

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.

Nevertheless, surveys of *A. thaliana* ecotypes have identified loci that may have undergone recent positive selection or selective sweeps (Box 2)¹⁰. Certain other loci show unusually high levels of amino-acid variation or too many intermediate-frequency nucleotide polymorphisms, which exceed neutral expectation and may have been maintained by natural selection^{11–13}. The selective mechanisms maintaining such polymorphisms remain an important open question. Allelic polymorphisms may be maintained within populations by frequency-dependent selection (where common genotypes have low fitness, and rare types are favoured), by temporal variation in selection across seasons and years, or by epistatic selection in which the fitness of an allele depends on genetic background¹⁴. Alternatively, different alleles might be maintained in different populations by local adaptation to geographically divergent selection, resulting in excess polymorphism species-wide, but little variation within individual populations. Wide surveys of *A. thaliana* ecotypes cannot distinguish between these different ecological mechanisms, because information is also needed about patterns of variation within undisturbed natural populations. However, human disturbance and admixture of different source populations may complicate such studies in most parts of the *A. thaliana* range. It is therefore important to understand the geographic population structure of natural variation within the species, as well as the distribution of molecular variation within and among populations across the species range.

Genome-wide polymorphism data for ecotypes across the range of *A. thaliana* have made it possible to investigate the geographic structure of molecular variation. The genetic divergence and geographic distance between pairs of populations is positively correlated across the native range^{3,15}. Such ‘isolation by distance’ reflects long-term geographical isolation with limited gene flow. The pattern of population ancestry of European ecotypes suggests that much of the native range was colonized from several glacial refugia, with admixture in the zones colonized from more than one source. However, recent human disturbance tends to homogenize variation among populations, especially in agricultural regions of Europe and introduced populations in North America^{15–17}.

How is molecular polymorphism distributed within and among natural populations of *A. thaliana*? Recent studies have estimated the

proportion of total molecular polymorphism that is found within populations to be near 45% in western Europe and North America^{16–18}, but only 12% in less-disturbed parts of Norway¹⁹. Although part of this difference may be attributable to loss of variability as *A. thaliana* migrated north from circum-Mediterranean refugia, comparisons across a 900-km north–south transect in Norway find no latitudinal changes in genetic variation. Therefore, low variation within Norwegian populations is probably not attributable to post-Pleistocene founder events. Instead, high variability within western European populations probably reflects admixture resulting from human disturbance¹⁸. Such admixture complicates our ability to understand the evolutionary forces shaping genetic variation within *A. thaliana* populations.

Finding the genes underlying trait variation

Natural selection operates on complex trait phenotypes. In order to understand the evolution of ecologically important variation, we

Box 1 | Association studies

Association studies use linkage disequilibrium to identify polymorphisms that may be responsible for complex trait variation. Linkage disequilibrium quantifies statistical correlations between polymorphic sites: when linkage disequilibrium is present then information from one locus provides some predictive information about the genotype at a second locus. Association studies take advantage of recombinations accumulated over thousands of generations, and hence may aid identification of individual genes responsible for QTL. In the past (top panel), a new mutation (triangle) occurs at a quantitative trait locus. The original population contains an ancestral chromosome on which the new mutation occurred (black line), as well as other chromosomes (grey lines). After many generations of recombination, the present-day population (lower panel) contains short chromosome regions derived from the original population. Molecular markers near the causal polymorphism (*b*) will be correlated with phenotype, whereas distant markers (*a*, *c*) are uncorrelated because they have been reshuffled by recombination. On average, in *A. thaliana* polymorphisms separated by >50 kb are usually statistically independent, whereas linkage disequilibrium is substantial between sites separated by <50 kb (ref. 71). Because these regions in disequilibrium typically contain in the order of 10 genes, linkage disequilibrium mapping and association studies with candidate genes bring us close to the level of individual loci. These approaches may be useful when pedigrees are small, as in human populations, or when genotyped populations can be used by many researchers for analysis of multiple phenotypes, as in *A. thaliana*. Several challenging statistical issues remain for association studies in *A. thaliana*. Population and pedigree structure can be incorporated into analytical models⁷² or homogeneous populations may be identified. However, if ancestral populations have diverged for the trait of interest, then statistical adjustment for population structure may remove the very effects of interest⁷³. Linkage disequilibrium mapping and association studies also result in false negatives and false positives⁷⁴. Consequently, results from association studies represent hypotheses that must be tested by independent experiments⁴³.

