Arabidopsis thaliana and its wild relatives: a model system for ecology and evolution

Thomas Mitchell-Olds

The postgenomics era will bring many changes to ecology and evolution. Information about genomic sequence and function provides a new foundation for organismal biology. The crucifer *Arabidopsis thaliana* and its wild relatives will play an important role in this synthesis of genomics and ecology. We discuss the need for model systems in ecology, the biology and relationships of crucifers, and the molecular resources available for these experiments. The scientific potential of this model system is illustrated by several recent studies in plant–insect interactions, developmental plasticity, comparative genomics and molecular evolution.

Thomas Mitchell-Olds
Dept of Genetics and Evolution, Max-Planck Institute of Chemical Ecology, Winzerlaer Strasse 10, 07745 Jena, Germany.
e-mail: tmo@ice.mpg.de

There is a rich history of field studies on many species in diverse environments in plant ecology and evolution. Molecular methods are sufficiently advanced that we can now study the function and evolution of genes that are important in ecology; for example, those that control responses to temperature, drought, pathogens, insects and other environmental challenges. However, studies in molecular ecology and evolution require large initial investments in molecular technologies. To be cost effective, these investments must be shared among multiple laboratories and funding sources. Equally important, individual investigators can achieve more in-depth biological understanding when they build upon previous studies from a research community that shares the tools and resources of model organisms. Clearly, no single species or model system can address the diverse range of ecological and evolutionary questions of current interest. Nevertheless, the crucifer *Arabidopsis thaliana* and its wild relatives provide a model system that has a vast array of molecular tools[1], genetic resources and biological information that can be used to address fundamental questions in ecology and evolution (Table 1).

Systematics

Arabidopsis thaliana belongs to the Cruciferae (family Brassicaceae, Capparales). Over the past 40 million years, ~3000 crucifer species have radiated to many habitats worldwide[2]. Recent molecular and morphological studies have outlined systematic relationships across the Brassicaceae (Fig. 1), and across the closest relatives of *Arabidopsis*[2–4]. The genus *Arabidopsis* contains about ten species that are native to Eurasia, North Africa and North America. The closest wild relatives of *A. thaliana* (Fig. 2) include *A. lyrata* and *A. halleri*, which are self-incompatible diploids with eight chromosome pairs. Evolutionary and ecological studies have focused on *A. lyrata* ssp. *lyrata* in North America, *A. lyrata* ssp. *petraea* and *A. halleri* in Europe, and *A. halleri* ssp. *gemmifera* in Japan. (Earlier literature refers to these taxa as *Arabis lyrata*, *Cardaminopsis petraea*, *C. halleri* and *Arabis gemmifera*, respectively.) Species are interfertile within this group of closely related $N = 8$ taxa. Chromosome number is reduced to $N = 5$ in *A. thaliana*, so diploid mapping crosses with wild relatives are impossible[5]. However, *Arabidopsis suecica* is believed to be an allotetraploid derived from *A. thaliana* and *A. arenosa*.

Previous taxonomic treatments of *Arabis* do not constitute a natural classification, because various distantly related species have been included in this grouping[2–4], and future taxonomic work is needed in this area. *Arabis* refers to at least two distantly related but natural groups: $N = 7$ species in North America, and $N = 8$ species of *Arabis sensu stricto* in Eurasia and North America. *Arabis alpina* is the type species of the genus *Arabis*, which is distantly related to *Arabidopsis*. This group of $N = 8$ *Arabis* spp. is found in Eurasia, East Africa, northern California and southern Oregon[2,4].

Approximately 50 named species of *Arabis*, such as *A. holboellii* and *A. drummondii*, are found in North America. This group has a basal number of $N = 7$ chromosomes, and might be reclassified as genus *Boechera* based on taxonomic considerations[4]. The group is closely related to *Arabidopsis*, and reproduces by self-pollination, outcrossing or APOMIXIS[6,7] (see Glossary). Taxa from diverse habitats can be cross fertilized, so it is feasible to map quantitative trait loci (QTLs) that control local adaptation to biotic and abiotic factors.

Agriculturally important crucifers include *Brassica*[8] and *Raphanus*. Oilseed rape (canola) consists of varieties of *B. napus* and *B. rapa*. Horticultural crops include European cabbage, cauliflower and brussel sprouts (*B. oleracea*), Asian vegetables such as Chinese cabbage and bok choy (*B. rapa*), and radish (*Raphanus* spp.). Because these weedy annuals are widely distributed in agricultural and other disturbed habitats, they have featured in many ecological and evolutionary studies. Rapid cycling genotypes are also useful for

http://tree.trends.com 0169-5347/01/$ – see front matter © 2001 Elsevier Science Ltd. All rights reserved. PII: S0169-5347(01)02291-1

Table 1. Examples of ecological and evolutionary questions that can be addressed using model systems

