

# Genomics for the ecological toolbox

Michael A. Thomas<sup>1</sup> and Rebecca Klaper<sup>2</sup>

<sup>1</sup>Department of Biological Sciences, Campus box 8007, Idaho State University, Pocatello, ID 83209, USA

<sup>2</sup>Great Lakes WATER Institute, University of Wisconsin – Milwaukee, 600 East Greenfield Ave, Milwaukee, WI 53204, USA

**Genomics technologies have expanded the types of question that can be addressed in human genetics and health and in fields such as ecology. Genome-scale approaches provide ways to examine physiological changes that occur when a pathogen invades a host, the response of an organism to a change in its environment, and the way in which changes in the microbial community affect ecosystem function. Here, we examine successful applications of genomics to ecological science to date and describe ways that classic ecological research disciplines might benefit from genomic approaches. We also address some of the challenges of using this methodology, and discuss how ecological researchers embracing these approaches enhance its effectiveness in applications such as gene hunting and gene expression analysis.**

Genomics technologies have previously seemed inaccessible to ecologists owing to the focus of such techniques on human health. Here, we demonstrate that these technologies can provide a tool with which to address ecological questions that involve the genetic basis of complex phenotypes, especially those responding to interactions between species. Currently, researchers use these techniques to address questions about the genetic mechanisms of ecological responses to interactions with other species, the environment and environmental perturbations. Ecologists are now beginning to take advantage of the potential power of these tools to explore (and, in some cases, re-examine) the central questions and concepts of their field. Recently, this emerging discipline has received an increasing amount of attention [1–4] and has been the focus of several conferences\*. However, further discussion among ecologists and genomics researchers will be necessary to advance the field of ecological genomics.

## Genomic tools in search of ecological problems

Ecologists interested in using genomic tools currently have two options: (i) use existing model organisms that already have significant genomic resources available. Theoretical questions that can be answered using model systems are most amenable to this option. This approach is already finding broad application, especially in microbial ecology

and plant physiology; and (ii) use tools that are currently available and apply them to a species of interest. Our emphasis here is on the second option, given that it is the approach that most ecologists might favor. Many of the approaches borrowed from more traditional research applications of genomics for use in ecological genomics require substantial existing resources. Such functional genomics approaches provide a picture of a process being perturbed rather than a simple list of genes (Box 1). For ecologists, this genetic perspective on a perturbation provides an integrated understanding:

- A group of genes is implicated;
- The pathways and systems in which those genes operate are uncovered;
- Variation in those genes (and implications for the systems) can be compared among populations;

### Box 1. Beyond the transcriptome

Most of the work described here uses the transcriptome, or functional genomics. The transcriptome encompasses all of the expressed genes of an organism, in the form of mRNA. There are typically multiple transcript isoforms per gene owing to alternative splicing, so the transcriptome is much larger than the genome itself. A given cDNA library represents all transcripts (reverse transcribed into complementary, or cDNA) produced by a specific tissue, conditions, and point in time (which is why there are so many libraries for each of the major model organisms).

Whereas transcriptome investigations were the first large-scale technologies used to look at cell and organismal function, many steps lie between transcript abundance and cell and organismal functional. There is little biological variation in transcript abundance reflected in the corresponding variation in the amount of a functional protein [44]; that is, any given genetic element that is differentially expressed in a microarray experiment is more likely to have no functional significance than researchers tend to acknowledge. Because of this, genomics approaches might give way to proteomics, in which protein abundance is directly assessed.

The proteome comprises all active peptides in a cell. Researchers are developing methods to characterize peptides involved in key processes and interactions with other organisms and the environment. Much of the work to date has been comparative [45], and the promise of proteomics for ecologists awaits progress in other disciplines. High-throughput proteome investigations attempt to assign process and function on a large scale, identifying the presence of proteins in particular physiological conditions or identifying interaction among proteins. These studies place the experimenter one step closer to the phenotype, removing some of the ambiguity and assumptions that are inherent in transcriptome-level studies, while continuing to provide a connection to genetic variation. Researchers can combine information about protein and transcript abundance with profiles of metabolites to develop an integrated understanding of state of a biological system under experimental conditions. This has been recently accomplished in studies of secondary metabolites in *Arabidopsis* [46].

Corresponding author: Michael A. Thomas (mthomas@isu.edu).

\* The Gordon Research Conference 'Evolutionary and Ecological Functional Genomics' (August 2003; <http://www.grc.uri.edu/programs/2003/evolecol.htm>); 'At the Crossroads of Genomics and Ecology: The Potential for a Canary on a Chip' Symposium at the Ecological Society of America Annual Meeting (August 2003; <http://www.esa.org/savannah>); and the 'Genes in Ecology, Ecology in Genes' Symposium hosted by the Ecological Genomics in Kansas NSF EPSCoR grant program (November 2003; <http://www.ksu.edu/ecogen>).

