OPINION

Evolutionary and ecological functional genomics

Martin E. Feder and Thomas Mitchell-Olds

A unique combination of disciplines is emerging — evolutionary and ecological functional genomics — which focuses on the genes that affect ecological success and evolutionary fitness in natural environments and populations. Already this approach has provided new insights that were not available from its disciplinary components in isolation. However, future advances will necessitate the re-engineering of scientific attitudes, training and institutions, to achieve extensive multidisciplinarity.

Wild organisms flourish in nature despite severe challenges from their biotic and abiotic environments. Indeed, every living organism can be viewed as an evolutionary success story[1]. The emerging field of evolutionary and ecological functional genomics (EEFG) seeks to understand how this success is achieved. To accomplish this goal, the biological mechanisms that influence or underlie ecologically important traits must be studied. Also, it is necessary to investigate how these traits affect evolutionary fitness in nature, and to examine the evolutionary processes through which specific traits arise and persist. This makes EEFG a multidisciplinary endeavour. Because the mechanisms of each trait of interest are manifested at lower levels of biological organization and the significance of a trait is only apparent at higher levels[1], understanding a given trait usually requires the simultaneous use of molecular, cellular, organismal, population and ecological approaches.

This outlook, rather than genomics *per se*, is both the defining feature of EEFG and the source of its primary challenge. The molecular tools and functional understanding that are required to accomplish the goals of the field are beyond the capacity of any single investigator, which necessitates sustained interactions among research communities[2]. However, these communities are often organized around classical laboratory-based model organisms, which can be poor exemplars of the wild organisms that flourish in challenging natural environments. So, should EEFG emphasize the focal species of model organism communities at the potential expense of ecological and evolutionary realism? Or should it eschew existing model species and their research assets in favour of more ecologically appropriate non-classical models, even if these are less tractable and it takes time to develop the necessary resources and tools? The EEFG meta-community is not monolithic, and is evolving rapidly despite the apparent constraints inherent in these options. Here, we review progress in this field, which points to several resolutions of its central challenge.

Goals

Seemingly everyday, the suffix '-omic' appends to yet another discipline, which signifies the expansion of POST-GENOMIC science. Most of this growth is in the fields of biomedical and agricultural functional genomics, which aim to improve the health, longevity, productivity and well being of humans and agricultural species. For EEFG, by contrast, the focus is on organisms that inhabit natural environments and the goal of researchers is to explain variation in DARWINIAN FITNESS in populations, and variation in size, range, longevity and diversity among populations, species and higher taxa.

Nonetheless, biomedical and agricultural functional genomics, as well as EEFG, share a dominant research motif: finding the genes and polymorphisms that affect traits of interest and characterizing the mechanisms that underlie these effects[3]. The first step — identifying genes of interest — is being helped by post-genomic technologies that allow the high-throughput discovery of candidate genes[4,5]. However, proving that a candidate is consequential still depends on studies of mutants[6].

Genes and polymorphisms that might be of evolutionary significance can also be identified from theoretical population genetics, by using algorithms that infer which nucleotides evolve non-neutrally[7]. However, these algorithms provide little insight into the molecular mechanisms or ecological consequences of fitness differences, or the probable impact of evolutionary adaptations. So, we still need to characterize the mechanisms that cause particular genes and polymorphisms to impact on ecologically and evolutionarily significant traits. Such insights require mechanistic biology (biochemistry, physiology and so on), ultimately under realistic cellular and environmental conditions — not just '-omics', but 'functional -omics'.

EEFG therefore seeks to carry out the full spectrum of investigations of biomedical and agricultural functional genomics, but in wild organisms or their proxies (FIG. 1), with a primary focus on the evolutionary processes that determine and maintain genotypes and phenotypes. This is challenging for several reasons. For example, the genetic variation that segregates in natural populations often encodes alternative phenotypes of comparatively minor magnitude, which might be evident only under natural conditions. Also, the

