Functional selection associated with microbial genome size under polycyclic aromatic hydrocarbons (PAHs) pollution stress in soils

Yuzhu Dong  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Shanghua Wu  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Haonan Fan  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Xianglong Li  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Yijing Li  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Shengjun Xu  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Zhihui Bai  
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Xuliang Zhuang (✉ xlzhuang@rcees.ac.cn)  
Research Center for Eco-Environmental Science, Chinese Academy of Sciences  https://orcid.org/0000-0001-5902-4267

Research

Keywords: Pollution stress, Genome size, Polycyclic aromatic hydrocarbons, Soil microbial community

Posted Date: June 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34776/v1

License: ☝️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Bacteria can adapt to complex environmental conditions by expanding or reducing their genomes. Microbes with large genomes carry a more diverse gene pool and thus have more metabolic function than those with small genomes. Various environmental disturbances have been investigated and proved to be important in determining the community genome sizes. However, the impacts of environmental pollution remain poorly understood, which is a selective factor. Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous pollutants that cause great damage to the natural environment and health. Here, we conducted incubation experiments under different concentrations to evaluate the impacts of organic pollution stress on the community genome size. Besides, changes of community structure, functions and potential key genes in different concentrations were illustrated by 16S rRNA amplicon gene sequencing and quantitative-PCR.

Results

We found the distinct communities of significantly larger genome size with the increase of PAHs concentration gradients in soils. Highly contaminated soil notably enhanced functions relative to xenobiotics biodegradation and metabolism, environmental information processing and signaling molecules and interaction. Abundance of Proteobacteria and Deinococcus-Thermus with relatively larger genomes increased along with PAHs stress and well adapted to polluted environments, while abundance of Patescibacteria with a highly streamlined and smaller genome decreased. Co-occurrence network analysis indicated a decreased average clustering coefficient but an increased connection number per node positively related to nidA gene along the concentration gradients. Moreover, nidA gene, appearing as an indicator, showed positive correlation with community genome size and PAHs concentrations.

Conclusions

Our findings demonstrate the significant impact of pollution stress on genome size and clarify that the increase in genome size is due to the directional selection for microorganisms with larger genomes or carrying specific functional genes and the increased capacity for horizontal transfer of degrading genes between communities.

Background

Soils harbor diverse microbiomes on Earth and soil microorganisms are sensitive to environment disturbances [1, 2]. The indigenous microbiomes in soils have a capacity for self-adjusting to adapt to changing environmental stresses [3, 4]. Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous pollutants that cause great damage to the natural environment and health [5]. Soil is the reservoir and
transfer station of PAHs in the environment, bearing the major environmental load of PAHs [6]. PAHs stress makes some microorganisms enhance their ability to mutate and evolve in order to adapt, which led to changes of microorganisms at the genome level [7, 8].

In previous studies, some microorganisms have been isolated from PAHs contaminated soil [5, 9], and studies of isolated bacteria in single cell level have shown that some bacteria have high GC content and large genomes size [11, 12]. In addition, some researchers have demonstrated that PAHs stress altered soil community structure and composition and enriched specific functional genes (i.e. polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenases (PAH-RHD) encoding genes) [5, 10]. However, at the community level, the detailed mechanisms of how the genome responds to environmental stress are not well clear. Especially, microbial communities are better able to survive in diverse environments than individual species because of their diverse metabolic patterns [13, 14]. Therefore, it is necessary to elucidate the reply of genomes to pollution stress at the community level.

Compared with many eukaryotes, bacterial genomes are relatively small, and tightly packed with protein-coding genes are tightly bound [15, 16]. Bacterial genomes vary in size by more than two orders of magnitude, ranging from 112 kb [17] to 16.04 Mb [18]. Several major evolutionary mechanisms in the bacterial genome: mutation and horizontal gene transfer, cause changes in genome size [15, 19–23]. Mutation is generally random, resulting in an increase or loss of bases [20, 21], while horizontal gene transfer brings new genes into the bacterial genome [22, 23]. In general, free-living bacteria have a larger genome than most parasitic bacteria, which discard some biosynthetic pathways and genes that could benefit from the host [24, 25]. Hence, genome size is important for microbial adaptation to environmental disturbances because it directly affects microbial functions.

Genome size of free-living bacteria is influenced by many factors, such as niche [7, 26] and nutrient availability [27, 28], but is not usually determined by replication efficiency [15]. Bacterial niche changes with the change of environmental conditions and the intrinsic ability of bacteria to adapt to the changing environment is related to genome size [27, 29]. In recent years, some studies have shown that environmental factors, including poor nutritional conditions [27, 30], depth [31], environmental stability [32], high temperature [33, 34], alkaline milieu [35] and symbiotic lifestyle [24], affect the microbial genome size to some extent. When exposed to a new environmental stress, the bacterial genome size may be increased by acquiring adaptive gene modules [22, 23], or reduced by discarding related genes for additional metabolic burdens [24, 34]. That is, the variation in the size of bacterial genomes is largely due to the acquisition and loss of functional helper genes. Generally, bacteria with large genomes size have greater metabolic diversity and are better able to cope with a wide range of environmental conditions [25, 32]. In addition, bacteria can also improve their ability to survive by expanding their gene pool [29].

In this study, we harness an experimental evolution approach in order to decipher how the community genome size are affected by the pollution disturbances, and how the community was selected and changed under pollution stress. We tested the potential change mechanisms of genome size based on niche and function. We speculated that the community genome size may increase when the community
stable state was broken by environmental pollution and hypothesized that (1) the pollution stress would select for bacteria with large genomes with degrading genes or a wide metabolic range; (2) a potential higher occurrence of horizontal gene transfer between resident species account for the changes of genome size in community level.