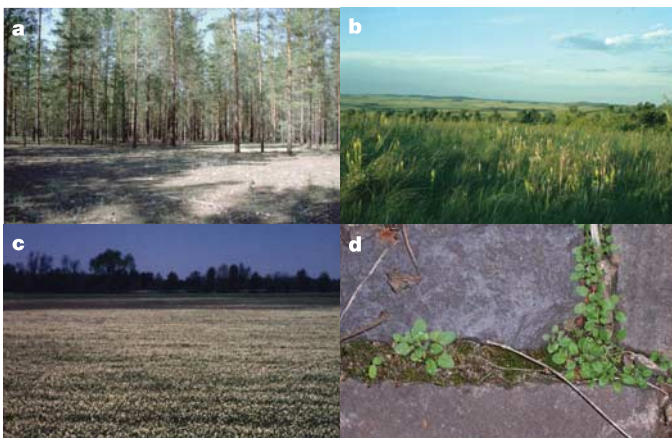
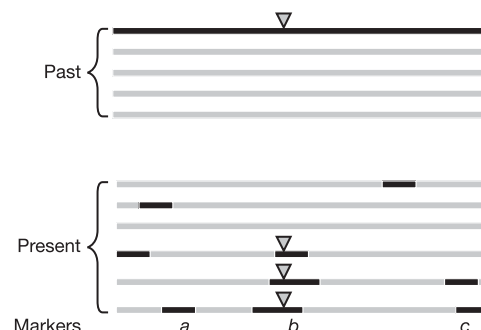


Figure 1 | *Arabidopsis* habitats. **a, b**, Although commonly found in disturbed sites, *A. thaliana* also grows in pine forests on sandy soils (**a**; Bastan, western Siberia) and perennial grasslands in sites with sparse vegetation (**b**; Stepnoje, western Siberia). **c**, Introduced populations in North America may number in the millions. Here, the white flowers are another species, but the interspersed vegetation is predominantly *A. thaliana* (Michigan, USA). **d**, Even in western Europe it can be a long-term resident of stone walls and other sites (Cologne, Germany). (Russia photographs by M. Hoffmann; USA photograph by J. Bergelson; Germany photograph by A. Wilczek.)

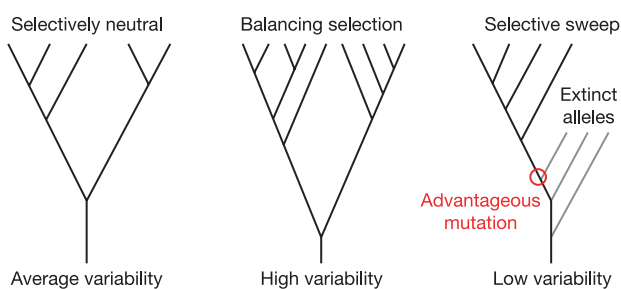


must identify nucleotide polymorphisms with functional effects on phenotypic differences. Such polymorphisms are also of great interest to plant molecular biologists seeking to elucidate gene function and genetic pathway structure²⁰. Identification of natural variants is particularly important for functional studies because some knockout mutations may be lethal, or may lack detectable phenotypic effects due to genetic redundancy²¹. These efforts will be greatly aided by databases on genome-wide patterns of nucleotide polymorphism³, gene expression²², and developmental phenotypes²³.

The identification of natural polymorphisms controlling variation in complex traits begins with screens of trait variation among accessions. Once such variation is identified, the next step is to identify polymorphic genomic regions (quantitative trait loci, or QTL) associated with that variation. Recent advances in characterizing genome-wide polymorphism and population structure for a wide range of *A. thaliana* ecotypes^{3,15} may facilitate identification of QTL through linkage disequilibrium mapping (Box 1). Recombinant inbred lines (RILs) between divergent parental ecotypes have already

