

Selective Sweeps Reveal Candidate Genes for Adaptation to Drought and Salt Tolerance in Common Sunflower, *Helianthus annuus*

Nolan C. Kane^{*,†,1} and Loren H. Rieseberg^{*,†}

^{*}Department of Biology, Indiana University, Bloomington, Indiana 47405 and [†]Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Manuscript received November 1, 2006
Accepted for publication January 9, 2007

ABSTRACT

Here we report the results of an analysis of variation at 128 EST-based microsatellites in wild *Helianthus annuus*, using populations from the species' typical plains habitat in Kansas and Colorado, as well as two arid desert and two distinct brackish marsh areas in Utah. The test statistics lnRV and lnRH were used to find regions of the genome that were significantly less variable in one population relative to the others and thus are likely to contain genes under selection. A small but detectable percentage (1.5–6%) of genes showed evidence for selection from both statistics in any particular environment, and a total of 17 loci showed evidence of selection in at least one environment. Distance-based measures provided additional evidence of selection for 15 of the 17 loci. Global F_{ST} -values were significantly higher for candidate loci, as expected under divergent selection. However, pairwise F_{ST} -values were lower for populations that shared a selective sweep. Moreover, while spatially separated populations undergoing similar selective pressures showed evidence of divergence at some loci, they evolved in concert at other loci. Thus, this study illustrates how selective sweeps might contribute both to the integration of conspecific populations and to the differentiation of races or species.

BECAUSE plants are sessile during a significant fraction of their life cycle, they are subject to strong selection for optimal performance under local environmental conditions. As a consequence, most plant populations are expected to exhibit significant local adaptation, an expectation that has been corroborated by a century of reciprocal transplant and common garden studies (TURESSON 1922; CLAUSEN *et al.* 1940, 1948; SCHEMSKE 1984; STANTON and GALEN 1997; DONOHUE *et al.* 2000; RAMSEY *et al.* 2003; VERHOEVEN *et al.* 2004; ANGERT and SCHEMSKE 2005). Significant progress has been made toward identifying phenotypic traits that contribute to adaptation and estimating the strength of selection on these traits (ARNQVIST 1992; DUDLEY 1996; DUDLEY and SCHMITT 1996; NAGY 1997; PETIT and THOMPSON 1998; WEINIG 2000; KINGSOLVER *et al.* 2001; CALLAHAN and PIGLIUCCI 2002; UNGERER *et al.* 2002; MCKAY *et al.* 2003; ETTERTSON 2004; GROSS *et al.* 2004; GRIFFITH and WATSON 2005). However, little is known about the genetic basis of locally adapted differences, particularly outside of the model plant, *Arabidopsis thaliana* (MITCHELLOLDS 1996; PURUGGANAN and SUDDITH 1998; UNGERER *et al.* 2002; CORK and PURUGGANAN 2005; STINCHCOMBE *et al.* 2005). Important issues that remain to be addressed include the number, kinds, and effects

of genes and mutations that contribute to local adaptation, the source of adaptive variants (new mutation, standing variation, or migration), the effects of genetic background and environment on allelic fitness, and the geographic extent of selective sweeps.

Providing answers to all of these questions, even within a single model system, is a formidable task. Nonetheless, progress is being made on several fronts. For example, QTL studies have been useful for estimating a minimum number of genes underlying fitness traits, either under laboratory (SARIGORLA *et al.* 1992; MITCHELLOLDS 1996; BRADSHAW *et al.* 1998; FRY *et al.* 1998; KIM and RIESEBERG 1999; UNGERER *et al.* 2003) or under field (SCHEMSKE and BRADSHAW 1999; LEXER *et al.* 2003; WEINIG *et al.* 2003b) conditions. QTL approaches also have been employed to estimate the fitness effects of particular genotypes across environments and have shown that different QTL may contribute to fitness when the environment changes (WEINIG *et al.* 2003a). While these kinds of studies provide insight into the complex and conditional nature of adaptation to local environmental conditions, they rarely lead to a detailed understanding of the genetic basis of adaptation because a given QTL typically contains dozens or hundreds of genes, any one (or several) of which might be under selection. Only for major QTL is cloning and functional analysis likely to be feasible (although see (KROYMANN and MITCHELLOLDS 2005), yet these represent a biased subset of the genes and mutations contributing to adaptation.

¹Corresponding author: Department of Biology, Indiana University, Jordan Hall, Box B055, 1001 E. Third St., Bloomington, IN 47405.
E-mail: nkane@indiana.edu

Another strategy that has become increasingly popular is the search for correlations between sequence and expression variation at candidate genes and phenotypic variation at fitness-related traits (or fitness itself). This approach is likely to be most successful in model organisms (PIGLIUCCI and SCHMITT 1999; KOHN *et al.* 2000, 2003; HARR *et al.* 2002; WAYNE and MCINTYRE 2002; STORZ 2005), where the gene of interest has been functionally characterized. However, candidate gene lists are necessarily biased toward developmentally important or well-studied genes. These genes, while showing large phenotypic effects in the lab, may not be the most important sites involved in adaptation in nature.

While QTL and candidate gene genetic approaches provide insight into the number and kinds of genes that contribute to local adaptation, a new type of multi-locus genetic screen may hold a more general route for the dissection of recent adaptation (HARR *et al.* 2002; SCHLOTTERER 2002; SCHLOTTERER and DIERINGER 2005). Multiple individuals from two or more populations are genotyped, typically at microsatellite markers, and the test statistics $\ln RV$ (HARR *et al.* 2002; SCHLOTTERER 2002) and $\ln RH$ (SCHLOTTERER and DIERINGER 2005) are used to identify loci that are likely to have recently experienced strong selection. The rationale for this approach is that recent genetic changes driven by selection leave a record or “genetic signature” in the genome. These include a reduction of variation and/or increased differentiation at selected loci and tightly linked regions (SMITH and HAIGH 1974). This approach is particularly powerful when applied to the study of local adaptation, especially with multiple populations recently adapted to similar environmental conditions. The divergence of populations is recent and the cause of divergence from ancestral populations is clear—adaptation to different environmental conditions. This technique has the advantage of wide applicability to nonmodel systems, including those for which controlled crosses, and thus traditional genetics, are impossible. In addition, this type of forward-genetic screen may be somewhat less biased than an analysis based on obvious phenotypic differences between groups or known candidate genes. That is, selection may be detected for genes underlying traits that might otherwise not be measured.

Because this approach can be applied to natural populations, it is ideal for the study of genes that enable a species to adapt to new environments in the wild. This was used with some success in *Drosophila* (SCHLOTTERER 2002; SCHOFEL and SCHLOTTERER 2004). A plant species that also has a remarkably diverse range of habitats is *Helianthus annuus*, the common sunflower. This species is among the most widespread and ubiquitous of the wild sunflower species in the United States (HEISER *et al.* 1969) and has populations adapted to numerous habitats, throughout North America from Mexico to Canada, and spanning both coasts. In some regions, hybridization with other native sunflower species may

have allowed introgression of favorable alleles into the *H. annuus* background, thus facilitating the *H. annuus* invasion. This appears to have occurred in both Texas (RIESEBERG *et al.* 1990) and California (CARNEY *et al.* 2000), but it is unclear if this mechanism is of widespread importance in *H. annuus*. In addition to introgression, adaptation in many areas could have involved either standing variation or new mutations in existing *H. annuus* alleles.

A particularly interesting region for examining local adaptation is the dry deserts of the American southwest. In much of the southwestern United States, *H. annuus* can grow only in more favorable years; sunflowers may have low or no germination if there is not enough moisture, and sunflower seeds may persist in the seed bank for many years (ALEXANDER and SCHRAG 2003). However, sunflowers may instead obtain a more reliable source of water by living in or near a small stream or seep or living on the edge of a pond or marshy area. While plants in these environments escape desiccation, they are likely to encounter other, novel stresses, as water supplies may be contaminated by salts or other toxins. In addition, competition may be more important in habitats with a more reliable water supply.

The focus in this study is on a comparison of *H. annuus* from the fertile plains of Kansas and eastern Colorado with populations from the arid deserts of Utah. Additionally, multiple populations within Utah were compared, placing a special emphasis on the different adaptations to stresses specific to salt marshes surrounding the Great Salt Lake as compared to life in more arid habitats. For this study, we used microsatellites located within expressed sequence tags (ESTs), which have the advantage of being tightly linked to a known, expressed gene. Using a screen of 128 microsatellite markers, we looked for genes that may underlie adaptation to salt and drought stress in *H. annuus*. We also estimated gene flow (F_{ST}) between the different populations, overall and for individual loci, at both neutral and putatively selected loci. Several distance-based tests were used to look for additional evidence confirming the presence of selection and to infer its mode of action.