Results

Inhibition of abundance and diversity under pyrene stress

To analyze the effect of pyrene in in-situ soils on the total number of microbial community, we measured the residual pyrene concentration and quantified the abundance of microbial community. A negative correlation was found between degradation rate and spiked pyrene over the 35-day period. The abundance of total bacterial, archaeal and fungal communities were significantly inhibited by pyrene. The bacterial gene copy numbers were lowest in PYR500, indicating some inhibiting effect of high concentration on the bacterial growth (Fig. 1a). There was a statistically significant difference in bacterial gene copy numbers between the five treatments (ANOVA, P< 0.001). The abundance of bacteria in PYR500 was significantly lower than in non-contaminated and other spiked-pyrene soils (Fig. 1a; Tukey HSD post hoc test, P< 0.05). In addition, Archaeal and fungal rRNA gene copy numbers were about 1–2 orders of magnitude lower compared with bacteria (Additional file: Fig. S2). The abundance of archaea and fungi in PYR100 decreased dramatically, with no significant difference compared with PYR500 (P> 0.05), suggesting that archaea and fungi were more sensitive to high concentrations of pyrene.

To figure out the influence of different pyrene stress on species richness and evenness, we calculated different alpha diversity indices. Alpha diversity described the diversity of the community members within a single sample. Alpha diversity values decreased with the increasing concentrations of pyrene, regardless of diversity measure used (Fig. 1b, Additional file: Table S3). The Shannon index of species diversity ranged between 5.67 and 7.05 and differed significantly depending on pyrene concentration (P< 0.001). Furthermore, Tukey HSD post hoc comparison identified high concentration (PYR100 and PYR500) to have significantly lower bacterial diversity compared with control and low concentration treatment (P< 0.05).

Resistance index, an ecologically relevant parameter, was calculated based on α-diversity (Shannon index) to estimate the stability of microbial communities in response to different pyrene stress. Across all experiment treatments, resistance index was altered significantly (P< 0.05, Fig. 1c) and decreased with increasing pyrene concentrations. The microbial communities in PYR500 had significantly lower resistance compared with other treatments (P< 0.05), except for PYR100 (P> 0.05), which further suggested that 100 mg/kg may be the turning point for the bacterial community to reach another steady state.

Changes of microbial structure and composition
We investigated the community structure and composition of all treatments to determine whether there were differentially representative taxa, which might be conducive to account for the differences in diversity and resistance. Non-metric multidimensional scaling (NMDS) of OTU-based Bray-Curtis dissimilarity distances was measured to visualize the similarities and differences in community composition between different samples. The samples spiked with and without pyrene formed clearly delimited groups and also clustered separately (Fig. 2a). Treatment CK, PYR1 and PYR10 were separately clustered at a certain distance, whereas samples of PYR100 overlapped with PYR500 to a large extent, indicating that there was no/low effect on microbial community structure when the pyrene concentration in soil was more than 100 mg/kg dry soil. The bacterial community structures were significantly changed by pollution levels of pyrene based on Bray–Curtis dissimilarity (ANOSIM, $R = 0.895$, $P = 0.001$), suggesting pollution levels were a potential predictor of community composition (Additional file: Fig. S3).

A total of 25,260 distinct OTUs were annotated to 42 different phyla, and the samples of experimental treatments were divided into two concentration-level groups according to different community compositions: (1) the low and medium concentration group (LM group; PYR1 and PYR10); (2) the high concentration group (H group; PYR100 and PYR500). Across all samples, *Proteobacteria* was the most abundant phylum, accounting for 35.8–45.7% of the total effective sequences (Fig. 2a). The other predominant phyla were *Firmicutes* (9.6–12.9%), *Actinobacteria* (7.0–11.1%), *Acidobacteria* (6.9–10.1%), with smaller proportion of *Chloroflexi* (6.7–9.5%) and *Bacteroidetes* (5.3–8.8%). Among the dominant phyla, only *Actinobacteria* significantly decreased with increasing pyrene concentrations ($P < 0.05$). Two dominant phyla, *Proteobacteria* and *Deinococcus-Thermus*, were identified as significantly differentially abundant taxa between the LM and H groups, while *Patescibacteria* showed a drastically decrease. Compared with CK, the overall trend of *Firmicutes* and *Patescibacteria* was an increase and *Proteobacteria* had been slightly inhibited in LM group. In contrast, *Proteobacteria* was significantly enriched in H group compared with CK.

To explain the great differences in microbial composition caused by pyrene stress in soils among all samples, Random Forests analysis with 10-fold cross-validation of 5 repeats was performed at the phylum level. Finally, according to the cross-validation error curve, 12 phyla were defined as biomarkers to distinguish samples of different concentrations, among which *Fibrobacterae* (4.25%) and *Nitrospirae* (2.75%) might contribute most to the community differences under pyrene stress(Fig. 2b). Intriguingly, almost all the biomarkers picked by Random Forest accounted for only 1.5–4.2% of all taxa, most of which had a relative abundance of less than 1%, except *Proteobacteria* (35.8–45.7%).

