NEWS AND VIEWS

MEETING REVIEW

Two Decades of Molecular Ecology: where are we and where are we heading?

EVELINE T. DIEPEVEEN* and WALTER SALZBURGER*

Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

The twentieth anniversary of the journal *Molecular Ecology* was celebrated with a symposium on the current state and the future directions of the field. The event, organized by Tim Vines and Loren Rieseberg, took place on the opening day of the First Joint Congress on Evolutionary Biology organized by the American Society of Naturalists (ASN), the Canadian Society for Ecology and Evolution (CSEE), the European Society for Evolutionary Biology (ESEB), the Society for the Study of Evolution (SSE) and the Society of Systematic Biologists (SSB) in Ottawa (Canada) from 6–10 July 2012. The get together of these five societies created a truly international and exciting "Evolution conference" and the ideal framework for the *Molecular Ecology* symposium. Its thirteen talks were grouped into the five different subject areas of the journal: Speciation and Hybridization; Landscape Genetics, Phylogeography and Conservation; Ecological Genomics and Molecular Adaptation; Kinship, Parentage and Behaviour; Ecological Interactions. Each session was followed by a panel discussion on the future direction of the subfield. That more than 300 colleagues registered for this special symposium illustrates the broad interest in, and appreciation of, molecular ecology – both the field and the journal.

Next-generation (Sequencing in the field of) Molecular Ecology

The overlapping and dominating theme of almost all sessions (except maybe Kinship, Parentage and Behaviour) was next-generation sequencing (NGS). Taking the sympo-

Correspondence: Walter Salzburger, Fax: +41 61 267 0301; E-mail: walter.salzburger@unibas.ch

sium as a snapshot of what is going on in the community, then it seems as if NGS has 'conquered' the field of molecular ecology (see also Fig. 1). At the same time, the different talks exemplified the broad applicability of NGS and the various questions that can be addressed using NGS technologies. In the following, we summarize the symposium by focussing on how NGS is used in each of the subfields.

Alex Buerkle (University of Wyoming) presented some recent work on reproductive isolation in *Lycaeides* butterflies and *Manacus* birds based on genome-wide single-nucleotide polymorphism (SNP) screens of both parental species and hybrids (see Gompert *et al.* 2012). They found that many loci show signs of extreme introgression along hybrid zones and that there are many fine-scaled genomic footprints of differentiation and reproductive isolation, calling into question the genomic island view of speciation. Alex Widmer (ETH Zurich) talked about his work on a hybrid zone in *Silene.* Previous amplification fragment length polymorphism (AFLP) genomic scans revealed that outlier markers were typically found on the sex chromosomes. In the absence of a reference genome, they built a reference transcriptome to determine gene expression differences between sexes. This way they were able to show that dosage compensation does occur in *Silene* by overexpression of the X-chromosome in males when the Y-chromosome is down-regulated (Muyle *et al.* 2012). Tatiana Giraud (University of South Paris) used NGS to sequence expressed sequence tag (EST) libraries of four *Microbotryum* species – fungal pathogens specialized on different host plants (Caryophyllaceae). These libraries were subjected to a genome-wide dN/dS analysis, which revealed 42 loci under positive selection that could be implicated with functions in the host–parasite interactions (Aguileta *et al.* 2010).

In the second session, Rose Andrew (University of British Columbia) talked about her research on dune sunflowers (*Helianthus*) in which she applied RAD sequencing to twenty subpopulations to identify putative adaptive loci. Several strong peaks of F_{ST} outliers were detected, leading to the conclusion that seed mass and vegetation cover were both associated with the same genomic region. Next, Victoria Sork (University of California, Los Angeles) stated that NGS is the essential tool for the development of a new subfield: landscape genomics. With NGS, the whole genome can be screened for adaptive genetic variation in relation to geographic patterns, making candidate gene approaches somewhat redundant. Furthermore, massive NGS genome scans provide valuable information on neutral loci that can be used for, for example, inferring demographic processes.

