Geoderma Regional 14 (2018) e00180

Contents lists available at ScienceDirect

Geoderma Regional

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

DNA sequencing and metagenomics of cultivated and uncultivated chernozems in Russia

Maria A. Gorbacheva [a], Nataliya V. Melnikova [b], Vladimir R. Chechetkin [a], Yuri V. Kravatsky [a], Nickolai A. Tchurikov [a,*]

[a] Department of Epigenetic Mechanisms of Gene Expression Regulation, Engelhardt Institute of Molecular Biology, Moscow, Russia
[b] Group of Postgenomic Studies, Engelhardt Institute of Molecular Biology, Moscow, Russia

ARTICLE INFO

Article history:
Received 7 March 2018
Received in revised form 30 May 2018
Accepted 8 June 2018
Available online xxxx

Keywords:
Chernozems
Virgin soil
Cropland
Kursk chernozem
Microbiomes
Metagenome
Deep sequencing
Microorganisms
Fungi

ABSTRACT

In this study, we analyzed the microbiomes of uncultivated (virgin) and cultivated (cropland) soils from the southern part of Russia using DNA shotgun deep sequencing and metagenomic studies. The processed reads were mapped to the nucleotide sequences of bacteria, fungi, archaea, and viruses extracted from the non redundant database of NCBI. Our data indicate that Archaea and Fungi are more abundant in the virgin soil. The analysis of the top ten most abundant taxa from phylum to species levels revealed only slightly statistically significant differences between the microbial communities of the soils. In contrast, among the less abundant organisms at all taxa levels, we detected considerable differences between the virgin soil and the cropland microbiomes. The higher biodiversity of the virgin soil was confirmed by functional annotation and abundance analysis using the eggNOG and KEGG databases. Collectively, our data strongly indicate that the biodiversity of virgin soil is much higher, suggesting that intensive farming reduces the microbial biodiversity and causes specific changes in the pattern of less abundant microbiome taxa.

1. Introduction

Biological changes in soil are caused by many factors, including the mineral and organic initial materials, climate, landscape, and farming. Microorganisms are critical components of the soil because they are involved in mineral and nutrient cycling (Albertsen et al., 2012), as well as chemical reactions in different types of soils (Amann et al., 1995; Baldrian et al., 2011), including bioremediation (degradation of hazardous organic pollutants to safe levels in soils) (Singh et al., 2014). Microbial communities in soil represent the alive and active part of the soil that is very sensitive to natural and anthropogenic influences. The functional and taxonomic diversity of soil microbial communities is highly susceptible to environmental factors such as soil characteristics (pH, nutrient availability), climate changes, and biotic interactions (Bartram et al., 2011; Bergmann et al., 2011; Biddle et al., 2008; Wagg et al., 2014), indicating that the soil microbial communities could serve as indicators of ecosystem functioning.

Nowadays, deep sequencing methods and subsequent metagenomic analysis are used for the comprehensive studies of microbiomes in different groups of soils (Pennanen et al., 1999; Hattenschwiler et al., 2011; Simon et al., 2009; Wallander et al., 2010; Kanokratana et al., 2011; Rousk et al., 2010; Mackelprang et al., 2011; Courty et al., 2010; Delmont et al., 2012; Alzubaidy et al., 2016), their dynamic associations with plants (Zarraonaindia et al., 2015), the connections between crop yield and microbial communities (Xu et al., 2015), the role of microbiomes in soil ecology, and even the social behavior of bacteria, which includes cell-to-cell signaling via quorum sensing and other mechanisms (Mandic-Mulec et al., 2015). The full genome sequencing of many microorganisms provides a powerful tool for metagenomic studies. Analysis of microbiomes became possible without the cultivation of microorganisms because microbial DNA can be recovered directly from the soil samples and over the past few years, many new genomes have been sequenced for metagenomic studies.

