CHAPTER TWELVE

Population Genomics of Plant Species

Khaled M. Hazzouri[*], **Michael D. Purugganan**[*,†], **Jonathan M. Flowers**[*,†,1]
[*]Center for Genomics and Systems Biology, NYU Abu Dhabi Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[†]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, USA
[1]Corresponding author: e-mail address: jmf11@nyu.edu

Contents

Abstract

This chapter addresses fundamental questions in population genomics and molecular evolution of plants and their domesticated relatives. We review the evolutionary forces that shape genetic diversity within and among species and highlight recent exceptional examples from the literature to illustrate how whole genome sequencing of entire populations of plant genomes has accelerated the pace of discovery of the mechanisms of plant genome evolution. Throughout the chapter, we emphasize the role of mating system and population size in plant genome evolution and discuss why plant genomes frequently show limited evidence of adaptation and display features that are inconsistent with simple models of linked selection. We close by highlighting recent studies that illustrate a range of discoveries enabled by sequencing of populations of plant genomes.

Advances in Botanical Research, Volume 69
ISSN 0065-2296
http://dx.doi.org/10.1016/B978-0-12-417163-3.00012-3

© 2014 Elsevier Ltd
All rights reserved.

1. INTRODUCTION

High-throughput sequencing of whole genomes has forever altered all of biology, and plant biology is no exception. Plant molecular and cell biology, plant genomics, plant ecology, crop genetics, plant development, botany, and related fields have fundamentally changed as a result of the ability to acquire whole genome sequences. The field of plant population genomics—the sequencing and analysis of populations of individual genomes—has emerged as an exciting extension of population genetics. The emergence of this field offers the ability not only to obtain and analyze comprehensive DNA sequence and functional information from large numbers of genotypes but also to extend analyses formerly restricted to a few model and crop organisms to a broad array of taxonomically diverse plant species.

Why are studies of populations of genomes essential to our understanding of plant biology? The reasons are many. For the plant breeder, populations of genomes provide a means for genomic selection and the ability to create detailed genomic maps of breeding populations, which provide a means to identify the locations of key genes. For the plant molecular biologist, populations of genomes have facilitated entirely new fields of exploration such as population epigenomics, which characterizes the molecular mechanisms of epiallele formation and how natural selection may operate on inherited epigenetic variation to shape patterns of phenotypic diversity. For the plant pathologist, sequences from entire populations of cells offer the prospect of advancing our understanding of senescence and the molecular basis of plant disease.

In the field of plant population genomics, population-level analyses of genomes have illuminated in detail the genetic differences among individuals and facilitated a description of the patterns and mechanisms of evolution in various plant species. From the sequences of many closely related genotypes and their evolutionary relatives, we gain insight into many evolutionary phenomena, including the frequency and intensity of natural selection, the relative importance of adaptive and neutral processes in plant evolution, the genomic targets of selection, the historical movements of individuals and dynamics of plant populations, and the elusive connections between genes and the traits they control. Many of these subjects have been core areas of research in plant evolutionary and population genetics for decades, but questions once beyond reach are tractable in ways that they never have been before. In particular, comparative studies of populations of genomes between closely related individuals and species with known phylogeny are responsible for some of the most significant advances in our understanding of the factors at work in plant evolution (Flowers & Purugganan, 2008).

The field of plant population genomics benefits from integration of a wide array of experimental approaches. For example, field studies of plant fitness in their native environments are essential to bridging the divide between molecular population genetics and evolution in the wild (reviewed in Anderson, Willis, & Mitchell-Olds, 2011; Mitchell-Olds & Schmitt, 2006). Functional studies of plant populations including comparative transcriptomics, proteomics, and epigenomics (Schmitz et al., 2013) together with genetic mapping of ecologically significant traits provide a means to characterize the molecular mechanisms responsible for plant adaptation (reviewed in Bergelson & Roux, 2010). Integrating these approaches is essential to a comprehensive understanding of plant evolution.

In this chapter, we describe insights into the mechanisms of plant evolution gained from the study of genetic diversity and the operation of selection in natural populations of plants, including domesticated species. We begin by summarizing what is known about the factors that contribute to differences in nucleotide variation observed across plant species, describe how estimates of nucleotide diversity and the mutation rate provide an essential framework to understanding the mechanisms of plant evolution, and highlight questions of historical interest to plant population geneticists. We then identify ways in which whole genome sequences from populations of plant genomes are providing new insight into the molecular basis of adaptation and, in some cases, maladaptation in plant species. Throughout the chapter, we highlight recent advances that have shed light on mechanisms of evolution in plant genomes with emphasis on ways in which the mating system and effective size of populations influence their evolution.

2. GENETIC DIVERSITY IN PLANT POPULATIONS

A primary focus of population genetic studies of plant species has been to identify the neutral and selective mechanisms that determine the amount of genetic variation within and among populations. Central to this problem is the concept of the effective population size (N_e)—the size of an idealized population that experiences the same rate of genetic drift as the study population.

Why are population geneticists interested in N_e? First, under the neutral theory of molecular evolution, nucleotide diversity reflects a balance between the rate of spontaneous mutation and the loss of variation due to genetic drift. In a population at equilibrium between these processes, nucleotide diversity is proportional to the product of the effective population size and the per generation rate of mutation ($\theta = 4\mathrm{Ne}\mu$). Thus, the effective population size determines the amount of genetic diversity in a population (reviewed in Charlesworth, 2009) and forms the basis for comparative studies that aim to characterize the effects of life history variation and other factors on the evolution of plant species.

Second, the efficacy of selection is a function of the effective population size and the distribution of fitness effects of mutations. Mutations behave neutrally and can escape selection if their effect on fitness is on the order of 1/Ne or less, while those with stronger effects are subject to selection. Thus, the effective population size provides insight into the relative importance of genetic drift and natural selection and the efficiency of selection at either removing harmful variants from a population or increasing the frequency of beneficial ones.

What factors determine the level of genetic diversity in a population? As can be anticipated from the equation in the preceding text, under neutral theory, the amount of genetic variation is determined by the effective population size and the mutation rate. Yet, there are many factors that contribute to the effective population size that must be distinguished. Moreover, neutral mechanisms are unlikely to be the only factors contributing to genomic patterns of variation despite the conspicuous absence of clear evidence of adaptation in most plant protein-coding genes (see succeeding text). Nevertheless, interest in this long-standing question has recently been rekindled in a sweeping survey of plant and animal species (Leffler et al., 2012), which reported a narrower than expected range of nucleotide diversity across species given the range of census population sizes. This observation may be explained by a variety of factors including a correlation between the effective size and the mutation rate (Sung, Ackerman, Miller, Doak, & Lynch, 2012), pervasive selection (see the following text), or other factors, which we describe in some detail in the succeeding text.

