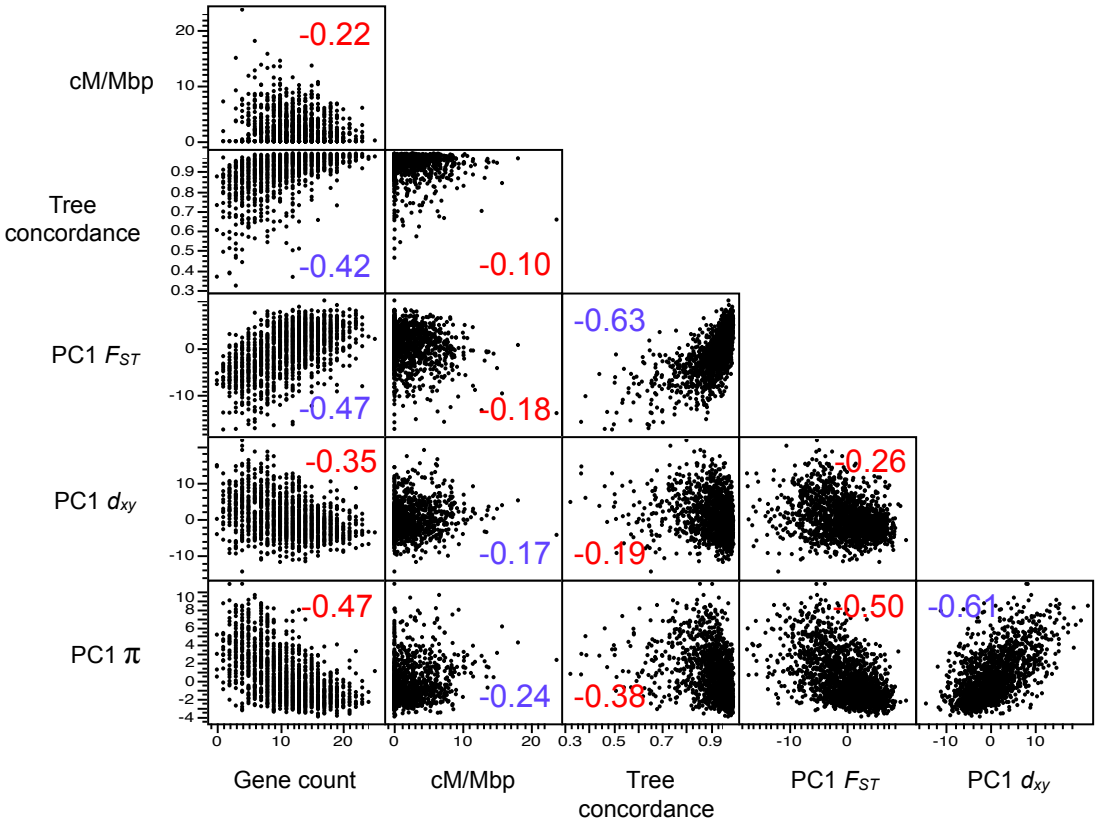

Figure S9. Bivariate plots among measures of variation and genomic features across 100 kb genomic windows. The number is the correlation coefficient. Positive correlation coefficients are colored blue and negative coefficients are colored red.