MATERIALS AND METHODS

Plant material: In July 2003, fresh, young leaf tissue was collected from 48 plants in each of eight populations of *H. annuus* (Figure 1), including two plains populations (PLA1, PLA2) and two Utah desert populations (DES1, DES2) collected from typical habitats near roadsides, two brackish marsh populations from the northeast side of the Great Salt Lake (SMN1, SMN2), and two brackish marsh populations from the southwest side of the Great Salt Lake (SMS1, SMS2). Brackish marshes were identified by their proximity to visible salt flats. In addition, we identified any species within 10 m known to favor saline environments. Plants included in the survey were: *Sarcobatus vermiculatus* (greasewood), *Atriplex* sp. (saltbush), *Salicornia utahensis* (glasswort), and *Allenrolfea occidentalis* (iodine

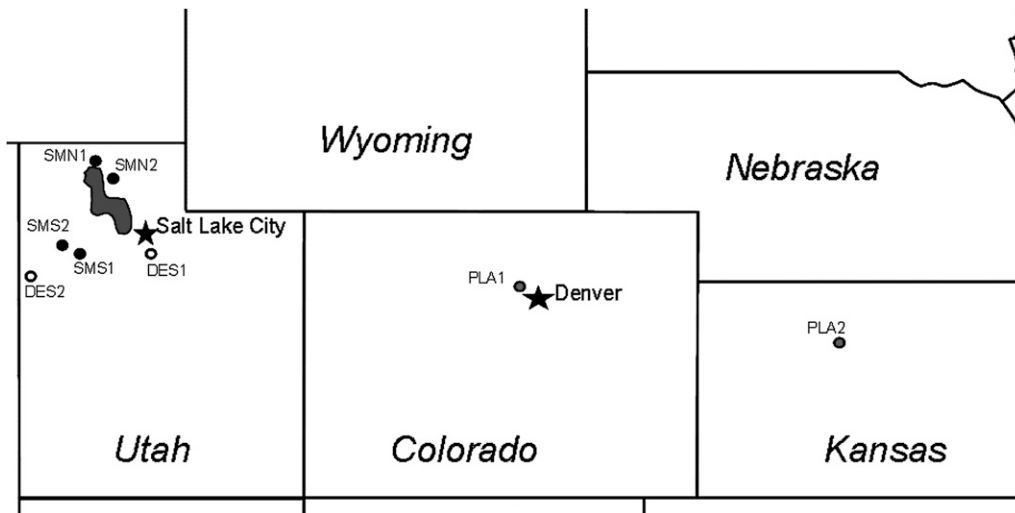


FIGURE 1.—Map of populations. Shaded circles, plains populations; open circles, Utah desert populations; solid circles, brackish marsh populations.

bush). All sites classified as saline had at least one of these four species growing nearby. DNA was extracted from 50 mg leaf tissue using DNeasy 96-well plate kits.

Microsatellite analysis: Screening was done using microsatellites derived from the sunflower *Compositae* Genome Project (CGP) database (<http://compgenomics.ucdavis.edu/>), which includes sequence information for ~70,000 sunflower ESTs. A total of 288 microsatellite markers previously developed from the CGP database (<http://cgpdb.ucdavis.edu/>) were screened for reliability and the presence of variation in two populations (PLA1 and SMN1), yielding 102 usable markers. To add to these genes, the CGP database was searched using perl scripts for additional genes with useful microsatellites. Microsatellites were defined as any occurrence of five or more repeats of di-, tri-, or tetramers. From this data set, 50 more microsatellites were chosen, and primers flanking the microsatellite were designed using Primer3, with an additional tail (GTCGTTTTACAACGTCGTG) added to the 5' end of the forward primer for hybridization with fluorescent probes (SCHUELKE 2000), and synthesized by Integrated DNA Technologies (<http://www.idtdna.com>). As before, these 50 markers were screened for reliability and variability, and 26 primer pairs were chosen, yielding a total of 128 microsatellite markers analyzed on the full data set. It should be noted that this screening method is likely to remove loci that are consistently invariant in all populations examined. Thus, loci with extremely low mutation rates or consistently high levels of selection in all populations screened would not be included. However, invariant loci cannot yield meaningful comparisons between populations, so there is little that researchers can do to avoid this effect.

PCR: During PCR, the 5' end of the forward primer was labeled using fluorescent dye (SCHUELKE 2000; WILLS *et al.* 2005). The PCR reactions were performed in volumes of 10 ml containing 2 ng DNA, 2 units Taq DNA polymerase, 0.2 μ M fluorescent dye 6FAM, VIC, NED, or PET, 0.2 μ M of each primer, 30 μ M Tricine, 50 μ M KCl, 2 μ M MgCl₂, and 100 μ M each dNTP. Fragments were amplified using a "touchdown" PCR protocol, developed to reduce nonspecific primer binding and fragment amplification (DON *et al.* 1991). An initial denaturing cycle of 3 min at 95° was followed by 10 touchdown cycles, starting with a 58° annealing temperature but dropping 1° each cycle, of 30 sec at 94°, 30 sec at the annealing temperature, and 45 sec at 72°. These 10 cycles were followed by 29 cycles of 30 sec at 95°, 30 sec at 48°, and 45 sec at 72°, with a final elongation period at 72° for 20 min.

Genotyping: Analysis of microsatellite fragment size was done on an ABI 3730 (Applied Biosystems, Foster City, CA). Multiple, nonoverlapping PCR fragments were pooled and diluted 1:60 with ddH₂O. One microliter of the diluted PCR product was added to 8.9 μ l ddH₂O and 0.1 μ l of the GenScan-500 Liz Size Standard (Applied Biosystems). Samples were denatured at 95°, snap cooled on ice, and then centrifuged to remove bubbles before loading onto the ABI 3730. For each marker, genotypes were scored using GeneMapper 3.7 (Applied Biosystems). Markers were scored automatically, with results manually corrected in cases of errors in the automatic scoring.

Selective sweeps: For each population, observed heterozygosity, expected heterozygosity, and variance in microsatellite length were calculated using the program Microsatellite Analyzer (DIERINGER and SCHLOTTERER 2003). Global F_{ST} and all possible pairwise F_{ST} -values were calculated using the Weir and Cockerham method (WEIR and COCKERHAM 1984) as implemented by GenePop 3.4 (RAYMOND and ROUSSET 1995). Additionally, global G_{ST} -values were calculated (NEI 1973) and yielded nearly identical results.

Populations were grouped into four pairs: the plains populations (PLA1, PLA2), the desert Utah populations (DES1, DES2), the southwestern Utah salt marsh populations (SMS1, SMS2), and the northeastern Utah salt marsh populations (SMN1, SMN2). These were verified as valid groupings using various algorithms including the neighbor-joining trees in PHILIP (FELSENSTEIN 2005) and Bayesian clustering methods using the program Structure (PRITCHARD *et al.* 2000), as described below. Each of the four groups was then used as a single population for the lnRH and lnRV tests described below. Additionally, for some tests, all six of the Utah populations were grouped together for some analyses (see below).