Questions	Refs
• What genes are responsible for ecologically important variation? Why are they polymorphic? Are quantitative trait locus alleles young, deleterious and spatially restricted, or are they old widespread polymorphisms contributing to local adaptation?	17
• How often do disease resistance genes evolve via an arms race of sequential SELECTIVE SWEEPS, versus ancient polymorphisms maintained by BALANCING SELECTION?	18
• What ecological processes influence breeding system evolution and the frequency of outcrossing, inbreeding, or asexual reproduction?	6
• Is plasticity adaptive? Do developmental responses to environmental conditions increase fitness? Have these responses been shaped by natural selection?	37
• Can expression profiling identify candidate genes that influence insect resistance, drought tolerance, flowering time and other ecologically important traits in natural populations? Does natural selection promote adaptive evolution of these loci?	42,44
• How do polyploids differ ecologically and physiologically from their diploid relatives? How are these differences shaped by natural selection?	5
• What is the genetic basis of speciation? Does sexual selection of pollen-stigma recognition genes influence reproductive isolation? What is the role of gene duplications in interspecific genomic incompatibility?	29,50
• What is the relative importance of changes in gene regulation versus protein function within and among species? Have these evolutionary changes resulted from neutral genetic drift or from natural selection?	14,34
• Do neutral molecular markers predict patterns of ecologically important variation in rare plants? What is the molecular basis of adaptively important polymorphisms in endangered species?	51

laboratory research. Brassicas are the closest agricultural relatives of *Arabidopsis*, so comparative genomics should provide applied benefits. However, because these species are ancient polyploids[8], molecular studies can be far more complex than in diploid *Arabidopsis*.

Geography and environments

A recent study of molecular markers in 142 accessions sampled from the native range of *A. thaliana* found statistically significant isolation by distance[9], probably reflecting Pleistocene migrations across Europe and Asia. Although human disturbance clearly influences the biogeography of *A. thaliana*, there is evidence for post-Pleistocene colonization of Europe from glacial refugia near the Mediterranean and in Asia[9]. Only a few ECOTYPES of *A. thaliana* are available from central Asia, and these show tantalizing patterns of genetic diversity for molecular and phenotypic traits.

There is no evidence for different subspecies among the *Arabidopsis* ecotypes that have been examined[9,10]. Furthermore, studies of amplified fragment length polymorphisms (AFLPs) and nucleotide variation find evidence for modest, but important historical recombination within and among genes, hence ecotype evolution does not proceed in a bifurcating, tree-like manner[10,11]. Instead, evidence for historical outcrossing and recombination is apparent in nucleotide polymorphisms segregating among individuals.

Species of *Arabidopsis* and *Arabis* have evolved a broad range of life-history, developmental and physiological traits that adapt them to a wide variety of habitats. Combined with molecular tools and biological information from *A. thaliana*, researchers can address many ecological questions in these species. Although *A. thaliana* is common in agricultural fields and disturbed sites in the temperate zones, it is also found in undisturbed rocky sites and forest openings (T. Mitchell-Olds and M. Clauss, unpublished). Significant isolation by distance across its native range suggests that populations of *A. thaliana* in Europe and Asia predate human agricultural disturbance[9], hence they have long ecological and evolutionary histories in association with insects, pathogens, competitors and abiotic environmental factors.

Among closely related taxa, *A. lyrata* grows on sand dunes, riverbanks, cliffs, or forest understorey in sites with low densities of competing species (Fig. 3). By contrast, *A. halleri* thrives in perennial meadows under intense competition with grasses and forbs, as well as in heavy metal-contaminated sites. These short-lived perennial species inhabit primarily mesic environments. Elsewhere, species of North American *Arabis* inhabit deserts and xeric grassland, whereas other *Arabis* taxa grow in cold, wet, or high elevation sites (Fig. 3). In some areas of western North America, *Arabis* spp. can be studied in pristine environments that are largely free of human disturbance.

A broad geographical range is important for model systems in plant ecology, so that field populations will be readily accessible to many researchers. *Arabidopsis* and its wild relatives fulfill this need[3,12]. *Arabidopsis thaliana* is widely distributed in North Africa, Eurasia and North America. *Arabidopsis lyrata* is found in Canada, the American Midwest and East Coast, UK, Scandinavia, central Europe, and parts of Asia. *Arabidopsis halleri* ranges from France to Taiwan. $N = 7$ *Arabis* (*Boechera*) spp. are found across North America[6], with vast diversity in the western USA. $N = 8$ species of *Arabis sensu stricto* occur in northern California and across Eurasia. Finally, *Brassica* and related species are common in Eurasia and have been introduced to most temperate agricultural areas. Importantly, seed collections of many species and genotypes are available from stock centers.

Genetic tools and biological resources

A model system for molecular ecology and evolution requires a diverse array of biological and molecular resources that are freely available to the research community. Many microsatellite markers have been optimized for use in the wild relatives of *A. thaliana*[13], and additional markers can be easily developed using genomic sequence information (M. Clauss, unpublished). Bacterial artificial chromosome (BAC) libraries are available from *A. lyrata* and *A. drummondii* (J. Nasrallah and T. Mitchell-Olds, unpublished),