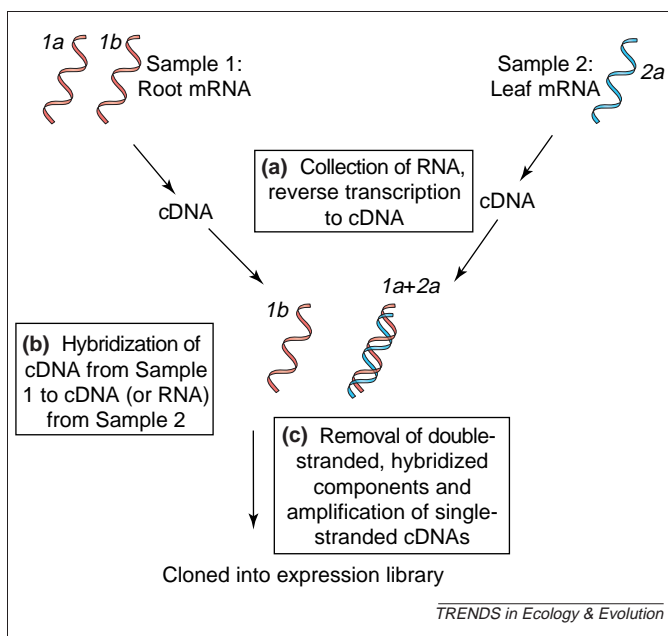
Available online 24 June 2004

- Protein and metabolite abundance can be profiled and related to transcript (expressed gene) abundance to verify and extend experimental findings;
- Associations of active genes and systems to other perturbations in the study organism and other organisms can be explored (e.g. the response to herbivory might also be linked to responses to UV light or drought);
- Novel hypotheses can be formulated and tested.

Several methods have been developed during the past decade to identify differences in gene expression across treatments, stages, tissues, or environmental conditions, including complementary DNA (cDNA) subtraction [5,6], SAGE (serial analysis of gene expression) [7] and microarrays [8].

#### cDNA subtraction

The basic concept of cDNA subtraction involves extracting mitochondrial RNA (mRNA) from two tissues, two organisms, the same tissue from organisms exposed to two environmental conditions, or similar comparisons (Figure 1). These expressed genes are then reverse transcribed into a more stable cDNA form, and the two samples are then subtracted, whereby an excess of single-stranded cDNA from one sample is exposed to the second treatment so that the matching cDNAs hybridize (anneal together). The remaining single-stranded cDNAs are unique to a particular treatment. These are isolated and



**Figure 1.** Simplified description of cDNA subtraction. (a) RNA from the two samples of interest is extracted and converted to complementary DNA (cDNA) by reverse transcriptase. (b) Single-stranded cDNA from one treatment is exposed to the cDNA from the other sample (or to RNA). The cDNAs that are the same in both treatments will hybridize (anneal) to each other and these double-stranded forms can be removed or selected against by only amplifying the single-stranded forms using small genetic primers and PCR amplification (c). In the example illustrated here, using an excess of Sample 2 in the hybridization step ensures that all of the mRNA that is expressed in both the leaves and roots is hybridized and removed. This leaves single-stranded cDNA that is unique to either Sample 1 or Sample 2. Different methods are used to select for one or the other, including labelling the cDNA from Sample 1 before step (b) with a primer so that, after hybridization of the similar cDNAs, the single-stranded cDNAs that are left can be amplified using PCR. An alternative is to use the single-stranded cDNAs to probe a library only from Sample 1.

used to either probe a cDNA library or create a cDNA library by ligation into plasmids and replication within bacteria. Macro- or microarrays can be created from the clones developed through several of the subtraction techniques. Hundreds to thousands of cDNAs can be spotted onto a slide or membrane and used as a tool to detect differences across individual samples in the same experiment or future related experiments. There are many companies that conduct these subtractions and construct arrays; using these services might be a viable option for ecologists to accelerate production of relevant transcriptome material for further experimentation. The cDNAs can also be sequenced to attempt to identify the function from known sequences in public databases. For a review of different subtraction methods see [9].

#### SAGE techniques

In the absence of existing genome data, exciting (and relatively inexpensive) discoveries can be made by using creative approaches. For example, a recently developed modification of the SAGE approach [7] requires no existing genetic database for the organism of interest. In SAGE, cDNA is reverse transcribed from mRNA isolated from the organism of interest and a 'tag' sequence of ~15 bp is excised from each cDNA. To identify the cDNAs, the short tags are sequenced and searched against genes [or annotated expressed sequence tags (ESTs)] in a database corresponding to the species examined. From the relative abundance of tags, the researcher can describe the gene expression profile for the tissue assayed. However, if there is no gene database for the species being examined, it might be difficult to identify the 15-bp tags. Matsumura and colleagues have developed a 'SuperSAGE' approach that results in a 25-bp SAGE tag [10]. This tag is large enough to use as a PCR primer to amplify the cDNA from which it was derived (which is then sequenced). The amplicon sequence is suitable for general gene database searches and provides the opportunity to identify the genetic element. The SAGE family of methods provides a quantitative assay of global gene expression in genetically uncharacterized organisms.