The field of ecological genomics and molecular adaptation has gained tremendously from the NGS revolution. Jon Slate (University of Sheffield), for instance, showed not
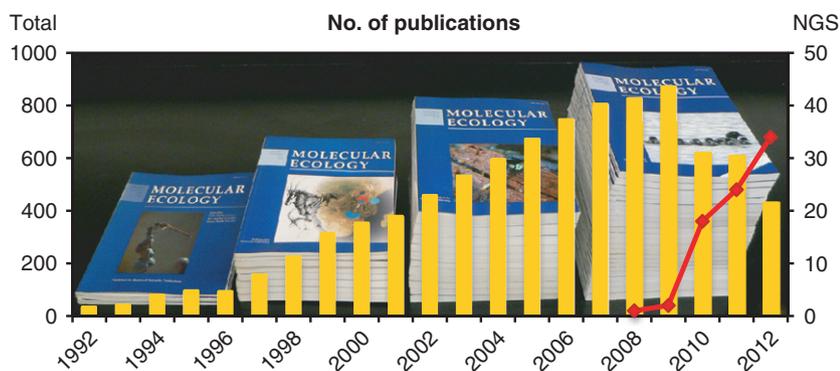
Fig. 1 The total number of publications (yellow bars; left *y*-axis) and the number of studies containing the keyword next-generation sequencing (NGS; red diamonds; right *y*-axis) in *Molecular Ecology* and *Molecular Ecology Resources* (previously *Molecular Ecology Notes*) per year. Numbers for 2012 are until August. The photograph was kindly provided by Louis Bernatchez.

only that NGS allows associating loci to phenotypes, but that we now have the tools to distinguish between patterns of genetic drift and natural selection. With 'gene-dropping' simulations, he and his co-workers recently showed that the frequency and excess of heterozygotes observed in Soay sheep cannot be explained by drift alone and that selection plays a role in shaping coat pattern (Gratten *et al.* 2012). That NGS is a rapid and effective way to identify candidate genes was demonstrated by Aurelie Bonin (LECA, Université Joseph Fourier). They combined NGS approaches (i.e. genome scans, transcriptomics) with admixture mapping and QTL analyses to narrow down the list of candidate genes for insecticide resistance in the yellow fever mosquito *Aedes aegypti*. These genes will now be studied in more detail.

Brent Emerson (IPNA-CSIC, Tenerife) gave a nice overview on the field of ecological interactions, where NGS is used to determine resource diversity. For example, diet analyses can now be performed by sequencing gut contents or faecal matter and, in some cases, even complete focal individuals. Several specialized techniques can be applied, such as sequencing distinct diagnostic fragments (with the aid of e.g. chloroplast- or invertebrate-specific primers, if the consumer is a vertebrate, or the usage of species-specific blocking primers). Furthermore, Graham Stone (University of Edinburgh) showed that besides prey species, also symbionts can reliably be recovered and identified using NGS techniques.

Limitations of NGS

Next-generation sequencing success stories, like the ones presented during the *Molecular Ecology* symposium, have the potential of giving a false impression about the presumed ease of applying the technique and/or analysing NGS data. Fortunately, most speakers also commented on some of the difficulties they encountered when applying NGS. Limitations of NGS techniques are widely accepted and discussed in the literature (see e.g. Ekblom & Galindo 2011; Harrison 2012). Besides the well-known issues such as short read lengths, occasional poor read quality, the sheer amount of data to be managed and analysed and/or the lacking user-friendliness of analytic tools, several other problems were emphasized by the speakers.

One major problem encountered in many NGS studies is the high percentage of nonannotated and/or un-mapped loci, which is especially the case with nonmodel organisms lacking reference genomes and other such resources. Often, a researcher can only speculate about functions of the discovered genes of effect or use indirect evidence from other organisms. Clearly, large-scale functional validation experiments and better comparative tools are needed to recover better annotations for nonmodel organisms.

Another widely discussed topic was brought forward by Alex Buerkle. His simulations on sequencing depth vs. allele frequencies estimates showed that the latter are easily biased, despite adequate coverage. He argued that for many population genetic studies, a coverage of $1\times$ is sufficient, as the individual genotype does not play that much of a role in most analyses. The great advantage of a $1\times$ approach is that many more individuals can be included in a study, reducing overall costs tremendously. It is needless to say that there are many other situations, where one would aim for high coverage. As Bryan Carstens (Louisiana State University) outlined in his talk, higher coverage does give a higher confidence when the data is, for instance, used for *de novo* assembly and SNP calling.

Bryan Carstens further pointed out that NGS is a great tool for phylogeographic studies, but that analyses should be performed more rigorously. He proposed a probabilistic method for model selection and suggested that an averaged approach of parameter estimates in relative proportions to the probabilities per model should be used.

Although the examples so far focussed on functional validation, study design and analytic methods, several speakers called for refined theories. Louis Bernatchez (Université Laval), for example, emphasized the need for an extended theory on the evolutionary causes and consequences of the molecular complexity that links the genotype not only to the phenotype, but also to phenotypic plasticity and nongenetic inheritance. In general, a more holistic approach would be necessary, said Louis.