3. MUTATION RATE VARIATION

In plants, the mutation rate has been inferred by calibrating the rate of mutation using known dates of divergence from the fossil record (i.e. the ‘molecular clock’, Gaut, Morton, McCaig, & Clegg, 1996). A significant recent advance, however, is the use of whole genome sequencing of mutation accumulation lines as a means to assess the per generation rate of mutation in a land plant (*Arabidopsis thaliana*, 7.1×10^{-9} mutations per base pair per generation; Ossowski et al., 2010) and a chlorophyte (*Chlamydomonas reinhardtii*, 2.08×10^{-10} mutations per base pair per generation, Ness, Morgan, Colegrave, & Keightley, 2012; 6.76×10^{-11}, Sung et al., 2012). These are at present the only direct estimates of mutation rate and the spectrum of spontaneous mutations in plants and their close relatives. As a consequence, direct estimates of a key parameter necessary for estimating the effective population size (from the relationship $\theta = 4\mathrm{Ne}\mu$) are unavailable for the vast majority of plant species. While estimates based on the molecular clock suggest that mutation rate variation may not be a primary factor contributing to differences in genetic diversity among species (Muse, 2000; Muse & Gaut, 1997), a rigorous test of this hypothesis awaits a more comprehensive understanding of mutation rate variation among species and the factors contributing to it. An intriguing observation is that across the tree of life, the mutation rate and the effective population size are negatively correlated, which could indicate that the mutation rate itself is subject to selection (Sung et al., 2012).

4. MATING-SYSTEM EFFECTS ON GENETIC DIVERSITY

A number of factors influence the rate of genetic drift and the long-term effective population size of plant species. These include variance in reproductive success, population expansion and contraction, distortion in the sex ratio, and the mating system. The effect of mating system has been of special interest in plant population genomics owing to a rich theory that predicts that components of life history, such as the outcrossing rate and seed dispersal mechanism, should have pronounced effects on the rate of genetic drift and the efficiency of selection (Glémin, 2007). For example, self-fertilization reduces the effective population size by up to twofold relative to an obligate outcrossing mode of reproduction owing to the effect of inbreeding (Charlesworth, Lyons, & Litchfield, 1993; Nordborg, 2000; Pollak, 1987). Selfing populations are also characterized by lower effective rates of recombination, which should increase losses of diversity due to linked selection (see succeeding text), and also frequently experience cycles of local extinction and recolonization, recurrent population bottlenecks, and reduced rates of migration—factors that further enhance the rate of genetic drift within, and the degree of isolation between, populations.

Early studies of mating systems using isozymes found that life history characteristics have clear and predictable effects on the heterozygosity within and among plant populations (reviewed in Hamrick & Godt, 1996), and nucleotide-based surveys have broadly supported a strong effect of mating system on genetic diversity in plants (Glémin, Bazin, & Charlesworth, 2006; Leffler et al., 2012). Furthermore, studies of DNA sequence variation have found strong population structure in selfing species compared to their outcrossing relatives (*Leavenworthia*, Liu, Charlesworth, & Kreitman, 1999; Liu, Zhang, & Charlesworth, 1998; *Arabidopsis*, Savolainen, Langley, Lazzaro, & Fr, 2000; Wright, Lauga, & Charlesworth, 2002; *Solanum*, Baudry, Kerdelhué, Innan, & Stephan, 2001; *Mimulus*, Sweigart & Willis, 2003; *Amsinckia*, Pérusse & Schoen, 2004; *Capsella*, Foxe et al., 2009; Guo et al., 2009; *Collinsia*, Hazzouri et al., 2013). In most cases, the associated effects of reduced census population size and the effect of self-fertilization jointly contribute to accelerated rates of genetic drift. Distinguishing among these effects is difficult, but in at least one well-documented example, low levels of diversity in populations of self-fertilizing *Eichornia paniculata* may be caused by founder events rather than the transition from an outcrossing to a selfing mode of reproduction (Ness, Wright, & Barrett, 2010). As we describe in the next section, the effect of mating system and its correlated effect on population structure may significantly impact the mechanisms of evolution in different plant groups.

5. ADAPTATION IN PLANT POPULATIONS

A major area of interest in evolutionary genomics is to identify whether genomes are shaped primarily by the action of genetic drift or adaptation. Recent advances have been made in quantifying the incidence of adaptive evolution by jointly estimating the distribution of fitness effects of mutations and the proportion of mutations fixed by positive selection (Fay, Wyckoff, & Wu, 2002; Smith & Eyre-Walker, 2002). Although estimates vary somewhat among studies, the emerging view is that a greater proportion of fixed mutations, or substitutions, are adaptive in Drosophila, mice, and enteric bacteria than in primates and plants (Eyre-Walker, 2006). In *A. thaliana*, there is a large class of rare and potentially harmful mutations (Bustamante et al., 2002; see the succeeding text) but limited evidence of fixation of adaptive variants, with the exception of certain classes of genes including disease resistance (Bergelson, Kreitman, Stahl, & Tian, 2001) and abiotic stress genes (Slotte et al., 2011). An absence of a correlation between protein divergence (i.e. ratio of nonsynonymous to synonymous substitutions between species, d*N*/d*S*) and nucleotide polymorphism further suggests that *A. thaliana* experiences limited adaptation (i.e. no evidence of recurrent hitchhiking) in protein-coding genes (Slotte et al., 2011). Moreover, a survey of 10 plant species found that only sunflowers showed evidence of adaptive evolution of protein-coding genes (Gossmann et al., 2010). This suggests that the conspicuous absence of adaptive substitution in Arabidopsis may be generally true for many plant species.

Why is there limited evidence of adaptation in plant genes? First, the proportion of mutations fixed by positive selection is dependent on both the proportion of adaptive and effectively neutral substitutions. The large fraction of neutral fixations in species with small effective sizes makes detection of relatively rare adaptive substitutions more difficult (Gossmann, Keightley, & Eyre-Walker, 2012). Second, populations with large effective sizes are expected to efficiently fix adaptive alleles, whereas genetic drift in small populations can lead to loss of beneficial alleles with weak or moderate effects on fitness (reviewed in Olson-Manning, Wagner, & Mitchell-Olds, 2012). This drift effect and the fact that smaller population size translates into a longer waiting time for beneficial mutations (i.e. they are mutation-limited to a greater degree than populations with large effective sizes) likely contribute to the limited signature of molecular adaptation in plant protein-coding genes (Gossmann et al., 2010; Wright & Andolfatto, 2008).

