Linking molecular insight and ecological research

Robert B. Jackson, C. Randal Linder, Michael Lynch, Michael Purugganan, Shauna Somerville and Susan S. Thayer

Significant environmental challenges, including the genetic and physiological effects of environmental pollutants, the rapid spread of diseases and invasive species, the release of transgenic organisms and global climate change, affect our daily lives and the sustainability of ecosystems. Managing these environmental problems will require new approaches that span the biology of genes, organisms, populations, communities and ecosystems. In parallel with these practical concerns is the basic need to study gene functions in their natural context. The *Arabidopsis* 2010 project, for example, seeks to understand the functions of all 25 000 *Arabidopsis* genes within a decade but, to do so, we must also understand the role of the environment in determining gene function. A new priority is evident – understanding the interplay of molecular mechanisms with organismal and ecosystem biology. Combining genomic and ecological research perspectives will answer crucial unresolved questions, but will require significant new multidisciplinary resources, infrastructure and training.

Published online: 11 July 2002

Robert B. Jackson*
Dept of Biology and Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708-0340, USA.
*e-mail: jackson@duke.edu

C. Randal Linder
Section of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA.

Michael Lynch
Dept of Biology, Indiana University, Bloomington, IN 47405, USA.

Michael Purugganan
Dept of Genetics, North Carolina State University, Raleigh, NC 27695, USA.

Shauna Somerville
Susan S. Thayer
Carnegie Institution of Washington, Dept of Plant Biology, Stanford, CA 94305, USA.

Throughout the 20th century, individual disciplines in biology succeeded by specializing in different scales of size and organization. Effective in many respects, this traditional partitioning sometimes hindered our understanding of how organisms function and interact with each other and their environment. For example, a complete understanding of organisms at the molecular level requires an account of how genotype, phenotype and environment interact. Conversely, molecular methods can suggest and, in some cases, identify candidate genes underlying ecologically relevant traits, providing creative new tools with which to address ecological questions.

Areas of inquiry that are historically ecological include whole-organism physiology, species interactions, the organization of communities and ecosystem functioning. Although conceptual advances and technological breakthroughs in ecology have traditionally come hand-in-hand, the inclusion of genetic and cellular methods and processes in ecology has been slow. This is true even though ecological genetics is essential for our understanding of the distribution and interactions of organisms in space and time, and for predicting how populations respond to changes in their environment. Identifying the molecular mechanisms that underlie ecological processes is also of increasing interest to physiological ecologists.

At the other end of the scale, molecular biologists are making important advances in understanding the molecular and cellular processes required for cell division, photosynthesis and changes in gene expression that accompany development. Nonetheless, the functions of thousands of genes in animals, microbes and flowering plants identified by high-throughput genomics remain a mystery [1–3]. Furthermore, many ecological processes arise from complex, interacting systems that may not be explained by examining gene function alone; identifying which ecological processes key genes control is one challenge. Because the expression patterns and functions of current genes evolved in response to the natural world, studying molecular processes in their ecological context provides a perspective that is crucial for understanding organismal development and performance.

Here, we highlight opportunities and existing impediments for integrating ecological and molecular research. We begin by discussing current model species, including the need for increased ecological research that takes advantage of the genetic variation within those species, as well as an improved understanding of their ecology. We also discuss ecological attributes that could guide the choice of new model species. We then examine methods for identifying specific genes whose variation underlies diversity in ecologically relevant traits. We close with recommendations for the future, outlining some challenges and the infrastructure and training needed to promote a synergy between ecologists and molecular biologists.

Model systems and the diversity of life

The recent advances in molecular, cellular and developmental biology have been achieved primarily using a handful of model species (e.g. *Escherichia coli*, yeast *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, fruit fly *Drosophila melanogaster*, nematode *Caenorhabditis elegans*, corn *Zea mays*, mouse *Mus musculus* and zebrafish *Danio rerio*). Molecular techniques developed for these species provide new tools for answering a range of questions, from the molecular basis of evolutionary adaptation to the mechanisms of phenotypic plasticity, interspecific interactions and chemical communication. However, model species were selected on the basis of particular genetic and developmental features (e.g. clonal propagation, self fertilization and short generation times) and for ease of growth in the laboratory, rather than for their ecological or evolutionary importance or their applicability to ecological questions. Consequently, many ecologically important pathways, processes and structures are not currently represented in model organisms.

