The Genomic Signal of Partial Sweeps in *Mimulus guttatus*

John K. Kelly*, Boryana Koseva, and Julius P. Mojica[1]

Department of Ecology and Evolutionary Biology, University of Kansas

[1]Present address: Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO.

*Corresponding author: E-mail: jkk@ku.edu.

Abstract

The molecular signature of selection depends strongly on whether new mutations are immediately favorable and sweep to fixation (hard sweeps) as opposed to when selection acts on segregating variation (soft sweeps). The prediction of reduced sequence variation around selected polymorphisms is much stronger for hard than soft sweeps, particularly when considering quantitative traits where sweeps are likely to be incomplete. Here, we directly investigate the genomic signal of soft sweeps within an artificial selection experiment on *Mimulus guttatus*. We first develop a statistical method based on Fisher's angular transformation of allele frequencies to identify selected loci. Application of this method identifies about 400 significant windows, but no fixed differences between phenotypically divergent populations. With two notable exceptions, we find a modest average effect of partial sweeps on the amount of molecular variation. The first exception is a polymorphic inversion on chromosome 6. The increase of the derived haplotype has a broad genomic effect due to recombination suppression coupled with substantial initial haplotype structure within the population. Second, we found significant increases in nucleotide variation around selected loci in the population evolving larger flowers. This suggests that "high" alleles for flower size were initially less frequent than "low" alleles. This result is consistent with prior studies of *M. guttatus* and illustrates how molecular evolution can depend on the allele frequency spectrum at quantitative trait loci.

Key words: soft sweeps, *Mimulus*, Illumina, partial sweeps.

Introduction

The most immediate purpose of population genetics is to predict allele frequency change. Throughout the 20th century, theorists developed increasingly detailed models to predict the simultaneous action of mutation, selection, migration, and genetic drift under various demographic and ecological scenarios. For most of this interval, there was little relevant data to evaluate the theory (Lewontin 1974). Direct observations of allele frequency change were largely limited to visible polymorphisms with a simple genetic basis (Ford 1971) or major chromosomal changes (Dobzhansky and Levene 1948). However, sequencing technologies now afford an unprecedented view of genetic change within populations.

Second-generation sequencing has enabled "population genomics," the characterization of whole-genome divergence among both natural (e.g. Emerson et al. 2010; Hohenlohe et al. 2010a) and experimental populations (e.g. Johansson et al. 2010; Turner et al. 2011). The method of population genomics typically involves two stages (Hohenlohe et al. 2010b; also see Pavlidis et al. 2008; Oleksyk et al. 2010). First, one estimates neutral demographic processes from a genome-wide integration of signal. Second, one identifies genomic outliers to be considered putative targets of selection. For example, polymorphisms that exhibit the greatest divergence in allele frequency among populations (F_{ST} outliers) may represent loci subject to divergent selection (Lewontin and Krakauer 1973; Beaumont and Nichols 1996). While polymorphism or "SNP-specific" analyses are intuitive, there are both population genetic and statistical reasons to seek evidence of selection within genomic windows (sets of closely linked polymorphisms). The population genetic reason is hitch-hiking. Selection on a polymorphism will cause allele frequency change at closely linked neutral sites (Maynard Smith and Haigh 1974). The statistical justification is that allele frequencies are estimated with error. As a consequence, aggregating information across polymorphisms within a window of sites can provide a more accurate characterization of divergence.

The hitch-hiking effect is contingent on the nature of genetic variation and how selection acts on that variation. The simplest case is a so-called "hard sweep" where a new mutation is immediately favorable and its increase in

frequency carries a single haplotype (a window of sites around the selected polymorphism) to fixation in the population (Maynard Smith and Haigh 1974). As a consequence, variation is eliminated around the selected site. The genomic extent of this effect depends on the rate of recombination, and also on the strength of selection, which determines how rapidly the favorable mutation fixes. In contrast, "soft sweeps" occur when extant alleles become favorable, say due to change in environmental conditions, and then increase in frequency. Here, the effect on linked neutral variation is more subtle because the favored allele occurs in a diversity of genetic backgrounds when it begins to increase in frequency (Innan and Kim 2004; Hermisson and Pennings 2005; Pennings and Hermisson 2006; Chevin and Hospital 2008).

Hitch-hiking may be most difficult to detect when considering selection on quantitative traits. With multi-locus inheritance, large changes in traits means can be generated through incremental changes in allele frequency at many loci. As a consequence, sweeps are likely to be both soft and incomplete. Interestingly, this kind of selection can actually increase nucleotide diversity within a genomic region. If the less common allele at a Quantitative Trait Locus (QTL) is favored, allele frequencies at this site will become more intermediate due to selection. If allele frequencies at neighboring sites also become more intermediate, then the level of sequence variation will be inflated for the entire genomic window. A simple calculation (Appendix A) indicates that this will occur if the uncommon favored allele tends to be positively associated with the less common base at neighboring single-nucleotide polymorphisms (SNPs).

This article investigates genomic variation in populations subject to divergent selection on a quantitative trait. Variation in flower size of *Mimulus guttatus* is highly polygenic (Kelly 2008; Lee 2009). Two independent populations (Low and High) were derived from a common source, a large and highly variable ancestral population, and subject to divergent artificial selection on flower size for nine generations. After nine generations, Low and High are largely non-overlapping in flower size (fig. 1) and substantially different in numerous morphological and life history traits (see table 2 of Kelly 2008). As in previous studies of fruit flies (Burke et al. 2010; Turner et al. 2011; Remolina et al. 2012), cattle (Hayes et al. 2009), and chickens (Johansson et al. 2010), we here investigate genomic divergence among artificially selected populations. However, the present study differs from previous experiments importantly in time scale; considering nine generations of evolution instead of hundreds (e.g. Burke et al. 2010). Divergence of our High and Low populations must be overwhelmingly due to the recruitment of standing genetic variation in flower size. Selective sweeps, if evident at all, must be soft. We also expect that most of the change in phenotype will be due to loci that have not yet fixed the favorable allele.
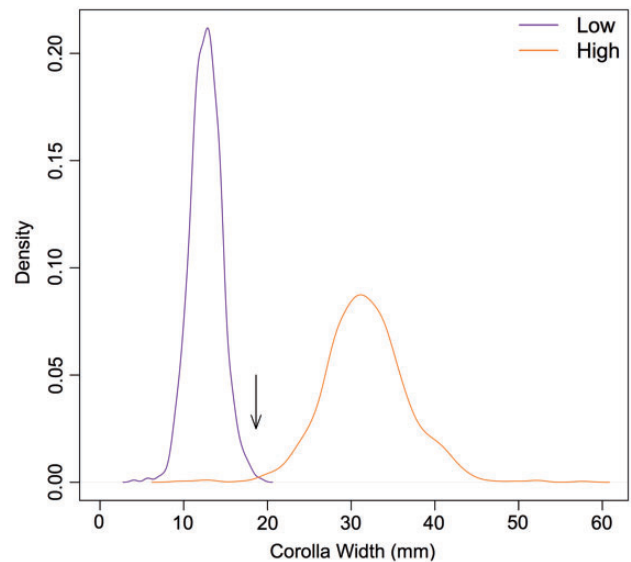

Fɪɢ. 1.—The density functions of corolla width (flower size) are depicted for the Low and High populations, respectively. The arrow denotes the mean of the Ancestral population.

This study is based on Illumina sequence data from pooled independent samples within the Low, High, and Ancestral populations. We first develop a simple statistical model based on classical population genetic methods to analyze whole genome divergence in allele frequency. Calibration of the model provides parameter estimates that are informative about the drift/sampling process and also thresholds for distinguishing significant outliers. Outliers are putative targets of selection. Surprisingly, we find that the contrast between the most divergent populations (Low vs High) indentifies fewer outliers than the comparisons of each selected population distinctly to the Ancestral population. The ancestral contrasts also identify a striking example of parallel evolution within Low and High.

This experiment also confirms two predictions from the population genetic theory of hitch-hiking. First, the relatively weak genomic signal of soft, partial sweeps is expected to be more pronounced if 1) the selected region exhibits substantial linkage disequilibrium when selection is initiated and 2) recombination is suppressed. These conditions are satisfied by a polymorphic inversion previously mapped by Lee (2009). We find that evolution at this locus affects nucleotide diversity patterns over much of chromosome 6. Apart from the inversion on chromosome 6, we find that changes in nucleotide diversity near selected loci (divergent regions) are highly variable. However, there is a clear and significant tendency for sequence diversity to increase near selected loci within the High population. This suggests that selection has made allele frequencies more intermediate within High; a result consistent with previous quantitative genetic inferences about the allele frequency spectrum at flower size loci (Kelly 2008).

