Ecological genomics, from genes to ecosystems

(Detlef Weigel and Diethard Tautz
detlef.weigel@tuebingen.mpg.de
tautz@evolbio.mpg.de)

At a glance

- *"Nothing in evolutionary biology makes sense except in the light of ecology." (Rosemary and Peter Grant* in *How and Why Species Multiply)*
- *To understand not only current patterns of biological diversity, but also to fully understand why organisms today are the way they are, we need to comprehend how they have been shaped by their natural environment.*
- *Until now, it has not been possible to connect our advanced knowledge of the basic principles of cellular functions to insights from ecology. The current technological revolution in genetics promises to change this, and in turn will lead to unification of disparate biological disciplines.*

Definition of research

The goal of the newly emerging discipline of ecological genomics is to bridge the current gap between genetic studies in the laboratory, largely focused on understanding basic cellular and developmental processes, with systems-level analyses of genetic adaptations and interactions between organisms in their natural setting.

Status of the field

Contemporary biology covers an enormous scale: On the one hand, structural biologists are discovering how basic cellular processes occur at the atomic level. On the other hand, the discipline of global ecology is trying to predict how climate change will affect broad patterns of biodiversity across the entire planet. Unfortunately, these fields of research are not continuous, and there has been a glaring gap trying to connect genetic processes in individual organisms of a single species with ecological processes resulting from interaction of many individuals drawn from the entire species inventory of an ecosystem. One reason for this has been that model species for evolutionary and ecological research have typically suffered from a lack of genetic tools, while conversely generally very little, if anything, has been known about the ecology of well-established genetic models.

The power of genetics -- which has been bolstered in the past decade by a broad range of genomic tools -- in dissecting basic cellular and developmental processes is undisputed. It has also been increasingly exploited to understand the interaction between different organisms, particularly between pathogens and their hosts. Because of the technological investments required, genetic and genomic approaches have, however, traditionally been applied only to a small set of carefully chosen model species. These species, such as nematode worms, fruit flies, or the small cabbage relative Arabidopsis, are typically adapted to the laboratory environment and are usually genetically inbred. They necessarily represent abstractions of biological systems, and it is difficult, if not impossible, to infer from these studies how species adapt to their natural environment. But this situation is now changing. A multitude of new approaches are becoming available to begin to address the

genetics of adaptations and ecological interactions in natural populations. It is today much easier to develop genetic approaches for ecological model systems than trying to understand the ecology of established genetic model organisms.

In contrast to laboratory-induced mutant phenotypes, which are easily distinguished from non-mutant phenotypes, values for adaptive traits are normally continuous, reflecting the effect of several loci on the phenotype. Despite their genetically more complex architecture, such phenotypes can be studied by forward genetic techniques such as quantitative trait locus (QTL) analysis of experimental populations. An alternative is the use of genome-wide association methods, which takes advantage phenotypic and genotypic information for natural populations, either at the local or global scale. In addition to these bottom-up approaches, evolutionary signatures in the genome can be exploited to screen systematically for genes involved in recent adaptations, without knowing the particular phenotype under selection. All of these approaches benefit enormously from the current revolution in sequencing technologies, which allows us to take an inventory of genetic variation at an unprecedented scale.

Together, these methods now enable us to bridge the gap between modern genetics and systems-level ecological studies, as classical models of evolutionary research have become amenable to genetic dissection. Using QTL mapping and genomic comparisons, it has already been possible to identify specific genes involved in adaptive traits like coat color in deer mice, plate armor in sticklebacks, beak shape in Darwin finches and flowering time in plants (Ellegren & Sheldon 2008). Genes responsible for reproductive isolation among divergent lineages and hence with possible causal roles in speciation have been identified as well (Noor & Feder 2006). In addition, selective sweep screens have provided a rich set of candidate loci for recent adaptations that await further genetic analyses (Turner et al. 2005). Second (and soon third) generation sequencing technologies provide genomic access to almost any species and its natural genetic variation, regardless of whether the species can be kept in the laboratory (Mardis 2008). With these technologies, one can also screen environmental samples at a large scale, to identify novel genes and ecologically relevant genetic pathways (Rokas & Abbot 2009).

The advances in genetic technologies are being paralleled by dramatic improvements in imaging and remote sensing, with which one can capture both spatial and temporal components of dynamic interactions between individuals and their natural environment. So far, remote sensing, which includes satellite based, aerial, as well as land-based capturing of electromagnetic information, has mostly been used to record broad trends in ecosystems. However, as imaging sensors, e.g., digital cameras, become cheaper and cheaper, and can be integrated through wireless networks, we can look forward to obtaining very detailed images at very high spatial and temporal resolution (Gigapix Systems LLC 2009). Similar to the sequencing revolution we are witnessing, data acquisition is becoming less and less of an issue, with the challenges being primarily in data analysis.

The combined advances we envision will be similar to the dramatic breakthroughs in human genetics, where genome-wide association studies have completely revolutionized the field. Before 2006, less than a handful of genes that were risk factors for common diseases such as diabetes were known. Since then, using very detailed SNP genotyping of very large number of patients, initially thousands, but now routinely tens of thousands, combined with disease information for these, hundreds of genes that contribute to common health problems have been identified (Kruglyak 2008).