To evaluate the significance of reductions in variance or heterozygosity, we used the statistics lnRH and lnRV, developed by SCHLOTTERER and DIERINGER (2005) and SCHLOTTERER (2002), respectively. The two statistics lnRH and lnRV were calculated using the equations $\ln RV = \ln(V_{pop1}/V_{pop2})$ (SCHLOTTERER 2002) and $\ln RH1 = \ln(((1/(1 - H_{pop1}))^2 - 1)/((1/(1 - H_{pop2}))^2 - 1))$ (SCHLOTTERER and DIERINGER 2005), respectively. These statistics are generally normally distributed, and simulations have confirmed that outliers are likely to be caused by selection (SCHLOTTERER and DIERINGER 2005). Additionally, independent, sequence-based measures of selection [*e.g.*, HKA tests (HUDSON *et al.* 1987)] have confirmed the existence of adaptive alleles within regions identified as

candidates using these statistics (HARR *et al.* 2002; VIGOUROUX *et al.* 2002; DU MONT and AQUADRO 2005), as have SNP-based measures of reductions in nucleotide diversity (IHLE *et al.* 2006). It is important to note that these statistics are merely screens to emphasize candidate markers likely to be linked to genes under selection and do not indicate how close the marker is to the selected locus. Another caution in interpreting these tests is that they produce a low level of false positives (SCHLOTTERER and DIERINGER 2005). However, because the variance in repeat number and the heterozygosity of a population measure different aspects of the variation at a locus, using both statistics together lowers the rate of false positives by a factor of three (SCHLOTTERER and DIERINGER 2005). There is also a reduction in power, especially for detecting older sweeps. However, for detecting the most recent and strong selective sweeps, the combination of lnRH and lnRV statistics is as powerful as lnRV on its own, but with one-third the false-positive rate (SCHLOTTERER and DIERINGER 2005). Thus, for each locus, we calculated lnRH and lnRV statistics for several evolutionarily significant groupings. These groupings were based on the presumed relatedness among populations, as indicated by both neighbor-joining and Bayesian clustering algorithms (below). Specifically, we compared the plains to all Utah populations (desert and salt marsh) to test for differential selection associated with the invasion of the arid southwest. Within Utah, we compared the desert populations to each pair of salt marsh populations, pooled, to identify genes under selection, looking for potentially independent evolution of salt tolerance in each salt habitat. Significance of lnRH and lnRV for each comparison was calculated according to standard methods (SCHLOTTERER 2002; SCHLOTTERER and DIERINGER 2005).

Distance-based tests of selection: Several different methods use population divergence, rather than heterozygosity or variance differences, to test for selection. To test for additional evidence of selection, we used the coalescent-based simulation methods of BEAUMONT and NICHOLS (1996) and VITALIS *et al.* (2001), as well as a more robust Bayesian test (BEAUMONT and BALDING 2004).

For the BEAUMONT and NICHOLS (1996) method, we used the program FDIST2 (<http://www.rubic.rdg.ac.uk/~mab/software.html>) to calculate *P*-values for each locus. We used the program to generate 100,000 simulated loci, providing an expected, neutral distribution of F_{ST} -values. To match our experimental design, each simulation had 48 individuals per population, eight populations, and 128 loci. This method provides evidence for divergent selection by looking for outliers with higher than expected F_{ST} -values, controlling for heterozygosity (BEAUMONT and NICHOLS 1996).

The VITALIS *et al.* (2001) coalescence-based approach was performed using the program DETSEL 1.0 (<http://www.univ-montp2.fr/~genetix/detsel.html>). Null distributions were generated using the following parameters: population size before the split $N_0 = 50, 500$; mutation rate $\mu = 0.01, 0.001, 0.0001$; ancestral population size $N_c = 500, 5000, 50,000$; time since bottleneck $T_0 = 50, 500, 5000$; and time since divergence $t = 50, 500$. This test defines an expected distribution of data points, on the basis of a defined *P*-value (in this case $P = q - 1 = 0.01$), and defines those loci outside of this expected region as showing evidence for recent selection. It has the advantage of being able to take into consideration a wide range of potential parameter values simultaneously and giving results that are robust to these starting assumptions.

BEAUMONT and BALDING's (2004) hierarchical-Bayesian method was performed using the program newfst.c, which generates 2000 Markov chain Monte Carlo (MCMC) simulated loci on the basis of the distribution of F_{ST} given the data. Outliers from our data set were identified on the basis of this distribution following BEAUMONT and BALDING (2004).

Gene function: Genes for which a significant reduction in variability was found were investigated further to determine their predicted function and expression patterns. Blasts against GenBank using tBlastn on NCBI were used to find significant hits (ALTSCHUL *et al.* 1990, 1997). These genes were categorized according to known cellular and molecular functions, using gene ontology (GO) functional groups. Gene ontology is a standardized vocabulary used to describe the biological function of a gene and is an important tool for relating the putative functions of different genes in separate cells, tissues, or organisms (ASHBURNER *et al.* 2000; HARRIS *et al.* 2006). Genes were assigned to six broad functional groups: GO:0003824 (catalytic activity), GO:0030528 (transcription regulator activity), GO:0019538 (protein metabolism), GO:0005575 (cellular component), GO:0006281 (DNA repair), and GO:0005554 (molecular function unknown). In a separate analysis of biological function, genes were categorized according to whether or not they are known to be involved in response to stress (GO:0006950). On the basis of the GO functional groups, chi-square tests were used to determine whether the genes under selection have a different distribution of function than expected by chance.

Expression patterns, gene flow, and marker characteristics: Because the sunflower EST libraries were not normalized, the number of redundant ESTs should be correlated with expression level. For each gene, the number of ESTs found and the tissues they were extracted from were calculated from the sunflower EST database (<http://cgpdb.ucdavis.edu/>). Using Statistical Package for Social Sciences, chi-square tests were used to determine whether the patterns of expression in the loci with significant selective sweeps were significantly different from those expected by chance given the distribution of expression patterns in all genes analyzed. Additionally, chi-square tests were used to examine whether microsatellite markers were in different parts of selected genes (5'-UTR, coding, or 3'-UTR) than in genes with neutral patterns of evolution. One-way ANOVAs were used to look at the association between selected loci and motif size, mean repeat number, chromosomal location, and global F_{ST} -values for each marker.

Population structure: $(\delta\mu)^2$ (GOLDSTEIN *et al.* 1995; SLATKIN 1995) was used as a measure of genetic distance because of its expected linear relationship to time of divergence (GOLDSTEIN *et al.* 1995). The $(\delta\mu)^2$ measure of divergence was calculated using the software Microsatellite Analyzer (DIERINGER and SCHLOTTERER 2003) for all population pairs, and the resulting genetic distance matrix was used to build a neighbor-joining tree using the program NEIGHBOR in PHYLIP (FELSENSTEIN 2005). Additionally, we investigated the population structure in further detail, using the Bayesian clustering method of PRITCHARD *et al.* (2000), as implemented by the program Structure2.1. The number of populations (*K*) was estimated using the "no admixture" ancestral model with correlated alleles, allowing maximal population resolution, with *K* ranging from 2 to 8. Three independent runs of 100,000 MCMC generations and 20,000 generations of "burn in" were used for each value of *K*. The recent migration status of individuals within the populations was then assessed by incorporating the population information into the model.

RESULTS

Number of inferred selective sweeps: In all three comparisons, the averages for both lnRH and lnRV were very close to zero and values were normally distributed, as expected for populations of roughly equal sizes. There were no significant genomewide reductions in

variation or heterozygosity in any given population or region, although slightly higher diversity was seen in the plains populations than in Utah. Including all three comparisons, 19 genes were found with significant ($P < 0.05$) lnRH. Note that this is approximately the number expected by chance. A partially overlapping set of 19 genes showed significant levels of reduction in one or more comparisons ($P < 0.05$) using lnRV, and 17 of these loci were significant for both lnRH and lnRV. These loci thus had significant reductions in both allelic diversity and variance in allele size. For all of these loci, in the populations thought to be under selection, one allele was nearly completely fixed, with a few other alleles at very low frequency. Simulations have indicated that lnRH and lnRV measure different aspects of variation, despite high levels of correlation between the statistics, and that using both statistics together eliminates at least two-thirds of the false positives (SCHLOTTERER and DIERINGER 2005). Thus, of the 17 genes with both lnRV and lnRH significant ($P < 0.05$), 7 are expected to be false positive under conservative assumptions, given the sample size and number of comparisons (SCHLOTTERER and DIERINGER 2005). Using a higher threshold for significance ($P < 0.01$), for which only one false positive would be expected, 13 genes showed reduced variation indicating selection. Five genes showed such dramatic reductions in variation that they were significant under almost any threshold for both statistics ($P < 0.001$). In addition to the possibility of statistical false positives, these candidate genes may of course not be the targets of selection, but merely tightly linked to them. However, linkage disequilibrium is known to be extremely low in wild sunflowers (LIU and BURKE 2006), so the region swept due to selection is likely to be quite small. Despite the possible presence of a few false positives, the full set of 17 significant genes was used in further analyses, as these are at the very least good candidate loci for further study.

Of these 17 genes, 7 have been mapped in at least one study (S. KNAPP, personal communication). These 7 microsatellites map to six different chromosomes. The 2 that map to the same linkage group (Figure 2) map to opposite ends of the linkage group. Moreover, in all seven cases, the nearest markers showed no reduction of lnRH (Figure 2) or lnRV (not shown). Thus, no evidence was found for linkage between markers under selection. It is thus likely that the 17 reductions in variation and heterozygosity represent 17 distinct selective sweeps.