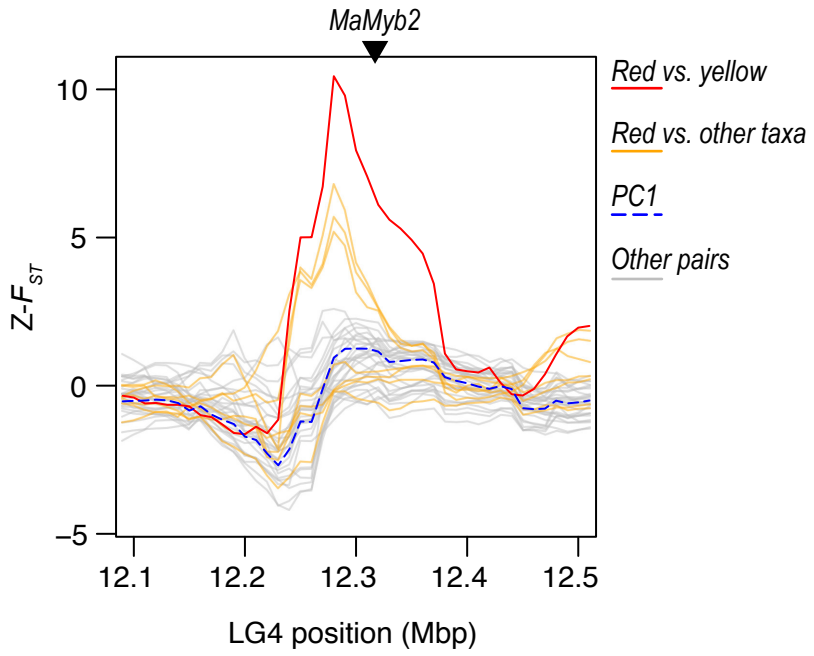

Figure S10. A large-effect adaptive locus shows a lineage-specific signature of positive selection. Plots of Z-transformed F_{ST} across the genome, estimated in 100 kb sliding windows (step size 10 kb). The red line shows values between the red and yellow ecotypes of subspecies *puniceus*. The orange lines show other comparisons with the red ecotype of *punicues*, and the gray lines show the values of all other comparisons. The dashed blue line shows the first PC calculated across all of the comparisons. The triangle marks the position of the gene *MaMyb2*. A *cis*-regulatory mutation that is tightly linked to this gene is responsible for the shift from yellow to red flowers (Corresponds to Figure 4 in the main text).

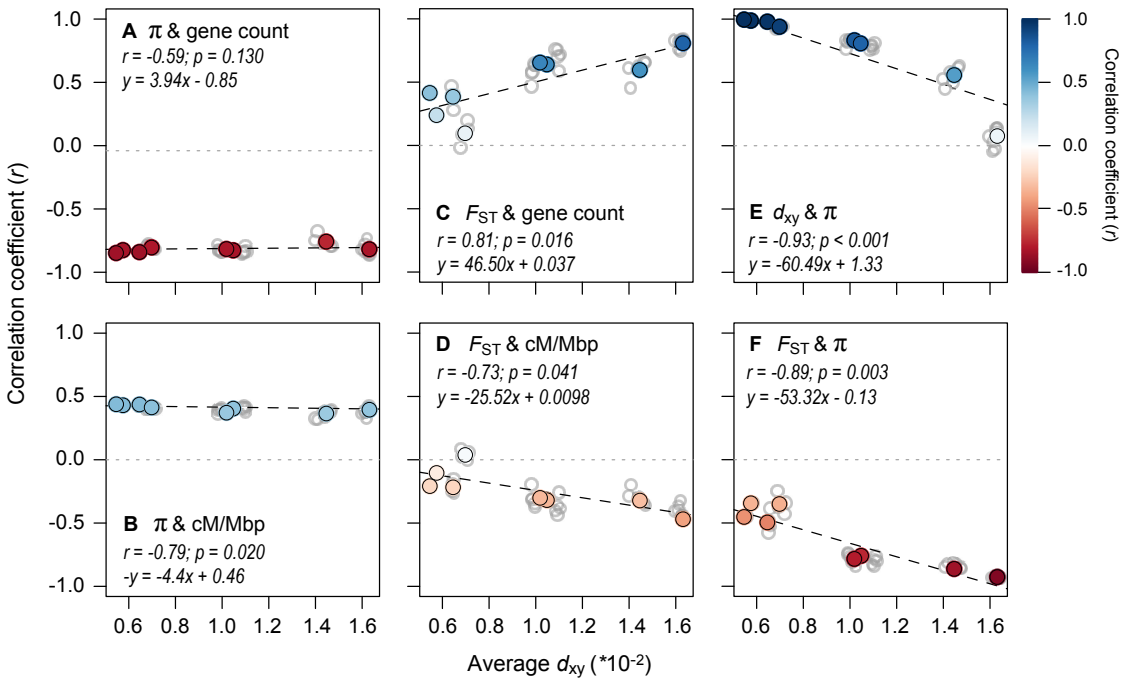

Figure S11. The range of divergence times reveals static and dynamic signatures of recurrent indirect selection. Correlations between variables (500 kb windows) for all 36 taxonomic comparisons (gray dots) plotted against the average d_{xy} as a measure of divergence time. The left panels show how the relationships between π (each window averaged across a pair of taxa) and (A) gene count and (B) recombination rate vary with increasing divergence time. The middle panels (C & D) show the same relationships, but with F_{ST}. The right panels show the relationships between (E) d_{xy} and π and (F) F_{ST} and π. The regressions (dashed lines) in each plot are fitted to the eight independent contrasts (colored points) obtained using a phylogenetic correction, with the regression equation and strength of the correlation given in each panel. The color gradient shows the strength of the correlation.

