American Journal of Botany 93(7): 953–962. 2006.

SPECIAL PAPER

THE MOLECULAR GENETIC BASIS OF PLANT ADAPTATION[1]

IAN M. EHRENREICH[2] AND MICHAEL D. PURUGGANAN[3,4]

[2]Department of Genetics, Box 7614, North Carolina State University, Raleigh, North Carolina 27695 USA; and [3]Department of Biology and Center for Comparative, Functional Genomics, New York University, 100 Washington Square East, New York, New York 10003 USA

How natural selection on adaptive traits is filtered to the genetic level remains largely unknown. Theory and quantitative trait locus (QTL) mapping have provided insights into the number and effect of genes underlying adaptations, but these results have been hampered by questions of applicability to real biological systems and poor resolution, respectively. Advances in molecular technologies have expedited the cloning of adaptive genes through both forward and reverse genetic approaches. Forward approaches start with adaptive traits and attempt to characterize their underlying genetic architectures through linkage disequilibrium mapping, QTL mapping, and other methods. Reverse screens search large sequence data sets for genes that possess the signature of selection. Though both approaches have been successful in identifying adaptive genes in plants, very few, if any, of these adaptations' molecular bases have been fully resolved. The continued isolation of plant adaptive genes will lead to a more comprehensive understanding of natural selection's effect on genes and genomes.

Key words: evolutionary genetics; genetic variation; genomics; natural selection; plant molecular evolution; polymorphisms.

THE NATURE OF ADAPTATIONS

Since Charles Darwin first postulated adaptation's central evolutionary role, much has been learned about how adaptations occur (1859). A major gap in our understanding remains, however, in connecting adaptive traits to their underlying molecular bases. Bridging this gap by isolating adaptive genes is important not only for discerning the evolutionary histories of individual traits, but also for clarifying how selection on phenotypes influences genetic and genomic changes.

The precise nature of the genetic architecture of adaptation—the number and effect of the genetic changes underlying adaptive traits—has proven both theoretically and empirically challenging to estimate. Ronald Fisher was the first to address this topic theoretically through his "geometric model" of phenotypic change and adaptation (Fisher, 1930). He concluded that the probability a mutation will be adaptive is nearly 50% for mutations of infinitesimally small effects and approximately zero for mutations of very large effects. Fisher and others used these results to argue that adaptation occurs through the accumulation of many beneficial mutations of small effect (Orr, 2005a). Motoo Kimura (1983) later challenged Fisher's findings, noting that the substitution rate of advantageous mutations under positive selection is not just dependent on the probability that a mutation is advantageous, but also on the probability of the mutation's fixation. He proffered that, relative to major effect mutations, minor effect mutations are more likely to be beneficial, but less likely to fix in a population. Kimura's results supported intermediate effect mutations as the most likely mutational class to underlie adaptations.

More recently, researchers have attempted to model the genetic basis of adaptation by focusing on DNA or protein sequence evolution (see Orr, 2005a, b for a description of this research area). These studies have provided evidence for a small number of sequence changes occurring during the adaptive evolution of a gene (Gillespie, 1991), as well as support for large relative fitness increases after the substitution of a beneficial mutation (Orr, 1998). Together, these results imply that major effect mutations may be important to adaptive evolution.

Empirical results, primarily from quantitative trait locus (QTL) mapping experiments, have been largely consistent with theoretical predictions (this has been discussed elsewhere, e.g., Remington and Purugganan, 2003). QTL mapping experiments have shown that the number and effect of loci controlling adaptive plant traits are variable with anywhere from a couple to many QTLs detectable. However, these QTL results cannot be taken entirely at face value because the degree to which QTLs represent single or multiple loci in plants is unresolved.

Though the genetic architecture of adaptation has received much discussion, the genetic dissection of more adaptive traits is necessary to expose any generalities about the molecular basis of adaptive evolution. Whether adaptation typically proceeds through changes in regulatory or structural genes, whether certain types of mutations are most commonly utilized by natural selection, and whether the number and effect of loci that underlie adaptations are typically large or small are questions that cannot be satisfactorily answered with existing data. Pinpointing the genes that underlie adaptations will elucidate how present adaptations have evolved and will facilitate a broader understanding of how adaptations arise at the molecular level.

[1] Manuscript received 25 February 2006; revision accepted 1 May 2006.

The authors would like to thank members of the Purugganan laboratory and two anonymous reviewers for a critical reading of the manuscript. This work was funded in part by a GAANN Fellowship from the U.S. Department of Education to I. M. E. and grants from the U.S. National Science Foundation's Frontiers in Integrated Biological Research and Plant Genome Research Programs to M. D. P.

[4] Author for correspondence (e-mail: michaelp@unity.ncsu.edu), phone: (919) 515-1761

IDENTIFYING PLANT ADAPTIVE GENES

A variety of methods exist for mapping genes involved in plant adaptations. Typically the methods used to map these genes attempt to detect natural selection at the molecular level or to find statistical associations between polymorphisms and adaptive traits. These techniques span many levels of genomic scale and can be used to connect adaptive traits to specific genes and polymorphisms (Fig. 1). Technologies for studying gene expression have become an additional tool for mapping plant adaptive genes, especially those that exhibit presence/absence polymorphisms in nature. We discuss these methods and their merits for use in plants.

Detecting selection from molecular data—Adaptations are shaped by selection, which can leave a distinctive imprint on the levels and patterns of nucleotide variation in an organism's genome. Numerous statistical tests that use molecular variation data to identify genes that bear the signature of selection exist (Nielsen, 2001). The null hypothesis for these tests is often that the observed genetic variation is consistent with selective neutrality at the locus of interest (Kimura, 1983) and that significant departures from this neutral expectation may be indicative of the action of selection.

One class of tests for selection examines the frequencies of single nucleotide polymorphisms (SNPs) in a sample of sequenced alleles. Sequences that have evolved neutrally are expected to display a different pattern of SNPs than those that have experienced selection (Tajima, 1989). For example, positively selected alleles that have recently swept to fixation should possess an excess of low-frequency SNPs (Maynard-Smith and Haigh, 1974; Kaplan et al., 1989). In contrast, some modes of balancing selection create an excess of intermediate frequency SNPs (Hudson and Kaplan, 1988). These tests, such as those proposed by Tajima (1989), Fu and Li (1993), and Fay and Wu (2000), examine whether the pattern of SNPs at a given gene is consistent with neutrality.

Another class of tests compares levels of polymorphism within a gene to levels of divergence at that gene between the species of interest and a closely related outgroup species. These tests are founded in the expectation that polymorphism and divergence levels at a locus should be proportional under neutrality. The Hudson–Kreitman–Aguade test (the HKA test; Hudson et al., 1987), which compares polymorphism and divergence at a gene of interest to one or more neutral reference loci, is the most commonly used form of this test. A similar test in this class, the McDonald–Kreitman (MK; McDonald and Kreitman, 1991) test, differentiates between polymorphisms or substitutions that affect the protein sequence encoded by a gene and those that do not.

An indicator of selection similar to the MK test examines $d_n : d_s$ (or $K_a : K_s$) ratios in protein-coding genes, where d_n (or K_a) is the nonsynonymous substitution rate and d_s (or K_s) is the synonymous substitution rate for a particular gene (Nei and Gojobori, 1986). Under neutrality, the $d_n : d_s$ ratio of a gene is expected to equal 1, and departures from this expectation can be indicative of selection (see Nielsen, 2001, for a more detailed discussion). The $d_n : d_s$ ratios can be constructed for orthologous sequences obtained from multiple species or individuals or for duplicate loci. More sophisticated tests using codon-based models have also been successful in identifying specific amino acid positions in a protein that show a history of positive selection (Yang, 1997).

