

*Exploring the dynamics of bacterial  
community composition in soil: the pan-  
bacteriome approach*

**Giovanni Bacci, Maria Teresa  
Ceccherini, Alessia Bani, Marco  
Bazzicalupo, Maurizio Castaldini, Marco  
Galardini, Luciana Giovannetti, et al.**

**Antonie van Leeuwenhoek**  
Journal of Microbiology

ISSN 0003-6072  
Volume 107  
Number 3

Antonie van Leeuwenhoek (2015)  
107:785-797  
DOI 10.1007/s10482-014-0372-4



**Your article is protected by copyright and all rights are held exclusively by Springer International Publishing Switzerland. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Exploring the dynamics of bacterial community composition in soil: the pan-bacteriome approach

Giovanni Bacci · Maria Teresa Ceccherini · Alessia Bani · Marco Bazzicalupo · Maurizio Castaldini · Marco Galardini · Luciana Giovannetti · Stefano Mocali · Roberta Pastorelli · Ottorino Luca Pantani · Paola Arfaioli · Giacomo Pietramellara · Carlo Viti · Paolo Nannipieri · Alessio Mengoni

Received: 1 August 2014 / Accepted: 26 December 2014 / Published online: 7 January 2015  
© Springer International Publishing Switzerland 2015

**Abstract** We performed a longitudinal study (repeated observations of the same sample over time) to investigate both the composition and structure of temporal changes of bacterial community composition in soil mesocosms, subjected to three different treatments (water and 5 or 25 mg kg<sup>-1</sup> of dried soil Cd<sup>2+</sup>). By analogy with the pan genome concept, we identified a core bacteriome and an accessory bacteriome. Resident taxa were assigned to the core bacteriome,

while occasional taxa were assigned to the accessory bacteriome. Core and accessory bacteriome represented roughly 35 and 50 % of the taxa detected, respectively, and were characterized by different taxonomic signatures from phylum to genus level while 15 % of the taxa were found to be unique to a particular sample. In particular, the core bacteriome was characterized by higher abundance of members of *Planctomycetes*, *Actinobacteria*, *Verrucomicrobia* and *Acidobacteria*, while the accessory bacteriome included more members of *Firmicutes*, *Chlamydiae* and *Proteobacteria*, suggesting potentially different responses to environmental changes of members from these phyla. We conclude that the pan-bacteriome model may be a useful approach to gain insight for

Giovanni Bacci and Maria Teresa Ceccherini have equally contributed to the work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10482-014-0372-4) contains supplementary material, which is available to authorized users.

G. Bacci · A. Bani · M. Bazzicalupo · A. Mengoni (✉)  
Department of Biology, University of Florence, Via  
Madonna del Piano 6, 50019 Sesto Fiorentino, Firenze,  
Italy  
e-mail: alessio.mengoni@unifi.it

G. Bacci  
Consiglio per la Ricerca e la Sperimentazione in  
Agricoltura, Centro di Ricerca per lo Studio delle  
Relazioni tra Pianta e Suolo (CRA-RPS), Via della  
Navicella 2/4, 00184 Rome, Italy

M. T. Ceccherini · L. Giovannetti · O. L. Pantani ·  
P. Arfaioli · G. Pietramellara · C. Viti · P. Nannipieri  
Department of Agrifood Production and Environmental  
Science, University of Florence, P.le delle Cascine 28,  
50144 Florence, Firenze, Italy

M. Castaldini · S. Mocali · R. Pastorelli  
Consiglio per la Ricerca e la Sperimentazione in  
Agricoltura, Centro di Ricerca per l'Agrobiologia e la  
Pedologia (CRA-ABP), Piazza d'Azeglio 30,  
50121 Florence, Firenze, Italy

M. Galardini  
EMBL-EBI—European Bioinformatics Institute  
Wellcome Trust Genome Campus Hinxton,  
Cambridge CB10 1SD, UK

modeling bacterial community structure and inferring different abilities of bacteria taxa.

**Keywords** Core and accessory bacteriome · Bacterial communities · 16S rRNA metabarcoding · Soil · Cd

## Introduction

Understanding changes in bacterial community structure over time is still one of the major challenges in microbial ecology (Donn et al. 2014; Ge et al. 2008; Zhou et al. 2014). Indeed, environmental changes often affect taxonomic composition and abundance in microbial communities (Allison and Martiny 2008), which may have a strong effect on soil health and plant productivity (Chaparro et al. 2012). Several works have been performed inspecting bacterial community variation both in cross sectional (different sites at the same time) and longitudinal studies (the same site studied over time) (Bartram et al. 2014; Chen et al. 2013; Costello et al. 2009; Kuang et al. 2012; Logares et al. 2013; Pini et al. 2012; Smith et al. 2012).

A pan-genome is defined as the collection of all the genes of a set of bacteria which can be divided into core genome (the group of genes shared by all the selected bacteria), accessory genome (the group of genes present in some, but not all, of the selected bacteria) and unique genome (genes belonging to only one particular strain) (Tettelin et al. 2008). In analogy with this concept we can define a pan-bacteriome as the collection of all bacterial taxa present in a set of environmental samples. As recently reported (Haridom and Costa 2014), the pan-bacteriome can be divided into the core bacteriome (the pool of taxa shared by all the samples), the accessory bacteriome (the group of taxa found in some, but not all the samples) and the unique bacteriome (taxa detected only in a particular sample).

Several recent works (Aravindrajana et al. 2013; Bartram et al. 2014; Bowen et al. 2012; Dohrmann et al. 2013; Gibbons et al. 2013; Kim et al. 2013; Logares et al. 2013; Oh et al. 2013; Portillo et al. 2013; Sánchez et al. 2013; Székely and Langenheder 2014; Wegner et al. 2013), mainly performed with next-generation sequencing technologies, have demonstrated that most bacterial communities are composed

of a few abundant taxa and a plethora of rare ones (Pedrós-Alió 2012). Such different fractions may have in theory different taxonomic signatures (at the various taxonomic levels), in relation to the physiological features of the groups of taxa present. For these reasons, we define the “pan-bacteriome” concept, as the whole set of taxa present in a bacterial community analyzed through multiple samplings through time. The pan-bacteriome will then include both a core and an accessory fraction, which describe the taxa shared by all samples and the taxa present only in some samples, respectively. In addition, the so called “rare biosphere”, composed by taxa occurring only in a fraction of the community samples, and interpreted as the accessory fraction of the bacterial community, seems more readily respond to environmental changes (Campbell et al. 2011; Dohrmann et al. 2013; Gobet et al. 2012; Kim et al. 2013; Székely and Langenheder 2014).