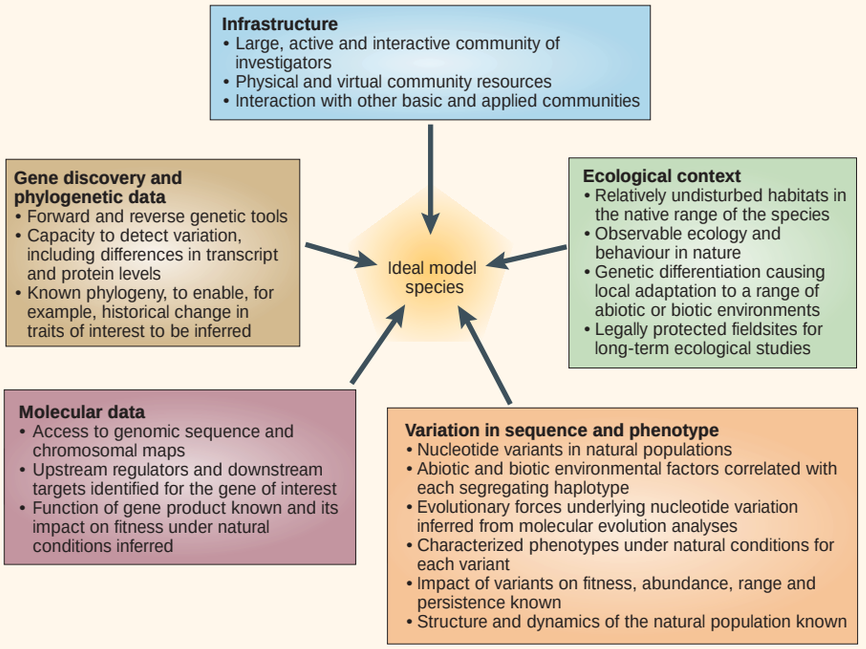

Figure 1 | **Criteria for model species in evolutionary and ecological functional genomics.** Ideally, study organisms should satisfy all of the criteria shown; however, at present, few eukaryotes do so. Classical model species, such as *Drosophila* and *Arabidopsis*, pose difficulties for ecological studies, whereas many popular ecological models are genetically intractable, have poorly-characterized genomes and lack large communities of investigators. This discrepancy should disappear with future advances in technology. Some microbial models, by contrast, satisfy all criteria and are yielding great progress.

biotic and abiotic environments in which wild organisms occur are often obscure, highly unpredictable and difficult to monitor, and techniques from standard model organisms can be difficult to apply to wild species. Altogether this is a challenging research agenda and is very much a work in progress. Data are now sufficient, however, to show how each of the component approaches (evolution, ecology, functional biology and genomics) can contribute to the whole, how each component benefits from the others and how EEFG can contribute to biomedical and agricultural functional genomics[8].

Approaches

Model organisms for EEFG. The dichotomy between ecologically and genetically tractable model systems that are suitable for EEFG studies is exemplified by two of the many genera of potential model organisms: *Daphnia* (FIG. 2a), which is a non-classical model, and *Arabidopsis* (FIG. 2b), which is a classical laboratory-based model.

Daphnia are aquatic crustaceans that have been the focus of ecological and evolutionary studies for many decades. Recently, an international *Daphnia* Genomics Consortium has formed to develop *Daphnia* as a model system for EEFG. This research and educational network is expanding genomic resources for *Daphnia*, including microsatellites, a linkage map, expressed sequence tags (ESTs) and microarrays, large-insert genomic clones, transformation and RNAi, genomic databases and an international stock centre. Parallel development of genomic tools for ecologically tractable species is underway for other eukaryotes (TABLE 1), including fish[4], honeybees[9] and tobacco[10], and is well advanced for Archaea and Eubacteria. At present, at least 140 complete genomes have been published and more than 700 other genome projects are underway (for further information see links to the Genomes Online Database, Genome News Network, TIGR Gene Indices and Organism-Specific Genome Databases in the online links box).

Arabidopsis thaliana is a genetically tractable laboratory model, the complete genomic sequence of which is known. Effort now focuses on determining the function of all of its 25,000 genes before the end of this decade. Many recent studies have examined the ecology and evolution of *A. thaliana* in nature (for example, REFS 11–13). Furthermore, because genomic tools are readily transferable to wild relatives of *A. thaliana*, a diverse research community has coalesced around several closely related species that grow in undisturbed natural habitats and differ from *A. thaliana* in breeding system, life history, physiology, genetics and developmental biology[3] (see link to Wild Relatives of *Arabidopsis* in the online links box). Similarly, parallel exploitation of laboratory species as ecological models is underway for yeast[14], *Caenorhabditis elegans* (see link to the NemATOL web site in the online links box), zebrafish and *Drosophila*[15–18].

Transcription profiling. Transcription profiling (TP) is a useful genomics tool for EEFG, and quantifies the expression of thousands of genes in a series of treatments, tissues or time points[19]. TP technology is relatively accessible for many laboratories, and is applicable to non-model organisms. This technology has had important successes in identifying genes the expression of which is correlated with ecologically important traits[4,20,21]. It has also spawned a vast literature of clustering, grouping and other forms of multivariate data massage to describe changes in gene expression. Only recently, however, has TP begun to focus on hypothesis testing in a rigorous statistical framework[22,23]. We suggest that TP is a useful first step in describing patterns of gene expression and finding candidate genes that might influence traits of interest. However, replication is often limited and hypothesis testing is usually provisional. So, TP should be regarded as exploratory data analysis in advance of the manipulative experiments that are needed to provide rigorous verification[24]. Accordingly, we see TP as a useful tool for EEFG, but not as a research goal in itself.

Two recent studies illustrate the power of TP in EEFG. Oleksiak *et al.*[4] examined gene expression in and between natural populations of *Fundulus heteroclitus*, a fish that is found along the Atlantic Coast. Northern and southern populations showed significant divergence for 27 genes that are expressed in heart tissue. In a similar approach, Rifkin *et al.*[5] found differences in gene expression between *Drosophila melanogaster* and two closely related species. Both studies indicate that changes in gene regulation might be important in the evolutionary divergence of populations or species, and provide a list of genes that might have been influenced by natural selection. However, neither study has taken the next step of showing the ecological or evolutionary significance of variation in transcript levels.

By contrast, in their study of the evolution of GEOTAXIS in *Drosophila*, Toma *et al.*[25] have followed TP with studies of mutant lines to show that some of the genes nominated by

Figure 2 | **Model organisms for evolutionary and ecological functional genomics**. **a** | *Daphnia minnehaha*, photograph courtesy of Paul Hebert, University of Guelph, Canada. **b** | *Arabidopsis lyrata* ssp. *petraea*, photograph courtesy of Thomas Mitchell-Olds.