Materials and Methods

Experimental Procedures

The ancestral population (Zia-1) and selection experiment are described in detail in Kelly (2008). Briefly, the ancestral population was synthesized from a random sample of more than 1000 plants from a single natural population located at Iron Mountain in Oregon, USA. Distinct Low and High populations were established from this common source. In each generation, 1000 plants were grown to maturity within each population and measured for corolla width (the largest width of the lower lip of the flower). In each generation, we selected 200 plants (the smallest flowers in Low and largest in High), randomly paired these plants (within populations) assigning one as sire and the other as dam, and hand pollinated the dam. This process was repeated for nine successive generations.

We grew progeny of the last generation simultaneously with plants germinated from the Ancestral population. We collected equal tissue per plant from 49 Low plants, 78 High plants, and 75 Ancestral population plants; each plant from an independent family within the population. We extracted DNA from each bulked tissue sample, made a single genomic DNA library for each population sample, and then sequenced each library within a distinct lane using the Illumina HiSeq 2000 instrument (v3 chemistry, 100 bp paired-end sequencing) at the University of Kansas Medical Center Genomics facility.

The QTL mapping study of Lee (2009) identified a large inversion on chromosome 6 that is segregating in the Iron Mountain population of *Mimulus guttatus*. We conducted a distinct tissue collection from 96 plants of each population for individual genotyping of markers on chromosome 6. We extracted DNA from these samples using our standard protocol (Marriage et al. 2009) followed by polymerase chain reaction (PCR) amplification of three distinct markers: MgSTS_431, MgSTS_229, and MgSTS_480. Each of these loci is length polymorphic and a particular set of alleles—240 bp at MgSTS_431, 201 bp at MgSTS_229, and 270 bp at MgSTS_480—is diagnostic of the derived inversion haplotype on chromosome 6 (Scoville et al. 2009). Allele lengths were determined using capillary electrophoresis of PCR products using an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). We sized fragments using GENEMAPPER 4.0 software (Applied Biosystems) calibrated with the ROX500 size standard (Applied Biosystems).

Processing of Sequence Data

Prior to read mapping, we trimmed low-quality ends and discarded reads reduced to fewer than 20 bases using Sickle (https://github.com/najoshi/sickle, last accessed July 30, 2013). For the High population, this process reduced the initial set of 394 million reads (197 million read pairs) to 360 million reads. For the Low population, read count was reduced from 358 to 333 million and the Ancestral population from 404 to 374 million. To map reads, we used BWA with default parameter values (Li and Durbin 2009) and a modified version of the *M. guttatus* reference genome (http://www.phytozome.net/, last accessed July 30, 2013) in which the repetitive regions were masked. After mapping, we removed PCR duplicates with Picard tools (http://picard.sourceforge.net, last accessed July 30, 2013). Duplicate removal followed by elimination of reads with a mapping quality less that 29 further reduced the number of reads to 110 million (High population), 96 million (Low), and 71 million (Ancestral), respectively.

We identified candidate polymorphisms (hereafter SNPs) using Samtools 0.1.8 (Li et al. 2009). This version of Samtools-pileup does not assume the true allele frequency in the sample is ½ or 1. Candidates were subsequently filtered based on allele frequency information from the entire dataset with High, Low, and Ancestral samples considered jointly. To eliminate false SNPs caused by sequencing and/or mapping errors, candidate SNPs were eliminated from subsequent consideration if 1) the minor allele was not present in at least two copies and at a frequency of 1% or greater within the populations-combined sample, 2) there was significant ($G > 3.84$) and substantial ($>20\%$) allele frequency divergence between strands, or 3) three bases were segregating at the site. For criterion (2), strand refers to whether a read aligned with the reference genome (plus) or as the reverse complement (minus). Without sequencing or mapping error, we expect that allele frequency should be the same in $+$ and $-$ strands; an expectation that we evaluate with a G test (Sokal and Rohlf 2000). Finally, we included a SNP within a particular contrast of populations (for example, High vs Low) only if each population had a minimum of $30\times$ coverage and a maximum of $1000\times$ coverage at the SNP. With these filters, the average coverage per SNP was $48\times$ in the Ancestral population, $61\times$ in Low, and $64\times$ in High.

After identifying valid SNPs, we recoded all read pairs to note location and allelic information at valid SNPs using custom Python scripts. These programs preserve haplotype information at the scale of read-pairs allowing estimation of linkage disequilibria (LD) between closely linked SNPs (2–500 bp). We wrote a series of C programs to calculate divergence statistics (see Statistical Model of Divergence section) and to estimate LD between all SNP pairs for which there were at least 20 read pairs within a population with valid base calls at each site. From these read pairs, we directly estimated the four "gamete frequencies" (y_{00}, y_{10}, y_{01}, and y_{11}), and then calculated $p_1 = y_{10} + y_{11}$, $p_2 = y_{01} + y_{11}$, $D = y_{11} - p_1 p_2$, and $r^2 = D^2/(p_1 p_2 [1 - p_1][1 - p_2])$. Our estimators for D and r^2 are equivalent to Method 1 (direct inference) of Feder et al. (2012). When calculating D between two sites (A/a and B/b), we defined A and B as the most common bases at each site in the Ancestral population sample. Polarizing alleles this way does not affect r^2, but is relevant to hitch-hiking effects on levels of variation. The variance increase condition of Appendix A requires positive D when alleles are scored this way.

Divergence among populations was estimated for each SNP and also in "genomic windows." For the latter, we defined windows as non-overlapping sets of S contiguous SNPs. In each window, we calculated B (eq. 2 below) as well as $\overline{pq}$, the average of $p(1-p)$ across SNPs within a window. This is essentially the pooled sample equivalent of expected heterozygosity. Conditional on S, $\overline{pq}$ is directly proportional to nucleotide diversity (the average pairwise number of differences between sequences in a sample) and is thus a natural measure of sequence diversity within windows defined to have a fixed S. The full analytical pipeline is summarized in supplementary table S3, Supplementary Material online.

Statistical Model of Divergence

Figure 2 illustrates the populations of this study and the processes/events generating allele frequency differences among populations. Inference of evolutionary processes is complicated by the fact that allele frequencies are not directly observed. Let p_A be the true allele frequency in the ancestral population; p_L and p_H for Low and High, respectively. The raw data are sequencing read counts; quantities that are removed from the allele frequencies by multiple sampling events. The first sampling event is collection of a finite number of individuals to form the "bulk" or "pool" for each population. Second, the various stages of library construction and events during the sequencing process may cause some genomes within a bulk to be overrepresented relative to others (library construction events in fig. 2). Finally, there is finite coverage of reads for any particular SNP.

Here, we employ Fisher's angular transformation (the arcsin square-root transformation) of allele frequency as it affords an

elegant statistical treatment of the successive sampling events (fig. 2). Following Fisher and Ford (1947), we let $x_A = 2 \sin^{-1}(\sqrt{P_A})$, $x_H = 2 \sin^{-1}(\sqrt{P_H})$, and $x_L = 2 \sin^{-1}(\sqrt{P_L})$. Values of x range from 0 to π radians. If divergence is due entirely to drift, x_H and x_L should be independent and approximately normally distributed random variables; each with mean x_A and variance $t/(2N_e)$. In the present experiment, $t = 9$. N_e is more difficult to estimate, but the census N was 200 within both High and Low. The effect of subsequent sampling events are additive with respect to the variance, e.g.

$$Var[x_H'|x_H] = \frac{1}{2n_H} \text{ and } Var[x_H^*|x_H''] = \frac{1}{m_H},$$

where n_H is the number of plants in the high bulk and m_H is the number of reads (at this SNP) in the high sample. It is difficult to know how to model the intermediate sampling events that occur during library construction and the sequencing process. As a consequence, we aggregate the variance effects of genetic drift, sampling of individuals to bulks, and library construction events into a single quantity (v) for each population:

$$Var[x_H^*|x_A] = v_H + \frac{1}{m_H} = u_H \tag{1a}$$

$$Var[x_A^*|x_A] = v_A + \frac{1}{m_A} = u_A \tag{1b}$$

$$Var[x_L^*|x_A] = v_L + \frac{1}{m_L} = u_L \tag{1c}$$

The population specific v terms are common to all SNPs within each population. The u terms have SNP-specific values because read depth (m values) varies among SNPs. At a neutral SNP, the difference in x^* between populations is predicted to be normally distributed with mean zero and variance equal to sum of relevant u values from equations (1).