Contaminated soils have been often used as models for inspecting bacterial evolution over time (Mengoni et al. 2010; Porter and Rice 2013), since different bacterial taxa show different responses to contamination. The most oxidized state (and most frequent) of cadmium ( $\text{Cd}^{2+}$ ) is known to be toxic for soil microbial biomass and activity (Renella et al. 2002, 2005; Xu et al. 2013) and several studies have shown that  $\text{Cd}^{2+}$  concentrations affect soil microbial diversity under short-term cadmium stress (Chien et al. 2008; Duan and Huang 2008; Fritze et al. 2000; Gomes et al. 2010; Lazzaro et al. 2008; Lorenz et al. 2006; Sheoran et al. 2008; Zhang et al. 2009). The aim of this work was to apply the pan-bacteriome model, as sum of a core and an accessory assemblage of bacteria taxa present in bacterial communities, to evaluate the dynamics of bacterial community composition in three soil mesocosms exposed to slightly different environmental conditions (concentrations of 0, 5 and 25  $\text{mg kg}^{-1}$  of dried soil  $\text{Cd}^{2+}$ ) through time. A metabarcoding approach was applied on 16S rRNA gene based on Illumina sequencing technology.

## Materials and methods

### Experimental setup

The top layer (0–15 cm) of a soil located near Romola (Florence, Italy 43.696240N, 11.153894E) was sampled

(about 20 kg), air-dried overnight at room temperature and sieved at 2 mm. The sieved air-dried soil mass was repeatedly split to ensure the representativeness of the initial material, which was finally divided into masses of about 1 kg each. One mass was analyzed for relative humidity (RH 7.30 %), particle size (sand 81.9 %; silt 6.7 %; clay 11.4 %), water holding capacity (WHC 14.85 g 100 g<sup>-1</sup> of dried soil), pH (5.4), total organic C (0.7 %), and organic N (0.07 %) after the air-drying process and just before the constitution of the mesocosms.

Three of the remaining masses were used to constitute the mesocosms. Each mesocosm was formed by a flat-bottomed plexiglas cylinder (10 cm height, 15 cm of diameter) previously disinfected with 70 % ethanol, which contained the soil mass. Its base was perforated to permit aeration and water flow. The cylinder was supported by a similar one not perforated, intended to collect any possible leak of soil or solution. The upper part of the mesocosm was weighted, to keep track of the water losses which were reconstituted weekly throughout the experiment. Then an excess 350 ml of sterile distilled water was thoroughly and slowly dripped on the dried soil: the mesocosms were then brought to a thermostatic clean room at 22 °C into which they were maintained and their position rotated every third day for 2 months. When the mesocosms reached about the 25 % of the WHC, they were randomly assigned to one of the following treatment: (A) sterile distilled water, (B) 3CdSO<sub>4</sub>·8H<sub>2</sub>O (Sigma-Aldrich) sterile solution as to reach a final concentration of 5 mg kg<sup>-1</sup> of dried soil, (C) 3CdSO<sub>4</sub>·8H<sub>2</sub>O sterile solution as to reach a final concentration of 25 mg kg<sup>-1</sup> of dried soil (Figure S1). The volume and the concentration of Cd solutions were adjusted to reach both the 50 % of the soil WHC and the Cd concentration in mg kg<sup>-1</sup> of dried soil. The mesocosms were sampled at 0, 1, 4, 8, 36 and 60 days (t<sub>0</sub>, t<sub>1</sub>, t<sub>4</sub>, t<sub>8</sub>, t<sub>36</sub>, t<sub>60</sub>) after the initial Cd solution spiking. The first sampling time (t<sub>0</sub>) took place as soon as the soil was visibly saturated with the Cd solution. On each date, a brass pipe was used to withdraw three replicates. DNA extraction was performed on each of the three soil cores (1.2 cm diameter, 6 cm deep). The location of the cores on the mesocosm surface was selected by randomly generating 1,000 groups of 6 × 3 polar coordinates (i.e. an angle and a distance for each date and replica): if the area of the cores in two or more locations/dates overlapped or were

tangent, the next group (among the 1,000 generated) with more distant locations was selected. An R script was used to generate and print 1,000 cardboard notched disks which were used to unambiguously locate the position of the cores, as previously described (Ceccherini et al. 2007). A total of 54 samples (3 mesocosms, 6 sampling dates, 3 soil cores for DNA extraction (replicates) were collected (Figure S1). All steps were performed under sterile conditions.

#### DNA extraction and 16S rRNA metabarcoding

Each soil core was well mixed before DNA extraction. DNA was extracted from 0.5 g of the cores using the bead-beating method as previously described (Ascher et al. 2009). The extracted DNA was checked by electrophoresis on 1 % (w/vol) agarose gel and quantified by Picodrop spectrophotometer (Picodrop Limited, UK).

For metabarcoding, each extracted DNA sample was amplified using primers targeting the variable V6 region of 16S rDNA (V6-967F 5'-CAACGCGAAGA ACCTTACC-3' and V6-1046R 5'-CGACAGCCATG CACACCT-3' (Huse et al. 2008). PCR conditions were those previously described by Sogin and co-workers (Sogin et al. 2006). Ten independent PCR reactions for each of the 54 samples were done. Products were resolved by agarose gel (1.5 % w/vol) electrophoresis and bands were purified with MinElute Gel Extraction Kit (Qiagen, Inc.). Quality and quantity of products were assessed spectrophotometrically (Biophotometer, Eppendorf). The amplicons from each sample were pooled together and a total of 54 samples were sequenced. Massive parallel sequencing was performed by Illumina-Solexa technology (Bartram et al. 2011; Gloor et al. 2010) with the pair-end protocol on an Illumina HiSeq 2000 machine by Beijing Genome Institute sequencing service ([www.genomics.cn](http://www.genomics.cn)). Sequences are deposited in SRA database under the BioProject accession number SRP038532.

#### Bioinformatic processing of 16S rRNA metabarcoding data: data pretreatment

Sequences were analyzed in order to identify and remove low quality regions. Since raw sequences had a length of 100 bp, the quality control step had to be very conservative, to prevent the removal of

taxonomically significant data. As a consequence, a dynamic trimming algorithm was used (Bacci et al. 2014) setting the cutoff parameter as the mean quality of the whole files minus the standard deviation value (in this case an average quality cutoff of 33 Phred was used for all files (Ewing and Green 1998; Ewing et al. 1998). Collected sequences were subjected to a further quality control step and assembled using PANDAseq (Masella et al. 2012). Finally, a set of 18,778,601 sequences (mean length 112 bp and overlapping for more than 70 bp, data not shown) was achieved and used for successive analyses. In order to assign each read to specific taxa we used the RDP multiclassifier trained on the default RDP dataset (16S training set 9). An assignment cutoff of 0.5 was used as reported in the RDP Classifier pipeline for sequences shorter than 250 bp (Wang et al. 2007).

#### Data treatment and statistical analysis

RDP assignments at all taxonomic levels (from phylum to genus) were collected. The average number of assignments in all samples was evaluated. RDP assignments were collected for each sample to generate a community data matrix ( $X$ ) at genus level (rows = samples; columns = detected genera) (Table S1).  $X$  was used as the input/source for all statistical analyses. In order to detect genera belonging to the core and to the accessory bacteriome,  $X$  was transformed into a Boolean (presence/absence) matrix ( $X_{p/a}$ ). Abundance values greater than or equal to 1 were rounded to 1 while abundance values equal to 0 were left unchanged. Each genus was assigned to one of the core or to the accessory bacteriome fractions using these criteria: if a genus was detected in all samples (its abundance value was greater than 0 in the 54 samples) it was assigned to the core bacteriome; otherwise it was assigned to the accessory bacteriome. In addition, if a genus was detected only in a single sample, it was considered as “unique assignment”.

