Population Genomics for Understanding Adaptation in Wild Plant Species

Detlef Weigel[1] and Magnus Nordborg[2]

[1]Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; email: weigel@weigelworld.org

[2]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna Biocenter, 1030 Vienna, Austria; email: magnus.nordborg@gmi.oeaw.ac.at

Annu. Rev. Genet. 2015. 49:315–38

First published online as a Review in Advance on October 2, 2015

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev-genet-120213-092110

Keywords

natural selection, evolution, adaptation, population genomics, plants, *Arabidopsis*

Abstract

Darwin's theory of evolution by natural selection is the foundation of modern biology. However, it has proven remarkably difficult to demonstrate at the genetic, genomic, and population level exactly how wild species adapt to their natural environments. We discuss how one can use large sets of multiple genome sequences from wild populations to understand adaptation, with an emphasis on the small herbaceous plant *Arabidopsis thaliana*. We present motivation for such studies; summarize progress in describing whole-genome, species-wide sequence variation; and then discuss what insights have emerged from these resources, either based on sequence information alone or in combination with phenotypic data. We conclude with thoughts on opportunities with other plant species and the impact of expected progress in sequencing technology and genome engineering for studying adaptation in nature.

INTRODUCTION

This article is about how population genomics—the analysis of whole-genome polymorphism data from large population samples—can help us understand adaptation. There is obviously nothing plant-specific about this: The title primarily reflects the typical pigeonholing of researchers by organism rather than question. That said, we argue that, for the purpose of studying adaptation, wild plants are ideally suited and seriously underutilized.

Why Genomics?

As a prelude, it is worth asking why we need genomics to study adaptation. After all, Darwin deduced much about adaptation without even the knowledge of genetics, and there is a considerable body of fieldwork that is exclusively focused on measurements of fitness and the phenotypes that influence it, without any consideration of genotype. As noted by Lande & Arnold in the opening statement of their seminal 1983 paper:

> "Natural selection acts on phenotypes, regardless of their genetic basis, and produces immediate phenotypic effects within a generation that can be measured without recourse to the principles of heredity or evolution" (87).

However, they continue:

> "In contrast, evolutionary response to selection, the genetic change that occurs from one generation to the next, does depend on genetic variation."

They concluded that it is necessary to characterize the genetic variation, in particular to determine the extent to which variation for a trait is due to additive effects that can readily respond to selection. This can be done in a classical quantitative genetics framework, modeling genes statistically and without considering the effects of individually known genes, essentially ignoring any advances in genetics since the 1940s.

Many studies were inspired by their analysis (the paper is one of the most highly cited in evolutionary biology), but it soon became apparent that the approach had limitations. Unfortunately, even with excellent natural history knowledge of a species, it is next to impossible to know a priori which traits should be measured (and on what scale), and, with respect to field studies, what the appropriate temporal and spatial dimensions of the experiments should be. It also became clear that the genetic details could matter greatly (6), contributing to a long-running debate about the importance of epistasis and a large number of publications justified by the need to understand genetic architecture. Perhaps more importantly, the pure quantitative genetics approach was overtaken from the 1980s on by the molecular biology revolution, which raised the possibility of identifying individual genes and polymorphisms involved in adaptation, at least if they had large effects. Although it is hotly debated how finding such—probably atypical—genes would advance our general understanding of adaptation (91, 126), for better or worse, studies discussing real genes proved rather more influential than those merely containing tables of ANOVA (analysis of variance) results. Today, however, the quantitative genetics approach is experiencing something of a revival, owing to the limited success of large-scale genome-wide association studies (GWASs) in humans. For most traits, these studies have not managed to identify more than a tiny fraction of causal alleles: the problem of the so-called missing heritability. This has served as a dramatic reminder of the possibility that a trait can be highly heritable without any of the underlying

polymorphisms having a marginal effect large enough to be measured (90, 102, 160). Needless to say that this is making many colleagues rather uncomfortable, as it implies an unbridgeable gap between complex traits and traditional molecular biology.

In evolutionary biology, this identifiability problem is made worse by the fact that selection acts on large populations and over long periods of time, thus the classical result that selection (rather than drift) determines the fate of an allele (once it has reached appreciable frequency) as long as its effect on fitness is greater than the inverse of the effective population size (75). Even for large organisms with small population sizes, this means that there can be strong selection for (or against) alleles that increase (or decrease) fitness through effects that could never be directly seen as developmental, morphological, or physiological phenotypes. Furthermore, selection is influenced by both the biotic and abiotic environments, neither of which is constant over time. All things considered, it is no surprise that connecting genotype (in the sense of particular polymorphisms) to phenotypes, and ultimately to fitness, has been so difficult. Similarly, large fractions of genomes may remain unannotated, because even when knocked out, the genes or regulatory elements have phenotypic effects too small to be measured in experiments, yet easily large enough to be maintained by selection—another example of a possible unbridgeable gap to molecular biology.

However, the fact that something may be impossible according to accepted theory does not imply that it is impossible in practice. In many cases, it will surely be possible to characterize adaptively important polymorphisms molecularly (e.g., whether they affect gene expression or coding regions, whether they knock out gene function or merely change it, what types of proteins the affected genes encode, etc.), and as long as we are careful not to overgeneralize, these examples will teach us important details about the genetic basis of evolutionary change. Furthermore, although the effects of individual causative polymorphisms are often so small that it is difficult to study them in isolation, it should be possible to study them in aggregate, relying on their joint effect in the Fisherian sense. Although loci identified by GWASs may explain in some cases little of the genetic variation, it may still be possible to predict phenotypes based on the collection of all genome-wide variants (160). From this, we can learn a great deal about the molecular mechanisms underlying these phenotypes at the systems level by considering correlations between GWAS signals and other forms of annotation. This mirrors trends toward systems approaches in functional biology, in which the function of entire networks rather than individual genes is studied. In a similar manner, it may be much easier (and more informative) to make statements about selection on sets of polymorphisms that are correlated with body size, rather than trying to elucidate the selective history of particular loci (11, 145). Indeed, the continuing fascination with trying to prove, statistically, that particular loci are under selection is, to a large extent, simply a legacy of the old neutralist-selectionist controversy (80, 111).

A final argument in favor of approaches that start with genomes is that the genetic composition of a population reflects its history and thus must, by definition, contain the footprints of adaptation. This is in contrast to field experiments, which, regardless of whether they utilize genomic methods or not, may simply not have the power to detect long-term adaptive changes because the temporal variation in selective pressures is too large. A paradigm for this is the unpredictable evolution of Darwin's finches reported by Peter and Rosemary Grant: Short-term trait fluctuations would not have been predicted from estimates for long-term rates of change, owing to episodes of selection in one direction alternating with episodes of selection in the opposite direction (49). This limitation should not be taken as an argument against field studies, however, but rather as an argument for combining them with approaches that exploit signals of selection over large temporal and spatial scales—as well as with laboratory experiments, which reveal the molecular basis for the action of different alleles.

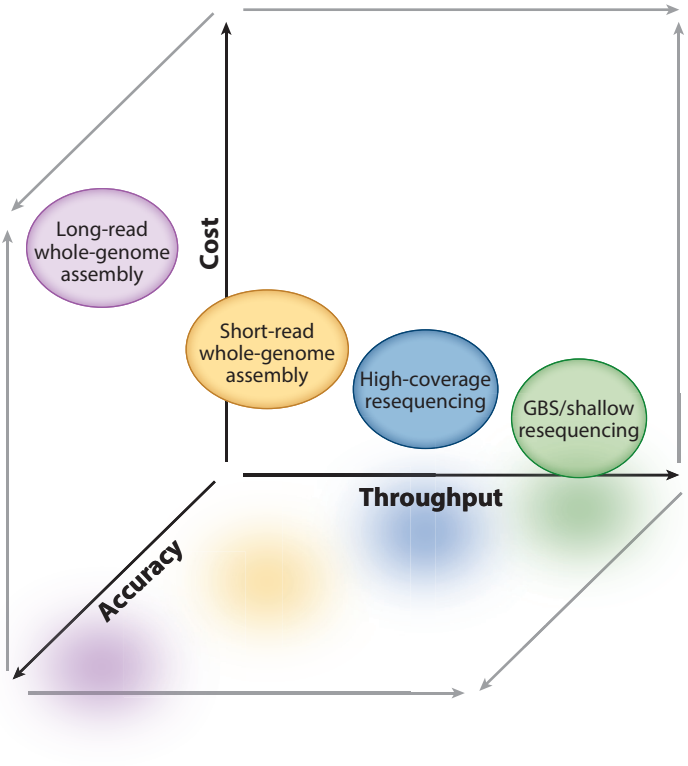

Figure 1

Sequencing approaches and their cost, accuracy, and scalability for throughput. Abbreviation: GBS, genotyping-by-sequencing.

