Data in Brief 46 (2023) 108860

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Data on the temporal changes in soil properties and microbiome composition after a jet-fuel contamination during the pot and field experiments

Ivan N. Semenkov[a,†,*], Pavel V. Shelyakin[b,c,†], Daria D. Nikolaeva[d,†], Maria N. Tutukina[b,d,†], Anna V. Sharapova[a,†], Sergey A. Lednev[a], Yuliya V. Sarana[d], Mikhail S. Gelfand[d,b,‡], Pavel P. Krechetov[a,‡], Tatiana V. Koroleva[a,‡]

[a] *M.V. Lomonosov Moscow State University, 119991 Moscow, Russia*
[b] *Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, 127051 Moscow, Russia*
[c] *Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia*
[d] *Skolkovo Institute of Science and Technology, 121205 Moscow, Russia*

ARTICLE INFO

Article history:
Received 2 August 2022
Revised 20 December 2022
Accepted 21 December 2022
Available online 27 December 2022

ABSTRACT

The soil response to a jet-fuel contamination is uncertain. In this article, original data on the influence of a jet-fuel spillage on the topsoil properties are presented. The data set is obtained during a one-year long pot and field experiments with Dystric Arenosols, Fibric Histosols and Albic Luvisols. Kerosene loads were 1, 5, 10, 25 and 100 g/kg. The data set includes information about temporal changes in kerosene

Abbreviations: AL, Albic Luvisols; ASV, amplicon sequence variant; CA, cellulolytic activity; CEC, cation exchange capacity; DA, Dystric Arenosols; DNA, deoxyribonucleic acid; EDTA, Ethylenediaminetetraacetic acid; FH, Fibric Histosols; Kav, available potassium; NH_4^+, exchangeable ammonium; NO_3^-, water-soluble nitrate; Pav, available phosphorus; PCR, polymerase chain reaction; qPCR, real-time polymerase chain reaction; rRNA, ribosomal ribonucleic acid; SOM, soil organic matter; WMO, World Meteorological Organization.

* Correspondence.
E-mail address: semenkov@geogr.msu.ru (I.N. Semenkov).
† These authors share the first authorship.
‡ These authors share the last authorship.

https://doi.org/10.1016/j.dib.2022.108860

Keywords:
Soil metagenome
Soil pollution
Ecological indicators
Gasoline
Total petroleum hydrocarbons
Bearing capacity
Xenobiotic compounds

concentration; physicochemical properties, such as pH, moisture, cation exchange capacity, content of soil organic matter, available P and K, exchangeable NH_4^+, and water-soluble NO_3^-; and biological properties, such as biological consumption of oxygen, and cellulolytic activity. Also, we provide sequencing data on variable regions of 16S ribosomal RNA of microbial communities from the respective soil samples.

Specifications Table

Subject	Environmental Science Environmental Genomics and Metagenomics Environmental Chemistry Health, Toxicology and Mutagenesis Pollution
Specific subject area	Soil metagenome, biodiversity, microbiome, soil pollution, one health, soil properties, metagenomics
Type of data	Figures, Tables, and Fastq files
How the data were acquired	A-horizon soil samples contaminated with ~~a~~ jet-fuel were collected from Albic Luvisols, Dystric Arenosols and Fibric Histosols during the pot and field experiments. In topsoil (0 – 10 cm) samples, physicochemical properties (pH, moisture, cation exchange capacity, content of kerosene soil organic matter, available P and K, exchangeable NH_4^+, and water-soluble NO_3^-) and biological properties (biological consumption of oxygen, cellulolytic activity) were controlled in 3, 90, 180 and 360 days after a jet-fuel treatment. Paired-end sequencing of metagenomic DNA isolated from the topsoil samples was performed on the Illumina MiSeq platform. Bioinformatics analysis was carried out using the following programs and databases: FastQC, DADA2, decontam, phyloseq, IdTaxa, SILVA, AlignSeqs, FastTree, Picrust2
Data format	Raw Filtered
Description of data collection	The A-horizon topsoil samples were collected from Albic Luvisols (Kaluga region, Russia) and Dystric Arenosols (cosmodrome Baikonur, Kazakhstan) to conduct a pot experiment. Samples without kerosene were used as a control. The remaining samples were treated with the jet-fuel loads of 1, 5, 10, 25, and 100 g/kg. In the field experiment (Kaluga region, Russia), Fibric Histosols and Albic Luvisols were contaminated with the same jet-fuel loads. Then, soil samples were collected for analysis in 3, 90, 180 and 360 days after treatment. Total DNA from the topsoil samples was isolated using DNeasy PowerSoil kit (Qiagen, Germany). Variable V3V4 and V4V5 16S rRNA regions were amplified using the Phusion polymerase (New England Biolabs, USA). Index PCR was made using the Phusion polymerase and the Nextera XT Index kit. The libraries were sequenced on Illumina MiSeq with the read length of 250 bp (MiSeq Reagent Kit v2). The average number of reads was 75,000 per sample.
Data source location	Albic Luvisol topsoil samples: Institution: M.V. Lomonosov Moscow State University Region: Kaluga region Country: Russia Latitude and longitude for collected samples: N 55°11′05″ E 36°25′05″ Dystric Arenosol topsoil samples: Institution: Cosmodrome Baikonur City: Baikonur Country: Kazakhstan, Russia Latitude and longitude for collected samples: N 45°43′20″ E 63°11′40″

(*continued on next page*)

	Fibric Histosol topsoil samples: Institution: M.V. Lomonosov Moscow State University Region: Kaluga region Country: Russia Latitude and longitude for collected samples: N 55°11′03″ E 36°24′58″
Data accessibility	Data on the kerosene content, soil physicochemical properties (pH, moisture, cation exchange capacity (CEC), content of soil organic matter (SOM), available phosphorus (Pav), potassium (Kav), exchangeable ammonium (NH_4^+), and water-soluble nitrate (NO_3^-)), as well as on biological consumption of oxygen and cellulolytic activity of soil are deposited at figshare (https://doi.org/10.6084/m9.figshare.19609503.v1). Sequencing data have been deposited in the Sequence Read Archive with Bioproject accession PRJNA786393 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA786393). Decontaminated and singleton filtered amplicon sequence variant (ASV) tables with assigned taxonomy and phylogenetic tree are deposited at figshare (https://doi.org/10.6084/m9.figshare.19609503.v1) in phyloseq format. Community functional profiles predicted with Picrust2 are deposited at figshare.
Related research article	P.V. Shelyakin, I.N. Semenkov, M.V. Tutukina, D.D. Nikolaeva, A.V. Sharapova, Yu.V. Sarana, S.A. Lednev, A.D. Smolenkov, M.S.Gelfand, P.P. Krechetov, T.V. Koroleva, The influence of kerosene on microbiomes of diverse soils, Life. 12 (2022) 221. https://doi.org/10.3390/life12020221

Value of the Data

The presented dataset is the first report of the influence of a jet-fuel contamination on the soil microbiome composition. These data characterize the responses of soil microbiomes to kerosene contamination under humid and semi-arid climates.