A final class of tests relies on strong divergence between particular genes in populations or species as an indicator of selection. This is the basis of the Lewontin–Krakauer test, which tests whether the variance of F_{ST} estimates from different loci sampled from multiple populations is larger than what might be expected by chance (Lewontin and Krakauer, 1973). Significant results from this test may be indicative of selection-driven population divergence.

Many of the tests described in this section must be used with caution as patterns suggestive of selection can also arise from demographic effects (Hein et al., 2004). It may be possible to control for demographic effects by using empirical distributions of test statistics obtained from genome-wide sequencing projects, rather than distributions from theoretical models, to assess statistical significance (Luikart et al., 2003). This is

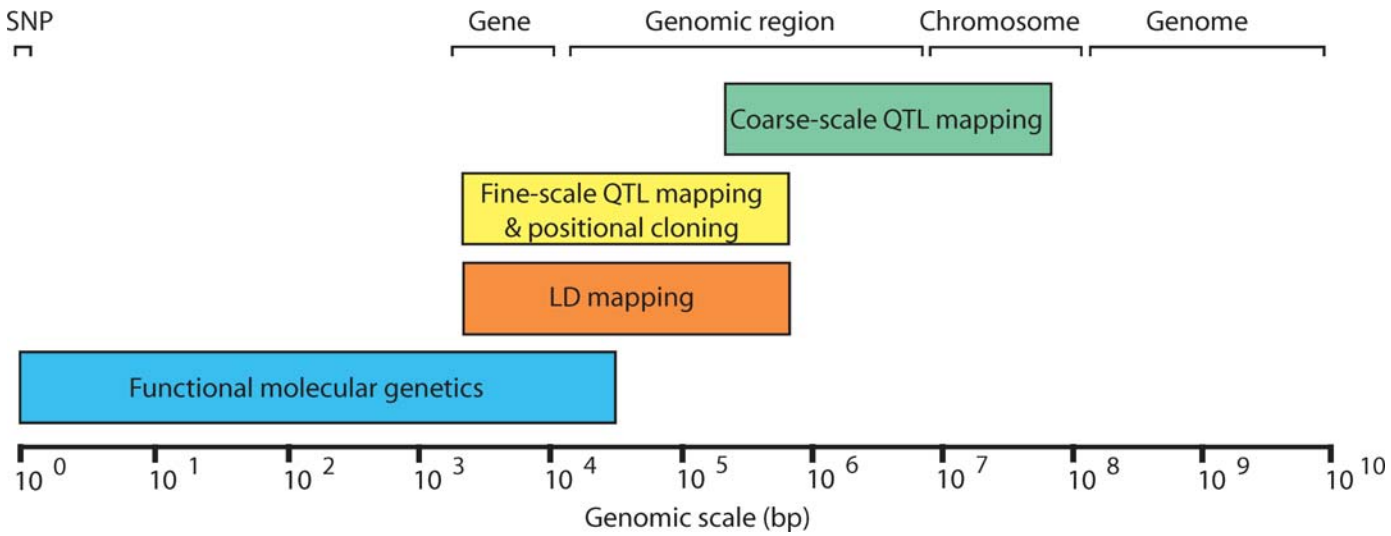

Fig. 1. The genomic scale of different genetic methods to isolate adaptive trait genes. A challenge exists in connecting broad genomic associations to specific adaptive polymorphisms. Combining genetic methods is the most likely approach to be successful in connecting adaptive traits to specific genes and polymorphisms.

sometimes referred to as adaptive trait locus mapping. Loci that exist in the tails of these empirical distributions may be regarded as candidate adaptive genes subject to further examination. It should be noted that the value of a test statistic for a gene may fall in the tails of these empirical distributions by chance, so additional experimentation is necessary to prove that a gene is indeed adaptive.

An alternative to the aforementioned tests is to use simulation-based approaches to assess selection. In particular, coalescent theory (Kingman, 1982; Hudson, 1991) has provided a powerful opportunity to detect selection at the molecular level under a variety of evolutionary scenarios (Nordborg, 2001). Coalescent simulation can be used to generate distributions of genealogies against which samples of alleles can be statistically compared (Rosenberg and Nordborg, 2002; Hein et al., 2004). Additionally, other non-coalescent methods, such as Poisson random field (PRF)-based methods, have been constructed to estimate the selection coefficients of sampled genes (Bustamante et al., 2002). These methods have proven to be of great utility in the search for genes under selection.

Genetic mapping of plant adaptive genes—Multiple techniques, including QTL mapping and linkage disequilibrium mapping methods, exist for mapping genes underlying adaptive traits based on marker-trait associations (as reviewed in Mackay, 2001; Remington et al., 2001b; Whitt and Buckler, 2003; Phillips, 2005, and elsewhere). These methods have proven successful for mapping genes for trait variation in several plant species, such as *A. thaliana*, tomato, and maize (see Remington et al., 2001b for discussion).

In QTL mapping, loci controlling trait variation between two individuals are mapped to specific genomic regions. Initially, individuals that differ in traits of interest are crossed, and their progeny are inbred and backcrossed to generate populations of recombinant inbred lines (RILs). These RILs are typically homozygous throughout the majority of their genomes with different genomic regions being descended from each parent. When grown in a controlled setting or a common garden, phenotypic differences can be mapped back to the genome based on trait associations with parental markers.

QTL mapping has played a prominent role in mapping genomic regions that control phenotypic variation in many species of plants. Most of these studies have resulted in the characterization of large-sized QTLs (>500 kb) that span hundreds of genes (e.g., Ungerer et al., 2002). Subsequent fine-mapping using nearly isogenic lines (NILs) is generally necessary to localize the gene(s) of effect within identified QTLs (Tanksley, 1993). The resolution of QTL mapping experiments can be improved by increasing marker density and the number of RILs, but rarely have QTLs been localized to regions of fewer than 10 genes.

An additional caveat with QTL mapping is that the typical biparental origin of most mapping populations may lead to the identification of QTLs that are found only in the parental lines and are not prevalent in the general population or species. This almost certainly has been the case for some QTLs that have been identified in *A. thaliana*, which due to its species history possesses a high number of rare polymorphisms throughout the genome that are often present in only a single ecotype. For other plant species that have higher outcrossing rates and less population structure, QTLs are more likely to be representative of genetic variation segregating within the population or species.

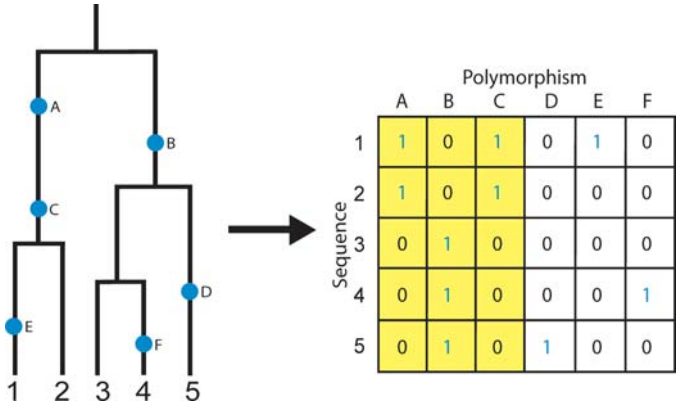

Fig. 2. A genealogical history without recombination. Polymorphisms A, B, and C are in complete linkage disequilibrium (LD) with each other. Here, LD is generated by the shared history of linked positions.