In order to investigate whether these significant community differences at different pyrene stresses affected soil genomes, an average genome size for each sample was calculated based on the relative abundance of the community at phylum level. A strong positive correlation was found between the logarithm of the concentrations and average genome size (Fig. 2c; Pearson's $r = 0.5164$, $P = 0.0082$). Compared with CK, the soil average genome size of PYR500 increased by 0.43 Mbp.
Linking functional gene and genome size to potential community functions

To explore how these differences in biodiversity and community composition affect the functions to change, we analyzed PICRUSt based on KEGG Orthology database to predict the functional composition of soil bacterial communities. A total of 298 level-3 pathways affiliating with 7 components in KEGG were predicted, and metabolism accounted for about 53%, followed by genetic information processing (16%). Functional pathways at KEGG level 2 revealed that differences in pyrene stress led to a significant shift differences across all samples in 39 functional categories, except “biosynthesis of other secondary metabolites”, “Carbohydrate metabolism” and “Membrane transport” (ANOSIM, R = 0.392, P = 0.001). According to the functional hotspot, these changing KEGG modules were divided into two functional groups: internal functions within microorganism (defined as Self-functional group) and external functions between microorganisms (defined as Community functional group) (Fig. 3a).

In Self-functional group, the relative abundances of two main categories relating to Metabolism (i.e. “Xenobiotics biodegradation and metabolism” and “Metabolism of other amino acids”) and Cellular Processes (i.e. “Cell motility” and “Cell growth and death”) were more abundant in samples spiked with pyrene. However, not all categories relating to Metabolism had a consistent trend of increasing. For instance, abundance of “Glycan biosynthesis and metabolism”, “Nucleotide metabolism” and “Energy metabolism” were significantly lower in Experiment Treatments compared with CK (Fig. 3a, P < 0.05). The abundance of Environmental information processing KEGG Modules including “Replication and repair” and “Translation” gradually eliminated as the pyrene stress increased. For functional categories of Community functional group, “Signal transduction” and “Signaling molecules and interaction” showed an increasing trend along with increasing pyrene concentrations, suggesting pyrene stress enhanced the ability of microorganisms to communicate with each other.

To further test how the specific xenobiotics (Pyrene) biodegradation and metabolism function changed at different pyrene stresses, we chose the nidA gene, which was a key dioxygenase gene for pyrene degradation, to evaluate the abundance of pyrene-degrading bacteria. No nidA gene was detected over the course of 35 days in CK. After 35 days of incubation, the nidA gene copy numbers across spiked-pyrene samples significant changed (P< 0.05) from $6.85 \times 10^3$ to $8.74 \times 10^4$. The number of nidA gene copies increased quickly with pyrene concentrations, reaching the maximum value at PYR100. Thereafter, there was no significant decrease in PYR500 compared with PYR100 (P > 0.05). Correlation analysis between unique nidA gene and metabolic functional categories at KEGG level 2 showed that the indicator gene for pyrene metabolism were positively correlated with, metabolism of other amino acids and xenobiotics biodegradation and metabolism (Fig. 3b; P< 0.05). A significant negative correlation was revealed between nidA gene and Nucleotide metabolism (“Sulfur metabolism”, “Nitrogen metabolism”, “Methane metabolism”).

Previous studies have shown that the increase in genome size was proportional to the number of metabolic genes [28]. To test whether this consequence could also be observed in our experiments, we
calculated the correlation between genome size and metabolic functions. The average genome size was only positively correlated with “Xenobiotics biodegradation and metabolism” (Fig. 3b; \( P < 0.05 \)). Focusing on the specific degradation gene (\( nidA \)), we discovered that there was also a significant positive correlation with genome size (\( P < 0.05 \)).

Co-occurrence network under pyrene stress

In order to determine the effect of pyrene on soil microbial ecological links between different species, 5 networks were constructed at the OTU level based on spearman correlation. The numbers of nodes (979–1987) decreased from 1987 to 979 with the increase of pyrene concentration, and edges between microorganisms reduced from 3720 to 1413. Compared with CK (maximum value, 0.539), the average clustering coefficient in experimental treatments was decreased by 0.011, 0.086, 0.018 and 0.097 (Fig. 4), respectively, which indicate pyrene decreased the soil microbial associations.

To obtain a better understanding of the cooperative relationships of pyrene-degrading consortiums, we performed Spearman correlation analysis to link the unique \( nidA \) gene and bacterial taxa profiles. A total of 151 OTUs belongs to 14 phylum in spiked-pyrene samples were significantly positively correlated with \( nidA \) gene copy numbers (\( r > 0.8, P < 0.05 \)), which was supposed as the “potential degraders”. \( Proteobacteria \) and \( Chloroflexi \) were the most abundant potential degraders, accounting for about 60%. There were 43, 56, 79 and 61 “potential degrader” in experimental treatments. In addition, 142,132,180 and 134 nodes in PYR1, PYR10, PYR100 and PYR500 were associated with the “potential degrader”, respectively, which might become the “potential horizontal gene transfer receptors”. An average of 3.33, 2.43, 2.61 and 2.36 links in each network was correlated with each potential degrader, with an average of 1.24, 1.18, 0.99 and 1.05 positive correlation links. In general, the addition of pyrene resulted in an increase in the number of OTUs positively correlated with the \( nidA \) gene copies, and the same trend was observed for the edges positively correlated with these OTUs. In experimental treatments, OTUs positively correlated with the potential degraders mainly belong to \( Proteobacteria \), \( Chloroflexi \), \( Actinobacteria \) and \( Acidobacteria \) (Additional file: Fig. S4).