These data will help to mitigate the negative consequences of jet-fuel leakages. Soil scientists, policy makers, government official and stakeholders can benefit from these data.

From this data set, it may be possible to find out potentially beneficial soil bacteria able to metabolize hydrocarbons and to survive the jet-fuel pollution.

1. Data Description

The dataset described here contains soil properties and amplicon sequencing data of soil bacterial community.

1.1. Regional setting

We investigated 3 contrast soils. During a laboratory pot experiment, the Dystric Arenosols (N 45°43′20″ E 63°11′40″) sampled in Kazakhstan and Albic Luvisols (N 55°11′5″ E 36°25′5″) of Russia were treated. And a field experiment was conducted with Fibric Histosols (N 55°11′03″ E 36°24′58″) and the same Albic Luvisols in Russia (N 55°11′5″ E 36°25′5″; Fig. 1).

The soils of the Kaluga region could be considered as the background for the Moscow region [1] where five international airports located including the Sheremetyevo Alexander S. Pushkin International Airport that is the second-busiest airport in Europe. Dystric Arenosols are the most vulnerable soils in the Baikonur Cosmodrome area, where 'Soyuz' vehicles running on kerosene are launched, and the airport named Krainiy is located [2,3].

Dystric Arenosols were characterized by an alkaline environment, low content of available phosphorus and soil organic matter (Fig. 2). Fibric Histosols were highly acidic, rich in SOM and available phosphorus. Albic Luvisols occupied an intermediate position in terms of physicochemical properties.

Fig. 1. Location of soil sampled: 1 – Albic Luvisols and Fibric Histosols of the Kaluga region, 2 – Dystric Arenosols of the Baikonur Cosmodrome area.

1.2. Description for a dataset characterized soil properties

Data on soil physicochemical properties, as well as on biological consumption of oxygen and cellulolytic activity are deposited at figshare (https://doi.org/10.6084/m9.figshare.19609503.v1). File 'Soil_physicochemical_properties.tsv' contains 17 columns.

The first three columns ('soil_experiment', 'soil', 'experiment') provides information about:

-soils studied (Dystric Arenosols, Albic Luvisols and Fibric Histosols) and
-the variant of the experiments (pot or field).

The forth column named 'sample_id' is a primary (unique) key coding data on:

-soils treated (DA, Dystric Arenosols; AL, Albic Luvisols and FH, Fibric Histosols),
-kerosene loads in g/kg,
-a day after kerosene treatment and
-the variant of the experiments (p, pot or f, field).

The concentration in the target sample is presented in the column entitled 'Kerosene, g/kg'. A day after contamination can be found in the last column. We controlled:

-a soil pH value in a water solution,
-soil moisture content (in %),
-cation exchange capacity (CEC, in mM(+)/100 g), and
-content of soil organic matter (SOM, in%),
-content of available phosphorus (Pav, in mg/kg),
-content of available potassium (Kav, in mg/kg),
-content of exchangeable ammonium (NH_4^+, in mg/kg),
-content of water-soluble nitrate (NO_3^-, in mg/kg),
-biological consumption of oxygen (in mM O_2/100 g),
-cellulolytic activity (in mg/g of readily hydrolysable organic matter).

Other columns in a 'Soil_physicochemical_properties.tsv' file provides information about soil physicochemical, chemical and biological properties studied.

1.3. Soil microbiome characteristics

Raw sequences are deposited in the Sequence Read Archive with Bioproject accession PRJNA786393. The 16S rRNA gene amplicon sequencing was performed using the Illumina MiSeq platform (2 × 250-bp paired ends). A total of 22,002,216 reads were classified out of 52,504,639 initially obtained (Table 1).

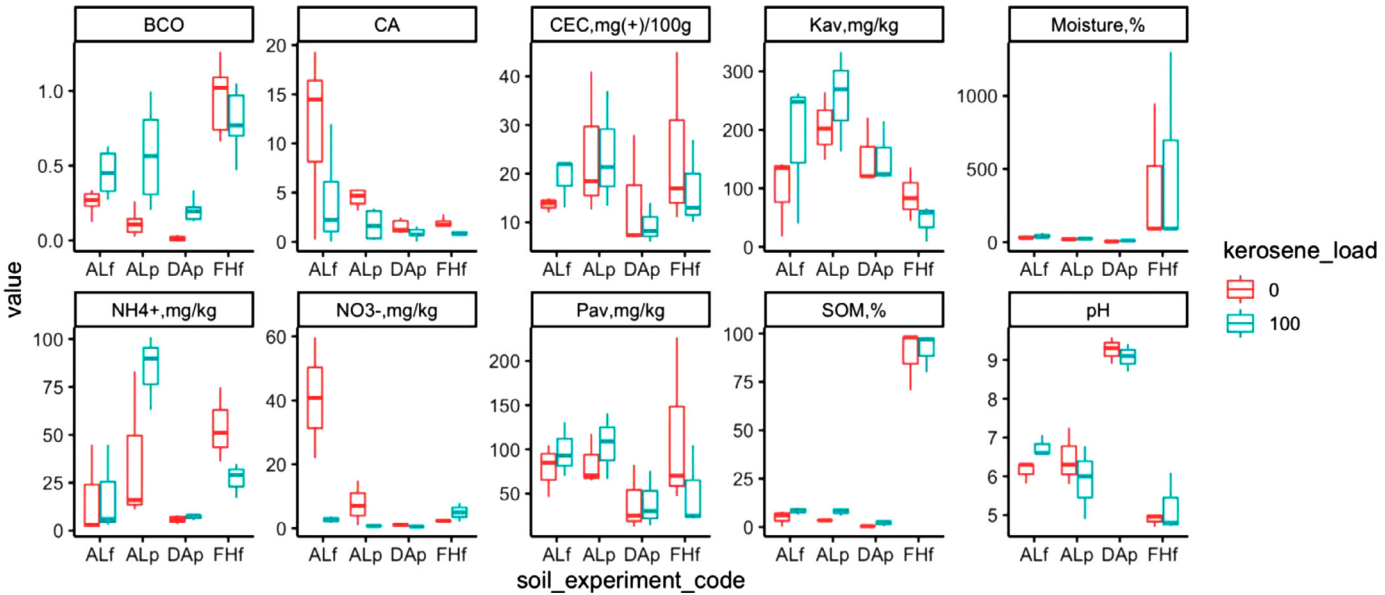

Fig. 2. Soil physicochemical properties (pH, moisture, CEC, content of SOM, Pav, Kav, NH_4^+, and NO_3^-), as well as on biological consumption of oxygen (mM O_2/100 g) and cCA (linen cloth size [cm]; weight [g]). On x axis: ALf – Albic Luvisols (field experiment), Alp – Albic Luvisols (pot experiment), DAp – Dystric Arenosols (pot experiment), Fhf – Fibric Histosols (field experiment).