A more phylogenetically and functionally diverse group of model species will be needed to provide a mechanistic basis for some ecological processes. Many molecular tools developed using model systems can now be applied to other organisms, removing the technical barriers to examining biological mechanisms in nature. Acquiring a detailed genomic understanding

 0169-5347/02/$ – see front matter PII: S0169-5347(02)02571-5

Box 1. The 2010 project

A decade ago, plant biologists first established 'The Multinational Coordinated *Arabidopsis thaliana* Genome Research Project', a collaboration resulting in the complete sequence of the *Arabidopsis* genome [a]. The new 'Multinational Coordinating *Arabidopsis* 2010 Project' proposes to use these sequence data as a foundation for determining the functions, within their cellular, organismal and evolutionary contexts, of all ~25 000 *Arabidopsis* genes and the proteins that they encode by the end of 2010 [b,c].

The 2010 Project identifies eight key areas that would derive predictive power from a whole-systems approach to *Arabidopsis* biology [b]. At least three of them are central to plant ecology and evolution: enhanced understanding of the genetic basis of phenotypic plasticity, understanding of the genetic basis of plant evolution, and an understanding of interactions between plants and other organisms in their environment, up to the level of ecosystems. Fully understanding the adaptive functions of an organism requires an account of how genotype, phenotype and environment interact; abstracting plants from this context could impede a deeper understanding of the origin and maintenance of particular adaptations. Ecology at its various scales of investigation provides essential insight into such processes.

For incorporating knowledge gained from the genome projects, genes cloned from *Arabidopsis* (and from bacteria, fungi and animals) will provide the foundation for broad genetic knowledge in all plants. Between 48% and 60% of genes in *Arabidopsis* have counterparts in the other eukaryotic (nonplant) genomes sequenced to date, suggesting at least some highly conserved gene functions [a]. Similarities within higher plants will probably be much greater, although important genetic differences remain between groups, such as the monocots and dicots [d] and among species with C4 photosynthesis, nitrogen fixation and other important processes that are missing in *Arabidopsis*.

References

a *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815

b Chory, J. *et al.* (2000) National Science Foundation-sponsored workshop report, 'The 2010 Project'. Functional genomics and the virtual plant: a blueprint for understanding how plants are built and how to improve them. *Plant Physiol.* 123, 423–426

c Somerville, C. and Dangl, J. (2000) Plant biology in 2010. *Science* 290, 2077–2078

d van Dodeweerd, A.M. *et al.* (1999) Identification and analysis of homologous segments of the genomes of rice and *Arabidopsis thaliana*. *Genome* 42, 887–892

of each species in every ecosystem is unrealistic, but a judicious choice of new model species for ecological and evolutionary genomics would improve our understanding of life and its response to natural and human-induced disturbances. In addition, a larger set of model organisms could help ecologists examine phylogenetic constraints and organismal redundancy; convergent biochemical phenotypes and behaviors in different species might have different genetic bases [4], and the same gene might have more than one function in different organisms and environments.

This new range of model species should embrace diverse phylogenetic lineages, life-history strategies, and ecological and physiological attributes. For example, C3, C4 and Crassulacean acid metabolism (CAM) are the three pathways of plant photosynthesis. *Arabidopsis* and rice, the plant species for which genomic sequences are available, are both C3. New model species that use the C4 and CAM pathways could help biologists understand more fully the link from gene expression to the unique structures used by such species, and from structure to function. At the biogeochemical level, we know that many processes involve symbioses, including phosphorus acquisition (plants with mycorrhizal fungal partners) and nitrogen fixation (plants with bacterial partners). Therefore, key species of such interactions merit closer scrutiny. For animals, model species should encompass diverse trophic and developmental strategies, with special consideration given to keystone species.

The current set of model species could also be used more extensively in field research. An understanding of the adaptive significance of traits and gene functions will come more quickly by studying organisms in the environments that shaped their phenotypic evolution, because the fitness effects of genes depend upon the environment in which they are expressed. Mutations that appear innocuous in a benign laboratory environment might be harmful or helpful in nature depending on the ecological context. At higher scales, emergent properties occur at levels of organization that cannot be predicted from the molecular analysis of a single species [5,6]. Community-level interactions, for example, occur among multiple organisms living in the same spatial environment.

Box 2. Microarrays in ecology

Hundreds to thousands of genes can be assayed simultaneously for their expression levels using microarrays [a]. Thus, microarrays provide an integrated view of many different responses and have attracted the interest of ecologists wanting to investigate the molecular processes operating in their experiments. For example, the changes in gene expression of a plant responding to multiple factors, such as nitrogen level, water availability and herbivory, could be evaluated simultaneously on the same microarray. Also, knowledge of gene expression patterns in diverse environments will help elucidate the function of the many genes for which there is no known function.

Two examples illustrate the application of microarrays to ecological questions. Using a full genome microarray, Hihara *et al.* [b] identified changes in transcript levels in cyanobacterium *Synechocystis* spp. on transfer from low to high light conditions. In addition to genes previously implicated in high light acclimation, several novel genes were identified. In a second example in *Arabidopsis*, wounding and insect feeding could be distinguished by the transcript profile elicited by each treatment, supporting the notion that feeding by cabbage butterfly larvae species does not elicit or possibly suppresses nonspecific stress responses in plants [c].