Box 2 | Sequence signatures

Until recently, cloning of QTL has been confined to laboratory model systems, hence it has been impossible to measure the fitness effects of QTL alleles in natural environments. As an alternative approach, adaptive significance has been inferred from the sequence signature of nucleotide polymorphism⁵. Population genetics theory can predict patterns of nucleotide polymorphism on the basis of standard neutral models for ideal, equilibrium populations. Statistical hypothesis tests can then compare predictions from these standard neutral models to observed variation at genes of interest. For example, the figure shows 'gene trees' portraying the historical relationship among alleles at a given locus. A typical selectively neutral locus is shown on the left, with average levels of nucleotide polymorphism. In contrast, balancing selection (centre) can maintain high levels of variability for long periods of time. Such polymorphisms are unusually old, and have accumulated high levels of molecular variation. Finally, an advantageous mutation may out-compete existing alleles in a 'selective sweep' (right), reducing extant variation below neutral levels. Although these theoretical predictions apply to ideal populations, non-standard or non-equilibrium demographic conditions such as gene flow or changing size⁵ also influence allelic variation in real-world populations. With data from only a single locus, statistical tests often cannot determine whether differences between observed and predicted patterns are due to natural selection, or instead reflect neutral effects of demographic history. Fortunately, genome-wide patterns of variation contain a signature of historical demographic processes, because demography influences variation across the genome. In contrast, the effects of natural selection are confined to the target locus and adjacent regions. Consequently, putative natural selection at genes of interest may be identified by comparison to a genome-wide 'empirical null distribution' summarizing variability at many loci, or by simulations based on plausible non-standard models. Recently, several studies^{3,4} have shown that variation in *A. thaliana* does not conform to standard demographic models of ideal populations. Instead, gene flow among populations, changing population size, extinction, and recolonization evidently have influenced polymorphism in *A. thaliana*.



proved extremely valuable for mapping QTL for a variety of ecologically important complex traits²⁰. QTL of moderately large effect have been identified in a number of studies, but chromosomal regions with large phenotypic effects may actually contain multiple linked QTL of smaller effect²⁴. A substantial portion of genetic variation for complex traits may actually correspond to polygenic models of complex trait variation, with many genes of small effect^{25,26}.

Recent advances in genomics make it possible to treat genome-wide expression patterns as complex traits and screen for natural variation in those patterns. Schmid *et al.*²² studied expression of 22,000 genes in many developmental stages and tissues, providing invaluable data on regulatory patterns in *Arabidopsis* and enabling further functional²⁷ and evolutionary analyses. Patterns of gene expression differ among *Arabidopsis* accessions^{28,29}, and QTL influencing *cis*- or *trans*-regulatory polymorphisms can be identified³⁰. *Cis*-acting allele-specific transcriptional differences are apparently common in *Arabidopsis*^{30,31} and other organisms³². Progress in this field requires experiments that relate transcriptional polymorphisms to phenotypic variation at the whole-plant level.

One important lesson from QTL studies with *A. thaliana* is that the expression of phenotypic differences between QTL alleles may depend strongly on environmental conditions and genetic background. QTL-environment interactions are common; often, certain QTL are expressed in some environments but not in others, or differ in the magnitude of their allelic effects. Although such observations are common in *A. thaliana*^{33,34}, their molecular basis and relationship to genotype-environment interactions at the whole-organism level have rarely been elucidated. To date, antagonistic pleiotropy—in which the alleles of a QTL change phenotypic rank across environments³⁵—seems to be rare in *A. thaliana*. Clearly, more work is needed to understand the genetic basis of genotype-environment interaction.

Epistasis, or interaction effects among different loci, is a fundamental force in many aspects of adaptive evolution³⁶. At its simplest, epistasis occurs when the genotype at one locus influences the magnitude of allelic effects at another locus. More intriguingly, with sign epistasis the genotype at one locus may reverse the direction of allelic effects at another gene. There is now strong evidence that both kinds of epistasis are important in *A. thaliana*^{34,37-40}. In some cases, the mechanisms of epistasis are known from functional studies^{41,42}. At the level of individual genes, sign epistasis influencing juvenile growth rate (an important component of fitness) is caused by genetic polymorphism at a serine/threonine protein kinase, and patterns of nucleotide polymorphism suggest that balancing selection maintains elevated levels of genetic variation²⁵. These observations suggest that epistasis may have a fundamental role in the evolution of *A. thaliana* and other inbreeding species.

