

# Patterns of Selection in Plant Genomes

Josh Hough, Robert J. Williamson,  
and Stephen I. Wright

Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada, M5S 3B2; email: josh.hough@utoronto.ca, robert.williamson@utoronto.ca, stephen.wright@utoronto.ca

Annu. Rev. Ecol. Evol. Syst. 2013. 44:31–49

First published online as a Review in Advance on August 28, 2013

The *Annual Review of Ecology, Evolution, and Systematics* is online at [ecolsys.annualreviews.org](http://ecolsys.annualreviews.org)

This article's doi:  
10.1146/annurev-ecolsys-110512-135851

Copyright © 2013 by Annual Reviews.  
All rights reserved

## Keywords

plant evolution, genome-wide selection, effective population size, recombination, dominance, polyploidy

## Abstract

Plants show a wide range of variation in mating system, ploidy level, and demographic history, allowing for unique opportunities to investigate the evolutionary and genetic factors affecting genome-wide patterns of positive and negative selection. In this review, we highlight recent progress in our understanding of the extent and nature of selection on plant genomes. We discuss differences in selection as they relate to variation in demography, recombination, mating system, and ploidy. We focus on the population genetic consequences of these factors and argue that, although variation in the magnitude of purifying selection is well documented, quantifying rates of positive selection and disentangling the relative importance of recombination, demography, and ploidy are ongoing challenges. Large-scale comparative studies that examine the relative and joint importance of these processes, combined with explicit models of population history and selection, are key and feasible goals for future work.

---

**Effective population size ( $N_e$ ):** the size of an otherwise ideal population experiencing the same amount of genetic drift as the population in question

**Negative selection:** selection of alleles that have deleterious effects on the fitness of their carriers (i.e., selection against deleterious alleles)

**Positive selection:** selection of alleles that confer fitness benefits to their carriers (i.e., selection on beneficial alleles)

**Purifying selection:** also termed negative selection; the selective removal of deleterious alleles

---

## INTRODUCTION

Population genetics theory suggests that several key factors should be important in influencing patterns of genome-wide selection, including effective population size ( $N_e$ ), population structure, recombination, and ploidy. The extent to which variation in these factors can drive variation in selection within and between species is an important issue that connects diverse study systems and research questions into a unifying framework. Although our understanding of the causes of variation in genome-wide selection is limited, recent developments in genomics have started to shed light on this issue.

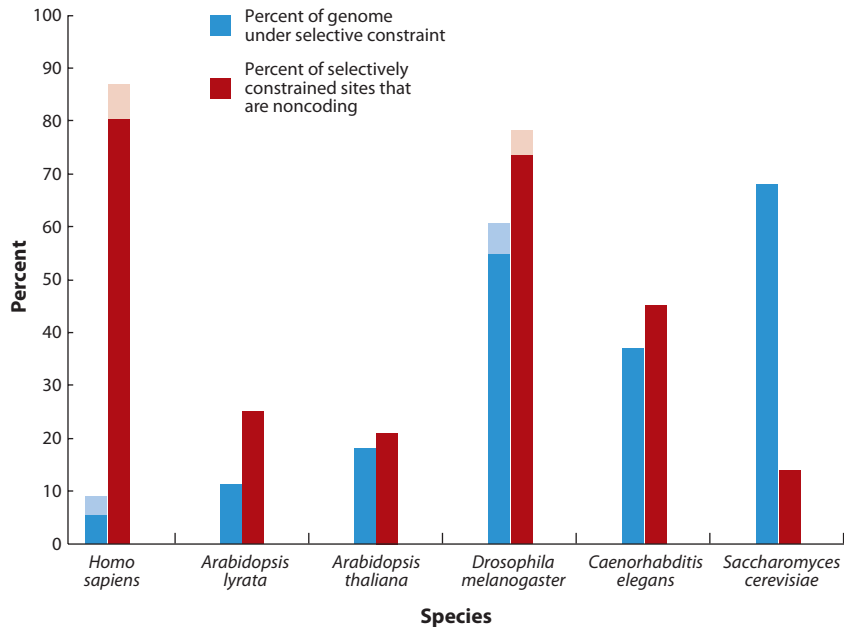
Plants, with their unique life-cycle characteristics and diversity of mating systems, ploidy levels, and demographic histories, provide an ideal testing ground for exploring empirically the importance of different population genetic processes in shaping patterns of selection. For example, the frequent occurrence of mating system transitions in plants makes it possible to test how changes in effective population size and recombination rates can affect the extent and efficacy of selection across the genome (Haudry et al. 2008, Hazzouri et al. 2012), how elevated homozygosity can influence the genetics of adaptive evolution (Glémin & Ronfort 2013), and how self-fertilization can affect population structure and demographic patterns (Platts et al. 2010). In addition, the prevalence of ploidy transitions in plants (Otto & Whitton 2000) enables investigations into the effects of genome duplication and genetic redundancy on the rates of fixation of deleterious and advantageous alleles (Otto & Yong 2002), and results to date highlight the importance of gene dosage and dominance as evolutionary factors affecting patterns of selection in polyploid lineages (Otto 2007). Finally, widespread gene expression and selection during the haploid stage of plant life cycles (Walbot & Evans 2003, Borg et al. 2009) offer insight into the unique characteristics of plants and may explain some of the differences between plant and animal species in the patterns of genome-wide variation and selection.

Here, we review recent developments in our understanding of the extent and nature of selection in plant genomes. We begin by documenting recent results regarding the extent of selection in both coding and noncoding DNA and by discussing how and why the patterns in plants differ from those in other organisms. We then review the expected and observed differences in the patterns of selection in plants as they relate to differences in effective population size and structure, recombination rate, mating system, and ploidy level.

## GENOME-WIDE SELECTION IN MODEL SYSTEMS

With the increasing availability of whole-genome sequence data, several methods have been applied to infer the genome-wide extent of positive and negative selection. These methods involve a variety of approaches, including analyses of selective constraint across species (Siepel et al. 2005), comparisons of polymorphism and divergence at selected and neutral sites (McDonald & Kreitman 1991, Boyko et al. 2008, Eyre-Walker & Keightley 2009), and measurements of levels of neutral diversity surrounding fixed functional mutations (Sattath et al. 2011). These methods have been reviewed recently (Sella et al. 2009, Zhen & Andolfatto 2012), and here we focus on the patterns that have emerged when they have been applied to estimate the extent and strength of positive and negative selection.