In spite of their promise, microarrays have limitations for ecological research. To date, they have been constructed for a limited number of model species, such as humans, *Arabidopsis* and yeast, restricting their usefulness in the study of complex ecosystems. However, interspecies microarray trials suggest that unique arrays will not have to be constructed for every species. Using *Arabidopsis* microarrays, expression profiles for developing seeds from *Arabidopsis* and *Brassica napus* correlated well [d]. One caveat to interspecific microarray experiments is that data on novel genes that are not present in the model organism will be lacking. A general limitation of microarrays is that many regulatory changes (e.g. protein synthesis, enzyme activation and metabolite sequestration) do not occur at the level of mRNA accumulation and thus cannot be monitored with this technology. Appropriate tools for data analysis are still being developed, adding to the challenge of obtaining useful information from microarrays. Finally, their cost is high and must be balanced against the insights expected from their application to ecological problems.

References

a Brown, P.O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21(Suppl.), 33–37

b Hihara, Y. *et al.* (2001) DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* 13, 793–806

c Reymond, P. *et al.* (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12, 707–719

d Girke, T. *et al.* (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124, 1570–1581

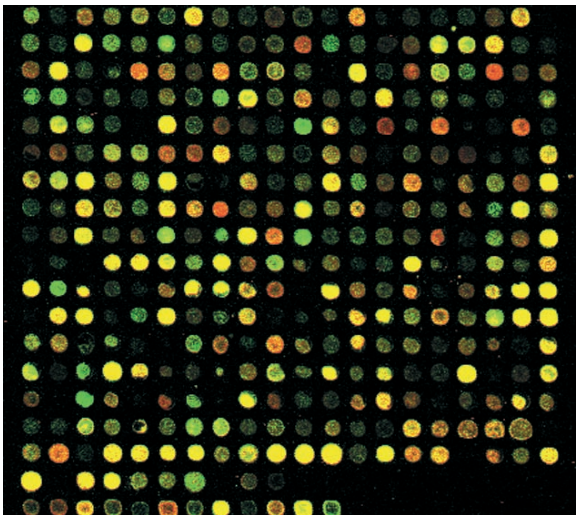

Fig. 1. False color image of a laser scan of a portion of an 11 500-element *Arabidopsis* two-color microarray hybridized with two cDNA samples (one from plants grown under ambient CO_2 levels and the other from plants grown under levels of 800 ppm CO_2). Green spots represent genes that are more highly expressed under ambient than in elevated CO_2 conditions. Red spots depict genes that are more highly expressed in high CO_2. Genes that are similarly expressed under the two conditions are in yellow. Contributed by D. Finkelstein, Expt ID 7562, Stanford Microarray Database.

However, theory and empirical work suggest that much of the variation seen at the molecular level (e.g. DNA and protein sequences) might be neutral or nearly so. Because fitness measurements of most genetic variation are made in a narrow range of laboratory conditions, it is unclear how often variation appears neutral because the appropriate environment has not been considered. Here, ecological knowledge of the conditions in which many species live has an important role in our understanding of the gene functioning of model systems. For variation that is neutral or nearly so, sequence data in combination with coalescent theory can sometimes be used by population ecologists to infer the dynamics of population demographics, although the models that are commonly used have some restrictive assumptions [7].

Hunting for ecological genes

One area where the merging of ecology and molecular biology has begun is identifying genes of ecological and evolutionary relevance. *Arabidopsis* geneticists, for example, have identified the pathways that control plant responses to light quality, photoperiod, temperature and pathogens. By characterizing the genes that encode constituent members of such pathways, molecular biologists provide ecologists with a mechanistic understanding (at the molecular scale) of the phenotype of the organism and its response to ecological signals. Using genes from well-characterized models to isolate homologous genes from other species is increasingly routine, and gene functions in new organisms can then be compared using this method.

Natural selection acts on variation in phenotypes, and understanding the origins and maintenance of this variation is the focus of ecological genetics. Some ecologists might not be as interested in every gene or protein underlying a trait, but rather in identifying specific genes whose variation underlies diversity in ecologically relevant traits among individuals, populations or species. The modern tools of molecular genetics and genomics, combined with the use of polymorphic molecular markers, provide new approaches to finding such genes. Once identified, the role of genetic diversity in ecology can be tested in ways that were previously impossible.