Verification of genome size and \( nidA \) gene in soils of different latitudes

In order to further verify the universality of our conclusion that genome size increased under pyrene stress, we selected soils at two other latitudes to conduct microcosmic incubation experiments. The average genome sizes of soils with added contaminants in HZ-PYR500 and GZ-PYR500 were 3.69 ± 0.033 Mb and 3.94 ± 0.038 Mb, increasing by 0.08 and 0.18 Mbp compared with HZ-CK and GZ-CK (Fig. 4), respectively, which showed a consistent trend with the soil in Beijing. In addition, the \( nidA \) gene was also not detected in uncontaminated soil and the average \( nidA \) gene copy numbers of HZ-PYR500 and GZ-PYR500 were \( 2 \times 10^5 \) and \( 8.2 \times 10^5 \), respectively (Additional file: Fig. S5).

Discussion
In this study, we found for the first time that PAHs increased the average genome size of soils. Pyrene, as a selective pressure and a special carbon source, on the one hand, specifically enriched larger genome microorganisms with stronger adaptability, and on the other hand, enhanced the ability of microbial communities to communicate and potential horizontal gene transfer between species.

Based on previous studies, genome size was affected by many different factors. Firstly, species in stable community tend to have small genomes \[7, 25\]; secondly, shift in community composition and structure can change the community genome size and microorganisms with larger genome size have better adaptation to environmental stress \[25, 32\]; thirdly, strong ability of community communication and horizontal transfer can improve the viability of bacteria in a changing environment \[22, 29\].

For the first factor, in our work, we discovered significant changes in community stable state under different pyrene stresses. The Shannon and resistance index decreased with the increasing concentration, indicating that different concentrations of pyrene may be used as environmental filtering factors in different degrees to screen microorganisms with specific niche. Moreover, in the $\beta$-diversity analysis, it was found that other treatments except the high-concentration treatment accumulated separately, suggesting that the low-concentration pyrene broke the stable state of the original microbial community, while the high-concentration pyrene made the community enter a new stable state and reshaped the community niche. As reported in previous studies, species living in unchanging niches tend to have small genomes \[7, 26\]. In our study, it was found that pyrene not only led to the change of niche but also caused the increase of average genome size, and this result was confirmed in three soils of different latitudes, indicating that the positive correlation between the genome size and pollutant concentration had a certain universality. Overall, significant changes in species diversity and niche led to an increase in community average genome size.

Besides, it is generally believed that microorganism with larger genomes are better able to cope with a wide range of environmental conditions \[15, 24\]. We then focused on the change of community composition under the selection of pollutant. Results showed that among the abundant phyla (> 1%), Proteobacteria and Deinococcus-Thermus were significantly enriched but the relative abundance of Actinobacteria and Patescibacteria decreased. We checked the genome size of these species and found that Proteobacteria and Deinococcus-Thermus had genome sizes of 0.16–16.04 Mb and 2.02–6.65 Mb according to the Integrated Microbial Genomes ( IMG ) database, respectively. Previous studies have shown that bacterial genome distribution peaks at about 2 Mb \[16\] and average bacterial genome size was ~ 3.87 Mb \[36\]. Proteobacteria and Deinococcus-Thermus, with average genome sizes of 4.4 and 3.1 Mb, had relatively larger genome sizes compared with other phyla. Moreover, Proteobacteria had been widely demonstrated to be dominant in PAH-contaminated soils, and most of the isolated PAH-degrading bacteria were phylogenetically belonged to Proteobacteria \[5, 37\]. These suggested that Proteobacteria might have their own unique mechanisms for PAHs-tolerance, possibly by carrying genes with specific functions such as $nida$ gene. The result that OTU with positive correlation with $nida$ gene copy numbers mainly assigned to Proteobacteria proved that Proteobacteria had the potential to utilize pyrene as carbon and energy source. The enhancement of degradation capacity further promoted...
xenobiotics biodegradation and metabolism. The increase in genetic information processing compared with the control treatment may be due to *Deinococcus-Thermus*, which carried unique carotenoid genes or enzymes. Carotenoids were helpful to protect genetic material from stress and thus made bacteria better adapt to harsh environments [38]. Interestingly, *Patescibacteria*, with a decreasing relative abundance, had a highly reduced genome size of $1.1 \pm 0.2$ Mb and significantly streamlined cellular activities (cell motility; signal transduction) and metabolic potentials (carbohydrate, protein, and lipid metabolisms; sulfur and nitrogen compound metabolisms) [39]. To some extent, the increasing phyla including *Proteobacteria* and *Deinococcus-Thermus* had bigger genome sizes and did well in adaptation to PAHs-contaminated soils, while the decreasing phyla containing *Patescibacteria* owned simplified genomes and reduced function related to metabolisms and cellular activities. In a word, the increase of large-genome microbes and the decrease of small-genome microbes eventually resulted in the increase of community genome size.

Last, the increase of *Proteobacteria* provided more genes in the community to handle with the changes of circumstances. Some microbes can choose other strategies to improve their own adaptation to PAHs selective pressure, in addition to carrying degradation or anti-stress genes [37]. Many previous studies have found that microorganisms in a community have ability to degrade and tolerate xenobiotic compounds through horizontal transfer of mobile genetic elements [40, 41]. Our results have shown that the addition of pyrene in uncontaminated soils activates pyrene degradation genes. Furthermore, in the network, it could be discovered that with the increasing concentration, the proportion and connectivity of potential degraders and OTUs connected to them improved. The results suggested that there was a strong link between the microbes. What was interesting was that these OTU belonged mainly to *Proteobacteria*, *Chloroexi* and *Actinobacteria*. Moreover, currently, the microorganisms carrying catabolic mobile components are mainly *Proteobacteria* and *Actinomyces*, such as *Pseudomonas* sp., *Sphingomonas* sp., and *Mycobacterium* sp. [37, 40, 42]. These degradation genes, whether on plasmids or chromosomes, are usually transferred in the form of an insertion machines-elements. In addition, the increase of environmental adaptability, signal transduction and signaling molecules and interaction indirectly proved that there was a strong cooperative function between microbial communities.