Uses of Genomics

We have chosen to focus on the utilization of whole-genome polymorphism data from large population samples, the generation of which is now practicable for individual researchers, even if one first has to produce a reference genome assembly (**Figure 1**; see sidebar, Current Advances in Genome Sequencing and Genotyping). Such data provide an estimate of the current genetic composition of the population, which can be used to:

- Infer its history and structure.
- Infer the history of particular loci (in particular with respect to selection, see **Figure 2** and sidebar, Footprints of Selection).
- Reveal correlations between polymorphism and phenotypes or other characteristics, such as location or climate of origin (52, 88) (see sidebar, GWASs and Population Structure).

The close connection between these various uses is not always appreciated. For example, consider two populations that differ strongly with respect to some adaptive traits. To identify the genes that underlie these differences, we can either use methods that rely on genomic signals of selection (**Figure 2**) or we can view this as a mapping problem and identify the adaptive alleles using GWAS approaches. Either way, it is a challenge to distinguish true from false signals. The GWAS field is considerably more advanced than selection scans when it comes to dealing with

CURRENT ADVANCES IN GENOME SEQUENCING AND GENOTYPING

In the past decade, the short-read sequencing revolution has dramatically lowered the price of genome assemblies, although generally at the expense of assembly quality. A second revolution is currently under way. There are new physical methods for arranging assembly scaffolds along chromosomes without genetic linkage information (19, 71, 84). A further improvement comes from long-read sequencing, which can generate from shotgun data highly contiguous assemblies (73). Many plant species have haploid genomes of less than 1 Gb (93). Generating approximately 50× sequence coverage for such genomes with long-read technologies costs much less than 100,000 US dollars, and even less for species with smaller genomes. As before for short-read sequencing, it is almost certain that these costs will continue to drop steeply in the next few years. Thus, high-quality reference genome assemblies are within reach for many of the species long preferred by evolutionary biologists and ecologists (**Figure 1**). Short-read assemblies may continue to be cost-effective for a larger number of individuals, and this can be complemented with low-coverage sequencing for even more individuals. Finally, where individuals are closely related, genotyping-by-sequencing of thousands of markers is very cost-effective, costing a few US dollars per sample (119).

these issues, mostly because it is easier to specify realistic genetic models (152). When looking for footprints of selection, about the best that can be done is to use genome-wide data to try to estimate a selectively neutral null model, but this only makes sense if most of the genome is indeed evolving neutrally (see sidebar, Footprints of Selection). To use a well-known quote of unknown origin, it is important to avoid "using statistics as a drunk uses a lamppost—for support rather than illumination." Regardless of the approach taken, scans for selection can only generate lists of genes that are likely to be important. Unless these results can be combined with independent evidence, they are arguably not very interesting.

One important source of independent evidence comes from direct estimates of allele frequency change. If we have population genomics data from multiple time points or generations (which can include paleontological remains or herbarium specimens; 162), we can directly identify polymorphisms under selection using allele frequency changes. Analogously, we can confirm a genotype-phenotype association using linkage mapping. In both cases, the advantage is that the signal is not confounded by history, and the disadvantage is the lack of resolution in pinpointing individual loci due to extensive linkage disequilibrium. However, when combined with the approaches mentioned above, we can get the best of both worlds (**Figure 3**), provided the experiments are carefully designed. A comparison between genotypes with very distant geographic origins may not be that meaningful, whereas differences between local genotypes may be too small to yield meaningful results in a reasonable time frame.

Population genomics approaches can also be classified according to whether we start with a phenotype or not. Scans for footprints of selection (see sidebar, Current Advances in Genome Sequencing and Genotyping), ongoing or historical, are agnostic with respect to the traits under selection, and the same is true for mapping studies that use fitness as a phenotype. These approaches have the advantage that they do not make assumptions about the importance of specific traits for selection in the wild—-but suffer from the obvious disadvantage that the ease with which one can make a connection to a specific trait very much depends on the alleles and genes identified. And if we do not do that, all we have accomplished is demonstrating that allele frequencies change as a result of selection—hardly a major insight. Field and laboratory studies are needed, but it is important to remember that even if there is evidence from laboratory studies for the involvement of a gene in a certain biological process, it may not be the one that is responsible for selection in nature. Furthermore, functional annotation is usually based on the effects of knockout mutations,

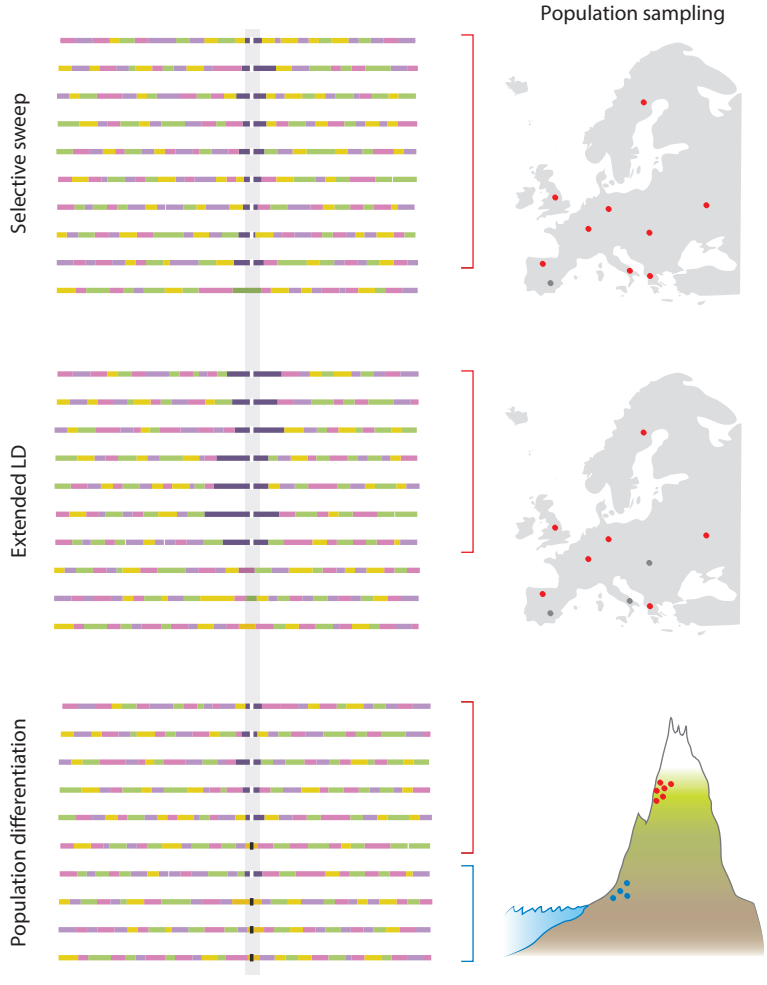

Figure 2

Principal types of selection scans. Colors symbolize ancestral chromosomes. Mutation(s) selected for are denoted by short white and black vertical lines in the center. Population sampling illustrated on the right.

whereas natural variants often have more subtle effects on gene function, or generate a new function all together. Also note that selection does not necessarily equal adaptation (54). Major genetic differences between populations may arise because of Red Queen effects—arms races with pathogens that certainly involve strong selection but may be irrelevant for the adaptive differences between the populations.

The opposite approach is to draw on knowledge about the natural history of a species and to begin with explicit hypotheses about traits that are under selection. Variation in phenotypes can be assessed in uniform conditions indoors, which more easily reveals the effects of genotype, or in naturally variable conditions outdoors, which captures ecologically meaningful genotype-environment interactions. After having measured the traits of interest, causal genes can be identified through standard genetic mapping approaches, often with a combination of experimental crosses and GWAS approaches. Once causal variants have been confirmed using fine mapping,

FOOTPRINTS OF SELECTION

There are three main approaches to scanning genomes for evidence of selection (**Figure 2**): frequency spectrum-based methods that look for either a reduction in sequence diversity or an increase in high frequency–derived alleles as a consequence of recent directional selection, or for an excess of intermediate-frequency polymorphisms as a sign of balancing selection; linkage disequilibrium (LD)-based methods to reveal unusually long tracts of linked polymorphisms that have not yet been broken down by recombination; and approaches that identify polymorphisms with very different frequencies in different populations, indicative of local adaptation (153) (**Figure 1**). These were used already in the pregenomic era, but without whole-genome information it is very difficult to gauge how the polymorphism patterns have been shaped by other forces than selection, especially confounding demographic factors, although it is still unclear how well one can estimate confounders (e.g., 67, 140). In any case, these methods can only generate lists of candidate alleles that need to be tested experimentally. This may include biochemical assays if the molecular function of the candidate is known, phenotypic analyses if involvement in specific biological processes is known, and ultimately tests for fitness effects under natural conditions (with all the caveats discussed in the main text).

complementation crosses and/or transgenic analyses, one can return to the population genomics data to ask relevant questions about long-term selection. Limitations of this approach are that traits that appear to be obvious to us as being of ecological and evolutionary importance may not be the ones actually under selection, or that selection might have been too weak to leave a strong DNA sequence signal, so that one cannot easily make statements about the adaptive value of a specific variant.