Russian chernozem soils are an interesting example of fertile soils, which are widely recognized as the etalon soils. Chernozem occupies about 6% of the country and is mostly used for intensive farming agriculture, which provides up to 60% of all agricultural products in Russia (Kovda, 1983; Avetov et al., 2011). To date, the study of microbiomes of the Russian chernozem and its seasonal dynamics has been limited to 16S rRNA sequence analysis (Chernov et al., 2015; Ivanova et al., 2017). Russian chernozem contains two predominant phyla, *Proteobacteria* and *Actinobacteria*, which comprise about 48% and 28% of total microbiome, respectively. These studies provide data regarding the microbial communities on the basis of 16S rRNA gene sequencing. However, this approach cannot examine the functional characteristics

* Corresponding author.
E-mail address: tchurikov@eimb.ru (N.A. Tchurikov).

https://doi.org/10.1016/j.geodrs.2018.e00180
2352-0094/

of microorganisms with high activity, since the homology of sequences in the 16S rRNA gene in different organisms is very high and thus, although it can confidently identify the organisms of higher taxonomic levels (family or above), it cannot identify the species (Will et al., 2010).

The sequencing and metagenomic study of total DNA preparations isolated from soil provides much greater potential for the analysis of microbiomes. Therefore, we used this approach to compare the microbiomes of cropland and virgin soils from the southern part of Russia where the chernozem soils are located.

2. Materials and methods

2.1. Sample collection

Soil samples were collected in the Kursk region of Russia in March 2013 (Fig. S1). Two virgin soil samples were collected at Alekhin Biosphere Reserve (http://zapoved-kursk.ru/) (51°31′02.8″ N, 36°16′59.7″ E). Two virgin soil-surface samples (depth to 10 cm) located 45–50 m apart were collected. The soil has been classified as Typical Chernozem (CH) virgin land and is under a wild vegetation grassland steppe. Typical plants for the virgin soil include wild cereals, *Fitillaria ruthenica*, *Fitillaria meleagris*, *Paeonia tenuifolia*, *Iris aphylla*, and a number of *Stipa sp*.

Two cropland samples were collected from a field of the Scientific Research Institute of Agricultural Production, which is also located in the Kursk region (51°36′58.1″N, 36°16′51.4″ E), about 12 km northeast of the site where the virgin soil samples were collected. The field in the previous year (2012) was used to grow potatoes and then later in the same year to grow legumes (holy clover and Lucerne). Earlier (2011), the field was under the winter crop. In each case, two soil-surface samples (depth to 10 cm) were collected and these replicates were located at a distance of 45–50 m apart. No specific permissions were required for each location activities because the field studies did not involve endangered or protected species. The samples were stored in sterile bags on dry ice at the time of sampling, and then were stored at −70 °C until DNA extraction.

The Kursk region is located in the southwestern part of the middle-Russian plateau. The annual average air temperature for the region is +4.6–6.1 °C and the annual precipitation is 475–640 mm. Both types of soil are characterized by neutral pH (6.8 and 6.2 for the virgin soil and the cropland, respectively) and by high organic carbon content (3.96% and 2.1% for the virgin soil and the cropland, respectively). The colloidal fraction of both soil types is saturated mainly with 17.6% CaO and 3.2% MgO. The A horizon (top soil, 0–12 cm) for both soils is very loose, has a fine-powdery structure, and is dry and dark gray.

2.2. DNA extraction

Total DNA samples were extracted from all soil samples using a Power Soil DNA Isolation Kit (MO BIO, USA). A 0.25-g sample of soil was used from each sample. The final DNA preparations were dissolved in 50–100 μL of sterilized deionized water and stored at -70 °C before use. The DNA yields were 7–17 μg per sample, as tested by a Nanodrop 2000.

