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

Sean Stankowski[¤a][*][☯], Madeline A. Chase[¤b][☯], Allison M. Fuiten[¤c], Murillo F. Rodrigues, Peter L. Ralph, Matthew A. Streisfeld[*]

Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon, United States of America
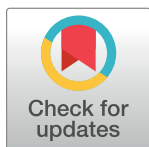
☯ These authors contributed equally to this work.
¤a Current address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom
¤b Current address: Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden
¤c Current address: Department of Dermatology, Oregon Health and Science University, Portland, Oregon, United States of America
* s.stankowski@sheffield.ac.uk (SS); mstreis@uoregon.edu (MAS)

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 radiation of monkeyflowers. After inferring the phylogenetic relationships among the 9 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 (BGS; 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. Combined with previous evidence for adaptation in this system, we conclude that the correlation in F_{ST} among these taxa informs us about the processes contributing to adaptation and speciation during a rapid radiation.

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 [1–3]. 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 [2]. 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 [4].

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 [1–3]. Numerous studies have revealed highly heterogeneous differentiation "landscapes" between pairs of taxa at different stages in the speciation process [5–13]. 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 [14]. Peaks in the differentiation landscape were interpreted as genomic regions containing loci underlying reproductive barriers, whereas valleys were thought to reflect regions that were homogenized by ongoing gene flow [6, 15, 16]. 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 [17]. 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 [17–23]. 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, because this is where LD breaks down most gradually [24]. 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 [25, 26] (but see [20, 22]), recent empirical studies indicate that they can have a strong impact on the topography of the differentiation landscape [1, 17, 27–29]. 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 [1, 28, 29]. For example, Burri and colleagues [30] 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 [12], *Heliconius* butterflies [10, 31], Darwin's finches [32], and other birds [33, 34].

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 [29] argued that selection against deleterious mutations (i.e., background selection [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 2 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 [29]. 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 [29].

Even though it is often argued that correlated differentiation landscapes reflect the action of recurrent BGS, 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 [35, 36]. Moreover, if adaptation is highly polygenic, positive selection will inevitably impact most of the genome [37]. 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 in which the recombination rate is low [38, 39]. 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 8 taxa of *Mimulus aurantiacus* distributed mainly throughout California (Fig 1; [40]). Together with their sister species *M. clevelandii*, they span a range of divergence times over the past approximately 1 million years. The plants inhabit a range of environments, including temperate coastal regions, mountain ranges, semi-arid habitats, and offshore islands [41]. Crossing experiments have shown that all of these taxa are at least partially interfertile [42], and many hybridize in nature in narrow regions where their distributions overlap [43–47]. Despite their close evolutionary relationships and opportunities for gene flow, these taxa show striking phenotypic differentiation [40]. 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 [45, 47]. In his seminal works on plant speciation, Grant [48–50] 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 [43–47].

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
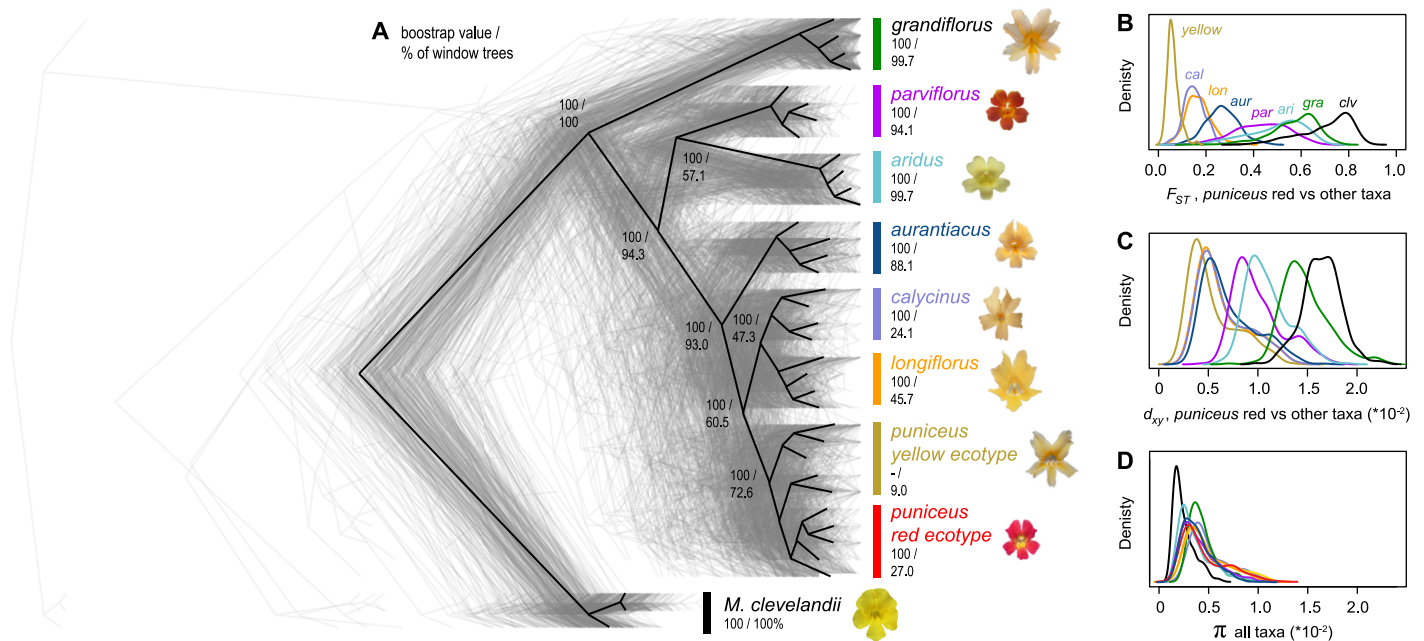
Fig 1. Evolutionary relationships and patterns of genome-wide variation across the radiation. (A) The black tree was constructed from a concatenated alignment of genome-wide SNPs and is rooted using *M. clevelandii*. The 387 gray trees were constructed from 500-kb genomic windows. The first number associated 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. Flower photos were taken by the authors. (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 S7 Fig for the distributions of F_{ST} and d_{xy} across all pairs of taxa.

https://doi.org/10.1371/journal.pbio.3000391.g001

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 BGS (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.

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 segregating markers across 10 linkage groups (LGs); S1 Fig; S1 Table). Contig building and scaffolding yielded 1,547 scaffolds, with an N50 size of 1.6 Mbp and a total length of 207 Mbp. The high-density map allowed us to anchor and orient 94% of the assembled genome onto 10 LGs, which is the number of chromosomes inferred from karyotypic analyses in all subspecies of *M. aurantiacus* and *M. clevelandii* [51]. Analysis of assembly completeness based on conserved gene space [52] revealed that 93% of 1,440 universal single-copy orthologous genes were completely assembled, with a

further 2% partially assembled (S2 Table). Subsequent annotation yielded 23,018 predicted genes.

Phylogenetic relationships among taxa and patterns of discordance across the genome

To infer phylogenetic relationships among the taxa in this radiation, we sequenced 37 whole genomes from the 7 subspecies and 2 ecotypes of *Mimulus aurantiacus* (*n* = 4–5 per taxon) and its sister taxon *M. clevelandii* (*n* = 3; S2 Fig; S3 Table). Close sequence similarity allowed us to align reads from all samples to the reference assembly with high confidence (an average of 91.7% reads aligned; S3 Table). After mapping, we identified 13.2 million variable sites that were used in subsequent analyses (average sequencing depth of 21× per individual; S3 Table). Relationships were then inferred among the 9 taxa from a concatenated alignment of genome-wide SNPs using maximum-likelihood (ML) phylogenetic analysis in *RAxML* [53].

The tree topology obtained from this analysis (Fig 1) confirmed the same phylogenetic relationships as previous analyses based on reduced-representation sequencing and 5 different methods of phylogenetic reconstruction [40, 54] and was supported by patterns of clustering from principal components analysis (PCA; S3 Fig). Individuals of each of the 7 subspecies formed monophyletic groups with 100% bootstrap support (Fig 1). Relationships within subspecies *puniceus* were more complex, because the red ecotype formed a monophyletic subclade within the paraphyletic yellow ecotype. This is consistent with the recent origin of red flowers from a yellow-flowered ancestor [54].

Although the whole-genome phylogeny provides a well-supported summary of the relationships among these taxa, concatenated phylogenies can obscure phylogenetic discordance in more defined genomic regions [55] (Fig 1). To test for fine-scale phylogenetic discordance, we next constructed ML phylogenies for 500-kb and 100-kb genomic windows. We then calculated a "concordance score" for each tree by computing the correlation between the distance matrix generated from each window-based tree and the whole-genome tree, with a stronger correlation indicating that 2 trees have a more similar topology.

At the 500-kb scale, only 22 (6%) trees showed the same taxon branching order as the whole-genome tree. However, concordance scores tended to be very high for all of the trees (mean = 0.964, SD = 0.039; minimum = 0.719), suggesting that variation in the topologies was due primarily to minor differences in branching order. This was confirmed by quantifying how often each node in the genome-wide tree was recovered in the set of window-based trees (Fig 1). Specifically, we found that differences in branching order were associated with the most recent splits, which included pairs of closely related taxa. For example, even though subspecies *puniceus* was monophyletic in the majority of trees (72.6%), individuals from the red ecotype only formed a monophyletic group in 27% of the trees. Similarly, the closely related subspecies *longiflorus* and *calycinus* were monophyletic in only 45.7% and 24.1% of trees, respectively. Higher discordance among these closely related, geographically proximate taxa is likely due to 2 factors. First, only a short time has passed since they shared a common ancestor, meaning that ancestral polymorphisms have had little time to sort among lineages. Second, ongoing gene flow between some taxa may have opposed sorting, prolonging the retention of ancestral variants among them.