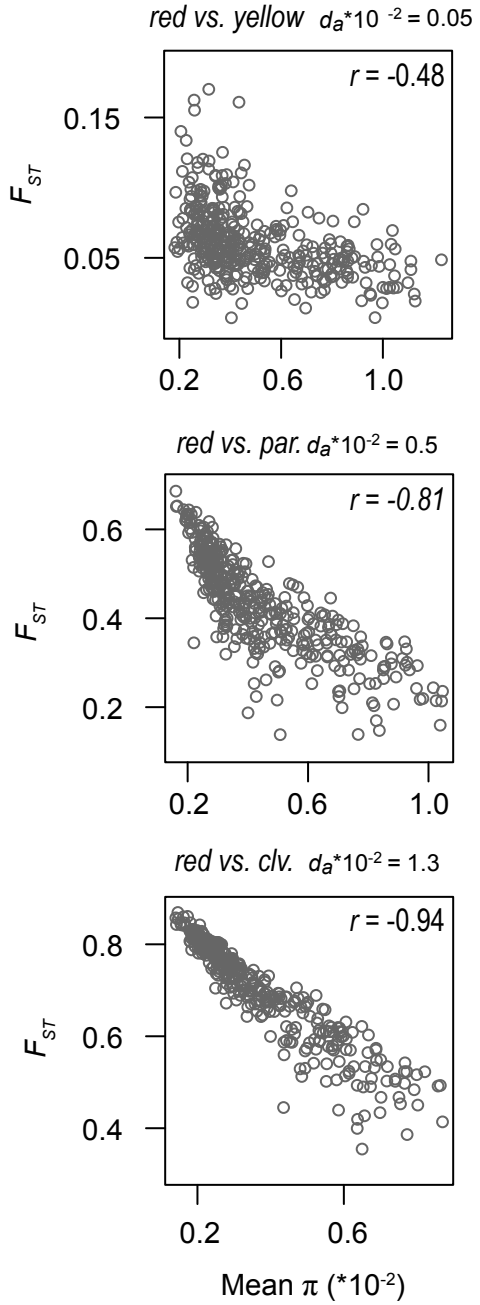

Figure S12. Negative correlation between nucleotide diversity and differentiation becomes stronger with increasing divergence time. Bivariate plots of the correlation between F_{ST} and π at varying levels of sequence divergence (d_{xy}).

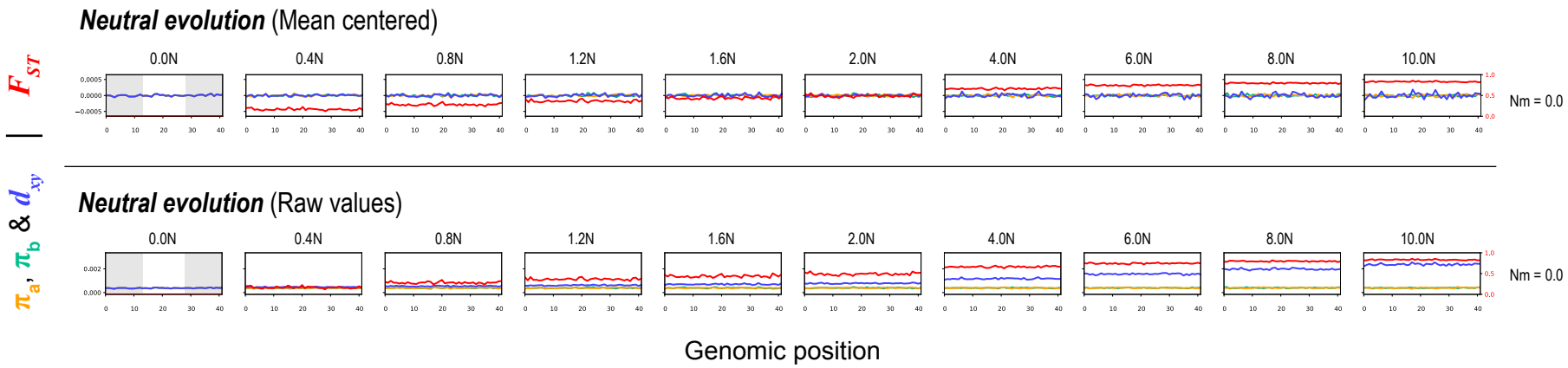

Figure S13. Genomic landscapes simulated under different different divergence histories. Each row of plots shows patterns of within- and between-population variation (π, d_{xy}, and F_{ST}) across a 21Mb chromosome (500 kb windows) at ten timepoints (in N generations, where N = 10,000) for one parameter combination of six scenarios: neutral divergence, background selection, Bateson-Dobzhansky-Muller incompatibilities, positive selection, background selection and positive selection, and local adaptation. The grey boxes in the first column show the areas of the chromosome that are constrained by selection. Mean centered (above line) and raw values (below line) of π, d_{xy}. The parameter Ns modulates the average selective coefficient (where s = Ns/N) while Prop is the proportion of new mutations that are not neutral, Nm is the average number of migrants per generation.