2.3. Preparation of shotgun DNA libraries for deep sequencing

DNA of each sample was sonicated to obtain 300–500-bp fragments. About 1 μg of each DNA was used for the library preparation. Shotgun DNA libraries were prepared according to Illumina's instructions that accompany the DNA Sample Kit or the NEBNext Ultra DNA Library Prep Kit for Illumina. Briefly, DNA was end-repaired using a combination of T4 DNA polymerase, *E. coli* DNA Pol I large fragment (Klenow polymerase), and T4 polynucleotide kinase. The blunt, phosphorylated ends were treated with Klenow fragment and dATP to yield a protruding 3′ A-base for ligation of Illumina adapters, which have a single T-base overhang at the 3′ end. After adapter ligation, DNA was PCR amplified with Illumina primers for 15 cycles. Deep sequencing (two biological replicates for each sample) was performed using an Illumina Genome Analyzer IIx. The raw and processed data were deposited in NCBI (accession number SUB2560707).

2.4. Metagenomic analysis

Bioinformatic analysis was carried out commercially by Novogen. After the data quality control step, the clean data were used for taxonomic annotation (the workflow is illustrated in Fig. S2). The SOAPdenovo (Luo et al., 2012) assembly package was used to perform metagenomic assembling with different K values (default 49, 55, and 59). The assembly result of Scaffold that had the largest N50 was then selected for the subsequent analysis. For species annotation, the CD-HIT was used to cluster Scaftigs derived from the assembly with a default identity of 0.95. In order to analyze the relative abundance of Scaftigs further in each sample, the clean reads after pre-processing were mapped with the non-redundant Scaftigs dataset by SoapAligner (Liu et al., 2011). Then the Scaftigs of total depth equal to 0 were filtered, and the abundance table of filtered Scaftigs was obtained. The corresponding Scaftigs were mapped to the mass of Bacteria, Fungi, Archaea, and Viruses data extracted from the NR database of NCBI. The LCA algorithm (lowest common ancestor, applied in the MEGAN (Huson et al., 2011) software system) was used to ensure the annotation significance by picking out the lowest common classified ancestor for final display. The number of reads that were mapped to a particular taxa level and used for analysis was >20.

Functional annotation and abundance analysis were performed using the KEGG and eggNOG databases and BLAST, and the results with a minimum value were selected. The BCR (BLAST Coverage Ratio of Gene) of the reference and query genes was selected with cutoff ≥ 40%. The index of BCR was: BCR (Ref.) = (Match/Length(R)) × 100%; BCR (Que.) = (Match/Length (Q)) × 100%, where Match is the available alignment length between the reference and query genes. Length(R) is the length of the reference gene. Length (Q) is the length of the query gene. For each result of the annotation, functional profiling was performed based on KEGG (Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al., 2006); Version: 58), eggNOG (Evolutionary genealogy of genes: Non-supervised Orthologous Groups (Powell et al., 2014); Version: 4.0).

2.5. Statistical analyses

In our experiments, a crosscheck was performed by repeating deep sequencing twice for each of the cropland and virgin soil samples. The mean and dispersion for the fractions of reads corresponding to particular microorganisms in the patterns of each type were calculated as previously described (Johnson and Leone, 1977; Weir, 1996):

$$\overline{f}_{species} = (n_1 f_{s1} + n_2 f_{s2})/(n_1 + n_2) \quad (1)$$

$$\sigma^2\left(\overline{f}_{species}\right) = \frac{n_1 f_{s1}(1-f_{s1})}{(n_1+n_2)^2} + \frac{n_2 f_{s2}(1-f_{s2})}{(n_1+n_2)^2} \quad (2)$$

where n_1 and n_2 are the complete numbers of filtered-out reads (Table S1) in two repeating deep sequencing rounds, while f_{s1} and f_{s2} refer to the fractions of reads corresponding to particular taxa defined in each round. The relevant parameters were calculated separately for the cropland and virgin soil patterns. The statistical cutoff for the small fractions can be determined by the standard 5% threshold of statistical significance:

$$\overline{f}_{species} \geq 1.96\,\sigma\left(\overline{f}_{species}\right) \quad (3)$$

The significance of the difference between the mean fractions corresponding to the counterpart microorganisms for the cropland