TP actually function in geotaxis. A possible next step would be to discover the polymorphisms that segregate in natural populations and regulate differences in gene expression, and to determine the evolutionary forces (mutation, migration, natural selection and drift) that maintain them. However, experiments that are designed to answer such questions are often still only at the planning stage. Consequently, despite the usefulness of TP for EEFG, we focus on progress towards understanding the functional and evolutionary bases of ecologically important variation. Necessarily, successes so far involve individual genes rather than genome-wide patterns.

Ecology

Benefits from EEFG. Genomic and molecular tools have revolutionized our ability to identify the organisms that are present in any community or ecosystem under study, which, in the past, was only possible using phenotypic criteria. Indeed, genomic sequencing of DNA from environmental samples, such as seawater or soil, now shows that more than 99% of microbial species were not detected by earlier methods[26]. These techniques are discovering previously unrecognized but crucial components of biological communities, such as the SAR 11 clade, which is a group of poorly known uncultured marine bacteria that might represent more than 10% of the marine prokaryotic biomass worldwide[27]. Pre-genomic ecology emphasized conspicuous if not charismatic species; post-genomic ecology has the potential to escape this bias.

Large-scale molecular identification, which is becoming crucial in describing ecological pattern and process[28], has only been feasible with the advent of genomic techniques and databases. For example, ecosystem studies of rooting-depth patterns in forest tree communities have used molecular markers to determine the species identity of roots[29]. Also, genes

Table 1 | Non-classical model eukaryotes used in evolutionary and ecological functional genomics studies

Genera	Common name	Web sites	Reference
Vertebrates			
Oryzias	Medaka	http://mbase.bioweb.ne.jp/~dclust/medaka_top.html	65
Gasterosteus	Sticklebacks	http://cegs.stanford.edu	35
Tilapia, Astatotilapia and others	Cichlids	http://hcgs.unh.edu/cichlid http://cgr.harvard.edu/hans/html/research.html	66
Salmo and others	Salmon	http://www.salmongenome.no	–
Cyprinus	Carp	http://sphere.bioc.liv.ac.uk:8080/bio/research/legr/index_html	67
Oncorhynchus	Trout	http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/intro2.pl?BASE=rainbow	–
Ambystoma and others	Salamanders	http://salamander.uky.edu	–
Invertebrates			
Bombyx, Heliothis and others	Lepidoptera	http://www.ab.a.u-tokyo.ac.jp/lep-genome http://www.ucl.ac.uk/taxome/intro.html	–
Apis	Honeybee	http://www.hgsc.bcm.tmc.edu/projects/honeybee	68
Daphnia	Water flea	http://daphnia.cgb.indiana.edu	–
Amblyomma	Tick	http://www.genome.ou.edu/tick.html	–
Strongylocentrotus	Sea urchin	http://sugp.caltech.edu http://www.hgsc.bcm.tmc.edu/projects/seaurchin	69
Plants			
Populus	Poplar	http://genome.jgi-psf.org/poplar0/poplar0.home.html	70
Pinus and others	Other forest trees	http://dendrome.ucdavis.edu	–
Helianthus	Sunflowers	http://compgenomics.ucdavis.edu/index.htm http://www.tigr.org/tdb/tgi/hagi	–
Mesembryanthemum	Iceplant	http://www.tigr.org/tdb/tgi/mcgi	–
Nicotiana and *Solanum*	Tobacco and nightshade	http://www.ice.mpg.de/departments/Ecol/moleculartools/moleculartools.html	–

and transgenes the expression of which is sensitive to environmental conditions can provide continuing environmental monitoring in organisms for which invasive instrumentation is not feasible. For example, diverse protein-damaging stresses induce the expression of genes that are under the control of heat-shock promoters, and so report the microenvironment of the host organism[30].

So far, most ecological experimentation has been comparatively gross, involving wholesale alterations of entire habitats, communities or ecosystems, the inclusion or exclusion of species and GUILDS, and the surgical manipulation of individual organisms. Although it is possible to experimentally manipulate traits one gene or even one nucleotide at a time, at least in genetically tractable species, the effects of such changes in nature have not been widely assessed. Nevertheless, several first studies show that such experiments are feasible[13,30]. Similarly, although there has been progress in identifying specific naturally occurring polymorphisms, the number of studies testing the effects of these phenotypes in nature is small[9,11,13,20,31].

"…knowing the details of how genetic variation affects molecular, cellular and organismal function is often essential to understanding genetic variation and its evolutionary impact…"

Experimental work on model organisms in the laboratory, moreover, can be extended into complementary work on non-classical models in nature. For comparisons between *Drosophila* and honeybees, or laboratory mice and wild rodent species, polymorphisms or mutations in laboratory studies have counterparts in wild or exotic species, which can be analysed as NATURAL EXPERIMENTS[9,32]. Insights from such fine-scale experimentation might be essential to predict the outcome of several unintentional large-scale ecological experiments that are now underway, including global climate change, anthropogenic ecosystem destruction, biodiversity depletion and the introduction of invasive species. Indeed, evolved or engineered genomes are already crucial components of human responses to habitat deterioration and invasive species, and will probably increase in importance.