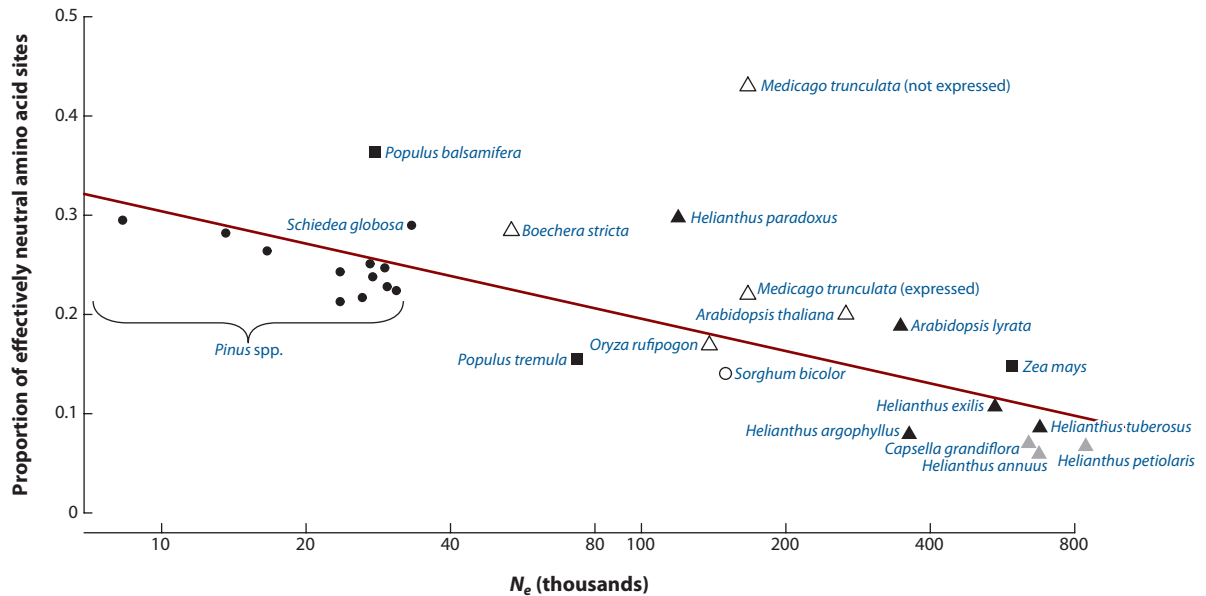
In *Drosophila*, two particular results have emerged from recent studies that are in marked contrast to previously held views about the nature of molecular evolution (Sella et al. 2009). First, comparative and population genomics analyses of selective constraint have found that a large fraction of the *Drosophila* genome, including noncoding regions, is subject to purifying selection (Andolfatto 2005, Sella et al. 2009). Second, analyses of genome-wide polymorphism



**Figure 1**

Estimates of the percent of the genome under selective constraint (*blue*) and the percent of selectively constrained bases that are noncoding (*red*) in model species from comparative genomics approaches (analysis of between-species constraint). Species are shown in decreasing order of genome size. For humans and *Drosophila*, light bars indicate the increase in the estimates when population genomics approaches (analysis of within-species diversity and allele frequencies) were used. Comparative genomics estimates were obtained for humans (Lindblad-Toh et al. 2011), *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae* (Siepel et al. 2005), *Arabidopsis thaliana*, and *Arabidopsis lyrata* (Haudry et al. 2013); population genomics estimates were obtained for humans (Ward & Kellis 2012) and *Drosophila* (Sella et al. 2009).

and divergence suggest that positive selection in coding and noncoding DNA is occurring at a remarkably high rate (Eyre-Walker & Keightley 2009, Sella et al. 2009, Sattath et al. 2011). In contrast, and more in line with expectations of neutral theory, such high rates of species-wide positive selection from new mutations have not been detected in the human genome (Eyre-Walker & Keightley 2009, Hernandez et al. 2011), which also appears to contain a larger proportion of DNA subject to little or no selective constraint (Eory et al. 2009, Eyre-Walker & Keightley 2009, Lindblad-Toh et al. 2011) (**Figure 1**). This latter result is in apparent contrast to recent results from the ENCODE project (ENCODE Project Consortium 2013), which suggest that a very high proportion of sites are involved in biochemical activity. However, although the integration of ENCODE results with population genetics approaches has resulted in increased estimates of contemporary constraint that was undetected from comparative approaches (Ward & Kellis 2012) (**Figure 1**), many sites falling under the ENCODE definition of function may not be subject to purifying selection (Graur et al. 2013). This does not imply that the human genome has few functional noncoding regions, however, as it may simply contain more neutral or nearly neutral DNA. Indeed, estimates of the amount of functional noncoding sequence in humans are higher than those in *Drosophila* (**Figure 1**). Nevertheless, analyses of protein-coding and noncoding sequences indicate weaker purifying selection, more neutral sites, and lower rates of positive



**Figure 2**

Estimate of the proportion of effectively neutral sites ( $N_e s < 1$ ) in various plant taxa calculated using the methods of Eyre-Walker & Keightley (2009) plotted against estimates of their effective population size ( $N_e$ ): Symbols indicate (*filled symbols*) predominant outcrossers, (*empty symbols*) predominantly selfing, (*squares*) species with a recent whole-genome duplication (synonymous substitution rate among duplicates  $K_s < 0.3$ ), (*triangles*) species with ancient whole-genome duplications ( $K_s > 0.3$ ), (*circles*) species for which the age of the last whole-genome duplication is unknown, and (*gray symbols*) species in which  $\alpha$  (the proportion of fixations driven by positive selection) is significantly greater than zero. The regression line shown was calculated excluding the data from *Medicago trunculata* (not expressed). Estimates of the proportion of effectively neutral sites are from the following for the various species indicated: Paape et al. (2013) for *M. trunculata*; Slotte et al. (2010) for *Capsella grandiflora* and *Arabidopsis thaliana*; Eckert et al. (submitted manuscript; see end of caption for details regarding the submitted manuscript referenced here) for *Pinus* spp.; Strasburg et al. (2011) for all *Helianthus*; and Gossmann et al. (2010) for the remaining species. Estimates of  $N_e$  were obtained from Gossmann et al. (2010), Gossmann et al. (2012), and Strasburg et al. (2011), with the following exceptions:  $N_e$  estimates for *Pinus* species were calculated using the diversity levels reported in Eckert et al. (submitted manuscript), a per-year mutation rate of  $10^{-9}$  (Willyard et al. 2007), and assuming a generation time of 75 years [based on arguments in Eckert et al. (submitted manuscript)]. For *Arabidopsis lyrata* and *Boecheera stricta*,  $N_e$  estimates were updated using direct estimates of the per-generation substitution rate in *Arabidopsis thaliana* following Gossmann et al. (2012). For *M. truncatula*, we used the estimate of per-generation substitution rate from Young et al. (2011) and synonymous diversity estimates from Branca et al. (2011). *M. trunculata* (not expressed) reflect data from sites that were not expressed in any of six tissues sampled, whereas *M. trunculata* (expressed) reflect data in genes that were expressed in at least one tissue (Paape et al. 2013). Duplication-age estimates were obtained from Blanc & Wolfe (2004), Barker et al. (2008), and Sterck et al. (2005). Submitted manuscript: (Eckert et al.) A.J. Eckert, A.D. Bower, K.D. Jermstad, J.L. Wegrzyn, B.J. Knaus, J.V. Syring & D.B. Neale.

selection in the human genome than in *Drosophila* (Eory et al. 2009, Eyre-Walker & Keightley 2009, Hernandez et al. 2011).

In plants, early analyses of protein-coding sequences in *Arabidopsis* found little evidence of positive selection on amino acid substitutions but did indicate a large proportion of slightly deleterious amino acid substitutions segregating in populations (Bustamante et al. 2002). More recent larger-scale analyses have confirmed these results (Foxe et al. 2008, Gossmann et al. 2010, Slotte et al. 2011). Indeed, approximately 20% of new amino acid mutations are effectively neutral in *Arabidopsis thaliana* (Figure 2), and estimates of species-wide positive selection on amino acid variants are approximately zero.

Although evidence for purifying selection on noncoding sites has been obtained from several plant genomes (Wright & Andolfatto 2008), the first studies of conservation in such regions suggested that there are considerably fewer conserved noncoding sites in plant than in animal genomes (Lockton & Gaut 2005). This initial conclusion has been confirmed by a recent study of genome-wide constraint in *Arabidopsis* (Haudry et al. 2013) (**Figure 1**). This suggests that plant genomes in the *Brassicaceae* contain many fewer selectively constrained noncoding sites than the genomes of mammals (Lindblad-Toh et al. 2011), *Drosophila* (Andolfatto 2005, Sella et al. 2009), and nematodes (Siepel et al. 2005), though estimates are higher in plants than in yeast (Siepel et al. 2005) (**Figure 1**). If low amounts of functional noncoding sequence are generally found in plant genomes, plants may have fewer (or smaller) regulatory regions or selection may be weaker in these regions compared with animals. If so, modulation of gene expression in plants may occur less often through complex noncoding regulatory mechanisms and more often through gene duplication (see Lockton & Gaut 2005).