Once QTL regions have been identified, the next step is to find the underlying genes and nucleotide polymorphisms causally associated with trait variation. Rigorous standards of proof for determination of the molecular basis of a QTL have been established⁴³, and several recent studies have used methods such as fine mapping and transgenic complementation to identify causal genes^{25,44,45}. However, until homologous recombination techniques for allelic replacement become available for *A. thaliana*⁴⁶, the variation introduced by position effects may limit the power to detect minor allelic effects. This is well illustrated by a study of the fitness effects of inserting a transgene for *Rpm1*-mediated pathogen resistance⁴⁷, where comparisons among five independent insertion lines found that fitness differed by 37% among lines that varied only in the genomic location of the transgene.

What sorts of molecular polymorphisms control quantitative trait variation? Examples from *A. thaliana* span the full range of gene function, including photoreceptor protein polymorphisms⁴⁴, transcription factors⁴⁸, *cis*-regulatory polymorphisms³⁰, insertion/deletion polymorphisms⁴², and copy-number variation in tandem

gene families¹². However, current examples are insufficient to draw generalizations about the importance of regulatory versus coding polymorphisms, especially because the QTL that have been cloned so far have larger than average effects, and may not be representative of small-effect QTL.

Natural selection for complex traits

How does natural selection act on trait variation in the real world? The native range of *A. thaliana* spans a broad range of climates and habitats. *A. thaliana* has evolutionarily expanded its range to warmer climates from the cool temperate ‘climate space’ inhabited by other *Arabidopsis* species⁴⁹. This climatic range expansion may result from *A. thaliana*’s evolutionary transition to an annual life history from the perennial strategy shown by its close relatives. The annual strategy may increase tolerance of warmer climates by allowing plants to survive summer conditions as seeds rather than as juvenile or adult plants. However, the annual strategy necessitates germinating and flowering during favourable growing conditions, and the seasonal timing of these events varies across the climatic range of the species. Natural populations of *A. thaliana* are thus likely to experience very different selective pressures across the species range. If geographic variation in climate has resulted in local adaptation, we should expect genetic associations between the phenotype of an accession and the climate in its site of origin.

How much natural variation exists for ecologically important traits in *A. thaliana*? Experiments with geographically diverse ecotypes under common garden conditions reveal remarkable genetic variation in many ecologically important traits^{20,23,48,50,51}. Much less is known about the phenology and demography of natural *A. thaliana* populations *in situ*, but it is clear that the species shows variation in life history timing across its range. In particular, there is a life history polymorphism between winter annual ecotypes—which overwinter as small vegetative rosettes and flower in spring—and rapid cycling ecotypes—which germinate and flower in late summer and again in spring to produce two generations per year (ref. 52). This polymorphism has been attributed to loss of a vernalization requirement conferred by mutation at the major flowering time loci, *FRIGIDA* (*FRI*)^{42,53} and *FLOWERING LOCUS C* (*FLC*)⁵⁴, with additional variation due to other loci^{23,48,51}.

Does the observed genetic variation among natural populations represent adaptive differentiation in response to heterogeneous natural selection? This hypothesis can be tested with several different lines of evidence (Table 1). First, is trait variation associated with geographical variation in climate or other ecological factors? Second, is the degree of quantitative variation among populations greater than expected on the basis of differentiation at neutral markers? Third, do local genotypes have higher fitness than foreign genotypes in reciprocal transplant experiments, and if so, is this ‘home-court advantage’ attributable to selection favouring the local phenotype of the candidate trait? The first two questions have recently been addressed by quantitative genetic studies in common garden environments, and provide some support for the adaptive differentiation hypothesis. However, the third question will require evidence from reciprocal transplant experiments and field studies of selection.

Several studies have detected latitudinal clines in *A. thaliana* growth and morphology, seedling traits, flowering time, vernalization sensitivity, and circadian period^{23,55–58}. Such gradients in complex traits may provide evidence of local adaptation, especially if the direction of the cline is consistent with functional arguments⁵⁸. However, clinal variation may also result from non-adaptive phenomena such as isolation by distance or admixture between two divergent founder populations. Therefore, geographic tests of trait adaptation are most compelling when combined with other forms of evidence.

