News and views

Soil biodiversity and DNA barcodes: opportunities and challenges

CrossMark

Alberto Orgiazzi [a, *], Martha Bonnet Dunbar [a], Panos Panagos [a], Gerard Arjen de Groot [b], Philippe Lemanceau [c]

[a] European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 2749, I-21027 Ispra, VA, Italy
[b] ALTERRA — Wageningen UR, P.O. Box 47, 6700 AA, Wageningen, The Netherlands
[c] INRA, UMR 1347 Agroécologie, 17 rue Sully, BP 86510, 21065, Dijon, Cedex, France

ABSTRACT

Soils encompass a huge diversity of organisms which mostly remains to be characterized due to a number of methodological and logistical issues. Nonetheless, remarkable progress has been made in recent years toward developing strategies to characterize and describe soil biodiversity, especially thanks to the development of molecular approaches relying on direct DNA extraction from the soil matrix.

Metabarcoding can be applied to DNA from any environment or organism, and is gaining increasing prominence in biodiversity studies. This approach is already commonly used to characterize soil microbial communities and its application is now being extended to other soil organisms, i.e. meso- and macro-fauna.

These developments offer unprecedented scientific and operational opportunities in order to better understand soil biodiversity distribution and dynamics, and to propose tools and strategies for biodiversity diagnosis. However, these opportunities also come with challenges that the scientific community must face. Such challenges are related to i) clarification of terminology, (ii) standardisation of methods and further methodological development for additional taxonomic groups, (iii) development of a common database, and (iv) ways to avoid waste of information and data derived from metabarcoding. In order to facilitate common application of metabarcoding in soil biodiversity assessment, we discuss these opportunities and challenges and propose solutions towards a more homogeneous framework.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Soil biodiversity represents a huge underground world containing a wide range of organisms, from archaea, bacteria and fungi to nematodes, insects and earthworms. These organisms interact with each other and affect the functioning of the soil ecosystem (Wagg et al., 2014). The study of soil biodiversity is continuously gaining importance in the environmental sciences due to its significant interlinkages with many other areas, such as agriculture and climate change (Wall et al., 2012). Indeed, soil biodiversity and its functioning deliver many ecosystem services that impact, both directly and indirectly, human wellbeing (Van der Putten et al., 2004; De Vries et al., 2013). However, despite the value of soil biodiversity, this diversity remains to be better explored also in relation to the major threats that it is subjected to; so much so that international initiatives, such as the EU project EcoFINDERS

(Lemanceau, 2011) and the Global Soil Biodiversity Initiative (www.globalsoilbiodiversity.org) have been established. These initiatives call for a better understanding of soil biodiversity and better soil and land management in order to preserve and value soil biodiversity and functioning. In order to reach this goal the development of innovative strategies of characterization shared by the scientific and non-scientific communities is required.

The development of molecular tools for biodiversity characterization based on DNA extraction from the soil matrix — applied so far mostly to microorganisms — or from organisms initially extracted from soils — mainly fauna, but also microorganisms through previous in vitro cultivation — represents unprecedented opportunities (Ogram, 2000). The first Next-Generation Sequencing (NGS) based study on soil biodiversity was published in 2006 (Leininger et al., 2006) and the impact of such technologies on the study of soil biota is now clear, leading to the description of a much larger below-ground diversity than originally expected (Buée et al., 2009). Current NGS platforms yield millions of DNA sequences in a relatively short period of time, and the sequencers' performance improves every year (Glenn, 2011). Application of NGS technologies has resulted in an increasing number of

metabarcoding surveys on soil biodiversity conducted in a wide range of environments (e.g. grasslands, agricultural fields and forests, but also deserts and the Arctic and Antarctic) (Mardis, 2008; Nielsen and Wall, 2013) from small to large scale, including national surveys (Griffiths et al., 2011; Ranjard et al., 2013). The resulting large data sets yield invaluable reference data allowing a more general specification of biodiversity variation in relation to different factors such as soil type, climate and land use.

Applications of these molecular methodologies allow for either characterization of Operational Taxonomic Units (OTUs) by targeting specific fragments of the genome (Roesch et al., 2007; Bates et al., 2011; Orgiazzi et al., 2013) or for extensive sequencing in an untargeted way aiming at in depth screening for functional genes, community structure and phylogenetic diversity (Vogel et al., 2009; Fierer et al., 2012). The former approach is referred to as metabarcoding (Taberlet et al. 2012a,b) and the latter as metagenomics (Simon and Daniel, 2011) (Fig. 1). Metabarcoding is a molecular approach based on the assumption that each OTU can be unequivocally identified through a specific sequence of DNA (barcode). The general strategy consists of (i) extracting DNA from soil or organisms, (ii) amplifying a specific DNA sequence chosen for its taxonomic value, (iii) sequencing the corresponding DNA amplicons, (iv) analysing the sequences using proper pipelines, and finally (v) assessing the taxonomic diversity of the analysed soil or identifying the organism from which DNA has been extracted (Taberlet et al., 2012b). However, this common procedure is rapidly evolving towards new and innovative approaches. An example is the current tendency to move towards methods that bypass PCR amplification of a single DNA fragment by applying shotgun sequencing of e.g. the entire mitochondrial genome (Zhou et al., 2013), thereby introducing genomic approaches into the classic metabarcoding framework.