The hypothesis that small population sizes limit the rate of adaptive substitution in plants is supported by studies of species with large effective population sizes. For example, *Capsella grandiflora*, an outcrossing species with little population structure and a large effective population size (~500,000), shows a high rate of adaptive fixation, with 40% of amino acid substitutions inferred to have been fixed by positive selection (Slotte, Foxe, Hazzouri, & Wright, 2010). Sunflowers similarly have a large effective population size and a large fraction of amino acids fixed by positive selection (Gossmann et al., 2010; Strasburg, Scotti-Saintagne, Scotti, Lai, & Rieseberg, 2009), and comparison of rates of protein evolution shows a positive correlation between the rate of protein evolution (i.e. d*N*/d*S*) and the effective population size across species (Strasburg et al., 2011). These observations suggest that effective population size contributes significantly to molecular adaptation, or lack thereof, in many plant species. A wrinkle in this explanation is that some species including maize have large effective population sizes but show little evidence of adaptation in protein-coding genes (Gossmann et al., 2010). Thus, other factors such as population structure among isolated populations or reduced effective rates of recombination likely constrain the probability of fixation of beneficial mutations (Wright & Andolfatto, 2008). This area of research will benefit greatly from additional surveys of variation from related species with different effective population sizes and mating systems and expansion of comparative analyses to noncoding DNA.

Limited evidence for adaptive substitution using the approaches adopted in the studies in the preceding text does not preclude the significance of molecular adaptation in plants, including those species with small effective sizes. In fact, it is in situations where molecular evidence of adaptive evolution is rare that application of genome-wide approaches may be particularly effective. For example, widespread plant species, that is, those with geographic distributions that span a wide range of climatic conditions, are excellent models for exploring the effect of climate on local adaptation. In a genome-wide survey of SNPs genotyped in 948 accessions of *A. thaliana*, Hancock et al. (2011) found that SNPs whose population frequencies showed the strongest correlations with climatic variables were significantly enriched for nonsynonymous SNPs relative to their overall proportions in the genome, while noncoding SNPs were not. This suggests that amino acid-altering variants improve plant fitness under local environmental conditions and provide strong support for climatic adaptation. In another example, Lee and Mitchell-Olds (2012) reported that after accounting for the effects of genomic factors, a relatively large percentage of the variation in nonsynonymous polymorphism is explained by the environment. Their analysis supports a model where spatially varying selection maintains polymorphism at environmentally responsive genes through the action of antagonistic pleiotropy. While examples such as this are at present limited to a few model plant species, the prospect of large samples of genotypes from different environments should provide a fruitful means to detect genes responding to selection.

A question of great interest in molecular population genetics is the importance of balancing selection in molecular evolution (Charlesworth, 2006; Wright & Andolfatto, 2008). This form of selection, which may include, but is not limited to, heterozygote advantage (i.e. overdominance), can lead to maintenance of variation for periods far longer than expected for neutral alleles. Well-known examples in plants include maintenance of conditionally adaptive alleles at self-incompatibility loci (Castric, Bechsgaard, Schierup, & Vekemans, 2008; Castric & Vekemans, 2007; Roux et al., 2013; Schierup, Mable, Awadalla, & Charlesworth, 2001; Wright, 1939) and disease resistance alleles over periods spanning millions of years, which in extreme cases can lead to maintenance of highly diverged alleles and sharing of polymorphisms among species (Gos, Slotte, & Wright, 2012; Liu et al., 2010; Meyer, Nelson, Clement, & Ramakrishnan, 2010; Rose, Michelmore, & Langley, 2007; Stahl, Dwyer, Mauricio, Kreitman, & Bergelson, 1999; Tian, Traw, Chen, Kreitman, & Bergelson, 2003). The population genetic signature of such 'balanced' polymorphisms—increases in coalescence times at sites tightly linked to the selected locus—has been detected in genomic scans of polymorphism in plants most prominently in regions enriched for disease resistance genes (Clark et al., 2007). Such whole genome resequencing studies facilitate a description of outlier regions with high levels of polymorphism and provide a means for identifying genes subject to long-term balancing selection.

While examples of long-term balancing selection are interesting cases, details concerning the mechanism (e.g. overdominance, frequency-dependent selection, or spatially varying selection) are frequently obscure. Moreover, there are likely to be many cases where various forms of diversity-maintaining selection do not operate for long enough to leave a diagnostic signature of balancing selection. In order to elucidate both the target(s) and mechanisms of selection in such cases, it is essential to use a combination of experimental approaches including genomic scans for selection, association studies, functional analyses, and field experiments. For example, using 39 years of aphid abundance data, Züst et al. (2012) correlated allele frequencies at *GS-ELONG*, a gene controlling the defence response in *A. thaliana,* with the abundance of two specialist aphid species. In addition to reporting a strong correlation between herbivore intensity and allele frequency, they carried out a multigenerational selection experiment and demonstrated a causal link between aphid abundance and variation at the defence locus. In another example, Fournier-Level et al. (2011) conducted common garden experiments at four field sites across Europe with *A. thaliana* ecotypes with different geographic origins. Using genome-wide association mapping (GWAS; see the succeeding text), they mapped components of lifetime fitness in the field and found that high-fitness alleles identified in their study were distributed closer to the site where they increase fitness compared with genomic controls. These and related studies offer clear insight into the mechanisms of plant evolution and help to close the gap between inferences drawn purely from population genetic patterns and the operation of selection in the wild.

6. DELETERIOUS VARIATION IN PLANT POPULATIONS

Small population sizes are relatively permissive of deleterious variation owing to increased rates of genetic drift and the effective neutrality of weakly selected alleles (Charlesworth, 2009; Gossmann et al., 2012; Pollak, 1987; Wright, Ness, Foxe, & Barrett, 2008). Therefore, species of plants with small effective population sizes, including many domesticated crops, are subject to the accumulation of deleterious variants whose selective effect is on the order of the reciprocal of the effective population size (i.e. the so-called effectively neutral polymorphisms). In *A. thaliana*, there is a large class of potentially deleterious amino acid polymorphisms segregating in natural populations (Wright & Andolfatto, 2008; Bustamante et al., 2002). A similar pattern is observed in the outcrossing relative *Arabidopsis lyrata*, suggesting that the excess of replacement polymorphisms in *A. thaliana* is not the result of reduced efficiency of selection in a selfing species (Foxe et al., 2008), although patterns of selection on codon usage are consistent with an effect of mating system (Qiu, Zeng, Slotte, Wright, & Charlesworth, 2011). These studies point to a large class of potentially harmful mutations segregating both in *A. thaliana* and *A. lyrata*, but the effect of mating system is only apparent for very weakly selected sites. By contrast, in a comparison between the primarily outcrossing *C. grandiflora* and its selfing relative *Capsella rubella*, there is a strong signature of relaxed purifying selection in the self-fertilizing species (Hazzouri et al., 2013; Slotte et al., 2013).