If population differentiation in a trait is adaptive, the degree of quantitative genetic differentiation in the trait among populations should be greater than the genetic differentiation among populations in neutral molecular markers¹⁸. (Typically, the proportion of genetic variation found among populations for phenotypic traits and neutral molecular markers is quantified by Q_{st} and F_{st} , respectively.) Using this approach, Le Corre¹⁸ measured trait variation within and among natural populations in France. The Q_{st} for bolting time was significantly higher than the overall F_{st} estimated from neutral molecular markers, suggesting adaptive divergence among populations in reproductive timing. Moreover, the F_{st} for functional polymorphism at the principal flowering time gene *FRI* was also significantly higher than F_{st} at marker loci, suggesting adaptive differentiation at that locus. Because *FRI* confers a vernalization requirement, and is known to be associated with flowering time in non-vernalized plants, these

Table 1 | Evidence for adaptation

Method	Known gene?	Advantages	Disadvantages	Reference	Time frame
Transgenic comparison of candidate allele effects	Yes	The gold standard for functional verification and fitness consequences	Transgene position effects Regulatory complications with field trials	47	Current
QTL, RILs, NILs	No	Does not require known gene Modest regulatory requirements	Effects may be confounded with flanking genes	40	Current
Sequence signature	Yes	Capable of detecting very weak selection Does not require phenotypic information	Must control for demographic history Phenotype experiencing selection is unknown	5	Historical
F_{st}/Q_{st}	No	F_{st} can be calculated from relatively few neutral loci Does not require genome-wide molecular markers	Difficult to detect local adaptation when F_{st} is large	18	Historical
Hitchhiking mapping	No	May localize chromosomal region experiencing local adaptation	Phenotype experiencing selection is unknown	75	Historical
Genotype–environment association	No	Can be applied when causal genes are unknown	May be confounded with gene flow or admixture	58	Historical
Multivariate selection analyses	No	Can be applied when causal genes are unknown	May be confounded with unmeasured traits	61	Current
Reciprocal transplants	No	With proper replication, may detect local adaptation Can be applied when causal genes are unknown	Requires access to original populations and environments Further experiments needed to establish mechanism	59	Historical and current

Abbreviations: QTL, quantitative trait loci; RILs, recombinant inbred lines; NILs, near-isogenic lines; F_{st} , the proportion of molecular marker polymorphism found among populations; Q_{st} , the proportion of quantitative genetic variation found among populations.

results suggest that the adaptive divergence of flowering time occurred through selection for loss of *FRI* function in certain populations.

Common garden studies may provide evidence for adaptive population differentiation in complex traits, and suggest hypotheses about selective mechanisms for such differentiation. However, direct support for the local adaptation hypothesis requires reciprocal transplants between natural populations⁵⁹, and selective mechanisms are best tested by measuring natural selection on traits of interest in such experiments. Although this approach has proved to be very powerful in other plant species, it has rarely been attempted for *A. thaliana*, and evidence for local adaptation has been equivocal⁶⁰. A reciprocal-transplant approach could be very valuable for testing hypotheses about local adaptation to climate across the native range of *A. thaliana*.

Although direct experimental tests of local adaptation are still lacking for *A. thaliana*, several field experiments provide strong evidence for real-time natural selection on complex traits and specific loci. Multivariate selection analysis⁶¹ quantifies how natural selection acts on variation in complex traits. Such studies are most valuable when combined with ecological manipulations to test hypotheses about the causes of selection⁶². For example, Donohue *et al.*⁶³ demonstrated geographical differences in selection on germination timing for seeds of the same genotypes experimentally dispersed in Rhode Island and Kentucky, as well as seasonal differences in selection between seeds dispersed in spring and autumn.

How does natural selection on complex traits act on specific loci? Recent field experiments with RILs have identified QTL influencing fitness in different environments^{40,64,65}. The same set of lines showed largely different QTL for fitness in Georgia⁴⁰, North Carolina, and Rhode Island⁶⁵, suggesting that natural selection acts on different genomic regions in different geographic locations and seasonal environments. These studies also revealed pervasive epistatic selection. For example, Malmberg and colleagues⁴⁰ found epistatic QTL to be more common and more important than non-epistatic QTL, and found a high frequency of sign epistasis for fecundity. Such epistatic interactions can maintain genetic variation and constrain evolutionary responses to natural selection³⁶, and may provide a possible mechanism for the maintenance of polymorphism in *A. thaliana*.