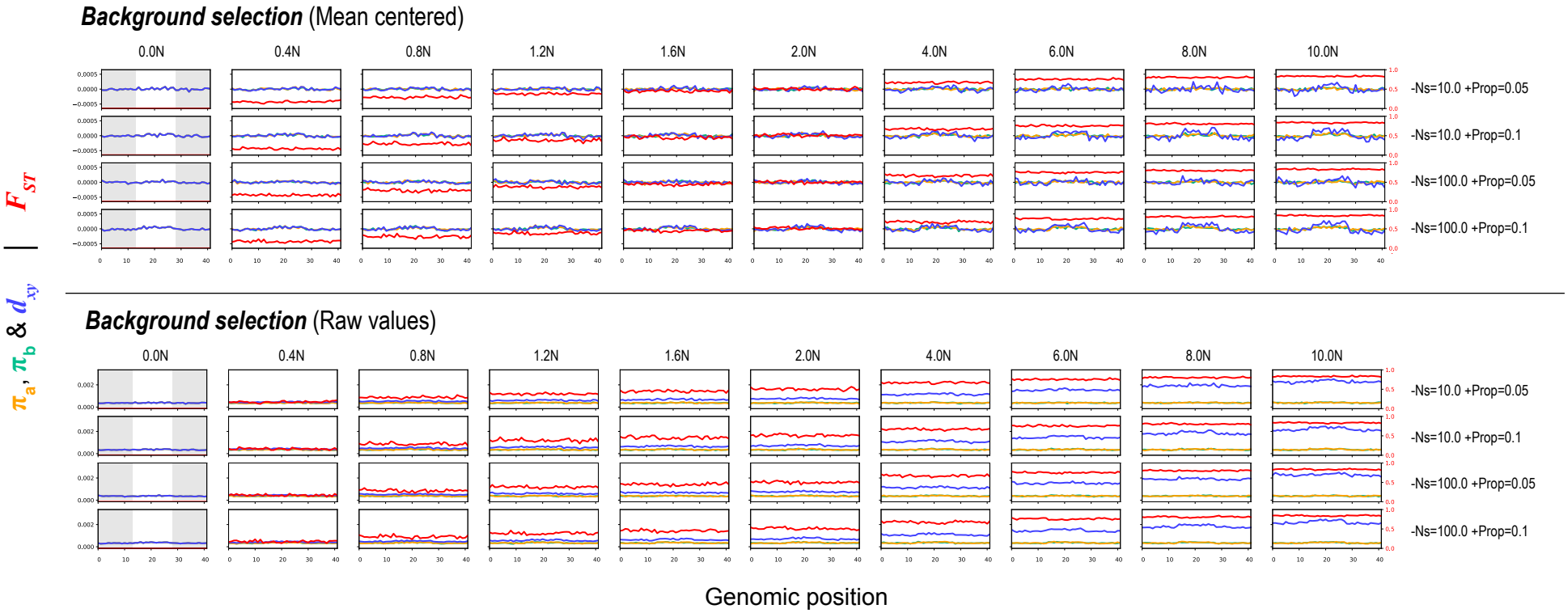

Figure S13 continued.