In other cases, population size contractions may be responsible for the release of formerly deleterious mutations into the effectively neutral class. This has received some empirical support in crop plants, where *tropical* and *temperate japonica* rice segregate proportionally more effectively neutral nonsynonymous mutations than their wild relative *Oryza rufipogon* (Gossmann et al., 2010), and there is evidence of genome-wide relaxation of purifying selection in *japonica* (Yang et al., 2012). The ability of effectively neutral mutations to escape the effect of purifying selection is expected to increase the fixation probability of deleterious mutations and may account for the higher rate of amino acid substitution in rice cultivars with smaller effective size (Lu et al., 2006; Yang et al., 2012). As illustrated by these examples, reductions in population size associated with life history transitions to self-fertilization or population bottlenecks may allow harmful alleles to accumulate in plant populations. In some cases, this may contribute to declines in mean population fitness (Kondrashov, 1984; Phillips, Otto, & Whitlock, 2000; Schultz & Lynch, 1997).

Differences in the efficiency of selection against harmful mutations may also have profound consequences for many aspects of plant genome evolution. For example, it has been postulated that variation in genome size across the web of life is the result of differences in the effective size among species

(Lynch & Conery, 2003). This hypothesis proposes that mutations that increase genome size (e.g. TE proliferation) have a weak fitness cost and natural selection will work more efficiently to constrain genome expansion in species with larger effective size, whereas mutations of the same effect in species with small effective size can drift to fixation. This hypothesis has received mixed support both overall and within plants. For example, genomes of self-fertilizing plant species are smaller than those of their outcrossing relatives (Wright et al., 2008) and species with restricted geographic ranges can have larger genome sizes compared with their more common relatives (Vinogradov, 2003), although this may not be a universal effect (Grotkopp, Rejmánek, Sanderson, & Rost, 2004). However, in a comparison across populations of *A. lyrata*, there is evidence that TEs are subject to purifying selection in large populations, whereas they evolve neutrally in smaller populations (Lockton, Ross-Ibarra, & Gaut, 2008). A study of 10 species in the genus *Oryza* found no relationship between the effective population size and genome size using lineage-specific mutation rates to estimate the effective size and phylogenetic independent contrasts to control for phylogenetic nonindependence (Ai, Wang, & Ge, 2012). Examination of 205 species of seed plants similarly found no support for the Lynch and Conery (2003) model when phylogenetic relationships were factored into the analysis (Whitney et al., 2010). Studies such as these, which consider genome size in the context of phylogeny, control for phylogenetic nonindependence, and incorporate lineage-specific mutation rates to estimate the effective population size, provide the best framework for addressing these and related questions concerning the accumulation of deleterious variation and the effect of population size on genome evolution (Charlesworth & Barton, 2004; Flowers & Purugganan, 2008).

7. THE EFFECTS OF LINKED SELECTION

In previous sections, we highlighted work on the incidence of positive and negative selection in plant genomes. A related approach is to study the effects of natural selection on patterns of variation at neutral sites across the genome. Analysis of linked neutral sites can greatly enrich our understanding of selection and genes targeted by selection can be identified. Moreover, studies of neutral variation can provide a possible answer to the riddle of why genetic diversity varies so little among species that differ so dramatically in census population size (Leffler et al., 2012).

The “hitchhiking” model is perhaps the best-known model of linked selection that describes the rise in frequency of a new beneficial mutation due to positive selection and the concurrent loss of genetic diversity at linked neutral sites (Maynard Smith and Haigh 1974). This model predicts that the loss of variation in regions subject to positive selection is proportional to the recombination rate. Under this model, fixation of beneficial mutations in regions of suppressed recombination (e.g. heterochromatic regions in many plants, Gaut, Wright, Rizzon, Dvorak, & Anderson, 2007) will lead to losses of diversity over greater physical distances than fixation of an equivalent allele in regions with normal rates of crossing over. Charlesworth et al. (1993) discovered that “background selection”—loss of genetic diversity at linked sites due to purifying selection—will similarly operate to eliminate variation to a greater extent in regions with reduced rates of crossing over. In principle, both genetic hitchhiking and background selection should yield a positive correlation between estimates of nucleotide diversity and the recombination rate provided that the rate of deleterious or beneficial mutations is roughly the same in different genomic regions.

This expectation is met in many organisms including Drosophila and Caenorhabditis, but studies of plant populations suggest that this pattern is absent in Arabidopsis, *O. rufipogon*, and *Oryza sativa* (Flowers et al., 2012; Nordborg et al., 2005) and is rare in plant species studied to date (reviewed in Cutter & Payseur, 2013). Why does polymorphism in plant species frequently not conform to expectations of simple models of linked selection? In *O. rufipogon*, significant negative associations are observed between polymorphism and gene density, but a correlation with recombination rate is not apparent. The probable cause of this pattern is the joint effect of an uneven distribution of deleterious mutations across the genome (i.e. enrichment of deleterious mutations in gene dense regions) and a reduced outcrossing rate (Flowers et al., 2012), although other factors including dominance of deleterious mutations and the distribution of fitness effects may contribute to the pattern in rice and other species (Cutter & Payseur, 2013). It remains to be seen what factors underlie the different patterns observed in different plant species, but it is clear that the physical organization of genes on recombining chromosomes, mating system, population structure, and mutation rate effects can generate patterns that differ significantly from the classic Drosophila-like association between polymorphism and the rate of crossing over (Begun & Aquadro, 1992).

Genome scans of nucleotide polymorphism also provide a means to identify regions targeted by selection. In plants, genome scans have been applied most frequently in crop species with the primary goals of identifying the number and function of genes subject to selection (reviewed in Alonso-Blanco et al., 2009; Olsen & Wendel, 2013) and quantifying the strength and frequency of selection during the course of domestication (reviewed in Olsen & Wendel, 2013). A recent study of maize domestication inferred that 7.6% of the genome showed evidence of linked selection during the "domestication" phase with large regions showing evidence of selective sweeps and relatively large estimates of the strength of selection ($s = 0.015$). Regions subject to selection during the "improvement phase" were subject to less intense selection ($s = 0.003$) and were smaller in size and contained fewer genes. While up to 1–2% of the genes in the genome may have been the direct targets of selection for domestication or improvement traits, another 3000 genes not directly targeted are genetically impoverished of variation due to linkage with genes subject to selection (Hufford et al., 2012) above and beyond the losses of variation associated with population bottlenecks. Similar whole genome resequencing studies of domesticated *indica* and *japonica* rice reported more than 700 10 kb regions containing on the order of a thousand genes that are severely depleted of variation in both groups of cultivars (Xu et al., 2011) and a genotyping study of sorghum landraces identified selective sweeps of up to several megabases (Morris et al., 2013). These studies suggest a mating-system effect on the breadth of selective sweeps with self-pollinating species of rice and sorghum (lower effective rate of recombination) having larger sweep regions on average than outcrossing maize (higher recombination rate; Morris et al., 2013).