Table 1
Summary statics table. Data for all samples except negative controls are shown.

Reads	Count	
	V3V4	V4V5
Total analyzed reads	22,891,629	29,613,010
Reads after quality filtration, decontamination and filtration of singleton ASV	10,561,330	12,988,144
Filtered reads classified as bacterial	9768,701	12,233,515

Phyloseq files entitled 'kerosene_ASV_V3V4' and 'kerosene_ASV_V4V5' contain meta table including data on:

-the kerosene load (0, 1, 5, 10, 25 or 100 g/kg),
-the sample name (the primary key),
-abbreviation of the soil name (DA, AL, FH),
-number of biological replica (1 – 3),
-a day after kerosene treatment (3, 30, 90, 180, 360),
-the experiment variant (pot or field),
-the 16S rRNA region (V3V4 or V4V5) sequenced,
-a run batch number (1 – 4).

Moreover, filtered ASVs and assigned taxonomic labels to ASVs, as well as table of OTUs and a phylogenetic tree can be found in these files, too. Files 'kerosene_V3V4_pathways.tsv' and 'kerosene_V4V5_pathways.tsv' contain information about pathways predicted based on the data on the V3V4 and V4V5 regions of 16S rRNA sequenced, respectively. All these data were presented as taxonomic and functional profiles, as shown in Fig. 3 and Fig. 4, respectively.

In unpolluted samples, dominant bacterial phyla were Proteobacteria, Actinobacteriota, Acidobacteriota and Verrumicobiota. Pollution with kerosene lead to a pronounced expansion of Proteobacteria in all soils (Fig. 3).

Among more than 170 bacterial families present in unpolluted Dystric Arenosols, Beijerinckiaceae, Phormidiaceae, Rubrobacteriaceae, and Sphingomonadaceae were the most abundant. In the most contaminated (a load of 100 g/kg) samples of Dystric Arenosols, Caulobacteraceae, Nocardiaceae, and Sphingomonadaceae were widespread (Fig. 3).

From more than 180 bacterial families present in unpolluted Albic Luvisols sampled during the field experiment, Chthoniobacteraceae, Xanthobacteraceae, unclassified Acidobacteriales and subgroup 2 of Acidobacteriota were the most abundant. After contamination of samples with high kerosene load Burkholderiaceae, Mycobacteriaceae and Yersiniaceae became prevalent (Fig. 3).

Among more than 160 bacterial families present in unpolluted Albic Luvisols sampled during the pot experiment, Bryobacteraceae, Chthoniobacteraceae, Flavobacteriaceae, Nitrosomonadaceae, and Solibacteraceae prevailed together with Chthoniobacteraceae, Xanthobacteraceae and unclassified subgroup 2 of Acidobacteriota found during the field experiment with the same soil, too. In the most contaminated samples of Albic Luvisols (the pot experiment), the most abundant families were Caulobacteraceae, Chthoniobacteraceae, Moraxellaceae, Pseudomonadaceae, and Rhodocyclaceae (Fig. 3).

Among more than 130 bacterial families present in unpolluted Fibric Histosols, Rubrobacteriaceae was the most dominant, followed by Bacillaceae, Micromonosporaceae, Phormidiaceae, Pirellulaceae, Pseudonocardiaceae, Pyrinomonadaceae, and Sphingomonadaceae. In the most contaminated samples dominant family was Sphingomonadaceae, followed by Alcaligenaceae, Azospirillaceae, Caulobacteraceae, Pseudomonadaceae, and (Pseudo)nocardiaceae (Fig. 3).

In all soils prior to contamination and after it, more than 70% of all determined pathways were related to biosynthesis, followed by degradation and generation of precursor metabolites, and energy metabolism. After contamination, relative abundance of degradation pathways slightly increased due to increase in number of aromatic compound degradation pathways (Fig. 4).