An alternative approach to mapping adaptive genes is to perform linkage disequilibrium mapping. Linkage disequilibrium (LD) is the nonrandom association of polymorphisms within a population (Pritchard and Przeworski, 2001; Fig. 2). Because polymorphisms that are in LD with a functionally important polymorphism will also be associated with any phenotypic differences caused by that polymorphism, LD can be exploited to map the genomic regions that underlie adaptations. In practice, LD mapping requires a sample of genotyped and phenotyped individuals taken from a natural population or from a family with a known pedigree. Correlations between observed genetic variants and trait variation in this sample can then be measured, leading to the identification of specific polymorphisms or haplotypes that explain adaptive trait variation.

In LD mapping, the mapping resolution is primarily influenced by the rate of LD decay. In *A. thaliana*, LD usually decays within 20 to 50 kb (Nordborg et al., 2005), which suggests that LD mapping may have a resolution approximately an order of magnitude higher than QTL mapping (Aranzana et al., 2005). In maize, LD decays much faster than in *A. thaliana*, oftentimes within a couple kilobases (Remington et al., 2001a). Because LD patterns vary substantially across plant species, the utility of this method for non-model plants has yet to be determined.

LD mapping is not without its caveats. In particular, this method is prone to spurious results based on population structure. Techniques exist for assessing the extent of cryptic population structure and accounting for it in association tests (Pritchard and Rosenberg, 1999; Pritchard et al., 2000a, b). A recent survey in *A. thaliana* found that including estimates of population structure as covariates in association tests dramatically reduced the number of false positives throughout the genome (Aranzana et al., 2005). Another concern for LD mapping is its lack of power for identifying associations at loci with very low minor allele frequencies. Increased sampling can reduce this problem, but not entirely solve it in some plant species. Overall, LD mapping has proven successful in plants, such as in maize (e.g., Thornsberry et al., 2001; Wilson et al., 2004) and *Arabidopsis* (e.g., Caicedo et al., 2004; Olsen et al., 2004; Aranzana et al., 2005; Gibson and Weir, 2005).

Expression mapping—An alternative approach to adaptive gene mapping has come from advances in technologies for studying gene expression. Either cDNA- or oligonucleotide-based whole genome microarrays are available for many plant species. Whole genome transcriptional analysis with these microarrays can be used to identify loci that control transcriptional differences between individuals. One increasingly used method is to analyze the transcriptomes of RILs in order to define expression QTLs (eQTLs)—marker intervals correlated with transcriptional variation (Gibson and Weir, 2005). Such screens allow for the differentiation of *cis* and *trans* eQTLs based on the position of the significant marker relative to the transcribed gene.

In conjunction with genetic mapping, gene expression analysis provides a powerful tool for connecting genetic variation to adaptive trait differences. Not only will this technology help in the identification of evolutionarily important genes and polymorphisms, but it may also help to determine the functional molecular basis of adaptive trait changes.

CONTEMPORARY STUDIES OF ADAPTATION IN PLANTS

Although few studies of adaptation have spanned all relevant levels of biological organization (Wright and Gaut, 2005b), several plant adaptations have been extensively examined from ecological, evolutionary, and molecular perspectives. In the following section, we will summarize some of these notable examples in which specific genes responsible for putatively or established adaptive phenotypic variation have been identified. These studies represent the application of the techniques described in the previous sections.

Adaptive trait locus mapping using selection signatures in plant genomes—For *A. thaliana*, there have been several moderate- to large-scale screens for adaptive genes based on selection tests. A sequence-based screen of 334 randomly distributed genomic regions among 12 ecotypes, for example, led to the identification of 28 loci that were in the tails of the empirical distribution of various test statistics (Schmid et al., 2005). In a similar screen of rapidly evolving genes between *A. thaliana* and *A. lyrata*, 14 genes among 304 compared orthologues were shown to have K_a values exceeding K_s values (Barrier et al., 2003). Six of these genes were examined further by comparing within- to between-species patterns of nucleotide change in the coding regions of these loci, and these rapidly evolving genes were demonstrated to have a higher average selection intensity than previously studied genes in *A. thaliana*.

One can also use genomic screens to identify loci that have been under balancing selection or that have been involved in local adaptation. Genes subject to these types of selection pressures are expected to have high levels of intraspecific variation (Hudson and Kaplan, 1988). A recent screen for high diversity genes in *A. thaliana* found three genomic regions with higher variation than the rest of the genome (Cork and Purugganan, 2005). One of these genomic regions harbored a member of a class of disease resistance genes commonly associated with balanced polymorphisms, while the putative reasons for selection for high diversity on the other loci remain unclear.

The genomic screen approach has also been used in maize to identify loci with reduced levels of diversity due to selection associated with domestication and crop diversification (Wright et al., 2005a). In this study, 774 loci were sampled from 14 maize and 16 teosinte inbred lines to estimate the severity of the bottleneck associated with the domestication of maize from teosinte. Using a bottleneck scenario to model the demographic effects of domestication, they identified two classes of genes—one class of genes whose sequence variation is consistent with the bottleneck and one for genes that are putative domestication loci. Two to four percent of the studied loci were estimated to belong to the latter group, and extending these results to the entire maize genome, up to 1200 genes could have been responsible for maize domestication.

Floral adaptations in Ipomoea—Floral color variation in the American morning glory (*Ipomoea*), has been implicated in adaptive evolution (Clegg and Durbin, 2003). These adaptive changes are believed to be driven by complex interactions between these plants and their pollinators and are based on pollinator preferences for particular floral colors. Because the molecular pathways underlying floral pigmentation have been well characterized, researchers have been successful in identifying the molecular bases for much of the floral color variation in these species.

Ipomoea purpurea, which is indigenous to Mexico and was likely introduced into the southeastern U.S. concomitant with maize culture, possesses three main floral colors—blue, red, and white—that are often found in different color blends and patterns. In Mexico, most populations are fixed for blue flowers, but in the U.S. white and red flowers are not uncommon. Selection has been implicated in the maintenance of the white polymorphism, but no evidence has been found for selection on the other floral variants (Clegg and Durbin, 2000). Ecological studies have suggested that white flowers are discriminated against by pollinators when rare but that they increase in frequency through self-fertilization (Clegg and Durbin, 2000).

Floral color in *Ipomoea* is determined primarily by the relative concentrations of two anthocyanin derivatives—cyanidin, which produces blue flowers, and pelargonidin, which produces red flowers. Two anthocyanin pathway genes have been identified that control floral color and patterning variation in *I. purpurea* (Clegg and Durbin, 2003). *Flavonoid 3′-hydroxylase* (*F3′H*) has been shown to confer the dominant blue phenotype and the recessive red phenotype (Zufall and Rausher, 2003). In addition, *chalcone synthase-D* (*CHS-D*), which possesses many natural alleles, was found to control floral color patterning through epistasis with other loci (Habu et al., 1998). Though the cloning of these genes represents a major achievement, no selective basis has been demonstrated for the maintenance of these polymorphisms in nature.

