MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 3515–3517

NEWS AND VIEWS

PERSPECTIVE

Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research

ANNA PAPADOPOULOU,* PIERRE TABERLET† ‡ and LUCIE ZINGER§

*Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109, USA; †Laboratoire d'Ecologie Alpine (LECA), Centre National de la Recherche Scientifique, Grenoble F-38000, France; ‡Laboratoire d'Ecologie Alpine (LECA), Université Grenoble Alpes, Grenoble F-38000, France; §CNRS, ENFA, UMR 5174 EDB, Université Toulouse 3 Paul Sabatier, Toulouse F-31062, France

It is now well recognized that considering species evolutionary history is crucial for understanding the processes driving community assembly (Cavender-Bares *et al.* 2009). Considerable efforts have been made to integrate phylogenetics and community ecology into a single theoretical framework. Yet, assessing phylogenetic structure at the community scale remains a great challenge, in particular for poorly known organisms. While DNA metabarcoding is increasingly used for assessing taxonomic composition of complex communities from environmental samples, biases and limitations of this technique can preclude the retrieval of information on phylogenetic community structure. In this issue of *Molecular Ecology*, Andújar *et al.* (2015) demonstrate that shotgun sequencing of bulk samples of soil beetles and subsequent reconstruction of mitochondrial genomes can provide a solid phylogenetic framework to estimate species diversity and gain insights into the mechanisms underlying the spatial turnover of soil mesofaunal assemblages. This work highlights the enormous potential of 'metagenome skimming' not only for improving the current standards of DNA-based biodiversity assessment but also for opening up the application of phylogenetic community ecology to hyperdiverse and poorly known biota, which was heretofore inconceivable.

Keywords: genome skimming, metabarcoding, NGS, organellar metagenomics, phylogenomics, soil biodiversity

Received 9 May 2015; accepted 29 May 2015

Preventing biodiversity decline requires efficient techniques to understand community assembly and dynamics across multiple spatial and temporal scales. DNA metabarcoding has been proposed as a promising alternative to classical taxonomic approaches for high-throughput biodiversity assessment and is currently becoming a standard in biodiversity research. Yet, it still suffers from several limitations including PCR biases, which may lead to an overestimation/distortion of species diversity and abundances, a possible lack of taxonomic/phylogenetic resolution, the need to select suitable barcode regions and design versatile primers, as well as the requirement for extensive barcode reference databases. It has been three years since Taberlet *et al.* (2012) predicted that rapid advances in sequencing technology would overcome DNA metabarcoding limitations through shallow shotgun sequencing of specimens or environmental DNA. This approach would take advantage of the high proportion of plastid, mitochondrial or nuclear ribosomal DNA in a (meta)genome and would hence maximize the reliability and representativeness of sample diversity while minimizing sequencing depths and costs. It hence differs from classical (meta)genomics, which rather aim at characterizing full phylogenetic and functional information of the sample, and therefore remains hardly applicable from a high-throughput perspective as they require much deeper sequencing coverage (Fig 1).

These predictions are now fulfilled: genome skimming on specimens allows cost-effective construction of reference libraries of whole organellar genomes and repetitive ribosomal nuclear DNA (Straub *et al.* 2012; Malé *et al.* 2014). Recent studies have further confirmed the possibility to recover and assemble correctly partial or full mitochondrial genomes from bulk arthropod samples (Zhou *et al.* 2013; Gillett *et al.* 2014; Tang *et al.* 2014), allowing to reconstruct well-supported phylogenetic trees, even for deeper evolutionary relationships (Gillett *et al.* 2014; Crampton-Platt *et al.* 2015) and avoiding false positives in species detection (Gómez-Rodríguez *et al.* 2015). Additionally, shotgun sequencing of bulk arthropod samples has revealed a strong correlation between the recovered number of reads and the total biomass of each species (abundance x body length; see Zhou *et al.* 2013; Gómez-Rodríguez *et al.* 2015). By providing increased power for species identification and phylogenetic reconstruction, as well as the potential to assess relative species abundances, 'metagenome skimming' (Linard *et al.* 2015; Fig. 1) hence opens a new era in high-throughput biodiversity assessment.

In this issue of *Molecular Ecology*, Andújar *et al.* (2015) demonstrate the applicability of metagenome skimming for evaluating phylogenetic composition and turnover of poorly known arthropod communities. Their methodology attempts to circumvent the taxonomic impediment and the absence of comprehensive reference databases by combining read-based analyses (i.e. mapping of sequence reads against known reference databases of full mitochondrial

Correspondence: Anna Papadopoulou, Fax: (734) 763 4080; E-mail: apapadop@umich.edu