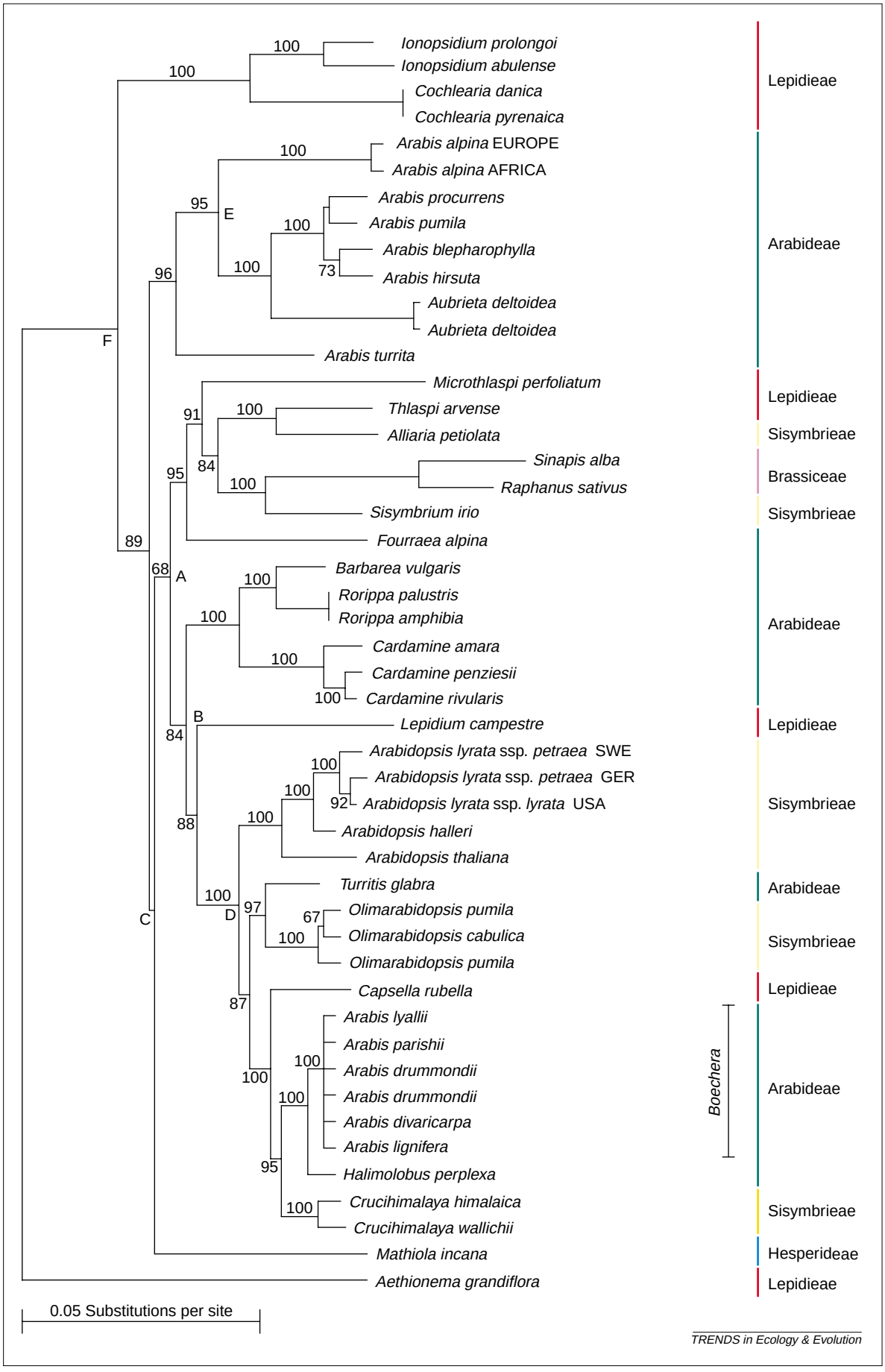

Fig. 1. Neighbor-joining distance tree based on *matK* and *Chs* sequences[2]. Percentage bootstrap values from 1000 replicates are shown above branches. Tribal assignments are given on the right. Approximate divergence dates (million years before present) for nodes A–F are: A, 16–21; B, 13–19; C, 19–25; D, 10–14; E, 15–17; and F, 26–32. Several taxa mentioned in the text are not shown in this figure: *Arabidopsis arenosa* and A. *suecica* are closely related to *A. lyrata*. *Brassica* is near *Raphanus* and *Sinapis*. Finally, *Leavenworthia* is related to *Barbarea*[32]. Reproduced, with permission, from Ref. 2.

Fig. 2. Wild relatives of *Arabidopsis thaliana*. (a) *Arabidopsis lyrata* ssp. *lyrata*. (b) *Arabis holboellii*. (c) *Arabis fecunda*. Figure 2c is reproduced, with permission, from Ref. 52.

providing large insert clones (>80 kb) for analyses of gene families, molecular evolution and comparative genomics. A linkage map comparing *A. thaliana* and *A. lyrata* is approaching completion (O. Savolainen, unpublished). Levels of nucleotide identity are generally high within the genus *Arabidopsis* (>90% within coding regions)[14–16], so carefully chosen PCR primers can be used to amplify most genes of interest. At greater evolutionary distances within the Cruciferae, DNA or amino acid sequences can be compared from other dicotyledenous families to identify conserved regions suitable for primer design. Genetic variation at known loci influences life histories[17], pest resistance[14,18], floral architecture[19] and other ecologically important traits. Consequently, genes of interest are readily accessible from the close wild relatives of *A. thaliana*. More information on the wild relatives of *Arabidopsis* can be found at http://www.ice.mpg.de/departments/Gen/wild.htm.