To assess taxa richness, a rarefaction analysis was performed using the R package Vegan (Dixon 2003; Oksanen et al. 2013), based on the genus assignments in  $X$ .

The variation in the composition of bacterial communities with respect to conditions and time was analyzed by canonical correlation analysis (CCA) using the R package Vegan on  $X$ . The different conditions and the sampling times have been fitted

onto previously developed ordination analyses using the *envfit* function of the R package Vegan with 10,000 permutations.

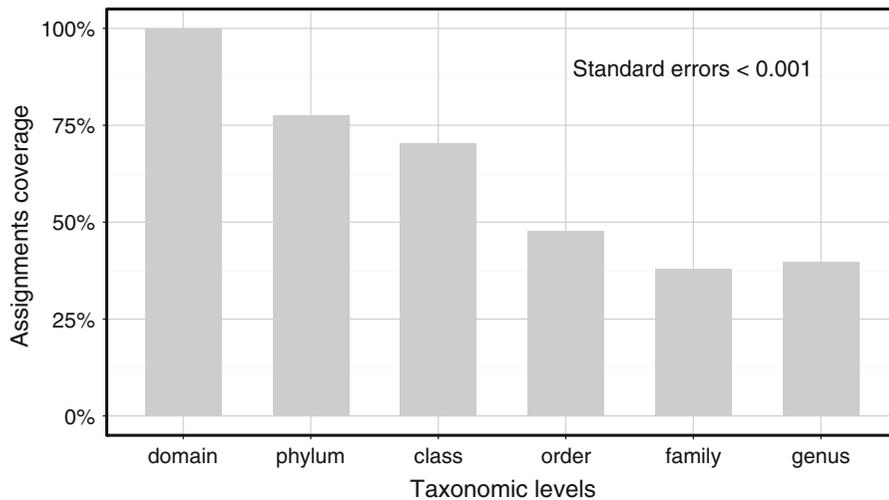
## Results

### Description of sequence data and coverage

A total of 18,778,601 reads were processed for the 54 sequence files representing each of the samples (Figure S1). The reads in the sequence files ranged from a minimum of 339,328 to a maximum of 356,164. A preliminary analysis showed that the percentage of reads assigned to each taxonomic level was similar (standard error <0.001) among all files and that a mean of 39.6 % of total reads were assigned to the genus level (Fig. 1), with a total of 901 genera (with a similarity cutoff of 95 %). To describe the richness of the samples at the genus level, a rarefaction analysis was performed (Fig. 2). Since the curves approached asymptotes, the richness of each sample was similar regardless of simulated environmental conditions or sampling time. Finally, an analysis of similarities (ANOSIM) was performed using all genus assignments in matrix  $X$  (Table S1) as response variables. Both conditions and sampling times were used for grouping together samples from the same replicate (see Figure S1). As a result, the ANOSIM analysis showed that samples belonging to the same triplicate differed significantly between each group ( $p$  values <0.05 obtained with the *anosim* function from R package *vegan*, Figure S2). When calculating the diversity indices (Richness, Shannon and Simpson index) also, all samples showed similar values (Table S2).

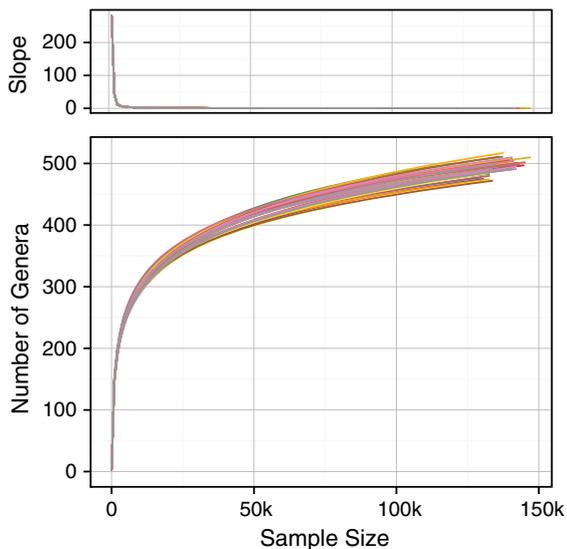
### Differential taxonomic assemblage in core and accessory bacteriome

To check whether the differences found in the composition of the bacterial community at genus level were also represented at higher taxonomy levels (Table S3), all genus assignments were collected and collapsed from phylum level to order level. Then, taxonomic differences between core and accessory bacteriomes were inspected (Fig. 5). Beginning from phylum level, core and accessory groups displayed significant differences in almost all phyla (only 2



**Fig. 1** Percentage of reads assigned at each taxonomic level. The number of reads assigned at each taxonomic level was calculated for each sample. The average number of assigned

reads for each taxonomic level from phylum to genus is reported. The standard error is shown at the *top right* of the plot



**Fig. 2** Rarefaction curves. Reported curves have been created by randomly re-sampling the pool of assignments for each sample multiple times and then plotting the average number of genera found in each sample. An increasing step of 1,000 assignments has been chosen in order to draw smooth curves. The *slopes* for each sample are shown in the *top panel* while rarefaction *curves* are shown in the *bottom panel*

phyla out of 23 did not differentiate between the two groups). In particular, six phyla (*Gemmatimonadetes*, *Elusimicrobia*, OD1, TM7, WS3) were only detected in core bacteriome. Six phyla were present in the accessory bacteriome (*Cyanobacteria*, *Aquificae*,

*Deinococcus-Thermus*, *Fusobacteria*, *Lentisphaerae*, OP11). The presence of such phyla exclusive of the two pan-bacteriome fractions highlight the presence of large taxonomic differences between them, which was also clear from a higher abundance of members of *Planctomycetes*, *Actinobacteria*, *Verrucomicrobia* and *Acidobacteria* in the core bacteriome, while the accessory bacteriome comprised more *Firmicutes*, *Chlamydiae* and *Proteobacteria*. Similar results were obtained by inspecting different taxonomic assemblage at class level. In fact, 16 classes out of 19 showed different occurrences between core and accessory bacteriomes, with 4 (*Erysipelotrichia*, *Epsilonproteobacteria*, *Holophagae*, Subdivision 5) and 2 (*Acidobacteria*, Subdivision 3) classes only present in the accessory and in the core bacteriome, respectively. Finally, almost all detected orders showed a differential occurrence patterns between accessory and core bacteriome, with 7 orders found exclusively in the core dataset and 12 orders found exclusively in the accessory dataset.

Finally, to give an insight into the genus-level distribution inside each sample (Table S4), the number of sequences attributed to each genus was transformed into a relative abundance value. These values were used to display bacterial taxonomic composition in each sample and in each bacteriome (Fig. 6). Results confirmed the overall representation of a conserved core group and of a more scattered accessory genus

distribution. Moreover, from this analysis the core group seemed to vary (a little) in relation to the sampling time, while the accessory bacteriome displayed variation with respect to sampling time to treatment and among repetitions.