Although field experiments are not a panacea, they are clearly fundamental to any claim of adaptation, and many of the difficulties discussed above can be addressed by carrying out experiments over longer periods of time and in multiple locales. Moreover, owing to the dramatic fall in genotyping costs (see sidebar, Current Advances in Genome Sequencing and Genotyping), it is no longer necessary to follow individuals of a known genotype, which is normally done by transplanting plants germinated in the greenhouse. Instead, one can seed field sites directly, which more closely mimics the natural life cycle, and determine genotypes after phenotyping. This also

GWASs AND POPULATION STRUCTURE

GWAS associations may not reflect causal relationships, which are sometimes, but not always, attributable to population structure (152). A naïve test of association of a particular allele with a phenotype implicitly ignores other loci that (also) affect the trait (3). This is not a reasonable assumption in a structured population, where physically distant loci are in linkage disequilibrium, and also not for selected traits, because causal loci may be co-selected.

The classical "animal model" of quantitative genetics was developed to deal with this issue (56). Maize geneticists (163) have led the way in introducing this model to GWAS approaches, and human genetics is following suit (152). Broadly speaking, one first estimates how much of the phenotypic similarity between individuals can be explained by their relatedness, which includes genome-wide sharing of causal alleles, before estimating contributions from individual variants. The theoretical underpinnings date back to Fisher's model of correlations between relatives for quantitative traits (39). A similar approach can be used when scanning the genome for correlation with climate or similar variables (51, 52). The approach of course works only to the extent that Fisher's model of many alleles with small effects is right.

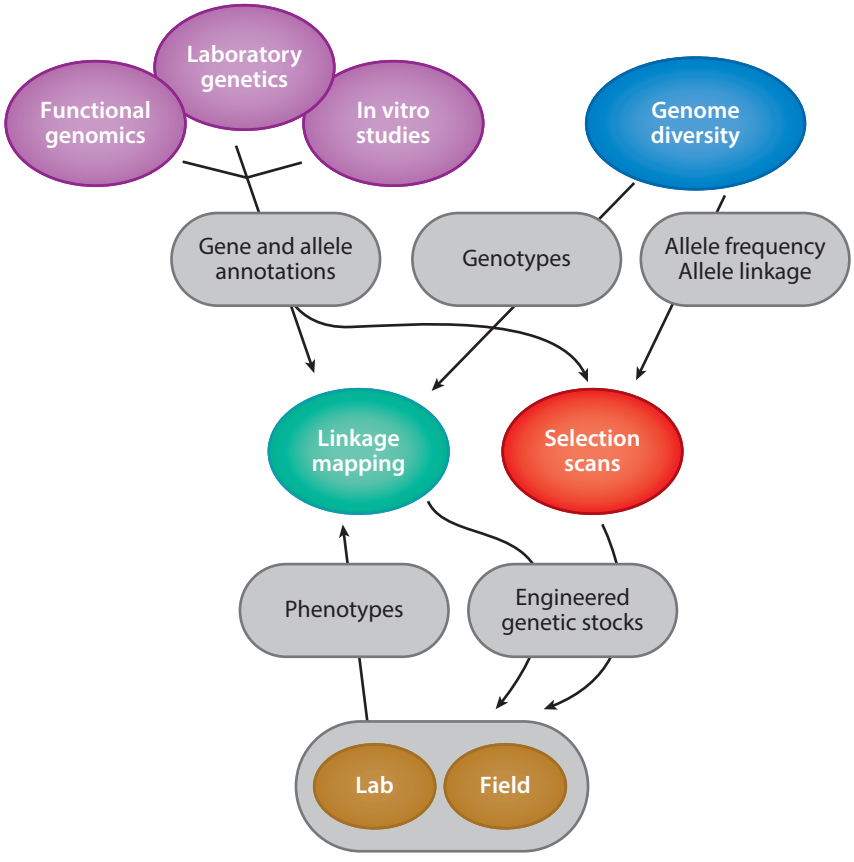

Figure 3

Dependencies among data types and approaches for the identification of functionally differentiated alleles. Engineered genetic stocks: transgenic lines, near isogenic lines, allele replacement lines, etc.

opens the door to long-term experiments under natural conditions, using genomic scans to map genes and alleles under selection.

In summary, it is obvious to us that if we wish to advance our understanding of adaptation beyond that of Darwin, genomic approaches are indispensable. We know the record is in the genome—the challenge is to decode it.

Why Plants?

The advantages of plants for studying adaptation are obvious. "Plants stand still [. . .] to be counted" (53, p. 515), making field studies much easier than with most animals, especially very small animals such as flies and worms. Understanding selective pressures is considerably more difficult if you are not even able to observe the organism in its natural environment. Furthermore, in perennial species, the same individuals can be easily studied over many years, which affords opportunities to study genotype-specific responses to a changing environment. Because of their limited mobility, plants are also almost certainly more likely to be locally adapted. It is no coincidence that some of the first rigorous studies of local adaptation were carried out in plants (31, 146, 147).

This is especially true for populations that thrive in contrasting habitats, such as normal and serpentine soils (17). Clear evidence for strong local adaptation is even found in plants such as wind-pollinated forest trees that appear to come close to the population genetics ideal of a large, random-mating population (81, 106, 125). Two further practical advantages are that many plant species, at least when they are herbaceous, are quite easily maintained in greenhouses or gardens, especially when they propagate by self-fertilization, and that it is often relatively straightforward to set up experimental crosses, including between closely related species.

From the point of view of evolutionary genomics, plants also have an advantage in that closely related taxa can differ greatly in mating system, ploidy level, and life history, providing opportunities for addressing many important questions of evolution. For example, within the genus *Arabidopsis*, we have an annual selfer with a global distribution (*Arabidopsis thaliana*), a heavy-metal tolerant, perennial, obligate outcrosser (*Arabidopsis halleri*), a selfing, allotetraploid interspecific hybrid (*Arabidopsis suecica*), and ploidy variants of the same species (*Arabidopsis lyrata*, *Arabidopsis arenosa*). Both selfing and polyploidy are predicted to affect many aspects of molecular evolution (29, 116), predictions that can be tested particularly well in comparisons of close relatives that have easily alignable genomes. Because of the enormous diversity in mechanisms of seed and pollen dispersal, plants also differ greatly in population structure. This, again, is predicted to have important implications for the genetic basis of local adaptation (123).

Given these advantages, the dominance of studies of adaptation in animals is almost shocking. Although much of the work in animals has addressed questions related to speciation and reproductive isolation (e.g., 36, 136), there is an increasing number of studies focusing on adaptation, e.g., to new habitats in sticklebacks (69), to brackish water in herring (86), or to different food sources in Darwin's finches (85). The repeated evolution of mimicry in *Heliconius* butterflies (124) is another example. In contrast, although there are many population genomics studies in plants, essentially all have focused on crop plants or their close relatives, including maize/teosinte, rice, soybean, tomato, chickpeas, pepper, and even cucumber (see the recent review in 64). The major exception is *Arabidopsis thaliana*, which was not only the first plant but also one of the first species overall for which rich genome-wide and subsequently whole-genome polymorphism data became available (22, 30, 45, 61, 74, 99, 112, 114, 131, 132). Information on other wild species that are not closely related to crops is emerging only very slowly (18, 37, 43, 44, 47, 59, 121, 161). This presumably reflects funding priorities, which emphasize new resources for breeding, but it represents a missed opportunity if one wants to understand how wild species adapt to natural environments. Indeed, when we agreed to write this review, we had hoped to cover a broad array of species; as things stand, this review is dominated by *A. thaliana*—and note that these data could only be generated because a few enlightened granting agencies have fortunately funded blue sky research. We are of course fully aware that the mating system of *A. thaliana*, which appears to outcross on average only about once in 10 to 100 generations (13), has consequences for population diversity and evolutionary signatures. It therefore remains to be seen how good a model *A. thaliana* is for the majority of plant species, which are outcrossing.