Next, we examined how the level of phylogenetic discordance varied across the bush monkeyflower genome. If the discordance was due to the stochastic effects of neutral processes, variation in tree concordance scores should be distributed randomly across the genome [56]. To test this prediction, we plotted the tree concordance scores across the 10 LGs (Fig 2A; see S4 Fig for results from 100-kb windows and S5 Fig for plots along each chromosome). Rather
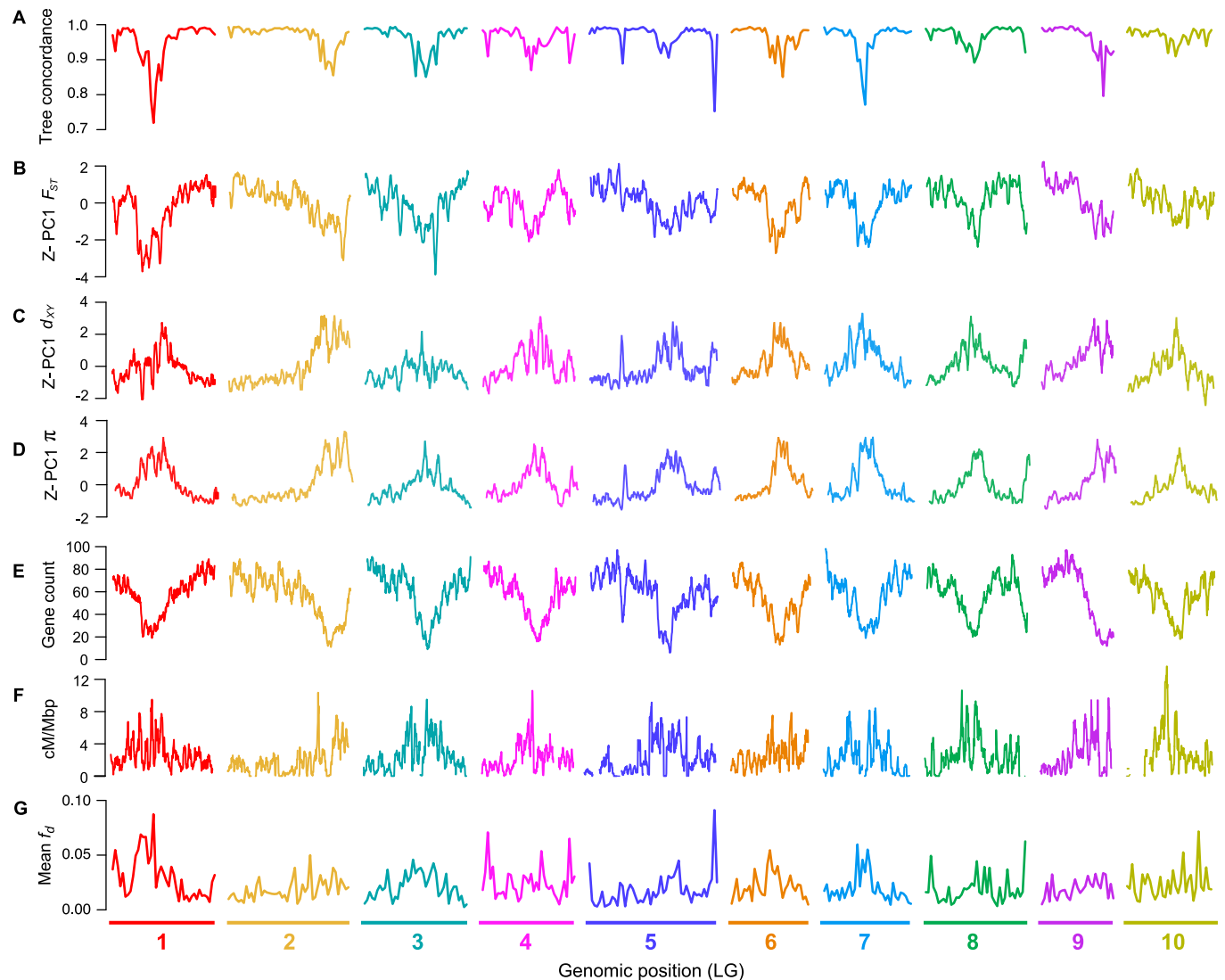
Fig 2. Common genomic landscapes mirror variation in the local properties of the genome. (A) Tree concordance scores for 500-kb nonoverlapping genomic windows plotted across the 10 bush monkeyflower chromosomes. (B–D) Plots of PC1 for F_{ST}, d_{xy}, and π in overlapping 500-kb windows (step size = 50 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 nonoverlapping genomic windows. See S4 Fig for the same plot made for 100-kb windows. LG, linkage group; PC1, first principal component.

https://doi.org/10.1371/journal.pbio.3000391.g002

than being randomly distributed, trees with lower concordance scores tended to cluster in relatively narrow regions of all 10 chromosomes (Fig 2A; autocorrelation analysis permutation tests $p = 0.001–0.023$; S6 Fig). This nonrandom pattern indicates that the rate of sorting varies along chromosomes, which could be due to variation in the strength of indirect selection across the bush monkeyflower genome.

Correlated patterns of genome-wide variation across the radiation

To gain deeper insight into the evolutionary processes that have shaped patterns of genome-wide variation during this radiation, we used 3 summary statistics to quantify patterns of diversity (π), divergence (d_{xy}), and differentiation (F_{ST}) within and between these taxa (the

same 500-kb and 100-kb sliding windows as above). π, or nucleotide diversity, was used to quantify the level of genetic variation within each taxon. It is defined as the average number of nucleotide differences per site between 2 sequences drawn from the population [57]. Second, we calculated d_{xy} as a measure of sequence divergence between all 36 pairs of taxa. In principle, π and d_{xy} are the same measure, except that the former is calculated between sequences within a taxon and the latter between pairs of taxa; they have the same units and both are proportional to the average coalescence time of the sequences being compared multiplied by mutation rate. Third, we calculated F_{ST} between each pair of taxa. Unlike d_{xy}, which is a measure of divergence between sequences drawn from different taxa, F_{ST} measures differentiation between the taxa relative to the total diversity in the sample. Between samples x and y, F_{ST} is roughly equal to $1 - \pi/d_{xy}$, where π is the mean diversity in the 2 taxa [58]. F_{ST} is strongly influenced by current levels of within-taxon diversity, whereas d_{xy} is strongly influenced by the level of ancestral variation when divergence times are small. The use of the term "divergence" to describe both d_{xy} and F_{ST} has caused some confusion in the literature, leading to alternative naming schemes. Here, we follow a recently proposed convention, referring to d_{xy} as "divergence" and F_{ST} as "differentiation" [2].

The variation in F_{ST} among all 36 pairs of taxa highlights the continuous nature of differentiation across the group (Fig 1B; S7 Fig), with mean window-based estimates ranging from 0.06 (red versus yellow ecotypes of *puniceus*) to more than 0.70. Distributions of divergence (d_{xy}) show a similar pattern (Fig 1C), with mean values ranging from 0.54% (red versus yellow ecotypes) to 1.6% (yellow ecotype versus *M. clevelandii*). Broad distributions of window-based estimates indicate high variability in levels of differentiation and divergence among genomic regions (Fig 1B and 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 in-group 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 nonrandomly distributed across broad regions of each chromosome ($p < 0.005$; Fig 2; S4 Fig, S5 Fig and S6 Fig). To account for the magnitude of variation in these statistics across all 9 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 (S5 Fig). Visual inspection of these data revealed that patterns of genome-wide variation in each statistic were qualitatively similar in all comparisons. We therefore used PCA to quantify their similarity and extracted a single variable (first principal component [PC1]) that summarized the common pattern (S5 Fig).

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, SD = 0.18; S4 Table 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 PC1 (mean loading for d_{xy} = 0.78, SD = 0.18; for π, 0.91, SD = 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 3 statistics).

Common genomic landscapes have been shaped by heterogeneous indirect selection

Observing highly similar genomic landscapes suggest that a common pattern of heterogeneous selection has shaped variation in all 9 taxa. Indeed, if a region experiences recurrent indirect
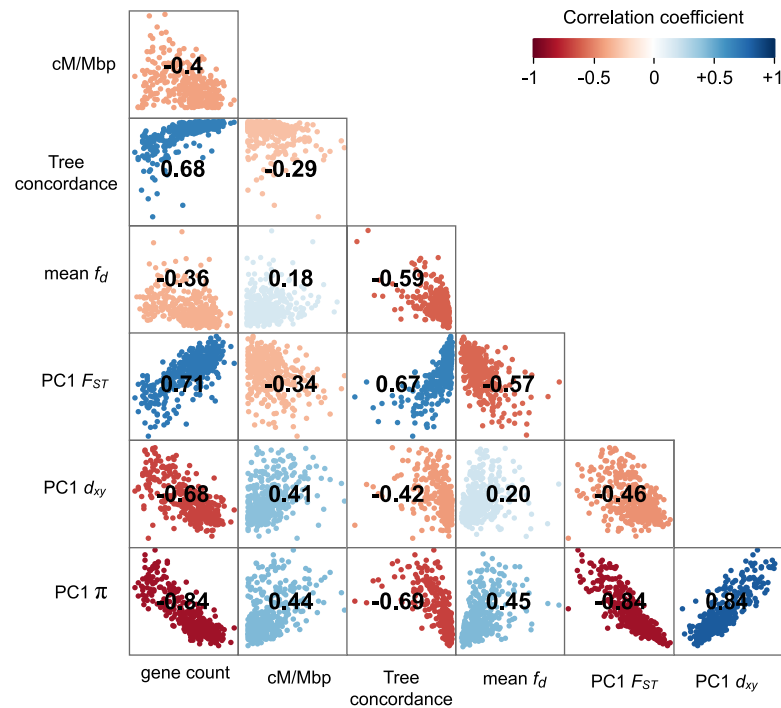
Fig 3. Correlations between measures of diversity and intrinsic features reveal the impact of heterogeneous indirect selection. Matrix of pairwise correlations between PC1 F_{ST}, PC1 d_{xy}, PC1 π, tree concordance, mean f_d, gene density, and recombination rate, all estimated in 500-kb nonoverlapping windows. The heat map indicates the strength of the correlation and its sign. All correlations are statistically significant at $p < 0.001$. (See S8 Fig for a more detailed correlation matrix and S9 Fig for the correlation matrix from 100 kb windows.) PC1, first principal component.

https://doi.org/10.1371/journal.pbio.3000391.g003

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 [19, 20, 29]. 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, S8 Fig for scatterplots and S9 Fig 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 [59].

In models of recurrent indirect selection, variation in its impact on associated sites is determined by the distribution of genomic features, including the local density of functional elements and the recombination rate [18, 19]. To test these theoretical predictions, we used our genome annotation and genetic map to calculate the number of protein coding genes and the average recombination rate (cM/Mbp) in each 500-kb window (Fig 2E and 2F; S4 Fig and S5 Fig). Regions of the genome with more functional elements tended to have a lower recombination rate ($r = −0.40$; $p < 0.0001$), leading to large variation in the predicted strength of indirect selection among regions. Consistent with the theoretical predictions outlined above, we found strong correlations between PC1 π and gene count ($r = −0.84$; $p < 0.0001$) and PC1 π and recombination rate ($r = 0.44$; $p < 0.0001$; Fig 3; S8 Fig and S9 Fig), indicating that diversity is indeed lower in regions where selection is predicted to have stronger indirect effects. However,

we did not observe a significant interactive effect of gene count and recombination rate on diversity (p = 0.057). This may be because the distribution of these features is correlated, making it difficult to tease apart their relative impacts.

Taken together, our results indicate that a common pattern of indirect selection has caused correlated genomic landscapes to evolve across this radiation. As predicted by theory [18, 19, 60] [21] and observed in diverse taxa [30, 59, 61], genome-wide variation in the strength of indirect selection is caused by the heterogeneous distributions of intrinsic genome features, namely, the density of functional elements and the local recombination rate.

### A known adaptive locus shows a strong deviation from the common pattern of differentiation

Because widespread signatures of indirect selection are often assumed to reflect the impact of heterogeneous BGS across the genome, several authors have proposed that the correlation in F_{ST} across multiple pairs of taxa can be used as a baseline for detecting genomic regions that have been affected by positive selection [29, 62]. More specifically, loci that have contributed to adaptation or speciation should be detectable as a positive deviation from the common pattern of differentiation (i.e., PC1), which is considered to reflect processes that are unrelated to ecologically relevant positive selection [29].

We have a unique opportunity to test the effectiveness of this comparative genomic approach by examining patterns of differentiation around a locus that is known to contribute to adaptation and speciation in subspecies *puniceus*. Using a candidate gene approach, Streisfeld and colleagues [63] showed that the shift from yellow to red flowers in *puniceus* was caused by a *cis*-regulatory mutation in the R2R3-MYB transcription factor *MaMyb2*. Patterns of haplotype variation within the gene indicate that the red allele was subject to strong positive selection and rapidly swept to fixation in what is now the geographic range of the red ecotype [54].

To test for a lineage-specific signature of differentiation at this locus, we examined patterns of Z-F_{ST} in a relatively narrow region (approximately 3 Mbp) surrounding *MaMyb2* (Fig 4; S10 Fig). At the 500-kb scale, the comparison between the red and yellow ecotypes shows a sharp peak centered near the flower color locus. The peak is strongly elevated above PC1 Z-F_{ST} (3.33 SDs), 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 SDs 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 (S10 Fig).