Breeding system and genetic variation

Breeding system influences experimental tractability in *A. thaliana* and its relatives. High levels of self pollination (e.g. *A. thaliana* and *A. drummondii*) facilitate QTL mapping and progeny testing in advanced generation crosses, which can be hindered by inbreeding depression in outcrossing species[20]. However, many population genetic models assume random mating, which is approximated more closely in populations of outcrossing species (e.g. *A. lyrata* and *A. halleri*). Outcrossing species are predicted to maintain higher levels of genetic variation within populations[21]. This pattern has been observed among *Leavenworthia* spp. at the *PGI* (phosphoglucose isomerase) locus[22], but is less clear when the *ADH* (alcohol dehydrogenase) gene is compared between *A. lyrata* and *A. thaliana*[15]. Data from additional loci are needed to resolve this discrepancy. Breeding system also influences transposon abundance and mobility[23], which, in turn, can affect genetic variation for quantitative traits. Consequently, breeding system can influence the genetic architecture of quantitative traits.

Sequence variation at neutral genes can allow estimation of important population genetic parameters, such as population growth rate, or $N_e\mu$ and $N_e m$, the products of effective population size and mutation or migration[24]. Analysis of DNA sequence variation within and between species can compare observed patterns of variation with predictions of neutral equilibrium models. In *Arabidopsis* and related species, neutrality has been rejected by some studies (e.g. Refs 18,25,26) but not by others (e.g. Ref. 16). Several studies find patterns of genetic variation indicative of recent population expansion in *A. thaliana*[16,27]. A large study of nucleotide polymorphism at 400 loci (K. Schmid and T. Mitchell-Olds, unpublished) finds genome-wide evidence for population expansion, which corresponds to the large population size of *A. thaliana* resulting from human disturbance.

Plants are ideal for molecular ecological genetics because they are sessile and are confronted by measurable environmental challenges, such as temperature, drought, pathogens and insects. Many known genes function in response to these environmental challenges, so these loci could be influenced by natural selection. Ecologically important polymorphisms have already been identified at known loci in *A. thaliana* (e.g. Refs 17,18) and corresponding genes can be isolated readily in

Fig. 3. Habitats of *Arabidopsis* and *Arabis*. (a) Populations of *Arabidopsis lyrata* grow adjacent to montane glaciers in the Pacific Northwest, USA. (b) *Arabidopsis halleri* grows under highly competitive conditions in perennial meadows, Thüringerwald, central Germany. (c) *Arabis perennans* grows in xeric prairie and desert environments, AZ, USA. (d) *Arabis drummondii* grows in meadow, forest and high-elevation habitats in the Rocky Mountains, USA.

Box 1. Expression profiling and the molecular ecology of plant defense

Expression profiling simultaneously quantifies expression (mRNA abundance) for hundreds or thousands of genes (Fig. I). *Arabidopsis thaliana* 'DNA chips' with thousands of sequenced genes are available commercially and from public-sector resource centers. Alternatively, expression of known genes of special interest (e.g. drought or disease blots) can be assayed on nylon or glass at reasonable cost[a,b] with adequate replication[c]. These tools can be applied to many ecological questions, for example: are so-called 'inducible defenses' actually expressed under normal field conditions, even in the absence of insects and pathogens?

Many researchers are examining possible costs of plant resistance to insect herbivores[d]. Although this issue can be approached using quantitative genetic methods, it is desirable to understand the functional and molecular basis of plant defenses and putative costs. For example, manipulation of defensive physiology with plant hormones, mutants, or transgenics can up- or downregulate the jasmonic-acid signal transduction pathway. Several studies have shown that jasmonate-induced responses are costly, resulting in reduced growth or fitness[e–g]. However, jasmonates have extensive pleiotropic effects on growth and metabolism, and further research is required to determine whether fitness costs are attributable to either plant defense, or to confounded pleiotropic effects of the jasmonate signal transduction pathway[d,g]. Expression profiling (Fig. I) shows extensive crosstalk between the jasmonate, salicylate, and ethylene response pathways, thus affecting many aspects of plant physiology[h]. Consequently, manipulation of distal defense mechanisms (downstream from signaling components responsible for crosstalk) is necessary to determine whether costs are actually attributable to plant defense[d]. This implies that detailed understanding of signal transduction pathways is required before costs of defense can be addressed by either transgenic or physiological manipulations.