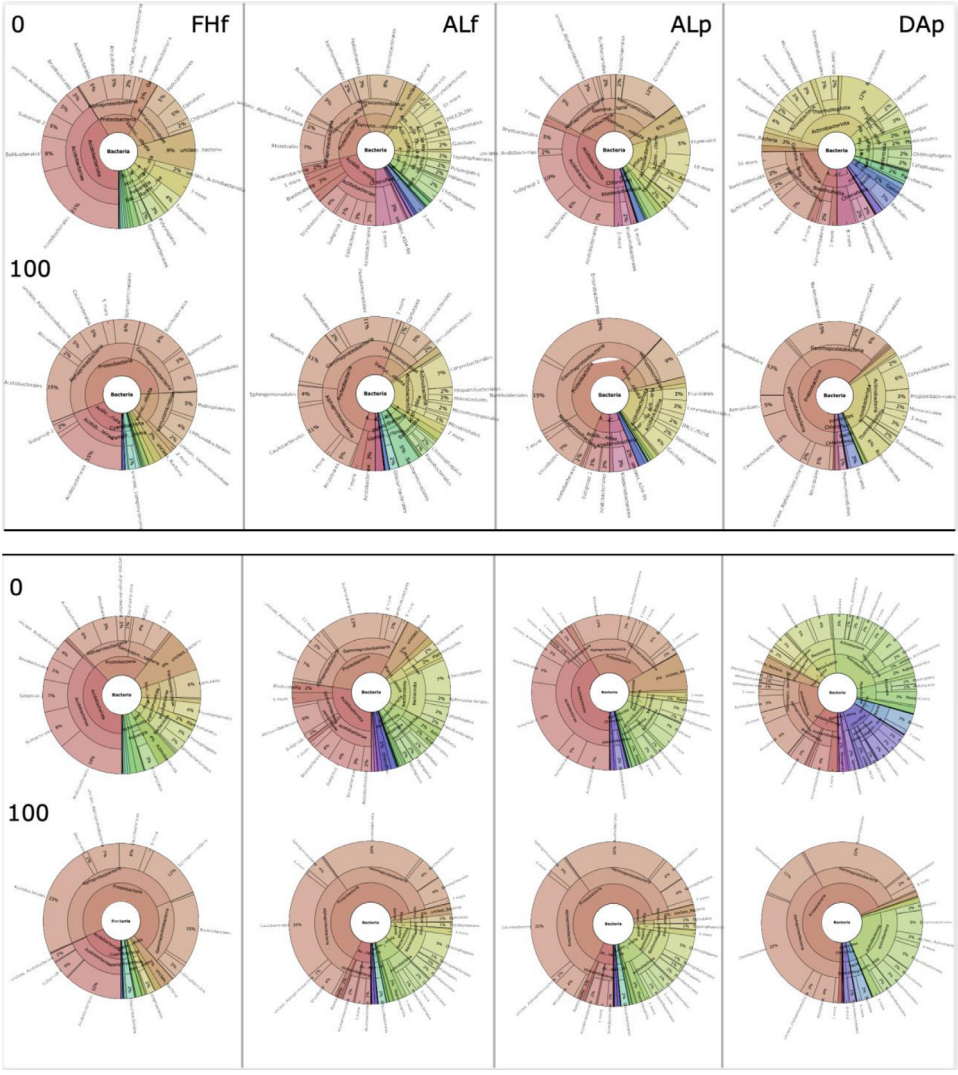

Fig. 3. Taxonomic profile based on the 16S rRNA gene amplicon sequencing of the soil samples collected during the pot and field experiments. The upper two row contain data on the V3V4 region, the lower two row – data on the V4V5 region. ALf – Albic Luvisols (field experiment), Alp – Albic Luvisols (pot experiment), DAp – Dystric Arenosols (pot experiment), Fhf – Fibric Histosols (field experiment). 0 and 100 mark rows that show taxonomic composition of control samples (0 g / kg) and highly contaminated samples (100 g kerosene / kg), respectively. Data on the samples collected on day 90 and day 180 are presented. Two samples of each soil/experiment combination were taken.

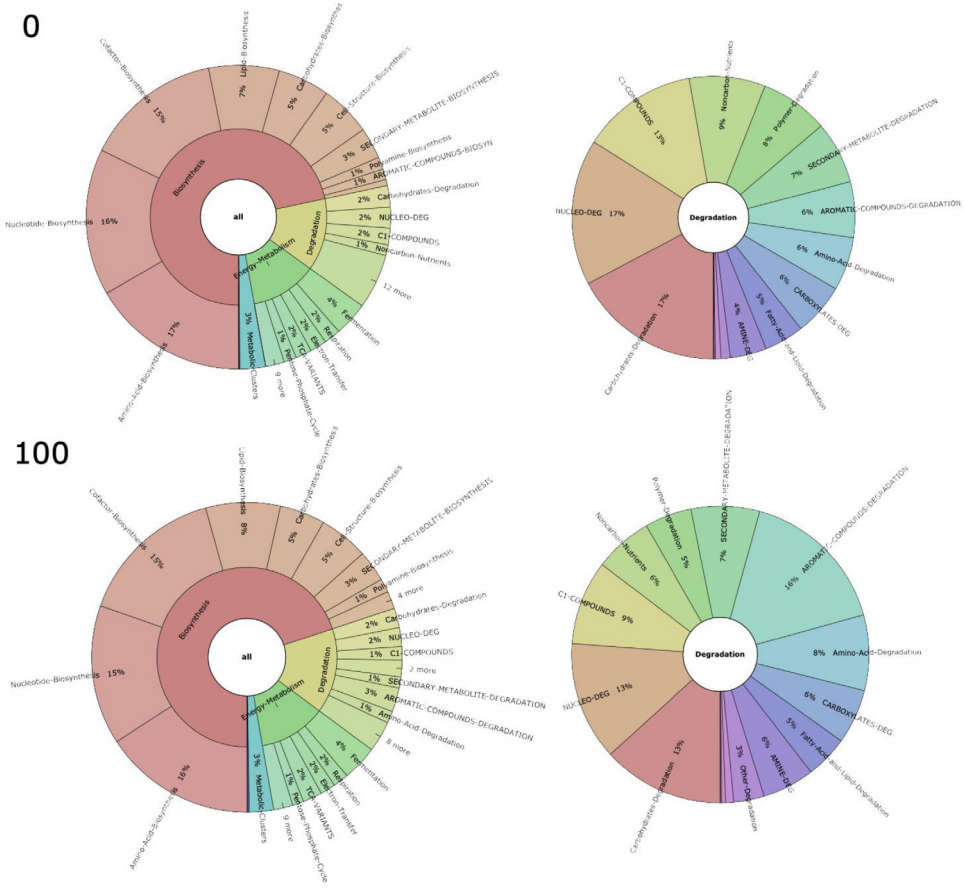

Fig. 4. Functional profile based on the V3V4 region of 16S rRNA gene amplicon sequencing of the soil samples collected during the pot and field experiments. 0 and 100 mark rows that show taxonomic composition of control samples (0 g / kg) and highly contaminated samples (100 g kerosene / kg), respectively. Data on the samples collected on day 90 and day 180 are presented. Two samples of each soil/experiment combination were taken.

2. Experimental Design, Materials and Methods

We conducted two experiments (pot and field) considering the response of soil physicochemical and biological proxies on a jet-fuel contamination (Table 2).

2.1. Sampling design and collection

The pot and field experiments were conducted from October 2019 to October 2020 and from June 2020 to June 2021, respectively.

2.1.1. Pot experiment

Two soil samples (20 kg, natural moisture) of the A-horizons of Dystric Arenosols and Albic Luvisols were sieved using sieves with a mesh of 3 mm, cleaned of roots and other coarse fraction, dried to the air-dry state in the lab. Then, they were moistened by distilled water added in small portions to a level of 60% of the maximum field capacity using a dispenser [4]. Addition of water was estimated by gravimetrically. The soil was thoroughly mixed after each water application for uniform water absorption. The soil moisture before a jet-fuel contamination, determined gravimetrically in samples drying at 105 °C, was 22 – 23% and 5% for Albic Luvisols and Dystric Arenosols, respectively (Table 3).

Table 2
Experimental design.

Parameters		Pot	Field	Total number of samples analyzed
Dystric Arenosols		Yes	No	–
Albic Luvisols		Yes	Yes	–
Fibric Histosols		No	Yes	–
Kerosene loads, g/kg		0, 1, 5, 10, 25, 100		–
Sampling day after a jet-fuel treatment		3, 10[a], 30 [a], 90 [b], 180 [b], 360 [b]		–
Replicas of soil proxies controlled	Biological consumption of oxygen	3	3	348
	Cellulolytic activity (linen cloth size [cm]; weight [g])	3 (6 × 4; 0.7 ± 0.1)	3–5 (10 × 20; 4–6)	234
	V3V4 and V4V5 regions of 16S rRNA gene	3	3	1152
	Kerosene	3	3	288
	pH	1	1	96
	Moisture	1	1	96
	SOM	1	1	96
	Kav	1	1	96
	Pav	1	1	96
	NO_3^-	1	1	96
	CEC	1	1	96
	NH_4^+	1	1	96

[a] just for control of the biological consumption of oxygen during the pot experiment.
[b] periods for control of cellulolytic activity were 0 – 90, 90 – 180 and 180 – 360 days after a jet-fuel treatment.

Table 3
Soil moisture before pretreatment with a jet-fuel during the pot experiment.

Sample	Dystric Arenosols	Albic Luvisols
1	5.18	22.98
2	4.98	23.05
3	5.48	22.71

Fig. 5. Soil subsamples in containers and in fruit jars. Upper row, remains of initial soil samples with slightly disturbed (natural) structure. Two rows of jars are for taking aliquots for soil analysis. The two lower rows of jars (with a yellow net) are for putting test objects (linen) to study a cellulolytic activity.