#### Microarray techniques

Gene expression experiments using microarray techniques are a way of determining the identity of every genetic element expressed in particular tissues under specific experimental condition [8]. Given that microarray experiments are comparative by design, the experimenter is provided with information about which genes are differentially expressed (i.e. expression is increased or decreased in one treatment relative to the other). The experimental conditions can be different tissues in the same species, the same tissue experiencing two treatments, the same tissue in two different species, and so on. In a typical microarray experiment, mRNA from each of the two tissues are extracted, reverse transcribed to cDNA, tagged with different fluorescent markers, and hybridized to an array containing thousands of known probes (composed of either oligonucleotides or cDNAs, depending on the type of array). Relative rates of hybridization are detected by comparing intensities from the two fluorescent

markers. Once a group of differentially expressed genes is identified, appropriate follow-up and verification experiments are conducted; these steps provide crucial quality control.

The value of microarrays is that thousands of genes can be examined in a single experiment, providing a list of genes or pathways for further study with more specific approaches. However, this feature of gene expression data is also one of its primary difficulties. For example, how should one interpret a situation in which 275 genes have increased expression and 182 genes have decreased expression (especially considering that many of these differentially expressed genes are unlikely to be involved in process being investigated)? To address this complication, approaches have been developed that determine which systems and processes are altered by the experiments, rather than focusing on individual genetic elements [11].

For ecologists, microarray difficulties are also compounded by the requirements of microarray research, primarily the amount of genetic information that must already exist for the experimental organisms, the variability among organisms in a population, and the cost of the experiments. For each element on the array, information about its sequence, identity, molecular function, biological role and other characteristics, must be known. There are few commercial array options available for ecologically important organisms, much less arrays with annotated elements. Creating an array is also expensive and requires specialized equipment. If an array is commercially available, the cost of conducting the experiment can be covered by a typical NSF grant budget, with certain steps contracted out to core facilities.

Affordable commercial arrays are largely available only for model organisms with sufficiently characterized genomes, such as *Rattus* and *Arabidopsis*. An alternative for researchers studying non-model organisms is to use cross-species hybridizations to microarrays designed for other species [12,13]. In this approach, an array for the model organism that is an evolutionary relative of the target species is treated with mRNA extracted from that species. Differentially expressed genes are interpreted in the same way as with same-species microarray experiments, using the differentially expressed genes as an indication of the systems and pathways that are being expressed in the organism of interest. The assumption is that orthologous pairs (genes sharing homology from a speciation event and, typically, with similar sequence and function) will hybridize, whereas hybridizations involving paralogs (homologs from a gene duplication event, often with lower sequence similarity and less conserved function) and nonhomologs will be minimal, or at least no greater than in same-species experiments. However, the degree of sequence divergence at which reliability decreases to unacceptable levels is unknown. For example, using a nylon human cDNA microarray containing 4324 genetic elements, Moody *et al.* hybridized pig skeletal muscle RNA and reported reproducibility rates similar to using human RNA on the same array [12], in spite of the human–pig minimum divergence time of 65 million years ago (MYA) [14,15].

Although commercial applications are beginning to realize the power of this approach (<http://www.agilixcorp.com/gencompass.php>), much work remains to determine the expected error rates from cross-species microarray experiments before ecologists can begin extracting mRNA from their research organisms. Because we do not fully understand microarray hybridization [16], it is difficult to create a model that enables us to appreciate the sources of error. Assuming that sequence similarity is the primary component of hybridization success, we can predict the amount of sequence divergence that would make a given transcript less likely to hybridize with its cross-species ortholog compared with a similar but non-orthologous gene on that array. Whereas the time of human–pig minimum divergence was 65 MYA [14,15], that for the fruit fly *Drosophila* and the apple maggot fly *Rhagoletis* (Box 2) was 55 MYA [17]: does this mean that *Rhagoletis* researchers can assess gene expression differences between host races using readily available *Drosophila* arrays as successfully as researchers comparing pig versus human microarrays [12]? Perhaps so, but expected error rates first need to be estimated and verified.