POPULATION GENOMICS OF *ARABIDOPSIS THALIANA*

Describing Genome Variation

Arabidopsis thaliana was the first plant and one of the first multicellular species with a high-quality reference genome sequence, and the initial genome paper already disclosed results from low-pass shotgun sequencing of a different accession (1). Although this paper underestimated the extent of sequence differences, it highlighted that such differences were not restricted to single nucleotide

polymorphisms (SNPs) and small insertion-deletions (indels) but could encompass fragments that were dozens of kilobases long. The next milestone was the use of nearly 2,000 evenly spaced sequence tags throughout the genome to determine sequence diversity in a larger set of accessions (112), which revealed genetic isolation by distance, even though intermediate-frequency alleles are typically found across most of the species range. It also set the stage for one of the first whole-genome studies of sequence variation in any organism, using massive oligonucleotide arrays (30). One of the findings from these early studies was that polymorphism is nonrandomly distributed, with polymorphism rates near the centromeres being higher near the centromeres than on the chromosome arms. These efforts also documented a surprising number of apparent knockout mutations in different accessions, with a heavy bias toward genes involved in interactions with the biotic environment, indicative of tradeoffs linked to an active immune system (30).

Arabidopsis thaliana was also the first plant for which short-read resequencing data were generated, in parallel with similar efforts in humans (114), and the early success with this technology led to calls for a 1001 Genomes project for the species (113). The resequencing studies provided a richer picture of variants, including detection of sequences absent from the reference, better estimates of copy number variation, and clearer evidence for isolation as well as evidence for reduced selection in populations at the edge of the native species range. Some of the additional insights were that large differences in rDNA copy number can lead to massive genome size variation and that mutation spectra for single base substitutions differ between the wild and the laboratory (22, 45, 68, 99, 115, 132).

A more or less dirty secret of "complete" genome sequences obtained from resequencing has been that large-scale variants are either entirely ignored or, at best, only partially analyzed. Even the isolated consideration of small-scale variants can be misleading, as there are often compensatory mutations that make up for frameshifts, splice-site mutations, and so on (45, 99, 132). In addition, the presence of paralogs in the nonreference genomes can lead to erroneous inferences about polymorphisms, including heterozygosity and copy number variation, especially when analyzing segregating populations (98, 114, 122). Assembly of so-called leftover sequencing reads that could not be mapped to the reference has shown that *A. thaliana* accessions contain on the order of 2 Mb of single-copy sequences that are not represented in the reference genome (99). Often, matches for such fragments can be found in *A. lyrata* (63), indicating that they represent ancestral sequences that were lost in the reference strain. The reference bias is not evenly distributed, with genes involved in disease resistance being most prominent among those for which mere resequencing fails to describe the full extent of variation. The same is true for transposons and other repetitive sequences (22, 99).

Describing Epigenome Variation

Scientists interested in adaptation differ greatly on the importance of epigenetic variation (34). Most epigenetic modifications, such as histone modifications and DNA methylation, appear unchanged from generation to generation, either because of stable inheritance or because of erasure and resetting in the germ cells or early zygote of any alterations that have accumulated during somatic development. Dense DNA methylation in all three sequence contexts known in plants (CG, CHG, and CHH, where H is any base but G) is typically associated with transposon and repeat silencing, and involves a complex interplay with small RNAs and chromatin remodeling (103). Silencing of transposons or repeats in turn can have knock-on effects on the expression of adjacent genes, leading to an evolutionary trade-off between the obvious advantages of transposon silencing and the collateral damage caused by altered, usually reduced expression of nearby genes (60). It is thus likely that there is selection for reduced silencing of transposon sequences that have

become inactivated by mutation, and an outcome of this might be that some sequences are more likely than others to spontaneously cycle between methylated and unmethylated states in different generations (9, 21, 50, 55, 130).

Although transposon and repeat methylation are much less variable between accessions than gene body methylation (the function of which remains somewhat mysterious), differences do exist (150), and this motivated the large-scale investigation of whole-genome methylation patterns in natural accessions (35, 50, 131). In one study (131), many more differentially methylated regions (DMRs) that had hallmarks of gene silencing were linked to local DNA sequence variants than to *trans*-controlling loci, with more than 10% having evidence for polygenic control. Unfortunately, because detection of structural variants from resequencing data is still imperfect, it was not possible to provide exact estimates of how many DMRs are completely independent of DNA sequence changes. However, investigation of a near-clonal population of wild accession confirmed that spontaneous DMRs in the absence of apparent sequence variation are common (50). One conclusion from this other study was that certain regions of the genome are apparently privileged for the spontaneous occurrence of DMRs, in line with such changes reflecting cycling of DNA methylation at transposon or repeat sequences (21, 55).

Importantly, both induced and spontaneously occurring epigenetic variation at discrete loci in the genome can lead to heritable phenotypic differences that are similar in magnitude to those caused by natural genetic variation (32, 127, 135). An interesting question is therefore whether spontaneous DMRs contribute to evolution. In one scenario, attendant gene expression changes would amplify the effects of a specific allele so that it can now be selected for. In another scenario, the increased mutation rate at methylated sites (22) increases the chances of a genetic change that stabilizes the effects of the epigenetic effect, in a process reminiscent of genetic assimilation (155).

In addition to local DNA methylation differences, wholesale differences in CHH methylation throughout the entire genome have been documented that are apparently due to a change at a single *trans*-acting locus that encodes the CMT2 methylase (35, 133). Natural knockout alleles at *CMT2* are found in approximately 10% of all accessions, and they appear to affect a broad suite of traits, from heat-stress tolerance to immune response and leaf morphology. Moreover, not only is the distribution of these alleles geographically skewed, with *CMT2* knockout alleles being more frequent in places with a greater temperature range, but genome-wide CHH methylation is reduced when plants are grown in a cool environment, which may reflect selection for stronger silencing of transposons at higher temperatures (35, 133).

Genome-First Approaches for Identifying Loci Under Selection

Somewhat disappointingly, whole-genome selection scans (**Figure 2**; see sidebar, Footprints of Selection) that focused on the global population of *A. thaliana* have turned up remarkably few candidate regions for strong, species-wide sweeps. By far the most obvious such region is one that was already identified in the first 20 accessions with whole-genome information (22, 30, 61, 99). This sweep is only several tens of thousands of years old (99), which is much younger than the species itself, estimated to be at least a million years old (139). Not only is the sweep allele found in well over 90% of all accessions throughout the worldwide range, but it also stands out because it is several hundred kilobases long. Recombination in this region between accessions with the sweep allele and other alleles is suppressed because it is associated with the transposition of an approximately 300-kb fragment to a location some 0.5 Mb away on the same chromosome (99). Unfortunately, none of the genes in this interval have been assigned functions in laboratory studies that would provide insights into their potential roles in adaptation to natural environments.

Selection scans were more successful with a local population from northern Sweden, which produced much stronger signals. This was not simply due to the particular composition of the sampled population but reflected apparently much stronger selection in this geographic area (66, 99). Because of the extensive linkage disequilibrium in this population, the signals extended over large regions, and individual candidate loci were therefore difficult to discern. Nevertheless, these findings highlight the importance of appropriate population sampling in unbiased selection scans.

A less broad approach is to search for polymorphism patterns that are correlated with environmental variables (41, 51, 88, 89); when these have an explicit geographic component, one speaks of landscape genomics (101). Although such studies have had limited success in pinpointing the individual loci that underlie adaptation, they have clearly confirmed that environmental factors, and in particular a host of climate parameters, shape the pattern of genomic variation in *A. thaliana*. Confidence in the conclusions comes from field trials, which have shown that alleles associated with local climate variables are also correlated with proxies of fitness, such as number of fruits or seeds (41). Conversely, by growing the same set of accessions in different sites across the range and then comparing the results with GWASs on fruit number it was found that locally favorable alleles tended to be most frequent near the sites where they were positively associated with fitness (51). Most loci appeared to be favorable in only one location, with effects at other locations having a tendency to be negative. Although there was a significant association with local climate, a caveat is that it is difficult to completely disentangle correlations between climate variables and geographic location. In support of climate itself having a more limited effect, there was little evidence for these loci having experienced recent strong selection, as one might expect given that the climate in Eurasia, the native range of the species, has changed dramatically over the past tens of thousands of years (33). Finally, there are also biotic variables that vary with geography, an impressive example being the distribution of different aphid species that prey on *A. thaliana* and that have apparently driven a longitudinal gradient of variation at two loci responsible for defensive chemicals of the glucosinolate class. Support for this pattern being indeed due to a causal relationship comes from the rapid responses of different *A. thaliana* accessions when they were exposed over several generations to these two aphids in the greenhouse (165).