Fine-mapping and map-based cloning of genes

In recent years, ecological geneticists have succeeded in coarse mapping of genes that underlie variation in ecologically relevant traits, including pollinator attraction in *Mimulus* [8], floral morphology in *Arabidopsis* [9], skeletal morphology in sticklebacks [10] and behavior in honeybees [11]. Coarse mapping of continuously distributed phenotypes (quantitative traits) can localize genes to ~10 CENTIMORGANS (see Glossary) [12] (although *Arabidopsis* has ~500–600 genes in such an interval). Fine-scale mapping, applying closely spaced markers to large populations, can help refine gene positions to the point where identifying specific genes that underlie variation in an ecological trait is feasible [13]. In fine mapping, large populations are established to create lines that show recombination between closely spaced molecular markers and the gene of interest. Closely spaced markers can be used to identify rare recombinant events, ultimately helping to identify and isolate the gene of ecological interest. In addition to providing a wealth of markers, whole-genome sequences define the ultimate physical guide to gene identification via map-based cloning.

Candidate-gene association studies

This approach uses functional information on specific proteins to identify candidate genes that are responsible for variation in ecological phenotypes. A researcher obtains phenotypic information from individuals of a population and identifies their genotypes with respect to molecular polymorphisms found in the candidate genes [14]. In systems that are amenable to genetic manipulation, the function of candidate genes can be further studied by producing individuals with particular variants of the genes in otherwise genetically uniform backgrounds. Such association studies could, in principle, provide evidence that polymorphisms at a specific candidate gene are associated with an ecological phenotype. The success of this approach depends on the extent of linkage disequilibrium in the genome and the age of mutations that cause the phenotype. By using association studies, geneticists have identified genes involved in naturally occurring variation in *Drosophila* bristle number [14] and in maize flowering time [15]. Linkage disequilibrium mapping is also used for mapping human disease genes [16], including cardiovascular disease and other illnesses.

The utility of such methods for ecological studies depends on the availability of genetic resources for the study species and the time invested. In the short term, many of these studies will be fruitful in a few model genetic organisms for which a strong base of knowledge, genetic tools and resources are available. For other species, a commitment of several years by a community of investigators might be needed to develop the genetic resources and knowledge for such

Acknowledgements
We gratefully acknowledge support from NSF (IBN-0129616 to R.B.J. and C.R.L.) and from USDA, the National Institute of Global Environmental Change/DOE, and the Andrew W. Mellon Foundation. We especially thank participants of the NSF-sponsored workshop Linking Molecular Insights and Ecological Research (August 2001) who co-authored the report on which some of this paper is based, including D. Ackerly, F Assaad, C. Field, L. Jackson, T. Juenger, R. Mauricio, H. Mooney, A. Schmitt, C. Somerville, V. Walbot, and J. Verbeke (NSF observer). We also thank participants of an earlier symposium Physiological ecology: linking the organism to scales above and below (J.S. Coleman, T.E. Dawson and R.B. Jackson, organizers, Ecological Society of America 2000). D. Robinson, T. Mitchell-Olds and two anonymous reviewers provided helpful suggestions.

Box 3. Molecular techniques and ecological questions

We present two examples to help illustrate the promise of using molecular approaches in ecology (Fig. I). The first is the use of ANTISENSE GENE technology to understand plant responses to the environment. Results from studies of rice and tobacco plants transformed with an antisense rbcS (the small subunit of Rubisco, ribulose-1,5-bisophosphate carboxylase/oxygenase) address questions of photosynthetic downregulation under conditions of elevated CO_2, the subject of many ongoing field studies [a]. Such molecular approaches can be used to manipulate the concentrations of functional Rubisco along a range of leaf nitrogen concentrations [b,c] and provide a mechanistic framework for understanding why some species downregulate photosynthesis when exposed to long-term changes in levels of CO_2. In the study illustrated in Fig. I, antisense line 77 (AS 77) had 65% wild-type Rubisco content and a second line, AS-71, had 40%. Antisense plants were smaller than wild-type plants when grown at ambient CO_2 levels but had similar biomass at elevated CO_2 levels (100 Pa; Fig. Ia,b). Such studies illustrate not only the utility of molecular approaches, but also the many compensatory mechanisms that operate in plants (in this case, changes in leaf area and allocation) that make predicting the organismal consequences of single-gene changes difficult. Reproduced, with permission, from [c]