### Box 2. *Rhagoletis*, meet *Drosophila*

The study of speciation of the apple maggot fly *Rhagoletis pomonella* has a long history [47–49] and continues to generate substantial contributions (and controversy) to our understanding of ecology and evolution [50–53]. *Rhagoletis* is the ‘poster child’ for sympatric speciation: subtle changes in pupal diapause depth, host preference and oviposition-related behavior [47,49] have led to the emergence of *Rhagoletis* ‘host races’, even within the ranges of sister host races. Host race formation, by definition, is the first stage in sympatric speciation [48,53]. *Rhagoletis* host race formation began by taking advantage of genetic diversity in *Rhagoletis* for traits involved in host selection when the fly shifted from using native hawthorn *Crataegus* fruit to using introduced apple *Malus* fruit [53]. Although much work has described genetic markers that enable the inference of host race development [47,52], we still lack fundamental information about the genetic basis of speciation, the identity of genes involved and the intensity of selection acting on those genes [51].

The fully sequenced and annotated genome of *Drosophila* makes it a useful system for the study of *Rhagoletis*, just as it does for complex human traits [54–56], even though the minimum divergence time between insects and human is great (~933 MYA [57]). More insect-like traits are of interest to *Rhagoletis* biologists, such as *Drosophila* mutations that affect behavior (e.g. courtship, locomotion and larval hatching), including *homer* [58] and others [59–61]. These characteristics represent the behaviors characterizing host shifts and, thus, ‘speciation genes’.

For *Rhagoletis* researchers, using *Drosophila* as a model or leveraging *Drosophila* genomics to understand the *Rhagoletis* system will yield fruit. *Rhagoletis* researchers can identify or create the *Drosophila* mutants with appropriate behavioral variation with respect to pupal diapause, food preference, or oviposition-related characters and identify genes involved using RNAi gene knockout [63] or other approaches. A leveraged approach would begin with candidate genes mapped on the *Drosophila* genome, identify *Rhagoletis* homologs through comparative mapping, and manipulate *Rhagoletis* genes and phenotypes. Examining nucleotide substitution rates (relative to *Drosophila*) will provide a measure of evolutionary rates in these genes and testing hypotheses about speciation processes. *Drosophila* is also being used in this way for studies of bee behavior [62].

### Ecological genomics success stories

Recent publications have made substantial contributions to ecological research by using novel and innovative approaches and are helping develop the discipline of ecological genomics.

#### 'Micro' success stories

In a recent study [18], a whole-genome comparison and gene expression experiments were conducted with microbial thermophilic communities to identify genes and pathways involved in thermoregulation, DNA repair, and other activities necessary for life in extreme environments. This study is the beginning of a quest to understand the origins and regulatory processes of ecosystems. Using genome-based approaches is common in the microbial ecology research community, and researchers studying at the 'macro' level can gain insights from these efforts. Recent progress has been made using hydrothermal vent ecosystems [18] and other studies of microbial diversity [19,20], investigations into microbial community structure [21], and alterations in gene regulation of microbial symbiont communities to changes in host diet [22]. These studies examine population- and community-level phenomena and explore ecosystem-level implications from a genome perspective.

#### Gene expression in non-model organisms

Two recent reports have shown that genome-scale expression analysis might soon be a useful, widely used tool for ecological studies. In two studies, Baldwin and colleagues constructed a custom cDNA microarray for the native tobacco plant *Nicotiana attenuata* containing 161 and 241 unique elements, respectively [23,24]. Using these microarrays, they compared transcriptional responses in *Nicotiana* treated with regurgitants from the tobacco hornworm *Manduca sexta*, a lepidopteran herbivore of this plant. Halitschke and colleagues showed that the most abundant fatty acid–amino acid conjugates in *Manduca* regurgitant cause most of the herbivory-induced response in *Nicotiana* (measured by the degree of gene expression perturbation) [23]. Izaguirre and colleagues demonstrated that differentially expressed genes in *Nicotiana* in response to solar UV light were similar to the expression of genes involved in *Manduca* herbivory [24]. These studies begin to solve the longstanding ecological question about the nature of herbivore damage and the response of the plant to that damage. The use by the authors of custom cDNA microarrays paves the way for other such applications (especially at a larger scale). Such studies will be enhanced by adhering to the Minimum Information About a Microarray Experiment (MIAME) standards [25] established to maximize the potential for sharing, combining and interpreting microarray data.

For ecological studies without the cDNA resources required for custom-made microarrays, there are other options. Reineke and colleagues conducted a genome-wide analysis of gene expression for two clonal lines of the parasitic wasp *Venturia canescens* (Ichneumonidae) that exist sympatrically by using two different host species [26]. The clonal lines were differentiated by two allelic forms of a virus-like particle protein (VLP1). Using a transcript profiling approach (cDNA-AFLP) used widely in plant

systems [27], Reineke and colleagues assayed 2150 wasp ovary-expressed transcripts and found seven that were expressed at different rates in the two wasp strains [26]. The expression of seven genes was confirmed by conventional laboratory experiments [28]; these genes included several with a role in stress responses during larval development. This work not only opens the door to understanding the genetic basis of phenotypic differences involved in host race formation, but also highlights an exciting course of study that is applicable to other non-model organisms involved in similar interactions.