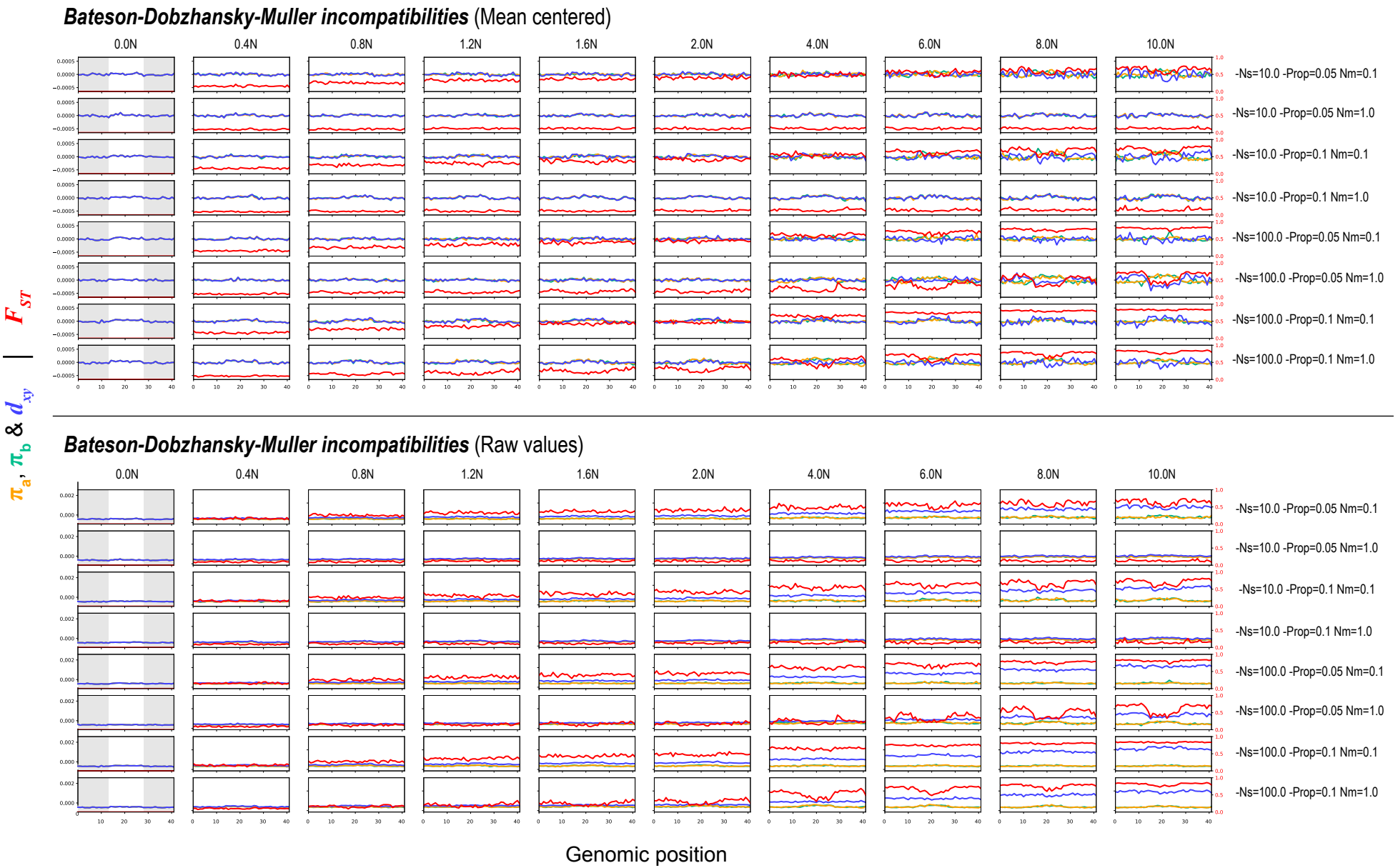

Figure S13 continued.

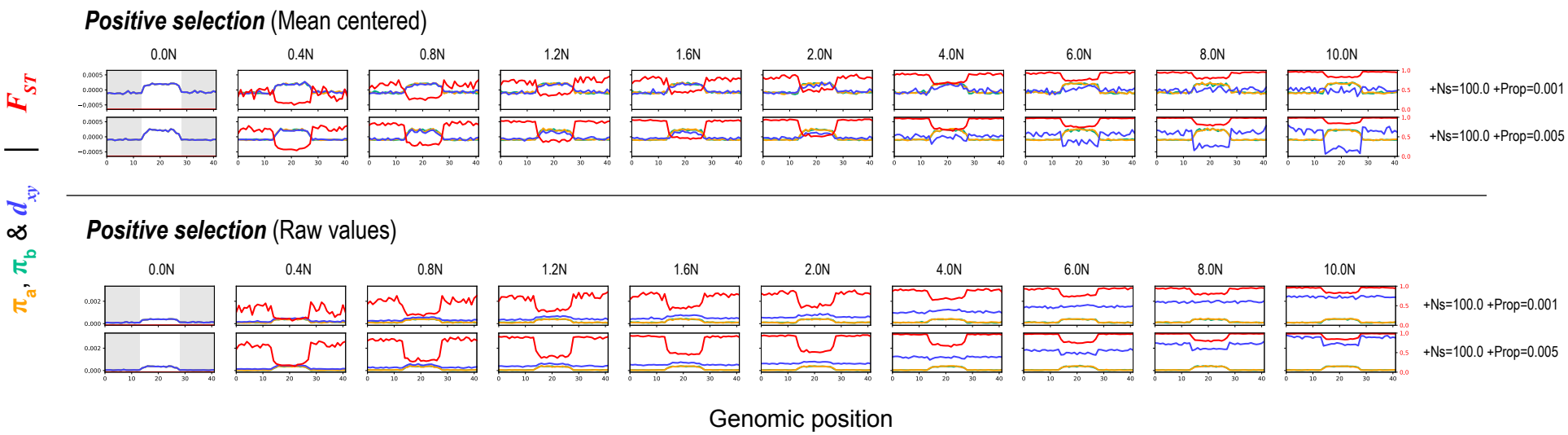

Figure S13 continued.