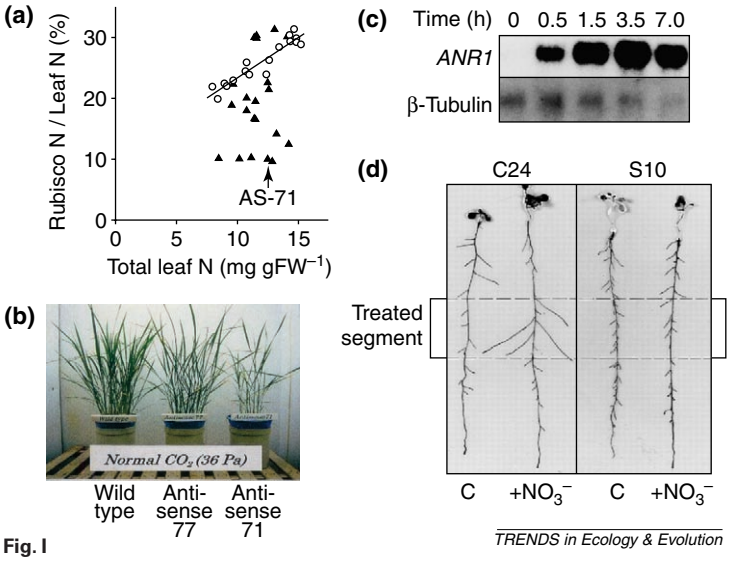

Fig. I

The second example is the use of genetically engineered plants to understand adaptive plastic responses. Many ecological studies have examined the factors that control root proliferation for water and nutrients [d], and new molecular studies shed light on the mechanism of such responses. A NO_3^--inducible gene (*ANR1*) in the MADS BOX family of transcription factors was recently shown to help control root proliferation and developmental plasticity in *Arabidopsis* [e]. *ANR1* is rapidly induced by NO_3^- within a matter of hours (Fig. Ic). In control plants (C24), lateral root growth in a NO_3^--enriched segment was stimulated two- to threefold, but plants in the *ANR1*-repressed line (S10) were insensitive to locally supplied NO_3^- (Fig. Id). The stimulation of lateral root elongation by localized applications of NO_3^- is dependent on expression of the *ANR1* gene. Understanding the significance of such genetic controls for plants in the field will require studies that integrate competition for nitrate among individuals in populations and communities. Reproduced, with permission, from [e].

References

a Wand, S.J.E. *et al.* (1999) Responses of wild C_4 and C_3 grass (Poaceae) species to elevated atmospheric CO_2 concentration: a meta-analytic test of current theories and perceptions. *Global Change Biol.* 5, 723–741

b Fichtner, K. *et al.* (1993) Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with 'antisense' rbcS: relationship between photosynthetic rate storage capacity, biomass allocation and vegetative plant growth at three different nitrogen supplies. *Planta* 190, 1–9

c Makino, A. *et al.* (2000) Whole-plant growth and N allocation in transgenic rice plants with decreased content of ribulose-1,5-bisphosphate carboxylase under different CO_2 partial pressures. *Aust. J. Plant Physiol* 27, 1–12

d Robinson, D. *et al.* (1999) Plant root proliferation in nitrogen-rich patches confers competitive advantage. *Proc. R. Soc. Lond. Ser. B* 266, 431–435

e Zhang, H. and Forde, B.G. (1998) An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root structure. *Science* 279, 407–409

studies. Other caveats are also relevant. Some species can be readily transformed, including those in the Brassicaceae and the Solanaceae, but most have never been, or are more difficult to transform (e.g. grasses and most animals). Woody plants, with longer life cycles and greater need for growing space, are often not practical. Genetic mapping methods work best in organisms that can be inbred, but methods have been developed for outcrossing organisms. For the future, when high-throughput sequencing methods become sufficiently inexpensive, genetic mapping could give way to physical mapping approaches.

Arabidopsis and other examples

The complete sequencing of the *Arabidopsis* genome [2], along with the recent sequencing of the indica- [17] and japonica- [18] type rice genomes, provides an opportunity for biologists to understand how key genes and groups of genes control interactions among plants, pathogens and the environment [19,20]. Nonetheless, molecular tools alone will be unlikely to predict the complex interactions of plants growing in the wild. A synergy between molecular biology and ecological and evolutionary research will be more successful.

Ecologists can use new molecular tools to help understand the genetic controls and limitations that influence organism distributions and environmental interactions. Within the many accessions of model species, there is extensive phenotypic and genotypic variation to be exploited [21,22], including traits that are central to ecology, such as water-use efficiency [23], enzyme activity [24], light sensitivity [25] and nutrient uptake [26]. Increased use of *Arabidopsis* and rice in greenhouse and field research will enhance knowledge of the physiological controls of interactions with the environment.

In addition to complete genome sequences, CDNA and DNA sequence information is already available electronically for many species. At present, >100 000 expressed sequence tag (EST) sequences are available for each of 17 species, and DNA sequences are available for >55 000 different organisms in GenBank [27]. The genomes of a few other plant species will be sequenced in the next decade, with sunflower, banana, lettuce, lotus and alfalfa already underway [28]. However, in the short term, the number of organisms with fully sequenced genomes will be limited, and ecologists will need new ways to apply molecular tools to the range of organisms with which they work. As additional genes are cloned and their functions identified in model organisms, potential ORTHOLOGS in other species will be identifiable through data base searches and molecular methods. The corresponding genes could then be cloned using sequence information [27,29]. As an example of the kind of gene function information that will be available from plants, the *Arabidopsis* 2010 project seeks to determine the function of all ~25 000 *Arabidopsis* genes by the year 2010 (Box 1).