#### Gene hunting

Gene hunting (or discovery) is the process of identifying genetic elements associated with a given phenotype such as a complex trait involving a biological process of interest (e.g. a response to herbivory). Given that complex traits are typically controlled by multiple genes, each influenced by other genes and environmental factors, gene hunting is extremely difficult. The payoff, however, can be huge, leading to an understanding of the genetic components regulating the complex trait, the genetic variation generating variation in the expression of the trait and a deeper appreciation of the environmental factors modifying expression of that trait.

Several ecological research groups now use approaches for gene discovery that are used for model organism systems in the laboratory [29] and, increasingly, with compelling field components [30]. These exciting discoveries use model organisms as they were intended (as stand-ins for species that are more relevant for ecological processes but are long lived, genetically uncharacterized, or otherwise intractable) and by applying gene-hunting approaches directly to organisms of interest.

The use of model organisms to study complex traits is more amenable to experimental conditions that are more familiar in a genomics research setting. Advances in our understanding of the genetic component of human diseases comes courtesy of the laboratory rat *Rattus norvegicus*, for which there is a complete and well-annotated genome sequence [31,32]. Another model organism, *Arabidopsis thaliana*, also has a complete and well annotated genome [33] in addition to having a long history in ecological research [34].

The identification of *Arabidopsis* quantitative trait loci (QTLs) that are associated with herbivory resistance and tolerance [35] is having a significant effect among ecologists. Weinig and colleagues examined recombinant inbred *Arabidopsis* lines in a field setting and, although they found no evidence to support a QTL for tolerance, they did find evidence for at least one QTL that affected plant resistance and fitness [35]. *Arabidopsis* has also proved useful for studies of the genetic basis of tolerance and resistance of herbivory [36–38] and, using *Nicotiana*, researchers have identified proteins mediating interactions with pathogens [10,39].

Other recent work has used non-model (or, at least, less well characterized model) organisms. Rolfe and colleagues conducted peptide mass fingerprinting (PMT) of all proteins involved in legume–microbe interactions [40], using the barrel medic legume *Medicago truncatula* and a

### Box 3. *Sapium*, meet *Aedes*

Recent publications [64–66] explore differences in preference of grasshoppers for invasive plants growing in native and invaded host plant habitats. Siemann and Rogers examined grasshopper preference, plant phenotypic plasticity for tolerance, and plant post-invasion genetic differences, linking their results to broader issues such as the importance of predators and other environmental factors for limiting or enhancing invasive plant success. In one set of experiments involving *Sapium* (Chinese tallow tree, an invasive plant in the southeastern USA) [65], the authors showed that grasshoppers prefer *Sapium* saplings from the native range of the species to invasive saplings from Texas. Understanding this post-invasion genetic difference provides a new understanding of invasion ecology. It could also lead to innovative approaches to preventing problems caused by invasive species.

Identifications of genetic underpinnings has been successful in other systems: the Center For Tropical Disease Research and Training at the University of Notre Dame (<http://ctdrt.bio.nd.edu>) has uncovered midgut-expressed genes in the *Aedes* mosquito that are associated with susceptibility to malarial infection by *Plasmodium* [67,68]. This work was motivated by the need to control vector infection, a goal of malaria-related genome projects [69,70]. This group identified *Aedes* genes interfering with *Plasmodium* development (in resistant *Aedes* populations) by preventing *Plasmodium* from forming oocysts [67]. The genes identified mediate the interactions between parasite and resistant host, similar to interactions between grasshoppers and native and invasive *Sapium*.

Both projects explore complex phenotypes mediating an intimate interaction between two species. On the one hand, Siemann and Rogers ask, 'What is the plant, in its native setting, doing differently than its invasive sister that changes grasshopper preference?'; whereas, on the other, Severson asks, 'What is the resistant mosquito strain doing differently than its wild sister that changes the ability of the parasite to use it as a host?' In both cases, the answer involves a heritable propensity for susceptibility. The key is understanding the identity of the genes involved and the degree and timing of their expression. For *Sapium*, the work has just begun: the genetic contribution to the differences found by Rogers and Siemann must be verified and compared in native and invasive habitats. Phenotypes mediating differences in susceptibility must be identified. Gene expression differences can then be explored and consequences of these differences for the phenotypic components can be determined. The systems involved in susceptibility to grasshopper herbivory can then be identified and exploited to control this (and, potentially, other) invasive species.

nitrogen-fixing root nodule bacterium *Sinorhizobium meliloti*. PMF is a whole-proteome approach involving the determination of peptide masses by spectrometric techniques and the matching of these masses against peptide libraries to create a list of probable protein identifications. Rolfe and colleagues identified proteome constituents involved in the interaction, although identification of the genetic elements involved is currently difficult. This approach can also lend itself to traditional ecological questions (Box 3). The molecular basis of flowering time variation [41], parasite resistance in salmon and reproductive changes in trout [42] have also been characterized. These types of experiment provide information that can link responses to different experimental conditions and can generate testable hypotheses that help extend our understanding of the interactions.