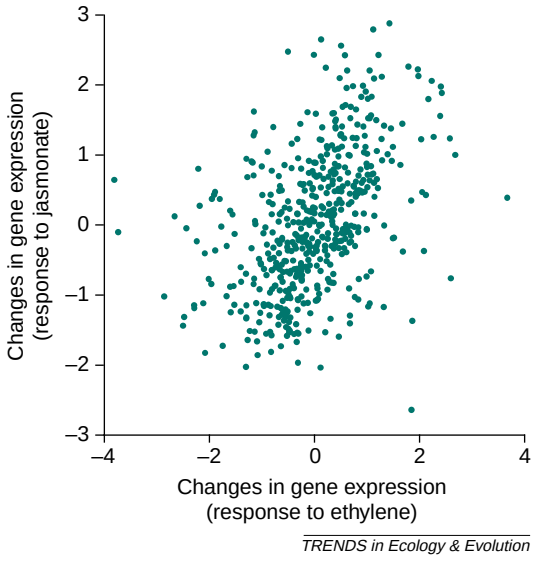

Fig. I. Changes in gene expression controlled by the methyl jasmonate wound response pathway are positively correlated with effects of another plant hormone, ethylene. Schenk *et al.*[h] examined 529 differentially regulated genes with repeatable expression levels in control experiments. Application of ethylene or jasmonate hormones results in correlated changes in gene expression ($r = 0.45$; $n = 529$, $P < 0.001$). Although these hormones influence gene expression in many aspects of plant physiology, some of these responses have clear effects on plant resistance to pest species[i]. Data on expression levels of responsive genes was obtained from http://www.pnas.org. Expression levels were log transformed, and changes in response to hormone treatments (Z-scores) were calculated as residuals from a third-order regression on expression levels in the control treatments. Of the genes, 529 had repeatable expression in replicated control experiments ($-1.5 < Z < 1.5$). Axes show changes in gene expression in standard deviation units, as a deviation from the regression line that summarizes average changes between control and treatment plants.

References

a Reymond, P. *et al.* (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12, 707–720

b Sekia, M. *et al.* (2001) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13, 61–72

c Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res. Camb.* 77, 123–128

d Purrington, C.B. (2000) Costs of resistance. *Curr. Opin. Plant Biol.* 3, 305–308

e Baldwin, I.T. (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8113–8118

f Agrawal, A.A. (1999) Induced responses to herbivory in wild radish: effects on several herbivores and plant fitness. *Ecology* 80, 1713–1723

g Redman, A.M. *et al.* (2000) Fitness costs of jasmonic acid-induced defense in tomato, *Lycopersicon esculentum*. *Oecologia* 126, 380–385

h Schenk, P.M. *et al.* (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11655–11660

i Stotz, H.U. *et al.* (1999) Plant–insect interactions. *Curr. Opin. Plant Biol.* 2, 268–272

wild relatives growing in undisturbed natural populations. Thus, it will be possible to determine the ecological and evolutionary forces influencing quantitative genetic variation in the wild, and to examine geographical distribution of non-neutral allelic variants. Such studies can determine whether QTL alleles are young, deleterious and spatially restricted, or whether they are old widespread polymorphisms contributing to local adaptation.

Evolutionary genomics

Plant genomes are shaped by a dynamic balance between rates of gene duplication and gene loss. Several studies have examined gene families and

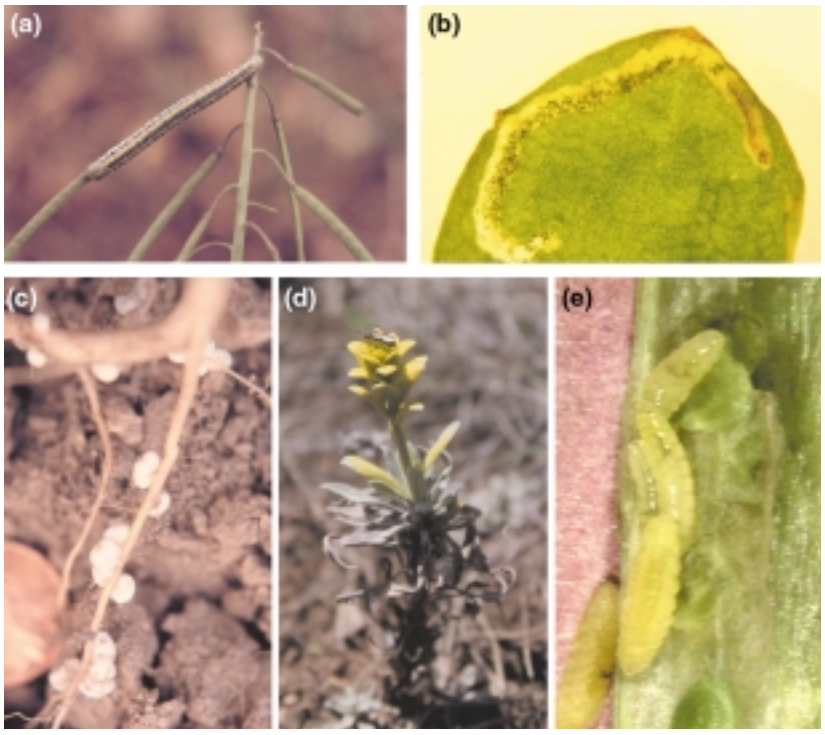

Fig. 4. Insects and pathogens of *Arabidopsis* and *Arabis*. (a) *Pontia* (Pieridae) consuming *Arabis holboellii* fruits in Montana. (b) Dipteran leaf miner on *Arabidopsis halleri*. (c) Insect root galls on *A. holboellii*. (d) *Puccinia* infection of *Arabis perennans* causes elongation of pseudo flowers emitting fungal spores. (e) Insect seed predators destroy many developing seeds in *Arabis* populations. Fig. 4d reproduced, with permission, from J-Z. Lin.

segmental duplications in the *Arabidopsis* genome[1,28,29]. Extensive duplicated regions suggest that a polyploidization event occurred ~100 million years ago (Mya)[28]. Since then, progressive loss of individual genes from duplicated segments has been a fundamental feature of genome evolution. The half-life of duplicated genes in *Arabidopsis* is estimated at approximately 3.2 My (Ref. 29). Continuing gene loss is opposed by individual gene duplications, often in tandem gene families. During population differentiation, loss of gene function at alternative, duplicated loci provides a plausible speciation mechanism consistent with existing models[29].