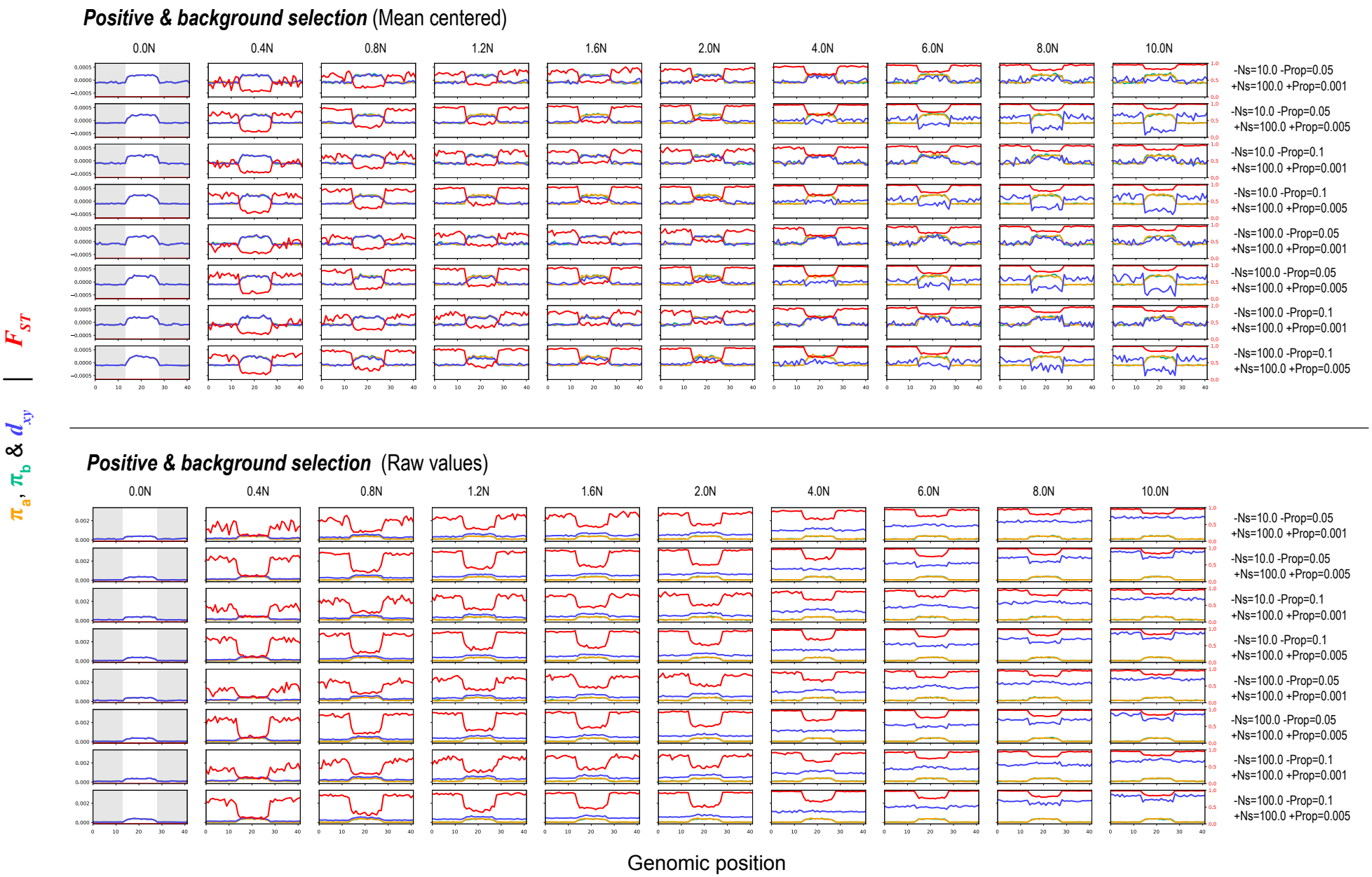

Figure S13 continued.