To activate the soil microbiota, the samples were placed in plastic boxes for three days and were carefully and periodically stirred to homogenize while preserving soil micro aggregates [18]. During this period, the temperature was 18–22°C.

After 3 days (time moment 0), both soil samples were divided into six subsamples. One subsample was used as a control (uncontaminated soil subsample). The remaining five subsamples were treated with various loads (1, 5, 10, 25, and 100 g/kg of soil) of a jet-fuel, i.e. TS-1 kerosene, which is the fuel most commonly used for commercial aviation in Russia [2]. Low jet-fuel loads (1, 5, and 10 g/kg) were applied as a spray from a portable device with periodic stirring. High loads (25 and 100 g/kg) were applied from a watering can, again with periodic stirring. The loads were selected based on previous researches on the response of vegetation [5] and cultivated soil microorganisms [6] to a jet-fuel contamination.

All subsamples (12 in duplicate) were placed in 400 cm^3 fruit jars (24 in total: one group of subsamples for collecting of aliquots for soil analyses mentioned earlier, and other group for incubation of linen fragments; Figs. 5 and 6) with iron lids to the bulk density of 1.47±0.04 and 0.92±0.09 kg/dm^3 for Dystric Arenosols and Albic Luvisols, respectively. These levels are typical of natural soils.

The experiment lasted one year in 2019–2020 at a temperature of 18–22 °C (daily monitoring of the air temperature with a contact thermometer). Every 5 days, the containers were opened for ventilation and, if needed, moisturizing. Distilled water from a dispenser was added to samples with decreased weight.

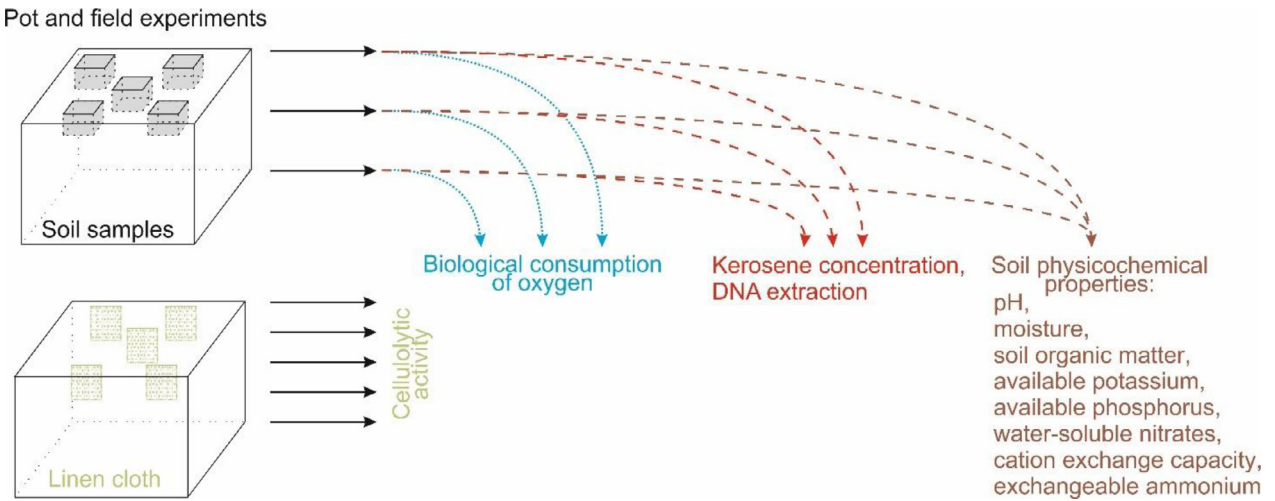

Fig. 6. Sampling design for different proxies controlled during the pot and field experiments. The number of lines corresponds to the number of replicas.

Table 4
Dominant species of vegetation at study sites (projective cover specified in parentheses).

Layer		Spruce–aspen forest at Albic Luvisols	Subshrubs–sphagnum pine forest at Fibric Histosols
A	Tree	*Picea abies* (L.) H. Karst. (40%) *Populus tremula* L. (20%) *Sorbus aucuparia* L. (<1%)	*Pinus sylvestris* L. (20%) *Betula pubescens* Ehrh. (3%)
B	Shrub	*Corylus avellana* L. (20%) *Lonicera xylosteum* L. (3%)	*Frangula alnus* Mill. (<1%)
C	Herb	*Ajuga reptans* L. (25%) *Athyrium filix-femina* (L.) Roth. (2%) *Lysimachia nummularia* L. (1%)	*Vaccinium oxycoccos* L. (30%) *Rhododendron tomentosum* Harmaja (20%) *Chamaedaphne calyculata* (L.) Moench (10%) *Eriophorum vaginatum* L. (10%)
D	Moss	*Atrichum undulatum* (Hedw.) P. Beauv. (15%) *Rhythidiadelphus triquertus* (Hedw.) Warnst. (1%) *Plagionmnium undulatum* (Hedw.) T.J.Kop. (1%)	*Sphagnum* spp. (90%): *S. angustifolium* (C.E.O.Jensen ex Russow) C.E.O.Jensen *S. squarrosum* Crome *S. fallax* (H.Klinggr.) H.Klinggr.

Vascular plant spaces are named according to https://powo.science.kew.org. Mosses are named according to http://arctoa.ru/en/Flora-en/general-en.php

2.1.2. Field experiment

Field experiment was performed in the Kaluga region (Fig. 1) in 2020–2021 on Albic Luvisols under a spruce–aspen forest and Fibric Histosols under a subshrubs–sphagnum pine forest (Table 4).

Experimental plots of 50 × 50 cm in size were marked as in our previous work in the Russian Far East [7]. The plots were selected taking into account the microtopography, comprising horizontal homogeneous microsites without visible microslopes of the soil surface. The surface of each plot was cleared of plant litter to reduce possible redistribution of a jet-fuel over the soil surface and for better absorption into the soil. Plots were contaminated with the same jet-fuel as in the pot experiment. Loads of a jet-fuel for the 0–10 cm topsoil layer were 1, 5, 10, 25, and 100 g/kg, as in the pot experiment.

The low jet-fuel loads of 1 and 5 g/kg were applied as spray from a portable device. The medium and high loads (10, 25, and 100 g/kg) were applied from a watering can. We tried to distribute a jet-fuel evenly over the soil surface within the plots.

The experiment lasted one year in 2020–2021 under the natural conditions (Table 5). During the field experiments, soil samples were collected in summer (3 and 360 days after treatment), autumn (90 days), and early winter (180 days) to take into account seasonal variability of the microbiome composition. At the same days, linen fragments were withdrawn from the soil and replaced with a new batch according to techniques described in details in [8].