Genomic regions depauperate of variation are frequently caused by "hard sweeps" that lead to losses of variation due to linkage with new advantageous mutations (Clark, Linton, Messing, & Doebley, 2004; Olsen et al., 2006; Wang et al., 2008). An interesting area of future research will be to examine whether the numbers of genes subject to selection have been underestimated because of current methodological limitations in detecting "soft sweeps"—incomplete or partial loss of variation at linked sites owing to selection on standing variation (Pritchard, Pickrell, & Coop, 2010). Nevertheless, studies such as these not only provide insight into the operation of selection under domestication but also suggest regions of the genome where trait mapping approaches will provide novel and powerful insights into the evolutionary history of crop plants (Chia et al., 2012; Hufford et al., 2012; Jiao et al., 2012; Mandel et al., 2013).

8. MAPPING OF QUANTITATIVE TRAIT LOCI

Traditionally, QTL mapping in plants was the method of choice to identify loci responsible for the phenotypic changes associated with crop evolution (Burke, Tang, Knapp, & Rieseberg, 2002; Doebley, Stec, Wendel, & Edwards, 1990; Xiong, Liu, Dai, Xu, & Zhang, 1999) and the evolution of mating systems (Fishman, Kelly, & Willis, 2002; Goodwillie, Ritland, & Ritland, 2006; Grillo et al., 2009; Lin & Ritland, 1997; Slotte, Hazzouri, Stern, Andolfatto, & Wright, 2012). The traditional method or single-family QTL approach is limited by small mapping population sizes and issues in model construction (Beavis, 1998). Family-based QTL mapping approaches use existing breeding families and improve both precision and accuracy in QTL detection, compared with the single-family approach (Rosyara, Gonzalez-Hernandez, Glover, Gedye, & Stein, 2009). GWAS, on the other hand, uses correlations between molecular polymorphism and phenotypic variation in assemblages of individuals drawn from natural populations (Yu & Buckler, 2006). In contrast to traditional mapping approaches, GWAS uses linkage disequilibrium between markers and traits of interests and therefore improves mapping resolution by capturing many generations of historical recombination. Moreover, GWAS removes limitations frequently encountered in breeding of species with delayed maturity and long generation times and is therefore a valuable approach to mapping in perennial species of plants. In recent years, genome-wide association approaches have been shown to be an especially powerful tool for identifying the genetic contributors to complex traits.

In plant genomics, numerous other tools have been developed to map quantitative traits. For example, nested association mapping (NAM) utilizes a combination of linkage and association analyses of a large multifamily recombinant mapping population that takes advantage of both historical and recent recombination events to improve resolution and statistical power (Yu, Holland, McMullen, & Buckler, 2008). This strategy has been deployed in maize, where efforts over the last decade have led to the development of robust NAM approaches (Buckler et al., 2009; Tian et al., 2011). Another area of development has been to use multiple generations of intercrossing between several individuals, which can both increase the number of recombination events and broaden the base of genetic diversity for mapping traits (Liu, Kowalski, Lan, Feldmann, & Paterson, 1996). These have led to the development of multiple advanced generation intercrossed lines in *A. thaliana* (Kover et al., 2009) and wheat (Huang, George, et al., 2012; Huang, Kurata, et al., 2012). These and other approaches (such as bulk segregant analysis mapping/extreme QTL mapping) (Ehrenreich et al., 2009) promise to increase the ease and resolution of QTL mapping.

9. ADVANCES THROUGH WHOLE GENOME RESEQUENCING

The previous sections have described areas of research that were already well developed from both theoretical and empirical perspectives prior to the emergence of population genomics as a field. However, the ability to analyze entire populations of genomes not only has facilitated refined analysis of the evolutionary dynamics at work in plant genome evolution but also has ushered in an era of discovery in areas of investigation that were virtually impossible prior to the new era of whole genome sequencing (Table 12.1). For example, population genomic analyses facilitated the resolution of a long-standing riddle in the study of sex determination in date palm (*Phoenix dactylifera*). Al-Dous et al. (2011) used whole genome sequencing of male and female date palm cultivars to identify a small region of the genome (~600 kb) with SNP genotypes consistent with males being the heterogametic sex. Additional research has found this to be a nonrecombining XY-like region (Cherif et al., 2013). Now, breeders who formerly had to wait up to 6 years to determine the sex of seedlings now have the ability to do so with a simple PCR assay.

In other examples, whole genome resequencing efforts have illuminated important details about the domestication process through comparative analysis of domesticated and progenitor populations. A striking example is the study of Guo et al. (2013), which resequenced domesticated watermelons and their wild progenitor and discovered that disease resistance genes had been lost during domestication. They reported that members of the NBS-LRR class of resistance genes showing presence/absence variation among genomes in the progenitor population were completely absent in domesticated cultivars. This could indicate that resistance alleles have a cost in domesticated species, potentially through their effects on fruit quality or yield.

Finally, whole genome sequencing of entire populations of genomes has been extended beyond conventional analysis of DNA sequence variation. For example, Schmitz et al. (2013) resequenced the epigenomes (i.e. genome-wide characterization of methylated cytosines) of 152 *A. thaliana*

Table 12.1 Whole genome sequencing of populations of plant genomes (excluding studies using genotyping arrays)

Species	Number of genomes	Interesting finding
Arabidopsis thaliana (thale cress)	80	Accumulation of deleterious mutations at higher rates in marginal populations (Cao et al., 2011)
Citrullus lanatus (watermelon)	20	Loss of disease resistance genes during domestication (Guo et al., 2013)
Phoenix dactylifera (date palm)	9	Determination of males as the heterogametic sex (Al-Dous et al., 2011)
Oryza sativa	1083	Single origin of domestication in southern China (Huang, Kurata, et al., 2012)
Oryza rufipogon	446	Limited number of individuals used during domestication (loss of genetic diversity) and mapping of two traits leaf sheath colour and tiller angle (Huang, Kurata, et al., 2012)
Oryza sativa	40	Identification of candidate regions selected during domestication (Xu et al., 2011)
Oryza rufipogon	5	A complex evolutionary history between *O. indica* and *O. nivara* (Xu et al., 2011)
Oryza nivara	5	Independent domestication of *O. japonica* from the Chinese *O. rufipogon* population (Xu et al., 2011)
Zea mays (domesticated maize)	58	Recovery of diversity after domestication and important number of genes under strong selection for important phenotypes (Hufford et al., 2012)
Zea parviglumis (teosinte)	14	Introgression into the domesticated maize (Hufford et al., 2012)
Prunus persica (peach)	11	Evidence for selection for different maturation times among varieties (International Peach Genome Initiative, 2013)
Glycine max (soybean)	31	Mapping soybean agronomic traits as well as cloning devastating resistance gene in Asian soybean rust (Schmutz et al., 2010)
Cicer arietinum (chickpea)	90	Candidate genes for disease resistance genes and agronomic traits highlighted (Varshney et al., 2013)
Capsella rubella	2	Genomewide relaxation of purifying selection in *C. rubella* and a shift in expression of flowering genes in the transition to selfing (Slotte et al., 2013)

accessions and generated a genomic map of SMPs (single methylation polymorphisms). From this map of SMPs, the authors characterized the extent of natural epigenomic variation in a population, quantified the rate of decay of linkage disequilibrium among SMPs, and mapped methylation quantitative loci (i.e. genes controlling variation in differentially methylated regions). These and many other examples of whole genome resequencing (Table 12.1) illustrate the diversity of applications and questions to which population genomic approaches are now being applied.