Although extensive physiological and ecological information obtained using model organisms will

Box 4. Needs and recommendations

The objectives outlined in this article will require new training programs and resources. Some suggestions and examples include:

Intellectual capital: Training opportunities for interdisciplinary approaches

- Postdoctoral fellowships for training scientists jointly in ecology and molecular biology
- Training grants and training centers to provide graduate students with cross-disciplinary research opportunities, and to facilitate training of more senior scientists
- Dissertation improvement grants of sufficient size to cover the costs of molecular studies
- Efforts to coordinate training programs within and among institutions, including websites that collate courses bridging molecular biology and ecology, and joint training in both ecology and/or evolution and cell and/or molecular biology at the undergraduate and graduate levels
- Short summer courses for interdisciplinary training, including hands-on training in molecular techniques for ecologists and ecological techniques for molecular biologists
- Establish links with research networks concerned with organismal diversity, including programs such as 'The Tree of Life' (http://www.nsf.gov/pubs/2002/nsf02074/nsf02074.html#DESC) that seek to classify organisms; maintain training in animal, plant and microbial taxonomy

Enabling resources and infrastructure: Developing the tools and infrastructure for integrating molecular and ecological research

- Emphasize whole-genome approaches for analysing function and fitness, community composition and ecosystem functioning; relevant tools include microarrays, proteomics, QTL ANALYSIS and confirmation studies
- Study model organisms and their wild relatives in the field
- Exploit natural variability in traits; known accessions could be resampled to obtain additional information about their ecology and habitat
- Harness the benefits of diversity by studying a broader range of organisms across habitats, life histories, structures and physiological pathways; characterize full-length expressed sequence tags for such species and complete sequencing and genomic analysis of a subset as sequencing speed increases and costs decrease (http://www.nerc.ac.uk/funding/thematics/envgen/)
- Create comprehensive collections of DNA, tissues and organisms from the species chosen for diversity
- Use improved controlled-environment facilities for molecular studies
- Use stock centers for storage and distribution of materials
- Improve capabilities for high-throughput analyses of molecular and ecological samples, including high-throughput genotyping, phenotyping and physiology using automated approaches, as well as precise and rapid phenotyping, such as rapid seed weighing and image analysis tools

apply to all higher organisms, not all orthologs in other species will function identically. Ecologists need to consider how best to apply what can be learned using model organisms to other species. For example, how trees modify the expression of genes controlling their structure and physiology (e.g. for water transport) will not always be a simple extrapolation from studies with *Arabidopsis* or tobacco plants [30,31].

Another challenge for ecologists is how to link information about gene expression with process-based information. Using soil nitrification as an example, changes in the expression of ammonia oxidizer genes affected by ecosystem management [32] might reveal important differences in the diversity and identity of microbial communities. Nonetheless, ecologists interested in nitrate fluxes must still find a way to predict what happens at the ecosystem scale using the community-based information, otherwise that information is only qualitatively useful to them.

High-throughput technologies in genomics, PROTEOMICS and METABOLOMICS are being applied to model organisms as exploratory tools. These technologies can be used to determine when, where in the organism and in response to what environmental and developmental stimuli genes are expressed and molecules occur [33]. Such technologies need to be adapted more broadly to ecological research. Of them, the genomics technology of MICROARRAYS (Box 2) is currently the most accessible. Microarrays can be used to examine changes in expression for thousands of genes in parallel and thus lend themselves to exploring the network of interacting responses to environmental stimuli. For example, a researcher might alter NO_3^- or CO_2 availability and then examine the cascade of altered gene expression over time (Fig. 1). Genetically distinct individuals can also be examined, either from the same population or across a geographical gradient. DNA-fragment arrays can, in theory, be made for any species, including those whose genomes have not been sequenced. However, microarray experiments present significant statistical and production challenges [34,35] (Box 2), and there are technical limitations to follow-up experiments in some organisms (e.g. inability to transform).

Opportunities for the future

Genetic transformations and mutational knockouts of model organisms can generate traits and trait combinations that do not necessarily occur in nature. They thus provide a means for identifying genes that play important ecological roles. Capitalizing on these discoveries, ecologists and evolutionary biologists have recently addressed a range of questions, including the fitness costs of heat-shock protein expression in *Drosophila* [36], the adaptive significance of phytochrome-mediated shade avoidance in plants [37] and the costs of herbicide resistance in the field [38] (Box 3).