Correlated differentiation landscapes emerge rapidly following a population split

Although correlated landscapes may evolve due to the indirect effects of widespread BGS, it is also conceivable that they could evolve due to other evolutionary processes. To gain insight into the role that BGS 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.
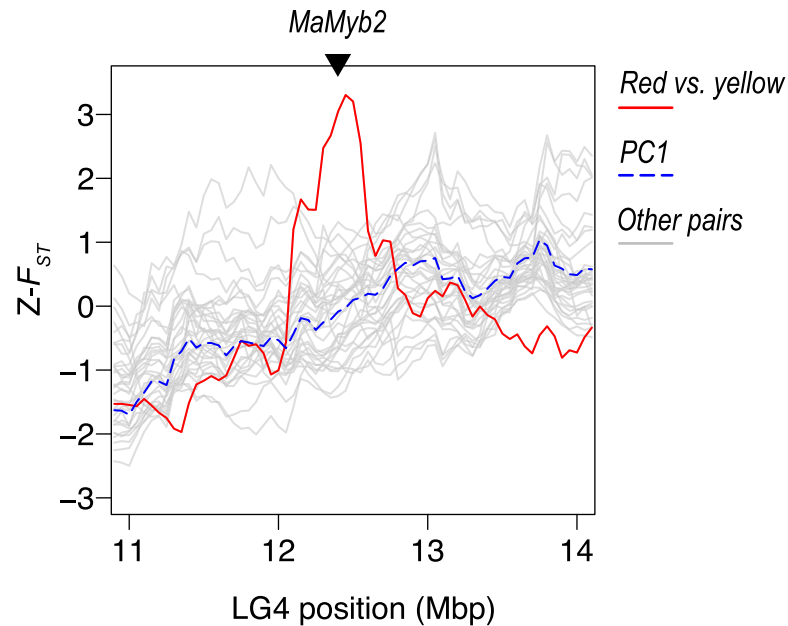
Fig 4. A large-effect adaptive locus shows a lineage-specific signature of positive selection. Plots of Z-transformed F_{ST} across the genome, estimated in 500-kb sliding windows (step size 50 kb). The red line shows values between the red and yellow ecotypes of subspecies *puniceus*, whereas the gray lines show the values of all other comparisons. The dashed blue line shows PC1 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 S10 Fig for the same plot made for 100 kb windows. LG, linkage group; PC1, first principal component.

https://doi.org/10.1371/journal.pbio.3000391.g004

In a verbal model, Burri [29] 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, because 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} − 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 S11 Fig for results with d_{xy} as a proxy for time). First, the presence of strong correlations between π (mean of the 2 taxa) and the distribution of genomic features barely changed with increasing divergence time between populations (Fig 5A and 5B; S5 Table). This is consistent with the prediction that diversity
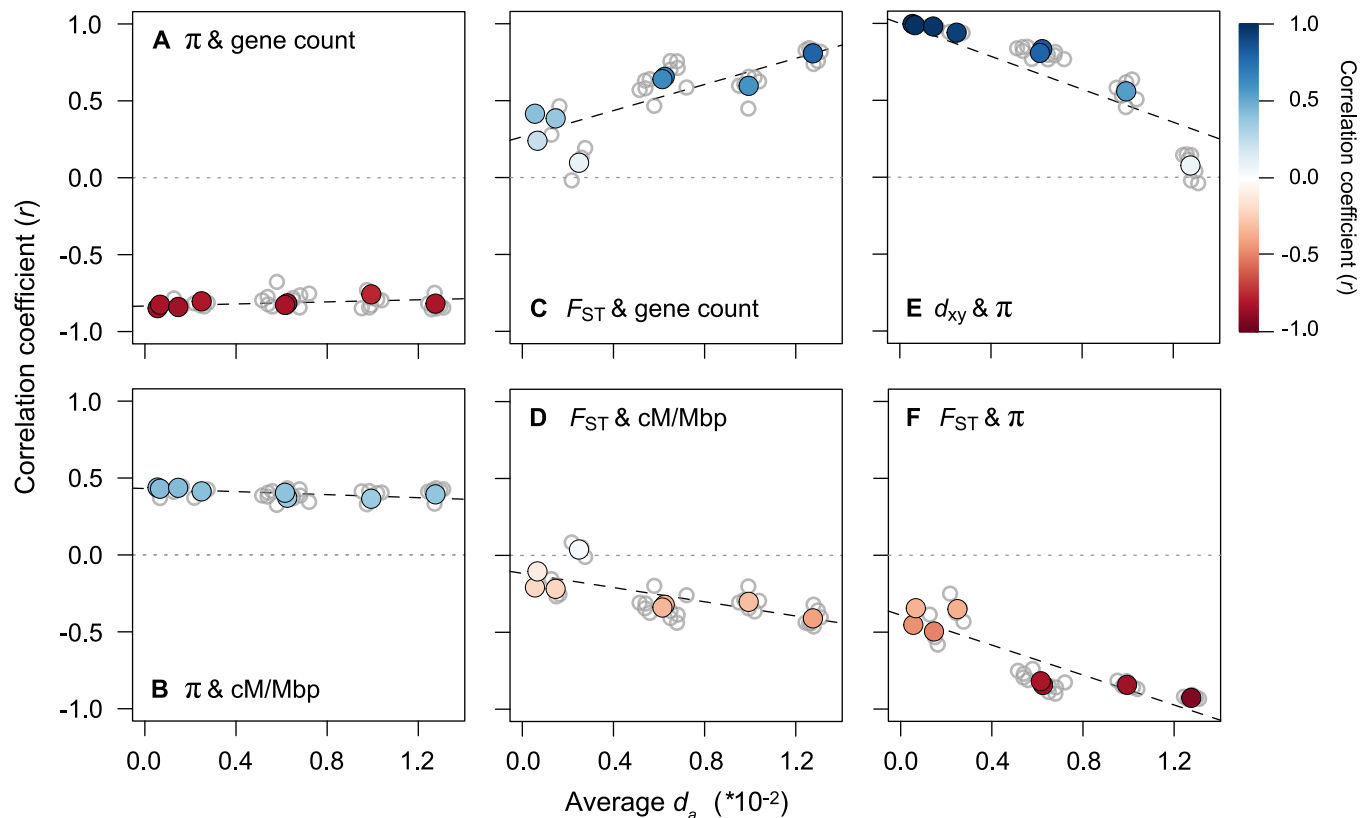
Fig 5. The range of divergence times reveals static and dynamic signatures of recurrent indirect selection. Correlations between variables (500-kb windows) for all 36 taxonomic comparisons (gray dots) plotted against the average d_a as a measure of divergence time. The left panels show how the relationships 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 and 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 8 independent contrasts (colored points) obtained using a phylogenetic correction. The color gradient shows the strength of the correlation.

https://doi.org/10.1371/journal.pbio.3000391.g005

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 and 5F; S5 Table). Thus, as predicted, the differentiation landscape increasingly reflects the distribution of intrinsic features as divergence time increases.

However, 2 of the observed patterns differed markedly from existing predictions [29]. 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; S12 Fig). 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.

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 [64, 65]. Our basic model consisted of an ancestral population (N = 10,000) that evolved for 10N generations and then split into 2 daughter populations that diverged for a further 10N generations. Each individual carried a 21 Mbp chromosome (similar to the physical size of a bush monkeyflower chromosome), partitioned into 3 equally sized regions. In the central third, all mutations were neutral, whereas some mutations in the 2 distal ends could affect fitness, depending on the scenario. Although simple, this partitioning scheme generates broad-scale variation in the strength of indirect selection across the chromosome and roughly approximates the distributions of genomic features in this system (e.g., LG 7 in Fig 2).

We implemented 6 modifications to this basic model: (i) Neutral divergence: no mutations affect fitness; (ii) BGS: non-neutral mutations are deleterious; (iii) Bateson-Dobzhansky-Muller incompatibility (BDMI): As (ii), except that after the split, a fraction of non-neutral mutations is deleterious in one population and neutral in the other (randomly chosen). (iv) Positive selection: non-neutral mutations have positive fitness effects. (v) BGS and positive selection: (ii) and (iv) combined; (vi) Local adaptation: As (iv), but after the split some non-neutral mutations are beneficial in one population and neutral in the other. Migration between populations was allowed only in scenarios (iii) and (vi). See Materials and methods for more details and S6 Table for parameter values. To summarize the results, measures of variation (π, d_{xy}, and F_{ST}) were calculated in 500-kb regions at 10 time points for each simulation.

This broad range of scenarios generated quite different patterns of variation, both through time and across the chromosome (Fig 7; S13 Fig). As expected, the neutral model did not produce heterogeneous genomic landscapes. BGS produced a diversity landscape characterized by higher diversity in the unconstrained central third of the chromosome (Fig 7; S13 Fig and S14 Fig). Consistent with the predictions of Burri (2017a), π and d_{xy} remained highly correlated over the course of the simulations. However, even with a high proportion of deleterious mutations (10%) and substantial fitness effects (1%), the variation in π and d_{xy} was modest, as was variation in the differentiation landscape (Fig 7).

Under some conditions, the BDM incompatibility model produced patterns of variation that were more similar to our empirical findings than the BGS model (Fig 7, S13 Fig and S14 Fig). In all simulations, patterns of diversity and divergence were relatively homogenous at the time of the split. However, selection against gene flow induced by incompatibility loci on the chromosome ends caused patterns of variation to become highly structured over time. Indeed, higher π, lower d_{xy}, and lower F_{ST} were observed in the neutrally evolving chromosome center due to the homogenizing effects of higher gene flow in this region. As in our empirical data, the correlation between π and d_{xy} decayed over the course of some simulations and even became negative depending on rates of gene flow and selection (S14 Fig). Although we simulated this scenario starting with undifferentiated populations (i.e., primary contact), widespread BDMIs should produce the same patterns for populations that diverged in allopatry and came back into secondary contact. This is because gene flow will have a stronger homogenizing effect in regions of the genome that are not associated with incompatibilities [66].

The 3 models with positive selection (positive, BGS and positive, and local adaptation) all produced highly heterogeneous genomic landscapes across the range of parameter values
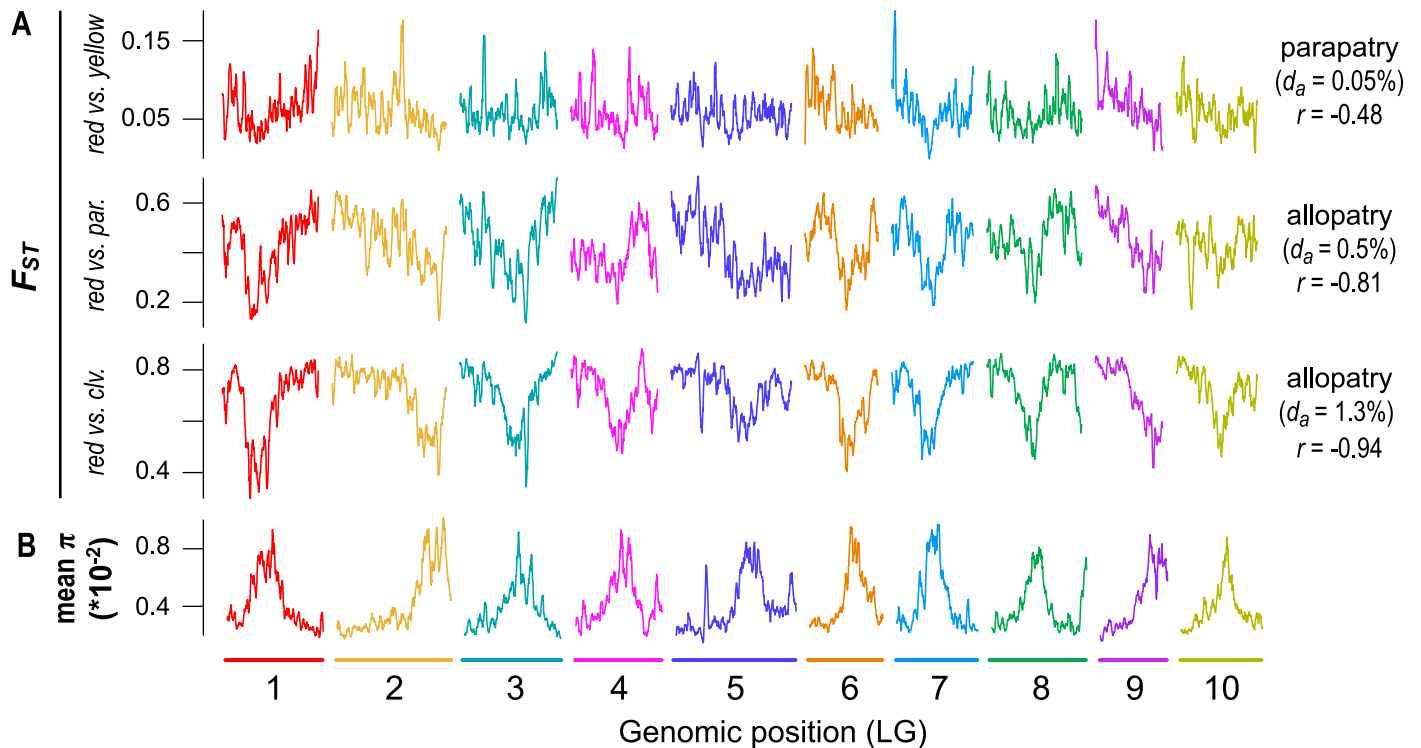
Fig 6. Emergence of a heterogeneous differentiation landscape across 1 million years of divergence. (A) Plots of F_{ST} (500-kb windows) across the genome for pairs of taxa at early (red versus yellow), intermediate (red versus *parviflorus*), and late stages (red versus *M. clevelandii*) of divergence. (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. LG, linkage group.