2.2. Soil sampling and chemical analysis

In total, 288 topsoil samples (aliquots) of 50 g were sampled from 400 cm^3 glass jars and the 50 × 50 cm plots at a distance of at least 10 cm from the edge of the plot (Fig. 6). For the chemical analysis, each of the triplicates was placed into a glass jar with metal lids and stirred thoroughly to homogenize. Then, to isolate the total DNA, subsamples were taken.

2.2.1. Chemical analysis

In 288 soil samples, kerosene concentration was determined using the method thoroughly described in [4,9] and briefly in Table 6. In 96 soil samples (one mixed sample of the triplicate replicas, Fig. 5), chemical analyzes were performed immediately after soil sampling using routine standardized techniques as described in Table 6 for pH, moisture, CEC, content of SOM, Pav, Kav, NH_4^+, NO_3^-.

Table 5
Weather conditions at the weather station Maloyaroslavets (WMO ID=27,606), which is the nearest to the place of the field experiment.

Year	Month	T, °C	Tn, °C	Tx, °C	P, mm	U,%	FF, m/s	RRR, mm	sss, cm
2020	1	−0,5	−8,1	3,9	760	87	2	91	4
	2	−0,7	−14,9	8,0	755	82	3	56	5
	3	3,5	−10,5	17,7	762	66	3	79	2
	4	4,6	−11,3	16,8	758	60	3	40	4
	5	10,9	−0,2	25,1	759	69	2	315	0
	6	18,4	7,2	30,4	761	73	2	368	0
	7	18,0	7,2	30,0	759	76	2	211	0
	8	16,5	6,2	30,4	761	76	1	83	0
	9	13,2	2,4	25,0	763	72	2	128	0
	10	9,1	−3,3	20,5	764	73	2	69	0
	11	1,6	−9,2	11,2	768	87	2	104	5
	12	−4,5	−15,3	1,6	771	87	2	65	6
2021	1	−6,2	−27,3	2,8	760	89	2	132	30
	2	−11,0	−28,1	6,8	763	82	2	145	57
	3	−2,2	−26,3	11,5	761	74	2	62	32
	4	6,7	−3,2	22,4	760	66	2	76	7
	5	13,2	−1,0	29,9	759	66	2	198	0
	6	19,0	1,6	33,6	762	70	2	127	0
	7	20,7	8,8	31,9	760	70	1	94	0
	8	18,7	7,1	31,1	760	76	2	73	0
	9	9,3	−3,0	23,3	762	83	2	227	0
	10	5,3	−4,4	14,9	768	78	2	61	0
	11	1,8	−9,4	11,8	760	85	3	102	0
	12	−7,0	−22,7	1,5	760	88	2	107	19

T, mean air temperature at 2 m height above the earth's surface (AES). P, atmospheric pressure reduced to mean sea level. U, relative air humidity at a height of 2 m AES. FF, mean wind speed at a height of 10–12 m AES over the 10-minute period immediately preceding the observation. Tn, minimum air temperature. Tx, maximum air temperature. RRR, amount of precipitation. sss, snow depth. Based on the data available at https://rp5.ru/Weather_archive_in_Maloyaroslavets (date of access 20.02.2022).

Apart from physicochemical soil properties and kerosene concentration soil biological activity was controlled based on the biological consumption of oxygen and cellulolytic activity. Biological consumption of oxygen is determined on the basis of the method reported in [8]. This technique has been tested in coal mining areas and is based on the determination of changes in the oxygen content in the soil suspension. Soil samples of 1–5 g were placed in a container with a hermetically screwed lid, after which the containers were completely added with distilled water (50–53 ml) and hermetically sealed. For homogenization, the suspension was stirred on a rotator for 1 hour, after which it was placed in a dark place for 5 days. Simultaneously with the test samples, blank samples (a distilled water) were incubated. In five days, the content of dissolved oxygen in the suspension was measured using a Clark electrode (DKTP-02.3, Russia) and an oximeter. Biological consumption of oxygen was calculated in mM O_2/100 g as the difference between the oxygen content in the analyzed and blank samples.

2.2.2. DNA extraction, library preparation, and sequencing

Three independent samples were collected from each pot and plot, and either put immediately into a PowerBead tube or in a sterile 1.5 ml eppendorf tube. Eppendorf tubes filled in during the field experiment were put into the thermos and delivered to the lab for 3 – 4 h. Then, tubes were frozen and stored at –20 °C prior to DNA extraction. Total DNA from the topsoil subsamples of 1–2 g was isolated (Table 7).

Then, soil sample of 300 mg was put into a PowerBead tube and 60 μl of the C1 buffer were added. The tube was inverted 4–5 times to mix the reagents. The samples were then disrupted and DNA was purified. Its concentration (6.6 ± 4.4 ng/μl; Table 8) was measured on the Qubit 1 fluorimeter.

Table 6
Methods of soil chemical analyses used.

Parameters	Methods and equipment	Reference
Kerosene	Extraction with anhydrous sodium sulfate and methylene chloride for 15 min in an ultrasonic bath. Filtration. Gas chromatography: the Agilent 7890 V gas chromatograph by Agilent Technologies (the USA) equipped with the 5977 A quadrupole mass-spectrometric detector	[4,9]
Soil organic carbon	Wet dichromate oxidation and spectrocolorimetery	[10]
pH	Soil: water ratio 1:5. Potentiometry: pH-meter 'I-160MI' (Izmeritelnaya technika, Russia)	[10]
NO_3^-	Soil: water ratio 1:5. Photometry (salicylic acid; wavelength 410 nm)	[10]
NH_4^+	Extraction with 1 M KCl. Photometry (indophenol; wavelength 655 nm)	[10]
Pav	Extraction with 0.2 M HCl (Histosols and Luvisols) and 0.2 M $(NH_4)_2CO_3$ (soil: solution ratio 1:5 and 1:50 for A-horizons and O-horizons, respectively), spectrophotometry (wavelength 710 nm)	[10–12]
Kav	Extraction with 0.2 M HCl (Histosols and Luvisols) and 0.2 M $(NH_4)_2CO_3$ (soil: solution ratio 1:5 and 1:50 for A-horizons and O-horizons, respectively), spectrophotometry. The extracted potassium was measured by the inductively coupled plasma atomic emission spectrometry. Agilent 720 ICP-OES (Agilent Technologies, Malaysia).	[10–12]
CEC	Magnesium acetate method with complexometric determination of Mg with EDTA. Saturation: Saturation with 0.25 M $Mg(CH_3COO)_2$ (pH 7.0); agitation 30 min; decantation 5 min; filtration; Saturation with 0.5 M $Mg(CH_3COO)_2$ (pH 7.0); agitation 15 min; decantation 5 min; filtration; Saturation with 0.25 M $Mg(CH_3COO)_2$ (pH 7.0); shaking; filtration; Twice washing-out by distilled water Displacement: 0.5 mol M KCl.	[10,13]
Moisture	Gravimetry after drying at a temperature of 105 °C	[10]
Biological consumption of oxygen	Soil: water ratio 1:10(50). Difference in a dissolved oxygen content in soil suspension and distilled water after 5 days of incubation. A Clark electrode (DKTP-02.3, Russia) and an oximeter.	[8]
Cellulolytic activity of soil	Fragments of linen cloth air-dried and weighed in advance were placed into the soil (glass jars or field plots). At the end of each observation interval, test objects were withdrawn and replaced with a new batch. The withdrawn linen cloth fragments were thoroughly washed from soil particles, air-dried, and weighed. Cellulolytic activity was calculated as the mean rate of the loss in weight of test object relative to its initial weight over the selected observation interval in mg/(g day).	[1,8]

Variable 16S rRNA regions were amplified with two primer combinations, V3V4 and V4V5 (Table 9).