The overall distribution of differences across the genome is a mixture of neutral SNPs and SNPs affected by divergent selection. Selected SNPs and linked polymorphisms will likely exhibit greater allele frequency divergence and inflate the tails of the overall distribution of differences. To test for significantly divergent SNPs, we would like to calculate probabilities from the null model, the normal density associated with neutral SNPs. While the neutral fraction of the distribution is not known, a robust estimator for the variance should be only minimally affected by selected SNPs unless they are very abundant in the dataset. Our single SNP analyses employ the interquartile range (IQR) of the entire distribution to estimate the variance of the underlying null distribution, and subsequently the v terms of equations (1). We then standardize the observed divergences, e.g., $x_L^* - x_H^*$, by dividing by the estimated neutral standard deviation, e.g., $\sqrt{u_L + u_H}$. The resulting quantity can be compared with the standard normal Cumulative Density Function to calculate P values for each SNP. Finally, we employ a False Discovery Rate model
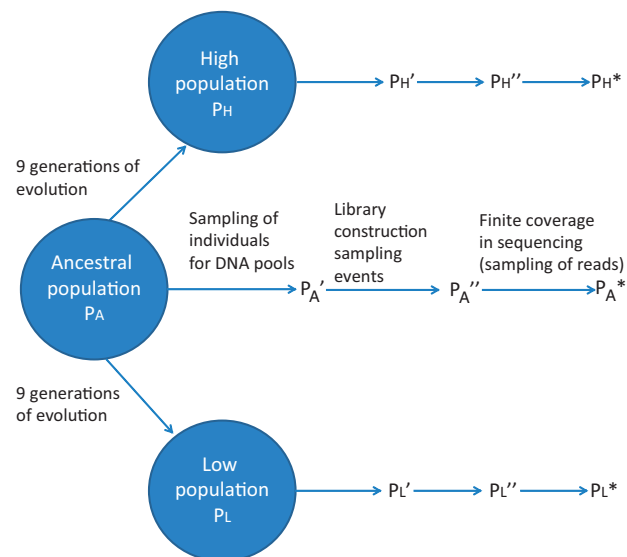

FIG. 2.—Allele frequency (p) within the Ancestral (subscript A), Low (subscript L), and High (subscript H) populations is illustrated at various stages of the experiment.

(Benjamini and Hochberg 1995 procedure), with FDR $= 0.1$, to identify SNPs that are genome-wide significant (e.g., Goldringer and Bataillon 2004).

With LD among closely linked sites, divergent selection will produce a signal beyond the causal site. For this reason, we test for divergence within genomic windows, sets of contiguous SNPs. Consider the following statistic

$$B = \sum_{i=1}^{S} d_i^2 \qquad (2)$$

where i indexes individual SNPs within a window containing S polymorphisms and d_i is the standardized divergence: $d_i = \frac{x_{Li}^* - x_{Hi}^*}{\sqrt{u_{Li} + u_{Hi}}}$ for the Low-High contrast, $d_i = \frac{x_{Hi}^* - x_{Ai}^*}{\sqrt{u_{Hi} + u_{Ai}}}$ for the High-Ancestral contrast, and $d_i = \frac{x_{Li}^* - x_{Ai}^*}{\sqrt{u_{Li} + u_{Ai}}}$ for the Low-Ancestral contrast. For a neutral SNP, $E[d_i^2] = 1$, and in a window with S polymorphisms, $E[B] = S$. Divergent selection should inflate B.

A null distribution for B can be constructed by considering the statistical properties of quadratic forms of normal variables (Pearson 1959; Imhof 1961; Kotz et al. 1967; Liu et al. 2009). The variance of B depends on the association of d_i values among different SNPs within a window (see Appendix B). If Ω is the variance/covariance matrix for d_i values, $Var[B] = 2\,Trace[\Omega^2]$ and the kth cumulant of B equals $2^{k-1}(k-1)!\,Trace[\Omega^k]$. The distribution for B is approximately proportional to a χ^2 distribution with the number of degrees of freedom chosen to match the predicted skew of B (Liu et al. 2009). Because divergence at a SNP is standardized (eq. 2), the diagonal terms of Ω are 1. The off-diagonal terms, $Cov[d_i, d_j]$, depends on the magnitude and direction of LD between SNPs i and j, and on the likelihood that both sites are contained within the same read pair (eqs. A5 and A6 of Appendix B).

We propose a testing procedure for divergence in windows analogous to the method for individual SNPs. Here, we require robust estimators for both the variance and skew of B. Bowley (1920) proposed a measure of skew based on quartiles: $skew = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$, where Q_1 is the 25th percentile of the distribution, Q_2 is the 50th percentile (median), and Q_3 is the 75th percentile. The procedure has the following steps:

1. Calculate the percentiles for the observed distribution of B across the genome for a contrast of populations. Given these, calculate Bowley's skew for the data.
2. Determine the value of δ, the degrees of freedom of a χ^2 distribution that yields an equivalent value for Bowley's skew, accepting that δ will not generally be an integer. In the present application, we determined δ iteratively using Mathematica8©.
3. Calculate $\sigma(B)$, the robust estimate for the standard deviation of B, as

$$\sigma(B) = \sqrt{2\delta}\frac{Q_3 - Q_1}{X_3 - X_1} \qquad (3)$$

where $Q_3 - Q_1$ is the IQR of the empirical distribution (B) and $X_3 - X_1$ is the IQR of a χ^2 distribution with δ d.f.
4. Calculate transformed values for B:

$$B^* = \delta + \sqrt{2\delta}\left(\frac{B - S}{\sigma(B)}\right) \qquad (4)$$

5. Compare B^* values with the Cumulative Density Function of $\chi^2[\delta]$ to determine P values for each window.
6. Given window-specific P values, apply an FDR model to establish windows that are significant genome-wide.

Results

Direct Genotyping of Inversion Markers

Given marker genotypes, we scored each individual as CC, Cc, or cc (homozygous normal, heterozygous, or homozygous inversion) and then calculated the frequency of the inversion (P) in each population (Ancestral $n = 91$, Low $n = 92$, and High $n = 96$). In the Ancestral population, $P = 0.15$ with 95% confidence interval (0.10, 0.21). In Low, $P = 0.66$ (0.59, 0.73) while in High, $P = 0.65$ (0.58, 0.72). These data indicate a striking increase in the frequency of the Inverted haplotype within both High and Low populations over the course of nine generations of selection.

Allele Frequency Divergence between Populations

After filtering, we identified 1,224,974 SNPs with at least $30\times$ coverage in the Low population, 1,405,249 in the High, and 836,385 in Ancestral. We calculated $x_L^* - x_H^*$, $x_L^* - x_A^*$, and $x_H^* - x_A^*$ for each SNP with sufficient coverage in the contrasted populations. Figure 3 illustrates that the distributions for these differences correspond closely to a normal distribution with mean zero and a contrast-specific variance (eqs. 1). The mean, variance, skew, and excess kurtosis are reported for all 14 of the main scaffolds (sequences corresponding to chromosomes 1–14) in supplementary table S1, Supplementary
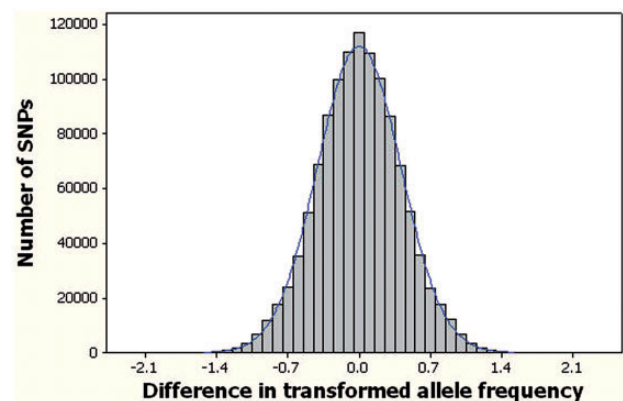
Fig. 3.—The distribution of $x_L^* - x_H^*$ is depicted for the entire genome ($n = 1,154,300$).