Phenotype-First Approaches for Identifying Individual Loci Affecting Ecologically Relevant Traits

An alternative to the largely agnostic approaches discussed above is to combine information on population genome variation with measurements of discrete phenotypes that are suspected of being adaptive, or at least under selection. In *A. thaliana*, such phenotype-driven GWAS analyses can work amazingly well. Prime examples of successes have been disease resistance, flowering, and abiotic stress responses. As in most other systems, studies of natural genetic variation in *A. thaliana* started with quantitative trait locus (QTL) mapping in F_2 or F_2-like populations, and **Figure 4** compares the power of QTL and GWAS mapping with that of naïve selection scans. Further discussion focuses on alleles that were discovered, or at least confirmed, in GWAS studies, which usually require that variants are reasonably common in the study population.

Disease resistance and microbe interactions. Some of the earliest and clearest successes in (re)discovering natural alleles with a large contribution to a specific trait came from GWAS analyses of disease resistance (2, 3, 72, 117). This was not particularly surprising given that disease resistance is a paradigm for balancing selection, with causal alleles often occurring at moderate frequencies throughout the global population (4, 5, 20, 40, 42, 137, 141, 158). Disease resistance has

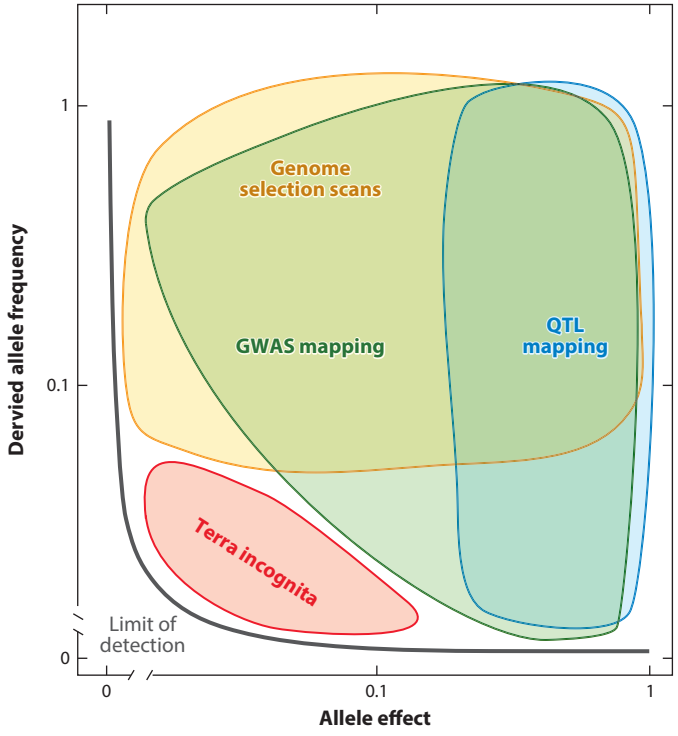

Figure 4

Power of different methods to detect functionally differentiated alleles depending on size of phenotypic effect and population frequency. Terra incognita indicates alleles inaccessible to current methods.

also provided examples for the productive combination of GWASs with mapping in experimental F_2 or F_2-like populations (65, 110, 143). In the case of the *RKS1* gene, a two-step GWAS analysis, focusing in the second step on accessions that carry an SNP associated with resistance, identified derivatives of the resistant allele with secondary mutations that led to a regaining of susceptibility (65).

Although the intermediate frequency of resistance alleles in both global and local populations is consistent with the expectation of balancing selection (4, 77, 143), it has remained a major challenge to determine the exact selective forces behind it. Elegant studies with isogenic lines that differed only for the presence or complete absence of a single NLR (nucleotide-binding leucine-rich repeat)-type resistance gene have demonstrated that the costs incurred in the absence of infection can be on the order of five to ten percent not only in the laboratory (46, 78) but also in field trials (72, 142). The most insightful study in this area has focused on the *RPS5* gene, which confers resistance to *Pseudomonas syringae* strains carrying *avrPphB* effectors (72). It has been estimated that the alternative alleles are more than two million years old and that they have been maintained since before the species arose. In several different genetic backgrounds, there are high fitness costs in absence of the pathogen, about half of what has been estimated as a fitness benefit in plants infected with *P. syringae* carrying *avrPphB* (46). Strikingly, the prevalence of the *RPS5* allele recognizing *avrPphB* is much greater than the incidence of *avrPphB* in *P. syringae* isolates. The conclusion was that *avrPphB* is only one of the factors responsible for maintenance of *RPS5* and that at the same time evolution of *avrPphB* is driven primarily by species other than

A. thaliana (72). These observations serve as a cautionary tale: Even in a case where we have an excellent mechanistic understanding of the interaction of a specific genotype with its environment, and where fitness effects can be clearly demonstrated, there may not be sufficient information to explain allele frequency.

The pronounced structural and copy number differences at many NLR loci are shared with loci responsible for variation in the content of insecticidal glucosinolates (25, 82). Importantly, whereas only about a quarter of glucosinolate-controlling loci known from QTL mapping in experimental crosses (76) have been detected by GWAS, the reverse is also true: GWAS detects many more loci than QTL mapping, especially when aided by additional information such as transcriptional networks (24, 25).

Flowering. The onset of flowering is the most intensively studied life history trait in *A. thaliana* because it is so critical for reproductive success. Even though a selfer does not have to coordinate flowering with other individuals, seeds must be produced before the weather becomes hot and dry in late spring or summer, or before snowfall in regions with a cold winter. The commonly accepted wisdom is that *A. thaliana* accessions fall into two groups, one that germinates and flowers in spring and the other germinating in fall and flowering the next spring, after having been exposed to vernalization, the winter-like temperatures that relieve an epigenetic flowering block; the reality, as usual, is more complex, and the same accession can show both behaviors, depending on when seeds germinate (157). Extensive laboratory work has shown that the two alternative life histories are primarily due to variation at the *FRI* and *FLC* loci (94, 156), although the effects of the *FRI-FLC* system under natural conditions are more limited (157). Nevertheless, there seem to be clear fitness effects, as specific allelic combinations at *FRI* and *FLC* can affect survival and seed set, depending on when plants germinate (79).

GWAS analyses under a variety of environments have confirmed both *FRI* and *FLC* as major flowering loci (2, 3, 164) and identified additional loci causing further variation in flowering (3, 14, 15, 95, 96). Direct evidence for the exact identity of the causal genes remains, however, an exception. Ideally, flowering loci would be mapped in field populations that germinate on their own. Because there is substantial genetic and phenological heterogeneity in local populations (13, 16, 48, 105, 107, 108, 134), a start would be to phenotype naturally occurring individuals, complemented with populations generated by experimental crosses between locally co-occurring individuals with contrasting phenotypes.

Element accumulation. The accumulation of several mineral nutrients and trace elements, so-called ionomic traits (128), is controlled by large-effect alleles as well. Traits for which common large-effect alleles have been identified by GWAS analyses include leaf concentrations of sodium, molybdenum, cadmium, and arsenic (7, 8, 27, 28). GWAS approaches are, however, not always successful, an example being the genetics of sulfur and selenium accumulation (26). The reason in this case is that causal alleles are rare and carry different mutations. The selective forces driving the emergence of alleles affecting variation in ionomic traits remain mostly unknown, except for sodium accumulation, which may be related to salinity tolerance. Similarly, the allele that is associated with higher leaf sodium concentrations is generally found closer to high-salinity environments than other alleles (7). The obvious next step must be to determine whether allelic variation is indeed linked to actual soil composition at natural *A. thaliana* sites.