10. CONCLUSION

The ability to explore the diversity of genomes in species and their evolutionary relatives has led to a greater understanding of the evolutionary forces that drive genome variation and genome structure. Population genomics provides a synthesis of genomics, molecular evolution, population genetics, and even molecular biology and ecology. As genomics technology continues to improve, we can look forward to providing in-depth analysis in unprecedented detail.

As the field of population genomics advances, there continues to be a need to explore several areas. First, we need to apply population genomic analyses to a wider array of taxonomically diverse species. Currently, genome-wide information is available primarily for members of the Brassicaceae and cereal crop species. By increasing the diversity of taxa, comparative analysis that consider phylogenetic context will become tractable that will greatly improve our understanding of the mechanisms of evolution in plant species. Second, integrating functional (e.g. transcriptomic, proteomic, and epigenomic) and structural (i.e. sequence) information in population genomic studies will provide opportunities to study functional polymorphism at a genome-wide scale and facilitate characterization of the mechanisms of evolution in entire networks of genes. These are exciting possibilities that we expect will attract much attention in the coming years.

REFERENCES

Ai, B., Wang, Z. S., & Ge, S. (2012). Genome size is not correlated with effective population size in the Oryza species. *Evolution*, *66*(10), 3302–3310.

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., et al. (2011). De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). *Nature Biotechnology*, *29*(6), 521–527.

Alonso-Blanco, C., Aarts, M. G., Bentsink, L., Keurentjes, J. J., Reymond, M., Vreugdenhil, D., et al. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, *21*(7), 1877–1896.

Anderson, J. T., Willis, J. H., & Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends in Genetics*, *27*(7), 258–266.

Baudry, E., Kerdelhué, C., Innan, H., & Stephan, W. (2001). Species and recombination effects on DNA variability in the tomato genus. *Genetics*, *158*(4), 1725–1735.

Beavis, W. D. (1998). QTL analyses: Power, precision, and accuracy. In A. H. Paterson (Ed.), *Molecular analysis of complex traits* (pp. 145–161). Boca Raton: CRC Press.

Begun, D. J., & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature*, *356*(6369), 519–520.

Bergelson, J., Kreitman, M., Stahl, E. A., & Tian, D. (2001). Evolutionary dynamics of plant R-genes. *Science*, *292*(5525), 2281–2285.

Bergelson, J., & Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics*, *11*(12), 867–879.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The genetic architecture of maize flowering time. *Science*, *325*(5941), 714–718.

Burke, J. M., Tang, S., Knapp, S. J., & Rieseberg, L. H. (2002). Genetic analysis of sunflower domestication. *Genetics*, *161*(3), 1257–1267.

Bustamante, C. D., Nielsen, R., Sawyer, S. A., Olsen, K. M., Purugganan, M. D., & Hartl, D. L. (2002). The cost of inbreeding in Arabidopsis. *Nature*, *416*(6880), 531–534.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., et al. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nature Genetics*, *43*(10), 956–963.

Castric, V., Bechsgaard, J., Schierup, M. H., & Vekemans, X. (2008). Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genetics*, *4*(8), e1000168.

Castric, V., & Vekemans, X. (2007). Evolution under strong balancing selection: How many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae? *BMC Evolutionary Biology*, 7, 132.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*(4), e64.

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195–205.

Charlesworth, B., & Barton, N. (2004). Genome size: Does bigger mean worse? *Current Biology*, *14*(6), R233–R235.

Charlesworth, D., Lyons, E. E., & Litchfield, L. B. (1993). Inbreeding depression in two highly inbreeding populations of Leavenworthia. *Proceedings of the Royal Society of London B: Biological Sciences*, *258*, 209–214.

Cherif, E., Zehdi, S., Castillo, K., Chabrillange, N., Abdoulkader, S., Pintaud, J. C., et al. (2013). Male-specific DNA markers provide genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. *New Phytologist*, *197*(2), 409–415.

Chia, J. M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics*, *44*(7), 803–807.

Clark, R. M., Linton, E., Messing, J., & Doebley, J. F. (2004). Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(3), 700–707.

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., et al. (2007). Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Science*, *317*(5836), 338–342.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews Genetics*, *14*(4), 262–274.

Doebley, J., Stec, A., Wendel, J., & Edwards, M. (1990). Genetic and morphological analysis of a maize-teosinte F2 population: Implications for the origin of maize. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9888–9892.

Ehrenreich, I. M., Hanzawa, Y., Chou, L., Roe, J. L., Kover, P. X., & Purugganan, M. D. (2009). Candidate gene association mapping of Arabidopsis flowering time. *Genetics*, *183*(1), 325–335.

Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*, *21*(10), 569–575.

Fay, J. C., Wyckoff, G. J., & Wu, C. I. (2002). Testing the neutral theory of molecular evolution with genomic data from Drosophila. *Nature*, *415*(6875), 1024–1026.

Fishman, L., Kelly, A. J., & Willis, J. H. (2002). Minor quantitative trait loci underlie floral traits associated with mating system divergence in Mimulus. *Evolution*, *56*(11), 2138–2155.

Flowers, J. M., Molina, J., Rubinstein, S., Huang, P., Schaal, B. A., & Purugganan, M. D. (2012). Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Molecular Biology and Evolution*, *29*(2), 675–687.

Flowers, J. M., & Purugganan, M. D. (2008). The evolution of plant genomes: Scaling up from a population perspective. *Current Opinion in Genetics and Development*, *18*(6), 565–570.

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A map of local adaptation in Arabidopsis thaliana. *Science*, *334*(6052), 86–89.

Foxe, J. P., Dar, V. U., Zheng, H., Nordborg, M., Gaut, B. S., & Wright, S. I. (2008). Selection on amino acid substitutions in Arabidopsis. *Molecular Biology and Evolution*, *25*(7), 1375–1383.

Foxe, J. P., Slotte, T., Stahl, E. A., Neuffer, B., Hurka, H., & Wright, S. I. (2009). Recent speciation associated with the evolution of selfing in Capsella. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5241–5245.

Gaut, B. S., Morton, B. R., McCaig, B. C., & Clegg, M. T. (1996). Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(19), 10274–10279.