Of the 17 candidate genes, 2 showed reduced variation only in the plains populations (Figure 3a). Four were found to be reduced significantly across all of the Utah populations examined (Figure 3a). Five and three loci had reductions in variation only in the northeastern (Figure 3b) and southwestern (Figure 3c) brackish marsh populations, respectively. Three loci, on the other hand, showed reductions in variation in both northeastern and southwestern salt populations, but not in desert populations. It should be emphasized that this result is

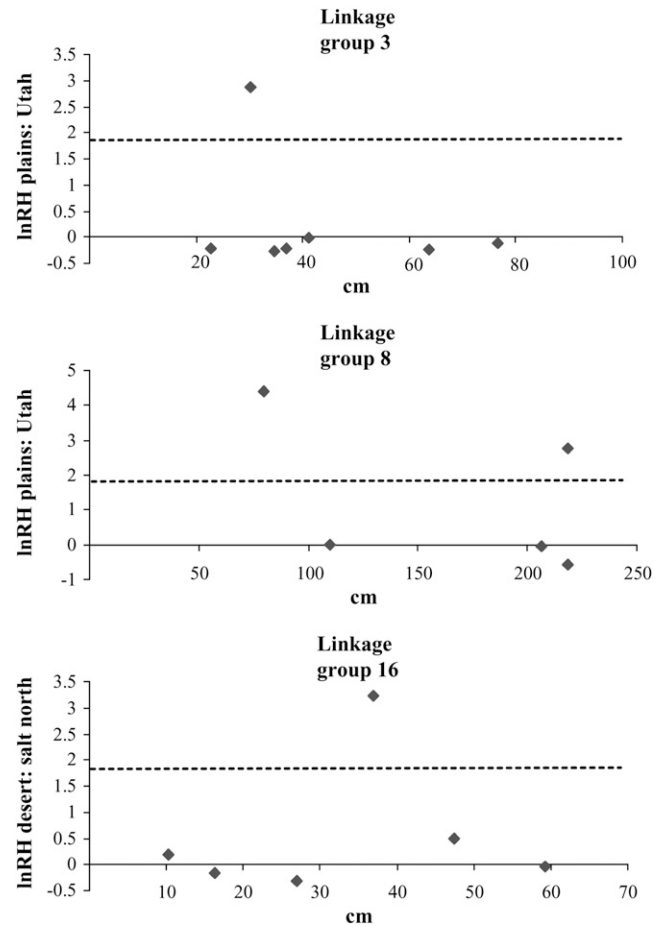


FIGURE 2.—Graphs of lnRH for mapped markers on three representative chromosomes, showing four examples of selective sweeps. Dashed line represents significance ($P < 0.05$).

not an artifact of spatial clustering of salt marsh populations; *DES1* was closer to each of the salt marsh populations than they were to each other (Figure 1), but did not show the same selection patterns. Nor is this result likely to be due to a close relationship between the two salt populations; these populations do not cluster together when grouped by genetic distance [*e.g.*, $(\delta\mu)^2$ (Figure 4) or fixation index (*i.e.*, F_{ST})]. Thus, spatially and genetically distinct populations thought to be undergoing similar selective pressures appear to have shared selective sweeps at several loci.

Whether these represent independent, parallel sweeps or a single sweep that passed through multiple populations remains to be determined. In all three cases, the same size allele was apparently swept in both populations, which would apparently make a single mutation more likely. However, this could be due to two independent mutations in the same ancestral allele as well.

Gene function: Candidate genes selected using the lnRH and lnRV statistics had a wide variety of predicted functions (Table 1). The 17 significant loci included genes with the following homology based on closest BLAST hits: four genes with catalytic or metabolic functions, five

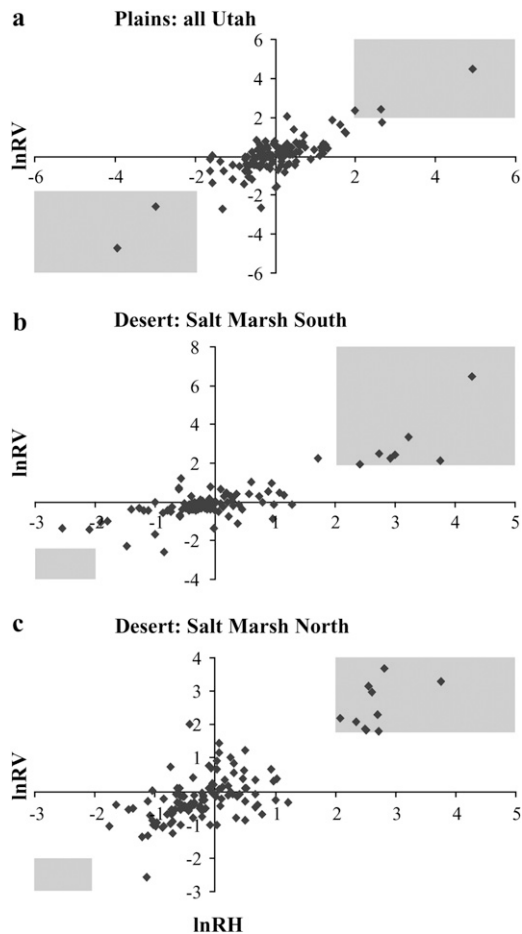


FIGURE 3.—Graphs of $\ln RV$ and $\ln RH$ for all three comparisons. Gray areas contain points significant ($P < 0.05$) for $\ln RV$ and $\ln RH$. (a) Plains populations *vs.* all Utah populations. (b) Utah desert *vs.* northeastern salt marsh populations (SMN). (c) Utah desert *vs.* southwestern salt marsh populations (SMS).

transcription factors, three cellular components, one DNA-repair gene, and four genes of unknown homology or function (Table 2). This was significantly different from the distribution expected by chance ($\chi^2 = 15.28$; 5 d.f.; $P < 0.01$), mainly due to fewer than expected genes with no known function or with kinase, ubiquitinase, or other protein metabolic function and more genes than expected with roles as transcription factors, cell structure, or catalytic functions. In addition to these molecular and structural functions, among these 17 candidate genes were 3 of the 4 known stress-response genes (GO: 0006950) from the 128 loci used, which was significantly different from that expected ($\chi^2 = 4.25$; 1 d.f.; $P < 0.05$). Of course, as stated above, it must be kept in mind that these genes are only candidate genes and may be tightly linked only to the actual locus under selection. However, the fact remains that their gene functions differ significantly from the expected distribution, indicating that they are not just a random set of genes.

Expression patterns, motifs: The expression patterns of the significant genes were found to be not significantly

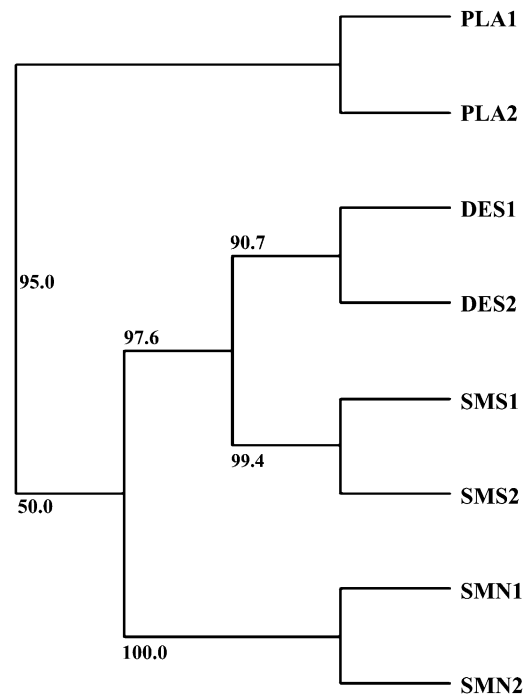


FIGURE 4.—Unrooted neighbor-joining tree showing the relationships between the eight populations used, based on genetic distances, $(\delta\mu)^2$. Numbers represent bootstrap support, from 1000 iterations, using the Phylip 3.63 program “neighbor.”

different than expected by chance, although there were marginally significantly more genes from the seed hull tissues than expected when that tissue was analyzed alone ($\chi^2 = 2.71$; 1 d.f.; $P < 0.1$). There was no evidence for any effect of motif size, motif location, mean repeat number, chromosomal location, diversity of tissue expression, or level of expression on likelihood of detecting selective sweeps.

