

## EDITORIAL

## Next Generation Molecular Ecology

DIETHARD TAUTZ,\* HANS ELLEGREN† and  
DETLEF WEIGEL‡

\*Max Planck Institute for Evolutionary Biology, Plön, Germany; E-mail: tautz@evolbio.mpg.de, †Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden; E-mail: Hans.Ellegren@ebc.uu.se, ‡Max Planck Institute for Developmental Biology, Tübingen, Germany; E-mail: weigel@tue.mpg.de

Calling a special issue of *Molecular Ecology* that deals with next-generation sequencing 'Next generation Molecular Ecology' seems to be all too obvious. However, as the contributions to this issue came in, it became increasingly clear that next-generation sequencing (NGS) is more than just a technical leap – it will result in a transformation of how we think about molecular ecology as a discipline. Progress in molecular ecology has always been driven by advances in DNA technologies, but the current revolution could not really be foreseen. 'Ecological genomics' has suddenly become a valid emerging discipline.

Contemporary biology covers already an enormous scale. On the one hand, structural biologists are discovering how basic cellular processes occur at the atomic level. On the other, ecologists are trying to predict how climate change will affect broad patterns of biodiversity across the entire planet. But these fields have heretofore not been continuous, and there has been a glaring gap in connecting genetic processes in individual organisms with ecological processes that result from the interaction of many individuals in 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 very little, if anything, has been known about the ecology of well-established genetic models.

The power of genetics in dissecting basic cellular and developmental processes is undisputed, particularly since the addition in the past decade of a broad range of genomic tools. Its power has also been increasingly exploited to understand the interaction between different organisms, such as 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

laboratory studies how species adapt to their natural, highly variable environment. But this situation is now changing. A multitude of new approaches are becoming available to address the genetics of adaptation 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. The NGS technologies are the key to this.

Within the coming years, it will be feasible to obtain genome scale insights into natural variation through comparative sequencing of thousands of individual genomes from a species. We are thus no longer limited to populations that have been created in the laboratory by crossing, but rather can directly analyse 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 content. Obtaining the complete genetic sequence information of a whole ecosystem may be in reach within the next 10 years. However, the more significant challenge will be to understand how the many genomes present in an ecosystem interact with each other. The speed of data generation will soon outstrip the capacity of data analysis unless radically new approaches are implemented. Computational resources will very likely become limiting and the efficiency of analysis algorithms will need to be greatly improved.

The main problem we will all be faced with in the coming years is finding 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 can only provide new biological insights if the samples are well chosen and studies are appropriately designed. 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.

It is therefore good to see that the contributions to this volume are all firmly based on significant ecological and evolutionary questions. They explore avenues of how NGS technology can be profitably applied to ecological questions. Although it is predictable that in only a few years the data sets presented here will seem to be tiny steps in ecological genomics, their ways of approaching the problems and of analysing the data may set future standards.

### Contributions

There are 21 contributions in this volume, spanning the broad field of molecular ecology from community analysis to population differentiation, from adaptation to speciation. They exemplify the broad range of experimental options and biological questions that come with the new technologies. Although each of the contributions has its own specific approach to using NGS data, they can nonetheless be grouped into major themes, such as taxonomy, single nucleotide polymorphism (SNP) detection and analysis, expression studies, tracing adaptations and bioinformatic developments.

### Taxonomy

Getting a grip on eukaryotic meiofaunal diversity is currently at the forefront of DNA based approaches in taxonomy. Four of the contributions deal with this theme. Creer *et al.* (2010) address it in a broad context and try to identify the major problems and pitfalls that are associated with the technology. Stoeck *et al.* (2010) use a multiple marker approach to identify an unexpected diversity of eukaryotic taxa in marine anoxic waters. Medinger *et al.* (2010) focus on alveolates in oligotrophic freshwater lakes to trace the seasonal community turnover. Finally, Jumpponen *et al.* (2010) estimate the fungal richness, diversity and community composition in ectomycorrhiza of oak trees in different ecological contexts.

Prokaryotic microbial taxonomy has always been based on DNA sequences, but can now be done at an ecological scale. Schütte *et al.* (2010) characterize bacterial diversity along two chronosequences in an Arctic glacier forefield. Coolon *et al.* (2010) provide an insight into bacterial community turnover on heavy metal contaminated sites – not only from the perspective of the soil bacterial communities, but also from the perspective of bacteria in the gastrointestinal tracts of rodents living on these soils. Roossinck *et al.* (2010) have developed a new approach to link plant viruses to their hosts and have discovered thousands of new such associations.

### Single nucleotide polymorphisms

The distribution and frequencies of SNPs are key to understand diversity between closely related populations and species. NGS can be used to discover these on a large scale. Van Bers *et al.* (2010) used short read sequences of the great tit (*Parus major*) genome to identify novel SNPs and to map them in reference to the zebra finch genome. Whittall *et al.* (2010) show that the acquisition of entire organelle sequences allows the identification of cryptic variation in pines, which would otherwise have been difficult to detect. Renaut *et al.* (2010) use a SNP survey between closely related whitefish to identify classes of genes that might have played a role in adaptive divergence. Buggs *et al.* (2010) have generated SNP markers from *Tragopogon* plant species to study differential expres-

sion of genes contributed by the two genomes that have combined to form an allopolyploid species. Cannon *et al.* (2010) use short read sequence data comparisons to identify population specific markers for endangered timber species.