Gaut, B. S., Wright, S. I., Rizzon, C., Dvorak, J., & Anderson, L. K. (2007). Recombination: An underappreciated factor in the evolution of plant genomes. *Nature Reviews Genetics*, *8*(1), 77–84.

Glémin, S. (2007). Mating systems and the efficacy of selection at the molecular level. *Genetics*, *177*(2), 905–916.

Glémin, S., Bazin, E., & Charlesworth, D. (2006). Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Biological Sciences*, *273*(1604), 3011–3019.

Goodwillie, C., Ritland, C., & Ritland, K. (2006). The genetic basis of floral traits associated with mating system evolution in Leptosiphon (Polemoniaceae): An analysis of quantitative trait loci. *Evolution*, *60*(3), 491–504.

Gos, G., Slotte, T., & Wright, S. I. (2012). Signatures of balancing selection are maintained at disease resistance loci following mating system evolution and a population bottleneck in the genus Capsella. *BMC Evolutionary Biology*, *12*, 152.

Gossmann, T. I., Keightley, P. D., & Eyre-Walker, A. (2012). The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biology and Evolution*, *4*(5), 658–667.

Gossmann, T. I., Song, B. H., Windsor, A. J., Mitchell-Olds, T., Dixon, C. J., Kapralov, M. V., et al. (2010). Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution*, *27*(8), 1822–1832.

Grillo, M. A., Li, C., Fowlkes, A. M., Briggeman, T. M., Zhou, A., Schemske, D. W., et al. (2009). Genetic architecture for the adaptive origin of annual wild rice, Oryza nivara. *Evolution*, *63*(4), 870–883.

Grotkopp, E., Rejmánek, M., Sanderson, M. J., & Rost, T. L. (2004). Evolution of genome size in pines (Pinus) and its life-history correlates: Supertree analyses. *Evolution*, *58*(8), 1705–1729.

Guo, Y. L., Bechsgaard, J. S., Slotte, T., Neuffer, B., Lascoux, M., Weigel, D., et al. (2009). Recent speciation of Capsella rubella from Capsella grandiflora, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5246–5251.

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nature Genetics*, *45*(1), 51–58.

Hamrick, J. L., & Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions: Biological Sciences*, *351*, 1291–1298.

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., et al. (2011). Adaptation to climate across the Arabidopsis thaliana genome. *Science*, *334*(6052), 83–86.

Hazzouri, K. M., Escobar, J. S., Ness, R. W., Killian Newman, L., Randle, A. M., Kalisz, S., et al. (2013). 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 system shift. *Evolution*, *67*(5), 1263–1278.

Huang, B. E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., Morell, M. K., et al. (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnology Journal*, *10*(7), 826–839.

Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, *490*(7421), 497–501.

Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nature Genetics*, *44*(7), 808–811.

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., et al. (2012). Genome-wide genetic changes during modern breeding of maize. *Nature Genetics*, *44*(7), 812–815.

Kondrashov, A. S. (1984). Deleterious mutations as an evolutionary factor. 1. The advantage of recombination. *Genetics Research*, *44*(2), 199–217.

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., et al. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. *PLoS Genetics*, *5*(7), e1000551.

Lee, C. R., & Mitchell-Olds, T. (2012). Environmental adaptation contributes to gene polymorphism across the Arabidopsis thaliana genome. *Molecular Biology and Evolution*, *29*(12), 3721–3728.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., et al. (2012). Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388.

Lin, J. Z., & Ritland, K. (1997). Quantitative trait loci differentiating the outbreeding Mimulus guttatus from the inbreeding M. platycalyx. *Genetics*, *146*(3), 1115–1121.

Liu, F., Charlesworth, D., & Kreitman, M. (1999). The effect of mating system differences on nucleotide diversity at the phosphoglucose isomerase locus in the plant genus Leavenworthia. *Genetics*, *151*(1), 343–357.

Liu, S. C., Kowalski, S. P., Lan, T. H., Feldmann, K. A., & Paterson, A. H. (1996). Genome-wide high-resolution mapping by recurrent intermating using Arabidopsis thaliana as a model. *Genetics*, *142*(1), 247–258.

Liu, X. Q., Wang, L., Liu, X. D., Liu, X. Q., Wang, D. B., Wang, C. T., et al. (2010). The molecular evolution of the rice blast resistance gene Pi36. *International Journal of Plant Sciences*, *171*(3), 235–243.

Liu, F., Zhang, L., & Charlesworth, D. (1998). Genetic diversity in Leavenworthia populations with different inbreeding levels. *Proceedings of the Biological Sciences*, *265*(1393), 293–301.

Lockton, S., Ross-Ibarra, J., & Gaut, B. S. (2008). Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 13965–13970.

Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., & Wu, C. I. (2006). The accumulation of deleterious mutations in rice genomes: A hypothesis on the cost of domestication. *Trends in Genetics*, *22*(3), 126–131.

Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science*, *302*(5649), 1401–1404.

Mandel, J. R., Nambeesan, S., Bowers, J. E., Marek, L. F., Ebert, D., Rieseberg, L. H., et al. (2013). Association mapping and the genomic consequences of selection in sunflower. *PLoS Genetics*, *9*(3), e1003378.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*(1), 23–35.

Meyer, S. E., Nelson, D. L., Clement, S., & Ramakrishnan, A. (2010). Ecological genetics of the Bromus tectorum (Poaceae)-Ustilago bullata (Ustilaginaceae) pathosystem: A role for frequency-dependent selection? *American Journal of Botany*, *97*(8), 1304–1312.

Mitchell-Olds, T., & Schmitt, J. (2006). Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature*, *441*(7096), 947–952.

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(2), 453–458.

Muse, S. V. (2000). Examining rates and patterns of nucleotide substitution in plants. *Plant Molecular Biology*, *42*(1), 25–43.

Muse, S. V., & Gaut, B. S. (1997). Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. *Genetics*, *146*(1), 393–399.

Ness, R. W., Morgan, A. D., Colegrave, N., & Keightley, P. D. (2012). Estimate of the spontaneous mutation rate in Chlamydomonas reinhardtii. *Genetics*, *192*(4), 1447–1454.

Ness, R. W., Wright, S. I., & Barrett, S. C. (2010). Mating-system variation, demographic history and patterns of nucleotide diversity in the Tristylous plant Eichhornia paniculata. *Genetics*, *184*(2), 381–392.

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, *154*(2), 923–929.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., et al. (2005). The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biology*, *3*(7), e196.

Olsen, K. M., Caicedo, A. L., Polato, N., McClung, A., McCouch, S., & Purugganan, M. D. (2006). Selection under domestication: Evidence for a sweep in the rice waxy genomic region. *Genetics*, *173*(2), 975–983.

Olsen, K. M., & Wendel, J. F. (2013). A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology*, *64*, 47–70.

Olson-Manning, C. F., Wagner, M. R., & Mitchell-Olds, T. (2012). Adaptive evolution: Evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, *13*(12), 867–887.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., et al. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science*, *327*(5961), 92–94.