Several large genomic regions have been sequenced from relatives of *Arabidopsis*[30,31]. The genera *Capsella* and *Brassica* diverged from *Arabidopsis* 10–14 and 16–21 Mya, respectively[32] (Fig. 1). In comparison to *Arabidopsis*, functional genes are highly conserved, whereas intergenic regions display insertions, deletions and higher levels of nucleotide substitution. Such sequence comparisons will be useful for genetic analysis of crop plants, and for annotation and regulatory studies of *Arabidopsis* genes[33,34]. Furthermore, evolutionary analysis of comparative sequence data can provide insights on the evolution of gene families, the molecular basis of adaptation and the rate of deleterious mutations[35].

Growth and phenology

Many ecological and evolutionary studies have examined genetic variation for growth form and age at first reproduction in *A. thaliana*[17,19,36,37]. Phytochrome-mediated light perception has received particular attention, because the red:far red light (R:FR) ratio provides an unambiguous indication of neighboring plants and community density[37]. As predicted, single locus mutants deficient in phytochrome activity display altered fitness responses to plant density[36]. Recently, a large manipulative study of *A. thaliana* genotypes from natural populations examined the adaptive significance of phenotypic plasticity in response to environmental conditions. Dorn *et al.*[37] measured plastic responses to light levels, R:FR, and plant density, and natural selection on phenology and growth form in these environments. Developmental changes in response to R:FR signals had beneficial effects on fitness, showing the adaptive importance of phenotypic plasticity to this environmental cue.

Insects and pathogens

Arabidopsis and *Arabis* are attacked by a wide variety of pest species, offering potential for studies of coevolution, costs of plant defense and other topics in plant–insect and plant–pathogen interactions (Box 1). Many insect species complete their life cycle on *Arabidopsis* relatives (Fig. 4), including the widely studied cabbage butterflies *Pieris* spp. QTL influencing insect resistance can be rapidly mapped in *Arabidopsis*[38]. Several leaf miners in the genus *Scaptomyza* attack *A. thaliana* and related species[39]. *Scaptomyza* are closely related to *Drosophila virilis* and have active P-element transposons[40], offering potential for a plant–insect model system combining the genetic tools of *Arabidopsis* and *Drosophila*. Likewise, microbial pathogens such as *Albugo*, *Peronospora*, *Pseudomonas* and *Puccinia* are commonly found on wild and agricultural crucifers[6,18].

Ecological models of plant–insect interactions have suggested that perennial species might invest heavily in chemical defenses against insect herbivores, whereas weedy and ephemeral species could lack common mechanisms of plant defense, because they might escape detection by herbivores[41]. However, recent molecular studies show that *A. thaliana* possesses typical mechanisms of plant defense that have been documented in many plant families. For example, ethylene, jasmonate and salicylate signal transduction pathways are conserved in *A. thaliana*, and influence plant resistance to insect and microbial pests[42,43]. Genomic analyses have recently shown that many familiar defense and signaling genes are induced by *Pieris* feeding on *A. thaliana*[44]. Likewise, glucosinolates are biologically active secondary metabolites that are widely implicated in plant–insect and plant–pathogen interactions[45–47]. Several glucosinolate biosynthetic and modification genes have recently been cloned in *A. thaliana*[48,49]. In summary, *A. thaliana* employs pathways of plant defense that are widely conserved in many plant families, hence *Arabidopsis* and *Arabis* provide

representative models of physiology and evolution of plant–pest interactions in herbaceous dicots.

Conclusion

Arabidopsis thaliana and its wild relatives provide a model system with a diverse array of molecular tools, genetic resources and biological information. Crucifers are geographically widespread, and are adapted to a wide range of biotic and abiotic environments. Molecular tools can identify genes that evolve in response to temperature, drought, herbivores, pathogens and other environmental factors. Such approaches can elucidate the ecological and evolutionary forces influencing quantitative genetic variation in the wild, and the history and geographical distribution of ecologically important alleles. Recent studies of plastic responses to light quality have shown the adaptive importance of environmentally mediated changes in plant development. Evolutionary analysis of comparative genomic sequence can provide insights on the evolution of gene families, the molecular basis of adaptation and the rate of duplication and loss in gene families. The synthesis of organismal biology and genomics will change our understanding of ecology and evolution.

Glossary

Apomixis: asexual reproduction, producing seed progeny identical to the maternal genotype.
Balancing selection: natural selection that maintains a polymorphism at a stable equilibrium. Examples include heterozygote advantage and frequency dependent selection.
Ecotype: in evolutionary biology, ecotypes are genetically differentiated and locally adapted populations. However, in the *Arabidopsis* literature, 'ecotype' is synonymous with 'accession', and does not necessarily imply local adaptation.
Selective sweep: when selection fixes a favorable mutation, linked, neutral polymorphisms can also become fixed via hitch hiking. Thus, nearby polymorphisms are swept away during selective fixation of advantageous mutations.