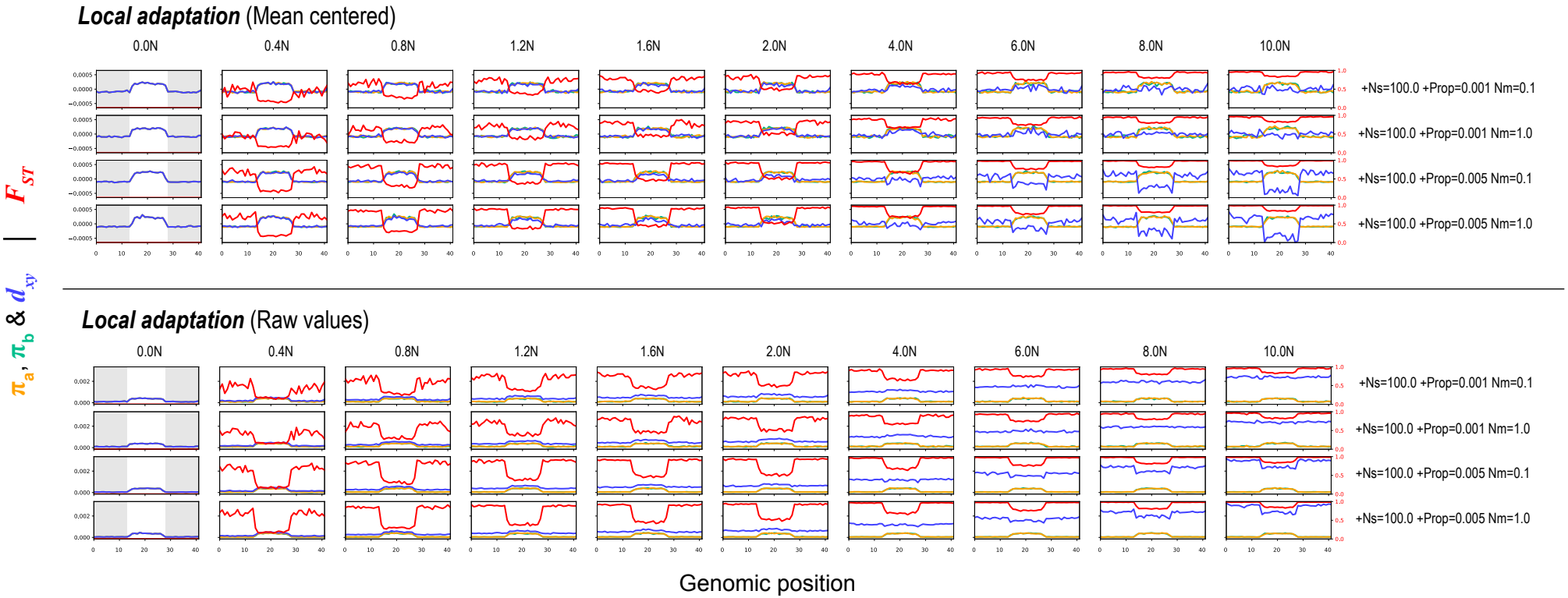

Figure S13 continued.