Similar to humans, but in contrast with *Drosophila*, a significant proportion of sites in plant genomes appears to evolve effectively neutrally. In particular, only 6% of the *Arabidopsis* noncoding sequence is estimated to be under purifying selection (Haudry et al. 2013). This proportion of constraint is expected to decrease in larger plant genomes with more repetitive elements (Tenaillon et al. 2002). Indeed, the *Arabidopsis lyrata* genome has considerably more repetitive elements (Hu et al. 2011), and estimates of the proportion of constrained sites are lower than in *A. thaliana* (**Figure 1**). In contrast, approximately two-thirds of the *Drosophila* noncoding sequence is estimated to be under purifying selection, implying a much lower proportion of selectively neutral sites in this group (Andolfatto 2005, Sella et al. 2009).

## POSITIVE AND NEGATIVE SELECTION IN PLANTS

With many data sets from a variety of plant species now available, it is becoming possible to assess the extent to which the patterns of selection in *Arabidopsis* are representative of other species. In **Figure 2**, we show estimates of the proportion of effectively neutral amino acid mutations and indicate those species with significantly positive estimates of the proportion of amino acid divergence fixed by positive selection ( $\alpha$ ). Estimates of  $\alpha$  and the proportion of effectively neutral sites are also shown in **Supplemental Table 1** (Follow the **Supplemental Material** link in the online version of this article or at <http://www.annualreviews.org/>). These estimates are based on extensions of the McDonald-Kreitman test (McDonald & Kreitman 1991), and we have restricted our data set to include only results obtained using the likelihood-based approach of Eyre-Walker & Keightley (2009).

Although parameter estimates in many of these species are similar to those in *Arabidopsis*, there is considerable variation in the proportion of neutral amino acid sites across species (**Figure 2**), ranging from ~5% to ~30%. In addition, three species, *Helianthus annuus* (Strasburg et al. 2011), *Capsella grandiflora* (Slotte et al. 2010), and *Helianthus petiolaris* (Strasburg et al. 2011), show evidence for significant genome-wide positive selection (**Figure 2**). Interestingly, estimates of the proportion of selectively neutral sites are in some cases strongly divergent across close relatives (e.g., in *Populus*). This highlights the significant variation that exists across plant species in the patterns and strength of selection.

Do differences in estimates of selection across plant species reflect biological differences or statistical noise? In some cases, uncertainty in parameter estimates can be very high, particularly in systems with relatively small numbers of sampled loci, and differences in the choice of loci can also contribute to this uncertainty if selection parameters vary strongly across genes. For example, recent population genomic comparisons of genes with and without evidence of gene expression

 Supplemental Material

in *Medicago truncatula* have shown strong heterogeneity in the strength of purifying selection (**Figure 2**) (Paape et al. 2013). Furthermore, differences in divergence time from the outgroup used can also affect estimates of positive selection (Keightley & Eyre-Walker 2012). However, comparisons among species at identical (Strasburg et al. 2011) or comparable (Slotte et al. 2010) loci, and with similar divergence times, have shown significant between-species differences in positive and negative selection. We now consider several possible factors that may contribute to this variation.

## EFFECTIVE POPULATION SIZE

Effective population size ( $N_e$ ) is a key parameter involved in determining between-species differences in the strength of selection. In particular,  $N_e$  describes the extent to which evolutionary change is caused by genetic drift; a lower  $N_e$  implies a greater effect of drift (Charlesworth 2009). Smaller populations will thus experience greater effects of drift relative to selection, and if a significant proportion of sites are subject to weak selection, as predicted by nearly neutral theory (Ohta 1992), slightly deleterious mutations will be more likely to segregate at higher frequencies and fix in small populations (Akashi et al. 2012). In addition, populations with small  $N_e$  are expected to have a lower probability of fixing beneficial mutations (Gossmann et al. 2011, Akashi et al. 2012, Gossmann et al. 2012). Differences in  $N_e$  have thus been suggested to have had important effects on the strength of positive and negative selection in many genomes, including humans, *Drosophila*, and mice (Eyre-Walker & Keightley 2009, Langley et al. 2012). In plants, there is considerable variation in levels of neutral diversity (Leffler et al. 2012), range size (Brown et al. 1996), and census population size (Ness et al. 2010). This suggests that there is also variation in  $N_e$ , which may have important effects on patterns of selection in this group.

Estimates of  $N_e$  from neutral polymorphism data negatively correlate with estimates of the proportion of unconstrained sites across diverse organisms, including comparisons across plant species (**Figure 2**) (Wright & Andolfatto 2008, Gossmann et al. 2010, Akashi et al. 2012). This implies that species with larger  $N_e$  generally experience stronger purifying selection. Although broad-scale comparisons may be complicated by changes in the strength of selection and assumptions about mutation rates and generation times used to estimate  $N_e$  [particularly for trees such as *Pinus* and *Populus* (**Figure 2**)], comparisons across species within genera have also demonstrated that  $N_e$  affects the strength of purifying selection in sunflowers (Strasburg et al. 2011) and soft pines (A.J. Eckert, A.D. Bower, K.D. Jermstad, J.L. Wegrzyn, B.J. Knaus, J.V. Syring & D.B. Neale, submitted manuscript) (**Figure 2**). In addition, comparisons of different *A. thaliana* populations have shown that major-effect amino acid changes occur at a high frequency in smaller populations, consistent with the effect of increased genetic drift (Cao et al. 2011). Although the number of species investigated is still small, the general pattern emerging from these studies is that variation in  $N_e$  has a considerable effect on the strength of purifying selection (**Figure 2**).

Similarly, studies have assessed the effects of  $N_e$  on the extent of positive selection (as estimated by  $\alpha$ ). Evidence for more positive selection in species with large  $N_e$  was found, for example, in pairwise comparisons in *Capsella* and *Arabidopsis* (Slotte et al. 2010), across six sunflower and four lettuce species (Strasburg et al. 2011), and across seven diverse plant taxa (Gossmann et al. 2012). In contrast, no signs of positive selection and no correlation with  $N_e$  were found in soft pine species (A.J. Eckert, A.D. Bower, K.D. Jermstad, J.L. Wegrzyn, B.J. Knaus, J.V. Syring & D.B. Neale, submitted manuscript).

Despite growing evidence for a significant effect of  $N_e$  on patterns of positive selection, there are several concerns regarding the interpretation of the  $N_e$ - $\alpha$  correlation. First, the vast majority

of point estimates of positive selection ( $\alpha$ ) are negative, counterintuitively implying fewer than zero positively selected substitutions. However,  $\alpha$  is estimated by fitting a model of purifying selection to polymorphism data and calculating the excess nonsynonymous divergence above the model's expectations (Eyre-Walker & Keightley 2009). A negative value indicates that fewer than expected amino acid mutations have fixed, which could be caused by various factors including local adaptation on amino acids, balancing selection, and/or an inaccurate account of purifying selection assuming either an incorrect distribution of selective effects (Kousathanas & Keightley 2013) or an inaccurate demographic history. Given the large number of negative estimates, part of the correlation between  $\alpha$  and  $N_e$  could result from model misspecification (A.J. Eckert, A.D. Bower, K.D. Jermstad, J.L. Wegrzyn, B.J. Knaus, J.V. Syring & D.B. Neale, submitted manuscript).