Phusion polymerase (New England Biolabs, USA) and the program described in Table 10 was used for amplification.

The product specificity was checked using electrophoresis in 2% agarose. To prepare sequencing libraries, amplicons were purified using the AMPure XP beads (Table 7). Amplicon concentrations were measured on the Qubit 1 fluorimeter. The final concentrations ranged from 0.8 to 40 ng/μl. Samples containing no DNA processed in the same laboratory were used as negative controls. Their concentrations ranged from 0.2 to 0.6 ng/μl.

To skip the adaptor ligation step, all primers already contained Illumina 1 (forward primer) or Illumina 2 (reverse primer) adaptors. Index PCR was made using Phusion polymerase and the Nextera XT Index kit according to the manufacturer's protocols. The library concentrations were measured on the Qubit 1 fluorimeter using the Qubit DNA HS kit. The libraries were sequenced on Illumina MiSeq with the read length of 250 bp (MiSeq Reagent Kit v2). The average number of reads was 75,000 per sample.

Table 7
The main steps of DNA extraction from thr soil samples, library preparation, and sequencing.

Step	Techniques and reagents used
Soil sampling	Knife
Transporting to the lab (if necessary)	By car
Storing in a fridge	A fridge, –20 °C
Total DNA isolation	DNeasy PowerSoil (Qiagen, Germany)
Samples disrupting	TissueLyser II or TissueLyser LT (Qiagen, Germany, 10 min, 30 Hz)
DNA purifying	The manufacturer's protocol
Measurement of DNA, the amplicon, and the library concentration	The Qubit 1 fluorimeter; the Qubit DNA HS kit (Thermo Fisher Scientific, USA)
Checking of the product specificity	Electrophoresis in 2% agarose (1x Tris-acetate-EDTA buffer, 100 V/14 mA)
Gel visualization	A transilluminator under the UV light
Amplicons of sequencing libraries	The AMPure XP beads (Beckman Coulter, USA); the manufacturer's protocol (90% V:V)
Index PCR	Phusion polymerase (New England Biolabs, USA); the Nextera XT Index kit (Illumina, USA); the manufacturer's protocols
The libraries sequencing	Illumina MiSeq with the read length of 250 bp (MiSeq Reagent Kit v2)
Quantitative PCR (qPCR)	341F–806R primers; qPCRMix HS-SYBR (Evrogen, Russia); DT-Lite machine (DNA Technology, Russia)

Table 8
DNA concentration in soil samples.

		Period after treatment, days			
Load, g/kg	Region	0	90	180	360
Dystric Arenosols (pot experiment)					
0	V3V4	na	5.44	3.8	14.3
0	V3V4	na	5.5	3.67	13.4
0	V3V4	na	na	4.2	10.2
1	V3V4	na	7.19	2.68	11.8
1	V3V4	na	5.49	2.81	8.22
1	V3V4	na	na	3.04	13.9
5	V3V4	na	4.88	3.19	14.3
5	V3V4	na	13.3	2.87	9.72
5	V3V4	na	na	3.05	15.8
10	V3V4	na	3.17	5.66	20.1
10	V3V4	na	12.9	5.41	10.8
10	V3V4	na	na	4.3	10.8
25	V3V4	na	8.66	11.2	3.09
25	V3V4	na	10.4	10.6	5.13
25	V3V4	na	na	12.5	5.82
100	V3V4	na	5.62	5.86	11.1
100	V3V4	na	20.8	4.76	8.6
100	V3V4	na	na	5.29	5.39
0	V4V5	na	5.92	0.8	9.5
0	V4V5	na	4.51	0.79	8.72
0	V4V5	na	na	3.41	8
1	V4V5	na	4.07	8.45	9.48
1	V4V5	na	4.6	9.05	18.2
1	V4V5	na	na	8.67	8.42
5	V4V5	na	0.79	15.1	7.52
5	V4V5	na	5.08	16.7	8.64
5	V4V5	na	3.52	11.3	11.5
10	V4V5	na	5.84	9.65	9.96
10	V4V5	na	5.79	10.4	6.96
10	V4V5	na	na	10.9	8.9
25	V4V5	na	7.8	16.8	8.31

(*continued on next page*)

Table 8 (*continued*)

Load, g/kg	Region	Period after treatment, days			
		0	90	180	360
25	V4V5	na	6.78	15.3	4.06
25	V4V5	na	na	11.9	6.95
100	V4V5	na	3.33	0.463	5.17
100	V4V5	na	4.1	0.584	6.79
100	V4V5	na	na	0.863	6.04
Albic Luvisols (pot experiment)					
0	V3V4	na	12.1	3.47	5.68
0	V3V4	na	8.69	1.89	4.97
0	V3V4	na	na	2.94	7.21
1	V3V4	na	4.47	3.83	4.5
1	V3V4	na	16.4	3.67	4.84
1	V3V4	na	na	3.39	4.4
5	V3V4	na	13.5	2.37	8.07
5	V3V4	na	5.51	2.08	5.75
5	V3V4	na	na	2.68	5.17
10	V3V4	na	6.75	5.58	3.76
10	V3V4	na	5.21	5.99	4.43
10	V3V4	na	na	5.38	4.77
25	V3V4	na	5.39	5.9	6.69
25	V3V4	na	5.72	3.4	7.25
25	V3V4	na	na	4.91	7.61
100	V3V4	na	6.06	1.71	0.253
100	V3V4	na	8.67	0.96	0.227
100	V3V4	na	na	1.15	0.258
0	V4V5	na	3.86	9.71	11.4
0	V4V5	na	4.08	10.8	6.74
0	V4V5	na	na	9.94	5.24
1	V4V5	na	6.49	11.5	4.91
1	V4V5	na	5.93	13.4	3.21
1	V4V5	na	na	13.7	5.26
5	V4V5	na	5.3	4.42	4.9
5	V4V5	na	5.12	4.13	4.31
5	V4V5	na	na	3.45	5.91
10	V4V5	na	6.65	4.85	6.68
10	V4V5	na	6.17	3.98	4.88
10	V4V5	na	na	5.04	7.16
25	V4V5	na	3.29	2.36	2.74
25	V4V5	na	3.83	2.87	10.5
25	V4V5	na	na	3.18	14.2
100	V4V5	na	3.74	0.98	0.503
100	V4V5	na	4.01	0.84	0.439
100	V4V5	na	na	0.87	0.413
Negative control					
NK	V3V4	na	0.069	0.153	too low
NK	V4V5	na	too low	0.138	0.173
Albic Luvisols (field experiment)					
0	V3V4	2.58	0.321	4.11	2.14
0	V3V4	11	2.14	0.793	2.98
1	V3V4	10.4	2.67	1.93	4.67
1	V3V4	5.78	6.21	4.61	3.77
5	V3V4	10.7	7.41	4.82	3.37
5	V3V4	9.07	6.98	2.71	1.26
10	V3V4	3.12	2.93	5.04	5.25
10	V3V4	0.75	1.24	3.33	3.32
25	V3V4	5.83	2.25	1	5.37