Material online. General agreement to the normal distribution is good for all chromosomes except chromosome 6 (supplementary fig. S1, Supplementary Material online). Skew is minimal. Excess kurtosis is typically small in magnitude, but on average positive across scaffolds. This suggests that a subset of SNPs have more extreme divergences than predicted from drift and sampling alone, consistent with genomically localized effects of selection.

The distributions for chromosome 6 are clearly non-normal with positive excess kurtosis for Low vs High, and negative excess kurtosis for each contrast against the Ancestral population (supplementary fig. S1B, Supplementary Material online). The standard deviations of $x_L^* - x_A^*$, and $x_H^* - x_A^*$ among SNPs on chromosome 6 were 0.56 and 0.59, respectively; much higher than for all other chromosomes (supplementary table S1, Supplementary Material online). These observations suggest that selection on the inversion, which suppresses recombination over a large portion of chromosome 6 (Lee 2009), has a genomically widespread effect on SNP variation. Given the evidence of a major selective event on chromosome 6, we excluded that data for subsequent model-fitting (e.g., estimating the v parameters of eqs. 1) and hypothesis testing.

Using IQR as a robust estimator for the variance of transformed allele frequency divergence, $\sigma^2 = (IQR/1.349)^2$, and then subtracting variance attributable to SNP-specific finite coverage, we estimate that $v_L + v_H = 0.13126$, $v_L + v_A = 0.07919$, and $v_A + v_H = 0.07602$. Solving, $v_L = 0.06722$, $v_H = 0.06404$, and $v_A = 0.01198$. These values, combined with the observed read counts, were employed to calculate SNP-specific d_i for each contrast, which were then compared with the standard normal distribution. For the Low versus High contrast, two SNPs were significantly divergent using a genome wide false discovery rate of 0.1. For High versus Ancestral, the 26 SNPs were significant, while no single SNP was significant for the Low versus Ancestral contrast. The significant SNPs are listed in supplemental table S2, Supplementary Material online.

Considering divergence in genomic windows of 10 SNPs ($S = 10$), figure 4 is the observed distribution of B^* (eq. 4) to the calibrated null distribution for Low versus High. There is excellent general agreement between observed and null, indicating that the estimation procedure effectively captures divergence owing to genetic drift and serial sampling (fig. 2). However, the window-based method is much more powerful than the single-SNP analysis for detecting genome-wide significant outliers. Figure 5 identifies significant outliers for each contrast: 304 windows for the High-Ancestral contrast (panel A), 50 windows for the Low-Ancestral contrast (panel B), and 40 windows for the Low-High contrast (panel A). The median size of windows was 1216 bp for High-Ancestral, 1221 bp for Low-Ancestral, and 837 bp for Low-High. We also performed the model fitting with larger windows ($S = 20$), but found fewer significant outliers than with $S = 10$.
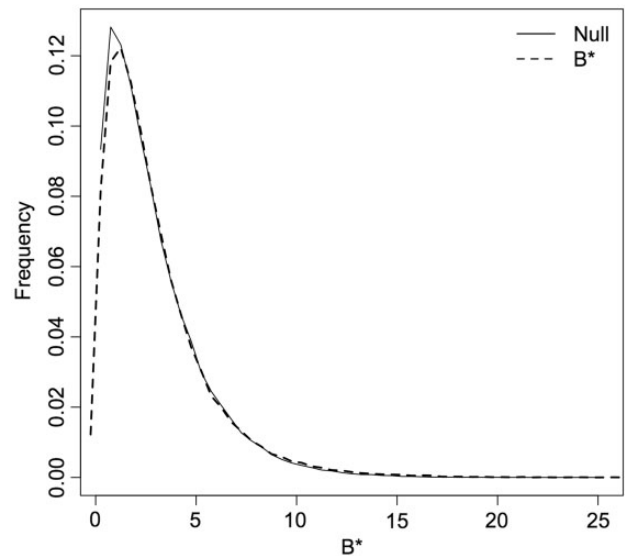

FIG. 4.—The observed density function for B^* (broken line) is compared with the fitted null density (solid line) with $S = 10$ for the Low-High contrast.

There is clear clustering of significant windows when considering their distribution across the genome, particularly noting the concentration of outliers on chromosome 8 (fig. 5). However, at the scale of 10–100 kb, the signal of selection is fairly localized. If one considers the 10 windows bracketing each significant outlier of the High-Ancestral contrast, the mean P value for these windows is 0.14. This is substantially lower than the mean of all windows ($P = 0.496$) but not near the genome-wide threshold required for an individual window to be significant ($P = 0.0004$ for High-Ancestral). Notably, all of these calculations were conducted after excluding chromosome 6. On chromosome 6, there is fairly consistent divergence between both derived populations and the ancestral populations across a large fraction of the chromosome (fig. 6).

The distributions for $\overline{pq}$ with $S = 10$ are similar across populations with nearly equivalent mean values (ca. 0.155). The minimal average change in $\overline{pq}$ from Ancestral to either Low or High indicates that genetic drift had a generally small effect on overall sequence variation in this experiment. However, there are interesting patterns if we classify windows on significance of B^* (fig. 7). Across the 304 significant windows of the High-Ancestral contrast, mean $\overline{pq}$ essentially doubled over the course of nine generations of selection. In these windows, change in $\overline{pq}$ was highly variable (actually negative in 41 of the windows), but substantially positive on average owing to the fact that selection made allele frequencies more intermediate in most windows. For the Low-Ancestral and Low-High contrasts, differences in mean $\overline{pq}$ are more modest. For Low-Ancestral, mean $\overline{pq}$ is reduced by 25% from Ancestral to
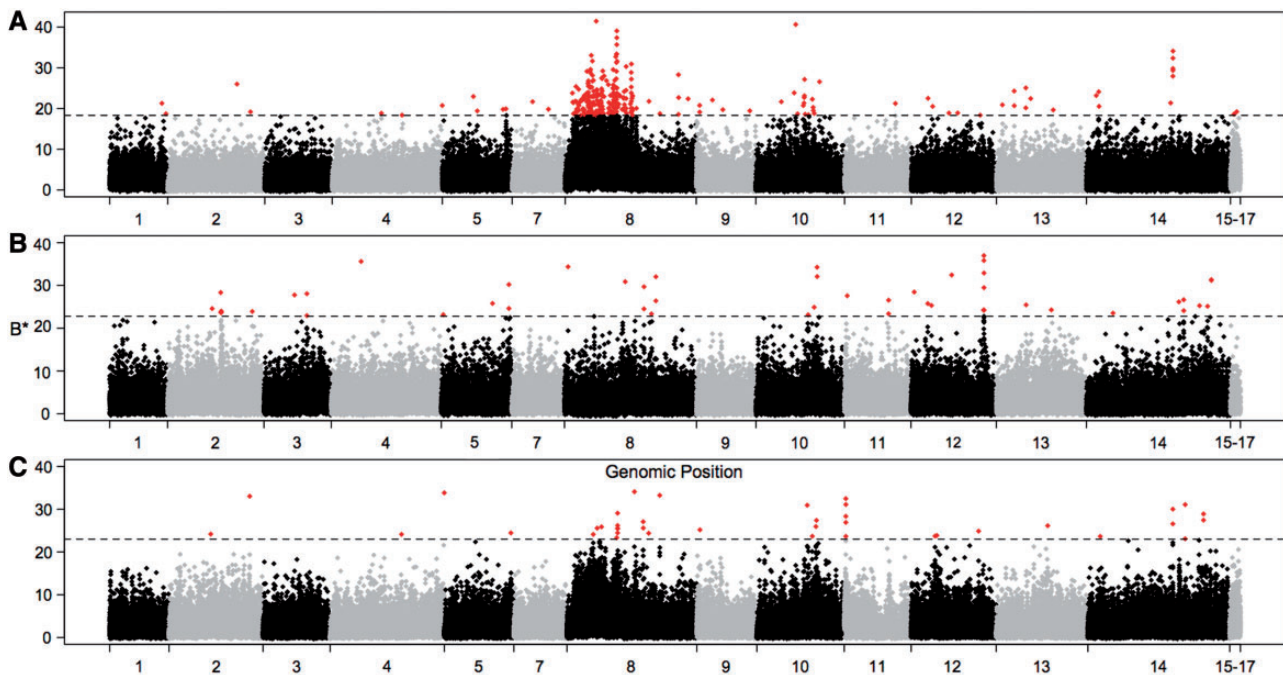
Fig. 5.—Observed B^* is plotted as a function of genomic location for the contrast of (*A*) Low-High, (*B*) Low-Ancestral, and (*C*) High-Ancestral. The broken horizontal line in each panel is the genome-wide significance threshold for each contrast (significant windows red). Scaffolds 15–17 are autosomal, but not currently located to the 14 major chromosomes in the genome build.
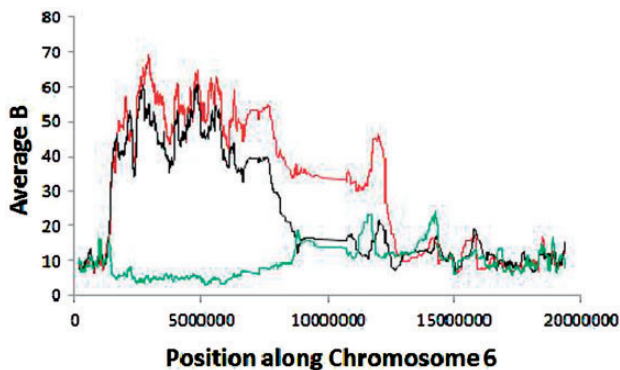