### Defining core and accessory bacteriome

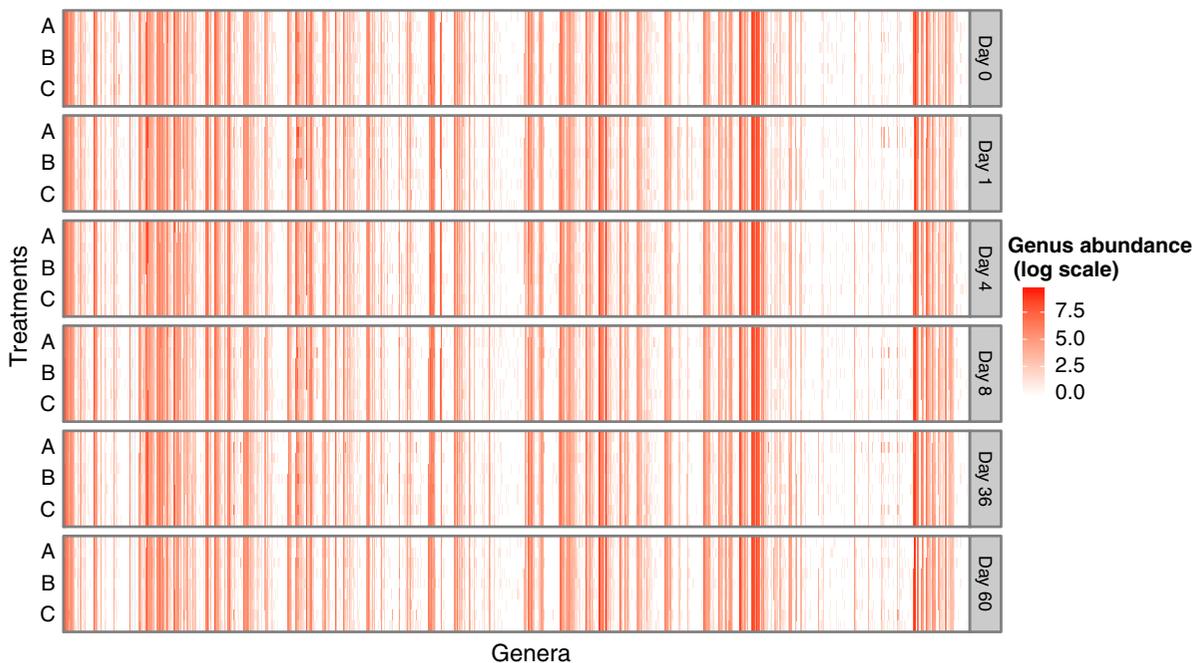
In order to describe the bacterial community structure of the soil under study, the  $X$  matrix was transformed using the logarithmic function. Thus, a heat map was produced displaying the abundance of all genera in the samples (Fig. 3; Table S4), which indicated the presence of highly represented and shared taxa (red regions over the whole map) and occasional taxa (regions of the heat map that do not display a uniform color inside the plot). These groups of taxa were defined (as detailed in the “Materials and methods” section) as core and accessory bacteriome, respectively. The core (shared) set of genera included those present in all samples (without considering changes in

their abundance), while the accessory bacteriome included those genera not present in all samples.

In terms of number of classified reads, the core and the accessory bacteriome contained 99.4 and 0.5 % of all sequences classified at genus level, respectively (data not shown). Furthermore, 120 genera were detected in one sample only (less than 0.1 % of the whole sequences); these genera were considered as “unique” ones and were excluded from the following analysis (Table S4). Regarding the core and the accessory bacteriome fractions we have collected 316 core genera (35.1 % of the total) and 465 accessory genera (51.6 % of the total) (Fig. 4).

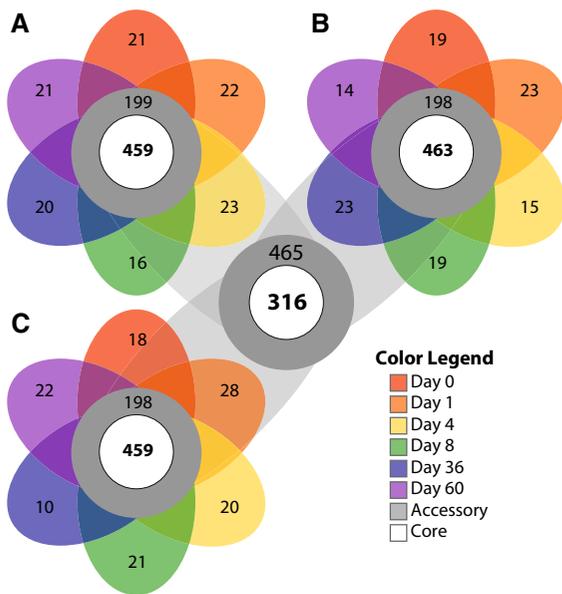
### Effect of sampling time and treatment on the whole pan-bacteriome components

We then considered the differential response of core and accessory bacteriome when comparing the three different mesocosms (treatment A, B, and C, with  $[\text{Cd}^{2+}]$  at 0, 5 and 25 mg kg<sup>-1</sup>, respectively). Two



**Fig. 3** Bacterial community structure at genus level. The number of assignments at genus level has been transformed using the logarithmic function in order to reduce the range of values. The obtained values have been reported using a heat map representation. As reported in the plot, red color represents high number of assignments ( $10^7$ ), while the white color indicates a low number of assignments (below  $10^1$ ). Samples were divided

into six blocks following sampling time (grey blocks to the right). Each block was divided into 9 lines, grouped into three parts following the three simulated environmental conditions (named A, B and C, corresponding to water, Cd 5 and 25 mg/kg, respectively). The 901 genera identified in the samples were ordered according to the taxonomic output of RDP Classifier (see Table S2)



**Fig. 4** The pan-bacteriome size. Pseudo-Venn diagrams drawn for each treatment dividing each sample according to sampling time. Generated plots report the number of genera found only at a specific time (unique assignments), the number of genera detected at least at two distinct times (accessory assignments) and the number of genera detected in all samples regardless sampling time (core assignments). The two groups reported in the *middle* of the three Venn diagrams are the accessory and the core groups defined with respect to the whole bacterial community, regardless sampling time and treatment. The three Venn diagrams marked with the letters *a*, *b* and *c* are referring to the 3 treatment conditions (0, 5, 25 mg/kg of Cd)

analyses were performed by CCA: one considering all samples collected for the analysis (replicates included) and one pooling the 3 replicates of each condition into a single observation. Both factors (sampling time and treatment) were fitted onto the two ordination analyses described above (Fig. 7). Results showed a tendency of the non-pooled samples to preferably cluster based on sampling time rather than on treatments. On the other hand, the pooled samples displayed an opposite behavior; in fact, they tended to cluster following the treatments (A, B, C) but not the sampling time. These contrasting results could be due to the high level of heterogeneity in the accessory bacteriome assignments. As shown in the bacterial communities analysis performed above (Fig. 6), the accessory group displays a very scattered bacterial distribution even inside the same temporal and treatment conditions. If this effect is minimized by pooling the 3 replicates inside a single sample even the effect of randomness is