Multiple subgenera of *Ipomoea*, including the clade containing *I. quamoclit*, another well-studied species in this genus, have undergone changes from blue to red flowers to facilitate adaptive pollinator shifts from bees to birds (Zufall and Rausher, 2004). Molecular analysis of anthocyanin pathway genes in *I. quamoclit* revealed that *F3′H* mRNA levels are dramatically reduced in this species relative to the predominantly blue *I. purpurea*, though biochemical analysis showed *I. quamoclit*'s M*F3′H* is still functional. In addition, *I. quamoclit*'s *dihydroflavonol reductase-B* (*DFR-B*), one of three paralogues of this gene in *Ipomoea*, was found to have

numerous insertions and amino acid substitutions, as well as a 59-bp upstream shift of the stop codon. Heterologous complementation tests in the model genetic system *A. thaliana* found that the *I. purpurea DFR-B* gene could complement *A. thaliana DFR* null mutants, but that *I. quamoclit*'s *DFR-B* gene could not. These results provide strong evidence that one or both of these molecular changes—the *F3′H* regulatory changes and the *DFR-B* mutations—are responsible for the adaptive floral color transition from blue to red in *I. quamoclit*.

Flowering time variation in A. thaliana—Flowering time is a major developmental transition in plants, and this trait is likely a strong determinant of fecundity (Simpson and Dean, 2002). In the model genetic species *A. thaliana*, the timing of flowering varies significantly between different accessions (Nordborg and Bergelson, 1999), although the adaptive significance of this variation is still under active exploration (see Engelmann and Purugganan, in press, for a discussion of this ongoing research). Flowering time in *A. thaliana* has been shown to exhibit a latitudinal cline, suggesting the possible adaptation of this trait to a geographical/climatic component (Stinchcombe et al., 2004).

Over 60 genes have been shown to regulate flowering time in *A. thaliana,* illustrating the complex molecular circuitry underlying this trait. Much less is known about the genetic controls of natural variation in flowering time, because only four genes have been cloned that contribute to flowering time differences across populations of this species. *CRYPTOCHROME 2* (*CRY2*), which is involved in blue light photoreception, was one of the first genes to be shown to contribute to flowering time variation in *A. thaliana* (El-Assal et al., 2001). Two amino acid polymorphisms were identified that result in altered *CRY2* protein levels during the circadian cycle, causing the early flowering of plants under short day conditions. These polymorphisms, however, were found only in an accession from the Cape Verde Islands (the Cvi ecotype), and it is unclear whether this allele is the product of rare mutations or a local adaptation. An association study suggested that more common haplotypes of *CRY2* also contribute to flowering time variation in this species (Olsen et al., 2004).

A similarly rare polymorphism has been observed in the *FLOWERING LOCUS M*/*MADS AFFECTING FLOWERING 1* (*FLM*/*MAF1*) gene of the Nd-1 accession of *A. thaliana*, which was collected from Niederzenz, Germany (Werner et al., 2005). Initially identified as a QTL controlling over 60% of the flowering time variation in a recombinant inbred line (RIL) population derived from the Nd-1 and Columbia (Col-3 and Col-5) ecotypes, sequencing of the *FLM* genomic region in Nd-1 found that it was entirely absent in this accession. Genotyping of a larger group of accessions showed that the Nd-1 *FLM* deletion was unique to plants sampled from Niederzenz.

The most significant contributor to flowering time variation that has been characterized in *A. thaliana* to date, both in terms of effect and frequency, is the *FRIGIDA* (*FRI*) gene (Johanson et al., 2000). Molecular analysis revealed that multiple loss-of-function *FRI* alleles possessing large deletions segregate in natural populations of *A. thaliana*, at least two of which are found at moderate frequency throughout the species range (Johanson et al., 2000; Hagenblad and Nordborg, 2002; Le Corre et al., 2002; Hagenblad et al., 2004; Stinchcombe et al., 2004). Among *A. thaliana* accessions carrying a functional *FRI* allele, there exists a *FRI* genotype-dependent latitudinal cline in flowering time under field conditions (Stinchcombe et al., 2004). Recent work has shown that *FRI*'s association with flowering time is detectable with markers $\geq$100 kb away from *FRI* (Aranzana et al., 2005). This atypically extensive LD around *FRI* may be due to *FRI*'s involvement in local adaptation and selective sweeps**.**

An epistatic effect of *FRI* on *FLOWERING LOCUS C* (*FLC*), which encodes a MADS box floral repressor that is known to be upregulated by *FRI*, may be responsible for this latitudinal cline in flowering time (Caicedo et al., 2004). Two major *FLC* haplotype groups have been detected in *A. thaliana*, and there is significant flowering time variation associated with *FRI-FLC* two-locus genotypes. *FLC* haplotypes also show a significant latitudinal distribution, but only in functional *FRI* backgrounds. Finally, *FRI* and *FLC* show significant intergenic linkage disequilibrium, even though the two genes are found in two different *A. thaliana* chromosomes. However, there is some question as whether this clinal pattern reflects some unrecognized cryptic population structure. These findings, however, suggest that epistatic selection may underlie flowering time variation in this species, although these associations need confirmation by molecular analysis of the different alleles at these loci (Caicedo et al., 2004; Stinchcombe et al., 2004). Although these four genes have been shown to contain polymorphisms that underlie natural variation in flowering time in *A. thaliana*, QTL mapping experiments suggest that there are many other loci that contribute to this trait variation (Ungerer et al., 2002, 2003; Weinig et al., 2002). Interestingly, the genes controlling flowering time variation appear to differ between laboratory and field conditions (Weinig et al., 2002). This emphasizes the importance of studying the genetics of adaptive traits in ecological settings. Work remains to identify these ecologically relevant genetic polymorphisms and their potential contributions to life history adaptation in this species.

The evolution of plant mating systems: self-incompatibility and selfing—Self-incompatibility (SI), the prevention of self-fertilization, has evolved multiple times across distantly related plant species (Matton et al., 1994). SI is commonly regarded as adaptive due to its role in promoting genetic diversity through the generation of heterozygosity and new allelic combinations. In many plants, the *S* (*Sterility*) locus, which is comprised of multiple allelic loci essential to self-recognition, is the primary controller of SI (Nasrallah, 1997). The maintenance of *S*-locus diversity by balancing selection has been the subject of intense investigation, and it remains one of the classic examples of frequency dependent selection, which is one type of balancing selection, in nature (Charlesworth et al., 2005).

While much is known about the genetic basis of self-incompatibility, much less is known about the genetic basis for the evolution of selfing. In *A. thaliana*, the *S* locus is present but nonfunctional, resulting in the loss of SI and a very high rate of self-fertilization relative to its outcrossing congeners. This species lacks functional copies of both *S-locus receptor kinase* (*SRK*) and *S-locus cysteine-rich protein* (*SCR*), the receptor and ligand, respectively, responsible for SI and encoded by the *S* locus. Transformation of *A. thaliana* with functional copies of the *SRK* and *SCR* genes from *A. lyrata* is sufficient to create self-incompatible *A. thaliana*, confirming that one of these genes is responsible for the loss of SI (Nasrallah et al., 2002). Recent work has provided evidence for positive selection on loss-of-function alleles at the *SCR* locus, suggesting that pseudogenization of *SCR* (and possibly *SRK*),

is the most likely explanation for the evolution of selfing in *A. thaliana* (Shimizu et al., 2004).

