

Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*

Thomas Mitchell-Olds¹ & Johanna Schmitt²

Genomic studies of natural variation in model organisms provide a bridge between molecular analyses of gene function and evolutionary investigations of adaptation and natural selection. In the model plant species *Arabidopsis thaliana*, recent studies of natural variation have led to the identification of genes underlying ecologically important complex traits, and provided new insights about the processes of genome evolution, geographic population structure, and the selective mechanisms shaping complex trait variation in natural populations. These advances illustrate the potential for a new synthesis to elucidate mechanisms for the adaptive evolution of complex traits from nucleotide sequences to real-world environments.

The beginning of the twenty-first century is an exciting time for biologists. Rapid advances in genomics have changed our view of the biological world and fostered new links between molecular biology, ecology and evolution. Genomic studies of natural variation in model organisms are a crucial ingredient in this new synthesis. Molecular biologists have begun to exploit natural variation to identify the genetic mechanisms underlying complex traits¹. Simultaneously, these new genomic tools make it possible for evolutionary biologists to study how ecologically important complex traits evolve in natural environments. These advances now make it possible to understand the adaptive evolution of complex trait variation from molecular mechanisms to geographic patterns of population structure and natural selection.

The diminutive weed *Arabidopsis thaliana* provides an ideal system for such interdisciplinary synthesis. This species—a close relative of *Brassica* crops such as mustard and broccoli—is a convenient genetic model because of its short generation time and small genome. The *A. thaliana* genome was the first plant genome to be sequenced, and the genes and developmental pathways controlling ecologically important traits such as germination, flowering time, pest resistance, and stress tolerance are rapidly being elucidated. It is therefore possible to identify 'candidate' genes for adaptive variation in natural populations. *A. thaliana* is a widespread annual weed of rocky places and disturbed sites, native to Europe and central Asia and naturalized in North America (Fig. 1). Across this geographic range, it experiences a broad range of climatic conditions² and selective pressures. Inbred stocks are available for many natural *A. thaliana* accessions ('ecotypes'), originating across the species' range. Because the species is habitually inbreeding, genomic and phenotypic data can be combined from multiple experiments with the same genotypes. These genomic tools and resources have enabled a number of important advances in molecular and evolutionary genetics. Here we focus on advances in three complementary areas: (1) genomic studies of molecular variation and population structure; (2) identification of genetic polymorphisms underlying natural variation in complex traits; and, (3) ecological and evolutionary studies of

natural selection and adaptation. Taken together, these advances now make it possible to identify the genetic mechanisms underlying the adaptive evolution of complex traits in natural populations.

Molecular variation and population structure

How much molecular polymorphism exists in *A. thaliana*? Recent genome-wide studies show that an average pair of alleles differs at about seven nucleotides per kilobase (kb; nucleotide diversity = 0.007; refs 3, 4). This is about 50% lower than polymorphism in the outcrossing congener *A. lyrata* ssp. *petraea*⁵, roughly the same as *Drosophila melanogaster*⁶, and nearly an order of magnitude higher than humans⁷. *A. thaliana* has a high frequency of self-pollination in the wild⁸, hence individuals are homozygous at most loci³. Such high rates of self-pollination may influence patterns of linkage disequilibrium, which provides the basis for association studies and linkage disequilibrium mapping in human genetics and plant breeding (see Box 1).

Patterns of nucleotide polymorphism contain a signature of historical demography and natural selection (see Box 2). Genome-wide information for *A. thaliana* has enabled fundamental advances in our understanding of the evolutionary processes that influence these patterns. In many species, a positive correlation exists between local recombination rates and levels of nucleotide diversity³. In regions of low recombination, genetic variation may be reduced owing to 'hitchhiking' of neutral variation with nearby selected sites, which will influence wider chromosomal regions when recombination is low. However, this reduction in variation could be due either to selective sweeps of advantageous mutations, or to background selection, which eliminates deleterious mutations and the haplotypes that carry them. Genome-wide polymorphism data make it possible to distinguish these possibilities. The recent observation³ that nucleotide polymorphism is negatively correlated with gene density supports background selection as the predominant mechanism. Gene-dense regions show little sign of rare variants attributable to recent selective sweeps, but the frequencies of non-synonymous polymorphisms indicate purifying selection against deleterious mutations⁹.

¹Department of Biology, PO Box 91000, Duke University, Durham, North Carolina 27708, USA. ²Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, Rhode Island 02912, USA.