#### Ecological genome databases

Genome project data are compiled in databases encompassing all relevant information about that organism. For

example, the Rat Genome Database [31] (RGD: <http://rgd.mcw.edu>) contains information about rat genes (and their homologs in human), disease models, genome maps, and bioinformatics tools designed to mine these data. Similarly, the Daphnia Genomics Consortium (DGC: <http://daphnia.cgb.indiana.edu>) is an international alliance of investigators using the freshwater crustacean *Daphnia* as a model system for evolutionary and ecological genetics and genomics. The *Daphnia* system is unique, because it is both a model organism and is ecologically interesting in and of itself.

Are genome databases likely for other ecologically relevant organisms? Currently, databases are likely to be restricted to those systems that have sufficient funding and, therefore, a suitable number of researchers, whatever that number might be. One such possibility is *Rhagoletis* (Box 2.) Cost is becoming less important as applications, such as the Generic Model Organism Database [43] 'construction set' (GMOD: <http://www.gmod.org>) become available. Using the GMOD, a research consortium can assemble genetic data, literature and phenotypic information in a customizable database with existing bioinformatics data-mining tools. For biomedical model organisms, the phenotypic information includes human disease-related data: genes and loci directly associated with disorders, genetic elements associated with systems and pathways involved in disorders, and other information linking the model organism to human disease (polymorphisms, comparative maps, quantitative trait loci, consomic strains, etc.).

#### Future directions

Genomic approaches will not replace traditional ecological methods but will enhance the type of information that can be derived from an experiment. Genomics techniques can be used with ecologically relevant species, and could provide new, unique insight into current models of ecological phenomena. Not all ecologists need to become genomics researchers, but ecology as a field will benefit by collaborations with genomic scientists. Genomics researchers provide a new set of tools with which to study the molecular basis of these characteristics.

It is a two-way street: according to Martin Feder (pers. commun.) loosely paraphrasing John F. Kennedy: 'Ask not what genomics can do for ecology, but also what ecology can do for genomics'. In turn, ecologists can provide a unique angle for genomics researchers and a significant contribution to the study of functional genomics. Ecological projects give a unique insight into the natural history, behavior, distribution and interactions of native species. Ecologists participating in (and, by doing so, improving) systems biology research will have a leading role in helping understand the ecological basis for interactions among cells and molecules.

#### Acknowledgements

We thank Marjorie Matocq and Alissa Salmore for reviewing the article and Martin Feder and two anonymous reviewers for providing comments and suggestions that greatly improved its delivery and content.