Coalescent analysis of the observed genetic variation within the *A. thaliana S* locus suggests that the transition from SI to self-compatibility arose very recently, likely having evolved during the population expansion of this species from post-glacial refugia in Europe (Shimizu et al., 2004). This expansion is believed to have occurred approximately 17 000 years ago and to have had a significant effect on species-wide genetic variation of *A. thaliana* (Sharbel et al., 2000). The evolution of self-fertilization in *A. thaliana* at this time may have been adaptive by allowing continued reproduction in this species during the population expansion despite a deficiency of reproductive partners. This hypothesis is supported by changes that have occurred in other floral characters in *A. thaliana* to facilitate autopollination (Shimizu and Purugganan, 2005).

Disease resistance in A. thaliana—The interaction of pathogens and their hosts has been a powerful model for the study of coevolution (Bent, 1996; Bergelson et al., 2001). Numerous disease resistance loci have been identified in the plant genetic model *A. thaliana* (e.g., Kunkel et al., 1993; Grant et al., 1995), and multiple studies of the molecular evolutionary dynamics of pathogen resistance loci in *A. thaliana* have been conducted (Caicedo et al., 1999; Stahl et al., 1999; Tian et al., 2002; Mauricio et al., 2003). These studies suggest that resistance (*R*) genes are often maintained as highly divergent alleles or presence/absence polymorphisms due to balancing selection driven by interactions between host plants and their pathogens (Bergelson et al., 2001).

One extensively studied *R* gene is *RPM1*, which encodes an NBS-LRR protein that confers resistance to *Pseudomonas syringae* strains carrying either the *AvrRPM1* or *AvrB* avirulence genes (Stahl et al., 1999). This gene exists as a presence/absence polymorphism across the *A. thaliana* species range, and the genomic region surrounding it possesses the molecular signature of balancing selection. The creation of transgenic lines differing only in the presence or absence of *RPM1* was used to demonstrate that a fitness tradeoff at *RPM1* could be responsible for the maintenance of this balanced polymorphism (Tian et al., 2003).

A larger-scale study of the 20-kb genomic region containing *RPS5*, another NBS-LRR protein-encoding *R* gene, also found evidence for balancing selection (Tian et al., 2002). This gene confers resistance to *Pseudomonas syringae* strains that express the *AvrPph3* avirulence gene. Like *RPM1*, this locus also possesses a widespread presence/absence polymorphism, but an additional susceptible allele that contains a frameshift mutation was discovered. The pattern of molecular variation and linkage disequilibrium at this locus suggests that balancing selection is acting upon *RPS5*. These *RPS2* and *RPS5* studies provide excellent examples of how ecologically based selection can be connected to molecular signatures at the genomic level.

The genomics of herbivore resistance and growth rate in A. thaliana—How plants defend themselves against their predators has also played a central role in the study of the molecular evolution of ecological interactions. Glucosinolates, a class of secondary metabolites produced by many plants, are thought to provide an important defense against herbivory in the Brassicaceae. Much is known about the biochemical pathways that produce glucosinolates in *A. thaliana* and a gene controlling variation in the types of glucosinolates synthesized, *methylthioalkylmalate synthase1* (*MAM1*), has been identified in this species (Kroymann et al., 2001).

Sequencing of the *MAM* genomic region in 25 ecotypes found that a closely linked paralogue of *MAM1*, designated *MAM2*, was often present (Kroymann et al., 2003). The presence of *MAM1*, *MAM2*, or both genes in the *MAM* genomic region was highly variable. A significant association was detected between the types of glucosinolates produced by an ecotype and its *MAM1*/*MAM2* genotype. In addition, the pattern of genetic variation at *MAM2* was indicative of balancing selection acting on this gene, suggesting that an ecological trade-off might influence this locus.

Unintentionally, further analysis of the upstream and downstream genomic regions surrounding the *MAM* cluster found two QTLs for growth rate within 100 kb of each other (Kroymann and Mitchell-Olds, 2005). One of these QTLs was fine-mapped to a single gene, at which balancing selection was detected from a molecular population genetic sample. The other QTL was refined to a 32-kb interval, but could not be mapped to a specific gene. Epistasis was shown to exert strong influence upon the effects of these loci on growth rate. The identification of these growth rate loci represents a major achievement, but whether they contribute to fitness in the natural habitat of *Arabidopsis* needs to be determined.

The signature of selection in domestication genes—Domestication has long been viewed as a model for adaptation (Darwin, 1859, 1897). Investigating the evolution of crop "domestication traits" favored by early farming cultures (e.g., the loss of seed dispersal mechanisms or an increased yield under agricultural field conditions), in addition to the subsequent diversification of other crop traits that have arisen from selective breeding to satisfy diverse human cultural preferences (e.g., different grain colors and tastes), can provide tremendous insight into the genomic signature of selection on specific traits (Doebley, 2004). Furthermore, though domestication may not precisely replicate adaptation, the characterization of the genes responsible for domestication traits may help to clarify the number and effects of genes that underlie the changes in selected traits.

The study of the molecular genetics of domestication is most clearly illustrated by studies of maize (*Zea mays* subsp. *mays*). Maize possesses a strikingly different morphology from its teosinte progenitor (*Z. mays* subsp. *parviglumis*), and QTL mapping studies have identified several genomic regions that harbor genes responsible for this species differentiation (Doebley, 2004). Two major phenotypic changes associated with maize domestication were a substantial decrease in axillary branching (Doebley et al., 1997; Gallavotti et al., 2004) and a reduction in glume size that exposed kernels on the corn ears (Dorweiler et al., 1993; Wang et al., 2005). Substantial effort has resulted in the characterization of the genes responsible for both of these maize domestication traits.

The increase in apical dominance in maize relative to teosinte has been mapped to the *teosinte branched1* (*tb1*) gene, which has higher expression in maize than teosinte (Doebley et al., 1997). This gene has a significant depression in genetic variability in the 5′ non-transcribed region of maize *tb1* compared to teosinte *tb1*, suggesting that this region could have been the target of positive selection that resulted in a change in gene regulation (Wang et al., 1999). More extensive analysis of the genomic region surrounding this region has shown that the selective sweep associated with the fixation of *tb1* in maize

extends for more than 60 kb of the upstream intergenic region (Clark et al., 2004). The *barren stalk1* (*ba1*) gene, an additional domestication locus that is epistatic to *tb1*, has been characterized, and has been proposed to be a target of selection during postdomestication improvement of maize (Gallavotti et al., 2004).

A major effect gene has also been characterized for glume reduction in maize. This gene, *teosinte glume architecture 1* (*tga1*), which was also initially identified as a QTL for domestication traits between maize and teosinte, explains up to 50% of the glume reduction associated with domestication (Dorweiler et al., 1993). The pattern of variation at *tga1* suggests that this locus has been the target of a selective sweep (Wang et al., 2005), again showing that the signature of artificial selection is often apparent at the molecular level in domesticated species. In addition to *tb1*, *ba1*, and *tga1*, hundreds of other domestication genes may exist in maize (as described in an earlier section [Wright et al., 2005a]).