Although the strategy described above seems easy to apply, its diffusion in the study of soil biodiversity, as well as in other areas, has awakened hidden issues while simultaneously creating new ones. The increased use of this kind of approach reveals the urgent need to establish reference points in the methodology and management of this field of research. However, the optimal use of the unprecedented opportunities described above requires certain preconditions. Metabarcoding-based surveys on soil biodiversity are only now reaching a significant number, thus the time is rife to (i) give an overview of the offered opportunities and (ii) to present and address the challenges that must be faced. Starting from what has been published so far in terms of soil metabarcoding studies (Fig. 1), we first present the opportunities presented by the application of metabarcoding to surveys on soil biodiversity and, secondly, we describe challenges that we consider relevant for the study of soil biodiversity through metabarcoding and propose possible solutions in order to obtain more accurate and comparable results and, consequently, valuable discussions. In particular, we identified two categories of challenges; the first refers to what is needed in order to obtain a reliable assessment of soil biodiversity and what is required for the application of the corresponding research. The second category refers to the strategies to obtain and manage data in such a way that they can be properly processed, compared and eventually used to develop a better management of land and soils.

2. Scientific and operational opportunities

Soil metabarcoding is increasingly applied in scientific studies with a progressing number of published papers following this approach (Fig. 1). This is most certainly due to the relative ease with which this technique can be applied together with the continuous reduction of time and costs involved in using NGS platforms and the development of new bioinformatics pipelines to analyse the data (Schmidt et al., 2013; Yang et al., 2013). Researchers are increasingly able to adapt the application of metabarcoding in order to shed light on several unanswered questions. Any soil biodiversity study can have a dual objective: the first refers to more basic and scientific research, the aim of which is to obtain in depth knowledge of the structure and the functions, i.e. the ecological roles, of soil biodiversity. The second purpose is more operational and directed toward decision makers, and aims to assess the level and possible fluctuations of soil biodiversity in different environmental conditions in order to obtain a diagnosis and establish
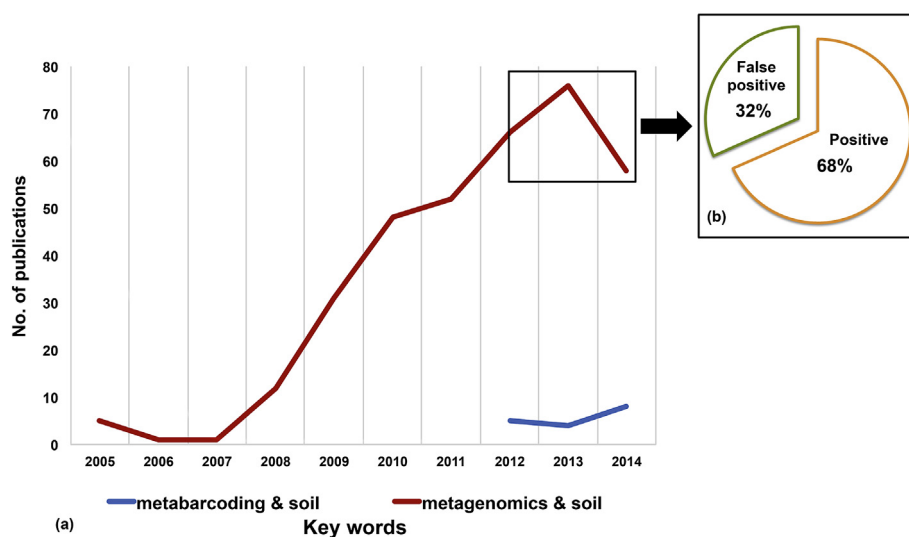

Fig. 1. (a) Number of articles (published and in press) in peer-reviewed journals with the keywords and/or the article title and/or the abstract containing the words "metabarcoding and soil" (blue) and "metagenomics and soil" (red), respectively. The values refer to an online search conducted in the *Scopus Database* (values recorded on September 15th, 2014; search parameters: all standard except). The growing trend of "metagenomics and soil" studies will likely be confirmed at the end of 2014. (b) Distribution of authentic metagenomics-based articles (positive – orange) and metabarcoding-based approach articles (false positive – green). The false positives are calculated as the articles with the keywords "metagenomics and soil", where they were, in fact, based on a metabarcoding strategy. The year 2012 has been chosen as a time threshold as it is the year when the term "metabarcoding" was proposed for the first time.