These studies could not test hypotheses about local adaptation, as the parents of the RILs were not native to the field sites. Rather, they provide insights into how natural selection will act on the segregating variation in different novel environments, a potentially relevant scenario for this colonizing species. Such experiments also can test whether fitness QTL co-localize with QTL for complex traits experiencing natural selection, thus providing insights into selective mechanisms⁶⁴. The next step is to take RILs from locally adapted parental ecotypes and conduct selection experiments with the RILs in the parental sites of origin. Such experiments can elucidate the selective mechanisms and specific loci contributing to local adaptation⁶⁶.

Near-isogenic lines (NILs) containing alternate alleles of polymorphic candidate genes provide another promising tool for testing hypotheses about how selection acts on polymorphisms in ecological time. Such studies will be particularly useful for polymorphisms that show the molecular signature of balancing selection in genes of known ecological function. For example, *R*-genes in *A. thaliana* confer resistance to pathogens, and are often polymorphic for resistant or susceptible alleles, with evidence for balancing selection at the molecular level^{13,67,68}. Experiments with NILs⁶⁹ and pairs of transgenics that differed in the presence or absence of a resistant allele⁴⁷ were able to demonstrate fitness costs of resistance alleles, which provide an ecological mechanism for the maintenance of these polymorphisms. This approach may be valuable for future studies of the selective mechanisms acting to maintain variation at other candidate loci.

Future prospects

Future advances in our understanding of *A. thaliana* natural variation depend, in part, on improvements in resources and technology.

Among the highest priorities are expanded collections from well-characterized environments, as well as possible undisturbed populations from Eurasia. On the technical side, the greatest need is for straightforward protocols to reduce transgene position effects, such as gene targeting⁴⁶ or alternative approaches^{47,70}.

Arabidopsis research is poised to address fundamental questions in evolution and ecology, including: What is the role of epistasis, and how does it influence the evolution of developmental pathways? How do ecological processes influence balancing selection within and among populations? What are the frequencies and effects of QTL alleles, and how is this influenced by evolutionary forces? What are the ecological and molecular mechanisms of local adaptation within the native range of *A. thaliana*? How do real-time ecological processes influence adaptive evolution in introduced and invasive populations? As the field matures, research will focus increasingly on related species, which will greatly improve our understanding of the genetic basis and evolutionary mechanisms of speciation.

Equally as important, *Arabidopsis* biology encourages (and requires) interdisciplinary research and training. A new synthesis of functional genomics with evolution and ecology will benefit each component discipline, and will bring fundamental advances in our understanding of biological mechanisms and processes. Such training provides an essential component of science education for the future.

1. Alonso-Blanco, C., Mendez-Vigo, B. & Koornneef, M. From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int. J. Dev. Biol.* **49**, 717–732 (2005).
2. Hoffmann, M. H. Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *J. Biogeogr.* **29**, 125–134 (2002).
3. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
4. Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. & Mitchell-Olds, T. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169**, 1601–1615 (2005).
5. Wright, S. I. & Gaut, B. S. Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* **22**, 506–519 (2005).
6. Ometto, L., Glinka, S., De Lorenzo, D. & Stephan, W. Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol. Biol. Evol.* **22**, 2119–2130 (2005).
7. Akey, J. M. *et al.* Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* **2**, e286 (2004).
8. Abbott, R. J. & Gomes, M. F. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**, 411–418 (1989).
9. Bustamante, C. D. *et al.* The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531–534 (2002).
10. Shimizu, K. K. *et al.* Darwinian selection on a selfing locus. *Science* **306**, 2081–2084 (2004).
11. Cork, J. M. & Purugganan, M. D. High-diversity genes in the *Arabidopsis* genome. *Genetics* **170**, 1897–1911 (2005).
12. Kroymann, J., Donnerhacke, S., Schnabelrauch, D. & Mitchell-Olds, T. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl Acad. Sci. USA* **100**, 14587–14592 (2003).
13. Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11525–11530 (2002).
14. Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G × E interactions. *Genetics* **166**, 1053–1079 (2004).
15. Schmid, K. *et al.* Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor. Appl. Genet.* **112**, 1104–1114 (2006).
16. Jorgensen, S. & Mauricio, R. Neutral genetic variation among wild North American populations of the weedy plant *Arabidopsis thaliana* is not geographically structured. *Mol. Ecol.* **13**, 3403–3413 (2004).
17. Bakker, E. G. *et al.* Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol. Ecol.* **15**, 1405–1418 (2006).
18. Le Corre, V. Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol. Ecol.* **14**, 4181–4192 (2005).
19. Stenoien, H. K., Fenster, C. B., Tonteri, A. & Savolainen, O. Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. *Mol. Ecol.* **14**, 137–148 (2005).
20. Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172 (2004).