Efforts to identify genes underlying domestication traits in other species have also been successful. Heterotopic expression of the *MPF2* gene, which encodes a MADS box protein, is suggested to have led to the evolution of the fruit husks in the genus *Physalis* (He and Saedler, 2005). Positive selection on the *CAULIFLOWER* gene has been shown to have been important in the origin of cauliflower (*Brassica oleracea* subsp. *botrytis*) (and possibly broccoli [*B. oleracea* subsp. *italica*]) (Purugganan et al., 2000). In rice, selection on a mutation impairing proper pre-mRNA splicing of the *Waxy* gene transcript has been critical to the evolution of glutinous ("sticky") rice (Olsen and Purugganan, 2002) and possibly the major variety group referred to as temperate japonica (*Oryza sativa* subsp. *japonica*; Olsen et al., 2006). Lastly, multiple genes that may have been important to differences in fruit morphology between wild and domestic tomato have been identified (Frary et al., 2000; Liu et al., 2002). The study of the domestication of crop plants has provided many insights into the genetic targets of selection on specific traits.

CONCLUSION

Numerous adaptive genes have been characterized in plants. The cloning of adaptive and domestication genes thus far has provided substantial support for the importance of major effect mutations to plant adaptation (e.g., the *FRI* locus in *A. thaliana*, Johanson et al., 2000; *tb1* in maize Doebly et al. 1997). At present, no minor effect loci have been cloned in plants, but their existence is supported by QTL studies (a fact reported in Tanksley [1993] and elsewhere). However, for no polygenic adaptation has the entire genetic architecture been determined, making it difficult to infer how many genes are typically responsible for adaptations. In addition, whether certain types of genes or mutations are more frequently responsible for adaptation is an unanswered question.

To better understand the genetic basis of plant adaptation in general, a transition from *A. thaliana* and crop species into other non-model plants is necessary. Genomic resources are being developed for genera like *Mimulus* that have long-standing traditions in ecology and evolution, but poor genetic tools. In some cases, it may be possible to use present genetic models as platforms for closely related non-models. Much research is being done in the Brassicaceae, the family containing *A. thaliana*, based on this paradigm. More examples of adaptive genes are needed in non-model plant species.

The continued characterization of adaptive genes in plants will make it possible to answer some longstanding questions about adaptation. For instance, that epistasis in genetic networks and pathways can shape how adaptations evolve is generally accepted, but the effect of natural selection acts on such systems is unknown (Cork and Purugganan, 2004). In addition, why parallel and repeated adaptations, such as the transition to self-compatibility, have occurred numerous times in plants across both intraspecific and interspecific scales can only be fully understood once the genetic architectures for these traits have been identified (Wood et al., 2005). These are just two areas in evolutionary biology that will benefit from the mapping of adaptive genes.

The research described in this paper parallels efforts of the animal research community to identify adaptive genes and to understand the effects of natural selection at the genomic level. For instance, selection signature screens of genes throughout the human genome have identified loci that have likely been important to human adaptive evolution (Clark et al., 2003; Bustamante et al., 2005). Efforts to classify the proportion of adaptive and deleterious mutations throughout animal genomes have provided another avenue for exploring how genomes evolve under selection (Eyre-Walker and Keightley, 1999; Smith and Eyre-Walker, 2002). The extension of the methods used in these examples to plants will help to determine how natural selection has shaped plant genomes. Plant evolutionary genetics has experienced many successes, but much work remains in determining plant adaptive traits and their underlying genetic controls. Techniques exist to study adaptive evolution on many levels, making it possible to formulate comprehensive explanations for how plant evolution has occurred. The continued progress of this field will help to explain the extant condition of plant biodiversity and should provide important insights into how adaptation occurs at the genetic level.

LITERATURE CITED

ARANZANA, M. J., S. KIM, K. ZHAO, E. BAKKER, M. HORTON, K. JAKOB, C. LISTER, J. MOLITOR, C. SHINDO, C. TANG, C. TOOMAJIAN, B. TRAW, H. ZHENG, J. BERGELSON, C. DEAN, P. MARJORAM, AND M. NORDBORG. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* 1: e60.

BARRIER, M., C. D. BUSTAMANTE, J. YU, AND M. D. PURUGGANAN. 2003. Selection on rapidly evolving proteins in the Arabidopsis genome. *Genetics* 163: 723–733.

BENT, A. F. 1996. Plant disease resistance genes: function meets structure. *Plant Cell* 8: 1757–1771.

BERGELSON, J., M. KREITMAN, E. A. STAHL, AND D. TIAN. 2001. Evolutionary dynamics of plant *R*-genes. *Science* 292: 2281–2285.

BUSTAMANTE, C. D., R. NIELSEN, S. A. SAWYER, K. M. OLSEN, M. D. PURUGGANAN, AND D. L. HARTL. 2002. The cost of inbreeding in *Arabidopsis*. *Nature* 416: 531–534.

BUSTAMANTE, C. D., A. FLEDEL-ALON, S. WILLIAMSON, R. NIELSEN, M. T. HUBISZ, S. GLANOWSKI, D. M. TANENBAUM, T. J. WHITE, J. J. SNINSKY, R. D. HERNANDEZ, D. CIVELLO, M. D. ADAMS, M. CARGILL, AND A. G. CLARK. 2005. Natural selection on protein-coding genes in the human genome. *Nature* 437: 1153–1157.

CAICEDO, A. L., B. A. SCHAAL, AND B. N. KUNKEL. 1999. Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* 96: 302–306.

CAICEDO, A. L., J. R. STINCHCOMBE, K. M. OLSEN, J. SCHMITT, AND M. D. PURUGGANAN. 2004. Epistatic interaction between Arabidopsis *FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences, USA* 101: 15670–15675.

CHARLESWORTH, D., X. VEKEMANS, V. CASTRIC, AND S. GLEMIN. 2005. Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytologist* 168: 61–69.

CLARK, A. G., S. GLANOWSKI, R. NIELSEN, P. THOMAS, A. KEJARIWAL, M. J. TODD, D. M. TANENBAUM, D. CIVELLO, F. LU, B. MURPHY, S. FERRIERA, G. WANG, X. ZHENG, T. J. WHITE, J. J. SNINSKY, M. D. ADAMS, AND M. CARGILL. 2003. Positive selection in the human genome inferred from human-chimp-mouse orthologous gene alignments. *Cold Spring Harbor Symposia on Quantitative Biology* 68: 471–477.

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*. *Proceedings of the National Academy of Sciences, USA* 101: 700–707.

CLEGG, M. T., AND M. L. DURBIN. 2000. Flower color variation: a model for the experimental study of evolution. *Proceedings of the National Academy of Sciences, USA* 97: 7016–7023.

CLEGG, M. T., AND M. L. DURBIN. 2003. Tracing floral adaptations from ecology to molecules. *Nature Reviews Genetics* 4: 206–215.

CORK, J. M., AND M. D. PURUGGANAN. 2004. The evolution of molecular genetic pathways and networks. *Bioessays* 26: 479–484.

CORK, J. M., AND M. D. PURUGGANAN. 2005. High-diversity genes in the Arabidopsis genome. *Genetics* 170: 1897–1911.

DARWIN, C. 1859. On the origin of species. Reprinted by Harvard University Press, Cambridge, Massachusetts, USA.

DARWIN, C. 1897. The variation of animals and plants under domestication. D. Appleton, New York, New York, USA.

DOEBLEY, J. 2004. The genetics of maize evolution. *Annual Review of Genetics* 38: 37–59.

DOEBLEY, J., A. STEC, AND L. HUBBARD. 1997. The evolution of apical dominance in maize. *Nature* 386: 485–488.

