

'Ecosystemics': ecology by sequencer

Anthony M. Poole, Daniel B. Stouffer and Jason M. Tylianakis

School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

At some point in their career, most students of ecology have endured months to years of laborious and mundane sorting, identifying and documenting of diverse assemblages of species. Despite this accumulated effort, ecology as a field remains far from being able to list every element of a complex ecosystem, let alone understand their functional roles and interactions with other species. The emergence of powerful new tools within the field of metagenomics (the direct sequencing of genomic material from environmental samples [1]) has opened the door to accelerated characterization of microbial ecosystems and detailed analyses of individual microbial species within an ecosystem to a depth never before thought possible. As a consequence, metagenomics is rapidly advancing understanding of microbial ecology; however, the question remains: will it also change the way in which the multicellular constituents of ecosystems and their interactions are studied or will the application of genomic technologies amount to nothing more than genomic stamp-collecting for iconic multicellular species? In our opinion, the answer is clear: metagenomics is on the cusp of transforming ecology.

Conventional methods of sampling ecosystems are unquestionably biased towards the visible. Plants, vertebrates and insects currently receive nearly an order of magnitude greater research effort than do their unicellular cousins*. The lack of any systematic means to identify the hidden microbial component behind whole-ecosystem processes means that not even the most well-studied ecosystems can claim to have generated a complete picture of all key species. Indeed, microbial communities (in particular those outside of the soil) are often treated as a kind of ecological 'dark matter'. To illuminate this part of ecosystems, there is a current dependency on the painstaking efforts of individuals to search for a suspected microbial component. For instance, conventional approaches to quantify nitrogen fixation in boreal forest lichens and mosses failed to locate the leaf-residing cyanobacterial symbionts responsible for the highest nitrogen fixation rate; this symbiosis was only later uncovered through the application of fluorescence microscopy [2].

By contrast, metagenomics would reveal such cryptic relationships between uni- and multicellular organisms *en masse*. Such methods have the potential to identify every organism within a given sample, exposing the diversity of nature that is both visible and invisible to the human eye. Yet, diversity reveals neither what those species are doing nor how they interact. Metagenomic approaches are

already extending far beyond simple taxonomic description. For instance, Iverson and colleagues [3] recently developed methods that permit the reconstruction of the genomes of individual species from metagenomic data. Their method enabled the genomic characterization of an uncultured marine archaeon, which in turn allowed the authors to infer attributes of its functional ecosystem role directly from the genome sequence.

Thus, the capacity to identify and perform detailed characterization of individuals within a metagenome, and even infer attributes of their functional ecosystem role, represents a key advance. Another is metatranscriptomics, which generates a survey of the genes expressed across an ecosystem by sampling RNA, and so enables identification of active metabolic pathways [4]. These approaches establish a link between the physiological traits of organisms and their functional consequences, and may even be the next step forward for *in situ* biodiversity–ecosystem function research [5].

As researchers seek to go beyond function and understand the effects of global environmental changes on ecosystems [6], metagenomics will be essential. It has already helped to unlock the mechanisms for climate–carbon-cycle feedbacks [7] and, for simple microbial ecosystems, has illuminated the probable metabolic basis for key community interactions [8]. These examples underscore two crucial points. First, genomic knowledge is increasing the understanding of how simple organisms interact with their multicellular counterparts in an ecosystem context [9]. Second, the ability to zoom in on the functional roles of species within an ecological community [3] will make metagenomics indispensable for the future study of whole-ecosystem functioning.

Sequencing capacity is doubling every 5 months, massively outstripping Moore's law for computer processing power [10]. Moreover, nanopore sequencing (<http://www.nature.com/news/nanopore-genome-sequencer-makes-its-debut-1.10051>) promises to add USB sequencers and laptop genome assembly to the contents of every field biologist's backpack; and before long, even we might be forced to say 'ecosystemics' with a straight face. More seriously, this erosion of practical barriers to genomic understanding opens numerous doors and requires reassessment of the limits of reductionism in ecology. The looming challenge for ecologists is to establish how far metagenomics can productively be taken. A cell biologist might well argue that there is little to gain from mapping the location of every atom in a cell. Nevertheless, if one dared to take a hectare of rainforest, perform the molecular equivalent of placing it in a blender, and sequenced the resulting nucleic acids, what could be learned? One thing seems certain: the answer would not be 'nothing at all'.

Corresponding author: Poole, A.M. (anthony.poole@canterbury.ac.nz).

* A Web of Science search for the keywords 'ecology and (plant or animal or vertebrate or invertebrate or bird or mammal or fish or amphibian or reptile)' yielded 36 845 hits from the past 20 years, whereas a search of 'ecology and (bacteria or microb* or virus)' yielded only 8896 hits.