Distance-based test of selection: Distance-based tests of selection are prone to false positives because of sensitivity to demographic history (WHITLOCK and McCAULEY 1999) and heterogeneity among loci in mutation rate (SLATKIN 1995; BALLOUX and GOUDET 2002). Nonetheless, we expect the set of loci identified by distance-based tests to be enriched for true positives. This appears to be what we find here. Fifteen of the 17 “selected” loci from the $\ln RH$ and $\ln RV$ tests also showed evidence of selection in at least one of the three distance-based tests (Table 1). Because of the sensitivity of any F_{ST} -based tests to locus-specific phenomena not related to selection, we limit our discussion to those loci that show selection in the more robust $\ln RV$ and $\ln RH$ tests, which are specifically designed to control for locus-specific phenomena (SCHLOTTERER and DIERINGER 2005). Nevertheless, the additional evidence of selection from the F_{ST} -based tests is important in bolstering the results from the more robust tests. Since each of the five tests used relies on slightly different assumptions, loci that are repeatedly found to be outside of the range expected for neutral

TABLE 1

Results of selection analyses lnRH, lnRV, fdist2, detSel, and newfst (see text), along with protein homology, putative function, and chromosomal locations for mapped markers

Contig/EST	lnRH ^a	lnRV ^a	fdist2 ^b	detSel ^c	newfst ^d	Populations	Function	Protein homology	Chromosome	cM
Contig 0291	***	**	—	***	—	SMS	Catalytic/metabolism	Secretory peroxidase	15	36.2
Contig 1731	***	***	*****	—	*	SMN	DNA binding	DNA-binding protein	13	103
Contig 1891	***	**	**	****	*	Plains	DNA binding	DNA-binding protein	—	—
Contig 2016	***	***	—	—	*	SMN	DNA binding	Heat-shock protein	16	36.8
Contig 2135	***	**	***	****	***	SMS, SMN	DNA binding	Ovule development	—	—
Contig 2281	****	***	—	—	*	SMS, SMN	Catalytic/metabolism	Catalytic/metabolism	11	39.1
Contig 2293	****	****	—	—	**	Utah	Repair of DNA	Purple acid phosphatase	8	79
Contig 2727	***	***	—	—	—	SMN	Unknown	—	3	30
Contig 3403	***	****	**	***	*	SMN	Unknown	None	—	—
Contig 3590	***	**	—	—	—	Utah	Membrane	COPT5	—	—
Contig 5468	***	***	**	—	**	SMN	Catalytic/metabolism	Cellulose synthase like	—	—
QHA19L11	***	***	***	****	**	SMS	Unknown	None	—	—
QHA6L06	*****	*****	—	—	***	SMS	Membrane	Xyloglucan galactosyltransferase	—	—
QHB12L21	***	****	***	****	**	SMN	Cell wall	None	10	77.4
QHB33B17	***	**	—	***	****	Utah	DNA binding	Squamosa promoter binding protein	8	218
QHG7N09	****	****	—	****	**	Plains	Unknown	None	—	—
QHJ18D19	**	***	—	***	***	SMS, SMN	Catalytic/metabolism	Putative dehydrogenase	—	—

* $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$, **** $P < 0.001$, ***** $P < 0.0001$.

^a Tests were performed following SCHLOTTERER (2002) and SCHLOTTERER and DIERINGER (2005).

^b The results of the program fdist2, implementing the protocol from BEAUMONT and NICHOLS (1996).

^c The results from the program detSel, which performs the calculations outlined in VITALIS *et al.* (2001).

^d Summary of the results of the program newfst.c (BEAUMONT and BALDING 2004).

loci are extremely good candidates for genes under selection.

Consistent with the results of all five tests, global F_{ST} values were significantly higher for selected loci ($P < 0.05$), as expected under divergent selection (STORZ 2005), averaging 0.11 in the neutral loci and 0.15 in the selected loci. It should be noted that all global F_{ST} values calculated were significantly > 0 (thus indicating population structure).

TABLE 2

Functional categories of genes analyzed, for all genes, for those found to be under selection, and the expected distribution

GO category	Total	Observed selected	Expected selected
Catalytic/metabolism	17	4	2
Transcription	15	5	2
Protein metabolism	18	0	2
Cellular component	6	3	1
Repair of DNA	2	1	1
Unknown	79	4	9
Total	128	17	17

Population comparisons: Using the Bayesian methods implemented by the program Structure2.1 (PRITCHARD *et al.* 2000), we were able show that the eight populations collected from were indeed distinct genetic clusters and did not show evidence for substantial recent admixture or migration. With eight populations assumed, between 95 and 100% of individuals clustered together into their appropriate population groupings, with only three populations showing even one individual with substantial genetic material from other populations. The Structure analysis was not informative regarding the true number of populations [as the number of populations, K , in the model was increased the posterior probabilities $\Pr(X|K)$ increased progressively, with no clear discontinuities]. However, the results did reveal several values for K that captured important aspects of clustering within the data. When the model was set at lower numbers of assumed populations, the populations tended to cluster according to geographic regions. With only two populations assumed, individuals collected from all six Utah populations clustered together in all three iterations, while individuals from Kansas and Colorado (PLA1 and PLA2) clustered as a separate group. With more populations assumed, populations clustered into pairs. SMS1

and SMS2 clustered together in every analysis where more than three populations were assumed. Similarly, DES1 and DES2 clustered together in at least two of three of the runs for values of K from 3 to 7. SMN1 and SMN2 seemed to be more differentiated, but did cluster together in some analyses. The two pairs of salt marsh populations did not tend to cluster together; rather, SMS1 and SMS2 tended to cluster more closely with DES1 and DES2, particularly with the geographically closer DES2. On the basis of this analysis, we conclude that the four groupings used for the lnRH and lnRV tests are valid. Also, it appears that the two salt marsh regions show little genetic relationship or gene flow and show much closer genetic relationships to geographically proximal desert populations.

DISCUSSION

The geographic extent of selective sweeps: The origin and spread of adaptive mutations has implications not only for understanding local adaptation and population-level phenomena, but also for interpreting the cohesions and divisions at the species level. For example, some authors argue that species are products of rather than units of evolution because gene flow within species often is too low to prevent differentiation through drift or local adaptation (EHRlich and RAVEN 1969; MISHLER 1999). However, even very low migration rates are sufficient for the spread of advantageous alleles (MORJAN and RIESEBERG 2004), leading RIESEBERG and BURKE (2001) to propose that selective sweeps associated with favorable mutations may allow species to be held together at some loci, while simultaneously differentiating at others.

The results presented here are broadly consistent with this proposal. While some genes sweep to fixation only within local populations, other genes sweep across a broader range of the species. Some selective sweeps define a single salt-tolerant population, three occur in spatially separated salt marsh populations, while others appear to have reduced variation across all of the Utah populations examined. Unfortunately, our design did not allow us to detect sweeps that have occurred across all populations of *H. annuus*, but another analysis utilizing comparisons with a hybridizing congener, *H. petiolaris*, indicates that specieswide selective sweeps are common as well (Y. YATABE and L. H. RIESEBERG, unpublished data).

There was a clear pattern of higher F_{ST} -values, and thus lower effective migration rates, for genes under selection than for other loci. This shows that selection may maintain higher levels of population differentiation at swept loci ($F_{ST} = 0.15$) than at other loci ($F_{ST} = 0.11$). However, for the four genes that were swept across all of Utah, there were dramatic increases in effective gene flow, such that pairwise F_{ST} -values were not significantly >0 between any Utah populations for these

genes. Thus, selection has effectively obliterated all population structure within Utah at these loci.

Although spatially separated populations undergoing similar selective pressures show evidence of divergence at some loci, they evolve in concert at others. This is illustrated by three sweeps that are exclusive to the spatially separated salt marsh populations. In two of the cases, the same allele appears to be fixed in all salt populations examined. Those selective sweeps point to the importance of an allele that spreads through both clusters of salt marsh populations despite hundreds of miles of separation. In the third case, all salt populations examined show strong reductions in allelic diversity, but the northern salt populations are fixed for a different allele than the southern populations. This is likely due to two completely independent selective sweeps, each due to a separate adaptive mutation. For all three loci, whether the coincident patterns of variation occurred through long-distance dispersal of a new mutation or parallel evolution based on ancestral variation (WOOD *et al.* 2005), strong selection at each locus is likely to be the cause of reduced variation in all populations. This is consistent with previous work looking at parallel evolution of edaphic races in *Laesthenia californica* (RAJAKARUNA *et al.* 2003).