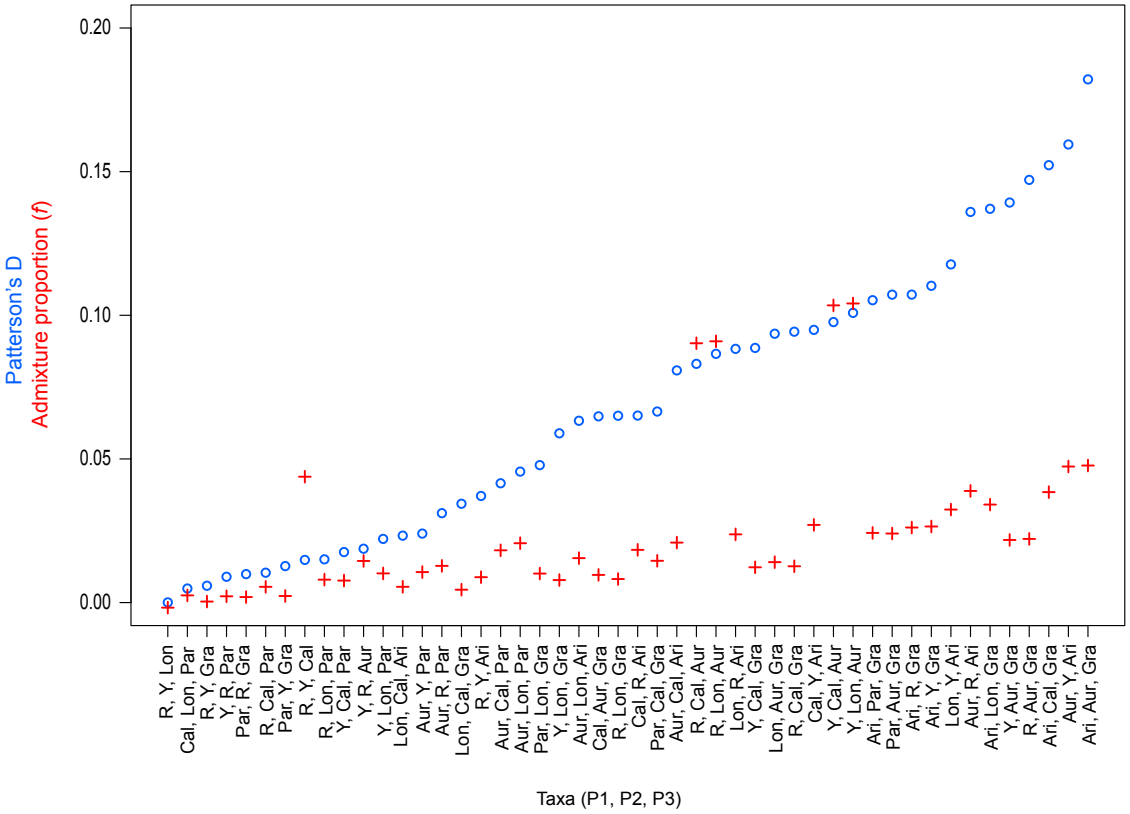

Figure S14. Evidence for widespread gene flow across the bush monkeyflower radiation. Genome-wide estimates of Patterson's *D* (blue circles) and the admixture proportion (*f;* red crosses) are shown for all 48 possible four taxon comparisons, with *M. clevelandii* as the outgroup in each test. The tests are ordered by increasing values of *D*, and each value is significant based on a block jackknife approach. The taxa included in each test are shown in the order P1, P2, P3 and *D* is always positive, meaning gene flow occurs between P2 and P3.

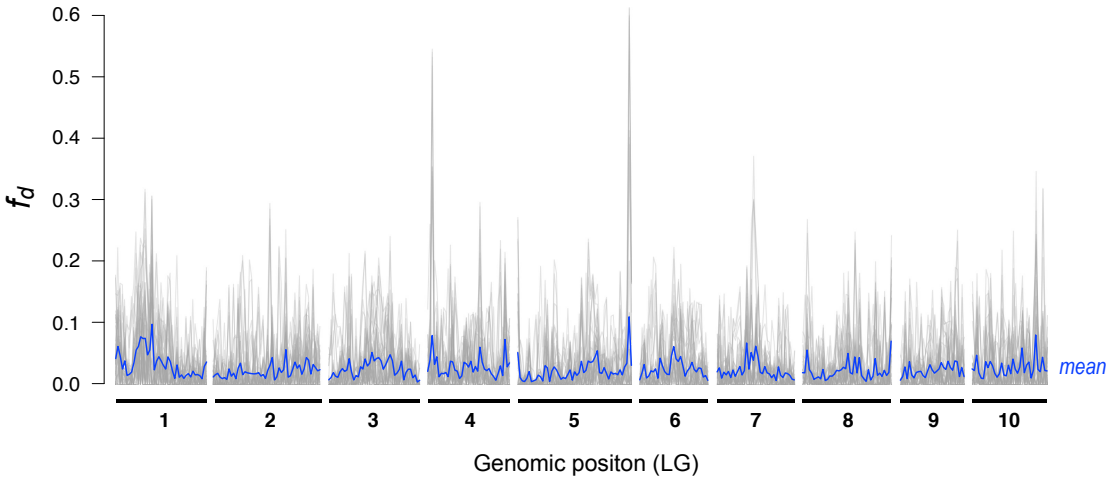

Figure S15. Variation in levels of admixture across the genome. The grey lines show measures of a modified test of the admixture proportion, f_d, estimated in 500kb non-overlapping windows for 48 different four-taxon comparisons, plotted across the 10 linkage groups of the bush monkeyflower genome. The blue line gives the mean value of f_d, calculated by taking the average value across all 48 tests in each genomic window.

Bonus Haiku!

Peaks and troughs of π,

Static, yet ever-changing.

Their reflection grows.