Fig. 6.—Averaged B is plotted versus genomic location on chromosome 6 (green = Low-High, red = High-Ancestral, red = High-Ancestral). *B* is calculated in windows with *S* = 10, but each trajectory is a moving average of 100 consecutive windows. The expected value for *B* under the null hypothesis is 10.

Fig. 7.—Mean $\overline{pq}$ within windows (*S* = 10) is reported (bands are a 95% CI) for various classifications of windows. The first bar is for the entire genome of the Ancestral population. The succeeding bars are in three sets, one set for each contrast. Within each set, A, L, and H refer to the population sample (Ancestral, Low, and High).

Low. For Low-High, mean $\overline{pq}$ increased in High and declined in Low relative to Ancestral.

LD among Closely Linked SNPs

Unfortunately, we cannot report LD estimates among the various SNPs within significant and non-significant windows identified in the previous section. This is because our
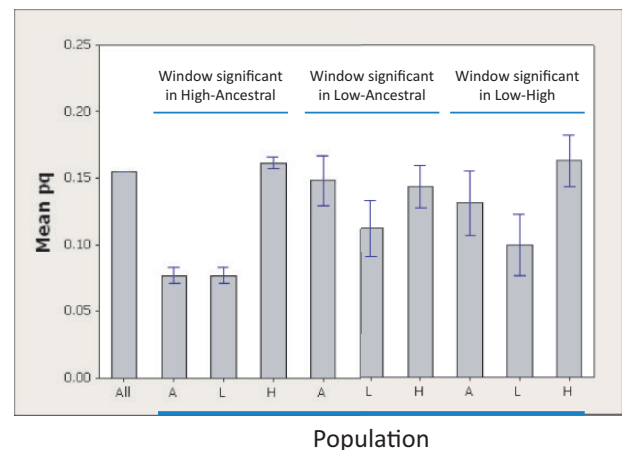
D estimates are limited to SNPs within read pairs (most estimates are for inter-SNP distances of less than 50 bp). Even at this small genomic scale, the data are very incomplete owing to idiosyncratic coverage of SNP pairs. However, the data are sufficient to identify general trends. As expected, the strength
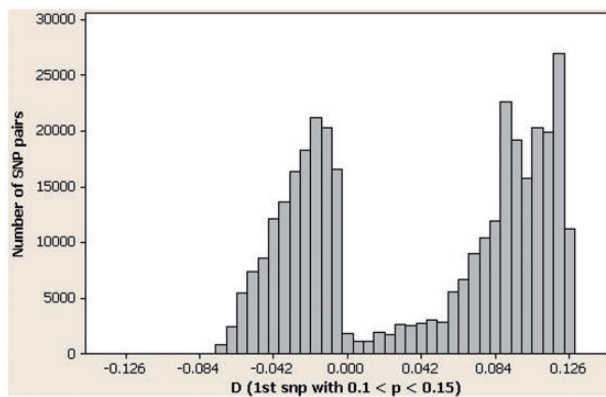
FIG. 8.—The distribution of D values for SNP pairs in the Ancestral population where one SNP has a minor allele frequency of 0.1–0.15.

of associations measured as r^2 declines with distance between SNPs (supplemental fig. S2, Supplementary Material online). There is also a tendency for rare alleles to be positively associated. Of the 1,222,058 SNP pairs with sufficient coverage in the Ancestral population, about 60% of D estimates are positive. Mean D is about 0.05 consistently across chromosomes. Figure 8 depicts the distribution of D values for all SNP pairs in the Ancestral population where one SNP has a minor allele frequency in the range of 0.1–0.15. Not only are most D values positive, but the average magnitude of positive D estimates is greater than the average magnitude of negative D estimates.

The level of LD is also relevant to our null model for window-based divergence (Appendix B). The variance of B depends partly on the pattern and strength of LD among SNPs within a window, and since LD will vary over the genome, the true neutral distribution for B (and B*) is itself a mixture distribution. Exaggerated values for B* could result from drift/sampling even in neutral regions if they exhibit unusually high LD. To evaluate whether this potential difficulty is relevant to the present dataset, we determined average r^2 in the Ancestral population for genomic regions that either include or do not include significantly divergent windows. We delineated significant regions as the 10 kb interval containing a window start site, plus the 10 kb on either side. These criteria capture 560 distinct 10 kb intervals across the significant windows of all three contrasts. The mean r^2 for these windows is 0.451 (SEM = 0.007). For the remainder of the genome (21,862 10 kb intervals), the mean r^2 is 0.474 (SEM = 0.001). This indicates that significant values for B* are not a spurious effect of genetic drift on high LD portions of the genome.

Discussion

The divergence of High and Low populations from this study is trait evolution owing to changes in allele frequency at many loci. Mean corolla width differs between these populations by nearly 3-fold (fig. 1), typical of interspecific differences in flower size within *Mimulus*. Low and High have also separated substantially in morphology (Kelly 2008), in physiology (Kelly et al. 2008), and perhaps most importantly, in flowering phenology. Low plants flower an average of 6 days earlier than High plants. Days-to-flower is a critical fitness determinant in the field (Mojica and Kelly 2010), and flowering asynchrony between populations is an important reproductive isolating mechanism, both in *M. guttatus* (Martin and Willis 2007) and many other plants (e.g., Petit et al. 1997; Husband and Schemske 2000; Borchsenius 2002). The quantitative trait divergence of Low and High represents polygenic adaptation (*sensu* Pritchard and Di Rienzo 2010) because we found not a single fixed difference between these populations across nearly a million polymorphisms.

The three different contrasts (High vs Low, High vs Ancestral, and Low vs Ancestral) identified nearly 400 outlier windows (fig. 5). Some of these are likely redundant; significant tests in closely linked windows reflecting the same selective event. Corroboration for these divergent regions comes from a QTL mapping study based on these same populations (Lee 2009). Lee performed three crosses between distinct High and Low plants sampled from generation 6 of the selection experiment. She selfed each F1 to create three replicate F2 mapping panels (approximately 380 plants per panel) and then measured genotype–phenotype associations. For 22 corolla width QTLs in Lee (2009), we could locate the marker that was nearest the LOD peak (estimated QTL location on the linkage map) to a scaffold of the genome sequence (here we are ignoring chromosome 6 because of the inversion). For 10 of 22 QTLs, there was a significantly divergent window within 100 kb. For 16 of 22 QTLs, a divergent window is within 2 mb.

To evaluate this level of congruence, we determined that there is about a 9% chance that a randomly selected location in the genome is within 100 kb of a significantly divergent window. A simple binomial calculation can then evaluate the null hypothesis that QTL and divergent windows are uncorrelated. The probability of getting 10 (or more) out of 22 QTLs within 100 kb is only 0.00000086. Thus, the correspondence is highly significant, although there are quite a few QTL markers that are not within 100 kb of a significantly divergent window. The reasons for this are multiple. There are likely some genuine differences, i.e., mapped flower size loci that did not contribute to response to selection. However, there are also several factors generating false negatives. For one, QTL mapping in F2s yields broad intervals for location. As a consequence, causal polymorphisms may be fairly distant from the QTL LOD peak. Second, the genome-wide significance threshold of the present study is quite stringent, requiring high divergence in allele frequency between populations. If one considers a less stringent threshold—windows with a

P value <0.001—we find that 21 of 22 QTL markers are within 1 mb of a divergent window. It is also notable that a large fraction of significantly divergent windows in this study are on chromosome 8 (fig. 5), which yielded more QTL than any other chromosome in Lee (2009) and for which subsequent fine mapping has identified numerous distinct flower size loci (Scoville et al. 2011).