Contributions to EEFG. Ecological knowledge is crucial for the interpretation of genomic and post-genomic data[10,11], particularly in establishing the consequences of genetic variation. Transcription profiles are exquisitely sensitive to the interaction and kinetics of environmental factors, which necessitates great care in establishing the physiological and ecological relevance of experimental conditions[22]. Similarly, the fitness consequences of genetic variants can differ substantially between laboratory and field experiments, with phenotypes that are absent in one venue being present in the other. In mice, for example, inbreeding is more harmful and major histocompatability complex (MHC) heterozygosity more beneficial in large semi-natural enclosures (which are environmentally complex and afford realistic social interactions) than in the laboratory[33]. Likewise, some quantitative trait loci (QTLs) for salt-tolerance traits in sunflowers differ substantially when measured in model soil in greenhouses and in the wild[34]. Also, QTLs that control *Arabidopsis* flowering time are detectable in both laboratory and field experiments, but other genomic regions have different influences on flowering under controlled versus natural conditions[11]. The genetic basis of flowering time in *Arabidopsis* is among the best understood traits; so, the limited predictive power of laboratory studies is sobering.

Finally, ecology and natural history are goldmines of species, mechanisms and genes that are ripe for scientific exploitation, as many examples from modern science and medicine attest, including DNA polymerase from *Thermus aquaticus* and other thermostable polymerases, squid giant axons, antibiotics, restriction enzymes, fluorescent proteins and taxol. Organisms that inhabit extreme environments have been especially valuable for BIOPROSPECTING, as the severe conditions often exaggerate their features. Also, broad knowledge of biodiversity allows the most suitable model organisms to be chosen: examples include genomics of the pufferfish (*Fugu*), which was chosen for its small genome, and evolutionary developmental biology of the *Bicyclus* butterfly and stickleback fish[35–37], the ecology and phylogeography of which are ideal for interpreting the phenotypic consequences of genetic variation.

Evolution

Genetics, genomics and evolutionary biology are already strongly embedded in one another, and the fruits of their interaction are obvious. Here, we present a few of the many excellent examples.

Figure 3 | **Variation in mouse coat colour.** Light and dark coloured pocket mice (*Chaetodipus intermedius*) shown on rocky substrates in Arizona. Reproduced with permission from REF. 32. © (2003) National Academy of Sciences, USA.

Positional cloning has successfully identified molecular polymorphisms that are responsible for variation in several complex traits[31,38,39]. The next objective is to understand the historical and evolutionary forces that influence this variation. Sensory bristles, which are components of the peripheral nervous system in *Drosophila*, are among the best understood quantitative traits[40], but nonetheless exemplify the difficulties that are involved in attaining this information. Bristle number clearly shows genotype–environment interaction between sexes and among growth environments, and seems to be under weak STABILIZING SELECTION. However, it is still unclear which of the particular nucleotide polymorphisms near bristle number QTLs have been targets of selection[41]. One possible explanation for these polymorphisms comes from a MUTATION–SELECTION BALANCE MODEL that incorporates both stabilizing selection and deleterious PLEIOTROPY[42]; this predicts that many QTL alleles might have substantial effects on quantitative traits, but little effect on fitness itself. Another explanation is that QTL interactions with other loci (EPISTASIS) or with the environment maintain genetic variation for bristle number (C. Langley, T. MacKay and M. Turelli, personal communication). Finally, since LINKAGE DISEQUILIBRIUM extends only a few hundred base pairs in *Drosophila*[41], the known polymorphisms might not be the nucleotides that affect fitness, but might simply be linked to those evolutionarily important sites. The genetic complexity of *Drosophila* bristle number will probably be evident for quantitative traits in wild populations, agricultural species and human biomedical research.

Polymorphism at disease resistance genes can be maintained by BALANCING SELECTION for long evolutionary periods[43]. For example, *Pseudomonas* species are natural pathogens

of *A. thaliana* in wild populations[44]. The resistance gene *RPM1* can detect pathogens carrying *AvrRpm1* or *AvrB* avirulence genes[45]. In both *A. thaliana* and *Brassica napus*, disease-susceptible plants have lost the *RPM1* gene, which is present and functional in resistant individuals. These plant species diverged ~18 million years ago, which indicates either that independent deletions occurred or an ancient *trans*-specific polymorphism existed. Does a cost of resistance maintain the susceptible *RPM1*-deletion allele? Genomic methods have only recently allowed a clear test of this hypothesis. In four independent pairs of *RPM1* insertion versus deletion lines, which differ only in the *RPM1*-encoding locus and its endogenous promoter, the presence of a functional *RPM1* allele caused a 9% reduction in total fecundity under disease-free conditions, providing strong evidence for a cost of resistance[13].

Colour polymorphisms have attracted biological interest for many years, and provide some of the best characterized examples of functionally and ecologically important polymorphisms[46,47]. In the pocket mouse *Chaetodipus intermedius*, coat colour is correlated with the colour of the soil that is inhabited by individual populations[32] (FIG. 3), presumably as an anti-predator adaptation. Nachman and colleagues[32] examined two candidate genes, *agouti* and *Mc1r*, which regulate coat colour in mammals. In one population, amino-acid polymorphisms in the *Mc1r* locus (encoding the melancortin-1-receptor) are associated with coat colour, and patterns of nucleotide polymorphisms indicate recent directional selection at this locus. In New Mexico populations, by contrast, the *Mc1r* locus shows no signature of natural selection, which indicates that other loci control the adaptive evolution of coat colour.