Acknowledgments

A.M.P. and J.M.T. are Rutherford Discovery Fellows supported by the Royal Society of New Zealand. We thank P.P. Gardner, W. Godsoe, D.A. Liberles and A. Varsani for helpful comments, and Wirströms, Stockholm for facilitating this work.

References

- 1 Keller, M. and Zengler, K. (2004) Tapping into microbial diversity. *Nat. Rev. Microbiol.* 2, 141–150
- 2 DeLuca, T.H. *et al.* (2002) Quantifying nitrogen-fixation in feather moss carpets of boreal forests. *Nature* 419, 917–920
- 3 Iverson, V. *et al.* (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335, 587–590
- 4 Helbling, D.E. *et al.* (2012) The activity level of a microbial community function can be predicted from its metatranscriptome. *ISME J.* 6, 902–904
- 5 Cardinale, B.J. *et al.* (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* 443, 989–992
- 6 Tylianakis, J.M. *et al.* (2008) Global change and species interactions in terrestrial ecosystems. *Ecol. Lett.* 11, 1351–1363
- 7 Zhou, J. *et al.* (2012) Microbial mediation of carbon-cycle feedbacks to climate warming. *Nat. Clim. Change* 2, 106–110
- 8 Tyson, G.W. *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43
- 9 Woyke, T. *et al.* (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443, 950–955
- 10 Stein, L.D. (2010) The case for cloud computing in genome informatics. *Genome Biol.* 11, 207

0169-5347/\$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
<http://dx.doi.org/10.1016/j.tree.2012.03.008> Trends in Ecology and Evolution, June 2012, Vol. 27, No. 6

Let the four freedoms paradigm apply to ecology

Duccio Rocchini and Markus Neteler

Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, Via E. Mach 1, 38010 S. Michele all'Adige (TN), Italy

In 1985, Richard Stallman, one of the most brilliant minds in computer science, founded the Free Software Foundation and launched the concept of ‘copyleft’, the opposite of copyright. The aim, outlined in the GNU Manifesto (<http://www.gnu.org/gnu/manifesto.html>, [1]), was to make software programs ‘free’ as in ‘freedom’.

The famous ‘four freedoms’ expounded by Stallman [1] are: (i) the freedom to run the program for any purpose; (ii) the freedom to study how the program works and adapt it to one’s own needs; (iii) the freedom to redistribute copies; and (iv) the freedom to make improvements to the program and release them to the public. Thus, the whole (scientific) community benefits from software development. These freedoms are also inherent in several free software licenses, the GNU General Public License (GPL) being one of the most popular.

Approximately a quarter of a century after Stallman put forward his ideas, William K. Michener and Matthew B. Jones, in an article in *TREE* [2] focusing on the analysis of ecological data, stated that: ‘analytical processes are fundamental to most published results in ecology’. Explicit reference to the analytical procedures adopted in generating scientific results is crucial for reproducibility, yet these processes are rarely documented in published ecological papers [2]. Scientific workflow applications, such as Kepler (<https://kepler-project.org>), attempt to address the problem [2], but are only partially successful because the underlying algorithms may still be opaque.

In our view, the explicit use of Free and Open Source Software (FOSS) with availability of the code is essential for completely open science: ‘scientific communication relies on

evidence that cannot be entirely included in publications’, but ‘anything less than the release of source programs is intolerable for results that depend on computation’ [3].

The idea of FOSS and the public availability of the code has been around for almost as long as software [4]. Nonetheless, as far as ecologists are concerned, the open source philosophy is only just taking off, as Stokstad has also pointed out [5].

The increasing availability of open ecological data through networks such as the Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>, [6]) or the Data Observation Network for Earth (DataONE) federated data archive (<http://www.dataone.org>, [7]) makes it increasingly possible to test cutting-edge ecological theories, such as dark diversity [8], evolutionary paths [9] and climate change scenarios [10]. In using a shared open-source code for testing these ecological theories, researchers can be sure that their results are reliable and also that the code they have used is robust [11]. This is particularly true when complex algorithms (or statistical approaches) are involved.

To avoid black box calculations and built-in user interfaces, criticized in [2], researchers have recourse to several examples of FOSS in areas of ecological research, such as ecological statistics (e.g. R Language and Environment for Statistical Computing, <http://www.R-project.org>, [12]) and spatial ecology [e.g. Geographical Resources Analysis Support System (GRASS) GIS, <http://grass.osgeo.org>, [4]). The modular design of such software means decentralized contributions can be made to the source code and allows different institutions and individuals around the world to improve the code base.

If FOSS were more widely employed in ecology and the code used in data analysis provided in scientific papers, more researchers [11] would be able to rely on and replicate

Corresponding author: Rocchini, D. (ducciorocchini@gmail.com), (duccio.rocchini@fmach.it).