QTL Allele Frequency and the Effect of Selection on Levels of Molecular Variation

We measured sequence variation in terms of $\overline{pq}$, the average of $p(1-p)$ within a window. $\overline{pq}$ is a distillation of the allele frequency spectrum at SNPs and is informative because short-term changes in sequence variation are caused by shifts in the relative abundance of distinct haplotypes. Conditional on the number of polymorphisms within a window (S), $\overline{pq}$ is directly proportional to nucleotide diversity (the average number of differences between sequences in a sample, Nei and Li 1979). However, nucleotide diversity is typically calculated in windows that vary in S. It thus integrates evolutionary processes over a much longer time scale over which mutation and the genealogical structure of the population influence the number of polymorphisms in a region.

We find that soft, partial sweeps have a complicated and variable effect on levels of molecular variation (fig. 7). This is to be expected given that the predicted change in $\overline{pq}$ due to selection depends on the initial frequency of the favored allele, the final frequency of the favored allele, and on the particular nature of LD between the selected site and neighboring SNPs (Appendix A). Despite variability, there are some statistically significant trends. In significant windows from Low versus High, mean $\overline{pq}$ increased by about 25% in High and declined comparably in Low relative to the Ancestral population over the nine generations of selection (fig. 7). As a consequence, mean $\overline{pq}$ differs significantly between Low and High. Importantly however, the magnitude of this difference (ca. 0.06) is not extreme when considered in relation to background variation. Across the entire genome, the standard deviation of differences in $\overline{pq}$ between Low and High is 0.04. Because the mean change within significant windows is only 1.5 SD of the background distribution, it is essentially impossible to assert an effect of selection strictly from changes in the amount of sequence variation.

The most substantial signal in mean $\overline{pq}$ occurred within the High population where sequence variation was substantially increased by selection (see High-Ancestral windows of fig. 7). We hypothesize this is due the fact that "high" alleles (those increasing flower size) are routinely less common than alternative "low" alleles in the ancestral population. If the high allele at a flower size QTL increased from 10% to 50% due to selection, the change in pq at this site would be 0.16 contributing positively to $\Delta\overline{pq}$ for the window in which it resides. Of course, $\Delta\overline{pq}$ is mainly determined by changes in allele frequency at the non-causal SNPs within the window. Changes at these sites will generally be smaller in magnitude than for the causal SNP owing to imperfect associations of allele frequencies among neighboring sites; a salient feature of selection on standing variation (Innan and Kim 2004; Hermisson and Pennings 2005). Despite this, the inflation of mean $\overline{pq}$ indicates that, for most windows, allele frequencies at non-causal sites became more intermediate with selection. This requires that, on average, the uncommon allele at the causal site is positively associated with the less common allele at neighboring sites (Appendix A). Our pooled sample experimental design does not allow us to calculate LD among the various SNPs defining each window, except for a small fraction of the relevant SNP pairs. However, the admittedly incomplete data from closely linked SNPs indicate the genome-wide pattern of linkage disequilibrium in our ancestral population. As required for positive $\Delta\overline{pq}$, linkage disequilibrium is usually positive (fig. 8).

The hypothesis that high flower size alleles are less frequent than their alternatives within the ancestral population is consistent with prior studies of this population. After six generations of the selection experiment, we conducted a breeding design to estimate V_A, the additive genetic variance of flower size within each population (Kelly 2008). V_A was slightly reduced in Low, but substantially elevated in High, relative to the Ancestral population. Response of the mean phenotype to selection was symmetric in early generations (equal magnitude of change in Low and High) but has become asymmetric over generations 5–9 with High more removed from Ancestral than Low (fig. 1). These phenotype-based observations are fully consistent with reduced mean $\overline{pq}$ at causal loci in Low and increased $\overline{pq}$ in High, but they do not conclusively indicate this model. Asymmetric response to selection can have numerous causes (Falconer and Mackay 1996, chapter 12). For example, epistasis can cause asymmetric changes in V_A and in trait means even with perfectly symmetric changes in allele frequency. The present data speak more directly to changes in allele frequencies.

Field experiments provide a potential explanation for why high alleles might be less frequent than low alleles at flower size loci. The ancestral population of this study is a large random sampling of genotypes from the Iron Mountain population of *M. guttatus* in central Oregon (Kelly 2008). We have conducted multiple field studies measuring the fitness consequences of Iron Mountain floral variants (Mojica and Kelly 2010; Mojica et al. 2012). High alleles for flower size tend to reduce viability (the probability that a plant survives to flower) but increase fecundity (seed set of plants that manage to flower). The relative strength of viability and fecundity selection vary both spatially and temporally. However, in most years and microsites, estimated lifetime fitness has been greater for small-flowered genotypes than large-flowered genotypes.

A Polymorphic Inversion Produces a Broad Genomic Signal of Selection

Lee (2009) identified the chromosome 6 inversion from recombination suppression (clustering of many markers on a genetic map) in the QTL mapping study described above. Subsequent genotyping of 93 outbred plants from the natural population showed that the "inversion" is a derived haplotype with a specific combination of alleles across 13 marker loci that span 4.2 mb of the reference genome (fig. 6 suggests that the inversion may be twice that size). The alternative ancestral gene order is composed of many distinct haplotypes as expected given historical recombination among marker loci. The increase of the derived inversion type from 15% in the Ancestral population to 65% in both Low and High produces a broad signal of divergence (fig. 6) because a single haplotype increased in frequency, largely unbroken by recombination.

Despite the large change in inversion frequency, levels of sequence variation within this region of chromosome 6 (from the 1 mb to 8 mb position) are only marginally different among populations. Mean $\overline{pq}$ of windows in the Ancestral population is about 0.17; about 0.13 in both Low and High. If the inversion harbored many new mutations, these specific alleles would have been uncommon in the Ancestral population and then at intermediate frequency in both Low and High (based on the observed frequency of inversion haplotype). We would then have expected mean $\overline{pq}$ to have increased with the inversion. However, the marker genotyping of the natural population noted above clearly indicates that the inversion haplotype is composed of alleles sampled from variation resident in the ancestral population (Lee 2009). Given that the "inversion allele" at a SNP is not predictably the minor allele in the remainder of the Ancestral population, a slight decline in $\overline{pq}$ is not surprising as the inverted haplotype rose from 15% to 65%. However, we do predict that $\overline{pq}$ should decline precipitously if the inversion haplotype increases toward fixation, at least if it remains internally homogeneous.

We do not yet know why selection generated parallel evolution of the inversion within both Low and High populations of this experiment. Previous mapping studies have found weak or inconsistent effects of the inversion on corolla width (Lee 2009; Scoville et al. 2009) which was our target of artificial selection. Moreover, the selection experiment had a contemporaneous control—a population maintained at the same size but without selection (Kelly 2008)—and the inversion is also at ca. 0.65 in this population (Kelly JK, unpublished results). The parallel increase of the inversion in three independent populations strongly implicates some sort of selection. This is surprising given that the methods for propagation in the artificial selection experiment substantially limited opportunities for incidental selection. Each surviving adult was mated to only one plant, and the number of progeny per family was approximately equal across populations and generations.

There was opportunity for uncontrolled fitness variation at the seed germination stage. It is possible that the derived inversion type increases seed germination under greenhouse growth conditions and that this is responsible for its increase in frequency within Low, High, and Control populations.

Identifying Selected Loci in Pooled Samples

Artificial selection is increasingly used as a tool for mapping QTL. When this is the purpose, having replicated populations for each direction of selection is clearly advantageous. Parallel divergence, in which the same SNP exhibits extreme allele frequency change across independent populations, is a strong evidence that the SNP affects the trait under selection (Turner et al. 2011; Remolina et al. 2012). Unfortunately, replicate populations that have evolved under exactly the same selection regime may not be available, particularly if one is sampling populations from nature. Moreover, even true replicate populations can respond to the same selection regime via different genetic loci (Cohan 1984). For these reasons, it is useful to have rigorous statistical procedures that identify significantly divergent sites or windows within individual populations.