"The molecular tools and functional understanding that are required to accomplish the goals of the field are beyond the capacity of any single investigator…"

These diverse findings exemplify several points. First, existing genomes and phenotypes can be the product of diverse evolutionary processes that interact in a complex fashion. This situation, rather than simple stabilizing or directional selection acting on a few genes of large effect, should be the starting point for research programmes in EEFG. Second, analysis of genomic and phenotypic complexity requires the joint application of tools from genetics, genomics and post-genomics, population biology and evolutionary biology. The disease-resistance genes of *Arabidopsis*, for example, required genomics for their discovery, population studies to document levels of polymorphism in the wild, evolutionary biology to frame robust hypotheses about the maintenance of these polymorphisms and experimental genetics to test them. These tools will frequently emerge from areas outside EEFG, such as biomedicine and model systems, which EEFG will need to include. Third, although research in this mode can be challenging, it is feasible and has already yielded significant insights.

Need for integration

Although evolutionary biology has welcomed genetics and genomics, it has sometimes been averse to functional biology (for exceptions see REFS 13,41). We contend, as have others[48,49], that knowing the details of how genetic variation affects molecular, cellular and organismal function is often essential to understanding genetic variation and its evolutionary impact, and frequently facilitates evolutionary genetic studies. The most pragmatic benefit of functional analysis is in excluding linkage as the cause of proposed genetic effects. Also, although exclusively bioinformatic analyses can infer which nucleotides have been targets of natural selection[50], they provide little information on the probable consequences of this selection. For example, knowing that haemoglobin-encoding genes have undergone balancing selection in a population would have been a dead end without knowledge of the functions of haemoglobin in respiratory-gas transport, pH buffering and malaria resistance. Indeed, if the functional consequences of each codon and non-coding nucleotide are understood, as are the interactions of these functions in the proteome, the power and resolution of functional insights increase accordingly. For example, Swanson and colleagues[51] inferred positive Darwinian selection acting specifically on functional domains of mammalian reproductive proteins (such as the 'sperm receptor' ZP3) by exploiting knowledge of which domains were functional. Similarly, Riley and colleagues[52] made use of the well-known Ras signalling pathway of *Drosophila* to show that upstream elements in particular were prone to selection.

Alternatively, comparative and ecological variation can implicate important nucleotides, genes and proteins for functional analysis[53]. For example, comparisons of metabolic enzymes, such as lactate dehydogenase (LDH) and phosphoglucose isomerase (PGI), among populations arrayed along environmental gradients have detected considerable genetic variation,

Glossary

BALANCING SELECTION
Natural selection that maintains higher levels of genetic variation than are expected under neutrality.

BIOPROSPECTING
The sampling of diverse organisms for genes, gene products and other compounds that are of value to humans.

DARWINIAN FITNESS
The expected reproductive contribution to future generations.

EPISTASIS
The influence of the interaction of multiple loci on variation in a single trait.

GEOTAXIS
Movement up or down, which requires the perception of and response to gravity.

GUILDS
Groups of species that use a common resource in similar ways.

LINKAGE DISEQUILIBRIUM
When genotype frequencies at several loci are correlated or non-independent.

MUTATION–SELECTION BALANCE MODEL
A population genetics model that assumes that a combination of mutation and balancing selection can explain present levels of genetic variation.

NATURAL EXPERIMENTS
The comparison of naturally arising variants of individual organisms, populations, species or higher taxa, which is similar to the way in which control and manipulated subjects are compared in anthropogenic experimentation.

PHYLOGENETIC FOOTPRINTING AND SHADOWING
Both approaches seek to identify conserved regulatory elements by comparing genomic sequences between related species. Phylogenetic footprinting uses one or a few relatively distant evolutionary comparisons, whereas phylogenetic shadowing examines a set of closely related species.

PLEIOTROPY
When a single gene or polymorphism influences two or more separate traits.

POST-GENOMIC
The era following the availability of complete genome sequences.

STABILIZING SELECTION
Natural selection that favours intermediate values of a quantitative trait.

which in turn has unexpectedly shown important functions for variants of these seemingly mundane enzymes[54]. A case in point concerns LDH in the fish *Gillichthys*, in which natural allelic variation in a presumed innocuous region of the protein, which was far from the active site and recognized functional domains, led to the discovery of a new functional role for this region[55].

Phylogenetic information is becoming especially important in these endeavours. A pre-existing phylogeny is a powerful tool for identifying adaptation and homology (for examples see REFS 56,57). In reverse, homology allows enough cDNAs to be identified to enable TP of many organisms that lack a fully sequenced genome (for examples see REFS 5,57). Similarly, new techniques such as PHYLOGENETIC FOOTPRINTING and PHYLOGENETIC SHADOWING can identify regulatory elements in cases in which standard bioinformatic algorithms might be uninformative[58]. Clearly, phylogenetic and comparative analyses provide many opportunities to advance the functional and evolutionary understanding of EEFG[59–61].

Conclusion

The contributions of genomics and post-genomic technologies to EEFG are several-fold, and include high-throughput tools, comparative databases that allow discovery in genomically obscure taxa and new experimental techniques such as RNAi. Indeed, genomics has enabled EEFG, rather than initiating or shifting its paradigm, and this is essential for any field that attempts the challenging task of integrating genes, function, ecology and evolution in its research programmes.