General considerations for genome-wide association studies. Although population structure can confound the unambiguous identification of causal variants in GWAS analyses, the most highly ranked variants tend to be enriched for genes that have a known role in the trait of interest based

on mutational genetics or other experimental evidence (3). Thus, focusing on variants with the highest statistical support can provide a useful guide, even if such variants are nominally not statistically significant. When extended to traits for which the underlying mechanisms have not been investigated, this can lead to the discovery of new genes, a nice example being work on proline accumulation in response to mild drought stress (151). A further consideration is the inclusion of the organellar genotype in GWAS analyses. It is unclear whether this is a concern for all traits, but it has been shown that variation in organellar genomes makes a small but measurable contribution to metabolite content, as one would expect, because the chloroplast and mitochondria are essential for both primary and secondary metabolism (70). Finally, it is natural to ask what population is optional for GWAS, but this is not a question with a simple answer. Depending on the trait of interest, one might want to use a global population, or contrasting local populations. Furthermore, population structure does not only affect false positives. GWAS only works if causative alleles have appreciable frequency in the sample. If variation is due to too many different alleles (at the same or different loci), it has no power, and the risk of this may (again depending on the trait) be greater in a very diverse sample. There are now several examples of allelic heterogeneity in *A. thaliana*, both where independent alleles have the same effect and where different alleles differ in effect size (3, 10, 26, 94, 120, 164).

With the increasing number of genes responsible for the discovery of inter- and intraspecific variation in the past decade, not only in *A. thaliana* but also in other species, some have been debating the relative importance of different types of mutations, such as *cis*-regulatory changes, coding changes, knockout alleles, or even epigenetic differences (23, 58, 138). We do not consider such discussions very productive, as any answer suffers from a strong ascertainment bias, given that coding region changes and DNA mutations are generally more easily found and confirmed than regulatory sequence changes and epigenetic differences. There is certainly evidence for all these different types contributing to naturally occurring phenotypic variation in *A. thaliana*. Perhaps the most important finding has been that clear loss-of-function mutations can apparently be adaptive, as seen for all three traits we have discussed in detail: flowering time, disease resistance, and element accumulation.

Beyond *Arabidopsis thaliana*

It might strike the uninitiated as surprising that *A. thaliana* is the forerunner in population genomics of wild plants, given that there is a much longer tradition of evolutionary and ecological work for many other, often much more iconic species. The situation is reminiscent of the one for animals, where *Drosophila melanogaster* was for many years the workhorse of quantitative and population genetics, despite a lack of information on its natural habitat before it became closely associated with humans (83). The reason for this is the sophisticated genetic tool kit available in *A. thaliana* and *D. melanogaster*, and the premier position of the two species in these areas was reinforced with the initial genome sequences published in 2000.

Change is under way, however. Serious population genomics begins by definition with at least one high-quality reference genome assembly, something that has become much more affordable, not just for individual labs but even as the basis for individual graduate student or postdoctoral projects (see sidebar, Current Advances in Genome Sequencing and Genotyping). Similarly, the costs for large-scale resequencing and genotyping continue to fall. Depending on the application, such data might be generated for fewer individuals at higher coverage or for more individuals at lower coverage (**Figure 1**). Practical limitations for investigating adaptation are therefore no longer the lack of genome and genome variation information but rather the lack of techniques for genetic manipulation that allow experimental tests of specific genes and alleles.

Some of the first studies of adaptation with whole-genome data focused on other *Arabidopsis* species. There is good evidence for local adaptation in *A. lyrata* from transplant experiments (e.g., 92), and initial population genomics analyses have suggested footprints of selection and climate associations in *A. halleri* (38). The *A. lyrata* reference genome (63) was used to identify candidate loci for adaptation to serpentine soils in this species (148) and for genetic adaptation to genome doubling in the close relative *A. arenosa* (59, 161). The strategy was similar in both cases, using contrasts between populations on serpentine and granite soils or diploid and tetraploid populations, and this approach should be informative for *A. halleri* as well, where there are many populations that have evolved metal hyperaccumulation and tolerance. Contrasts between populations growing on saline and nonsaline soils also supported the identification of candidate loci involved in salinity adaptation in *Medicago truncatula* (43, 44).

The most thorough and impressive characterization of genomic diversity in wild plants comes from *Populus trichocarpa*, which, almost 10 years ago, was the first tree and only the third plant for which a genome assembly became available (149). More than 400 accessions have been interrogated with a genotyping microarray (47, 104), and another set of more than 500 individual trees was resequenced to high coverage (37). These studies revealed the extent of population structure, including admixture with other species, identified candidate loci under selection (37, 47), and identified through GWASs candidates for loci affecting important biomass, ecophysiology, and phenology traits (37, 104). An encouraging result was that candidates from population genomics selection scans scored on average more highly in the GWAS scans than random markers, providing further evidence for an adaptive role for these loci (37).

A limitation of studies with *Arabidopsis* and other Brassicaceae is that the major types of root symbioses of flowering plants with bacteria or fungi have been lost in this family, although GWAS approaches have already been exploited to document genotype-dependent variation in the association with bacterial and fungal communities that colonize *A. thaliana* leaves (62). *Populus* species are known to differ in their metabolomic responses to colonization by the fungus *Laccaria bicolor* (144), suggesting that intraspecific GWAS analyses will be useful for identifying alleles that modulate symbiotic interactions, given that the selection for such interactions should depend on the availability of inorganic resources in the soil.

More generally, trees hold great promise for understanding adaptation because they feature some of the best documented geographical clines in traits, many of which are clearly adaptive (109, 129). One of these is growth cessation in fall, and molecular key regulators of this trait in *Populus* vary in their activity depending on latitudinal origin of the accessions (12). Of particular interest is structured quantitative variation and how it is maintained in the face of gene flow: Are there a few loci featuring alleles of different activity, or is a phenotypic gradient achieved through combination of many genes, each of which has qualitatively different alleles? Latitudinal clines are well known not only in angiosperm trees but also in gymnosperms (106), and there already have been attempts at GWASs in pine (118). However, because of the enormous genome size, it has been difficult to link genetic markers to annotated genes.

Plants also offer wonderful opportunities to study adaptive radiations in groups that are experimentally much more tractable than, for example, Darwin's finches or cichlids. Two such ongoing efforts focus on North American columbines, *Aquilegia* spp. (57), and monkeyflowers, *Mimulus* spp. (159). Although reference genomes have not yet been published, the *Mimulus guttatus* genome draft has already been exploited to produce interesting insights into whole-genome patterns in native and introduced populations of *M. guttatus*, including the identification of major selective sweeps in the introduced range (121).

Finally, successes in GWASs with domesticated sunflower and tomato accessions bode well for population genomics in their wild relatives, many of which have interesting ecologies (97, 100).

SUMMARY POINTS

1. Wild plant species have obvious advantages for studying adaptation but are amazingly underutilized.
2. The ability to phenotype and genotype truly wild individuals at large scales is transformative, especially for trees.
3. A powerful experimental approach consists of field experiments with near isogenic lines.
4. A drawback of field studies is that they are limited with respect to temporal timescales (years to decades with living material, and centuries when including, for example, herbarium material) and geography (because of logistics and expense).

FUTURE ISSUES

1. Although it has been difficult to engineer precise near-isogenic lines in the past, CRISPR/Cas9 technology should soon make precise allele swaps, down to individual causal nucleotides, the norm for fitness experiments in natural settings. The use of RNA viruses as vectors would circumvent the need to develop cumbersome genetic transformation protocols (154).
2. Inference from such studies becomes particularly strong when they include either close relatives that are distinguished by mating system or life history, or less closely related taxa that have similar mating systems and life histories.
3. Increasingly, these studies will consider plant species not in isolation, but include their interaction with other species, such as plant competitors, as well as animal and microbial predators and mutualists.
4. The acquisition of very high quality, very large population genomics data sets will become more and more affordable, and this can partially compensate for the limitations inherent in field and experimental studies. The acquisition of such data will be a sine qua non in studies of adaptation, to be complemented with data from the field and, depending on the organism, the laboratory.

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 thank members of our labs for discussion and comments. D.W. thanks Vincent Colot for discussions about genetic assimilation. Collaborative studies of our two labs are supported by the Austrian Science Fund and the Deutsche Forschungsgemeinschaft (Schwerpunktprogramm ADAPTOMICS). Our work on natural genetic variation is generously supported by the European Research Council (Projects MAXMAP and IMMUNEMESIS), the Austrian Academy of Sciences, and the Max Planck Society.