**Conclusions**

Overall, this is the first time to discover an increase in genome size under PAHs stress and associated with specific functions. In our study, different PAHs concentrations were found to inhibit microbial growth, alter microbial community structure, and significantly increase the copy numbers of *nidA* gene. Based on the positive correlation among genome size, functional genes and metabolic function, we propose a possible hypothesis for the increasing genome size. As a powerful selective factor, PAHs not only enriched the large genome microorganisms carrying degrading or anti-stress genes with stronger adaptability, but also promoted the horizontal transfer of mobile genetic elements with degrading genes, thus improving the tolerance of the whole community. However, much more follow-up work need to be done, not only more direct evidence is needed to confirm the role of horizontal gene transfer, but also
make sure whether there are other functional genes associated with genome size. Moreover, we should investigate whether genome size decreases with the disappearance of pollution stress.

**Methods**

**Soil sampling and experimental design**

Surface soil sample (from 0 to 20 cm) with unknown history of PAHs contamination was collected from an area (40°23' E and 116°40' N) in Changping District, Beijing, North China. One part of the sample was air-dried at room temperature, homogenized with a 2 mm sieve and stored in the dark at room temperature until the microcosm incubation, and the remaining part was stored in a self-sealing bag at -20°C to measure soil physicochemical properties. The original soil was one type of fluvo-aquic soil according to the FAO/UNESCO soil classification system and had a pH of 7.9. Additional details on soil characteristics are provided in the supplementary materials (Additional file: Table S1).

In order to better reveal the variation of genome size in soils with different pollution levels, microcosms under laboratory were constructed. All the microcosms were carried out by placing 10 g soils into 120 ml serum bottles which were sealed with rubber stoppers and aluminum caps, and sterile distilled water was added every 2 days to keep the soil moisture content at about 60%. The headspace in these microcosms was replaced daily with sterile filtered synthetic air (20% O₂, 80% N₂), then incubated at 25°C in the dark.

Five treatments were established with 12 replicates, including one Control Check: non-contaminated soil without Pyrene (denoted as CK) and four Experimental Treatments: contaminated soils with 1, 10, 100 and 500 mg Pyrene (denoted as PYR1, PYR10, PYR100 and PYR500, respectively). The contaminated soils in Experimental Treatments were conducted by following the procedure described by Brinch et al [43]. Briefly, the pyrene solution was mixed well with 25% air-dried soil in a glass container and then the container was placed in a fume hood overnight to let the acetone evaporate. Thereafter, the remainder of the soil (75%) was added to the spiked soil and blended thoroughly. The non-contaminated soil, in the meanwhile, was prepared using the same methods with pure acetone solution. During the whole incubation period, the soil CO₂ emission flux in all treatments were measured every day (Additional file: Fig. S1). When the CO₂ emission rate began to decline, suggesting inadequate carbon sources in soils, all the microcosms were harvested immediately after 35-days incubation and stored at -80°C for the following analyses. The soil samples were equally divided equally into two parts: one for phenanthrene quantification, and one for DNA extraction.

**Construction of another two latitudes soil microcosms**

To further confirm the results concluded in microcosms of Beijing, uncontaminated soils were collected from two other cities according to different latitudes to construct microcosms. One is Hangzhou (HZ, 30°26' N, 120°19' E) in the middle latitude of China, and the other is Guangzhou (GZ, 23°16' N, 113°23' E) in the low latitude. The HZ and GZ soils had a pH value of 6.25 and 5.46, respectively. The HZ and GZ soils was classified as Haplic Ferralsol soil based on the FAO/UNESCO soil classification system.
Two treatments were performed for each soil, including Control Check (denoted as HZ-CK, GZ-CK) and contaminated soils with 500 mg Pyrene (denoted as HZ-PYR500, GZ-PYR500). The whole incubation process and sampling were consistent with the microcosmos mentioned above.

**Pyrene extraction and quantification**

Residual pyrene concentrations in soils were detected to figure out the different rates of pyrene degradation in each treatment, which was an parameters to determine the reply of resident species to the exposure. Pyrene extractions were performed using the ASE 350 system (Dionex, Thermo Scientific, USA) by carefully loading freeze-dried soils (8 g) into sample cells containing pelletized diatomaceous earth (2 g). Sample cells were vertically hung in the tray slots in order and extracted twice with dichloromethane/acetone (1:1, v/v). Extract were concentrated to approximately 3 mL, purified via solid-phase extraction and then quantified by gas chromatography coupled to mass spectrometry (GC-MS, Shimadzu, Kyoto, Japan) with a HP-5MS capillary column (5% diphenyl-95% dimethyl polysiloxane, 30 m × 0.25 mm i.d., 0.25 μm film thickness, Agilent). The analysis procedure of GC-MS was slightly modified from the methods of Rombolà et al. [44]. High purity helium was used as a carrier gas with a flow rate of 1 ml/min. The GC oven procedure proceeded as follows: 50°C hold for 2 min, rise to 180°C at a rate of 20°C/min, increase at 10°C/min to 290°C, then hold for 10 min.