The EEFG community is in its youth and is still united more by a shared value system (that is, that interdisciplinary approaches are required to understand the success of wild organisms in natural environments) than by discoveries of law-like properties or general principles. Doubtless these discoveries will be forthcoming in due course, but at different rates for different portions of the community. For microbial systems in which genomes are more readily sequenced, evolution is rapid[62] and the economic and health consequences are stark, progress should be fast[63].

In studies of multicellular eukaryotes, the acceleration of progress in EEFG might mean overcoming several challenges. For example, the community is divided among many models, with a meta-divide between the customary model organisms of biomedical research and the charismatic non-classical models of natural history and comparative biology. For at least the next five years, manipulative genetics will be intractable in most multicellular eukaryotes. This intractability *per se* is not an impassable barrier to progress. The Human Genome Project and its successors show how a unified community of investigators can succeed even with a difficult and complex organism, and we have reviewed the substantial progress that has already been achieved with species other than the classical laboratory-based models. Rather, the challenge emerges from a continuing near-philosophical debate, the extremes of which are whether to make the customary biomedical model organisms do 'double duty' as model wild organisms, for which they are often not well suited, or to forego the advantages of massive community support to optimize the insights that are emerging from the study of non-classical model organisms. Our expectation is that technological advances will eventually make this debate moot.

Also, EEFG has set itself an ambitious further goal: the understanding of wild organisms *in situ* and their evolution. Such a goal, in principle, requires the application of every scientific discipline and model, if not the breaking-down of boundaries among disciplines and models. Achieving this goal will require new forms of training and education, and a greater role for collaborative research and supportive academic, private and governmental institutions (see summary in REF. 64). Ultimately, this challenge is concerned less with the technology than with the re-engineering of scientific attitudes, training and institutions.

Martin E. Feder is at the Committee on Evolutionary Biology, Committee on Genetics, and Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57th Street, Chicago, Illinois 60637, USA.

Thomas Mitchell-Olds is at the Department of Genetics and Evolution, Max-Planck Institute of Chemical Ecology, 07745 Jena, Germany.

Both authors contributed equally to this article. Correspondence to T.M.-O. e-mail: tmo@ice.mpg.de

doi:10.1038/nrg1128

1. Bartholomew, G. A. in *New Directions in Ecological Physiology* (eds. Feder, M. E., Bennett, A. F., Burggren, W. W. & Huey, R. B.) 11–35 (Cambridge Univ. Press, Cambridge, 1987).
2. Feder, M. E. in *New Directions in Comparative Developmental Physiology* (eds. Burggren, W. W. & Warburton, S.) (in the press).
3. Mitchell-Olds, T. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* **16**, 693–700 (2001).
4. Oleksiak, M., Churchill, G. & Crawford, D. Variation in gene expression within and among natural populations. *Nature Genet.* **32**, 261–266 (2002).
5. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
6. Glazier, A. M., Nadeau, J. H. & Aitman, T. J. Finding genes that underlie complex traits. *Science* **298**, 2345–2349 (2002).
7. Ford, M. J. Application of selective neutrality tests to molecular ecology. *Mol. Ecol.* **11**, 1245–1262 (2002).
8. Meagher, T. R. & Futuyma, D. Evolution, science, and society. *Am. Nat.* **158**, 1–46 (2001).
9. Ben-Shahar, Y., Robichon, A., Sokolowski, M. B. & Robinson, G. E. Influence of gene action across different time scales on behavior. *Science* **296**, 741–744 (2002).
10. Baldwin, I. T. An ecologically motivated analysis of plant–herbivore interactions in native tobacco. *Plant Physiol.* **127**, 1449–1458 (2001).
11. Weinig, C. *et al.* Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* **162**, 1875–1884 (2002).
12. Pigliucci, M., Pollard, H. & Cruzan, M. Comparative studies of evolutionary responses to light environments in *Arabidopsis*. *Am. Nat.* **161**, 68–82 (2003).
13. Tian, D. C., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. Pleiotropic cost of R-gene mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
14. Cavalieri, D., Townsend, J. P. & Hartl, D. L. Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl Acad. Sci. USA* **97**, 12369–12374 (2000).
15. Feder, M. E. Engineering candidate genes in studies of adaptation: the heat-shock protein Hsp70 in *Drosophila melanogaster*. *Am. Nat.* **154**, 55–66 (1999).
16. de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–89 (1998).
17. Parichy, D. M. & Johnson, S. L. Zebrafish hybrids suggest genetic mechanisms for pigment pattern diversification in *Danio*. *Dev. Genes Evol.* **211**, 319–328 (2001).
18. Kopp, A., Duncan, I. & Carroll, S. B. Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* **408**, 553–559 (2000).
19. Gibson, G. Microarrays in ecology and evolution: a preview. *Mol. Ecol.* **11**, 17–24 (2002).
20. Daborn, P. J. *et al.* A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253–2256 (2002).
21. Fowler, S. & Thomashow, M. F. *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* **14**, 1675–1690 (2002).
22. Churchill, G. A. Fundamentals of experimental design for cDNA microarrays. *Nature Genet.* **32**, 490–495 (2002).
23. Slonim, D. K. From patterns to pathways: gene expression data analysis comes of age. *Nature Genet.* **32**, 502–508 (2002).
24. Chuaqui, R. F. *et al.* Post-analysis follow-up and validation of microarray experiments. *Nature Genet.* **32**, 509–514 (2002).
25. Toma, D. P., White, K. P., Hirsch, J. & Greenspan, R. J. Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait. *Nature Genet.* **31**, 349–353 (2002).
26. Beja, O. *et al.* Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**, 630–633 (2002).
27. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810 (2002).
28. Jackson, R. B. *et al.* Linking molecular insight and ecological research. *Trends Ecol. Evol.* **17**, 409–414 (2002).
29. Jackson, R. B., Moore, L. A., Hoffmann, W. A., Pockman, W. T. & Linder, C. R. Ecosystem rooting depth determined with caves and DNA. *Proc. Natl Acad. Sci. USA* **96**, 11387–11392 (1999).
30. Roberts, S. P. & Feder, M. E. Changing fitness consequences of *hsp70* copy number in transgenic *Drosophila* larvae undergoing natural thermal stress. *Funct. Ecol.* 353–357 (2000).
31. Maloof, J. N. *et al.* Natural variation in light sensitivity of *Arabidopsis*. *Nature Genet.* **29**, 441–446 (2001).
32. Nachman, M. W., Hoekstra, H. E. & D'Agostino, S. L. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl Acad. Sci. USA* **100**, 5268–5273 (2003).
33. Meagher, S., Penn, D. J. & Potts, W. K. Male–male competition magnifies inbreeding depression in wild house mice. *Proc. Natl Acad. Sci. USA* **97**, 3324–3329 (2000).