https://doi.org/10.1371/journal.pbio.3000391.g006

explored (Fig 7; S13 Fig). 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, in which 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 (S13 Fig and S14 Fig). Highly heterogeneous differentiation landscapes, characterized by lower F_{ST} in the chromosome center, emerged rapidly in all 3 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.

Overall, the results of these simulations suggest that BGS is not primarily responsible for heterogeneous genomic landscapes. Although our simulations should be interpreted cautiously, because we have not thoroughly explored parameter space associated with each model, the results of other recent simulation studies also support this conclusion. For example, Matthey-Doret and Whitlock [67] simulated population divergence with BGS under different scenarios with simulations using parameters estimated from humans and stickleback. They found that BGS was unable to generate heterogeneous differentiation landscapes over short timescales, with or without gene flow. Similarly, Rettelbach and colleagues [68] simulated the
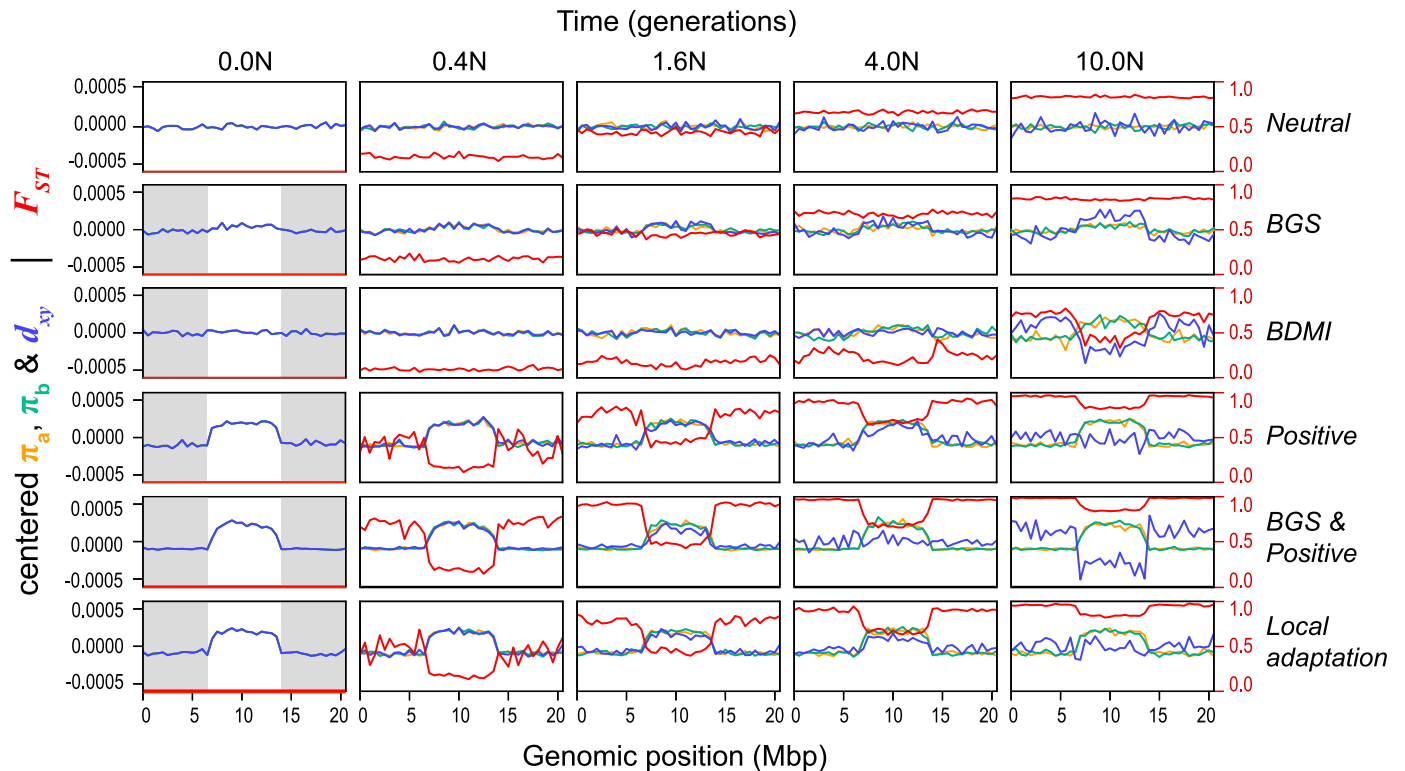
Fig 7. Genomic landscapes simulated under different divergence histories. Each row of plots shows patterns of within- and between-population variation (π, d_{xy}, and F_{ST}) across the chromosome (500-kb windows) at 5 time points (N generations, where N = 10,000) during one of the scenarios The selection 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) BGS (−Ns = 100; −prop = 0.1), (iii) BDMI (−Ns = 100, −prop = 0.05, Nm = 0.1), (iv) positive selection (+Ns = 100, +prop = 0.001), (v) BGS and positive selection (−Ns = 100, −prop = 0.1; +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. Uncentered values and additional simulations with different parameter combinations and more time points can be found in S13 Fig. BDMI, Bateson-Dobzhansky-Muller incompatibility; BGS, background selection.

https://doi.org/10.1371/journal.pbio.3000391.g007

evolution of diversity landscapes using empirical estimates of recombination rate and functional densities from the collared flycatcher genome. They found that BGS was able to generate modest variation in levels of diversity across chromosomes, but they concluded that it was not sufficient to explain the pronounced dips observed in empirical studies of flycatcher genomes (e.g., [30]). Both of these studies suggest that other processes (probably positive selection) are responsible for generating the observed patterns [67, 68].

Our simulations suggest that divergence histories involving positive selection and/or incompatibilities are plausible explanations for heterogeneous genomic landscapes that emerge rapidly after a population split. However, in order to account for the presence of strong correlations between measures of variation and genomic features across multiple comparisons, these scenarios assume that the genomic basis of adaptation and/or reproductive isolation is diffuse, shared across multiple taxa, and evolves rapidly. Although this was once considered unlikely, recent studies suggest that these conditions may be common. For example, adaptation and speciation are now thought to be highly polygenic [37, 39, 69], and recent evidence suggests that adaptation draws heavily on ancestral mutations, rather than those arising independently in multiple, related descendent lineages [35, 70, 71]. Similarly, the interaction between widespread hybrid incompatibilities (intrinsic or extrinsic) and intrinsic genomic

features is thought to have caused similar patterns of genome-wide variation to evolve between different pairs of hybridizing taxa [39].

Given the rapid and extensive trait diversification that has occurred during the bush monkeyflower radiation [40, 45, 72], positive selection has probably played an important role in shaping patterns of genome-wide variation across the group. This includes the striking divergence of floral traits (Fig 1A), which is thought to be due to divergent pollinator-mediated selection across southern California [48–50]. In the taxa that have been studied in detail, crossing experiments have shown that most of these traits are polygenic, and there is evidence that they are targets of selection [43, 45]. Intrinsic genetic incompatibilities may also influence patterns of variation between some of the taxa, but they are unlikely to explain the correlated landscapes that have evolved across the whole group, because most of these taxa show little or no evidence for intrinsic incompatibilities between them [42, 43]. However, local adaptation also can generate a porous isolating barrier [14], so it is likely that gene flow has contributed to genomic landscape evolution among some taxa, particularly those that are geographically proximate and/or hybridize in areas where their distributions overlap.

Radiation-wide evidence for selection against gene flow

To determine if gene flow has contributed to the evolution of these correlated genomic landscapes, we first tested for evidence of introgression among these taxa using D-statistics [73]. Patterson's D measures asymmetry between the numbers of sites with ABBA and BABA patterns (where A and B are ancestral and derived alleles, respectively) across 3 in-group taxa and an out-group (in our case, *M. clevelandii*) that have the relationship (((P1, P2), P3) O). A significant excess of either pattern gives a nonzero value of D, which is taken as evidence that gene flow has occurred between P3 and one of the in-group taxa [73]. Overall, these tests provide evidence that gene flow has occurred during this radiation. All 48 of the appropriate 4-taxon combinations yielded significant nonzero values of Patterson's D ($p < 0.0001$; S7 Table; S15 Fig). The admixture proportions (*f*) for each test indicate that an average of 2.5% of the genome has been transferred between pairs of in-group taxa, but the proportion varies among the tests, ranging from less than 1% to more than 10% (S7 Table).

Although gene flow initially involves the exchange of whole genomes between populations, selection against gene flow reduces effective migration (m_e) in regions of the genome that are associated with barrier loci [2, 69]. Therefore, porous isolating barriers should result in a heterogeneous pattern of introgression across the genome [14, 74]. To test this hypothesis, we calculated the f_d statistic, a version of the admixture proportion modified for application to genomic windows [75]. As predicted, estimates of f_d varied markedly among genomic regions (Fig 2; S16 Fig). Most windows showed levels of f_d at or near zero, indicating that foreign alleles have been purged from these regions by selection. However, a proportion of windows showed substantial admixture proportions. In some tests, values of f_d exceeded 0.5, indicating that more than half of the sites in some windows have been shared between taxa.

Because gene flow opposes differentiation, we would expect to see lower F_{ST} in regions with a higher proportion of introgressed variants. Consistent with this prediction, we observed a substantial negative correlation ($r = -0.57$; $p < 0.0001$) between mean f_d (for each window, averaged over the 48 tests) and PC1 F_{ST} (Fig 3). This result is not driven by a limited number of the 4-taxon tests, because 44 of the 48 comparisons showed the same significant negative relationship (S8 Table). Also consistent with models of widespread selection against gene flow, regions of the genome with higher f_d scores tended to have higher diversity ($r = 0.45$; $p < 0.0001$), lower tree concordance scores ($r = -0.59$; $p < 0.0001$), fewer functional genes ($r = -0.36$; $p < 0.0001$), and a higher recombination rate ($r = -0.18$; $p < 0.0001$; Fig 3).