**DNA extraction and real-time quantitative PCR**

DNA was extracted from 0.5 g of freeze-dried soils using the FastDNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA, USA) according to the manufacturer's instructions. The purity and concentration of extracted DNA were determined via agarose gel electrophoresis and a Nano Drop ND-2000 spectrophotometer (Wilmington, Delaware, USA).

Bacterial (16S rRNA gene) [45], archaeal (16S rRNA gene) [46], fungal (18S rRNA gene) [47] and nidA gene [48] copies were amplified using Premix Ex Taq (Takara, Dalian, China) and the gene-specific primer pairs respectively. Each soil DNA was quantified in triplicate in 25 μl reaction system on a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Details about primers sequences, amplification protocols and reaction mixtures were summarized in Additional file: Table S2. The PCR efficiency \((E)\) was 95-105% and all \(R^2\) values for standard curves were over 0.99.

**High-throughput sequencing and analysis**

Bacteria-specific fragments about 460 bp covering the V3/V4 region of 16S rRNA genes were amplified from samples using the primer pairs 338 F (5’-ACT CCT ACG GGA GGC AGC AG-3’) / 806R (5’-GGA CTA CHV GGG TWT CTA AT-3’) [49] by 26 cycles and sequenced on the Illumina MiSeq PE 300 × 2 sequencer at Majorbio Bio-Pharm Technology Co. Ltd. (Shanghai, China).

High-throughput sequencing data analysis was performed through the Galaxy pipeline [50] at the Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences.
In short, each sequence was split by detected barcodes, following by combing forward sequences and reverse sequences. Quality filtered sequences were clustered using UPARSE in Operational Taxonomic Units (OTUs) at 97% similarity level to pick represent sequences and OTUs. A total of 1,374,419 quality-filtered paired-end sequences from samples, ranging from 39,134 to 51,214 sequences, were clustered into 20,641 bacterial OTUs at 97% identity. Then OTU representative sequences were annotated using the RDP classifier with the SILVA database. A resample OTU table at 26,971 reads per sample was created for further statistical analysis based on the minimum value of reads numbers for all samples.

Bioinformatic analysis

The Shannon diversity index (α-diversity) and Non-metric multidimensional scaling (NMDS) analysis based on Bray-Curtis dissimilarity were performed using the Galaxy pipeline based on the resample OTU table. The resistance index [51] was calculated by comparing the α-diversity between spiked-samples and controls in order to evaluate the response stability of the microbial community to pollutant changes. An average genome size was calculated using genomes at each phylum level deposited in IMG and 16S rRNA gene composition following the method of Sorensen et al. [34]. Spearman correlation was used to evaluate correlations between nidA gene copies and the OTU relative abundance using SPSS 25.0 (IBM, USA), following by a network analysis to explore the co-occurrence patterns at http://ieg2.ou.edu/MENA. The metagenomic function of the microbiome was predicted based on 16S rRNA amplicon data through PICRUSt [52]. The output of PICRUSt showed the amount of each functional-gene in each sample by a functional-gene-count matrix.

Statistics analysis

All statistical analyses were carried out in R platform (3.6.2, http://www.r-project.org). The Shapiro–Wilk test and Bartlett test were used to check whether the data conformed to normality and homoscedasticity respectively. One-way ANOVA was used to determine whether there were any statistically significant differences (P < 0.05) among treatments in Shannon index and copies. The post-hoc Tukey HSD test were used to check if the relationship between two sets of data was statistically significant.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials
Raw sequences data of 16S rRNA high-throughput sequencing are available in NCBI Sequence Read Archive repository database (www.ncbi.nlm.nih.gov/sra) under accession number SRP266388: PRJNA638003.

Competing interests

The authors declare that they have no competing interests

Funding

This research was supported by National Key R&D Program of China (Nos. 2019YFC1805803), the National Natural Science Foundation of China (Nos. 41907273, 31670507 and 91951108), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA23010400) and Science and Technology Service Network Initiative of the Chinese Academy of Sciences (Grant KFJ-STS-ZDTP-064).

Authors' contributions

SHW designed the research. XLZ and SHW supervised the project. YZD, HNF, XLL YJL ZHB and SJX refined the experimental method. YZD performed the research and conducted the data analyses with assistance from SHW. YZD drafted the manuscript; XLZ and SHW reviewed and edited the manuscript.

Acknowledgements

Not applicable

Author information

Yuzhu Dong and Shanghua Wu contributed equally to this work.

Affiliations

Key Laboratory of Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

Yuzhu Dong, Shanghua Wu, Haonan Fan, Xianglong Li, Yijing Li, Shengjun Xu, Zhihui Bai & Xuliang Zhuang

College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

Yuzhu Dong, Shanghua Wu, Haonan Fan, Xianglong Li, Yijing Li, Shengjun Xu, Zhihui Bai & Xuliang Zhuang

References
1. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, et al. Structure and function of the global topsoil microbiome. Nature. 2018;560(7717):233--.

2. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol. 2017;15(10):579–90.

3. Jansson JK, Hofmockel KS. The soil microbiome - from metagenomics to metaphenomics. Curr Opin Microbiol. 2018;43:162–8.

4. Chen J, McIlroy SE, Archana A, Baker DM, Panagiotou G. A pollution gradient contributes to the taxonomic, functional, and resistome diversity of microbial communities in marine sediments. Microbiome 2019, 7.