In the present study, very few genes showed evidence for selection in the more benign and common plains populations, while relatively more genes appeared to be under selection in the more stressful and unusual environment of the salt marshes. The generality of this pattern of increased rates of molecular evolution in novel, unusual, or recently colonized environments remains to be determined. This may explain the widely reported pattern of increased phenotypic evolution in novel environments (SINGER *et al.* 1993; HUEY *et al.* 2000; REZNICK and GHALAMBOR 2001; SCHOFEL and SCHLOTTERER 2004; YEY 2004; BONSER and GEBER 2005).

Caveats: Because they are natural, not experimental, populations, the populations studied differ in many characteristics other than precipitation and salinity. Thus, some of the differentiation observed could have been due to some other selective force, such as herbivory, pathogens, or disturbance regime. Moreover, it is important to keep in mind that selection is not the only way that the distribution of variation can be changed at particular loci—reduced variation or increased differentiation can result from chance alone, *e.g.*, genetic drift, bottlenecks, or founder events. As in this study, false positives can be minimized by using tests that account for population differences, looking for locus-specific effects, and by using multiple measures of variability, each of which measures different parameters and relies on different assumptions, *e.g.*, heterozygosity and variance in allele size (SCHLOTTERER and DIERINGER 2005). Thus, loci with significant reductions in variability and/or increased differentiation are good candidates for further study and are quite likely to have been under recent selection (SCHLOTTERER 2002; VIGOUROUX *et al.* 2002).

Of course, the microsatellites themselves are not likely to be the target of selection, but are merely tightly linked to it. Since each of the microsatellite loci used is located within ESTs, this gives us one good candidate gene that is known to be expressed and is tightly linked to each locus. In many of the cases so far examined, selective sweeps have affected only a very small region, potentially containing only one or a few genes (CLARK *et al.* 2004; DUMONT and AQUADRO 2005; STORZ 2005), except in the case of extremely strong selection (KOHN *et al.* 2000; WOOTTON *et al.* 2002). Moreover, recent estimates of linkage disequilibrium in wild *H. annuus* as negligible after 200–300 bp (LIU and BURKE 2005) indicate that in most cases the region of reduced variability is likely to contain only a single gene or domain. Results for the domesticated *H. annuus* cultivars showed little linkage disequilibrium after 1800–1900 bp, which may be more reasonable for the more recently derived wild populations or for genes under strong selection. In any case, even if linkage disequilibrium is much higher, when a marker is found to be in a swept region, that marker is likely to be close to the actual locus under selection. Because of the small size of swept regions, however, even with the substantial number of loci we used, we are likely to have found only a very small fraction of the genes under selection in these *H. annuus* populations.

Nature of selected genes: The nature of this study precludes speculation on the type of genes that may sweep across an entire species range; for that a broader sampling of populations from multiple, closely related species would be necessary. However, this work provides insight into the percentage and types of genes involved in local and regional selective sweeps. In this study, we found that 1.5–6% of the loci examined, depending on the environment, showed significant reductions in heterozygosity and variance consistent with a recent selective sweep. Most of these loci showed additional evidence for selection on the basis of other statistical tests as well. Thus, while the large number of genes apparently under selection is somewhat surprising, our results are based on many independent tests of selection. Moreover, they are consistent with a growing body of evidence that a large percentage of a typical genome shows evidence for recent selection (DILLER *et al.* 2002; HARR *et al.* 2002; VIGOUROUX *et al.* 2002, 2005; KAUER *et al.* 2003; KAYSER *et al.* 2003; SCHOFEL and SCHLOTTERER 2004; STORZ *et al.* 2004; VASEMAGI *et al.* 2005; WRIGHT *et al.* 2005).

The genes under selection exhibit a wide range of putative functions, including DNA binding, DNA repair, formation of cell walls, and catalyzing biosynthetic or metabolic reactions, as well as a number of unknown functions. The phenotypes associated with alternate alleles remain to be determined, but these results imply that adaptation to salt and drought is polygenic. Additionally, a substantial proportion of the selected genes (18%) show homology to known stress-response genes, from the GO classification, thus indicating that responses to

abiotic stresses are indeed a major component of adaptation to the habitats examined. It should be emphasized, once again, that these genes are only candidate genes showing tight linkage to the markers identified as being under selection.

The three putative stress response genes apparently under selection have a variety of putative functions. One shows 85% identity to a secretory peroxidase thought to be involved in response to several different stresses (TOGNOLI *et al.* 2002) and that is highly homologous to a peroxidase upregulated in response to salt stress in cotton (RITTER *et al.* 1993). Another shows homology to a heat-shock protein from *Arabidopsis*. The third shows homology to a putative purple acid phosphatase, which is involved in response to phosphate starvation (DUFF *et al.* 1994) and potentially other stresses (OLCZAK *et al.* 2003). In addition to the genes classified as “stress response” genes in GO, some of the selected genes have known or putative functions in adaptation to salinity or drought. For instance, one candidate locus shows homology to tocopherol cyclase, which can enhance tolerance of oxidative stress (KANWISCHER *et al.* 2005).

In addition to these potential stress-response genes, a number of genes appear to be involved in determining the structure or function of the cell wall and cellular membrane. These include two ESTs with homology to membrane proteins (one of which is an ion channel) and one with homology to a cellulose-synthase-like protein. Other genes of interest include a *squamosa* promoter-binding protein, part of a family of DNA-binding proteins found only in plants (KLEIN *et al.* 1996), which regulates the expression of the MADS-box gene *squamosa* (HUIJSER *et al.* 1992), and AINTEGUMENTA (ELLIOTT *et al.* 1996; KLUCHER *et al.* 1996), an APETALLA-2 like gene involved in floral organ development.

In addition to homology, the expression patterns of the genes in our database give us some idea of the types of traits potentially involved in adaptation. The prevalence of seed hull genes as candidates for genes under selection is interesting, as several aspects of seed morphology are extremely variable between populations, and, on the basis of our observations, are likely to be under selection. Seed coat coloration is one of the most variable traits between wild sunflower populations, with seeds colored to match the soil surface in every population examined (R. RANDEL, personal communication). Germination characteristics, which also vary highly between populations of *H. annuus*, are also likely to be affected by seed coat morphology and composition. Thus, these traits merit further study in these populations, with particular emphasis on the role of allelic variation at the selected loci.

Conclusions and future directions: In summary, this study provides strong evidence that a large proportion of the genome is involved in adaptation and that selective sweeps play a major role in determining gene flow and genetic differences between populations. In

terms of the types of genes potentially involved in adaptation, transcription factors appear to be quite important, but so do genes involved in catalyzing biosynthetic or metabolic reactions. In contrast, we found no evidence for selection on any of the protein kinases or other genes involved in protein synthesis, degradation, or regulation in our data set, despite their prevalence as the most common functional group.

Currently underway, an effort to sequence alleles of these candidate genes from these populations should provide insight into the differences between adaptive and ancestral sequences and may help us understand the origin of these adaptive alleles—either as new mutants or as part of the standing variation prior to selection. By gaining a greater understanding of the molecular basis for local adaptation, we hope to understand the contributions of standing variation, mutation, and migration in allowing species to colonize new habitats.

We thank Sarah Elmendorf, Briana Gross, Daniel Ortiz-Barrientos, Jared Strassburg, Yuval Sapir, Yoko Yatabe, and three anonymous reviewers for helpful advice and comments. This research was funded by National Science Foundation (NSF) (DBI0421630) and National Institutes of Health (GM059065) grants to L.H.R., as well as by NSF predoctoral and Integrative Graduate Education and Research Traineeship fellowships to N.C.K.

