Biotechnology Advances 34 (2016) 1413–1426

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Metagenomics: Probing pollutant fate in natural and engineered ecosystems

Emna Bouhajja [a], Spiros N. Agathos [a,b], Isabelle F. George [c,*]

[a] Laboratoire de Génie Biologique, Earth and Life Institute, Université Catholique de Louvain, Place Croix du Sud 2, boite L7.05.19, 1348 Louvain-la-Neuve, Belgium
[b] School of Life Sciences and Biotechnology, Yachay Tech University, 100119 San Miguel de Urcuquí, Ecuador
[c] Université Libre de Bruxelles, Laboratoire d'Ecologie des Systèmes Aquatiques, Campus de la Plaine CP 221, Boulevard du Triomphe, 1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 14 March 2016
Received in revised form 1 October 2016
Accepted 12 October 2016
Available online 5 November 2016

Keywords:
Metagenomics
Pollutant biodegradation
Catabolic genes
Degradation pathways
Molecular microbial ecology

ABSTRACT

Polluted environments are a reservoir of microbial species able to degrade or to convert pollutants to harmless compounds. The proper management of microbial resources requires a comprehensive characterization of their genetic pool to assess the fate of contaminants and increase the efficiency of bioremediation processes. Metagenomics offers appropriate tools to describe microbial communities in their whole complexity without lab-based cultivation of individual strains. After a decade of use of metagenomics to study microbiomes, the scientific community has made significant progress in this field. In this review, we survey the main steps of metagenomics applied to environments contaminated with organic compounds or heavy metals. We emphasize technical solutions proposed to overcome encountered obstacles. We then compare two metagenomic approaches, i.e. library-based targeted metagenomics and direct sequencing of metagenomes. In the former, environmental DNA is cloned inside a host, and then clones of interest are selected based on (i) their expression of biodegradative functions or (ii) sequence homology with probes and primers designed from relevant, already known sequences. The highest score for the discovery of novel genes and degradation pathways has been achieved so far by functional screening of large clone libraries. On the other hand, direct sequencing of metagenomes without a cloning step has been more often applied to polluted environments for characterization of the taxonomic and functional composition of microbial communities and their dynamics. In this case, the analysis has focused on 16S rRNA genes and marker genes of biodegradation. Advances in next generation sequencing and in bioinformatic analysis of sequencing data have opened up new opportunities for assessing the potential of biodegradation by microbes, but annotation of collected genes is still hampered by a limited number of available reference sequences in databases. Although metagenomics is still facing technical and computational challenges, our review of the recent literature highlights its value as an aid to efficiently monitor the clean-up of contaminated environments and develop successful strategies to mitigate the impact of pollutants on ecosystems.

Contents

* Corresponding author.
E-mail address: igeorge@ulb.ac.be (I.F. George).

http://dx.doi.org/10.1016/j.biotechadv.2016.10.006
0734-9750/

1. Introduction

Pollutants such as polychlorinated biphenyls (PCBs), aromatic hydrocarbons, pesticides and heavy metals are continuously released in the environment through industrial activities, their associated disposal operations and accidental spills. Many of them are recalcitrant to natural degradation and are subject to accumulation especially in soils and sediments. Their hazardous and toxic effects on different living organisms have been demonstrated in many studies (e.g. (Armstrong et al., 2004; Schwarzenbach et al., 2006; Siegrist and Joss, 2012).

Bioremediation is defined as the use of living organisms (mostly plants or microbes) or enzymes to decontaminate sites polluted by hazardous substances. When a polluted site is hardly accessible, like aquifers, this ecofriendly and sustainable approach might be the only option available. To be successful, physicochemical and microbial characteristics should be taken into consideration before applying bioaugmentation (release of specific strains in the polluted environment), biostimulation (stimulation of indigenous microflora by amendement of the environment with limiting factors) (Fonti et al., 2015) or before relying on natural attenuation (recovery of the environment without any intervention) (Adetutu et al., 2015; Bento et al., 2005; Meckenstock et al., 2015). The indigenous microbial communities in polluted environments have high tolerance towards pollutants. They use pollutants as carbon or nitrogen source and/or energy source and degrade them into simpler intermediates up to complete mineralization, i.e. release of carbon dioxide, methane, ammonia and similar final products. In other cases indigenous bacteria cannot use pollutants as growth substrate but they can degrade them (but not mineralize them) by cometabolism. In this case, bacteria use another substrate to sustain their growth (for review see Nzila (2013)). Cometabolism of pollutants was mainly reported for PAHs with more than five aromatic rings, chlorinated biphenyls, chlorinated monoaromatics and chlorinated aliphatics (Nzila, 2013).