Second, species in which there is evidence for genome-wide positive selection have the lowest proportions of effectively neutral sites (**Figure 2**). Although this is expected if both are governed by differences in  $N_e$ , there is also an automatic effect from the measurement of  $\alpha$ ; populations with more effectively neutral mutations will experience a higher fixation rate of nonadaptive mutations, bringing down the proportion of positively selected fixations without necessarily changing the rate of positive selection (Gossmann et al. 2010). However, studies that correct for this by measuring the rate of positive selection, rather than  $\alpha$ , have made similar conclusions (Strasburg et al. 2011, Gossmann et al. 2012). Nevertheless, the signal of positive selection can be still be masked by high numbers of slightly deleterious mutations (Eyre-Walker & Keightley 2009, Gossmann et al. 2012) because the signal of excess between-species divergence due to positive selection can be “swamped out” by neutral and nearly neutral fixations. Thus, there is an unfortunate anticonservative loss of power to detect positive selection with decreasing  $N_e$ .

Third, estimates of positive selection using methods such as that of Eyre-Walker and Keightley (2009) are dependent on assumptions of constant population size since species divergence (McDonald & Kreitman 1991, Eyre-Walker 2002). Smaller past population sizes could have led to greater amino acid divergence than expected, possibly inflating estimates of  $\alpha$  in species with large present-day  $N_e$ . However, evidence for greater signals of selective sweeps in genomic regions experiencing higher rates of protein evolution in *C. grandiflora* provides independent support for genome-wide positive selection (E. Josephs, R. Williamson & S. Wright, unpublished results).

Thus, consistent evidence indicates an effect of  $N_e$  on the strength of purifying selection. There is also evidence that some species with large  $N_e$  experience high rates of positive selection (**Figure 2**). However, inferences regarding the role of  $N_e$  on rates of positive selection should be treated with caution, as they have been limited by our ability to quantify adaptive substitution rates, particularly in species with small  $N_e$ .

## POPULATION STRUCTURE, SELECTION ON STANDING VARIATION, AND POLYGENIC ADAPTATION

Another important contributor to selection variation in plant populations is population subdivision. Plants vary widely in their extent of between-population and between-species gene flow (Morjan & Rieseberg 2004, Renaut et al. 2013); this variation in gene flow can have several important effects on selection. In addition to effects on  $N_e$  (Whitlock 2003), strong subdivision can slow the spread of advantageous mutations, thereby weakening the effects of selective sweeps (Barton 2000, Kim & Maruki 2011). With strong subdivision, adaptive mutations may also be more likely to occur independently across distinct geographic locations, even when there is a global selection pressure (Ralph & Coop 2010). With environmental heterogeneity, adaptation may also often be local, as shown repeatedly through reciprocal transplant experiments in plants (Kawecki & Ebert

---

**Selective sweeps:**  
the process in which positive selection increases the frequency of beneficial alleles together with linked alleles that may not be direct targets of selection

---

2004). Thus, we expect that species with more subdivided populations will have lower rates of range-wide positive selection.

Given that the estimates of positive and negative selection shown in **Figure 2** have been obtained primarily using scattered samples across populations, there could be considerable undetected local positive selection due to population subdivision (Keller et al. 2012). In addition, local adaptive events may be wrongly inferred to be slightly deleterious variants in scattered population samples, because they will generate an excess of nonsynonymous polymorphism (Gossmann et al. 2010). One way to assess this is to apply McDonald-Kreitman tests to local population samples. Such tests have typically found estimates of positive and negative selection that are similar to those of species-wide samples (Foxe et al. 2008, Gossmann et al. 2010). However, migration events from other populations may continue to erode the signal of positive selection and inflate estimates of slightly deleterious mutations in within-population samples using this approach.

A more powerful approach is to examine the distribution of between-population differentiation across genes, with the expectation that loci subject to local adaptation will have elevated between-population differentiation. Such studies have successfully identified loci and genomic regions involved in local adaptation, and the results have highlighted that signals of local adaptation can be considerably stronger than signals of species-wide selection (Turner et al. 2010, Chen et al. 2012, Hufford et al. 2012, Keller et al. 2012). However, these studies typically rely on comparisons of candidate loci with assumed neutral loci or take an outlier approach. Thus, they are typically more useful for identifying underlying candidate targets of selection, such as flowering time genes (Chen et al. 2012, Keller et al. 2012) and genes important for heavy metal tolerance (Turner et al. 2010), than for quantifying and comparing rates of local adaptation.

Another possible source of undetected positive selection is adaptation involving many loci each with small effects on a phenotype and other forms of selection from standing genetic variation (“soft selective sweeps”) (Chevin & Hospital 2008, Pritchard et al. 2010). Under such scenarios, the footprint of positive selection on neutral variation can be weak or absent even within populations and may involve subtle allele frequency changes across many loci rather than high rates of differentiation and strong selective sweeps (Pritchard & Di Rienzo 2010, Le Corre & Kremer 2012). Such selection may be much more predominant than species-wide selection from new mutations and may be the predominant source of adaptive evolution in humans (Pritchard et al. 2010).

A potentially powerful approach for assessing signals of local adaptation that may arise from standing variation or polygenic selection is to test for correlations of single nucleotide polymorphism frequencies with environment, under the prediction that consistent environmental changes will drive significant allele frequency changes (Coop et al. 2010). This has recently been applied to whole-genome polymorphism data from worldwide *A. thaliana* samples (Hancock et al. 2011), and correlations with several environmental variables were found. Additional recent findings show that loci with signals of environmental adaptation exhibit excess neutral diversity, as expected if environmental heterogeneity is maintaining variation across the geographic range (Lee & Mitchell-Olds 2012). Significant environmental correlations have also been recently observed in other systems, including black spruce (Prunier et al. 2012) and loblolly pine (Eckert et al. 2010). Thus, for several plants showing little sign of species-wide selective sweeps, the potential for considerable local adaptation remains, although estimation and comparison of genome-wide rates of local adaptation is difficult partly because of possible residual confounding effects of population history (Coop et al. 2009) and because it is unlikely that highly polygenic adaptation will show sufficient allele frequency differentiation to be detected (Le Corre & Kremer 2012).



## RECOMBINATION RATE

Recombination rate variation within and between species is another well-recognized factor influencing selection (Nachman 2002). High recombination rates are expected to cause selection to act more independently across sites in the genome, whereas low recombination can cause interference between both selected and neutral sites owing to increased linkage (Hill & Robertson 1966). Neutral diversity and the efficacy of selection are thus expected to be reduced when recombination is low; this reduction can be caused by the fixation of beneficial mutations (selective sweeps) (Maynard Smith & Haigh 1974) and the elimination of deleterious mutations (background selection) (Charlesworth et al. 1993).

The effects of selective sweeps and background selection depend strongly on the rate of recombination and are expected to generate a positive correlation between recombination rate and both nucleotide diversity and the efficacy of natural selection. In agreement with this expectation, diversity is significantly reduced in genomic regions with low rates of recombination, including centromeres (Carneiro et al. 2008), telomeres (Savage et al. 2005), and Y chromosomes in animals (Bachtrog & Charlesworth 2002, Hellborg 2003). Indeed, variation in the rate of recombination is probably a major determinant of diversity throughout *Drosophila* genomes, and there is evidence that polymorphism increases with crossover rates in *D. melanogaster* (Andolfatto 2007), *D. simulans* (Begun et al. 2007), and *D. pseudoobscura* (Noor 2008). Similarly, there is evidence for a positive correlation between recombination rate and nucleotide diversity across the human genome (Lercher & Hurst 2002). Furthermore, several studies in *Drosophila* have provided evidence for a reduced efficacy of both purifying and positive selection in regions of low recombination, as expected if these regions experience more selective interference (Betancourt & Presgraves 2002, Campos et al. 2012).