Overall, these results indicate that widespread gene flow has played a key role in the formation of the genomic landscape in this system. In addition to reductions in diversity and increased differentiation owing to selection against gene flow, the persistence of introgressed variants has probably resulted in higher levels of diversity in regions with fewer genes and higher recombination rate. The 4-taxon tests show that the impact of gene flow is widespread across the radiation, though some caution should be exercised when interpreting the specific pattern of introgression events. Given that there is the potential for gene flow between so many taxa and ancestral lineages, it is difficult to infer the number, source, and timing of admixture events that have occurred during the radiation. For example, rather than reflecting recent gene flow between all pairs of taxa, some introgression events may have occurred deeper in history, and their consequences inherited by multiple taxa. Although more sophisticated methods will be needed to model gene flow across this group, these results clearly show that it has contributed to the rapid evolution of correlated genomic landscapes during this radiation.

Conclusions and implications for understanding genomic landscape evolution and the basis of adaptation and speciation

Facilitated by a new chromosome-level genome assembly, we have shed light on the causes of correlated genomic landscapes across a radiation of monkeyflowers. Adaptive divergence and gene flow are hallmarks of rapid radiations [76], and our data suggest that the indirect effects of selection resulting from these processes have contributed to a common pattern of differentiation among these taxa. Our "time-course" analysis shows that the common landscape emerged rapidly after populations split and has become more correlated with the distribution of genomic features as divergence time has increased. Although BGS may play some role, its effects are probably too subtle to have made a strong contribution to the correlated landscapes during the timeframe of this radiation.

In addition to enhancing our understanding of the processes that have shaped the genomic landscape during this radiation, our study contributes toward a more general understanding of the role that natural selection plays in shaping genome-wide variation. In line with the findings of other recent studies, our results indicate that little, if any, of the genome evolves free of the effects of natural selection [27, 61]. Moreover, our "time-course" analysis shows that between-taxon signatures of selection can emerge very rapidly after a population split and can be substantial, even between populations at the early stages of divergence. Overall, these results suggest that patterns of between-population variation cannot be understood without taking the effects of natural selection into account.

An important point arising from this work is that multiple divergence histories can generate heterogeneous differentiation landscapes that are correlated both among taxa and with the distribution of intrinsic genomic properties. When divergence is recent, possible explanations include polygenic adaptation across multiple populations and porous barriers to gene flow arising from divergent ecological selection and/or intrinsic incompatibilities. Although it has often been assumed that recurrent BGS is the primary cause of correlated landscapes, it is important to remember that all forms of selection can indirectly impact levels of variation at associated sites. In fact, our simulations show that alternative explanations may be more likely when divergence times are short and there is opportunity for gene flow among taxa. Thus, we advocate for a more nuanced approach when interpreting correlated differentiation landscapes, rather than assuming that they are caused by a single evolutionary process.

Finally, our results have important practical implications for studies attempting to identify the genomic basis of adaptation and speciation from patterns of genome-wide variation. Indeed, detecting these loci is a major goal of adaptation and speciation studies, and genome

scans are now commonly used to identify promising candidate regions [2]. In an effort to correct for the potentially confounding effects of BGS, it has been proposed that the correlation in F_{ST} among multiple population pairs can be used as a baseline for outlier detection [29, 62]. 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.

Materials and methods

Genome assembly

We used a combination of short-read Illumina and long-read SMRT sequencing to assemble the genome of a single individual from the red ecotype of *M. aurantiacus* subspecies *puniceus* (population UCSD; S3 Table). Genomic DNA was isolated from leaf tissue using either ZR plant/seed DNA miniprep kits (Zymo Research, Irvine, CA) or GeneJet Plant Genomic DNA purification kits (Thermo Fisher, Waltham, MA). Illumina libraries were generated following the *Allpaths-LG* assembly pipeline [77], which included a single fragment library with average 180 bp insert size and 3 mate-pair libraries (average insert sizes: 3.5–5 kb, 5–7 kb, and 7–13 kb). Libraries were sequenced on the Illumina HiSeq 2500 using paired-end 100 bp reads. An initial scaffold-level assembly was performed with *Allpaths-LG* using default parameters and the *haploidify* function enabled. This assembly yielded 11,123 contigs (N50 = 40.5 kb) and 2,299 scaffolds (N50 = 1,310 kb), for a total assembly size of 193.3 Mbp. Long-read sequencing was performed from the same individual using 12 SMRT cells sequenced on the Pacific Biosystems RS II machine at Duke University. We obtained a total of 6.4 Gbp of sequence, which corresponds to approximately 21× coverage of the genome. The PacBio reads were used to rescaffold the *Allpaths-LG* scaffolds using *Opera-LG* [78]. This reduced the number of scaffolds to 1,547 (N50 = 1,578 kb).

We then manually improved the scaffold containing the flower color gene *MaMyb2* [63]. We first aligned this scaffold to a previously published draft sequence assembly from this same individual [44], which was generated using Illumina short-reads and the *Velvet* assembler [79]. We used long range PCR and cloning to generate Sanger sequences across 3 regions within 20 kb of *MaMyb2* that did not assemble well. Genomic DNA was amplified using Phusion high fidelity polymerase (NEB, Ipswich, MA). PCR products were cloned into the pCR2.1 TOPO-TA vector (Life Technologies, Carlsbad, CA), and purified plasmids were sequenced with Sanger technology. Resulting sequences were aligned to the scaffold containing *MaMyb2*, and new PCR primers were designed to sequence internal fragments until the entire insert was sequenced. Using this approach, we sequenced a total of 9,824 bp across the 3 regions. The reference sequence in the assembly was corrected manually to match the Sanger data.

Finally, we gap filled the assembly using the PacBio data and the program *PBJelly* [80]. Resulting scaffolds were assembled into pseudomolecules using *Chromonomer* (http://catchenlab.life.illinois.edu/chromonomer/), according to the online manual. This software anchored and oriented scaffolds based on the order of markers in a high-density linkage map (see Construction of high-density linkage map) and made corrections to scaffolds when differences occurred between the genetic and physical positions of markers in the map. A final round of gap filling with *PBJelly* was performed to fill any gaps that were created by broken scaffolds in *Chromonomer*. To assess the completeness of the gene space in the assembly, we

used both the BUSCO and CEGMA pipelines to estimate the proportion of 956 single copy plant genes (BUSCO) or 248 core eukaryotic genes (CEGMA) that were completely or partially assembled [52, 81]. The proportion of these genes present in an assembly has been shown to be correlated with the total proportion of assembled gene space and thus serves as a good predictor of assembly completeness.

Construction of high-density linkage map

We generated an outbred F_2 mapping population by crossing 2 F_1 individuals, each the product of crosses between different greenhouse-raised red and yellow ecotype plants collected from one red ecotype and one yellow ecotype population (populations UCSD and LO, respectively; S3 Table). We then used restriction-site associated DNA sequencing (RADseq) to genotype F_1 and F_2 individuals. DNA was extracted from leaf material using Zymo ZR plant/seed DNA miniprep kits, and RAD library preparation followed the protocol outlined in Sobel and Streisfeld [43]. Libraries were sequenced on the Illumina HiSeq 2000 platform using single-end 100 bp reads at the Genomics Core Facility, University of Oregon.

Reads were filtered based on quality, and errors in the barcode sequence or RAD site were corrected using the *process_radtags* script in *Stacks* version 1.35 [82, 83]. Loci were created using the *denovo_map.pl* function of *Stacks*, with 3 identical raw reads required to create a stack, 2 mismatches allowed between loci for an individual, and 2 mismatches allowed when processing the catalog. SNPs were determined and genotypes called using a ML statistical model implemented in *Stacks* and a stringent χ^2 significance level of 0.01 to distinguish between heterozygotes and homozygotes. We then used the *genotypes* program implemented in *Stacks* to identify a set of 9,029 mappable markers. We specified a "CP" cross design (F_1 individuals coded as the parents), requiring that a marker was present in at least 75% of progeny at a minimum depth of 12 reads per individual, and we allowed automated corrections to be made to the data.

Linkage map construction was performed using *Lep-MAP2* [84]. The data were filtered using the *Filtering* module to include only individuals with less than 13% missing data and excluded markers that showed evidence for extreme segregation distortion (χ^2 test, $P < 0.01$). To assign markers to LGs, we used the *SeparateChromosomes* module with a logarithm of odds (LOD) score limit of 20 and no minimum size for LGs. This assigned 7,217 markers to 10 LGs, which matches the number of chromosomes in *M. aurantiacus*. The *JoinSingles* module was executed again with a LOD limit of 10 to join an additional 877 ungrouped markers to the 10 previously formed LGs. Fifty-seven singles that were not joined at this stage were discarded from the data set. Initial marker orders were determined using sex-averaged and sex-specific recombination rates using the *OrderMarkers* module. For each LG, we conducted 10 independent runs using the Kosambi mapping function (*useKosambi = 1*), with the data set split into 7 pseudofamilies to take advantage of parallel processing. When multiple markers had identical genotypes, only the duplicate marker with the least missing data was used in marker ordering. We retained the marker order from the run with the best likelihood. After removing markers with an error rate >0.05, the ML order was re-evaluated 10 times using the *evaluateOrder* flag and polished in 100 SNP windows. The map contained 8,094 informative loci from 269 F2 individuals, with an average of 3.5% ± SD 3.86 missing data per individual.

After the integration of our assembly and genetic map using the *Chromonomer* software [85], we made corrections to the map order based on the physical position of markers within assembled scaffolds. Using the output of *Chromonomer*, we identified markers that were out of order in the map compared to their local assembly order and aligned these markers to the assembly from *Chromonomer* using *Bowtie2* version 2.2.5 [86] with the *very_sensitive* settings

to obtain their physical order. We then re-estimated the map using the *evaluateOrder* flag in *Lep-MAP2* as described above but with the marker order constrained to the physical order (*improveOrder = 0*) and with all duplicate markers included in the analysis (*removeDuplicates = 0*). After initial map construction, we removed 17 markers with an estimated error rate greater than 5% and estimated the map one last time using the same settings. The final map contained 7,589 markers across the 10 LGs.

Genome annotation

Prior to genome annotation, the assembly was soft-masked for repetitive elements and areas of low complexity with *RepeatMasker* (RepeatMasker Open-4.0) using a custom *Mimulus aurantiacus* library created by *RepeatModeler* (RepeatModeler Open-1.0), Repbase repeat libraries [87] and a list of known transposable elements provided by *MAKER* [88]. In total, 30.99% of the genome assembly was masked by *RepeatMasker*. Repetitive elements were annotated with *RepeatModeler*. Hidden Markov Models for gene prediction were generated by *SNAP* [89] and *Augustus* [90] and were trained iteratively to the assembly using *MAKER*, as described by Cantarel and colleagues [91]. Training was performed on the 14.5-Mbp sequence from LG9. Evidence used by *MAKER* for annotation included protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum*, *Solanum tuberosum*, *Daucus carota*, *Vitis vinifera* (all downloaded from EnsemblPlants on 9 August 2016), *Salvia miltiorrhiza* (downloaded from Herbal Medicine Omics Database on 9 August 2016), *Mimulus guttatus* version *2* (downloaded from JGI Genome Portal on 9 August 2016), and all Uniprot/swissprot proteins (downloaded on 18 August 2016) [92–94] (Herbal Medicine Omics Database; Uniprot). We filtered the annotations with *MAKER* to include (1) only evidence-based information that also contained assembled protein support and (2) those ab initio gene predictions that did not overlap with the evidence-based annotations and that contained protein family domains, as detected with InterProScan [95].

Genome resequencing and variant calling

We collected leaf tissue from each taxon (collection locations available in S3 Table and S2 Fig) and extracted DNA from dried tissue using the Zymo Plant/Seed MiniPrep DNA kit following the manufacturer's instructions. We prepared sequencing libraries using the Kapa Biosystems HyperPrep kit, and libraries with an insert size between 400 and 600 bp were sequenced on the Illumina HiSeq 4000 using paired-end 150 bp reads at the Genomics Core Facility, University of Oregon.