21. Bouche, N. & Bouchez, D. *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* **4**, 111–117 (2001).
22. Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nature Genet.* **37**, 501–506 (2005).
23. Lempe, J. *et al.* Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genetics* **1**, e6 (2005).
24. Mackay, T. F. C. The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr. Opin. Genet. Dev.* **14**, 253–257 (2004).
25. Kroymann, J. & Mitchell-Olds, T. Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**, 95–98 (2005).
26. Fishman, L., Kelly, A. & Willis, J. Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution* **56**, 2138–2155 (2002).
27. Gachon, C. M. M., Langlois-Meurinne, M., Henry, Y. & Saindrenan, P. Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol. Biol.* **58**, 229–245 (2005).
28. Kliebenstein, D. J. *et al.* Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics* **172**, 1179–1189 (2006).
29. Vuylsteke, M., van Eeuwijk, F., Van Hummelen, P., Kuiper, M. & Zabeau, M. Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. *Genetics* **171**, 1267–1275 (2005).
30. DeCook, R., Lall, S., Nettleton, D. & Howell, S. H. Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* **172**, 1155–1164 (2006).
31. de Meaux, J., Goebel, U., Pop, A. & Mitchell-Olds, T. Allele-specific assay reveals functional variation in the *Chalcone Synthase* promoter of *Arabidopsis thaliana* that is compatible with neutral evolution. *Plant Cell* **17**, 676–690 (2005).
32. Gibson, G. & Weir, B. The quantitative genetics of transcription. *Trends Genet.* **21**, 616–623 (2005).
33. Ungerer, M. C., Halldorsdottir, S. S., Purugganan, M. D. & Mackay, T. F. C. Genotype–environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics* **165**, 353–365 (2003).
34. Juenger, T. E., Sen, S., Stowe, K. A. & Simms, E. L. Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica* **123**, 87–105 (2005).
35. Dilda, C. L. & Mackay, T. F. C. The genetic architecture of *Drosophila* sensory bristle number. *Genetics* **162**, 1655–1674 (2002).
36. Weinreich, D., Watson, R. & Chao, L. Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
37. Syed, N. H. & Chen, Z. J. Molecular marker genotypes, heterozygosity and genetic interactions explain heterosis in *Arabidopsis thaliana*. *Heredity* **94**, 295–304 (2004).
38. Hausmann, N. J. *et al.* Quantitative trait loci affecting delta C-13 and response to differential water availability in *Arabidopsis thaliana*. *Evolution* **59**, 81–96 (2005).
39. Kearsley, M., Pooni, H. & Syed, N. Genetics of quantitative traits in *Arabidopsis thaliana*. *Heredity* **91**, 456–464 (2003).
40. Malmberg, R. L., Held, S., Waits, A. & Mauricio, R. Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics* **171**, 2013–2027 (2005).
41. Michaels, S. D. & Amasino, R. M. Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* **13**, 935–942 (2001).
42. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347 (2000).
43. Weigel, D. & Nordborg, M. Natural variation in *Arabidopsis*. How do we find the causal genes? *Plant Physiol.* **138**, 567–568 (2005).
44. El-Assal, S. E.-D., Alonso-Blanco, C., Peeters, A. J. M., Raz, V. & Koornneef, M. A. QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genet.* **29**, 435–440 (2001).
45. Werner, J. D. *et al.* Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. *Proc. Natl Acad. Sci. USA* **102**, 2460–2465 (2005).
46. Puchta, H. & Hohn, B. Green light for gene targeting in plants. *Proc. Natl Acad. Sci. USA* **102**, 11961–11962 (2005).
47. Tian, D., Traw, M., Chen, J., Kreitman, M. & Bergelson, J. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
48. Werner, J. D. *et al.* *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* **170**, 1197–1207 (2005).
49. Hoffmann, M. H. Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* **59**, 1425–1436 (2005).