### Expression

Comparisons of gene expression patterns in population contexts have so far been limited by the availability of microarrays for the respective species. NGS is going to change this, since counting sequence reads from ESTs provides also expression information and can be applied to any species. Wolf *et al.* (2010) explore this approach to study an extremely young species pair of crows and suggest that expression divergence may be faster than neutral nucleotide divergence. Goetz *et al.* (2010) use transcriptional profiling to trace the genetics of the phenotypic differences of two ecologically distinct forms of a fish species. Elmer *et al.* (2010) look at a cichlid radiation in a crater lake and identify candidate genes for diversifying selection. Fontanillas *et al.* (2010) discuss the power of being able to detect allelic expression differences in heterozygous individuals for dissecting the genetic component of natural variation in gene expression. Harr & Turner (2010) address variation in alternative splicing of genes between subspecies of the house mouse as a source of expression change.

### Adaptation

A direct tracing of genes involved in adaptation can also be significantly accelerated by NGS. Ferguson *et al.* (2010) study the mimetic wing patterns of *Heliconius* butterflies as an example of an adaptive radiation and convergent evolution and use NGS to annotate and characterize mapped genes. Kloch *et al.* (2010) show that major histocompatibility complex allele associations with parasites can be studied at a population level scale. Künstner *et al.* (2010) use sequence information from a large number of bird species to demonstrate the care needed when inferring selection from the observed incidence of nonsynonymous and synonymous substitutions.

### Bioinformatics

One contribution deals with the computational challenge associated with making optimal use of the huge NGS data sets. Haubold *et al.* (2010) show that important population genetic parameters, such as population mutation and recombination rate, can be estimated from shotgun sequencing of individuals and they provide the necessary software for this.

### Outlook

Genomic analyses have so far been limited to a small number of model species. Soon, we will see a massive shift

towards the inclusion of the ecological and evolutionary context in genomic analysis. As such, genomics will move on from a largely biomedical perspective to an ecological perspective, with special relevance for global change questions. A full understanding of ecosystems, their services and stability will not be possible without understanding the genetics of adaptations and community interactions. The enormous power of genomic analysis based on NGS technologies promises to be a key factor in this endeavour.

### Conflicts of interest

The authors have no conflict of interest to declare and note that the sponsors of the issue had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

- Buggs RJA, Chamala S, Wu W *et al.* (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Molecular Ecology*, **19** (Suppl. 1), 132–146.
- Cannon CH, Kua C-S, Zhang D, Harting JR (2010) Assembly-free comparative genomics of short-read sequence data discovers the needles in the haystack. *Molecular Ecology*, **19** (Suppl. 1), 146–160.
- Coolon JD, Jones KL, Narayanan S, Wisely SM (2010) Microbial ecological response of the intestinal flora of *Peromyscus maniculatus* and *P. leucopus* to heavy metal contamination. *Molecular Ecology*, **19** (Suppl. 1), 67–80.
- Creer S, Fonseca VG, Porazinska DL *et al.* (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, **19** (Suppl. 1), 4–20.
- Elmer KR, Fan S, Gunter HM *et al.* (2010) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molecular Ecology*, **19** (Suppl. 1), 197–211.
- Ferguson L, Lee SF, Chamberlain N *et al.* (2010) Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the *HMYb/Sb* locus. *Molecular Ecology*, **19** (Suppl. 1), 240–254.
- Fontanillas P, Landry CR, Wittkopp PJ *et al.* (2010) Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Molecular Ecology*, **19** (Suppl. 1), 212–227.
- Goetz F, Rosauer D, Sitar S *et al.* (2010) A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Molecular Ecology*, **19** (Suppl. 1), 176–196.
- Harr B, Turner LM (2010) Genome-wide analysis of alternative splicing evolution among *Mus* subspecies. *Molecular Ecology*, **19** (Suppl. 1), 228–239.
- Haubold B, Pfaffelhuber P, Lynch M (2010) mlRho – A program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Molecular Ecology*, **19** (Suppl. 1), 277–284.
- Jumpponen A, Jones KL, Mattox J, Yaege C (2010) Massively parallel 454-sequencing of fungal communities in *Quercus* spp. Ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Molecular Ecology*, **19** (Suppl. 1), 41–53.
- Kloch A, Babik W, Bajer A, Siński E, Radwan J (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Molecular Ecology*, **19** (Suppl. 1), 255–265.
- Künstner A, Wolf JBW, Backström N *et al.* (2010) Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 birds species. *Molecular Ecology*, **19** (Suppl. 1), 266–276.
- Medinger R, Nolte V, Pandey RV *et al.* (2010) Diversity in a hidden world: potential and limitation of next generation sequencing for surveys of molecular diversity of Eukaryotic microorganisms. *Molecular Ecology*, **19** (Suppl. 1), 32–40.
- Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology*, **19** (Suppl. 1), 115–131.
- Roosinck MJ, Saha P, Wiley GB *et al.* (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology*, **19** (Suppl. 1), 81–88.
- Schütte UME, Abdo Z, Foster J *et al.* (2010) Bacterial diversity in a glacier foreland of the High Arctic. *Molecular Ecology*, **19** (Suppl. 1), 54–66.
- Stoeck T, Bass D, Nebel M *et al.* (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, **19** (Suppl. 1), 21–31.
- Van Bers NEM, van Oers K, Kerstens HHD *et al.* (2010) Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. *Molecular Ecology*, **19** (Suppl. 1), 89–99.
- Whittall JB, Syring J, Parks M *et al.* (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology*, **19** (Suppl. 1), 100–114.
- Wolf JBW, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D (2010) Nucleotide divergence versus gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, **19** (Suppl. 1), 162–175.