Unlike in animals, a major effect of recombination rate on sequence diversity was initially difficult to establish in plants, despite much interest and study (Baudry et al. 2001, Tenaillon et al. 2002, Nordborg et al. 2005, Wright et al. 2006, Haudry et al. 2008). For example, there is little evidence for a correlation between recombination rate and neutral diversity in *A. thaliana* or its outcrossing relatives, and low-recombining regions close to centromeres show evidence for elevated diversity (Nordborg et al. 2005, Wright et al. 2006). In wild tomato, there is some evidence for a positive correlation between recombination and polymorphism, although the relationship is weak (Baudry et al. 2001). Although early work in maize did not indicate a positive correlation between recombination and diversity (Tenaillon et al. 2002), more recent whole-genome polymorphism studies found a significant positive correlation (Gore et al. 2009). Diversity is also significantly positively correlated with recombination rates in *M. truncatula* (Branca et al. 2011). Finally, recombination correlates positively with the efficacy of purifying selection in the Triticeae (Escobar et al. 2010), and strongly reduced diversity and relaxed selection have been observed in the young Y chromosomes of a few plant species (Bergero & Charlesworth 2011, Chibalina & Filatov 2011).

Thus, although a growing number of studies have shown an effect of recombination rate on diversity and selection, this pattern is not universal. There are several possible explanations for the lack of association between recombination and diversity in some plant genomes, and an emerging picture is that gene density plays an important role. For example, in self-fertilizing *A. thaliana*, both recombination rates and gene density vary across chromosomes (Singer et al. 2006). Although there is little evidence that diversity correlates with recombination rate (Nordborg et al. 2005, Wright et al. 2006), it does correlate significantly with gene density (Schmid 2004, Nordborg et al. 2005). Similarly, using single-nucleotide polymorphism data from domesticated and wild rice, Flowers et al. (2012) did not find a positive correlation between nucleotide diversity and

---

### Background

**selection:** the process in which purifying selection eliminates deleterious alleles together with linked alleles that may not be direct targets of selection

---

recombination rate, but they did find a significant negative correlation between diversity and gene density. Evidence for lower recovery of polymorphism following domestication in gene-rich regions of maize (Hufford et al. 2012) is also consistent with this effect. Finally, whereas recombination and diversity are significantly correlated in *Medicago*, there is a stronger negative correlation with gene density (Branca et al. 2011).

The studies above highlight that, in addition to recombination rate, the density of selected mutations has important effects on the strength of interference. Accordingly, linked selective effects should most strongly affect diversity in regions with both high gene density and low recombination. Consistent with this, recent studies in *Caenorhabditis* have found strongly reduced diversity in regions with both low recombination and high gene density (Cutter & Choi 2010). In contrast, regions of low recombination in *Arabidopsis* are particularly gene poor, which likely erodes the correlation between recombination and diversity (Nordborg et al. 2005, Wright et al. 2006).

## MATING SYSTEM

In addition to recombination rate heterogeneity across the genome, the extensive mating system variability in plants (Barrett 2002) can lead to important between-species differences in effective rates of recombination. In particular, inbreeding populations are expected to have higher homozygosity and, therefore, lower effective rates of recombination (Nordborg 2000). As a result, levels of linkage disequilibrium are elevated, which should, in turn, reduce genetic diversity and the efficacy of selection (reviewed by Glémin et al. 2006).

An effective way to study the consequences of inbreeding on selection and diversity patterns is to compare DNA sequences from closely related species that differ in rates of self-fertilization. Studies using this approach have shown clear evidence for reduced nucleotide diversity in selfing species such as *Arabidopsis* (Nordborg et al. 2005, Ross-Ibarra et al. 2008), *Capsella* (Fuxe et al. 2009), *Leavenworthia* (Charlesworth & Yang 1998, Liu et al. 1998), *Eichbornia* (Ness et al. 2010), *Mimulus* (Sweigart & Willis 2003), *Lycopersicon* (Baudry et al. 2001, Roselius 2005), *Collinsia* (Hazzouri et al. 2012), and *Clarkia* (Pettengill & Moeller 2012). The patterns that emerge generally suggest a reduction in diversity greater than can be explained simply by including high levels of inbreeding in a neutral model, indicating that selective interference may play an important role, although demographic factors also likely contribute.

In contrast with patterns of diversity, early studies using small numbers of loci and comparisons of relative rates of nucleotide substitution provided few signs of reduced efficacy of natural selection in selfing species (Wright et al. 2002, Haudry et al. 2008, Escobar et al. 2010). However, more recent genome-wide studies have found evidence for reduced efficacy of selection on non-synonymous sites in selfing *Capsella* (Slotte et al. 2013), *Eichbornia* (Ness et al. 2012), and *Collinsia* (Hazzouri et al. 2012). Furthermore, recent analyses have suggested a decline in the efficacy of selection on codon usage bias in selfing *Arabidopsis* and *Capsella* compared with their outcrossing congeners (Qiu et al. 2011). Given that these comparisons include very recently derived selfing lineages (e.g., in *Eichbornia* and *Capsella*), detectable shifts in the efficacy of selection may happen rapidly.

There is less evidence for a reduced efficacy of positive selection in selfing lineages, however. Such a finding may not be expected if a large proportion of beneficial mutations are recessive, which can fix more rapidly in selfing populations (Glémin 2007), as discussed below. Nevertheless, plant species providing evidence for positive selection are all highly outcrossing (**Figure 2**), and a pairwise comparison of *A. thaliana* with the closely related obligately outcrossing *C. grandiflora* indicates a significant decline in the efficacy of positive selection in selfing species (Slotte et al. 2010).

The above patterns suggest that recombination rate variation caused by mating system evolution plays an important role in driving heterogeneity in diversity and selection, but it is difficult to distinguish the roles of linked selection from demographic history and population structure. For example, in colonizing selfing species such as *A. thaliana* and *Eichbornia paniculata*, population bottlenecks and increased subdivision associated with mating system differences may be a major cause of reduced effective population sizes (Ness et al. 2010). Furthermore, founder events associated with shifts to selfing (Foxy et al. 2009) can also be important factors reducing diversity and the efficacy of selection. Although genetic hitchhiking may still play an important role, demographic effects may also be major contributors to patterns of selection in selfing species.

In addition to the effects of selfing on recombination rate and demographic history, the dynamics of positive selection in selfers are also expected to be influenced by high levels of homozygosity (Glémin & Ronfort 2013). In particular, selfers are more likely to fix recessive beneficial mutations and less likely to experience selection on standing variation (Charlesworth 1992, Glémin 2007, Glémin & Ronfort 2013). Recent evidence that self-fertilizing crops have higher numbers of recessive quantitative trait loci contributing to phenotypic evolution is consistent with this prediction (Ronfort & Glémin 2013).