DORWEILER, J., A. STEC, J. KERMICLE, AND J. DOEBLEY. 1993. *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. *Science* 262: 233–235.

EL-ASSAL, S. E.-D., C. ALONSO-BLANCO, A. J. PEETERS, V. RAZ, AND M. KOORNNEEF. 2001. A QTL for flowering time in Arabidopsis reveals a novel allele of *CRY2*. *Nature Genetics* 29: 435–440.

ENGELMANN, K. E., AND M. D. PURUGGANAN. In press. The molecular evolutionary ecology of plant development: flowering time in *Arabidopsis thaliana*. *Advances in Botanical Research*.

EYRE-WALKER, A., AND P. D. KEIGHTLEY. 1999. High genomic deleterious mutation rates in hominids. *Nature* 397: 344–347.

FAY, J. C., AND C. I. WU. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.

FISHER, R. 1930. The genetical theory of natural selection. Oxford University Press, Oxford, UK.

FRARY, A., T. C. NESBITT, S. GRANDILLO, E. KNAAP, B. CONG, J. LIU, J. MELLER, R. ELBER, K. B. ALPERT, AND S. D. TANKSLEY. 2000. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289: 85–88.

FU, Y. X., AND W. H. LI. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.

GALLAVOTTI, A., Q. ZHAO, J. KYOZUKA, R. B. MEELEY, M. K. RITTER, J. F. DOEBLEY, M. E. PE, AND R. J. SCHMIDT. 2004. The role of *barren stalk1* in the architecture of maize. *Nature* 432: 630–635.

GIBSON, G., AND B. WEIR. 2005. The quantitative genetics of transcription. *Trends in Genetics* 21: 616–623.

GILLESPIE, J. 1991. The causes of molecular evolution. Oxford University Press, Oxford, UK.

GRANT, M. R., L. GODIARD, E. STRAUBE, T. ASHFIELD, J. LEWALD, A. SATTLER, R. W. INNES, AND J. L. DANGL. 1995. Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. *Science* 269: 843–846.

HABU, Y., Y. HISATOMI, AND S. IIDA. 1998. Molecular characterization of the mutable *flaked* allele for flower variegation in the common morning glory. *Plant Journal* 16: 371–376.

HAGENBLAD, J., AND M. NORDBORG. 2002. Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. *Genetics* 161: 289–298.

HAGENBLAD, J., C. TANG, J. MOLITOR, J. WERNER, K. ZHAO, H. ZHENG, P. MARJORAM, D. WEIGEL, AND M. NORDBORG. 2004. Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics* 168: 1627–1638.

HE, C., AND H. SAEDLER. 2005. Heterotopic expression of *MPF2* is the key to the evolution of the Chinese lantern of *Physalis,* a morphological novelty in Solanaceae. *Proceedings of the National Academy of Sciences, USA* 102: 5779–5784.

HEIN, J., M. SCHIERUP, AND C. WIUF. 2004. Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, Oxford, UK.

HUDSON, R. R. 1991. Gene genealogies and the coalescent process. *In* D. Futuyma and J. Antonovics [eds.], Oxford surveys in evolutionary biology, 1–44. Oxford University Press, Oxford, UK.

HUDSON, R. R., AND N. L. KAPLAN. 1988. The coalescent process in models with selection and recombination. *Genetics* 120: 831–840.

HUDSON, R. R., M. KREITMAN, AND M. AGUADE. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 153–159.

JOHANSON, U., J. WEST, C. LISTER, S. MICHAELS, R. AMASINO, AND C. DEAN. 2000. Molecular analysis of *FRIGIDA*, a major determinant of natural variation in Arabidopsis flowering time. *Science* 290: 344–347.

KAPLAN, N. L., R. R. HUDSON, AND C. H. LANGLEY. 1989. The "hitchhiking effect" revisited. *Genetics* 123: 887–899.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

KINGMAN, J. 1982. The coalescent. *Stochastic Processes and Their Applications* 13: 235–248.

KROYMANN, J., S. DONNERHACKE, D. SCHNABELRAUCH, AND T. MITCHELL-OLDS. 2003. Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. *Proceedings of the National Academy of Sciences, USA* (100 Supplement) 2: 14587–14592.

KROYMANN, J., AND T. MITCHELL-OLDS. 2005. Epistasis and balanced polymorphism influencing complex trait variation. *Nature* 435: 95–98.

KROYMANN, J., S. TEXTOR, J. G. TOKUHISA, K. L. FALK, S. BARTRAM, J. GERSHENZON, AND T. MITCHELL-OLDS. 2001. A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiology* 127: 1077–1088.

KUNKEL, B. N., A. F. BENT, D. DAHLBECK, R. W. INNES, AND B. J. STASKAWICZ. 1993. *RPS2,* an Arabidopsis disease resistance locus specifying recognition of *Pseudomonas syringae* strains expressing the avirulence gene *avrRpt2*. *Plant Cell* 5: 865–875.

LE CORRE, V., F. ROUX, AND X. REBOUD. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution* 19: 1261–1271.

LEWONTIN, R. C., AND J. KRAKAUER. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175–195.

LIU, J., J. VAN ECK, B. CONG, AND S. D. TANKSLEY. 2002. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the National Academy of Sciences, USA* 99: 13302–13306.

LUIKART, G., P. R. ENGLAND, D. TALLMON, S. JORDAN, AND P. TABERLET. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981–994.

MACKAY, T. F. 2001. The genetic architecture of quantitative traits. *Annual Review of Genetics* 35: 303–339.

MATTON, D. P., N. NASS, A. E. CLARKE, AND E. NEWBIGIN. 1994. Self-incompatibility: how plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences, USA* 91: 1992–1997.

MAURICIO, R., E. A. STAHL, T. KORVES, D. TIAN, M. KREITMAN, AND J. BERGELSON. 2003. Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. *Genetics* 163: 735–746.

MAYNARD-SMITH, J., AND J. HAIGH. 1974. The hitchhiking effect of a favorable gene. *Genetical Research* 23: 23–35.

MCDONALD, J. H., AND M. KREITMAN. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351: 652–654.

NASRALLAH, J. B. 1997. Evolution of the *Brassica* self-incompatibility locus: a look into *S*-locus gene polymorphisms. *Proceedings of the National Academy of Sciences, USA* 94: 9516–9519.

NASRALLAH, M. E., P. LIU, AND J. B. NASRALLAH. 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two *S* locus genes from *A. lyrata*. *Science* 297: 247–249.

NEI, M., AND T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3: 418–426.

NIELSEN, R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86: 641–647.

NORDBORG, M. 2001. Coalescent theory. *In* D. Balding, M. Bishop, and C. Cannings [eds.], Handbook of statistical genetics, 179–212. Wiley, Chichester, UK.

NORDBORG, M., AND J. BERGELSON. 1999. The effect of seed and rosette cold treatment on germination and flowering time in some *Arabidopsis thaliana* (Brassicaceae) ecotypes. *American Journal of Botany* 86: 470.

NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN, H. ZHENG, E. BAKKER, P. CALABRESE, J. GLADSTONE, R. GOYAL, M. JAKOBSSON, S. KIM, Y. MOROZOV, B. PADHUKASAHASRAM, V. PLAGNOL, N. A. ROSENBERG, C. SHAH, J. D. WALL, J. WANG, K. ZHAO, T. KALBFLEISCH, V. SCHULZ, M. KREITMAN, AND J. BERGELSON. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* 3: e196.