34. Lexer, C., Welch, M. E., Durphy, J. L. & Rieseberg, L. H. Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Mol. Ecol.* **12**, 1225–1235 (2003).
35. Peichel, C. *et al.* The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).
36. Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
37. Beldade, P., Brakefield, P. M. & Long, A. D. Contribution of Distal-less to quantitative variation in butterfly eyespots. *Nature* **415**, 315–318 (2002).
38. Frary, A. *et al. fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88 (2000).
39. El-Assal, S. E.-D., Alonso-Blanco, C., Peeters, A. J. M., Raz, V. & Koornneef, M. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genet.* **29**, 435– 440 (2001).
40. Dilda, C. L. & Mackay, T. F. C. The genetic architecture of *Drosophila* sensory bristle number. *Genetics* **162**, 1655–1674 (2002).
41. Robin, C., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. C. *hairy*: a quantitative trait locus for *Drosophila* sensory bristle number. *Genetics* **162**, 155–164 (2002).
42. Zhang, X. S. & Hill, W. G. Joint effects of pleiotropic selection and stabilizing selection on the maintenance of quantitative genetic variation at mutation-selection balance. *Genetics* **162**, 459–471 (2002).
43. Holub, E. B. The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature Rev. Genet.* **2**, 516–527 (2001).
44. Jakob, K. *et al. Pseudomonas viridiflava* and *P. syringae* — natural pathogens of *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* **15**, 1195–1203 (2002).
45. Grant, M. R. *et al.* Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **95**, 15843–15848 (1998).
46. Brakefield, P. M. & Liebert, T. G. Evolutionary dynamics of declining melanism in the peppered moth in the Netherlands. *Proc. Royal Soc. Lond. B* **267**, 1953–1957 (2000).
47. Clegg, M. T. & Durbin, M. L. Tracing floral adaptations from ecology to molecules. *Nature Rev. Genet.* **4**, 206–215 (2003).
48. Watt, W. B. Avoiding paradigm-based limits to knowledge of evolution. *Evol. Biol.* **32**, 73–96 (2000).
49. Feder, M. E. & Watt, W. B. in *Genes in Ecology* (eds. Berry, R. J., Crawford, T. J. & Hewitt, G. M.) 365–391 (Blackwell Scientific, Oxford, 1993).
50. Fay, J. C., Wyckoff, G. J. & Wu, C. I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026 (2002).
51. Swanson, W. J., Zhang, Z. H., Wolfner, M. F. & Aquadro, C. F. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl Acad. Sci. USA* **98**, 2509–2514 (2001).
52. Riley, R., Jin, W. & Gibson, G. Contrasting selection pressures on components of Ras-mediated signal transduction in *Drosophila*. *Mol. Ecol.* **12**, 1315–1323 (2003).
53. Schulte, P. M. Environmental adaptations as windows on molecular evolution. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **128**, 597–611 (2001).
54. Watt, W. B. & Dean, A. M. Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. *Ann. Rev. Genet.* **34**, 593–622 (2000).
55. Fields, P. A., Kim, Y. S., Carpenter, J. F. & Somero, G. N. Temperature adaptation in *Gillichthys* (Teleost: Gobiidae) A(4)-lactate dehydrogenases: identical primary structures produce subtly different conformations. *J. Exp. Biol.* **205**, 1293–1303 (2002).
56. Farrell, B. D. *et al.* The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodinae). *Evolution* **55**, 2011–2027 (2001).
57. Oleksiak, M. F., Kolell, K. J. & Crawford, D. L. Utility of natural populations for microarray analyses: isolation of genes necessary for functional genomic studies. *Marine Biotechnol.* **3**, 203–211 (2001).
58. Boffelli, D. *et al.* Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391–1394 (2003).
59. Zdobnov, E. *et al.* Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159 (2002).
60. Charlesworth, D., Charlesworth, B. & McVean, G. A. T. Genome sequences and evolutionary biology, a two-way interaction. *Trends Ecol. Evol.* **16**, 235–242 (2001).
61. Ureta-Vidal, A., Ettwiller, L. & Birney, E. Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nature Rev. Genet.* **4**, 251–262 (2003).
62. Cooper, T. F., Rozen, D. E. & Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100**, 1072–1077 (2003).
63. Elena, S. F. & Lenski, R. E. Microbial genetics: evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).
64. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life. *Annu. Rev. Genom. Human. Genet.* **2**, 343–372 (2001).
65. Wittbrodt, J., Shima, A. & Schartl, M. Medaka — a model organism from the far East. *Nature Rev. Genet.* **3**, 353–364 (2002).
66. Kocher, T. D., Lee, W.-J., Sobolewska, H., Penman, D. & McAndrew, B. A Genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* **148**, 1225–1232 (1998).
67. Pennisi, E. Recharged field's rallying cry: gene chips for all organisms. *Science* **297**, 1985–1987 (2002).
68. Whitfield, C. W. *et al.* Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res.* **12**, 555–566 (2002).
69. Davidson, E. H., McClay, D. R. & Hood, L. Regulatory gene networks and the properties of the developmental process. *Proc. Natl Acad. Sci.* **100**, 1475–1480 (2003).
70. Bradshaw, H. D., Ceulemans, R., Davis, J. & Stettler, R. F. Emerging model systems: poplar (*Populus*) as a model forest tree. *J. Plant Growth Reg.* **19**, 306–313 (2000).