## References

- 1 Stearns, S. and Magwene, P. (2003) The naturalist in a world of genomics. *Am. Nat.* 161, 171–180
- 2 Feder, M.E. and Mitchell-Olds, T. (2003) Evolutionary and ecological functional genomics. *Nat. Rev. Genet.* 4, 651–657
- 3 Klaper, R. and Thomas, M.A. (2004) At the crossroads of genomics and ecology: the promise of a canary on a chip. *Bioscience* 54, 403–412
- 4 Jackson, R.B. *et al.* (2002) Linking molecular insight and ecological research. *Trends Ecol. Evol.* 17, 409–414
- 5 Diatchenko, L. *et al.* (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U. S. A.* 93, 6025–6030
- 6 Rebrikov, D.V. *et al.* (2004) Suppression subtractive hybridization. *Methods Mol. Biol.* 258, 107–134
- 7 Velculescu, V.E. *et al.* (1995) Serial analysis of gene expression. *Science* 270, 484–487
- 8 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
- 9 Goetz, F.W. (2003) The ‘ups’ and ‘downs’ in using subtractive cloning techniques to isolate regulated genes in fish. *Integr. Comp. Biol.* 43, 16–23
- 10 Matsumura, H. *et al.* (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15718–15723
- 11 Draghici, S. *et al.* (2003) Global functional profiling of gene expression. *Genomics* 81, 98–104
- 12 Moody, D. *et al.* (2002) Cross-species hybridisation of pig RNA to human nylon microarrays. *BMC Genomics* 3, 27
- 13 Miller, N.A. *et al.* (2002) Cross-hybridization of closely related genes on high-density macroarrays. *Biotechniques* 32, 620–625
- 14 Hedges, S.B. and Kumar, S. (2002) Genomics. Vertebrate genomes compared. *Science* 297, 1283–1285
- 15 Burt, D.W. *et al.* (1999) The dynamics of chromosome evolution in birds and mammals. *Nature* 402, 411–413
- 16 Stillman, B.A. and Tonkinson, J.L. (2001) Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal. Biochem.* 295, 149–157
- 17 Roethele, J. *et al.* (2001) Evidence for broad-scale conservation of linkage map relationships between *Rhagoletis pomonella* (Diptera: Tephritidae) and *Drosophila melanogaster* (Diptera: Drosophilidae). *Ann. Entomol. Soc. Am.* 94, 936–947
- 18 Reysenbach, A.L. and Shock, E. (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* 296, 1077–1082
- 19 Torsvik, V. and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* 5, 240–245
- 20 Beja, O. *et al.* (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* 68, 335–345
- 21 Wellington, E.M. *et al.* (2003) Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr. Opin. Microbiol.* 6, 295–301
- 22 Moran, N.A. *et al.* (2003) A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14543–14548
- 23 Halitschke, R. *et al.* (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VI. Microarray analysis reveals that most herbivore-specific transcriptional changes are mediated by fatty acid-amino acid conjugates. *Plant Physiol.* 131, 1894–1902
- 24 Izaguirre, M.M. *et al.* (2003) Convergent responses to stress. Solar ultraviolet-B radiation and *Manduca sexta* herbivory elicit overlapping transcriptional responses in field-grown plants of *Nicotiana longiflora*. *Plant Physiol.* 132, 1755–1767
- 25 Brazma, A. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365–371
- 26 Reineke, A. *et al.* (2003) Differential gene expression in two strains of the endoparasitic wasp *Venturia canescens* identified by cDNA-amplified fragment length polymorphism analysis. *Mol. Ecol.* 12, 3485–3492
- 27 Donson, J. *et al.* (2002) Comprehensive gene expression analysis by transcript profiling. *Plant Mol. Biol.* 48, 75–97
- 28 Franz, O. *et al.* (1999) Verification of differential gene transcription using virtual northern blotting. *Nucleic Acids Res.* 27, e3
- 29 Ungerer, M.C. *et al.* (2002) Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* 160, 1133–1151
- 30 Weinig, C. *et al.* (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* 165, 321–329
- 31 Twigger, S. *et al.* (2002) Rat genome database (RGD): mapping disease onto the genome. *Nucleic Acids Res.* 30, 125–128
- 32 Stoll, M. *et al.* (2000) New target regions for human hypertension via comparative genomics. *Genome Res.* 10, 473–482
- 33 Zhu, W. *et al.* (2003) Refined annotation of the *Arabidopsis* genome by complete expressed sequence tag mapping. *Plant Physiol.* 132, 469–484
- 34 Nordborg, M. and Bergelson, J. (1999) The effect of seed and rosette cold treatment on germination and flowering time in some *Arabidopsis thaliana* (Brassicaceae) ecotypes. *Am. J. Bot.* 86, 470
- 35 Weinig, C. *et al.* (2003) QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Mol. Ecol.* 12, 1153–1163
- 36 Kroymann, J. *et al.* (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14587–14592
- 37 Clauss, M.J. and Mitchell-Olds, T. (2003) Population genetics of tandem trypsin inhibitor genes in *Arabidopsis* species with contrasting ecology and life history. *Mol. Ecol.* 12, 1287–1299
- 38 Kover, P.X. and Schaal, B.A. (2002) Genetic variation for disease resistance and tolerance among *Arabidopsis thaliana* accessions. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11270–11274
- 39 Kamoun, S. *et al.* (1998) Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. *Plant Cell* 10, 1413–1426
- 40 Rolfe, B.G. *et al.* (2003) Proteomic analysis of legume-microbe interactions. *Comp. Funct. Genomics* 4, 225–228
- 41 Gazzani, S. *et al.* (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* 132, 1107–1114
- 42 Thorgaard, G.H. *et al.* (2002) Status and opportunities for genomics research with rainbow trout. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 133, 609–646
- 43 Stein, L.D. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.* 12, 1599–1610
- 44 Smith, L. and Greenfield, A. (2003) DNA microarrays and development. *Hum. Mol. Genet.* 12, R1–R8
- 45 Zdobnov, E.M. *et al.* (2002) Comparative genome and proteome analysis of anopheles gambiae and *Drosophila melanogaster*. *Science* 298, 149–159
- 46 von Roepenack-Lahaye, E. *et al.* (2004) Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiol.* 134, 548–559
- 47 Bush, G.L. (1969) Mating behavior, host specificity, and the ecological significance of sibling species in frugivorous flies of the genus *Rhagoletis* (Diptera-Tephritidae). *Am. Nat.* 103, 669
- 48 McPherson, B.A. *et al.* (1988) Genetic differences between host races of *Rhagoletis pomonella*. *Nature* 336, 64–66
- 49 Feder, J. *et al.* (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proc. Natl. Acad. Sci. U. S. A.* 91, 7990–7994
- 50 Linn, C. Jr *et al.* (2003) Fruit odor discrimination and sympatric host race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11490–11493
- 51 Fry, J.D. (2003) Multilocus models of sympatric speciation: Bush versus Rice versus Felsenstein. *Evolution* 57, 1735–1746
- 52 Berlocher, S.H. and Feder, J.L. (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu. Rev. Entomol.* 47, 773–815
- 53 Feder, J.L. *et al.* (2003) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10314–10319
- 54 Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195
- 55 Borras, T. *et al.* (2003) Transcription profiling in *Drosophila* eyes that overexpress the human glaucoma-associated trabecular