5. Ghosal D, Ghosh S, Dutta TK, Ahn Y. Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): a review. Frontiers in microbiology. 2016;7:1369.

6. Nam JJ, Sweetman AJ, Jones KC. Polynuclear aromatic hydrocarbons (PAHs) in global background soils. J Environ Monit. 2009;11(1):45–8.

7. Andrei A-Ş, Salcher MM, Mehrshad M, Rychecký P, Znachor P, Ghai R. Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. ISME J. 2019;13(4):1056–71.

8. Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M: Drift barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences 2012, 109(45):18488–18492.

9. Vila J, Tauler M, Grifoll M. Bacterial PAH degradation in marine and terrestrial habitats. Curr Opin Biotechnol. 2015;33:95–102.

10. de Gannes V, Hickey WJ: Genetic Adaptations of Bacteria for Metabolism of Polycyclic Aromatic Hydrocarbons. In: Microbial Ecotoxicology. Springer; 2017: 133–164.

11. Zhao Q, Yue S, Bilal M, Hu H, Wang W, Zhang X. Comparative genomic analysis of 26 Sphingomonas and Sphingobium strains: dissemination of bioremediation capabilities, biodegradation potential and horizontal gene transfer. Science of The Total Environment. 2017;609:1238–47.

12. Kweon O, Kim S-J, Blom J, Kim S-K, Kim B-S, Baek D-H, Park SI, Sutherland JB, Cerniglia CE. Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus Mycobacterium. BMC evolutionary biology. 2015;15(1):21.

13. Thomas F, Corre E, Cebron A. Stable isotope probing and metagenomics highlight the effect of plants on uncultured phenanthrene-degrading bacterial consortium in polluted soil. Isme Journal. 2019;13(7):1814–30.

14. Lu C, Hong Y, Liu J, Gao Y, Ma Z, Yang B, Ling W, Waigi MG. A PAH-degrading bacterial community enriched with contaminated agricultural soil and its utility for microbial bioremediation. Environ Pollut. 2019;251:773–82.

15. Sela I, Wolf YI, Koonin EV: Theory of prokaryotic genome evolution. Proceedings of the National Academy of Sciences 2016, 113(41):11399–11407.
16. Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic acids research. 2008;36(21):6688–719.

17. Bennett GM, Abbà S, Kube M, Marzachi C. Complete genome sequences of the obligate symbionts “Candidatus Sulcia muelleri” and “Ca. Nasuia deltocephalinicola” from the pestiferous leafhopper Macrosteles quadripunctulatus (Hemiptera: Cicadellidae). Genome Announc. 2016;4(1):e01604–15.

18. Garcia R, Gemperlein K, Müller R. Minicystis rosea gen. nov., sp. nov., a polyunsaturated fatty acid-rich and steroid-producing soil myxobacterium. Int J Syst Evol MicroBiol. 2014;64(11):3733–42.

19. Ranea JA, Grant A, Thornton JM, Orenco CA. Microeconomic principles explain an optimal genome size in bacteria. Trends Genet. 2005;21(1):21–5.

20. Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. Trends Genet. 2001;17(10):589–96.

21. Lynch M. Evolution of the mutation rate. TRENDS in Genetics. 2010;26(8):345–52.

22. Nowack EC, Price DC, Bhattacharya D, Singer A, Melkonian M, Grossman AR: Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of Paulinella chromatophora. Proceedings of the National Academy of Sciences 2016, 113(43):12214–12219.

23. Bock R. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. Annual review of genetics. 2017;51:1–22.

24. Moran NA. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 2002;108(5):583–6.

25. Toft C, Andersson SG. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 2010;11(7):465–75.

26. Noack S, Baumgart M. Communities of niche-optimized strains: small-genome organism consortia in bioproduction. Trends in biotechnology. 2019;37(2):126–39.

27. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, DeLong EF. Environmental drivers of a microbial genomic transition zone in the ocean's interior. Nature microbiology. 2017;2(10):1367–73.

28. Bentkowski P, Van Oosterhout C, Ashby B, Mock T. The effect of extrinsic mortality on genome size evolution in prokaryotes. ISME J. 2017;11(4):1011–8.

29. Scheuerl T, Hopkins M, Nowell RW, Rivett DW, Barraclough TG, Bell T. Bacterial adaptation is constrained in complex communities. Nature communications. 2020;11(1):1–8.

30. Brewer TE, Handley KM, Carini P, Gilbert JA, Fierer N: Genome reduction in an abundant and ubiquitous soil bacterium ‘Candidatus Udaeobacter copiosus’. Nature Microbiology 2016, 2(2):1–7.

31. Ni J, Xu M, He Z, Sun G, Guo J. Novel insight into evolutionary process from average genome size in marine bacterioplanktonic biota. Appl Ecol Environ Res. 2016;14(2):65–75.

32. Kussell E, Leibler S. Phenotypic diversity, population growth, and information in fluctuating environments. Science. 2005;309(5743):2075–8.
33. Lear G, Lau K, Perchec AM, Buckley HL, Case BS, Neale M, Fierer N, Leff JW, Handley KM, Lewis G. Following Rapoport's Rule: the geographic range and genome size of bacterial taxa decline at warmer latitudes. Environ Microbiol. 2017;19(8):3152–62.

34. Sorensen JW, Dunivin TK, Tobin TC, Shade A. Ecological selection for small microbial genomes along a temperate-to-thermal soil gradient. Nature microbiology. 2019;4(1):55–61.