OLSEN, K., A. CAICEDO, N. POLATO, A. MCCLUNG, S. MCCOUCH, AND M. PURUGGANAN. In press. Selection under domestication: evidence for a sweep in the rice *Waxy* genomic region. *Genetics.*

OLSEN, K. M., S. S. HALLDORSDOTTIR, J. R. STINCHCOMBE, C. WEINIG, J. SCHMITT, AND M. D. PURUGGANAN. 2004. Linkage disequilibrium mapping of Arabidopsis *CRY2* flowering time alleles. *Genetics* 167: 1361–1369.

OLSEN, K. M., AND M. D. PURUGGANAN. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162: 941–950.

ORR, H. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52: 935–949.

ORR, H. A. 2005a. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* 6: 119–127.

ORR, H. A. 2005b. Theories of adaptation: what they do and don't say. *Genetica* 123: 3–13.

PHILLIPS, P. C. 2005. Testing hypotheses regarding the genetics of adaptation. *Genetica* 123: 15–24.

PRITCHARD, J. K., AND M. PRZEWORSKI. 2001. Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* 69: 1–14.

PRITCHARD, J. K., AND N. A. ROSENBERG. 1999. Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics* 65: 220–228.

PRITCHARD, J. K., M. STEPHENS, AND P. DONNELLY. 2000a. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

PRITCHARD, J. K., M. STEPHENS, N. A. ROSENBERG, AND P. DONNELLY. 2000b. Association mapping in structured populations. *American Journal of Human Genetics* 67: 170–181.

PURUGGANAN, M. D., A. L. BOYLES, AND J. I. SUDDITH. 2000. Variation and selection at the *CAULIFLOWER* floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics* 155: 855–862.

REMINGTON, D. L., AND M. PURUGGANAN. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. *International Journal of Plant Sciences* 164: S7–S20.

REMINGTON, D. L., J. M. THORNSBERRY, Y. MATSUOKA, L. M. WILSON, S. R. WHITT, J. DOEBLEY, S. KRESOVICH, M. M. GOODMAN, AND E. S. BUCKLER. 2001a. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences, USA* 98: 11479–11484.

REMINGTON, D. L., M. C. UNGERER, AND M. D. PURUGGANAN. 2001b. Map-based cloning of quantitative trait loci: progress and prospects. *Genetical Research* 78: 213–218.

ROSENBERG, N. A., AND M. NORDBORG. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3: 380–390.

SCHMID, K. J., S. RAMOS-ONSINS, H. RINGYS-BECKSTEIN, B. WEISSHAAR, AND T. MITCHELL-OLDS. 2005. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* 169: 1601–1615.

SHARBEL, T., B. HAUBOLD, AND T. MITCHELL-OLDS. 2000. Genetic distance by isolation in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology* 9: 2109–2118.

SHIMIZU, K. K., AND M. D. PURUGGANAN. 2005. Evolutionary and ecological genomics of Arabidopsis. *Plant Physiology* 138: 578–584.

SHIMIZU, K. K., J. M. CORK, A. L. CAICEDO, C. A. MAYS, R. C. MOORE, K. M. OLSEN, S. RUZSA, G. COOP, C. D. BUSTAMANTE, P. AWADALLA, AND M. D. PURUGGANAN. 2004. Darwinian selection on a selfing locus. *Science* 306: 2081–2084.

SIMPSON, G. G., AND C. DEAN. 2002. Arabidopsis, the Rosetta stone of flowering time? *Science* 296: 285–289.

SMITH, N. G., AND A. EYRE-WALKER. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415: 1022–1024.

STAHL, E. A., G. DWYER, R. MAURICIO, M. KREITMAN, AND J. BERGELSON. 1999. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400: 667–671.

STINCHCOMBE, J. R., C. WEINIG, M. UNGERER, K. M. OLSEN, C. MAYS, S. S. HALLDORSDOTTIR, M. D. PURUGGANAN, AND J. SCHMITT. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences, USA* 101: 4712–4717.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.

TANKSLEY, S. D. 1993. Mapping polygenes. *Annual Review of Genetics* 27: 205–233.

THORNSBERRY, J. M., M. M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN, AND E. S. BUCKLER. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* 28: 286–289.

TIAN, D., H. ARAKI, E. STAHL, J. BERGELSON, AND M. KREITMAN. 2002. Signature of balancing selection in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* 99: 11525–11530.

TIAN, D., M. B. TRAW, J. Q. CHEN, M. KREITMAN, AND J. BERGELSON. 2003. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423: 74–77.

UNGERER, M. C., S. S. HALLDORSDOTTIR, J. L. MODLISZEWSKI, T. F. MACKAY, AND M. D. PURUGGANAN. 2002. Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* 160: 1133–1151.

UNGERER, M. C., S. S. HALLDORSDOTTIR, M. D. PURUGGANAN, AND T. F. MACKAY. 2003. Genotype-environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics* 165: 353–365.

WANG, H., T. NUSSBAUM-WAGLER, B. LI, Q. ZHAO, Y. VIGOUROUX, M. FALLER, K. BOMBLIES, L. LUKENS, AND J. F. DOEBLEY. 2005. The origin of the naked grains of maize. *Nature* 436: 714–719.

WANG, R. L., A. STEC, J. HEY, L. LUKENS, AND J. DOEBLEY. 1999. The limits of selection during maize domestication. *Nature* 398: 236–239.

WEINIG, C., M. C. UNGERER, L. A. DORN, N. C. KANE, Y. TOYONAGA, S. S. HALLDORSDOTTIR, T. F. MACKAY, M. D. PURUGGANAN, AND J. SCHMITT. 2002. Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* 162: 1875–1884.

WERNER, J. D., J. O. BOREVITZ, N. WARTHMANN, G. T. TRAINER, J. R. ECKER, J. CHORY, AND D. WEIGEL. 2005. Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for

natural flowering-time variation. *Proceedings of the National Academy of Sciences, USA* 102: 2460–2465.

WHITT, S. R., AND E. S. BUCKLER. 2003. Using natural allelic diversity to evaluate gene function. *In* E. Grotewald [ed.], Plant functional genomics: methods and protocols. Humana Press, Totowa, New Jersey, USA.

WILSON, L. M., S. R. WHITT, A. M. IBANEZ, T. R. ROCHEFORD, M. M. GOODMAN, AND E. S. BUCKLER. 2004. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16: 2719–2733.

WOOD, T. E., J. M. BURKE, AND L. H. RIESEBERG. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* 123: 157–170.

WRIGHT, S. I., I. V. BI, S. G. SCHROEDER, M. YAMASAKI, J. F. DOEBLEY, M. D. MCMULLEN, AND B. S. GAUT. 2005a. The effects of artificial selection on the maize genome. *Science* 308: 1310–1314.

WRIGHT, S. I., AND B. S. GAUT. 2005b. Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution* 22: 506–519.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Compututational Applications to Biosciences* 13: 555–556.

ZUFALL, R. A., AND M. D. RAUSHER. 2003. The genetic basis of a flower-color polymorphism in the common morning glory, *Ipomoea purpurea*. *Journal of Heredity* 94: 442–448.

ZUFALL, R. A., AND M. D. RAUSHER. 2004. Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* 428: 847–850.