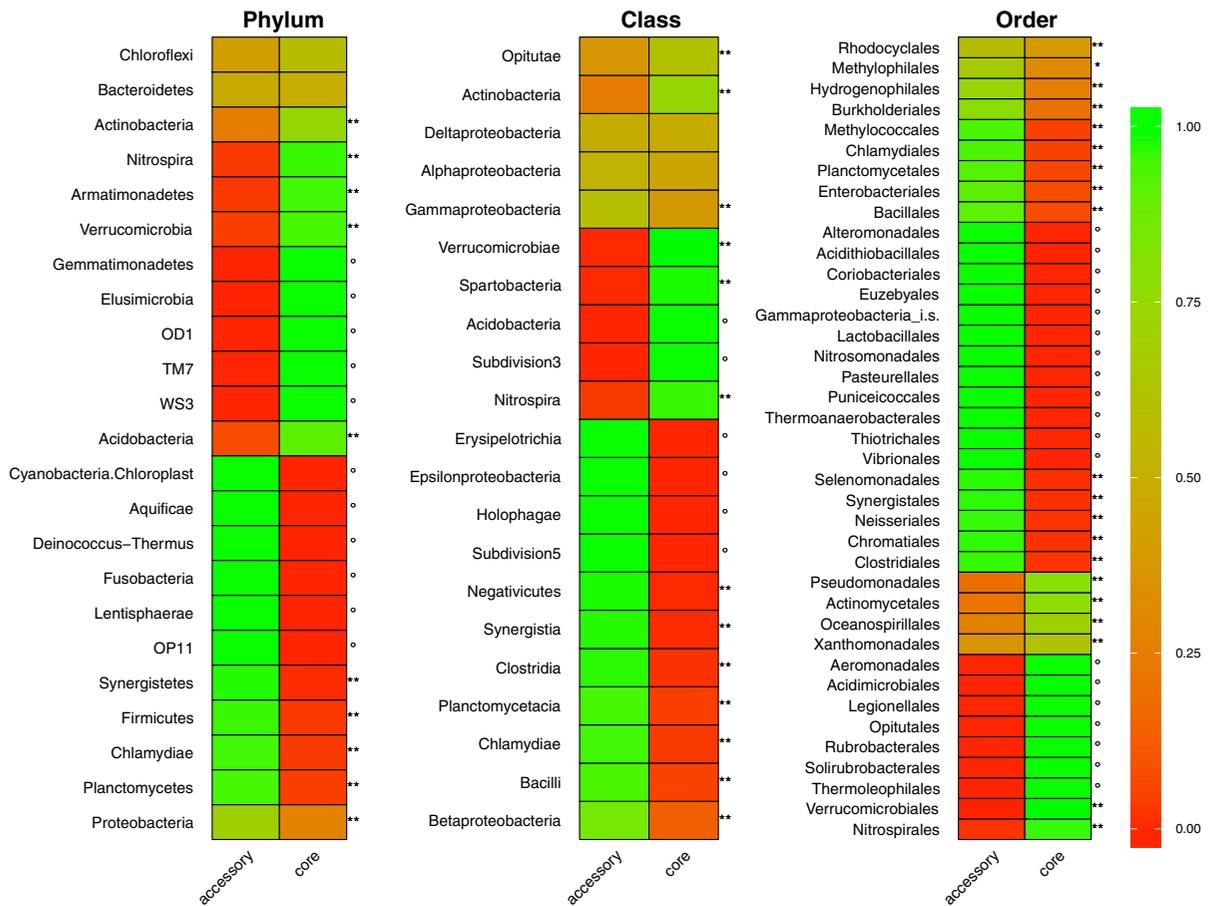
minimized. On the contrary, if this effect is not reduced, replicates of the same conditions result in a high degree of overlap in the ordination analysis.

## Discussion

It is now widely accepted that bacterial communities are composed by an assembly of resident taxa, those being slightly affected by environmental variables, and by occasional/fluctuating taxa, those varying among samples (Logares et al. 2013). Often resident taxa are the most abundant, while occasional taxa tend to be rare (Gobet et al. 2012). Recently, habitat specialization has been used to classify and provide insight into the ecological behavior of microbial taxa (Székely and Langenheder 2014).

To better describe the assemblage of bacterial communities as a sum of resident and occasional taxa, we have used a conceptual framework analogous to the pan genome concept (Tettelin et al. 2008), which is used for describing bacterial genomes belonging to the same species as sum of a common gene set (core genome) and of a dispensable/accessory gene set. In particular, we evaluated the presence of core and accessory bacteriomes in a set of three soil mesocosms followed over time.

The presence of two main fractions of the bacterial community was highlighted; one (the core), including the resident taxa while the other (accessory) composed of occasional taxa. The accessory taxa were by far fewer than the core taxa, considering the number of sequence assigned to each group. Since our experimental dataset included three different mesocosms (treatment A, B, and C, with  $[Cd^{2+}]$  at 0, 5 and 25 mg kg<sup>-1</sup>, respectively) followed for a relatively short time (60 days), occasional taxa could then include some fast growing/responding strains with rapid cycles of extinction (below amplification threshold)/colonization (above amplification threshold), as postulated for r-strategy taxa (Safriel and Ritte 1980), while the core bacteriome could include taxa with a k-strategy of growth. Treatment with high concentrations of Cd (A, B, C) did not substantially alter the diversity (Table S2) or the taxonomic assemblage of bacterial communities (Fig. 7). However, temporal shifts were detected which allowed to identify taxonomic signatures for the core and the accessory bacteriome. The core bacteriome was richer in