LITERATURE CITED

1. *Arabidopsis* Genome Initiat. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
2. Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, et al. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLOS Genet.* 1:e60
3. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–31
4. Bakker EG, Stahl EA, Toomajian C, Nordborg M, Kreitman M, Bergelson J. 2006. Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol. Ecol.* 15:1405–18
5. Bakker EG, Toomajian C, Kreitman M, Bergelson J. 2006. A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *Plant Cell* 18:1803–18
6. Barton NH, Turelli M. 1989. Evolutionary quantitative genetics: How little do we know? *Annu. Rev. Genet.* 23:337–70
7. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, et al. 2010. A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLOS Genet.* 6:e1001193
8. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, et al. 2008. Variation in molybdenum content across broadly distributed populations of *Arabidopsis thaliana* is controlled by a mitochondrial molybdenum transporter (*MOT1*). *PLOS Genet.* 4:e1000004
9. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480:245–49
10. Bentsink L, Jowett J, Hanhart CJ, Koornneef M. 2006. Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *PNAS* 103:17042–47
11. Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLOS Genet.* 10:e1004412
12. Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, et al. 2006. *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* 312:1040–43
13. Bomblies K, Yant L, Laitinen RA, Kim ST, Hollister JD, et al. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLOS Genet.* 6:e1000890
14. Brachi B, Faure N, Bergelson J, Cuguen J, Roux F. 2013. Genome-wide association mapping of flowering time in in nature: genetics for underlying components and reaction norms across two successive years. *Acta Bot. Gallica* 160:205–19
15. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, et al. 2010. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLOS Genet.* 6:e1000940
16. Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, et al. 2013. Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Mol. Ecol.* 22:4222–40
17. Brady KU, Kruckeberg AR, Bradshaw HD Jr. 2005. Evolutionary ecology of plant adaptation to serpentine soils. *Annu. Rev. Ecol. Evol. Syst.* 36:243–66
18. Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G. 2013. Genomic identification of founding haplotypes reveals the history of the selfing species *Capsella rubella*. *PLOS Genet.* 9:e1003754
19. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31:1119–25
20. Caicedo AL, Schaal BA, Kunkel BN. 1999. Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *PNAS* 96:302–6
21. Calarco JP, Borges F, Donoghue MT, Van Ex F, Jullien PE, et al. 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151:194–205
22. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, et al. 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43:956–63
23. Carroll SB. 2005. Evolution at two levels: on genes and form. *PLOS Biol.* 3:e245
24. Chan EK, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ. 2011. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLOS Biol.* 9:e1001125

25. Chan EK, Rowe HC, Kliebenstein DJ. 2010. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185:991–1007
26. Chao DY, Baraniecka P, Danku J, Koprivova A, Lahner B, et al. 2014. Variation in sulfur and selenium accumulation is controlled by naturally occurring isoforms of the key sulfur assimilation enzyme ADENOSINE 5′-PHOSPHOSULFATE REDUCTASE2 across the *Arabidopsis* species range. *Plant Physiol.* 166:1593–608
27. Chao DY, Chen Y, Chen J, Shi S, Chen Z, et al. 2014. Genome-wide association mapping identifies a new arsenate reductase enzyme critical for limiting arsenic accumulation in plants. *PLOS Biol.* 12:e1002009
28. Chao DY, Silva A, Baxter I, Huang YS, Nordborg M, et al. 2012. Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in *Arabidopsis thaliana*. *PLOS Genet.* 8:e1002923
29. Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. *Am. Nat.* 140:126–48
30. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, et al. 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317:338–42
31. Clausen J, Keck DD, Hiesey WM. 1940. *Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants.* Washington, D.C.: Carnegie Inst. Wash. Publ. 452 pp.
32. Cortijo S, Wardenaar R, Colome-Tatche M, Gilly A, Etcheverry M, et al. 2014. Mapping the epigenetic basis of complex traits. *Science* 343:1145–48
33. Dansgaard W, Johnsen SJ, Clausen HB, Dahljensen D, Gundestrup NS, et al. 1993. Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* 364:218–20
34. Dickins TE, Rahman Q. 2012. The extended evolutionary synthesis and the role of soft inheritance in evolution. *Proc. Biol. Sci. R. Soc.* 279:2913–21
35. Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, et al. 2014. DNA methylation variation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife* 4:e05255
36. Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756–60
37. Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, et al. 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat. Genet.* 46:1089–96
38. Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, et al. 2013. Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol. Ecol.* 22:5594–607
39. Fisher RA. 1918. The correlation between relatives on the assumption of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52:399–433
40. Flor HH. 1971. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9:275–96
41. Fournier-Level A, Korte A, Cooper M, Nordborg M, Schmitt J, Wilczek A. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334:86–89
42. Frank SA. 1993. Coevolutionary genetics of plants and pathogens. *Evol. Ecol.* 7:45–75
43. Friesen ML, Cordeiro MA, Penmetsa RV, Badri M, Huguet T, et al. 2010. Population genomic analysis of Tunisian *Medicago truncatula* reveals candidates for local adaptation. *Plant J.* 63:623–35
44. Friesen ML, von Wettberg EJ, Badri M, Moriuchi KS, Barhoumi F, et al. 2014. The ecological genomic basis of salinity adaptation in Tunisian *Medicago truncatula*. *BMC Genomics* 15:1160
45. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, et al. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477:419–23
46. Gao L, Roux F, Bergelson J. 2009. Quantitative fitness effects of infection in a gene-for-gene system. *New Phytol.* 184:485–94
47. Geraldes A, Farzaneh N, Grassa CJ, McKown AD, Guy RD, et al. 2014. Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution* 68:3260–80
48. Gomaa NH, Montesinos-Navarro A, Alonso-Blanco C, Picó FX. 2011. Temporal variation in genetic diversity and effective population size of Mediterranean and subalpine *Arabidopsis thaliana* populations. *Mol. Ecol.* 20:3540–54
49. Grant PR, Grant BR. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* 296:707–11

50. Hagmann J, Becker C, Müller J, Stegle O, Meyer RC, et al. 2015. Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLOS Genet.* 11:e1004920
51. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, et al. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334:83–86
52. Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, et al. 2011. Adaptations to climate-mediated selective pressures in humans. *PLOS Genet.* 7:e1001375
53. Harper JL. 1977. *Population Biology of Plants*. San Francisco: Academic. 892 pp.
54. Hartl DL, Taubes CH. 1996. Compensatory nearly neutral mutations: selection without adaptation. *J. Theor. Biol.* 182:303–9
55. Heard E, Martienssen RA. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157:95–109
56. Henderson CR. 1984. *Applications of Linear Models in Animal Breeding*. Guelph, Ont.: Univ. Guelph Press. 462 pp.
57. Hodges SA, Derieg NJ. 2009. Adaptive radiations: from field to genomic studies. *PNAS* 106(Suppl. 1):9947–54
58. Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016
59. 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:e1003093
60. Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19:1419–28
61. Horton M, Hancock AM, Huang YS, Toomajian C, Atwell S, et al. 2012. Genome-wide pattern of genetic variation in worldwide *Arabidopsis thaliana* accessions from the *RegMap* panel. *Nat. Genet.* 44:212–16
62. Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, et al. 2014. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat. Commun.* 5:5320
63. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43:476–81
64. Huang X, Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65:531–51
65. Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, et al. 2013. An atypical kinase under balancing selection confers broad-spectrum disease resistance in *Arabidopsis*. *PLOS Genet.* 9:e1003766
66. Huber CD, Nordborg M, Hermisson J, Hellmann I. 2014. Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana*. *Mol. Biol. Evol.* 31:3026–39
67. Jensen JD. 2014. On the unfounded enthusiasm for soft selective sweeps. *Nat. Commun.* 5:5281
68. Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. 2014. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Res.* 24:1821–29
69. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61
70. Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ. 2013. Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *eLife* 2:e00776
71. Kaplan N, Dekker J. 2013. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat. Biotechnol.* 31:1143–47
72. Karasov TL, Kniskern JM, Gao L, DeYoung BJ, Ding J, et al. 2014. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* 512:436–40
73. Kim KE, Peluso P, Babayan P, Yeadon PJ, Yu C, et al. 2014. Long-read, whole-genome shotgun sequence data for five model organisms. *Sci. Data* 1:140045
74. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, et al. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 39:1151–55
75. Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge Univ. Press. 384 pp.
76. Kliebenstein DJ. 2009. A quantitative genetics and ecological model system: understanding the aliphatic glucosinolate biosynthetic network via QTLs. *Phytochem. Rev.* 8:243–54