## PLOIDY

Another major source of variation expected to have important consequences for both positive and negative selection is differences in ploidy across species and the genome. Because of changes in both the effective population size and the extent of masking of mutations, differences in ploidy are predicted to have important effects on rates of positive selection, the efficacy of purifying selection, and the mutation load, as discussed by Otto & Whitton (2000). For example, species with higher ploidy levels may have lower mutation loads following polyploidization, but over time, deleterious mutations in their genomes may reach higher frequencies owing to the greater effects of masking. Similarly, if most beneficial mutations are recessive, masking can cause the rate of positive selection to be reduced in species with higher ploidy levels. By contrast, dominant beneficial mutations and selection from standing variation could lead to a greater extent of positive selection. The effects of ploidy on positive and negative selection thus depend in important ways on parameters such as dominance coefficients and the time since changes in ploidy.

In plants, gene-level differences in ploidy are likely to be important factors driving differences in selection. In particular, extensive gametophytic expression (Borg et al. 2009) means that a subset of genes will be expressed in the haploid phase, potentially contributing to differences in the efficacy of positive and negative selection across the genome. Evidence for greater selection on intron length in genes expressed in pollen (Seoighe et al. 2005) is consistent with this hypothesis. Furthermore, there is evidence that Y-chromosome degeneration may be significantly reduced at genes expressed in the haploid phase, perhaps contributing to a lower extent of relaxed selection in plants compared with that of animal Y chromosomes (Chibalina & Filatov 2011). Similarly, recent genome-wide estimates of positive and negative selection in *C. grandiflora* suggest that pollen-expressed genes show higher rates of both positive and negative selection (R. Arunkumar, E. Josephs, R. Williamson & S. Wright, unpublished results).

Transitions to higher ploidy levels are also common in plants (Bowers et al. 2003) and are associated with 15% of angiosperm and 30% of fern speciation events (Wood et al. 2009). Indeed, many of the model plant groups that have been used to study selection at the molecular level have undergone ancient polyploidization at various time points (Jiao et al. 2011), including *Arabidopsis* (*Arabidopsis* Genome Initiative 2000, Ku et al. 2000), wheat (Brenchley et al. 2012), maize (Messing 2009), and several species in the grass family (Levy & Feldman 2002). Furthermore, patterns of

duplicate gene divergence suggest that most plants have undergone multiple rounds of whole-genome duplication in their past (Blanc & Wolfe 2004). This means that differences in the extent and timing of polyploidization events may be important contributors to differences in the efficacy of selection among species.

To date, most studies examining selection associated with polyploidization have focused on cases of disomic inheritance caused by ancient allopolyploidization between species. In this case, polyploidization has led to gene duplication across the entire genome. In these systems, one signal of relaxed purifying selection following polyploidization is the ongoing silencing and loss of gene duplicates (Lynch & Conery 2000). Evidence from several systems suggests that gene duplicate retention and loss are not random with respect to biological function. In particular, retained gene duplicates are often dosage dependent and involved in transcription and signal transduction. This pattern is seen across independent polyploidization events, including *Arabidopsis* (Blanc & Wolfe 2004), yeast (Seoighe & Wolfe 1999), and *Paramecium* (Aury et al. 2006), indicating that gene dosage is an important factor determining the extent to which genes experience relaxed selection following duplication (Veitia 2004, Freeling & Thomas 2006, Hakes et al. 2007, Birchler & Veitia 2009; but see Barker et al. 2008).

The extent of gene retention and loss in polyploid lineages is expected to be an ongoing process reflecting the timing of genome duplication. For example, approximately 72% of duplicate genes have been retained over the 11 Ma following polyploid formation in maize (Gaut & Doebley 1997). In contrast, in *Arabidopsis*, three older rounds of whole-genome duplication events have occurred (>50 Mya) (Beilstein et al. 2010), with considerably fewer retained duplicates (less than 17%) (Maere et al. 2005). Accordingly, the maize genome may still be in the early stages of diploidization, with many duplicate genes experiencing relaxed selection prior to loss. Thus, differences across species in the timing of the most recent polyploidization event can have major effects on the proportion of effectively neutral substitutions and on rates of positive selection. Such effects could contribute to a lack of evidence for positive selection and less purifying selection in maize than in other outcrossing species with large  $N_e$  (Figure 2).

In addition to the direct effects of dominance, rates of beneficial mutation are also likely affected by changes in ploidy. Following polyploidization, extensive genome-wide structural rearrangements can occur (Song et al. 1995), potentially leading to changes in gene expression and increased genomic variability in polyploid populations (Osborn et al. 2003). With the relaxed selective constraint following genome duplication combined with a greater mutational input of beneficial mutations, positive selection for novel functions in gene duplicates may be increased (Walsh 1995, Otto & Whitton 2000, Otto & Yong 2002). In *A. thaliana*, there is evidence for extensive divergence and asymmetric rates of protein sequence evolution among ancient duplicate gene pairs (Blanc & Wolfe 2004, Hu et al. 2012), suggesting selection has been involved in driving functional divergence. Some studies have also identified a significant signal of positive selection on duplicates (Wendel 2000, Ho-Huu et al. 2012), and these results are consistent with a model in which the initial relaxation of selection caused by gene duplication is followed by functional divergence driven by positive selection (Innan & Kondrashov 2010, Ho-Huu et al. 2012).

Much less has been documented about the dynamics of positive and negative selection in autotetraploids with tetrasomic inheritance. In tetraploid maize, no effect of ploidy on either the effective population size or selection was found (Tiffin & Gaut 2001). In the autotetraploid *Arabidopsis arenosa*, the effective population size and the efficiency of selection appear to be higher than those of its diploid relatives (Hazzouri et al. 2008, Hollister et al. 2012). This species also appears to have experienced high rates of positive selection on genes involved in meiosis, possibly as a consequence of the genetic challenges associated with the shift in ploidy (Hollister et al. 2012).

## CONCLUSIONS AND FUTURE DIRECTIONS

The results to date suggest that plant populations vary greatly in the strength and genetic basis of positive and negative selection and that differences in effective population size, subdivision, mating system, recombination, and ploidy are playing important roles in determining this variation. However, recent results have also highlighted the difficulty of empirically distinguishing between the effects of different population genetic parameters in driving patterns of selection. In particular, the strongest signals of purifying and positive selection in plant genomes have been found in predominantly outcrossing species that have large effective population sizes, limited population subdivision, and no recent whole-genome duplication events, making the relative importance of each of these factors unclear. Furthermore, characterizing rates of positive selection in species with small  $N_e$  and strong population subdivision remains challenging.

To address some of these concerns, future large-scale genus and family-wide studies of positive and negative selection will be important. By acquiring a rich comparative data set of closely related species, researchers will be able to examine more quantitatively the relative and joint importance of changes in mating system, effective population size, ploidy, and subdivision. Of particular importance will be the examination of how changes such as shifts to selfing and polyploidy influence the strength of selection as a function of time since the transition. Future studies utilizing model-based approaches that integrate phylogenetics and population genomics will be important to quantify and compare changes in positive and negative selection across different lineages. Although such comparisons of selection across species are currently difficult to generate, recent work suggests that progress may be made by integrating large-scale comparative population genomics data sets with models that incorporate the effects of changes in demography, ploidy, and recombination.