International activities

The field is currently in an early phase of expansion. A Gordon Conference on Evolutionary & Ecological Functional Genomics was held in 2009, which covered many aspects of this emerging field. A special funding program on " Ecological and Evolutionary functional genomics" coordinated by the European Science Foundation (ESF) has been implemented in 2009, with funding starting in 2010. Next generation sequencing technologies have reached ecology, as testified in a special issue of *Molecuar Ecology*, the leading journal in the field. This issue will appear early in 2010 and will be called "Next generation molecular ecology" to mark this step forward.

A select number of research centers are exploiting the current sequencing revolution for evolutionary or ecological studies, such as the NIH funded Stanford Genome Evolution Center. Similarly, the NSF funded National Evolutionary Synthesis Center (NESCent) will likely facilitate many of the approaches discussed here. Finally, ihe DOE funded Joint Genome Institute (JGI) is increasingly turning its community sequencing program to organisms of ecological relevance.

Research opportunities and needs

The research options in ecological genomics have been dramatically increased with the explosive development of DNA sequencing power and genomic approaches in the past years, paralleled by imaging and remote sensing technologies, which allow monitoring of natural populations.

Within the next two years, it will be feasible to perform comparative sequencing of thousands of individual genomes from a species, to obtain genome scale insights into natural variation. We are therefore no longer limited to populations that are created in the laboratory by crossing, but rather can analyze directly natural populations that have been phenotyped in the wild. Thus, ecological and evolutionary model species that have not been previously tractable for genetic analysis can now be included in explicit molecular approaches. Similarly, complex communities of microorganisms can be efficiently sequenced to obtain insights into species composition and gene contents, but the challenge will be to understand how genomes interact with each other. The speed of data generation will soon outstrip the capacity of data analysis, unless radically new approaches are being implemented, through close collaboration of biologists, engineers and informaticians. Computational resources will very likely become limiting and the efficiency of analysis algorithms will need to be greatly improved.

The major challenge of the next five to ten years will clearly be to find a balance between the awesome power of data generation technologies and the depth of scientific questions that can be addressed with these data. At the same time, it will be necessary to keep in mind that genomic data on their own do not provide much biological meaning. Hence, biological concepts for comparative analysis and experimental strategies for functional analysis that optimally exploit the advances in data collection will have to be developed.

Scientific topics that will need particular attention include

- genes and molecular processes involved in adaptation
- genetic mechanisms of speciation
- the role of epistatic interactions in complex phenotypes
- the role of epigenetic changes in evolution and adaptation

Although these topics are firmly based on traditional genetic questions, all of them will require the further development of concepts and approaches in the light of evolutionary considerations. Successful exploration of these topics will form the basis for exploration of systems-levels questions such as

- predictability of evolutionary processes (inferred from studying the genetics of parallel adaptations)
- global change environmental genomics
- role of microbial communities in ecosystem stability
- interaction networks among organisms
- the genetics of nutrient fluxes in the environment

Two key compentencies from other disciplines will be required to approach these challenges. The first concerns a firm taxonomic knowledge, including species inventories and biodiversity censuses. This will in itself be bolstered by sequence based approaches (such as DNA barcoding and DNA taxonomy) as well as remote sensing technology. The second concerns a solid understanding of cooperation dynamics in interaction networks, based on game theoretical analysis.

Computational and experimental challenges

Until recently, the main repository for DNA sequences, GenBank, grew at about the same rate as computing power, following Moore's law and doubling every 18 months. The cost of sequencing has dropped from about 0.1 cents per base in 2005 to less than 0.001 cent today. The steep fall in sequencing cost and the concomitant increase in sequencing speed outpaces the improvement in computational power, and this will likely continue for several years. If the car industry was operating under a similar exponential rule, we'd now be driving cheap supersonic cars.

To look at this differently, in the middle of 2009, GenBank contained about 300 Gb of sequence. Already today, next generation sequencers can generate up to 2 Gb of sequence. Current algorithms for sequence data analysis have their roots in the early times of sequence acquisition, when data were scarce and computational efficiency not necessarily a large concern. But this is bound to change. For example, algorithms that are used for aligning genome sequences tend to have an exponential computing time requirement as the number of sequences in the analysis increases. New concepts and approaches will be required to reduce this into a linear requirement (Domazet-Loso & Haubold 2009). Bioinformatics will have to use latest computing technology to address the imminent challenges of data analysis in only a few years' time.

The experimental challenges to understand the functions of genes and their variants in adaptive processes may not be less. Every biological function is subject to selection – positive selection during its emergence and divergent or stabilizing selection thereafter. The power of selection is a function of effective population size, with the minimum selection coefficient to create or to stabilize a genetic function being $1/2N_e$, where N_e is the effective population size. The N_e of laboratory inbred organisms is usually much less than 10, while the N_e of a typical natural population will often be larger than 10^4. In other words, while laboratory selection experiments with metazoans can only identify new variants that provide a carrier a 10% advantage over its siblings, nature can easily select for improvements that are merely a hundredth of a percent better than the existing material. Accordingly, it has to be expected that a large part of genetic functions has escaped analysis, since their mutational effects can only be studied in population level experiments (Figure 1).