Molecular markers and DNA sequencing have revolutionized our understanding of community organization, including the discovery of surprising fungal and microbial diversity [39] (with whole new groups discovered, such as the Acidobacterium). Markers for species identification have been used to trace the roots of plants in the soil [40], to determine the host distribution of herbivorous insects and mycrorrhizae [41], and to discern the natural distributions of microbial species that are resistant to laboratory culture. Highly polymorphic markers have provided the tools to determine patterns of gene flow and population subdivision in marine organisms, such as whales [42], where barriers to movement are hard discern. They have also yielded fundamental information in conservation biology and agriculture for analysing pedigree and mating systems (e.g. of cheetahs [43]), stock identification, and the tracing of ancestral sources and spread of introduced species, such as zebra mussels and annual cheatgrasses [44]. High-throughput genotyping with molecular markers offers the possibility of estimating pairwise relatedness in the field and therefore of performing conventional

Glossary

Antisense gene: a genetically engineered gene that produces a RNA transcript complementary to a naturally occurring (sense) mRNA. In sufficient quantity, antisense RNA complexes with sense mRNA triggering the degradation of both sense and antisense mRNAs, decreasing the level of protein normally produced.
cDNA: complementary or copy DNA produced by reverse transcription of mRNA.
Centimorgan: a standard measure of recombination distance in a genetic map.
EST: expressed sequence tag: the cDNA sequence of the coding region of a gene expressed in an organism (derived from mRNA and devoid of introns). EST clones are used as probes for genomic sequences and in microarrays.
MADS box: a conserved sequence motif of 168 bp occurring in several transcription factors (and named for the first four members of the family: *MCM1, AGAMOUS, DEFICIENS* and *SRF* – serum response factor). MADS domain factors play important roles in developmental processes such as flower morphogenesis.
Metabolomics: studies of the metabolome, the entire metabolite complement of an organism.
Microarray: a glass slide or other substrate onto which thousands of characterized EST clones, each representing a different gene, have been individually spotted. Oligonucleotide-based microarrays, an alternate technology, consist of ~25-mer oligonucleotides synthesized *in situ* on silicon wafers. Each gene is represented by a set of oligonucleotides. Gene expression from a sample is determined by visualizing a population of fluorescently labeled mRNA-derived cDNAs or copy mRNAs that have hybridized to the DNAs (i.e. EST clones or a set of oligonucleotides) of the microarray. The intensity of the fluorescent signal associated with each DNA (gene) on the microarray is a reflection of the abundance of mRNA for that gene in the original sample.
Near-isogenic (lines): two or more lines that differ only for one gene and closely linked adjacent genes.
Ortholog: genes in two species that are similar (and might have a similar function) owing to their having been inherited from a common ancestor (i.e. they have the same evolutionary origin).
Proteomics: studies of the proteome, the entire protein complement of an organism.
QTL analysis: quantitative trait loci (QTL) analysis maps the genomic location of one or more genes that affect a trait for which phenotypic variation among genotypes is continuous rather than discrete.

quantitative genetic analyses in natural settings. The cost of many high-throughput techniques is decreasing quickly, opening the door to ecological studies with large numbers of samples. However, considerable time is still needed to process large numbers of samples before the high-throughput methods are used. Another impediment is the lack of software tools for exploring and analysing large data sets.

To maximize the benefits of integrating molecular and ecological research, new resources and infrastructure will be needed (Box 4). For the future, high-throughput genetic markers will be applied to population studies, including questions of dispersal and migration. Marker-assisted selection to create NEAR-ISOGENIC organisms can be used in field studies of adaptive characteristics. High-throughput phenotyping and physiology, as well as gene expression profiling using microarrays applied to a diverse set of organisms, will broaden our understanding of gene functioning in nature. Finally, molecular characterization of organismal diversity and activity will enable ecologists to address important ecosystem processes, including microbial interactions affecting plant nutrient uptake, the degradation of toxic chemicals and the effect of greenhouse gas emissions.