### SUMMARY POINTS

1. The increased availability of genomic data has enabled broad-scale comparisons of the extent and strength of natural selection, and a growing number of such comparisons are being done in plant species. These studies indicate extensive heterogeneity across species in the strength of positive and negative selection.
2. Compared with animals, plants have fewer functional noncoding sites and a relatively low proportion of selectively constrained sites across their genomes.
3. Plant population structure and effective population size are recognized as key factors affecting the strength and nature of selection, though they can also complicate our ability to make inferences about selective processes.
4. Insights from closely related, self-fertilizing, and outcrossing plant populations suggest that variation in the degree of inbreeding can have significant and detectable consequences on patterns of selection.
5. Both recombination and the density of selected mutations affect the efficacy of selection and are important for determining the effects of selective sweeps and background selection on surrounding sites.
6. Ploidy differences are common in plants, and the effects of whole-genome duplications in polyploid species can have significant consequences on the extent of both positive and negative selection. Studies of polyploid genomes suggest that gene retention and loss of gene duplicates are often nonrandom and that the timing of polyploidization events influences differences among species in the efficacy of selection.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We are very grateful to A. Eckert, T. Gossmann, P. Tiffin, L. Rieseberg, and S. Renaut for sharing their unpublished manuscripts; T. Gossmann for discussion; and S.P. Otto for helpful comments on the manuscript.

## LITERATURE CITED

- Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. *Genetics* 192(1):15–31
- Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437(7062):1149–52
- Andolfatto P. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res.* 17(12):1755–62
- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6):796–815
- Aury JM, Jaillon O, Duret L, Noel B, Jubin C, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–78
- Bachtrog D, Charlesworth B. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* 416:323–26
- Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, et al. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol. Biol. Evol.* 25(11):2445–55
- Barrett S. 2002. The evolution of plant sexual diversity. *Nat. Rev. Genet.* 3:274–84
- Barton NH. 2000. Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. Ser. B* 355(1403):1553–62
- Baudry E, Kerdelhué C, Innan H, Stephan W. 2001. Species and recombination effects on DNA variability in the tomato genus. *Genetics* 158(4):1725–35
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5(11):e310
- Beilstein MAM, Nagalingum NSN, Clements MDM, Manchester SRS, Mathews SS. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 107(43):18724–28
- Bergero R, Charlesworth D. 2011. Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Curr. Biol.* 21(17):5
- Betancourt AJ, Presgraves DC. 2002. Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 99(21):13616–20
- Birchler JA, Veitia RA. 2009. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 186(1):54–62
- Blanc GG, Wolfe KHK. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16(7):1667–78
- Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective. *J. Exp. Bot.* 60(5):1465–78
- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422(6930):433–38
- Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4(5):e1000083
- Branca AA, Paape TDT, Zhou PP, Briskine RR, Farmer ADA, et al. 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* 108(42):E864–70

- Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, et al. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426):705–10
- Brown JH, Stevens GC, Kaufman DM. 1996. The geographic range: size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Syst.* 27:597–623
- Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. 2002. The cost of inbreeding in *Arabidopsis*. *Nature* 416(6880):531–34
- Campos JL, Charlesworth B, Haddrill PR. 2012. Molecular evolution in nonrecombining regions of the *Drosophila melanogaster* genome. *Gen. Biol. Evol.* 4(3):278–88
- Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, et al. 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43(10):956–63
- Carneiro M, Ferrand N, Nachman MW. 2008. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* 181(2):593–606
- Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. *Am. Nat.* 140(1):126–48
- Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10(3):195–205
- Charlesworth B, Morgan MT, Charlesworth DD. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–303
- Charlesworth DD, Yang ZZ. 1998. Allozyme diversity in *Leavenworthia* populations with different inbreeding levels. *Heredity* 81(Pt. 4):453–61
- Chen J, Källman T, Ma X, Gyllenstrand N, Zaina G, et al. 2012. Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics* 191(3):865–81
- Chevin L-ML, Hospital FF. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* 180(3):1645–60
- Chibalina MV, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* 21(17):1475–79
- Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, et al. 2009. The role of geography in human adaptation. *PLoS Genet.* 5(6):e1000500
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185(4):1411–23
- Cutter AD, Choi JY. 2010. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. *Gen. Res.* 20(8):1103–11
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, et al. 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185(3):969–82
- ENCODE Project Consortium. 2013. An integrated encyclopedia of DNA elements in the human genome. *Nature* 488(7414):57–74
- Eory L, Halligan DL, Keightley PD. 2009. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol. Biol. Evol.* 27(1):177–92
- Escobar JSJ, Cenci AA, Bolognini JJ, Haudry AA, Laurent SS, et al. 2010. An integrative test of the dead-end hypothesis of selfing evolution in *Triticeae* (Poaceae). *Evolution* 64(10):2855–72
- Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics* 162(4):2017–24
- Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* 26(9):2097–108
- Flowers JM, Molina J, Rubinstein S, Huang P, Schaal BA, Purugganan MD. 2012. Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Mol. Biol. Evol.* 29(2):675–87
- Foxe JP, Dar VUN, Zheng H, Nordborg M, Gaut BS, Wright SI. 2008. Selection on amino acid substitutions in *Arabidopsis*. *Mol. Biol. Evol.* 25(7):1375–83
- Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proc. Natl. Acad. Sci. USA* 106(13):5241–45

- Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genes Dev.* 16(7):805–14
- Gaut BSB, Doebley JFJ. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. USA* 94(13):6809–14
- Glémin S. 2007. Mating systems and the efficacy of selection at the molecular level. *Genetics* 177(2):905–16
- Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc. R. Soc. B* 273(1604):3011–19
- Glémin S, Ronfort J. 2013. Adaptation and maladaptation in selfing and outcrossing species: new mutations versus standing variation. *Evolution* 67(1):225–40
- Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, et al. 2009. A first-generation haplotype map of maize. *Science* 326(5956):1115–17
- Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Gen. Biol. Evol.* 4(5):658–67
- Gossmann TI, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, et al. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* 27(8):1822–32
- Gossmann TI, Woolfit M, Eyre-Walker A. 2011. Quantifying the variation in the effective population size within a genome. *Genetics* 189(4):1389–402
- Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E. 2013. On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Gen. Biol. Evol.* 5:578–90
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, et al. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334(6052):83–86
- Haudry A, Cenci A, Guilhaumon C, Paux E, Poirier S, et al. 2008. Mating system and recombination affect molecular evolution in four *Triticeae* species. *Genet. Res.* 90(1):97–109
- Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, et al. 2013. An atlas of over 90,000 conserved noncoding sequences yields a detailed map of crucifer regulatory regions. *Nat. Genet.* 45:891–98
- Hazzouri KM, Escobar JS, Ness RW, Newman LK, Randle AM, et al. 2012. Comparative population genomics in *Collinsia* sister species reveals evidence for reduced effective population size, relaxed selection and evolution of biased gene conversion with an ongoing mating shift. *Evolution* 67(5):1263–78
- Hazzouri KM, Mohajer A, Dejak SI, Otto SP, Wright SI. 2008. Contrasting patterns of transposable-element insertion polymorphism and nucleotide diversity in autotetraploid and allotetraploid *Arabidopsis* species. *Genetics* 179(1):581–92
- Hellborg L. 2003. Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol. Biol. Evol.* 21(1):158–63
- Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, et al. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331(6019):920–24
- Ho-Huu J, Ronfort J, De Mita S, Bataillon T, Hochu I, et al. 2012. Contrasted patterns of selective pressure in three recent paralogous gene pairs in the *Medicago* genus (L.). *BMC Evol. Biol.* 12(1):195
- Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. 2012. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* 8(12):e1003093
- Hu CC, Lin S-YS, Chi W-TW, Charng Y-YY. 2012. Recent gene duplication and subfunctionalization produced a mitochondrial GrpE, the nucleotide exchange factor of the Hsp70 complex, specialized in thermotolerance to chronic heat stress in *Arabidopsis*. *Plant Physiol.* 158(2):747–58
- Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43(5):476–81
- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44(7):808–11
- Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11(2):97–108
- Jiao Y, Wickett NJ, Ayyampalayam S, Chandrabali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–100
- Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecol. Lett.* 7:1225–41