Bacterial degradation pathways of many pollutants which belong to the families of PCBs and aromatic hydrocarbons have been described in studies based on pure strains degrading single pollutants (Borja et al., 2005; Fischer et al., 2010; Haritash and Kaushik, 2009; Olsen et al., 1994). However, in contaminated environments such as old sites of petroleum refineries, gas stations and rivers near industrial manufacturing plants, pollution consists usually of a mixture of several compounds. Complete remediation might therefore require synergistic interactions between different degrading bacteria (Chang et al., 2006; Dastgheib et al., 2011; Ren et al., 2014; Tzintzun-Camacho et al., 2012), bacteria and fungi (Jacques et al., 2008) or bacteria and plants (Saiyood et al., 2010). In many reports, authors describe the isolation of biodegradative consortia rather than single species especially when the pollutant(s) belong to hardly degradable compounds such as heavy PAHs, highly chlorinated biphenyls and pesticides (Boonchan et al., 2000; Dastgheib et al., 2011; Ellegaard-Jensen et al., 2014; Pino and Peñuela, 2011; Rowe et al., 2008). Moreover, the majority of environmental bacteria are recalcitrant to cultivation in the laboratory and the mystery of their biology is far from solved (Stewart, 2012). Hence, metagenomics is currently one of the most advanced methods to discover and describe inaccessible environmental microbes in their whole complexity, and to provide a comprehensive overview of the biodegradation potential of microbial communities in polluted environments (Bell et al., 2014; George et al., 2011; Gillan et al., 2015).

Metagenomics is defined as the study of environmental microbial communities using a suite of genomic tools to directly access their genetic content (Thomas et al., 2013), i.e. without prior cultivation of microbes in the laboratory. In metagenomics, the first step consists in extracting total DNA from the environment, which is often challenging in polluted soils and sediments. Indeed, high concentration of pollutants (e.g., metals, aromatic hydrocarbons) and low cell density are the main factors that hamper successful DNA recovery. Metagenomic analysis of recovered DNA is based on (i) genetic and/or functional screening of cloned DNA (referred to as a metagenomic library) or on (ii) large-scale sequencing of environmental metagenomes without pre-cloning (referred to as “direct sequencing” in this paper, or “shotgun metagenomics” by other authors). The term “metagenomics” is also liberally applied in the literature to a third approach: the thorough analysis of the diversity of specific genes – primarily marker genes such as the 16S rRNA gene (Wang and Qian, 2009) or conserved single-copy genes (Lang et al., 2013) – through PCR amplification and high throughput sequencing of the amplicons. In all cases, metagenomic DNA is then sequenced by one of the currently available high throughput sequencing platforms (e.g., 454/Roche, Illumina, SOLiD). Lastly, the huge amount of data (scale of Gbp) generated by sequencing projects is analyzed using a panoply of bioinformatic tools to predict the microbial diversity and/or the functional potential of the studied environment.

This review paper discusses the latest methodological advances in metagenomics applied to contaminated environments, and highlights how this approach has allowed the recent discovery of new biodegradation genes/pathways and microbial adaptation and synergistic interactions for pollutant degradation.

2. Conducting metagenomic studies in polluted environments: an updated overview of existing solutions to overcome inherent difficulties

Fig. 1 illustrates the main steps of a metagenomic project design for the comprehensive study of pollutant degradation in the environment.

2.1. Sampling polluted, heterogeneous environments

Before starting a metagenomic study to address ecological questions, it is important to select the proper site, the method, the time, the size, and the number of samplings (biological replicates). In particular, biological replicates are a critical issue in metagenomics. They are necessary to accurately estimate spatial and temporal variability of microbial communities in heterogeneous environments like soils and sediments (e.g. estimation of β-diversity) by statistical analysis of data. Indeed, in the latter, physicochemical properties (texture, structure, pH) can vary locally, which affects microbial community structure (Stres et al., 2013; Van Horn et al., 2013). In