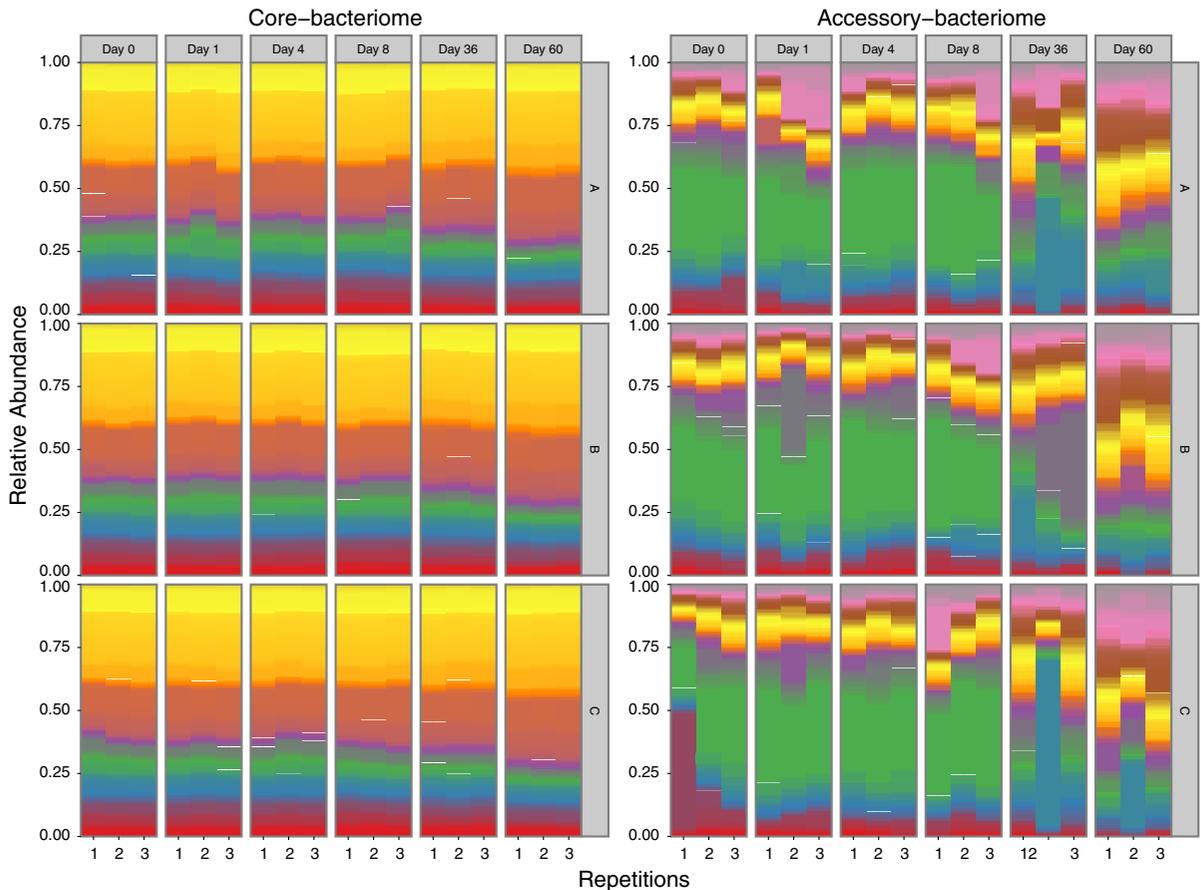
**Fig. 5** Phylogenetic dissection of variation of core and accessory bacteriome. Heatmaps report the difference of each taxonomic assignment between the core and the accessory bacteriomes. Differences have been tested using a *t* test (with the exclusion of presence/absence patterns). Colors are reflecting the level of the related genus in the heatmap. Green and red

colors correspond to high and low abundances values, respectively. The asterisks indicate the p value threshold in each comparison (one asterisk indicates a p value lower than 0.05 but higher than 0.001 while two asterisk indicates a p value lower than 0.001). The open circle indicates a strict presence/absence

members of phyla *Planctomycetes*, *Actinobacteria*, *Verrucomicrobia* and *Acidobacteria*, than the accessory bacteriome. The accessory bacteriome included more members of *Firmicutes*, *Clamydiae* and *Proteobacteria* (here in particular the order *Pseudomonadales*) than the core bacteriome. Moreover, the differential occurring taxa included 16S rRNA gene sequences from uncultured divisions, for which few or no functional information are available. In particular, *Armatimonadetes* (previously known as candidate phylum OP10) (Tamaki et al. 2010), OD1, TM7 and WS3 were only present in the core microbiome, suggesting they were scarcely or not affected by time in the three environmental condition tested

(mesocosms of A, B and C types). These phyla may include strains potentially being “seeds” for maintaining soil bacterial functionality and with low replication times, as k-strategists, similarly to the habitat generalist taxa, recently discussed for rock pools (Székely and Langenheder 2014).

Concerning the accessory bacteriome, several taxonomic orders were more represented, with respect to the core bacteriome, or exclusively found in the accessory bacteriome (Fig. 5). For instance, genera of the classes *Betaproteobacteria* (as *Burkholderiales* and *Methylophilales*), *Bacilli* (*Bacillales*) and *Clostridia* were more abundant in the accessory bacteriome. These taxa may include strains that are



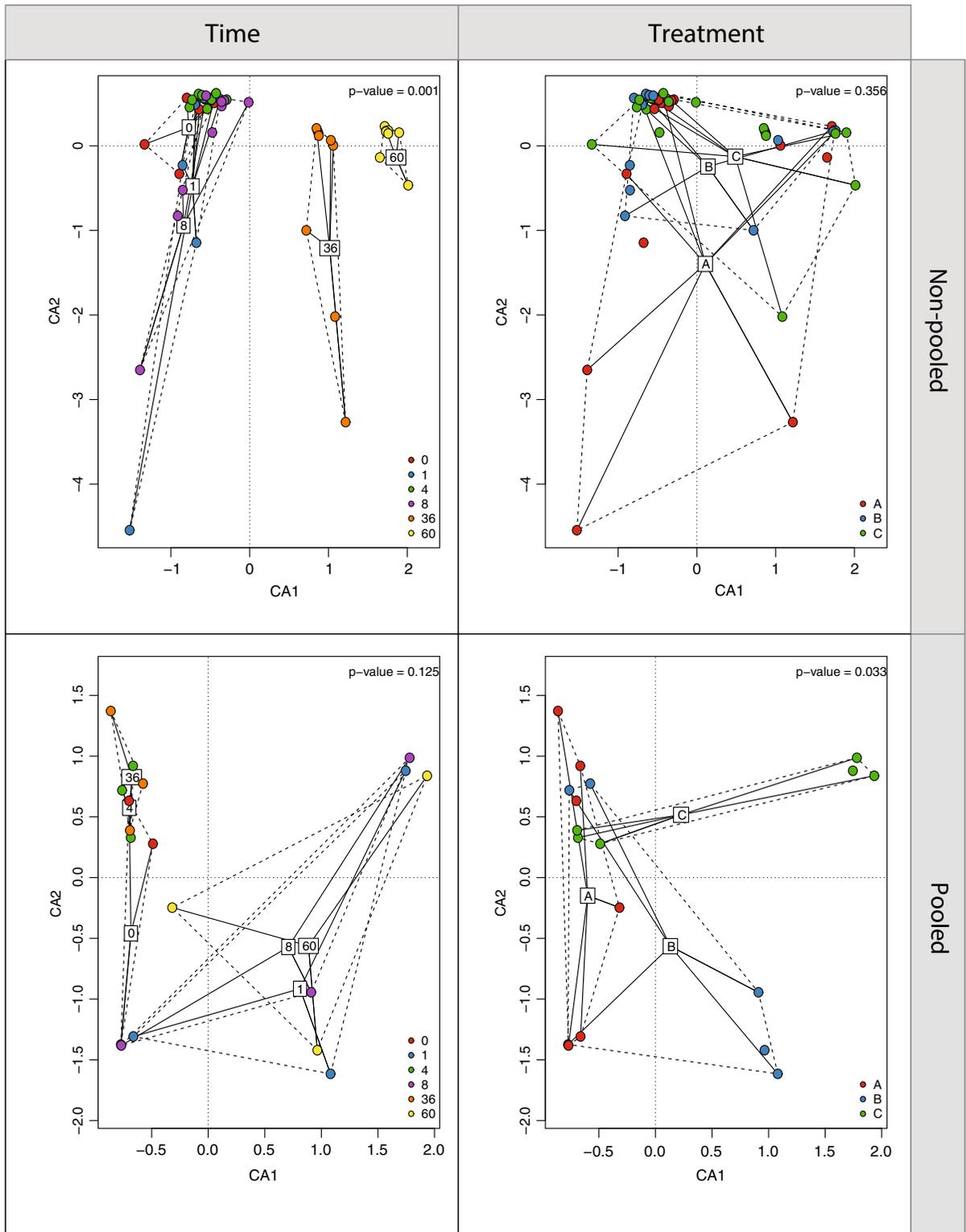
**Fig. 6** Relative abundances of genera detected in the core and accessory bacteriomes. Color patterns refer to distinct 901 genera (see Table S1). In the two panels, abundance values for