Future directions of the field

Next-generation sequencing obviously revolutionized the field (although, in our view, the number of publications seems to somewhat lack behind the large number of people

who have submitted samples for NGS or already obtained such data), and there is no reason to believe that method development will decelerate anytime soon. This leads to the question: where will we go from here?

The revolution in genotyping technologies, from isozymes and AFLPs to deep sequencing, is probably the best example for the advances that have been made in our field over the last twenty years, said Loren Rieseberg (University of British Columbia). Analytical methods developed and improved alongside the rise of new experimental techniques. And so did the journal *Molecular Ecology.*

Only four issues (258 pages in total) were published in its first year of existence, 1992, compared to over 5000 pages in the twenty-four issues since 2007 (see also Fig. 1). But we are not there yet, concluded Loren, as the gap between molecular biology and ecology is still substantial. We have learned more about evolution (and lately also ecology) by studying molecules. Now, it is time to increase our efforts of studying ecology as a way to increase our knowledge on the function of molecules. These goals can only be reached if we improve the integration of disciplines and methods, another important conclusion of the symposium put forward by several speakers.

The revolution of sequencing technology will continue, and we are on the doorstep of single-molecule sequencing or 'third-generation sequencing'. This method of sequencing single strands of DNA without prior amplification has great potential for the field in general and for population genetics/genomics in particular. Single-strand sequencing directly produces phased haplotypes allowing more accurate estimations of population genetic parameters and the determination of recombination rates and recombination breakpoints. The technique still needs to improve and to become more efficient, especially with respect to error rates that currently counterbalance the ultra-long reads produced, which would, in theory, simplify (*de novo*) assembly, shorten sequencing times and further reduce costs (see e.g. Schadt *et al.* 2010).

Technological and analytical advances are beginning to change the way in which ecology is studied. With available sequencing methods, it is already possible to genetically characterize ecologically divergent populations in detail (see e.g. Roesti *et al.* 2012). Today's multilocus data sets will soon be replaced by whole genome population samples (see e.g. Jones *et al.* 2012), making it possible to link, at a large scale, alleles throughout the genome to particular phenotypes, geographic patterns and ecological parameters. Importantly, new sequencing techniques will become more and more applicable to nonmodel organisms, which are the main targets of interest in ecology. Thus, in the field of molecular ecology, we can shortly start focussing on the effect of ecology on a single individual ('s genome) instead of studying alleles at a single locus or few loci only; and we can do so in nonmodel organisms and, hence, across a large range of taxa.

Furthermore, technical advances will open the opportunity to study topics that have stayed somewhat untouched or

isolated until now. In the next twenty years, we should, for example, focus on epigenomics and plasticity in relation to phenotype and/or genotype. Another challenge is the development of toolkits to study a variety of organisms. As discussed above, elucidating the link between genotype and phenotype is very difficult, if not impossible without proper genome annotation. The integration of ecological metadata with genomic data sets is another challenge, just like the development of better analytical tools for comparative population genomics, which will further support the development of new subfields like landscape genomics.

In conclusion, the field has challenging but very excited times ahead, and together with the ongoing revolution of technological, analytical and methodological tools, it will stay an exciting field for the next twenty years. We fully agree with Loren Rieseberg that the future of the field is bright indeed!

The videos and slides from the symposium as well as the Online Forum can be found on: www.molecularecologist.com/2012/10/molecular-ecology-online-forum-2012/

References

Aguileta G, Lengelle J, Marthey S *et al.* (2010) Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens. *Molecular Ecology*, **19**, 292–306.

Ekblom R, Galindo J (2011) Application of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Gratten J, Pilkington JG, Brown EA, Clutton-brock TH, Pemberton JM, Slate J (2012) Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Molecular Ecology*, **21**, 2977–2990.

Harrison RJ (2012) Understanding genetic variation and function-the application of next generation sequencing. *Seminars in Cell & Developmental Biology*, **23**, 230–236.

Jones FC, Grabherr M, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB (2012) Rapid De Novo Evolution of X Chromosome Dosage Compensation in *Silene latifolia*, a Plant with Young Sex Chromosomes. *PLoS Biology*, **10**, 4.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. *Human Molecular Genetics*, **19**, R227–R240.

Eveline Diepeveen is a PhD student in the group of Walter Salzburger and uses molecular genetic and genomic approaches to study the genetic basis of naturally and sexually

selected traits in cichlid fishes with special interest in the selective forces acting on genes involved. Walter Salzburger is Professor of Zoology and Evolutionary Biology at the University of Basel. The research of his team focuses on the genetic basis of adaptation, evolutionary innovation and animal diversification. The main model systems in the laboratory are three spine stickleback fish, Antarctic notothenioids and the exceptionally diverse assemblages of cichlid fishes.