Acknowledgements

We thank G. Bartholomew, J. Bergelson, B. Hill, C. Langley, M. Lynch, T. MacKay, M. Nachman, M. Turelli, W. Watt and three anonymous referees for helpful discussion and comments. We are grateful to J. Bergelson, J. Colbourne, M. Lynch, M. Nachman and C. Weinig for access to unpublished manuscripts and proposals. T.M.-O was supported by the European Union (contract number QLRT-2000-01097), the Bundesminesterium für Bildung und Forschung, the United States National Science Foundation and the Max-Planck Gesellschaft. M.E.F. was supported by National Science Foundation grants, which also supported the establishment of the evolutionary and ecological functional genomics (EEFG) community. In lieu of a trans-Atlantic coin flip, the order of authorship was determined by random fluctuation in the Euro/Dollar exchange rate.

Online Links

DATABASES

The following term in this article is linked online to:
TAIR: http://www.arabidopsis.org
RPM1

FURTHER INFORMATION

Daphnia Genomics Consortium: http://daphnia.cgb.indiana.edu
EEFG: http://pondside.uchicago.edu/~feder/EEFG.html
EEFG Conference: http://www.grc.uri.edu/programs/2003/evolecol.htm
Genome News Network: http://gnn.tigr.org
Genomes Online Database: http://www.genomesonline.org
Martin E. Feder's web site: http://pondside.uchicago.edu/~feder/Martin_Feder.html
Max Planck Institute of Chemical Ecology Department of Genetics and Evolution: http://www.ice.mpg.de/departments/Gen
NemATOL: http://nematol.unh.edu/pro_des.php
Organism-Specific Genome Databases: http://restools.sdsc.edu/biotools/biotools10.html
TIGR Gene Indices: http://www.tigr.org/tdb/tgi
Wild Relatives of *Arabidopsis*: www.ice.mpg.de/departments/Gen/wild.htm

Access to this interactive links box is free online.

TIMELINE

An everlasting pioneer: the story of *Antirrhinum* research

Zsuzsanna Schwarz-Sommer, Brendan Davies and Andrew Hudson

Despite the tremendous success of *Arabidopsis thaliana*, no single model can represent the vast range of form that is seen in the ~250,000 existing species of flowering plants (angiosperms). Here, we consider the history and future of an alternative angiosperm model — the snapdragon *Antirrhinum majus*. We ask what made *Antirrhinum* attractive to the earliest students of variation and inheritance, and how its use led to landmark advances in plant genetics and to our present understanding of plant development. Finally, we show how the wide diversity of *Antirrhinum* species, combined with classical and molecular genetics — the two traditional strengths of *Antirrhinum* — provide an opportunity for developmental, evolutionary and ecological approaches. These factors make *A. majus* an ideal comparative angiosperm.

"*Antirrhinum* has always allowed new, and frequently surprising, insights to be made into the nature, variability and manifestation of genetic substance and, even today, the rich variety of appearance in the genus *Antirrhinum* offers an inexhaustible resource for genetics-based studies in developmental biology, biochemistry and evolution". With this sentence, Hans Stubbe[1] justified his motivation to write a comprehensive monograph on *Antirrhinum* in 1966; this article shows that his assessment of *Antirrhinum* is as valid today as it was half a century ago.

It might seem surprising that such a familiar ornamental plant could be used as an experimental system. In fact, *Antirrhinum* was used in the earliest studies of inheritance by Darwin and Mendel, and became established as a model by Erwin Baur (FIG. 1) during the first decades of the twentieth century (TIMELINE). Interest in *Antirrhinum* declined after 1930, because of the emergence of