77. Kniskern JM, Barrett LG, Bergelson J. 2011. Maladaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae*. *Evolution* 65:818–30
78. Korves T. 2004. A novel cost of *R* gene resistance in the presence of disease. *Am. Nat.* 163:489–504
79. Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, et al. 2007. Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *Am. Nat.* 169:E141–57
80. Kreitman M. 2000. Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum. Genet.* 1:539–59
81. Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, et al. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol. Lett.* 15:378–92
82. Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T. 2003. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *PNAS* 100(Suppl. 2):14587–92
83. Lachaise D, Silvain JF. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster–D. simulans* palaeogeographic riddle. *Genetica* 120:17–39
84. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, et al. 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 30:771–76
85. Lamichhaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–75
86. Lamichhaney S, Martinez Barrio A, Rafati N, Sundstrom G, Rubin CJ, et al. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *PNAS* 109:19345–50
87. Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–26
88. Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. 2012. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol. Ecol.* 21:5512–29
89. Lee CR, Mitchell-Olds T. 2012. Environmental adaptation contributes to gene polymorphism across the *Arabidopsis thaliana* genome. *Mol. Biol. Evol.* 29:3721–28
90. Lee SH, DeCandia TR, Ripke S, Yang J, Schizophr. Psychiatr. Genome-Wide Assoc. Study Consort., et al. 2012. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* 44:247–50
91. Lee YW, Gould BA, Stinchcombe JR. 2014. Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* 6:pii:plu004
92. Leinonen PH, Sandring S, Quilot B, Clauss MJ, Mitchell-Olds T, et al. 2009. Local adaptation in European populations of *Arabidopsis lyrata* (Brassicaceae). *Am. J. Bot.* 96:1129–37
93. Leitch IJ, Chase MW, Bennett MD. 1998. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann. Bot.* 82:85–94
94. Li P, Filiault D, Box MS, Kerdaffrec E, van Oosterhout C, et al. 2014. Multiple FLC haplotypes defined by independent *cis*-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes Dev.* 28:1635–40
95. Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO. 2014. Genetic variation for life history sensitivity to seasonal warming in *Arabidopsis thaliana*. *Genetics* 196:569–77
96. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. 2010. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *PNAS* 107:21199–204
97. Lin T, Zhu G, Zhang J, Xu X, Yu Q, et al. 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46:1220–26
98. Liu S, Ying K, Yeh CT, Yang J, Swanson-Wagner R, et al. 2012. Changes in genome content generated via segregation of non-allelic homologs. *Plant J.* 72:390–99
99. Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, et al. 2013. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45:884–90
100. Mandel JR, Nambeesan S, Bowers JE, Marek LF, Ebert D, et al. 2013. Association mapping and the genomic consequences of selection in sunflower. *PLOS Genet.* 9:e1003378
101. Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18:189–97
102. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53

103. Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15:394–408
104. McKown AD, Klapste J, Guy RD, Geraldes A, Porth I, et al. 2014. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytol.* 203:535–53
105. Méndez-Vigo B, Gomaa NH, Alonso-Blanco C, Picó FX. 2013. Among- and within-population variation in flowering time of Iberian *Arabidopsis thaliana* estimated in field and glasshouse conditions. *New Phytol.* 197:1332–43
106. Mikola J. 1982. Bud-set phenology as an indicator of climatic adaptation of Scots pine in Finland. *Silva Fenn.* 16:178–84
107. Montesinos A, Tonsor SJ, Alonso-Blanco C, Picó FX. 2009. Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. *PLOS ONE* 4:e7213
108. Montesinos-Navarro A, Wig J, Picó FX, Tonsor SJ. 2011. *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytol.* 189:282–94
109. Morgenstern EK. 1996. *Geographic Variation in Forest Trees. Genetic Basis and Application of Knowledge in Silviculture.* Vancouver: UBC Press. 209 pp.
110. Nemri A, Atwell S, Tarone AM, Huang YS, Zhao K, et al. 2010. Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping. *PNAS* 107:10302–7
111. Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218
112. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLOS Biol.* 3:e196
113. Nordborg M, Weigel D. 2008. Next-generation genetics in plants. *Nature* 456:720–23
114. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. 2008. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* 18:2024–33
115. Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, et al. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327:92–94
116. Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* 131:452–62
117. Pagny G, Paulstephenraj PS, Poque S, Sicard O, Cosson P, et al. 2012. Family-based linkage and association mapping reveals novel genes affecting *Plum pox virus* infection in *Arabidopsis thaliana*. *New Phytol.* 196:873–86
118. Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* 21:2991–3005
119. Poland JA, Brown PJ, Sorrells ME, Jannink JL. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE* 7:e32253
120. Poormohammad Kiani S, Trontin C, Andreatta M, Simon M, Robert T, et al. 2012. Allelic heterogeneity and trade-off shape natural variation for response to soil micronutrient. *PLOS Genet.* 8:e1002814
121. Puzey J, Vallejo-Marin M. 2014. Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Mol. Ecol.* 23:4472–85
122. Qi J, Chen Y, Copenhaver GP, Ma H. 2014. Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *PNAS* 111:10007–12
123. Ralph P, Coop G. 2010. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* 186:647–68
124. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, et al. 2011. *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333:1137–41
125. Robledo-Arnuncio JJ. 2011. Wind pollination over mesoscale distances: an investigation with Scots pine. *New Phytol.* 190:222–33
126. Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66:1–17
127. Roux F, Colomé-Tatché M, Edelist C, Wardenaar R, Guerche P, et al. 2011. Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics* 188:1015–17

128. Salt DE, Baxter I, Lahner B. 2008. Ionomics and the study of the plant ionome. *Annu. Rev. Plant Biol.* 59:709–33
129. Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* 38:595–619
130. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, et al. 2011. Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334:369–73
131. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, et al. 2013. Patterns of population epigenomic diversity. *Nature* 495:193–98
132. Schneeberger K, Ossowski S, Ott F, Klein JD, Wang X, et al. 2011. Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *PNAS* 108:10249–54
133. Shen X, De Jonge J, Forsberg SK, Pettersson ME, Sheng Z, et al. 2014. Natural *CMT2* variation is associated with genome-wide methylation changes and temperature seasonality. *PLOS Genet.* 10:e1004842
134. Silady RA, Effgen S, Koornneef M, Reymond M. 2011. Variation in seed dormancy quantitative trait loci in *Arabidopsis thaliana* originating from one site. *PLOS ONE* 6:e20886
135. Silveira AB, Trontin C, Cortijo S, Barau J, Del Bem LE, et al. 2013. Extensive natural epigenetic variation at a de novo originated gene. *PLOS Genet.* 9:e1003437
136. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344:738–42
137. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400:667–71
138. Stern DL, Orgogozo V. 2008. The loci of evolution: How predictable is genetic evolution? *Evolution* 62:2155–77
139. Tang C, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, et al. 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science* 317:1070–72
140. Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* 16:702–12
141. Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. 2002. Signature of balancing selection in *Arabidopsis*. *PNAS* 99:11525–30
142. Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J. 2003. Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423:74–77
143. Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, et al. 2010. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465:632–36
144. Tschaplinski TJ, Plett JM, Engle NL, Deveau A, Cushman KC, et al. 2014. *Populus trichocarpa* and *Populus deltoides* exhibit different metabolomic responses to colonization by the symbiotic fungus *Laccaria bicolor*. *Mol. Plant-Microbe Interact.* 27:546–56
145. Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, et al. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44:1015–19
146. Turesson G. 1922. The genotypical response of the plant species to the habitat. *Hereditas* 3:211–350
147. Turesson G. 1922. The species and the variety as ecological units. *Hereditas* 3:100–13
148. Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42:260–63
149. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–604
150. Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, et al. 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLOS Biol.* 5:e174
151. Verslues PE, Lasky JR, Juenger TE, Liu TW, Kumar MN. 2014. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis*. *Plant Physiol.* 164:144–59
152. Vilhjálmsson BJ, Nordborg M. 2013. The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* 14:1–2
153. Vitti JJ, Grossman SR, Sabeti PC. 2013. Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47:97–120

154. Voytas DF. 2013. Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64:327–50
155. Waddington CH. 1961. Genetic assimilation. *Adv. Genet.* 10:257–93
156. Weigel D. 2012. Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* 158:2–22
157. Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez-Gallego C, et al. 2009. Effects of genetic perturbation on seasonal life history plasticity. *Science* 323:930–34
158. Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159
159. Wu CA, Lowry DB, Cooley AM, Wright KM, Lee YW, Willis JH. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100:220–30
160. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–69
161. Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, et al. 2013. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* 23:2151–56
162. Yoshida K, Burbano HA, Krause J, Thines M, Weigel D, Kamoun S. 2014. Mining herbaria for plant pathogen genomes: back to the future. *PLOS Pathog.* 10:e1004028
163. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203–8
164. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, et al. 2007. An *Arabidopsis* example of association mapping in structured samples. *PLOS Genet.* 3:e4
165. Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA. 2012. Natural enemies drive geographic variation in plant defenses. *Science* 338:116–19

TABLE OF CONTENTS FOR VOLUME 1:

Annual Review of Genetics

Volume 49, 2015

Contents

An online log of corrections to *Annual Review of Genetics* articles may be found at http://www.annualreviews.org/errata/genet