We sequenced population samples in which many plants were combined equally into a single DNA pool. This strategy has been applied to both experimentally evolved populations (e.g., Burke et al. 2010; Turner et al. 2011; Remolina et al. 2012) and to divergent natural populations (e.g., Emerson et al. 2010; Kolaczkowski et al. 2011). Pooling has numerous advantages (Futschik and Schlötterer 2010) but also important limitations (Cutler and Jensen 2010). One caveat regards proper statistical testing for allele frequency differences between populations. The data structure allows a standard contingency table test, e.g., χ^2 or Fisher's exact test, on allele counts from each population. However, the resulting P value is not generally valid. In pooled samples, alleles cannot be treated as independent samples from their respective populations unless the two intermediate sampling events of figure 2 (individuals into pools and library construction events) are inconsequential.

The statistical issue of serial sampling inherent to pooling has been treated in a number of ways depending on context (Futschik and Schlötterer 2010; Kolaczkowski et al. 2011; Magwene et al. 2011). The method applied here (eqs. 1–4) is intended not just to test whether populations are significantly different but whether differences are too large to be explained by genetic drift. An initially surprising result of our study is that the contrasts of Low and High populations distinctly to the Ancestral population identified more putatively selected windows than the contrast of Low to High. Relevant here is that our test statistics for both individual SNPs and genomic windows are essentially signal-to-noise ratios. Increased power can be achieved either by increased signal or reduced noise. The Low-High contrast exhibits the greatest phenotypic divegence, but is also encumbered by the greatest

amount of noise owing to the accumulation of drift effects within two lineages instead of one.

Our analysis differs from prior treatments of comparable genomic data in two major ways. First, we apply the arcsin square root transformation, a.k.a. the angular transformation, to allele frequencies prior to calculating divergence. This is an alternative to F_{ST} or other statistics that are functions of untransformed allele frequencies. The magnitude of changes in untransformed frequency due to genetic drift and sampling is contingent on the initial allele frequency (Fisher and Ford 1947), and as consequence, different SNPs have different null distributions. A common variance term applies genome-wide for divergence in transformed allele frequency, and as predicted by theory, the resulting distribution is well approximated by the normal curve (fig. 3).

Secondly, we adopt a "non-parametric" method for identifying significant outliers. We treat the overall distribution of a divergence statistic (either for individual SNPs or within windows) as a mixture of neutral and selected loci. We assume the former to be far more abundant than the latter and then calculate robust estimators of dispersion from the overall distribution. These robust estimators are used to calibrate the null distribution under the assumption that they will be minimally affected by outliers (selected loci). Alternatively, we could have simulated the entire process (fig. 2) under the assumptions that sites are neutral. This parametric option would involve specifying the genetic composition of the Ancestral population as well as all parameters governing random changes in allele frequency. As is typical of parametric/non-parametric contrasts, the potential difficulty with the non-parametric option is loss of power. If selection has pervasive effects on genomic divergence, the null variances (eqs. 1) will be overestimated, and the testing procedure may become excessively conservative. The potential difficulty with the parametric procedure is that a simulation may fail to accurately characterize dispersive processes, and depending of the nature of the error, tests may be either conservative or anti-conservative. Of course, if one uses the observed distribution of divergence to calibrate a parametric simulation (say by choosing an effective population size that reiterates the observed variance), then parametric and non-parametric procedures may converge.

Supplementary Material

Acknowledgments

Appendix A

Here, we consider the change in $\overline{pq}$ within a genomic window where one SNP is the target of selection. Let Q denote the frequency of the favorable allele at a selected SNP and P denote allele frequency at one of the other SNPs within the window. The change in $\overline{pq}$ is $\left(\delta_0 + \sum_{i=1}^{S-1} \delta_i\right)/S$, where δ_0 is the change at the selected SNP and δ_i is the change at the ith flanking SNP. Letting primes denote values after a duration of evolution, $\delta_0 = Q'(1-Q') - Q(1-Q)$ and $\delta_i = P_i'(1-P_i') - P_i(1-P_i)$. δ_0 will be positive if allele frequencies are more intermediate after selection, i.e., if Q' is closer to 0.5 than Q. With deterministic selection and no recombination,

$$P_i' = P_i + D\frac{Q'-Q}{Q(1-Q)}$$

where D is the standard linkage disequilibrium between the selected SNP and site i prior to the interval of selection. δ_i will be non-zero only if there is linkage disequilibrium given that we are ignoring drift. δ_i will be positive if $D(D(Q'-Q) - Q(1-Q)(1-2P)) < 0$. Given that allelic identity has not been specified, we are free to define P as the frequency of allele at the flanking SNP that is positively associated with the favorable allele at the selected SNP ($D > 0$). With this convention, it follows that δ_i will be positive if

$$P_i < \frac{1}{2} - \frac{D(Q'-Q)}{2Q(1-Q)}$$

Given that Q must be less than ½ for δ_0 to be positive, positive δ_i requires that the favored allele is positively associated with the minor allele at the flanking neutral SNP.

Appendix B

The covariance of standardized change at two linked sites, d_i and d_j, is derived here for the neutral case. These covariances are the off-diagonal terms of Ω which determines $\text{Var}[B]$. Most of the calculations are based on untransformed allele frequencies. However, we can calculate the moments in terms of x^* by using

$$\text{Cov}[x_1, x_2] = \frac{dx_1}{dp_1}\frac{dx_2}{dp_2}\text{Cov}[p_1, p_2]$$
$$= \frac{1}{\sqrt{p_1(1-p_1)p_2(1-p_2)}}\text{Cov}[p_1, p_2] \quad (A1)$$

Considering two linked SNPs, we can specify the ancestral frequencies of the "gametes" or haplotypes as y_{00}, y_{10}, y_{01}, and y_{11}. The respective frequencies after genetic drift, sampling of individuals into libraries, and subsequent library construction sampling events (fig. 2) is distinguished with a double prime, e.g., y''_{11}. If the sites are sufficiently close so that recombination is negligible over the course of the experiment and that there is no effect of selection, the $E[y''_{ij}] = y_{ij}$

and $Var[y''_{ij}] = \alpha(1 - y_{ij})y_{ij}$, where α is equal to the contributions from drift $\left(\frac{t}{2N}\right)$ plus individual sampling $\left(\frac{1}{2m}\right)$ plus an increment attributable to any differential representation of sampled genomes.

For any two distinct gametes, $Cov[y''_{ij}, y'_{kl}] = -\alpha y_{ij}y_{kl}$. Letting $p_1 = y_{10} + y_{11}$ and $p_2 = y_{01} + y_{11}$,

$$
\begin{aligned}
Cov[p''_1, p''_2] &= Var[y''_{11}] + Cov[y''_{10}, y''_{01}] + Cov[y''_{01}, y''_{11}] \\
&\quad + Cov[y''_{10}, y''_{11}] \\
&= \alpha[y_{11}(1 - y_{11}) - y_{10}y_{01} - y_{11}y_{01} - y_{11}y_{10}] \\
&= \alpha[y_{11}y_{00} - y_{10}y_{01}] = \alpha D
\end{aligned}
\tag{A2}
$$

where D is the linkage disequilibrium between these SNPs in the ancestral population. Using equation (A1), we find the covariance for transformed allele frequencies prior to sampling reads:

$$
Cov[x''_1, x''_2] = \alpha \frac{D}{\sqrt{p_1(1 - p_1)p_2(1 - p_2)}} = \alpha r
\tag{A3}
$$

where r is the correlation of allelic identity between SNPs (Hartl and Clark 1989, p. 53–54). To determine effects of finite read coverage, we need to elaborate the notation related to this stage of sampling. If m_1 is the read depth at SNP 1, then

$$
m_1 = m_{1[0*]} + m_{1[1*]} + m_{1[00]} + m_{1[01]} + m_{1[10]} + m_{1[11]}
$$

where $m_{1[0*]}$ is the count of reads with allele 0 at SNP 1 and no data at SNP 2, $m_{1[1*]}$ is the count of reads with allele 1 at SNP 1 and no data at SNP 2, and $m_{1[ij]}$ is the count of reads with gamete $[ij]$ across SNPs 1 and 2.

$$
Cov[p^*_1, p^*_2 | p''_1, p''_2] = Cov\left[\frac{m_{1[1*]} + m_{1[10]} + m_{1[11]}}{m_1}, \frac{m_{2[*1]} + m_{2[01]} + m_{2[11]}}{m_2}\right]
\tag{A4}
$$