the core bacteriome and the accessory bacteriome are reported on the left and the right, respectively. A, B and C refer to the three treatments

r-strategist of selection and/or strains more sensitive to environmental perturbation. Therefore, their number may vary consistently between samples, dropping below the PCR amplification threshold in some particular conditions. Moreover, since many of these taxa (as *Bacilli* and *Clostridia*) may produce spores, which have a lower DNA extraction yield than cells (Dineen et al. 2010), we cannot exclude that some of the fluctuations observed in the accessory bacteriome for *Bacilli* and *Clostridia* may indeed be due to the differential presence of spores and living cells along sampling times. Additionally, a possible higher level of stochasticity of accessory bacteriome fluctuation can be inferred when comparing A, B and C treatments. In fact, the different results obtained in the comparison of treatments in pooled and non-pooled samples (Fig. 7) could indeed be ascribed to

the higher variability of replicas for the accessory fractions (see also Fig. 6). However, from the presented data it is not possible to draw so far general conclusions about the taxa which may be part of the core and of the accessory bacteriome in soil, because nutrient and other environmental conditions may differentially affect the bacterial taxa. Other studies employing soil mesocosms with different nutrient availability (as for instance organic carbon and nitrogen) and different type of stressors (i.e. salinity, temperature, other heavy-metals, etc.) should be performed in order to better evaluate the extent of core versus accessory bacteriome and their respective taxonomic compositions.

Moreover, the metabarcoding method here described only included phylum- to genus-level comparisons. Therefore, it is possible that some species or



◀ **Fig. 7** Effect of environmental variables on the pan-bacteriome assemblage. Canonical correlation analysis of the whole bacterial community are reported. The four plots are divided according to the time and treatment factor (*vertical division*) and according to the dataset used (*horizontal division*). In particular, the two plots in the *bottom panel* report the ordination analysis based on the pooled dataset (samples with the same conditions were pooled together) whereas the two plots in the *top panel* report the ordination analysis based on the whole bacterial community (all samples). P values derived from an environmental fitting analysis onto the bacterial dataset were reported in the *top right* corner of each plot

strains, belonging to identified genera, may have different responses to the treatments that were not observed by the analyses described here.

Future studies focused on single genera could then be useful in providing insights into the biological interpretation of the core/accessory bacteriome model in a narrow taxonomic range.

In conclusion, the concept of the pan-bacteriome as the sum of all taxa from the same soil at different times is proposed. This approach can allow identifying a core (stable) and an accessory (variable) pool of taxa, with a response to time and to different environmental conditions (mesocosms of A, B and C types). Core and accessory bacteriomes represented roughly 1/3 and 1/2 of the taxa detected in our dataset, respectively. These two bacterial fractions show a divergent number of assigned sequences; in particular the core group contains the 99.4 % of the sequences assigned at genus level while the accessory group contained only the 0.5 %. These data suggest a large flexibility of the pan-bacteriome in our experimental conditions. Furthermore, the two groups of soil bacteriome show a different taxonomic assemblage, which may reflect a different ability of these pools to persist in soil. We propose that the pan-bacteriome model of the bacterial community dynamics could be a useful approach for highlighting potential ecological (and functional) differences among bacterial taxa thriving in soil.

**Acknowledgments** This work was supported by a grant from the Ente Cassa di Risparmio di Firenze (Grant No. 2010/4384 “Centro di Metagenomica del suolo”).

## References

- Allison SD, Martiny JBH (2008) Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA* 105:11512–11519. doi:[10.1073/pnas.0801925105](https://doi.org/10.1073/pnas.0801925105)
- Aravindraja C, Viszwapriya D, Karutha Pandian S (2013) Ultradeep 16S rRNA sequencing analysis of geographically similar but diverse unexplored marine samples reveal varied bacterial community composition. *PLoS One* 8:e76724
- Ascher J et al (2009) Sequential extraction and genetic fingerprinting of a forest soil metagenome. *Appl Soil Ecol* 42:176–181. doi:[10.1016/j.apsoil.2009.03.005](https://doi.org/10.1016/j.apsoil.2009.03.005)
- Bacci G, Bazzicalupo M, Benedetti A, Mengoni A (2014) StreamingTrim 1.0: a Java software for dynamic trimming of 16S rRNA sequence data from metagenetic studies. *Mol Ecol Resour* 14:426–434. doi:[10.1111/1755-0998.12187](https://doi.org/10.1111/1755-0998.12187)
- Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol* 77:3846–3852. doi:[10.1128/aem.02772-10](https://doi.org/10.1128/aem.02772-10)
- Bartram AK, Jiang X, Lynch MD, Masella AP, Nicol GW, Dushoff J, Neufeld JD (2014) Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiol Ecol* 87:403–415
- Bowen JL, Morrison HG, Hobbie JE, Sogin ML (2012) Salt marsh sediment diversity: a test of the variability of the rare biosphere among environmental replicates. *ISME J* 6:2014–2023
- Campbell BJ, Yu L, Heidelberg JF, Kirchman DL (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* 108:12776–12781. doi:[10.1073/pnas.1101405108](https://doi.org/10.1073/pnas.1101405108)
- Ceccherini M, Ascher J, Pietramellara G, Mocali S, Viti C, Nannipieri P (2007) The effect of pharmaceutical waste-fungal biomass, treated to degrade DNA, on the composition of eubacterial and ammonia oxidizing populations of soil. *Biol Fertil Soils* 44:299–306. doi:[10.1007/s00374-007-0204-z](https://doi.org/10.1007/s00374-007-0204-z)
- Chaparro J, Sheflin A, Manter D, Vivanco J (2012) Manipulating the soil microbiome to increase soil health and plant fertility. *Biol Fertil Soils* 48:489–499. doi:[10.1007/s00374-012-0691-4](https://doi.org/10.1007/s00374-012-0691-4)
- Chien C, Kuo Y, Chen C, Hung C, Yeh C, Yeh W (2008) Microbial diversity of soil bacteria in agricultural field contaminated with heavy metals. *J Environ Sci* 20:359–363
- Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R (2009) Bacterial community variation in human body habitats across space and time. *Science* 326:1694–1697
- Dineen SM, Aranda R, Anders DL, Robertson JM (2010) An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *J Appl Microbiol* 109:1886–1896. doi:[10.1111/j.1365-2672.2010.04816.x](https://doi.org/10.1111/j.1365-2672.2010.04816.x)
- Dixon P (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930
- Dohrmann AB, Kütting M, Jünemann S, Jaenicke S, Schlüter A, Tebbe CC (2013) Importance of rare taxa for bacterial diversity in the rhizosphere of Bt-and conventional maize varieties. *ISME J* 7:37–49
- Donn S, Kirkegaard JA, Perera G, Richardson AE, Watt M (2014) Evolution of bacterial communities in the wheat