Acknowledgements
I thank M. Clauss, D. Kliebenstein, J. Kroymann, K. Schmid, J. Schultz and four anonymous reviewers for comments. Figures 2c and 4d are generously provided by J-Z. Lin and D. Siemens, respectively. This work was supported by the Max-Planck-Gesellschaft, the US National Science Foundation (Grant DEB 9527725), the European Union, and the Bundesministerium für Bildung und Forschung.

References

1 The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
2 Koch, M. *et al.* (2001) Molecular systematics of the Cruciferae: evidence from coding plastidic *matK* and nuclear *CHS* sequences. *Am. J. Bot.* 88, 534–544
3 O'Kane, S.L. and Al-shehbaz, I.A. (1997) A synopsis of *Arabidopsis* (Brassicaceae). *Novon* 7, 323–327
4 Galloway, G.L. *et al.* (1998) Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. *Mol. Biol. Evol.* 15, 1312–1320
5 Nasrallah, M.E. *et al.* (2000) *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol.* 124, 1605–1614
6 Roy, B.A. (1993) Patterns of rust infection as a function of host genetic diversity and host density in natural populations of the apomictic crucifer, *Arabis holboellii*. *Evolution* 47, 111–124
7 Sharbel, T.F. and Mitchell-Olds, T. (2001) Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). *Heredity* 87, 59–68
8 Lagercrantz, U. (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150, 1217–1228
9 Sharbel, T.F. *et al.* (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and post-glacial colonization of Europe. *Mol. Ecol.* 9, 2109–2118
10 Miyashita, N.T. *et al.* (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics* 152, 1723–1731
11 Kuittinen, H. and Aguade, M. (2000) Nucleotide variation at the chalcone isomerase locus in *Arabidopsis thaliana*. *Genetics* 155, 863–872
12 Alonso-Blanco, C. and Koornneef, M. (2000) Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends Plant Sci.* 5, 22–29
13 van Treuren, R. *et al.* (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* 14, 230–238
14 Bishop, J.G. *et al.* (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5322–5327
15 Savolainen, O. *et al.* (2000) Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol. Biol. Evol.* 17, 645–655
16 Aguade, M. (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 18, 1–9
17 Johanson, U. *et al.* (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290, 344–347
18 Stahl, E.A. *et al.* (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400, 667–671
19 Juenger, T. *et al.* (2000) Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. *Genetics* 156, 1379–1392
20 Karkkainen, K. *et al.* (1999) Genetic basis of inbreeding depression in *Arabis petraea*. *Evolution* 53, 1354–1365
21 Charlesworth, D. and Charlesworth, B. (1995) Quantitative genetics in plants: the effect of the breeding system on genetic variability. *Evolution* 49, 911–920
22 Liu, F. *et al.* (1999) The effect of mating system differences on nucleotide diversity at the *phosphoglucose isomerase l* locus in the plant genus *Leavenworthia*. *Genetics* 151, 343–357
23 Charlesworth, D. and Charlesworth, B. (1995) Transposable elements in inbreeding and outbreeding populations. *Genetics* 140, 415–417
24 Beerli, P. and Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4563–4568
25 Filatov, D.A. and Charlesworth, D. (1999) DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus. *Genetics* 153, 1423–1434
26 Kawabe, A. *et al.* (2000) DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. *Genetics* 156, 1339–1347
27 Innan, H. and Stephan, W. (2000) The coalescent in an exponentially growing metapopulation and its application to *Arabidopsis thaliana*. *Genetics* 155, 2015–2019
28 Vision, T.J. *et al.* (2000) The origins of genomic duplications in *Arabidopsis*. *Science* 290, 2114–2117
29 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
30 Quiros, C.F. *et al.* (2001) *Arabidopsis* and *Brassica* comparative genomics: sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. *Genetics* 157, 1321–1330
31 Rossberg, M. *et al.* (2001) Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, *Arabidopsis*, and *Capsella* genomes. *Plant Cell* 13, 979–988
32 Koch, M.A. *et al.* (2000) Comparative evolutionary analysis of the chalcone synthase and alcohol dehydrogenase loci among different lineages of *Arabidopsis*, *Arabis* and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498
33 Wiehe, T. *et al.* (2000) SGP-1: Prediction and validation of homologous genes based on sequence alignments. *Genome Res.* 11, 1574–1583
34 Koch, M.A. *et al.* (2001) Comparative genomics and regulatory evolution: conservation and function of the *Chs* and *Apetala3* promoters. *Mol. Biol. Evol.* 18, 1882–1891
35 Charlesworth, D. *et al.* (2001) Genome sequences and evolutionary biology, a two-way interaction. *Trends Ecol. Evol.* 16, 235–242
36 Ballare, C.L. and Scopel, A.L. (1997) Phytochrome signalling in plant canopies: testing its population-level implications with photoreceptors mutants of *Arabidopsis*. *Funct. Ecol.* 11, 441–450
37 Dorn, L. *et al.* (2000) Plasticity to light cues and resources in *Arabidopsis thaliana*: testing for adaptive value and costs. *Evolution* 54, 1982–1994