Let $m_{12} = m_{1[00]} + m_{1[10]} + m_{1[01]} + m_{1[11]}$ be the number of reads overlapping SNPs 1 and 2 ($m_{12} \leq \min(m_1, m_2)$). Conditioning observed read counts, we can treat m_1, m_2, and m_{12} as constants while the other m terms are random variables. Given haplotype frequencies within the library, covariance of allele frequencies after accumulating reads is generated only by reads that overlap both SNPs. Thus,

$$
\begin{aligned}
Cov[p^*_1, p^*_2 | p''_1, p''_2] &= \frac{1}{m_1 m_2} Cov[m_{1[10]} + m_{1[11]}, \\
&\quad m_{2[01]} + m_{2[11]}] \\
&= \frac{m_{12}}{m_1 m_2} Cov[-y''_{10}y''_{01} - y''_{10}y''_{11} \\
&\quad - y''_{11}y''_{01} + y''_{11}(1 - y''_{11})] \\
&= \frac{m_{12}}{m_1 m_2} D''
\end{aligned}
\tag{A5}
$$

where D'' is the linkage disequilibrium between SNPs given gamete frequencies within the library. It follows that

$Cov[x^*_1, x^*_2 | x''_1, x''_2] = \frac{m_{12}}{m_1 m_2} r''$. We can obtain the overall (unconditioned) covariance of x^*_1 and x^*_2 from these parts:

$$
\begin{aligned}
Cov[x^*_1, x^*_2] &= E[Cov[x^*_1, x^*_2 | x''_1, x''_2]] \\
&\quad + Cov[E[x^*_1 | x''_1, x''_2], E[x^*_2 | x''_1, x''_2]] \\
&= E[Cov[x^*_1, x^*_2 | x''_1, x''_2]] + Cov[x''_1, x''_2] \\
&\approx r\left(\frac{m_{12}}{m_1 + m_2} + \alpha\right)
\end{aligned}
\tag{A6}
$$

This approximation assumes $E[r''] \approx r$. Drift and sampling do effect LD: $E[D''] = (1 - \alpha)D$. However, they also reduce the denominator of r in a comparable way. The final step is just to standardize, calculating d_i and d_j appropriate to a particular contrast. For the Low-High comparison,

$$
\begin{aligned}
Cov[d_i, d_j] &= Cov\left[\frac{x^*_{Li} - x^*_{Hi}}{\sqrt{u_{Hi} + u_{Li}}}, \frac{x^*_{Lj} - x^*_{Hj}}{\sqrt{u_{Hj} + u_{Lj}}}\right] \\
&= \frac{Cov[x^*_{Li}, x^*_{Lj}] + Cov[x^*_{Hi}, x^*_{Hj}]}{\sqrt{u_{Hi} + u_{Li}}\sqrt{u_{Hj} + u_{Lj}}}
\end{aligned}
$$

where the numerator covariances are given by equation (A6) with population/SNP-specific parameters.

Literature Cited

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B Biol Sci. 263: 1619–1626.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 57:289–300.

Borchsenius F. 2002. Staggered flowering in four sympatric varieties of Geonoma cuneata (Palmae). Biotropica 34:603–606.

Bowley AL. 1920. Elements of statistics. New York (NY): Scribner.

Burke MK, et al. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467:587–590.

Chevin L-M, Hospital F. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180: 1645–1660.

Cohan FM. 1984. Genetic divergence under uniform selection. I. Similarity among populations of Drosophila melanogaster in their responses to artificial selection for modifiers of ciD. Evolution 38:55–71.

Cutler DJ, Jensen JD. 2010. To pool, or not to pool? Genetics 186:41–43.

Dobzhansky TH, Levene H. 1948. Genetics of natural populations. XVII. Proof of operation of natural selection in wild populations of Drosophila pseudoobscura. Genetics 33:537–547.

Emerson KJ, et al. 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proc Natl Acad Sci U S A. 107: 16196–16200.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics. Essex (UK): Prentice Hall.

Feder AF, Petrov DA, Bergland AO. 2012. LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS One 7:e48588.

Fisher RA, Ford EB. 1947. The spread of a gene in natural conditions in a colony of the moth Panaxia dominula. Heredity 1:143–174.

Ford EB. 1971. Ecological genetics. London: Chapman and Hall.

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled dna samples. Genetics 186:207–218.

Goldringer I, Bataillon T. 2004. On the distribution of temporal variations in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection. Genetics 168: 563–568.

Hartl DL, Clark AG. 1989. Principles of population genetics. Sunderland (MA): Sinauer Associates.

Hayes BJ, et al. 2009. A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. Anim Genet. 40:176–184.

Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169: 2335–2352.

Hohenlohe P, et al. 2010a. Population genomic analysis of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6:e1000862.

Hohenlohe PA, Phillips PC, Cresko WA. 2010b. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. Int J Plant Sci. 171:1059–1071.

Husband BC, Schemske DW. 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. J Ecol. 88:689–701.

Imhof JP. 1961. Computing the distribution of quadratic forms in normal variables. Biometrika 48:419–426.

Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci U S A. 101: 10667–10672.

Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö. 2010. Genome-wide effects of long-term divergent selection. PLoS Genet. 6:e1001188.

Kelly JK. 2008. Testing the rare alleles model of quantitative variation by artificial selection. Genetica 132:187–198.

Kelly JK, Holeski LM, Arathi HS. 2008. The genetic correlation between flower size and water use efficiency in monkeyflowers. Evol Ecol Res. 10:147–152.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical australian populations of *Drosophila melanogaster*. Genetics 187:245–260.

Kotz S, Johnson NL, Boyd DW. 1967. Series representations of distributions of quadratic forms in normal variables. I. Central case. Ann Math Statist. 38:823–837.

Lee YW. 2009. Genetic analysis of standing variation for floral morphology and fitness components in a natural population of *Mimulus guttatus* (common monkeyflower). Durham (NC): Duke University.

Lewontin RC. 1974. The genetic basis of evolutionary change. New York (NY): Columbia University Press.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175–195.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760.

Li H, et al.; G. P. D. P. Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liu H, Tang Y, Zhang HH. 2009. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. Comput Stat Data Anal. 53:853–856.

Magwene PM, Willis JH, Kelly JK. 2011. The statistics of bulk segregant analysis using next generation sequencing. PLoS Comput Biol. 7: e1002255.

Marriage TN, et al. 2009. Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). Heredity 103:310–317.

Martin NH, Willis JH. 2007. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric Mimulus species. Evolution 61:68–82.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23:23–35.

Mojica JP, Kelly JK. 2010. Viability selection prior to trait expression is an essential component of natural selection. Proc R Soc B Biol Sci. 277: 2945–2950.

Mojica JP, Lee YW, Willis JH, Kelly JK. 2012. Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. Mol Ecol. 21: 3718–3728.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 76: 5269–5273.

Oleksyk TK, Smith MW, O'Brien SJ. 2010. Genome-wide scans for footprints of natural selection. Philos Trans R Soc Lond B Biol Sci. 365: 185–205.

Pavlidis P, Hutter S, Stephan W. 2008. A population genomic approach to map recent positive selection in model species. Mol Ecol. 17: 3585–3598.

Pearson ES. 1959. Note on an approximation to the distribution of non-central x2. Biometrika 46:364.

Pennings PS, Hermisson J. 2006. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet. 2:e186.

Petit C, Lesbros P, Ge X, Thompson JD. 1997. Variation in flowering phenology and selfing rate across a contact zone between diploid and tetraploid *Arrhenatherum elatius* (Poaceae). Heredity 79: 31–40.

Pritchard JK, Di Rienzo A. 2010. Adaptation—not by sweeps alone. Nat Rev Genet. 11:665–667.

Remolina SC, Chang PL, Leips J, Nuzhdin SV, Hughes KA. 2012. Genomic basis of aging and life-history evolution in *Drosophila melanogaster*. Evolution 66:3390–3403.

Scoville A, Lee YW, Willis JH, Kelly JK. 2009. Contribution of chromosomal polymorphisms to the G-matrix of *Mimulus guttatus*. New Phytol. 183: 803–815.

Scoville AG, Lee YW, Willis JH, Kelly JK. 2011. Explaining the heritability of an ecologically significant trait in terms of individual QTLs. Biol Lett. 7: 896–898.

Sokal RR, Rohlf FJ. 2000. Biometry. New York (NY): Freeman and Company.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. PLoS Genet. 7:e1001336.

Associate editor: Cécile Ané