- crop rhizosphere. *Environ Microbiol*. doi:[10.1111/1462-2920.12452](https://doi.org/10.1111/1462-2920.12452)
- Duan X, Huang C (2008) Effect of cadmium on diversity of microbial genes in paddy soil. *Chin J Appl Environ Biol* 14:510–513
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred II. Error probabilities. *Genome Res* 8:186–194
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred I. Accuracy assessment. *Genome Res* 8:175–185
- Fritze H et al (2000) Effect of Cd-containing wood ash on the microflora of coniferous forest humus. *Fems Microbiol Ecol* 32:43–51
- Ge Y, He J-Z, Zhu Y-G, Zhang J-B, Xu Z, Zhang L-M, Zheng Y-M (2008) Differences in soil bacterial diversity: driven by contemporary disturbances or historical contingencies? *ISME J* 2:254–264
- Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci USA* 110:4651–4655
- Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R, Reid G (2010) Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* 5:e15406
- Gobet A et al (2012) Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J* 6:542–553
- Gomes NCM, Landi L, Smalla K, Nannipieri P, Brookes PC, Renella G (2010) Effects of Cd- and Zn-enriched sewage sludge on soil bacterial and fungal communities. *Ecotoxicol Environ Saf* 73:1255–1263
- Hardoim CCP, Costa R (2014) Temporal dynamics of prokaryotic communities in the marine sponge *Sarcotragus spinosulus*. *Mol Ecol* 23:3097–3112. doi:[10.1111/mec.12789](https://doi.org/10.1111/mec.12789)
- Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, Sogin ML (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* 4:e1000255
- Kim TS, Jeong JY, Wells GF, Park HD (2013) General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. *Appl Microbiol Biotechnol* 97:1755–1765
- Kuang J-L et al (2012) Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J* 7:1038–1050
- Lazzaro A, Widmer F, Sperisen C, Frey B (2008) Identification of dominant bacterial phylotypes in a cadmium-treated forest soil. *FEMS Microbiol Ecol* 63:143–155
- Logares R et al (2013) Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J* 7:937–948
- Lorenz N, Hintemann T, Kramarewa T, Katayama A, Yasuta T, Marschner P, Kandeler E (2006) Response of microbial activity and microbial community composition in soils to long-term arsenic and cadmium exposure. *Soil Biol Biochem* 38:1430–1437
- L-X Chen, J-T Li, Y-T Chen, L-N Huang, Z-S Hua, Hu M, Shu W-S (2013) Shifts in microbial community composition and function in the acidification of a lead/zinc mine tailings. *Environ Microbiol* 15:2431–2444. doi:[10.1111/1462-2920.12114](https://doi.org/10.1111/1462-2920.12114)
- Masella A, Bartram A, Truszkowski J, Brown D, Neufeld J (2012) PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinform* 13:31
- Mengoni A, Schat H, Vangronsveld J (2010) Plants as extreme environments? Ni-resistant bacteria and Ni-hyperaccumulators of serpentine flora. *Plant Soil* 331:5–16
- Oh J et al (2013) The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. *Genome Res* 23:2103–2114
- Oksanen J et al. (2013) vegan: community ecology package. R package version 2.0-10. <http://vegan.r-forge.r-project.org/>
- Pedros-Alió C (2012) The rare bacterial biosphere. *Annu Rev Mar Sci* 4:449–466. doi:[10.1146/annurev-marine-120710-100948](https://doi.org/10.1146/annurev-marine-120710-100948)
- Pini F, Frascella A, Santopolo L, Bazzicalupo M, Biondi E, Scotti C, Mengoni A (2012) Exploring the plant-associated bacterial communities in *Medicago sativa* L. *BMC Microbiol* 12:78
- Porter SS, Rice KJ (2013) Trade-offs, spatial heterogeneity, and the maintenance of microbial diversity. *Evolution* 67:599–608
- Portillo MC, Leff JW, Lauber CL, Fierer N (2013) Cell size distributions of soil bacterial and archaeal taxa. *Appl Environ Microbiol* 79:7610–7617
- Renella G, Brookes PC, Nannipieri P (2002) Cadmium and zinc toxicity to soil microbial biomass and activity. *Dev Soil Sci* 28:267–273
- Renella G, Mench M, Landi L, Nannipieri P (2005) Microbial activity and hydrolase synthesis in long-term Cd-contaminated soils. *Soil Biol Biochem* 37:133–139
- Safriel UN, Ritte U (1980) Criteria for the identification of potential colonizers. *Biol J Linn Soc* 13:287–297
- Sánchez O, Ferrera I, González JM, Mas J (2013) Assessing bacterial diversity in a seawater-processing wastewater treatment plant by 454-pyrosequencing of the 16S rRNA and amoA genes. *Microb Biotechnol* 6:435–442
- Sheoran V, Sheoran AS, Poonam P (2008) Remediation techniques for contaminated soils. *Environ Eng Manag J* 7:379–387
- Smith BC et al (2012) The cervical microbiome over 7 years and a comparison of methodologies for its characterization. *PLoS One* 7:e40425
- Sogin ML et al (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA* 103:12115–12120. doi:[10.1073/pnas.0605127103](https://doi.org/10.1073/pnas.0605127103)
- Székely AJ, Langenheder S (2014) The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *Fems Microbiol Ecol* 87:102–112
- Tamaki H et al (2010) *Armatimonas rosea* gen. nov., sp. nov., of a novel bacterial phylum, *Armatimonadetes* phyl. nov., formally called the candidate phylum OP10. *Int J Syst Evol Microbiol* 61:1442–1447. doi:[10.1099/ijs.0.025643-0](https://doi.org/10.1099/ijs.0.025643-0)
- Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* 11:472–477
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. doi:[10.1128/aem.00062-07](https://doi.org/10.1128/aem.00062-07)

- Wegner KM, Volkenborn N, Peter H, Eiler A (2013) Disturbance induced decoupling between host genetics and composition of the associated microbiome. *BMC Microbiol* 13:252
- Xu XH, Liu CY, Li RY, Wang XR (2013) Impacts of cadmium stress on soil microbial and enzyme activities under laboratory conditions. *J Food Agric Environ* 11:1730–1734
- Zhang Y, Zhang X, Zhang H, He Q, Zhou Q, Su Z, Zhang C (2009) Responses of soil bacteria to long-term and short-term cadmium stress as revealed by microbial community analysis. *Bull Environ Contam Toxicol* 82:367–372
- Zhou J et al (2014) Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc Natl Acad Sci USA* 111:E836–E845. doi:[10.1073/pnas.1324044111](https://doi.org/10.1073/pnas.1324044111)