Given that the genes and alleles involved in such functions can now be uncovered by genomic approaches, it will be necessary to develop experimental approaches that study weak, but evolutionarily significant adaptations in a population context. This is most easily feasible with small organisms, such as yeast, where large populations can be grown in culture for extended times to assess competition of different genotypes. This approach has been used to identify possible functions for genes that lack an obvious knockout phenotype under standard experimental conditions (Giaever et al. 2002). Despite the power of these approaches, one of the limitations of analyzing single-gene effects is that epistatic interactions, in which different genes have combined effects that cannot be predicted based on the activities of individual genes alone, are extremely difficult to analyze, because there are so many possible combinations. It is here where the analysis of natural genotypes, which presumably harbor positively selected epistatic pairs of genes, will have a clear advantage, even though the statistical challenges are enormous (McKinney et al. 2009).

Expected outcome and benefit

The function of genes has so far largely been studied in a very limited number of species, and in the context of individual organisms. The next five to ten years will see a massive shift towards the inclusion of the ecological and evolutionary context in gene function analysis. As such, genetics will move on from a largely biomedical perspective to an ecological perspective, with special relevance for global change questions. A full understanding of the ecosystem, its services and its stability will not be possible without understanding the genetics of adaptations and community interactions.

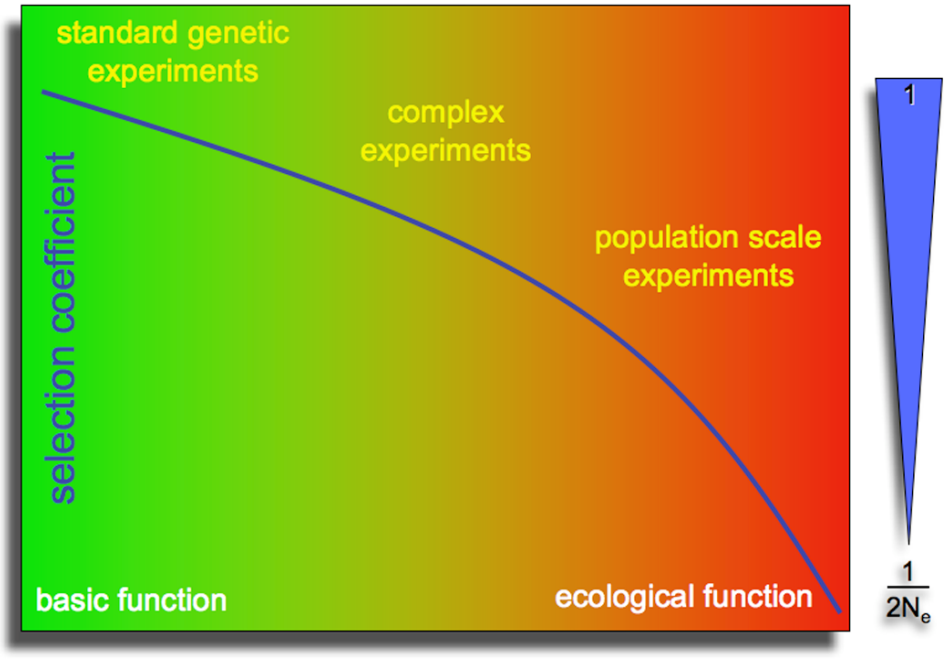

Fig. 1: Scheme of the relationship between selection coefficients and genetic functions. Standard laboratory experiments operate with selection coefficients of up to 1 (e.g. lethal alleles). However, the minimum selection coefficient that can maintain a genetic funtion is $1/2N_e$. Such genetic functions will be particularly relevant in an ecological context and tracing such functions will require complex experiments, up to population scale approaches.

References

Domazet-Loso, M. & Haubold, B. 2009 Efficient estimation of pairwise distances between genomes. *Bioinformatics*.

Ellegren, H. & Sheldon, B. C. 2008 Genetic basis of fitness differences in natural populations. *Nature* **452**, 169-75.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K. D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kotter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C. Y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W. & Johnston, M. 2002 Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-91.

Gigapix Systems LLC. 2009 Gigapan Systems Online.

Kruglyak, L. 2008 The road to genome-wide association studies. *Nat. Rev. Genet.* **9**, 314-8.

Mardis, E. R. 2008 Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**, 387-402.

McKinney, B. A., Crowe, J. E., Guo, J. & Tian, D. 2009 Capturing the spectrum of interaction effects in genetic association studies by simulated evaporative cooling network analysis. *PLoS Genet.* **5**, e1000432.

Noor, M. A. & Feder, J. L. 2006 Speciation genetics: evolving approaches. *Nat. Rev. Genet.* **7**, 851-61.

Rokas, A. & Abbot, P. 2009 Harnessing genomics for evolutionary insights. *Trends Ecol. Evol.* **24**, 192-200.

Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. 2005 Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, e285.