LITERATURE CITED

- ALEXANDER, H. M., and A. M. SCHRAG, 2003 Role of soil seed banks and newly dispersed seeds in population dynamics of the annual sunflower, *Helianthus annuus*. *J. Ecol.* **91**: 987–998.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. H. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- ANGERT, A. L., and D. W. SCHEMSKE, 2005 The evolution of species' distributions: reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. *Evolution* **59**: 1671–1684.
- ARNQVIST, G., 1992 Spatial variation in selective regimes—sexual selection in the water strider *Gerris odontogaster*. *Evolution* **46**: 914–929.
- ASHBURNER, M., C. A. BALL, J. A. BLAKE, D. BOTSTEIN, H. BUTLER *et al.*, 2000 Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**: 25–29.
- BALLOUX, F., and J. GOUDET, 2002 Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Mol. Ecol.* **11**: 771–783.
- BEAUMONT, M. A., and D. J. BALDING, 2004 Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**: 969–980.
- BEAUMONT, M. A., and R. A. NICHOLS, 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **263**: 1619–1626.
- BONSER, S. P., and M. A. GEBER, 2005 Growth form evolution and shifting habitat specialization in annual plants. *J. Evol. Biol.* **18**: 1009–1018.
- BRADSHAW, H. D., K. G. OTTO, B. E. FREWEN, J. K. MCKAY and D. W. SCHEMSKE, 1998 Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149**: 367–382.
- CALLAHAN, H. S., and M. PIGLIUCCI, 2002 Shade-induced plasticity and its ecological significance in wild populations of *Arabidopsis thaliana*. *Ecology* **83**: 1965–1980.
- CARNEY, S. E., K. A. GARDNER and L. H. RIESEBERG, 2000 Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (*Helianthus*). *Evolution* **54**: 462–474.
- CLARK, R. M., E. LINTON, J. MESSING and J. F. DOEBLEY, 2004 Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl. Acad. Sci. USA* **101**: 700–707.
- CLAUSEN, J., D. D. KECK and W. M. HIESEY, 1940 Experimental studies on the nature of species. I. The effect of varied environments on western North American plants. Pub. 520, Carnegie Institution, Washington, DC.
- CLAUSEN, J., D. D. KECK and W. M. HIESEY, 1948 Experimental studies on the nature of species. III. Environmental responses of climatic races of *Achillea*. Pub. 581, Carnegie Institution, Washington, DC.
- CORK, J. M., and M. D. PURUGGANAN, 2005 High-diversity genes in the Arabidopsis genome. *Genetics* **170**: 1897–1911.
- DIERINGER, D., and C. SCHLOTTERER, 2003 MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**: 167–169.
- DILLER, K. C., W. A. GILBERT and T. D. KOCHER, 2002 Selective sweeps in the human genome: a starting point for identifying genetic differences between modern humans and chimpanzees. *Mol. Biol. Evol.* **19**: 2342–2345.
- DON, R. H., P. T. COX, B. J. WAINWRIGHT, K. BAKER and J. S. MATTICK, 1991 Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**: 4008.
- DONOHUE, K., D. MESSIQUA, E. H. PYLE, M. S. HESCHEL and J. SCHMITT, 2000 Evidence of adaptive divergence in plasticity: density- and site-dependent selection on shade-avoidance responses in *Impatiens capensis*. *Evolution* **54**: 1956–1968.
- DUDLEY, S. A., 1996 The response to differing selection on plant physiological traits: evidence for local adaptation. *Evolution* **50**: 103–110.
- DUDLEY, S. A., and J. SCHMITT, 1996 Testing the adaptive plasticity hypothesis: density-dependent selection on manipulated stem length in *Impatiens capensis*. *Am. Nat.* **147**: 445–465.
- DUFF, S. M. G., G. SARATH and W. C. PLAXTON, 1994 The role of acid phosphatases in plant phosphorus metabolism. *Physiol. Plant.* **90**: 791–800.
- DUMONT, V. B., and C. F. AQUADRO, 2005 Multiple signatures of positive selection downstream of *notch* on the X chromosome in *Drosophila melanogaster*. *Genetics* **171**: 639–653.
- EHRlich, P. R., and P. H. RAVEN, 1969 Differentiation of populations. *Science* **165**: 1228–1232.
- ELLIOTT, R. C., A. S. BETZNER, E. HUTTNER, M. P. OAKES, W. TUCKER *et al.*, 1996 *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* **8**: 155–168.
- ETTERSON, J. R., 2004 Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. 1. Clinal patterns of selection along an environmental gradient in the great plains. *Evolution* **58**: 1446–1458.
- FELSENSTEIN, J., 2005 Phylogeny inference package (Version 3.2). *Cladistics* **5**: 164–166.
- FRY, J. D., S. V. NUZHIDIN, E. G. PASYUKOVA and T. F. C. MCKAY, 1998 QTL mapping of genotype-environment interaction for fitness in *Drosophila melanogaster*. *Genet. Res.* **71**: 133–141.
- GOLDSTEIN, D. B., A. R. LINARES, L. L. CAVALLISFORZA and M. W. FELDMAN, 1995 Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* **92**: 6723–6727.
- GRIFFITH, T. M., and M. A. WATSON, 2005 Stress avoidance in a common annual: reproductive timing is important for local adaptation and geographic distribution. *J. Evol. Biol.* **18**: 1601–1612.
- GROSS, B. L., N. C. KANE, C. LEXER, F. LUDWIG, D. M. ROSENTHAL *et al.*, 2004 Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *Am. Nat.* **164**: 145–156.
- HARR, B., M. KAUER and C. SCHLOTTERER, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **99**: 12949–12954.
- HARRIS, M. A., J. I. CLARK, A. IRELAND, J. LOMAX, M. ASHBURNER *et al.*, 2006 The gene ontology (GO) project in 2006. *Nucleic Acids Res.* **34**: D322–D326.