References

1 Goffeau, A. *et al.* (1996) Life with 6000 genes. *Science* 274, 546
2 *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
3 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
4 Gerhart, J. and Kirschner, M. (1997) *Cells, Embryos and Evolution*, Blackwell Science
5 Lidicker, W.Z. (1999) Responses of mammals to habitat edges: an overview. *Landscape Ecol.* 14, 333–343
6 Mulkey, S.S. *et al.* (1996) Plant physiological ecology of tropical forest canopies. *Trends Ecol. Evol.* 11, 408–412
7 Emerson, B.E. *et al.* (2001) Revealing the demographic histories of species using DNA sequences. *Trends Ecol. Evol.* 16, 707–716
8 Bradshaw, H.D. *et al.* (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* 376, 762–765
9 Juenger, T. *et al.* (2000) Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. *Genetics* 156, 1379–1392
10 Peichel, C.L. *et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature* 414, 901–905
11 Robinson, G.E. (1999) Integrative animal behaviour and sociogenomics. *Trends Ecol. Evol.* 14, 202–205
12 Lynch, M. and Walsh, J.B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer
13 Remington, D.L. *et al.* Map-based cloning of quantitative trait loci: progress and prospects. *Genet. Res.* (in press)
14 Mackay, T.F.C. (2001) Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* 2, 11–20
15 Thornsberry, J.M. *et al.* (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* 28, 286–289
16 Weiss, K.M. and Clark, A.G. (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet.* 18, 19–24
17 Yu, J. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92
18 Goff, S.A. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100
19 Stahl, E.A. *et al.* (1999) Dynamics of disease resistance polymorphism at the *RPM1* locus of *Arabidopsis*. *Nature* 400, 667–671
20 Mitchell-Olds, T. (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* 16, 693–700
21 Pigliucci, M. (1996) How organisms respond to environmental changes: from phenotypes to molecules (and vice versa). *Trends. Ecol. Evol.* 11, 168–173
22 Alonso-Blanco, C. and Koornneef, M. (2000) Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends Plant Sci.* 5, 22–29
23 Nienhuis, J. *et al.* (1994) Variance for water-use efficiency among ecotypes and recombinant inbred lines of *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot.* 81, 943–947
24 Mitchell-Olds, T. and Pedersen, D. (1998) The molecular basis of quantitative genetic variation in central and secondary metabolism in *Arabidopsis*. *Genetics* 149, 739–747
25 Maloof, J.N. *et al.* (2001) Natural variation in light sensitivity of *Arabidopsis*. *Nat. Genet.* 29, 441–446
26 Krannitz, P.G. *et al.* (1991) Correction for non-linear relationships between root size and short term P_i uptake in genotype comparisons. *Plant Soil* 133, 157–167
27 Benson, D.A. *et al.* (2000) GenBank. *Nucleic Acids Res.* 28, 15–18
28 Cyranoski, D. (2001) Japanese legume project may help to fix nitrogen problem. *Nature* 409, 272
29 Somerville, C. and Somerville, S. (1999) Plant functional genomics. *Science* 285, 380–383
30 Sederoff, R.R. (2000) Tree genomes: what will we understand about them by the year 2020 and how might we use that knowledge? In *Forest Genomics and Sustainability* (vol. 63) (Mátyás, C., ed.), pp. 23–30, Kluwer
31 Jackson, R.B. *et al.* (2000) Root water uptake and transport: using physiological processes in global predictions. *Trends Plant Sci.* 5, 482–488
32 Webster, G. *et al.* (2002) Grassland management regimens reduce small-scale heterogeneity and species diversity of β-proteobacterial ammonia oxidizer populations. *Appl. Environ. Microbiol.* 68, 20–30
33 Kehoe, D.M. *et al.* (1999) DNA microarrays for studies of higher plants and other photosynthetic organisms. *Trends Plant Sci.* 4, 38–41
34 Knight, J. (2001) When the chips are down. *Nature* 410, 860–861
35 Gibson, G. (2002) Microarrays in ecology and evolution: a preview. *Mol. Ecol.* 11, 17–24
36 Feder, M.E. *et al.* (2000) Evolutionary physiology. *Annu. Rev. Ecol. Syst.* 31, 315–341
37 Schmitt, J. *et al.* (1999) Manipulative approaches to testing adaptive plasticity: phytochrome-mediated shade-avoidance responses in plants. *Am. Nat.* 154, S43–S54
38 Cuthbert, J.L. (2001) Comparison of the performance of bromoxynil-resistant and susceptible near-isogenic populations of oilseed rape. *Can. J. Plant Sci.* 81, 367–372
39 Gordon, D.A. *et al.* (2000) Origin and phylogeny of microbes living in permanent Antarctic lake ice. *Microb. Ecol.* 39, 197–202
40 Jackson, R.B. *et al.* (1999) Ecosystem rooting depth determined with caves and DNA. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11387–11392
41 Bruns, T.D. *et al.* (1998) A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. *Mol. Ecol.* 7, 257–272
42 Avise, J.C. (1998) Conservation genetics in the marine realm. *J. Hered.* 89, 377–382
43 Kelly, M.J. (2000) Lineage loss in Serengeti cheetahs: consequences of high reproductive variance and heritability of fitness on effective population size. *Conserv. Biol.* 15, 137–147
44 Novak, S.J. and Mack, R.N. (2001) Tracing plant introduction and spread: genetic evidence from *Bromus tectorum* (Cheatgrass). *Bioscience* 51, 114–122