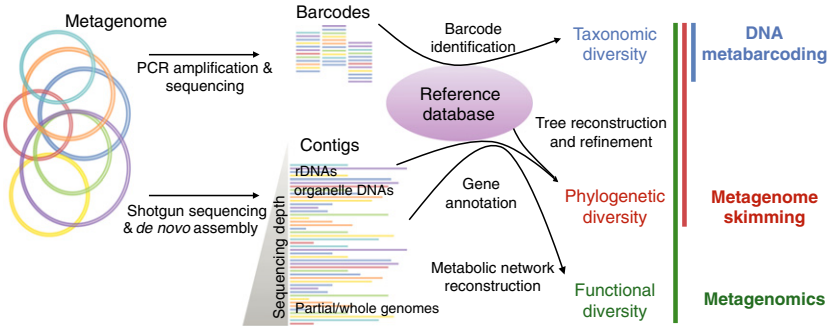

Fig. 1 Overview of NGS-based molecular methods for biodiversity assessment.

genomes or DNA barcodes) with contig-based analyses (i.e. *de novo* assembly of mitogenomes from a mixture of sequence reads). Long mitogenomic contigs serve as scaffold for generating well-resolved backbone phylogenetic trees and improving the phylogenetic placement of shorter contigs and barcode sequences. The resulting tree is first used for phylogeny-based species delimitation and subsequently for estimating total phylogenetic alpha-diversity for each community and phylogenetic beta-diversity among communities. Using this integrative framework, the authors assess species diversity and turnover, vertical stratification and phylogenetic clustering of soil beetle communities in the Southern Iberian Peninsula. Their analyses reveal a strong signature of vertical structuring, with higher phylogenetic beta-diversity in deep soil than superficial horizons and a pattern of strong phylogenetic clustering for exclusive deep-soil taxa. These patterns suggest the existence of specialist clades adapted to deep-soil conditions, which are particularly diverse and possibly more subjected to vicariant speciation than the superficial assemblages due to stronger dispersal limitation acting at regional scales. Thus, geographic isolation, as mediated by layer specialization and niche conservatism, could be a major driver of species diversity patterns and community assembly in soil arthropods.

The study shows the power of metagenome skimming for gaining insights into the processes shaping arthropod assemblages. But it also points out some weaknesses of the currently available methodologies and challenges that remain to be addressed. One consideration is the effect of intraspecific genetic variation in a sample, which may result in failure of contig formation or conversely, in chimeric contigs masking intrapopulation variation. These biases could in turn affect the results of coalescent-based species delimitation analyses, but should not affect greatly estimates of phylogenetic diversity. Also, retrieval of low-abundance species using *de novo* assembly requires much deeper sequencing coverage in comparison with the read-mapping approach, suggesting that constructing the reference set will pose a bottleneck in the metagenomic pipeline. This issue is, however, common to any DNA-based technique, and as the set of available long mitogenomic sequences is enlarging at a feverish pace, read-mapping will be eventually performed at a much more cost-effective way. Another consideration is that even if full organellar genomes and nuclear ribosomal regions provide higher resolution phylogenies, they still represent a limited number of loci for phylogenetic reconstruction and are therefore less powerful as compared to species tree analyses based on multiple unlinked single-copy genes (Maddison & Knowles 2006). Additionally, when working with mixed samples of poorly known or closely related taxa, it can be difficult to associate contigs corresponding to multiple unlinked loci (here organellar and nuclear ribosomal regions) with the genomes from which they originated. More generally, though, species tree methods are computationally intensive and thus inherently difficult to implement in high-throughput biodiversity pipelines.

While further development in bioinformatics and molecular biology is required, metagenome skimming constitutes a powerful high-throughput approach for community ecology. Scaling up from bulk specimens to environmental samples and complex matrices such as soils, waters and sediments, we can already see the possibility to retrieve phylogenetic structure across multiple trophic levels and hence to better clarify the interplay between ecology and evolution on biological interactions, or more globally between biodiversity and ecosystem functioning. The future of metagenome skimming is bright.

References

Andújar C, Arribas P, Ruzicka F *et al.* (2015) Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*, **24**, 3603–3617.

Cavender-Bares J, Kozak KH, Fine PV, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.

Crampton-Platt AL, Timmermans MJTN, Gimmel ML *et al.* (2015) Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. *Molecular Biology and Evolution*. doi:10.1093/molbev/msv111.

Gillett CP, Crampton-Platt A, Timmermans MJTN *et al.* (2014) Bulk *de novo* mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, **31**, 2223–2237.

Gómez-Rodríguez C, Crampton-Platt A, Timmermans MJ *et al.* (2015) Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution*. doi:10.1111/2041-210X.12376.

Linard B, Crampton-Platt A, Timmermans MJTN, Vogler AP (2015) Metagenome skimming of insect specimen pools: potential for

comparative genomics. *Genome Biology and Evolution*. doi:10.1093/gbe/evv086.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Malé PJG, Bardon L, Besnard G *et al*. (2014) Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Molecular Ecology Resources*, **14**, 966–975.

Straub SC, Parks M, Weitemier K *et al*. (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany*, **99**, 349–364.

Taberlet P, Coissac E, Pompanon F *et al*. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Tang M, Tan M, Meng G *et al*. (2014) Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research*, **42**, e166–e166.

Zhou X, Li Y, Liu S *et al*. (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience*, **2**, 4.

A.P., L.Z. and P.T. wrote the manuscript.

---

doi: 10.1111/mec.13263