- HEISER, C. B., D. M. SMITH, S. CLEVENGER and W. C. MARTIN, 1969 The North American sunflowers (*Helianthus*). Mem. Torrey Bot. Club **22**: 1–218.
- HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- HUEY, R. B., G. W. GILCHRIST, M. L. CARLSON, D. BERRIGAN and L. SERRA, 2000 Rapid evolution of a geographic cline in size in an introduced fly. *Science* **287**: 308–309.
- HUIJSER, P., J. KLEIN, W. E. LONNIG, H. MEIJER, H. SAEDLER *et al.*, 1992 Bracteomania, an inflorescence anomaly, is caused by the loss of function of the *MADS*-box gene *SQUAMOSA* in *Antirrhinum majus*. *EMBO J.* **11**: 1239–1249.
- IHLE, S., I. RAVAOARIMANANA, M. THOMAS and D. TAUTZ, 2006 An analysis of signatures of selective sweeps in natural populations of the house mouse. *Mol. Biol. Evol.* **23**: 790–797.
- KANWISCHER, M., S. PORFIROVA, E. BERGMULLER and P. DORMANN, 2005 Alterations in tocopherol cyclase activity in transgenic and mutant plants of *Arabidopsis* affect tocopherol content, tocopherol composition, and oxidative stress. *Plant Physiol.* **137**: 713–723.
- KAUER, M. O., D. DIERINGER and C. SCHLOTTERER, 2003 A microsatellite variability screen for positive selection associated with the “Out of Africa” habitat expansion of *Drosophila melanogaster*. *Genetics* **165**: 1137–1148.
- KAYSER, M., S. BRAUER and M. STONEKING, 2003 A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.* **20**: 893–900.
- KIM, S. C., and L. H. RIESEBERG, 1999 Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics* **153**: 965–977.
- KINGSOLVER, J. G., H. E. HOEKSTRA, J. M. HOEKSTRA, D. BERRIGAN, S. N. VIGNIERI *et al.*, 2001 The strength of phenotypic selection in natural populations. *Am. Nat.* **157**: 245–261.
- KLEIN, J., H. SAEDLER and P. HUIJSER, 1996 A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. *Mol. Gen. Genet.* **250**: 7–16.
- KLUCHER, K. M., H. CHOW, L. REISER and R. L. FISCHER, 1996 The *AINTEGUMENTA* gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene *APETALA2*. *Plant Cell* **8**: 137–153.
- KOHN, M. H., H. J. PELZ and R. K. WAYNE, 2000 Natural selection mapping of the warfarin-resistance gene. *Proc. Natl. Acad. Sci. USA* **97**: 7911–7915.
- KOHN, M. H., H. J. PELZ and R. K. WAYNE, 2003 Locus-specific genetic differentiation at *Rw* among warfarin-resistant rat (*Rattus norvegicus*) populations. *Genetics* **164**: 1055–1070.
- KROYMANN, J., and T. MITCHELL-OLDS, 2005 Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**: 95–98.
- LEXER, C., M. E. WELCH, J. L. DURPHY and L. H. RIESEBERG, 2003 Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Mol. Ecol.* **12**: 1225–1235.
- LIU, A., and J. M. BURKE, 2006 Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* **173**: 321–330.
- MCKAY, J. K., J. H. RICHARDS and T. MITCHELL-OLDS, 2003 Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* **12**: 1137–1151.
- MISHLER, B. D., 1999 Getting rid of species?, pp. 307–315 in *New Interdisciplinary Essays*, edited by R. WILSON. MIT Press, Cambridge, MA.
- MITCHELL-OLDS, T., 1996 Genetic constraints on life-history evolution: quantitative-trait loci influencing growth and flowering in *Arabidopsis thaliana*. *Evolution* **50**: 140–145.
- MORJAN, C. L., and L. H. RIESEBERG, 2004 How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* **13**: 1341–1356.
- NAGY, E. S., 1997 Selection for native characters in hybrids between two locally adapted plant subspecies. *Evolution* **51**: 1469–1480.
- NEI, M., 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**: 3321–3323.
- OLCZAK, M., B. MORAWIECKA and W. WATOREK, 2003 Plant purple acid phosphatases—genes, structures and biological function. *Acta Biochim. Pol.* **50**: 1245–1256.
- PETTI, C., and J. D. THOMPSON, 1998 Phenotypic selection and population differentiation in relation to habitat heterogeneity in *Arrhenatherum elatius* (Poaceae). *J. Ecol.* **86**: 829–840.
- PIGLIUCCI, M., and J. SCHMITT, 1999 Genes affecting phenotypic plasticity in *Arabidopsis*: pleiotropic effects and reproductive fitness of photomorphogenic mutants. *J. Evol. Biol.* **12**: 551–562.
- PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000 Inference of population structure from multilocus genotype data. *Genetics* **155**: 945–959.
- PURUGGANAN, M. D., and J. I. SUDDITH, 1998 Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: non-rare evolution and naturally occurring variation in floral homeotic function. *Proc. Natl. Acad. Sci. USA* **95**: 8130–8134.
- RAJAKARUNA, N., B. G. BALDWIN, R. CHAN, A. M. DESROCHERS, B. A. BOHM *et al.*, 2003 Edaphic races and phylogenetic taxa in the *Lasthenia californica* complex (Asteraceae: Heliantheae): an hypothesis of parallel evolution. *Mol. Ecol.* **12**: 1675–1679.
- RAMSEY, J., H. D. BRADSHAW and D. W. SCHEMSKE, 2003 Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**: 1520–1534.
- RAYMOND, M., and F. ROUSSET, 1995 Genepop (version-1.2). Population genetics software for exact tests and ecumenicism. *J. Hered.* **86**: 248–249.
- REZNICK, D. N., and C. K. GHALAMBOR, 2001 The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112**: 183–198.
- RIESEBERG, L. H., and J. M. BURKE, 2001 The biological reality of species: gene flow, selection, and collective evolution. *Taxon* **50**: 47–67.
- RIESEBERG, L. H., S. BECKSTROMSTERNBERG and K. DOAN, 1990 *Helianthus annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal-RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proc. Natl. Acad. Sci. USA* **87**: 593–597.
- RITTER, D., R. D. ALLEN, N. TROLINDER, D. W. HUGHES and G. A. GALAU, 1993 Cotton cotyledon cDNA encoding a peroxidase. *Plant Physiol.* **102**: 1351.
- SARIGORLA, M., M. E. PE, D. L. MULCAHY and E. OTTAVIANO, 1992 Genetic dissection of pollen competitive ability in maize. *Heredity* **69**: 423–430.
- SCHEMSKE, D. W., 1984 Population structure and local selection in *Impatiens pallida* (Balsaminaceae), a selfing annual. *Evolution* **38**: 817–832.
- SCHEMSKE, D. W., and H. D. BRADSHAW, 1999 Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci. USA* **96**: 11910–11915.
- SCHLOTTERER, C., 2002 A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**: 753–763.
- SCHLOTTERER, C., and D. DIERINGER, 2005 A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity, pp. 55–64 in *Selective Sweep*, edited by D. NURMINSKY. Landes Bioscience, Georgetown, TX.
- SCHOFEL, G., and C. SCHLOTTERER, 2004 Patterns of microsatellite variability among X chromosomes and autosomes indicate a high frequency of beneficial mutations in non-African *D. simulans*. *Mol. Biol. Evol.* **21**: 1384–1390.
- SCHUELKE, M., 2000 An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **18**: 233–234.
- SINGER, M. C., C. D. THOMAS and C. PARMESAN, 1993 Rapid human-induced evolution of insect host associations. *Nature* **366**: 681–683.
- SLATKIN, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- SMITH, J. M., and J. HAIGH, 1974 Hitch-hiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- STANTON, M. L., and C. GALEN, 1997 Life on the edge: adaptation versus environmentally mediated gene flow in the snow buttercup, *Ranunculus adoneus*. *Am. Nat.* **150**: 143–178.
- STINGHCOMBE, J. R., A. L. CAICEDO, R. HOPKINS, C. MAYS, E. W. BOYD *et al.*, 2005 Vernalization sensitivity in *Arabidopsis thaliana* (brassicaceae): the effects of latitude and FLC variation. *Am. J. Bot.* **92**: 1701–1707.
- STORZ, J. F., 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol. Ecol.* **14**: 671–688.

- STORZ, J. F., B. A. PAYSEUR and M. W. NACHMAN, 2004 Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**: 1800–1811.
- TOGNOLLI, M., C. PENEL, H. GREPPIN and P. SIMON, 2002 Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. *Gene* **288**: 129–138.
- TURESSON, G., 1922 The genotypical response of the plant species to the habitat. *Hereditas* **3**: 211–350.
- UNGERER, M. C., S. S. HALLDORS-DOTTIR, J. L. MODLISZEWSKI, T. F. C. MACKAY and M. D. PURUGGANAN, 2002 Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* **160**: 1133–1151.
- UNGERER, M. C., C. R. LINDER and L. H. RIESEBERG, 2003 Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. *Genetics* **163**: 277–286.
- VASEMAGI, A., J. NILSSON and C. R. PRIMMER, 2005 Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol. Biol. Evol.* **22**: 1067–1076.
- VERHOEVEN, K. J. F., T. K. VANHALA, A. BIERE, E. NEVO and J. M. M. VAN DAMME, 2004 The genetic basis of adaptive population differentiation: a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution* **58**: 270–283.
- VIGOUROUX, Y., M. McMULLEN, C. T. HITTINGER, K. HOUGHINS, L. SCHULZ *et al.*, 2002 Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl. Acad. Sci. USA* **99**: 9650–9655.
- VIGOUROUX, Y., S. MITCHELL, Y. MATSUOKA, M. HAMBLIN, S. KRESOVICH *et al.*, 2005 An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* **169**: 1617–1630.
- VITALIS, R., K. DAWSON and P. BOURSOT, 2001 Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**: 1811–1823.
- WAYNE, M. L., and L. M. MCINTYRE, 2002 Combining mapping and arraying: an approach to candidate gene identification. *Proc. Natl. Acad. Sci. USA* **99**: 14903–14906.
- WEINIG, C., 2000 Differing selection in alternative competitive environments: shade-avoidance responses and germination timing. *Evolution* **54**: 124–136.
- WEINIG, C., L. A. DORN, N. C. KANE, Z. M. GERMAN, S. S. HAHDORS-DOTTIR *et al.*, 2003a Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165**: 321–329.
- WEINIG, C., J. R. STINCHCOMBE and J. SCHMITT, 2003b QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Mol. Ecol.* **12**: 1153–1163.
- WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- WHITLOCK, M. C., and D. E. MCCAULEY, 1999 Indirect measures of gene flow and migration. *Heredity* **82**: 117–125.
- WILLS, D. M., M. L. HESTER, A. Z. LIU and J. M. BURKE, 2005 Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus*. *Theor. Appl. Genet.* **110**: 941–947.
- WOOD, T. E., J. M. BURKE and L. H. RIESEBERG, 2005 Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**: 157–170.
- WOOTTON, J. C., X. R. FENG, M. T. FERDIG, R. A. COOPER, J. B. MU *et al.*, 2002 Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**: 320–323.
- WRIGHT, S. I., I. V. BI, S. G. SCHROEDER, M. YAMASAKI, J. F. DOEBLEY *et al.*, 2005 The effects of artificial selection of the maize genome. *Science* **308**: 1310–1314.
- YEH, P. J., 2004 Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution* **58**: 166–174.

Communicating editor: A. H. PATERSON