- Keightley PD, Eyre-Walker A. 2012. Estimating the rate of adaptive molecular evolution when the evolutionary divergence between species is small. *J. Mol. Evol.* 74(1–2):61–68
- Keller SR, Levens N, Olson MS, Tiffin P. 2012. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* (L.). *Mol. Biol. Evol.* 29(10):3143–52
- Kim Y, Maruki T. 2011. Hitchhiking effect of a beneficial mutation spreading in a subdivided population. *Genetics* 189(1):213–26
- Kousathanas A, Keightley PD. 2013. A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* 193:1197–208
- Ku HM, Vision T, Liu J, Tanksley SD. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA* 97(16):9121–26
- Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. 2007. All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol.* 8(10):R209–R209
- Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8(3):269–94
- Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192(2):533–98
- Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.* 21(7):1548–66
- Lee C-R, Mitchell-Olds T. 2012. Environmental adaptation contributes to gene polymorphism across the *Arabidopsis thaliana* genome. *Mol. Biol. Evol.* 29(12):3721–28
- Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, et al. 2012. Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biol.* 10(9):e1001388
- Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* 18(7):337–40
- Levy A, Feldman M. 2002. The impact of polyploidy on grass genome evolution. *Annu. Rev. Plant Physiol.* 130(4):1587–93
- Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7):476–82
- Liu F, Zhang L, Charlesworth D. 1998. Genetic diversity in *Leavenworthia* populations with different inbreeding levels. *Proc. R. Soc. B* 265(1393):293–301
- Lockton S, Gaut BS. 2005. Plant conserved non-coding sequences and paralogue evolution. *Trends Genet.* 21(1):60–65
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–55
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, et al. 2005. Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. USA* 102(15):5454–59
- Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.* 23:23–35
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adb* locus in *Drosophila*. *Nature* 351(6328):652–54
- Messing J. 2009. The polyploid origin of maize. In *The Maize Handbook: Domestication, Genetics, and Genome*, ed. JL Bennetzen, SC Hake, pp. 221–38. New York: Springer Life Sci.
- Morjan CL, Rieseberg LH. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* 13(6):1341–56
- Nachman M. 2002. Variation in recombination rate across the genome: evidence and implications. *Curr. Opin. Genet. Det.* 12(6):657–63
- Ness RW, Siol M, Barrett SCH. 2012. Genomic consequences of transitions from cross- to self-fertilization on the efficacy of selection in three independently derived selfing plants. *BMC Genomics* 13:611
- Ness RW, Wright SI, Barrett SCH. 2010. Mating-system variation, demographic history and patterns of nucleotide diversity in the tristylous plant *Eichbornia paniculata*. *Genetics* 184(2):381–92
- Noor MAF. 2008. Connecting recombination, nucleotide diversity, and species divergence in *Drosophila*. *Fly* 2(5):255–56
- Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154(2):923–29

- Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3(7):e196
- Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23:263–86
- Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, et al. 2003. Understanding mechanisms of novel gene expression in polyploids. *Trend Genet.* 19(3):141–47
- Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* 131(3):452–62
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu. Rev. Ecol. Syst.* 34:401–37
- Otto SP, Yong P. 2002. The evolution of gene duplicates. *Adv. Genet.* 46:451–83
- Paape T, Bataillon T, Zhou P, Kono T, Briskine R, et al. 2013. Selection, genome-wide fitness effects and evolutionary rates in the model legume *Medicago truncatula*. *Mol. Ecol.* 22(13):3525–38
- Pettengill JB, Moeller DA. 2012. Tempo and mode of mating system evolution between incipient *Clarkia* species. *Evolution* 66(4):1210–25
- Platts A, Horton M, Huang YS, Li Y, Anastasio AE, et al. 2010. The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* 6(2):e1000843
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20(4):R208–15
- Pritchard JK, Di Rienzo A. 2010. Adaptation: not by sweeps alone. *Nat. Rev. Genet.* 11(10):665–67
- Prunier J, Gérardi S, Laroche J, Beaulieu J, Bousquet J. 2012. Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Mol. Ecol.* 21(17):4270–86
- Qiu S, Zeng K, Slotte T, Wright SI, Charlesworth D. 2011. Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Gen. Biol. Evol.* 3:868–80
- Ralph P, Coop G. 2010. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* 186(2):647–68
- Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, et al. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Commun.* 4:1827
- Ronfort J, Glémin S. 2013. Mating system, Haldane's sieve, and the domestication process. *Evolution* 67(5):1518–26
- Roselius K. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171(2):753–63
- Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, et al. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS One* 3(6):e2411
- Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. 2011. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. *PLoS Genet.* 7(2):e1001302
- Savage SA, Stewart BJ, Eckert A, Kiley M, Liao JS, Chanock SJ. 2005. Genetic variation, nucleotide diversity, and linkage disequilibrium in seven telomere stability genes suggest that these genes may be under constraint. *Hum. Mutat.* 26(4):343–50
- Schmid KJ. 2004. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* 169(3):1601–15
- Sella G, Petrov DA, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* 5(6):e1000495
- Seoighe C, Gehring C, Hurst LD. 2005. Gametophytic selection in *Arabidopsis thaliana* supports the selective model of intron length reduction. *PLoS Genet.* 1(2):e13
- Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* 2(5):548–54
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genes Dev.* 15(8):1034–50
- Singer T, Fan Y, Chang H-S, Zhu T, Hazen SP, Briggs SP. 2006. A high-resolution map of *Arabidopsis* recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet.* 2(9):e144
- Slotte T, Bataillon T, Hansen TT, St Onge K, Wright SI, Schierup MH. 2011. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Gen. Biol. Evol.* 3:1210–19
- Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol. Biol. Evol.* 27(8):1813–21

- Slotte T, Hazzouri KM, Ågren JA, Koenig K, Maumus F, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45(7):831–35
- Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. USA* 92(17):7719–23
- Sterck L, Rombauts S, Jansson S, Sterky F, Rouzé P, Van de Peer Y. 2005. EST data suggest that poplar is an ancient polyploid. *New Phytol.* 167(1):165–70
- Strasburg JL, Kane NC, Raduski AR, Bonin A, Michelmore R, Rieseberg LH. 2011. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Mol. Biol. Evol.* 28(5):1569–80
- Sweigart AL, Willis JH. 2003. Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution* 57(11):2490–506
- Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* 162(3):1401–13
- Tiffin P, Gaut BS. 2001. Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. *Genetics* 158(1):401–12
- Turner TL, Bourne EC, Wettberg Von EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42(3):260–63
- Veitia RA. 2004. Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. *Genetics* 168(1):569–74
- Walbot V, Evans MMS. 2003. Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* 4(5):369–79
- Walsh JB. 1995. How often do duplicated genes evolve new functions? *Genetics* 139(1):421–28
- Ward LD, Kellis M. 2012. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* 337(6102):1675–78
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol. Biol.* 42(1):225–49
- Whitlock MC. 2003. Fixation probability and time in subdivided populations. *Genetics* 164(2):767–79
- Willyard A, Syring J, Gernandt DS, Liston A, Cronn R. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Mol. Biol. Evol.* 24(1):90–101
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA* 106(33):13875–79
- Wright SI, Andolfatto P. 2008. The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu. Rev. Ecol. Evol. Syst.* 39(1):193–213
- Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Loseley M, et al. 2006. Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* 174(3):1421–30
- Wright SI, Lauga B, Charlesworth D. 2002. Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol. Biol. Evol.* 19(9):1407–20
- Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, et al. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480 (7378):520–24
- Zhen Y, Andolfatto P. 2012. Methods to detect selection on non-coding DNA. *Methods Mol. Biol.* 856:141–59