38 Jander, G. *et al.* (2001) The *TASTY* locus on chromosome 1 of *Arabidopsis* affects feeding of the insect herbivore *Trichoplusia ni*. *Plant Physiol.* 126, 890–898
39 Hering, E.M. (1957) *Blattminen von Europa*, Uitgeverij Dr W. Junk
40 Kwiatowski, J. and Ayala, F.J. (1999) Phylogeny of *Drosophila* and related genera: Conflict between molecular and anatomical analyses. *Mol. Phylog. Evol.* 13, 319–328
41 Feeny, P. (1976) Plant apparency and chemical defense. In *Biochemical Interaction Between Plants and Insects* (Wallace, I. *et al.*, eds), pp. 1–40, Plenum Press
42 Schenk, P.M. *et al.* (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11655–11660
43 Stotz, H.U. *et al.* (2000) Induced plant defense responses against chewing insects. Ethylene signaling reduces resistance of *Arabidopsis* against Egyptian cotton worm but not diamondback moth. *Plant Physiol.* 124, 1007–1017
44 Reymond, P. *et al.* (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12, 707–720
45 Mauricio, R. (1998) Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am. Nat.* 151, 20–28
46 Siemens, D. and Mitchell-Olds, T. (1998) Evolution of pest induced defenses in *Brassica* plants: tests of theory. *Ecology* 79, 632–646
47 Stowe, K.A. (1998) Experimental evolution of resistance in *Brassica rapa*: Correlated response of tolerance in lines selected for glucosinolate content. *Evolution* 52, 703–712
48 Wittstock, U. and Halkier, B.A. (2000) Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. *J. Biol. Chem.* 275, 14659–14666
49 Kliebenstein, D.J. *et al.* (2001) Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. *Plant Cell* 13, 681–693
50 Kliman, R.M. *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156, 1913–1931
51 McKay, J.K. *et al.* (2001) Local adaptation despite absence of marker diversity in the rare Sapphire Rockcress. *Proc. R. Soc. London B Biol. Sci.* 268, 1715–1721
52 Settles, A.M. and Byrne, M. (1998) Opportunities and challenges grow from *Arabidopsis* genome sequencing. *Genome Res.* 8, 83–85

Prospects for nuclear gene phylogeography

Matthew P. Hare

In phylogeography, an empirical focus on gene lineages enables the history of population processes to be inferred from the simultaneous analysis of temporal and spatial patterns. Rapidly evolving cytoplasmic DNA has been the empirical workhorse propelling the success of this nascent field. Now, as more sophisticated historical models are being tested, there is a growing need for phylogeography to expand from a largely marker-specific discipline to a more general analytical approach that can be applied across independent loci. Recent results using nuclear haplotypes to study phylogeography indicate that the anticipated technical and biological hurdles can be overcome in many taxa to achieve phylogeographical comparisons across unlinked loci. Although many challenges remain, a more complete understanding of the historical, demographic and selective processes shaping phylogeographical patterns is emerging.

Matthew P. Hare
Dept of Biology, University of Maryland, College Park, MD 20742, USA.
e-mail: matt.hare@umail.umd.edu

Over the past decade, phylogeography grew as a discipline because allelic phylogenies provided explicitly historical tools for the study of geographical subdivision among populations[1]. Analysing the relative ages and historical relationships of alleles in a geographical context could distinguish ongoing processes such as GENE FLOW (see Glossary) from previous events such as range expansion. The temporal resolution offered by genealogies was previously unavailable through classical population genetic analyses (e.g. hierarchical partitioning of variation using *F*-statistics)[2]. In addition, using alleles rather than populations as the basic unit in phylogenetic clustering promoted integrated analyses of history above and below the species level without the need to make *a priori* taxonomic distinctions. Animal mitochondrial (mt)DNA and chloroplast (cp)DNA have been the primary data sources making phylogeography such a productive empirical approach. Extending and generalizing phylogeography by using nuclear HAPLOTYPE data has been a desirable goal in efforts to test ever more sophisticated historical models. However, given the potential for technical and biological complications when analysing nuclear haplotype data at the intraspecific level, it has been uncertain whether a phylogeographical study design could generally be extended to noncytoplasmic markers[1]. An evaluation of achievements made to date in nuclear phylogeography arguably provides the most practical guide to its future potential.

A gene tree for a single cytoplasmic or nuclear (nDNA) LOCUS provides a slim and sometimes misleading representation of the population histories through which alleles were transmitted[3,4]. Because cytoplasmic loci are usually inherited uniparentally, they will not have genealogical patterns that are representative of the entire population history, especially when sex biases have affected fitness or dispersal behavior[5]. Loci under selection can also have genetic patterns that deviate from expectations based on population history and demography, yet the selected patterns often mimic alternative demographic histories. Testing alternative hypotheses about evolutionary forces must ultimately rely on the fact that selection acts locally within a genome, whereas population demography leaves a common signature across all neutral loci[1,6].

 0169-5347/01/$ – see front matter © 2001 Elsevier Science Ltd. All rights reserved. PII: S0169-5347(01)02326-6