- meshwork-inducible glucocorticoid response protein/myocilin (TIGR/MYOC). *Genetics* 163, 637–645
- 56 Feany, M.B. and Bender, W.W. (2000) A *Drosophila* model of Parkinson's disease. *Nature* 404, 394–398
- 57 Blair Hedges, S. and Kumar, S. (2003) Genomic clocks and evolutionary history. *Trends Genet.* 19, 200–206
- 58 Diagana, T.T. *et al.* (2002) Mutation of *Drosophila* homer disrupts control of locomotor activity and behavioral plasticity. *J. Neurosci.* 22, 428–436
- 59 Chan, B. *et al.* (2002) Courtship and other behaviors affected by a heat-sensitive, molecularly novel mutation in the cacophony calcium-channel gene of *Drosophila*. *Genetics* 162, 135–153
- 60 Caldwell, J.C. *et al.* (2003) Dynamic analysis of larval locomotion in *Drosophila* chordotonal organ mutants. *Proc. Natl. Acad. Sci. U. S. A.* 100, 16053–16058
- 61 Rayburn, L.Y. *et al.* (2003) *amontillado*, the *Drosophila* homolog of the prohormone processing protease PC2, is required during embryogenesis and early larval development. *Genetics* 163, 227–237
- 62 Grozinger, C.M. *et al.* (2003) Pheromone-mediated gene expression in the honey bee brain. *Proc. Natl. Acad. Sci. U. S. A.* 100 (Suppl. 2), 14519–14525
- 63 Kim, Y.-O. *et al.* (2003) A functional genomic screen for cardiogenic genes using RNA interference in developing *Drosophila* embryos. *Proc. Natl. Acad. Sci. U. S. A.* 101, 159–164
- 64 Rogers, W.E. and Siemann, E. (2003) Effects of simulated herbivory and resources on Chinese tallow tree (*Sapium sebiferum*, Euphorbiaceae) invasion of native coastal prairie. *Am. J. Bot.* 90, 243–249
- 65 Siemann, E. and Rogers, W.E. (2003) Reduced resistance of invasive varieties of the alien tree *Sapium sebiferum* to a generalist herbivore. *Oecologia* 135, 451–457
- 66 Siemann, E. and Rogers, W.E. (2003) Herbivory, disease, recruitment limitation, and success of alien and native tree species. *Ecology* 84, 1489–1505
- 67 Morlais, I. *et al.* (2003) A targeted approach to the identification of candidate genes determining susceptibility to *Plasmodium gallinaceum* in *Aedes aegypti*. *Mol. Genet. Genomics* 269, 753–764
- 68 Yan, G. and Severson, D.W. (2003) Dynamics of molecular markers linked to the resistance loci in a mosquito-*Plasmodium* system. *Genetics* 164, 511–519
- 69 Holt, R.A. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149
- 70 Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511

### Elsevier.com – Dynamic New Site Links Scientists to New Research & Thinking

Elsevier.com has had a makeover, inside and out. Designed for scientists' information needs, the new site, launched in January, is powered by the latest technology with customer-focused navigation and an intuitive architecture for an improved user experience and greater productivity.

Elsevier.com's easy-to-use navigational tools and structure connect scientists with vital information – all from one entry point. Users can perform rapid and precise searches with our advanced search functionality, using the FAST technology of Scirus.com, the free science search engine. For example, users can define their searches by any number of criteria to pinpoint information and resources. Search by a specific author or editor, book publication date, subject area – life sciences, health sciences, physical sciences and social sciences – or by product type. Elsevier's portfolio includes more than 1800 Elsevier journals, 2200 new books per year, and a range of innovative electronic products. In addition, tailored content for authors, editors and librarians provides up-to-the-minute news, updates on functionality and new products, e-alerts and services, as well as relevant events.

Elsevier is proud to be a partner with the scientific and medical community. Find out more about who we are in the About section: our mission and values and how we support the STM community worldwide through partnerships with libraries and other publishers, and grant awards from The Elsevier Foundation.

As a world-leading publisher of scientific, technical and health information, Elsevier is dedicated to linking researchers and professionals to the best thinking in their fields. We offer the widest and deepest coverage in a range of media types to enhance cross-pollination of information, breakthroughs in research and discovery, and the sharing and preservation of knowledge. Visit us at Elsevier.com.

**Elsevier. Building Insights. Breaking Boundaries.**