35. Han K, Li Z-f, Peng R, Zhu L-p, Zhou T, Wang L-g, Li S-g, Zhang X-b, Hu W. Wu Z-h: Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. Scientific reports. 2013;3:2101.

36. Finan TM. The divided bacterial genome: structure, function, and evolution. Microbiol Mol Biol Rev. 2017;81(3):e00019-00017.

37. Chen B, He R, Yuan K, Chen E, Lin L, Chen X, Sha S, Zhong J, Lin L, Yang L. Polycyclic aromatic hydrocarbons (PAHs) enriching antibiotic resistance genes (ARGs) in the soils. Environ Pollut. 2017;220:1005–13.

38. Tian B, Hua Y. Carotenoid biosynthesis in extremophilic Deinococcus–Thermus bacteria. Trends in microbiology. 2010;18(11):512–20.

39. Tian R, Ning D, He Z, Zhang P, Spencer SJ, Gao S, Shi W, Wu L, Zhang Y, Yang Y. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. Microbiome. 2020;8:1–15.

40. Top EM, Springael D. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Curr Opin Biotechnol. 2003;14(3):262–9.

41. Springael D, Top EM. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. Trends in microbiology. 2004;12(2):53–8.

42. Ma Y, Wang L, Shao Z. Pseudomonas, the dominant polycyclic aromatic hydrocarbon-degrading bacteria isolated from Antarctic soils and the role of large plasmids in horizontal gene transfer. Environ Microbiol. 2006;8(3):455–65.

43. Brinch UC, Ekelund F, Jacobsen CS. Method for spiking soil samples with organic compounds. Appl Environ Microbiol. 2002;68(4):1808–16.

44. Rombolà AG, Meredith W, Snape CE, Baronti S, Genesio L, Vaccari FP, Miglietta F, Fabbri D. Fate of soil organic carbon and polycyclic aromatic hydrocarbons in a vineyard soil treated with biochar. Environ Sci Technol. 2015;49(18):11037–44.

45. Fierer N, Jackson JA, Vilgalys R, Jackson RB. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl Environ Microbiol. 2005;71(7):4117–20.

46. Lee S-H, Sorensen JW, Grady KL, Tobin TC, Shade A. Divergent extremes but convergent recovery of bacterial and archaeal soil communities to an ongoing subterranean coal mine fire. ISME J. 2017;11(6):1447–59.

47. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J. 2010;4(10):1340–51.
48. Ren G, Teng Y, Ren W, Dai S, Li Z. Pyrene dissipation potential varies with soil type and associated bacterial community changes. Soil Biol Biochem. 2016;103:71–85.

49. Lee CK, Barbier BA, Bottos EM, McDonald IR, Cary SC. The inter-valley soil comparative survey: the ecology of Dry Valley edaphic microbial communities. ISME J. 2012;6(5):1046–57.

50. Feng K, Zhang Z, Cai W, Liu W, Xu M, Yin H, Wang A, He Z, Deng Y. Biodiversity and species competition regulate the resilience of microbial biofilm community. Molecular ecology. 2017;26(21):6170–82.

51. Liang Y, Xiao X, Nuccio EE, Yuan M, Zhang N, Xue K, Cohan F, Zhou J, Sun B. Differentiation strategies of soil rare and abundant microbial taxa in response to changing climatic regimes. Environmental Microbiology 2020.

52. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature biotechnology. 2013;31(9):814.

Figures

**Figure 1**

Bacterial 16S copies (a), Shannon index (b) and Resistance index (c) under different pyrene concentrations. Each sample has 6 replicates, and the bar represents the standard deviation of the mean from the 6 replicates. Values assigned with the same letter were not significantly different by Tukey HSD post hoc (P≤0.05)
Figure 2

Changes of the microbial taxonomic structure and genome size under changing pyrene concentrations. (a) nMDS was calculated based on Bray-Curtis dissimilarity and 95% confidence ellipse was added for each treatment. Stress value for the ordination was 0.086. (b) Mean relative abundance of bacteria at the phylum level. (c) TOP 12 phyla related to pyrene concentration were identified by Random Forest and ranked in descending order of importance to the accuracy of the model. (d) Correlation between mean genome size and logarithm of pyrene concentration at base E (Pearson's correlation $P = 0.5164$). 0.01 was used to instead the original pyrene concentration in the control treatment.
Figure 3

Comparison of potential community functions among treatments based on PICRUSt. (a) Heatmap of KEGG level 2 modules in all five treatments; (b) Correlation between twelve KEGG modules relative to carbon metabolic and nidA genes, and twelve KEGG modules and genome size. Color points indicate correlation coefficient: the bluer shows the stronger positive correlation, and the redder shows the stronger negative correlation. Asterisks indicate significance: * P<0.05.
Figure 4

Network of the bacterial communities at OUT level in soils. The network of (a) Treatment PYR1 (1 mg pyrene/kg dry soil), (b) Treatment PYR10 (10 mg pyrene/kg dry soil), (c) Treatment PYR100 (100 mg pyrene/kg dry soil) and (d) Treatment PYR500 (500 mg pyrene/kg dry soil) were visualized by Gephi. The red nodes represent OTUs significantly positively correlated with nida gene and the green nodes represent OTUs which are positively connected to the red nodes. The size of each node is proportional to its number of connections.
Figure 5

Changes in average genome size between Control and Treatment PYR500 in HZ and GZ. Asterisks indicate significance: * P-value < 0.05, ** P-value < 0.01 based on Tukey's HSD test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.tif
- supplement2.tif
- supplement3.tif
- supplement5.tif
- supplement6.docx