We filtered raw reads using the *process_shortreads* script in *Stacks* version 1.46 to remove reads with uncalled bases or poor quality scores. We then aligned the retained reads to the reference assembly using the BWA-MEM algorithm in *BWA* version 0.7.15 [96]. An average of 91.7% of reads aligned (range: 82.6%–96.0%), and the average sequencing depth was 21× (range: 15.16–30.86×). We then marked PCR duplicates using *Picard* (http://broadinstitute.github.io/picard). We performed an initial run of variant calling using the UnifiedGenotyper tool in *GATK* version 3.8 (McKenna and colleagues, 2010) and filtered the data to remove variants with a mapping quality <50, a quality depth <4, and a Fisher Strand score >50. We then used these variants to perform base quality score recalibration for each individual, before performing another run of the UnifiedGenotyper to call final variants. After the second run of variant calling, we removed variants with a mapping quality <40, a quality depth <2, and a Fisher Strand score >60. The final data set contained 13,233,829 SNPs across the 9 taxa. Finally, we ran UnifiedGenotyper with the EMIT_ALL_SITES option to output all variant and invariant genotyped sites.

Phylogenetic analyses and PCAs

We used *RAxML* version 8 to reconstruct the evolutionary relationships among the 9 taxa by concatenating variant sites from across the genome. To investigate patterns of phylogenetic discordance across the genome, we also built trees from windows across the genome. We phased SNPs using *BEAGLE* version 4.1 [97], using a window size of 100 kb and an overlap of 10 kb. We incorporated information on recombination rate from the genetic map and did not impute missing genotypes. After phasing, we used *MVFtools* (https://www.github.com/jbpease/mvftools) to run *RAxML* from 100-kb and 500-kb nonoverlapping windows, with the *M. clevelandii* samples set as the outgroups. We then visualized the window trees in *DensiTree* version 2.01 [98].

To assess concordance between the window-based trees and the whole-genome tree, we converted trees to distance matrices using the *Ape* package in R [99]. We then calculated the Pearson's correlation coefficient between the distance matrix from each window and the whole-genome tree, with a stronger correlation indicating higher concordance with the whole-genome tree. We used one-dimensional autocorrelation analysis to determine whether variation in tree concordance was randomly distributed across the genome. This involved estimating the autocorrelation between genomic position and tree concordance for each LG with a lag size of 2 Mbp. The significance of the observed value for each LG was determined from a null distribution of autocorrelation coefficients estimated from 1,000 random permutations of the genome-wide data.

We also conducted a PCA based on all variant sites from across the entire genome using *Plink* version 1.90 [100]. Initially, we ran the PCA with all 37 samples, but we consecutively reran the analysis by removing different taxa in order to assess clustering patterns among more closely related samples.

Population genomic analyses

To examine how genome-wide patterns of diversity, differentiation, and divergence varied among taxa, we calculated nucleotide diversity (π), between-taxon differentiation (F_{ST}), and between-taxon divergence (d_{xy}), respectively, in nonoverlapping and overlapping 100-kb (step size = 10 kb) and 500-kb (step size = 50 kb) windows using custom Python scripts downloaded from https://github.com/simonhmartin/genomics_general. We calculated measures of differentiation and divergence across all 36 pairwise comparisons among the 9 taxa, and diversity was estimated separately for each taxon. These scripts estimated π and d_{xy} by dividing the number of sequence differences within a window (either within or between taxa) by the total number of sites in that window. To account for missing data, the script counted the number of differences between each sample, divided by the total number of variant sites that were genotyped within those samples, and then averaged across all pairs of samples. To provide an unbiased estimate of diversity and divergence, we incorporated invariant sites into the calculation by dividing the number of pairwise differences (within and between taxa, respectively) by the total number of genotyped sites (variant and invariant) within a window. F_{ST} was calculated following the measure of K_{ST} [101], Eq 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 3 individuals for comparisons within the *M. aurantiacus* complex or at least 2 individuals for each comparison involving *M. clevelandii*.

We summarized the variation in each statistic across comparisons using a PCA, with taxon or taxon pair as the variables. Thus, across each window, PC1 for 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 Phylogenetic analyses and PCAs).

To examine the relationships among PC1 for diversity, differentiation, and divergence, we estimated Pearson's correlation coefficient among all 3 statistics across genomic windows. Further, we estimated correlations among these 3 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, because these represented unrealistically high values of recombination. A minimum of 3 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 2 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 and Coyne and Orr [102, 103] 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} [104, 105]. This value was then converted into years by multiplying by a generation time of 2 years.

Simulations

To assess the plausibility of different scenarios in producing heterogeneous genomic landscapes, we implemented forward-time Wright-Fisher simulations using SLiM [64, 65]. The basic model consisted of an ancestral population with a fixed population size of $N = 10,000$ that split after 10*N* nonoverlapping generations into 2 daughter populations, each with a fixed size of *N*. These populations were then allowed to evolve for a further 10*N* nonoverlapping generations. We simulated a 21 Mbp chromosome with a recombination rate of 1.5×10^{-8} and a mutation rate of 10^{-8}, both per base pair and per generation.

We explored the following 6 modifications of this basic model. (i) Neutral evolution: mutations did not impact fitness; (ii) BGS: mutations in the ancestor and daughter populations were neutral in the middle third of the chromosome but can be deleterious in the chromosome ends; (iii) BDMI: neutral and deleterious mutations occurred in the ancestor. After the split, we allowed migration between the 2 daughter populations. To simulate BDMIs, a fraction of the selected mutations was deleterious in each of the populations and neutral in the other. (iv) Positive selection: mutations in the ancestor and daughter population were neutral in the middle third of the chromosome but could be beneficial in the remaining regions. (v) Local adaptation: mutations in the ancestor and daughter population were neutral in the middle third of the chromosome but could be beneficial in the remainder. After the split, we allowed migration between the daughter populations. To simulate local adaptation, a fraction of the selected mutations was beneficial in each population and neutral in the other. Unlike the BDMI simulation, we allowed local adaptation from standing variation, so some of the mutations that were neutral in the ancestral population became beneficial after the split. BGS and positive

selection: mutations in the ancestor and daughter population were neutral in the middle third of the chromosome but could be either deleterious or beneficial in the tails.

Each scenario was simulated with varying proportions of beneficial/deleterious to neutral mutations, and varying mean selective coefficients and migration rates, where applicable (S6 Table). For each combination of parameters, we ran 5 replicate simulations. As described in Kelleher and colleagues [106], to improve simulation speed we recorded genealogies and non-neutral mutations in a tree sequence and added neutral mutations afterwards with *msprime* [107]. We used *scikit-allel* (version 1.1.8) [108] to calculate π, F_{ST}, and d_{xy} in 500-kb windows. All of the SLiM code used is available on GitHub (https://github.com/mufernando/mimulus_sims).

Tests for genome-wide admixture

We tested for introgression and quantified levels of admixture by calculating Patterson's *D* and the admixture proportion [73, 109]. Patterson's *D* takes 4 taxa with the relationship (((P1, P2), P3) O) and looks for an excess of either ABBA or BABA sites, where A and B represent the ancestral and derived alleles, respectively. We calculated Patterson's *D* for all possible groups of 3 in-group taxa based on the set of relationships inferred from the genome-wide (concatenated) data set. *M. clevelandii* was used as the out-group for all tests. The *D* statistic was calculated from allele frequency data for biallelic sites, with the ancestral allele called as the allele fixed within *M. clevelandii*. Therefore, only sites fixed within *M. clevelandii* were included in the analysis. We converted genotype data into allele frequency data using a Python script downloaded from https://github.com/simonhmartin/genomics. Patterson's *D* was calculated from the proportion of ABBA and BABA sites in R using a custom script. To evaluate the significance of each test, we performed block jack-knifing with a block size of 500 kb. We then calculated the genome-wide admixture proportion, *f*, for all tests with a significant value of Patterson's *D*. This test compares the excess of ABBA sites shared between P2 and P3 to the expected excess of ABBA sites between 2 completely admixed populations. The expected excess is calculated by splitting the samples from P3 into 2 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 [75]. Thus, we calculated the 4-taxon statistic f_d, which is a modified version of the admixture proportion developed for estimating local admixture [75]. We calculated f_d in 500-kb nonoverlapping 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 in which *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 mean value of f_d within each window, as well as the maximum value of f_d across the 48 different tests in each window, and estimated the Pearson's correlation coefficient between these values and other statistics.

Supporting information

S1 Table. 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 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. SDs are given in parentheses. LG, linkage group.
(DOCX)

S2 Table. 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 = 1,440).
(DOCX)

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

S4 Table. Loadings for PC1 calculated across all 36 pairwise comparisons (for F_{ST} and d_{xy}) or all 9 taxa (for π). PC1, first principal component.
(DOCX)

S5 Table. Details for the linear regressions presented in Fig 5 of the main text.
(DOCX)

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

S7 Table. 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 4-taxon tests, with *M. clevelandii* as the out-group for each test. All values of Patterson's D are statistically significant ($P < 0.0001$) based on a block jackknife approach.
(DOCX)

S8 Table. 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 500-kb nonoverlapping windows. *M. clevelandii* is the out-group for each test. PC1, first principal component.
(DOCX)

S1 Fig. Map distance (cM/Mbp) versus physical distance across the 10 LGs. Recombination for each marker is estimated relative to the start of the LG and plotted at its physical location on each chromosome in the reference assembly. LG, linkage group.
(EPS)

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

S3 Fig. Genome-wide 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. PCA, principal components analysis.
(EPS)

S4 Fig. 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. LG, linkage group; PC1, first principal component.
(EPS)

S5 Fig. Patterns of variation plotted across each bush monkeyflower LG. Z-transformed FST, *dxy*, 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 9 taxa (π), and the blue line is the Z-transformed score for PC1. Estimates of tree concordance, gene count, and recombination rate (cM/Mbp) are also shown. LG, linkage group; PC1, first principal component.
(PDF)

S6 Fig. Patterns of variation are nonrandomly 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 1,000 random permutations of the genome-wide values. The observed data are significant at $p = 0.001$ unless stated otherwise. LG, linkage group.
(EPS)

S7 Fig. 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.*
(EPS)

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

S9 Fig. 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.
(EPS)

S10 Fig. 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 PC1 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. PC1, first principal component.
(EPS)

S11 Fig. 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 and 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 8 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.
(EPS)

S12 Fig. 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}).
(EPS)

S13 Fig. 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 a 21-Mb chromosome (500-kb windows) at 10 time points (in N generations, where $N = 10,000$) for one parameter combination of 6 scenarios: neutral divergence, BGS, BDMIs, positive selection, BGS and positive selection, and local adaptation. The gray 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 and d_{xy}. The parameter Ns modulates the average selective coefficient (where s = Ns/N), whereas Prop is the proportion of new mutations that are not neutral. Nm is the average number of migrants per generation. BDMI, Bateson-Dobzhansky-Muller incompatibility; BGS, background selection.
(PDF)

S14 Fig. Correlations between measures of variation change over the course of the simulations. Each row of plots shows how the correlations between d_{xy} and π (left plot) and F_{ST} and π (right plot) change over 10 time points (in N generations, where $N = 10,000$) in one of 6 scenarios: neutral divergence, BGS, BDMIs, positive selection, BGS and positive selection, and local adaptation. The red line in each plot shows the correlation for the simulation shown in Fig 7 of the main text. The gray lines show how correlations change for different combinations of parameter values in each scenario. The parameter Ns modulates the average selective coefficient (where s = Ns/N), whereas Prop is the proportion of new mutations that are not neutral. Nm is the average number of migrants per generation. Correlations represent the mean values estimated from 5 replicate simulations of each set of parameter values. BDMI, Bateson-Dobzhansky-Muller incompatibility; BGS, background selection.
(PDF)

S15 Fig. 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 4-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. D is always positive, meaning gene flow occurs between P2 and P3.
(EPS)

S16 Fig. Variation in levels of admixture across the genome. The gray lines show measures of a modified test of the admixture proportion, f_d, estimated in 500-kb nonoverlapping windows for 48 different 4-taxon comparisons, plotted across the 10 LGs 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. LG, linkage group.
(EPS)

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 also thank people who contributed to insightful discussions at the 2017 Speciation Genomics Meeting, Cambridge, UK, the 2019 Gordon Speciation Conference, Ventura, CA, US, and on Twitter.