50. Pigliucci, M. in *The Arabidopsis Book* (eds Somerville, C. R. & Meyerowitz, E. M.) (American Society of Plant Biologists, Rockville, Maryland, 2002) doi:10.1199/tab.0009 (www.aspb.org/publications/arabidopsis/) (2002).
51. Shindo, C. *et al.* Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiol.* **138**, 1163–1173 (2005).
52. Thompson, L. The spatiotemporal effects of nitrogen and litter on the population dynamics of *Arabidopsis thaliana*. *J. Ecol.* **82**, 63–68 (1994).
53. Simpson, G. G. & Dean, C. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**, 285–289 (2002).
54. Michaels, S. D., He, Y., Scortecci, K. C. & Amasino, R. M. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **100**, 10102–10107 (2003).
55. Maloof, J. N. *et al.* Natural variation in light sensitivity of *Arabidopsis*. *Nature Genet.* **29**, 441–446 (2001).
56. Michael, T. P. *et al.* Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* **302**, 1049–1053 (2003).
57. Stenoien, H. K., Fenster, C. B., Kuittinen, H. & Savolainen, O. Quantifying latitudinal clines to light responses in natural populations of *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot.* **89**, 1604–1608 (2002).
58. Stinchcombe, J. R. *et al.* A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proc. Natl Acad. Sci. USA* **101**, 4712–4717 (2004).
59. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
60. Callahan, H. S. & Pigliucci, M. Shade-induced plasticity and its ecological significance in wild populations of *Arabidopsis thaliana*. *Ecology* **83**, 1965–1980 (2002).
61. Rauscher, M. D. The measurement of selection on quantitative traits—biases due to environmental covariances between traits and fitness. *Evolution* **46**, 616–626 (1992).
62. Wade, M. J. & Kalisz, S. The causes of natural selection. *Evolution* **44**, 1947–1955 (1990).
63. Donohue, K. *et al.* The evolutionary ecology of seed germination of *Arabidopsis thaliana*: variable natural selection on germination timing. *Evolution* **59**, 758–770 (2005).
64. Callahan, H. S., Dhanooolal, N. & Ungerer, M. C. Plasticity genes and plasticity costs: a new approach using an *Arabidopsis* recombinant inbred population. *New Phytol.* **166**, 129–139 (2005).
65. Weigand, C. *et al.* Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165**, 321–329 (2003).
66. Verhoeven, K. J. F., Vanhala, T. K., Biere, A., Nevo, E. & Van Damme, J. M. M. The genetic basis of adaptive population differentiation: a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution* **58**, 270–283 (2004).
67. Mauricio, R. *et al.* Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. *Genetics* **163**, 735–746 (2003).
68. Shen, J., Araki, H., Chen, L., Chen, J.-Q. & Tian, D. Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics* **172**, 1243–1250 (2006).
69. Korves, T. & Bergelson, J. A novel cost of R gene resistance in the presence of disease. *Am. Nat.* **163**, 489–504 (2004).
70. Siegal, M. L. & Hartl, D. L. Transgene coplacement and high efficiency site-specific recombination with the *Cre/loxP* system in *Drosophila*. *Genetics* **144**, 715–726 (1996).
71. Plagnol, V., Padhukasahasram, B., Wall, J. D., Marjoram, P. & Nordborg, M. Relative influences of crossing-over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. *Genetics* **172**, 2441–2448 (2005).
72. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet.* **38**, 203–208 (2006).
73. Remington, D. L. *et al.* Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. USA* **98**, 11479–11484 (2001).
74. Aranzana, M. *et al.* Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* **1**, e60 (2005).
75. Schlotterer, C. Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends Genet.* **19**, 32–38 (2003).

Acknowledgements We thank M. Noor, M. Nordborg and our collaborators and laboratory members for discussion and comments. T.M.-O. and J.S. were supported by Duke University and the National Science Foundation, respectively.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to T.M.-O. (tmo1@duke.edu).