Pérusse, J. R., & Schoen, D. J. (2004). Molecular evolution of the GapC gene family in Amsinckia spectabilis populations that differ in outcrossing rate. *Journal of Molecular Evolution*, *59*(4), 427–436.

Phillips, P. C., Otto, S. P., & Whitlock, M. C. (2000). Beyond the average; the evolutionary importance of gene interactions and variability of epistatic effects. In J. B. Wolf, E. D. Brodie, & M. J. Wade (Eds.), *Epistasis and the evolutionary process*. New York, NY: Oxford University Press.

Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, *117*(2), 353–360.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*(4), R208–R215.

Qiu, S., Zeng, K., Slotte, T., Wright, S., & Charlesworth, D. (2011). Reduced efficacy of natural selection on codon usage bias in selfing Arabidopsis and Capsella species. *Genome Biology and Evolution*, *3*, 868–880.

Rose, L. E., Michelmore, R. W., & Langley, C. H. (2007). Natural variation in the Pto disease resistance gene within species of wild tomato (Lycopersicon). II. Population genetics of Pto. *Genetics*, *175*(3), 1307–1319.

Rosyara, U. R., Gonzalez-Hernandez, J. L., Glover, K. D., Gedye, K. R., & Stein, J. M. (2009). Family-based mapping of quantitative trait loci in plant breeding populations with resistance to Fusarium head blight in wheat as an illustration. *Theoretical and Applied Genetics*, *118*(8), 1617–1631.

Roux, C., Pauwels, M., Ruggiero, M. V., Charlesworth, D., Castric, V., & Vekemans, X. (2013). Recent and ancient signature of balancing selection around the S-locus in Arabidopsis halleri and A. lyrata. *Molecular Biology and Evolution*, *30*(2), 435–447.

Savolainen, O., Langley, C. H., Lazzaro, B. P., & Fr, H. (2000). Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing Arabidopsis lyrata and the selfing Arabidopsis thaliana. *Molecular Biology and Evolution*, *17*(4), 645–655.

Schierup, M. H., Mable, B. K., Awadalla, P., & Charlesworth, D. (2001). Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of Arabidopsis lyrata. *Genetics*, *158*(1), 387–399.

Schmitz, R. J., Schultz, M. D., Urich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., et al. (2013). Patterns of population epigenomic diversity. *Nature*, *495*(7440), 193–198.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), 178–183.

Schultz, S. T., & Lynch, M. (1997). Mutation and extinction: The role of variable mutational effects, synergistic epistasis, beneficial mutations, and degree of outcrossing. *Evolution*, *51*, 1363–1371.

Slotte, T., Bataillon, T., Hansen, T. T., St Onge, K., Wright, S. I., & Schierup, M. H. (2011). Genomic determinants of protein evolution and polymorphism in Arabidopsis. *Genome Biology and Evolution*, *3*, 1210–1219.

Slotte, T., Foxe, J. P., Hazzouri, K. M., & Wright, S. I. (2010). Genome-wide evidence for efficient positive and purifying selection in Capsella grandiflora, a plant species with a large effective population size. *Molecular Biology and Evolution*, *27*(8), 1813–1821.

Slotte, T., Hazzouri, K. M., Agren, J. A., Koenig, D., Maumus, F., Guo, Y. L., et al. (2013). The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nature Genetics*, *45*(7), 831–835.

Slotte, T., Hazzouri, K. M., Stern, D., Andolfatto, P., & Wright, S. I. (2012). Genetic architecture and adaptive significance of the selfing syndrome in Capsella. *Evolution*, *66*(5), 1360–1374.

Smith, N. G., & Eyre-Walker, A. (2002). Adaptive protein evolution in Drosophila. *Nature*, *415*(6875), 1022–1024.

Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M., & Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature*, *400*(6745), 667–671.

Strasburg, J. L., Kane, N. C., Raduski, A. R., Bonin, A., Michelmore, R., & Rieseberg, L. H. (2011). Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Molecular Biology and Evolution*, *28*(5), 1569–1580.

Strasburg, J. L., Scotti-Saintagne, C., Scotti, I., Lai, Z., & Rieseberg, L. H. (2009). Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Molecular Biology and Evolution*, *26*(6), 1341–1355.

Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., & Lynch, M. (2012). Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(45), 18488–18492.

Sweigart, A. L., & Willis, J. H. (2003). Patterns of nucleotide diversity in two species of Mimulus are affected by mating system and asymmetric introgression. *Evolution*, *57*(11), 2490–2506.

The International Peach Genome Initiative (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, *45*, 487–494.

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., et al. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, *43*(2), 159–162.

Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M., & Bergelson, J. (2003). Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature*, *423*(6935), 74–77.

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nature Biotechnology*, *31*(3), 240–246.

Vinogradov, A. E. (2003). Selfish DNA is maladaptive: Evidence from the plant Red List. *Trends in Genetics*, *19*(11), 609–614.

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nature Genetics*, *40*(11), 1370–1374.

Whitney, K. D., Baack, E. J., Hamrick, J. L., Godt, M. J., Barringer, B. C., Bennett, M. D., et al. (2010). A role for non adaptive processes in plant genome size evolution? *Evolution*, *64*(7), 2097–2109.

Wright, S. (1939). The distribution of self-sterility alleles in populations. *Genetics*, *24*(4), 538–552.

Wright, S. I., & Andolfatto, P. (2008). The impact of natural selection on the genome: Emerging patterns in Drosophila and Arabidopsis. *Annual Review of Ecology, Evolution, and Systematics*, *39*, 193–213.

Wright, S. I., Lauga, B., & Charlesworth, D. (2002). Rates and patterns of molecular evolution in inbred and outbred Arabidopsis. *Molecular Biology and Evolution*, *19*(9), 1407–1420.

Wright, S. I., Ness, R. W., Foxe, J. P., & Barrett, S. C. H. (2008). Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences*, *169*, 105–118.

Xiong, L. Z., Liu, K. D., Dai, X. K., Xu, C. G., & Zhang, Q. (1999). Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between Oryza sativa and O. rufipogon. *Theoretical and Applied Genetics*, *98*, 243–251.

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*, *30*(1), 105–111.

Yang, C. C., Kawahara, Y., Mizuno, H., Wu, J., Matsumoto, T., & Itoh, T. (2012). Independent domestication of Asian rice followed by gene flow from japonica to indica. *Molecular Biology and Evolution*, *29*(5), 1471–1479.

Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, *17*(2), 155–160.

Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, *178*(1), 539–551.

Züst, T., Heichinger, C., Grossniklaus, U., Harrington, R., Kliebenstein, D. J., & Turnbull, L. A. (2012). Natural enemies drive geographic variation in plant defenses. *Science*, *338*(6103), 116–119.