Author Contributions

Conceptualization: Sean Stankowski, Madeline A. Chase, Peter L. Ralph, Matthew A. Streisfeld.

Formal analysis: Sean Stankowski, Madeline A. Chase, Allison M. Fuiten, Murillo F. Rodrigues, Matthew A. Streisfeld.

Methodology: Murillo F. Rodrigues.

Supervision: Matthew A. Streisfeld.

Visualization: Murillo F. Rodrigues.

Writing – original draft: Sean Stankowski, Madeline A. Chase.

Writing – review & editing: Sean Stankowski, Madeline A. Chase, Peter L. Ralph, Matthew A. Streisfeld.

References

1. Wolf JB, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. Nature reviews Genetics. 2017; 18(2):87–100. Epub 2016/11/15. https://doi.org/10.1038/nrg.2016.133 PMID: 27840429.

2. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovic M, et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J Evolution Biol. 2017; 30(8):1450–77. https://doi.org/10.1111/jeb.13047 ISI:000407280900001. PMID: 28786193

3. Campbell CR, Poelstra JW, Yoder AD. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biol J Linn Soc. 2018; 124(4):561–83. https://doi.org/10.1093/biolinnean/bly063 ISI:000442936200001.

4. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. J Evolution Biol. 2013; 26(2):229–46. https://doi.org/10.1111/j.1420-9101.2012.02599.x ISI:000313747600001. PMID: 23323997

5. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science. 2015; 350(6267):1493–8. https://doi.org/10.1126/science.aac9927 ISI:000366591100050. PMID: 26680190

6. Turner TL, Hahn MW, Nuzhdin SV. Genomic islands of speciation in Anopheles gambiae. PLoS Biol. 2005; 3(9):1572–8. ARTN e285 https://doi.org/10.1371/journal.pbio.0030285 ISI:000231820900008. PMID: 16076241

7. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS Genet. 2010; 6(2). ARTN e1000862 https://doi.org/10.1371/journal.pgen.1000862 ISI:000275262700012. PMID: 20195501

8. Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, et al. The genomic landscape of species divergence in Ficedula flycatchers. Nature. 2012; 491(7426):756–60. https://doi.org/10.1038/nature11584 ISI:000311606000045. PMID: 23103876

9. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, et al. Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. Science. 2014; 344(6185):738–42. https://doi.org/10.1126/science.1252136 ISI:000335912900039. PMID: 24833390

10. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Res. 2013; 23(11):1817–28. https://doi.org/10.1101/gr.159426.113 ISI:000326642500006. PMID: 24045163

11. Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science. 2014; 344(6190):1410–4. https://doi.org/10.1126/science.1253226 ISI:000337531700043. PMID: 24948738

12. Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, et al. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat Commun. 2013; 4. Artn 1827 https://doi.org/10.1038/ncomms2833 ISI:000320589900023. PMID: 23652015

13. Lamichhaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015; 518(7539):371–5. Epub 2015/02/18. https://doi.org/10.1038/nature14181 PMID: 25686609.

14. Wu CI. The genic view of the process of speciation. J Evolution Biol. 2001; 14(6):851–65. https://doi.org/10.1046/j.1420-9101.2001.00335.x ISI:000172905600001.

15. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. Trends Genet. 2012; 28 (7):342–50. https://doi.org/10.1016/j.tig.2012.03.009 ISI:000306252300006. PMID: 22520730

16. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. Mol Ecol. 2009; 18(3):375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x ISI:000262517400001. PMID: 19143936

17. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 2014; 23(13):3133–57. Epub 2014/05/23. https://doi.org/10.1111/mec.12796 PMID: 24845075.

18. Maynard-Smith J, Haigh J. Hitch-Hiking Effect of a Favorable Gene. Genet Res. 1974; 23(1):23–35. https://doi.org/10.1017/S0016672300014634 ISI:A1974T216900003. PMID: 4407212

19. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993; 134(4):1289–303. Epub 1993/08/01. PMID: 8375663; PubMed Central PMCID: PMC1205596.

20. Charlesworth B. Measures of divergence between populations and the effect of forces that reduce variability. Mol Biol Evol. 1998; 15(5):538–43. https://doi.org/10.1093/oxfordjournals.molbev.a025953 ISI:000073343800007. PMID: 9580982

21. Hudson RR, Kaplan NL. Deleterious Background Selection with Recombination. Genetics. 1995; 141 (4):1605–17. ISI:A1995TH41900031. PMID: 8601498

22. Gillespie JH. Genetic drift in an infinite population: The pseudohitchhiking model. Genetics. 2000; 155 (2):909–19. ISI:000087475100036. PMID: 10835409

23. Coop G, Ralph P. Patterns of neutral diversity under general models of selective sweeps. Genetics. 2012; 192(1):205–24. Epub 2012/06/21. https://doi.org/10.1534/genetics.112.141861 PMID: 22714413; PubMed Central PMCID: PMC3430537.

24. Baird SJE. Exploring linkage disequilibrium. Molecular ecology resources. 2015; 15(5):1017–9. https://doi.org/10.1111/1755-0998.12424 ISI:000359631600001. PMID: 26261040

25. Kimura M. Evolutionary rate at the molecular level. Nature. 1968; 217(5129):624–6. Epub 1968/02/17. https://doi.org/10.1038/217624a0 PMID: 5637732.

26. Ohta T. Slightly Deleterious Mutant Substitutions in Evolution. Nature. 1973; 246(5428):96–8. https://doi.org/10.1038/246096a0 ISI:A1973R229300031. PMID: 4585855

27. Hahn MW. Toward a selection theory of molecular evolution. Evolution. 2008; 62(2):255–65. https://doi.org/10.1111/j.1558-5646.2007.00308.x ISI:000253373900001. PMID: 18302709

28. Burri R. Dissecting differentiation landscapes: a linked selection's perspective. J Evolution Biol. 2017; 30(8):1501–5. https://doi.org/10.1111/jeb.13108 ISI:000407280900008. PMID: 28786187

29. Burri R. Interpreting differentiation landscapes in the light of long-term linked selection. Evolution Letters. 2017; 1:118–31.

30. Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 2015; 25(11):1656–65. https://doi.org/10.1101/gr.196485.115 ISI:000364355600007. PMID: 26355005

31. Kronforst MR, Hansen MEB, Crawford NG, Gallant JR, Zhang W, Kulathinal RJ, et al. Hybridization Reveals the Evolving Genomic Architecture of Speciation. Cell Rep. 2013; 5(3):666–77. https://doi.org/10.1016/j.celrep.2013.09.042 ISI:000328263400013. PMID: 24183670

32. Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. 2017; 27(6):1004–15. Epub 2017/04/27. https://doi.org/10.1101/gr.212522.116 PMID: 28442558; PubMed Central PMCID: PMC5453315.

33. Delmore KE, Ramos JSL, Van Doren BM, Lundberg M, Bensch S, Irwin DE, et al. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. Evolution Letters. 2018; 2(2):76–87. https://doi.org/10.1002/evl3.46 ISI:000446764700003. PMID: 30283666

34. Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M. Correlated patterns of genetic diversity and differentiation across an avian family. Mol Ecol. 2017; 26(15):3982–97. https://doi.org/10.1111/mec.14083 ISI:000406361200012. PMID: 28256062

35. Barrett RDH, Rogers SM, Schluter D. Natural selection on a major armor gene in threespine stickleback. Science. 2008; 322(5899):255–7. https://doi.org/10.1126/science.1159978 ISI:000259902300047. PMID: 18755942

36. Bassham S, Catchen J, Lescak E, von Hippel FA, Cresko WA. Repeated Selection of Alternatively Adapted Haplotypes Creates Sweeping Genomic Remodeling in Stickleback. Genetics. 2018; 209 (3):921–39. https://doi.org/10.1534/genetics.117.300610 ISI:000437171700021. PMID: 29794240

37. Rockman MV. The Qtn Program and the Alleles That Matter for Evolution: All That's Gold Does Not Glitter. Evolution. 2012; 66(1):1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x ISI:000298739100001. PMID: 22220860

38. Schumer M, Xu CL, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018; 360(6389):656–9. https://doi.org/10.1126/science.aar3684 ISI:000431790900046. PMID: 29674434

39. Martin SH, Davey JW, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol. 2019; 17(2). ARTN e2006288 https://doi.org/10.1371/journal.pbio.2006288 ISI:000460317100003. PMID: 30730876

40. Chase MA, Stankowski S, Streisfeld MA. Genomewide variation provides insight into evolutionary relationships in a monkeyflower species complex (Mimulus sect. Diplacus). Am J Bot. 2017; 104 (10):1510–21. https://doi.org/10.3732/ajb.1700234 ISI:000413976200014. PMID: 29885225

41. Thompson DM. Systematics of Mimulus subgenus Schizoplacus (Scrophulariaceae). Systematic Botany Monographs. 2005; 75:1–213.

42. McMinn HE. Studies in the genus Diplacus, Scrophulariaceae. Madrono. 1951; 11:33–128.

43. Sobel JM, Streisfeld MA. Strong premating reproductive isolation drives incipient speciation in Mimulus aurantiacus. Evolution. 2015; 69(2):447–61. https://doi.org/10.1111/evo.12589 ISI:000348916200013. PMID: 25545789

44. Stankowski S, Sobel JM, Streisfeld MA. Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. Mol Ecol. 2017; 26 (1):107–22. https://doi.org/10.1111/mec.13645 ISI:000391940600008. PMID: 27065228

45. Stankowski S, Sobel JM, Streisfeld MA. The geography of divergence with gene flow facilitates multi-trait adaptation and the evolution of pollinator isolation in Mimulus aurantiacus. Evolution. 2015; 69 (12):3054–68. https://doi.org/10.1111/evo.12807 ISI:000368249000003. PMID: 26514872

46. Streisfeld MA, Kohn JR. Environment and pollinator-mediated selection on parapatric floral races of Mimulus aurantiacus. J Evolution Biol. 2007; 20(1):122–32. https://doi.org/10.1111/j.1420-9101.2006.01216.x ISI:000242904600020. PMID: 17210005

47. Streisfeld MA, Kohn JR. Contrasting patterns of floral and molecular variation across a cline in Mimulus aurantiacus. Evolution. 2005; 59(12):2548–59. ISI:000234502400006. PMID: 16526503

48. Grant V. Origin of floral isolation between ornithophilous and sphingophilous plant species. Proceedings of the National Academy of Sciences of the United States of America. 1993; 90(16):7729–33.

Epub 1993/08/15. https://doi.org/10.1073/pnas.90.16.7729 PMID: 11607421; PubMed Central PMCID: PMC47216.

49. Grant V. Effects of hybridization and selection on floral isolation. Proceedings of the National Academy of Sciences of the United States of America. 1993; 90(3):990–3. Epub 1993/02/01. https://doi.org/10.1073/pnas.90.3.990 PMID: 11607361; PubMed Central PMCID: PMC45796.

50. Grant V. Plant Speciation. 2 ed. New York: Columbia University Press; 1981.

51. Vickery RK. Speciation by Aneuploidy and Polyploidy in Mimulus (Scrophulariaceae). Great Basin Nat. 1995; 55(2):174–6. ISI:A1995RH39300011.

52. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31 (19):3210–2. https://doi.org/10.1093/bioinformatics/btv351 ISI:000362845400018. PMID: 26059717

53. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 ISI:000336095100024. PMID: 24451623

54. Stankowski S, Streisfeld MA. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. P Roy Soc B-Biol Sci. 2015; 282(1814):154–62. Artn 20151666 https://doi.org/10.1098/Rspb.2015.1666 ISI:000363361800019.

55. Pease JB, Haak DC, Hahn MW, Moyle LC. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLoS Biol. 2016; 14(2). https://doi.org/10.1371/journal.pbio.1002379 ARTN e1002379. ISI:000371883600011.

56. Hudson RR. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology. 1990; 7:44.

57. Nei M, Li WH. Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. Proceedings of the National Academy of Sciences of the United States of America. 1979; 76 (10):5269–73. https://doi.org/10.1073/pnas.76.10.5269 ISI:A1979HR46000110. PMID: 291943

58. Slatkin M. Inbreeding Coefficients and Coalescence Times. Genetical research. 1991; 58(2):167–75. https://doi.org/10.1017/S0016672300029827 ISI:A1991GQ06500010. PMID: 1765264

59. Pease JB, Hahn MW. More Accurate Phylogenies Inferred from Low-Recombination Regions in the Presence of Incomplete Lineage Sorting. Evolution. 2013; 67(8):2376–84. https://doi.org/10.1111/evo.12118 ISI:000322329500017. PMID: 23888858

60. Langley CH, Stevens K, Cardeno C, Lee YC, Schrider DR, Pool JE, et al. Genomic variation in natural populations of Drosophila melanogaster. Genetics. 2012; 192(2):533–98. Epub 2012/06/08. https://doi.org/10.1534/genetics.112.142018 PMID: 22673804; PubMed Central PMCID: PMC3454882.

61. Corbett-Detig RB, Hartl DL, Sackton TB. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol. 2015; 13(4):e1002112. Epub 2015/04/11. https://doi.org/10.1371/journal.pbio.1002112 PMID: 25859758; PubMed Central PMCID: PMC4393120.

62. Berner D, Salzburger W. The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. 2015; 31(9):491–9. https://doi.org/10.1016/j.tig.2015.07.002 ISI:000361254100002. PMID: 26259669

63. Streisfeld MA, Young WN, Sobel JM. Divergent Selection Drives Genetic Differentiation in an R2R3-MYB Transcription Factor That Contributes to Incipient Speciation in Mimulus aurantiacus. PLoS Genet. 2013; 9(3). ARTN e1003385 https://doi.org/10.1371/journal.pgen.1003385 ISI:000316866700058. PMID: 23555295

64. Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Molecular ecology resources. 2019; 19 (2):552–66. https://doi.org/10.1111/1755-0998.12968 ISI:000459815200020. PMID: 30565882

65. Haller BC, Messer PW. SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. Mol Biol Evol. 2019; 36(3):632–7. https://doi.org/10.1093/molbev/msy228 ISI:000462585100017. PMID: 30517680

66. Durrett R, Buttel L, Harrison R. Spatial models for hybrid zones. Heredity. 2000; 84 (Pt 1):9–19. Epub 2000/02/26. https://doi.org/10.1046/j.1365-2540.2000.00566.x PMID: 10692006.

67. Matthey-Doret R, Whitlock MC. Background selection and the statistics of population differentiation: consequences for detecting local adaptation. BioRxiv. 2018. https://doi.org/10.1101/326256

68. Rettelbach A, Nater A, Ellegren H. How Linked Selection Shapes the Diversity Landscape in Ficedula Flycatchers. Genetics. 2019; 212(1):277–85. Epub 2019/03/16. https://doi.org/10.1534/genetics.119.301991 PMID: 30872320.

69. Jiggins CD, Martin SH. Glittering gold and the quest for Isla de Muerta. J Evolution Biol. 2017; 30 (8):1509–11. https://doi.org/10.1111/jeb.13110 ISI:000407280900010. PMID: 28786188

70. Nelson TC, Cresko WA. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evolution Letters. 2018; 2(1):9–21. https://doi.org/10.1002/evl3.37 ISI:000446762700002. PMID: 30283661

71. Marques DA, Meier JI, Seehausen O. A Combinatorial View on Speciation and Adaptive Radiation. Trends in ecology & evolution. 2019. Epub 2019/03/20. https://doi.org/10.1016/j.tree.2019.02.008 PMID: 30885412.

72. Sobel JM, Stankowski S, Streisfeld MA. Variation in ecophysiological traits might contribute to ecogeographic isolation and divergence between parapatric ecotypes of Mimulus aurantiacus. J Evol Biol. 2019. Epub 2019/03/19. https://doi.org/10.1111/jeb.13442 PMID: 30883990.

73. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010; 328(5979):710–22. Epub 2010/05/08. https://doi.org/10.1126/science.1188021 PMID: 20448178; PubMed Central PMCID: PMC5100745.

74. Barton NH, Hewitt GM. Analysis of Hybrid Zones. Annu Rev Ecol Syst. 1985; 16:113–48. https://doi.org/10.1146/annurev.es.16.110185.000553 ISI:A1985AUL3900006.

75. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol Biol Evol. 2015; 32(1):244–57. Epub 2014/09/24. https://doi.org/10.1093/molbev/msu269 PMID: 25246699; PubMed Central PMCID: PMC4271521.

76. Schluter D. The Ecology of Adaptive Radiation. Oxford: Oxford University Press; 2000.

77. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(4):1513–8. Epub 2010/12/29. https://doi.org/10.1073/pnas.1017351108 PMID: 21187386; PubMed Central PMCID: PMC3029755.

78. Gao S, Bertrand D, Chia BK, Nagarajan N. OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. Genome biology. 2016; 17:102. Epub 2016/05/14. https://doi.org/10.1186/s13059-016-0951-y PMID: 27169502; PubMed Central PMCID: PMC4864936.

79. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18(5):821–9. https://doi.org/10.1101/gr.074492.107 ISI:000255504600014. PMID: 18349386

80. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS ONE. 2012; 7(11):e47768. Epub 2012/11/28. https://doi.org/10.1371/journal.pone.0047768 PMID: 23185243; PubMed Central PMCID: PMC3504050.

81. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genornes. Bioinformatics. 2007; 23(9):1061–7. https://doi.org/10.1093/bioinformatics/btm071 ISI:000246773300003. PMID: 17332020

82. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013; 22(11):3124–40. Epub 2013/05/25. https://doi.org/10.1111/mec.12354 PMID: 23701397; PubMed Central PMCID: PMC3936987.

83. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3-Genes Genom Genet. 2011; 1(3):171–82. https://doi.org/10.1534/g3.111.000240 ISI:000312406900001. PMID: 22384329

84. Rastas P, Calboli FCF, Guo BC, Shikano T, Merila J. Construction of Ultradense Linkage Maps with Lep-MAP2: Stickleback F-2 Recombinant Crosses as an Example. Genome Biol Evol. 2016; 8(1):78–93. https://doi.org/10.1093/gbe/evv250 ISI:000370971600007. PMID: 26668116

85. Amores A, Catchen J, Nanda I, Warren W, Walter R, Schartl M, et al. A RAD-tag genetic map for the platyfish (Xiphophorus maculatus) reveals mechanisms of karyotype evolution among teleost fish. Genetics. 2014; 197(2):625–41. Epub 2014/04/05. https://doi.org/10.1534/genetics.114.164293 PMID: 24700104; PubMed Central PMCID: PMC4063920.

86. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9 (4):357–9. Epub 2012/03/06. https://doi.org/10.1038/nmeth.1923 PMID: 22388286; PubMed Central PMCID: PMC3322381.

87. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and genome research. 2005; 110(1–4):462–7. Epub 2005/08/12. https://doi.org/10.1159/000084979 PMID: 16093699.

88. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics. 2011; 12:491. Epub 2011/12/24. https://doi.org/10.1186/1471-2105-12-491 PMID: 22192575; PubMed Central PMCID: PMC3280279.

89. Korf I. Gene finding in novel genomes. BMC bioinformatics. 2004; 5:59. Epub 2004/05/18. https://doi.org/10.1186/1471-2105-5-59 PMID: 15144565; PubMed Central PMCID: PMC421630.

90. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003; 19:Ii215–Ii25. https://doi.org/10.1093/bioinformatics/btg1080 ISI:000207434300026. PMID: 14534192

91. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008; 18(1):188–96. https://doi.org/10.1101/gr.6743907 ISI:000251965300019. PMID: 18025269

92. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl Genomes 2016: more genomes, more complexity. Nucleic acids research. 2016; 44(D1):D574–80. Epub 2015/11/19. https://doi.org/10.1093/nar/gkv1209 PMID: 26578574; PubMed Central PMCID: PMC4702859.

93. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic acids research. 2012; 40(Database issue):D1178–86. Epub 2011/11/24. https://doi.org/10.1093/nar/gkr944 PMID: 22110026; PubMed Central PMCID: PMC3245001.

94. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic acids research. 2013; 42:D26–D31. https://doi.org/10.1093/nar/gkt1069 PMID: 24225321

95. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic acids research. 2005; 33:W116–W20. https://doi.org/10.1093/nar/gki442 ISI:000230271400020. PMID: 15980438

96. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;1303.3997v1.

97. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81(5):1084–97. https://doi.org/10.1086/521987 ISI:000250480900018. PMID: 17924348

98. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010; 26 (10):1372–3. https://doi.org/10.1093/bioinformatics/btq110 ISI:000277447500016. PMID: 20228129

99. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20(2):289–90. https://doi.org/10.1093/bioinformatics/btg412 ISI:000188389700026. PMID: 14734327

100. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4. Artn 7 https://doi.org/10.1186/s13742-015-0047-8 ISI:000365662000001. PMID: 25722852

101. Hudson RR, Boos DD, Kaplan NL. A Statistical Test for Detecting Geographic Subdivision. Mol Biol Evol. 1992; 9(1):138–51. ISI:A1992GX79400010. https://doi.org/10.1093/oxfordjournals.molbev.a040703 PMID: 1552836

102. Felsenstein J. Phylogenies and the comparative method. American Naturalist. 1985; 125:1–15.

103. Coyne JA, Orr HA. Patterns of Speciation in Drosophila. Evolution. 1989; 43(2):362–81. https://doi.org/10.1111/j.1558-5646.1989.tb04233.x ISI:A1989T799700009. PMID: 28568554

104. Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. Speciation and Introgression between Mimulus nasutus and Mimulus guttatus. PLoS Genet. 2014; 10(6). ARTN e1004410 https://doi.org/10.1371/journal.pgen.1004410 ISI:000338847700028. PMID: 24967630

105. Koch M, Haubold B, Mitchell-Olds T. Molecular systematics of Brassicaceae: evidence from plastidic matK and nuclear Chs sequences. Am J Bot. 2001; 88:534–44. PMID: 11250830

106. Kelleher J, Thornton KR, Ashander J, Ralph PL. Efficient pedigree recording for fast population genetics simulation. PLoS Comput Biol. 2018; 14(11). ARTN e1006581 https://doi.org/10.1371/journal.pcbi.1006581 ISI:000451835900035. PMID: 30383757

107. Kelleher J, Etheridge AM, McVean G. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. Plos Comput Biol. 2016; 12(5). ARTN e1004842 https://doi.org/10.1371/journal.pcbi.1004842 ISI:000379348100005. PMID: 27145223

108. Miles A, Harding N. cggh/scikit-allel: v1.1.8. 2017. https://doi.org/10.5281/zenodo.822784

109. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011; 28(8):2239–52. Epub 2011/02/18. https://doi.org/10.1093/molbev/msr048 PMID: 21325092; PubMed Central PMCID: PMC3144383.