(*continued on next page*)

Table 8 (*continued*)

Load, g/kg	Region	Period after treatment, days			
		0	90	180	360
25	V3V4	5.1	1.43	4.86	8.82
100	V3V4	6.07	5.79	2.87	4.84
100	V3V4	3.96	0.59	0.324	5.21
0	V4V5	16.7	2.35	6.72	10.2
0	V4V5	8.69	5.18	5.81	5.26
1	V4V5	8.37	13.4	4.3	3.46
1	V4V5	4.43	15	3.69	3.38
5	V4V5	12.9	2.63	3.76	3.84
5	V4V5	8.31	12.1	4.31	too low
10	V4V5	4.83	15.6	4.08	3.5
10	V4V5	5.27	12.6	3.78	3.07
25	V4V5	6.75	13.7	0.723	6.56
25	V4V5	11.1	17.5	4.09	4.89
100	V4V5	4.61	20.9	2.32	4.86
100	V4V5	6.39	8.3	0.454	4.42
Fibric Histosols					
0	V3V4	3.81	2.49	6.39	9.55
0	V3V4	1.86	6.76	0.692	na
1	V3V4	9.11	8.62	5.88	8.78
1	V3V4	4.83	3.71	3.56	na
5	V3V4	9.63	5.62	7.05	3.87
5	V3V4	12.1	7.97	3.91	na
10	V3V4	12.2	8.7	3.82	6.1
10	V3V4	7.12	0.135	0.194	na
25	V3V4	4.21	5.79	too low	7.98
25	V3V4	6.65	3.91	4.87	na
100	V3V4	0.058	3.64	3.34	8.86
100	V3V4	5.17	6.27	4.96	na
0	V4V5	7	10.5	3.75	5.11
0	V4V5	na	8.92	1.2	na
1	V4V5	18.8	14.6	6.1	13.1
1	V4V5	16.1	17.5	4.09	na
5	V4V5	16.6	14.3	13	13.7
5	V4V5	18.8	14.8	1.74	na
10	V4V5	18.3	18.6	2.42	5.07
10	V4V5	8.03	8.44	2.19	na
25	V4V5	9.86	15.5	1.98	3.59
25	V4V5	12.9	12.5	3.34	na
100	V4V5	7.29	17.4	7.17	4.49
100	V4V5	13.3	13.7	4.43	na
Negative control					
NK	V3V4	0.107			
NK	V4V5	0.096			

Table 9
Primers used for amplification in this research.

Primer	Sequence (5′–3′)
341F	CCTAYGGGRBGCASCAG
806R	GGACTACNNGGGTATCTAAT
515F	GTGCCAGCMGCCGCGGTAA
907R	CCGTCAATTCCTTTGAGTTT

Note: F, forward primer; R, reverse primer.

Table 10
Parameters of amplification used in this research.

Step	Temperature, °C	*Time, sec*	Remarks
1	95	120	initial DNA melting
24 cycles as follows:			
2	95	30	melting
3	58	30	primer annealing
4	72	40	synthesis
5	72	300	–

Table 11
The main steps of sequence data analysis used.

Step	Software	References
Analysis of read quality	FastQC	[14]
Quality filtering, denoising, pair reads merging, and chimera filtering	the R package DADA2 v. 1.14.1. Used parameters (to lose less than 50% of reads in the pipeline): maxEE=*c*(3,3), minOverlap=8, maxMismatch=1, minFoldParentOverAbundance=8	[15]
Analysis of the obtained ASV tables	R package phyloseq	[16]
The ASV tables filtering of potential contaminants taking into account the amplicon concentrations and ASVs found in negative controls	the R package decontam	[17]
Assigning of taxonomic labels to ASVs. Filtering out of ASVs assigned to "Chloroplast" and not classified at the domain level.	IdTaxa from the R package DECIPHER trained on 16S rRNA gene sequences from the SILVA database	[18–20]
Multiple alignment of ASVs and construction of phylogenetic trees	AlignSeqs from the DECIPHER package and FastTree v.2.1.11	[19,21]
Prediction of the metabolic potential of microbial communities	Picrust2 with MetaCyc pathways reconstruction	[22]
Taxonomical and functional piecharts	R package psadd	[16]

To ensure that changes in bacterial communities were not only due to their death under high kerosene concentrations, the relative amount of bacteria in soil samples with or without kerosene load at different time points was measured by quantitative PCR with 341F–806R primers.

2.2.3. Sequence data analysis

The main steps of sequence data analysis are in Table 11. Low total numbers of reads in negative controls and the data of the decontam analysis indicated low levels of possible contamination. Additionally, singleton ASVs were removed from the subsequent analysis. After filtration, the mean number of reads in the samples was approximately 52,000.

Filtered ASV tables in phyloseq format were deposited at figshare (https://doi.org/10.6084/m9.figshare.19609503.v1). Before subsequent analysis all samples were rarefied to a standard number of reads (10,000 reads) in order to account for differences in sequencing depth. Tables with functional predictions were deposited at figshare (https://doi.org/10.6084/m9.figshare.19609503.v1).

Ethics Statements

Not applicable. Research did not include human or animal subjects.

CRediT Author Statement

Ivan N. Semenkov: Conceptualization, Data curation, Methodology, Project administration, Resources, Supervision, Visualization, Writing- Original draft preparation, Writing- Reviewing and Editing. **Pavel V. Shelyakin:** Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing- Reviewing and Editing. **Daria D. Nikolaeva:** Data curation, Investigation, Software, Validation, Visualization, Writing- Reviewing and Editing. **Maria N. Tutukina:** Conceptualization, Data curation, Investigation, Laboratory work, Methodology, Project administration, Resources, Writing- Reviewing and Editing. **Anna V. Sharapova:** Conceptualization, Field work, Investigation, Laboratory work. **Sergey A. Lednev:** Field work, Investigation. **Yuliya V. Sarana:** Investigation, Laboratory work. **Mikhail S. Gelfand:** Conceptualization, Methodology, Project administration, Writing- Reviewing and Editing. **Pavel P. Krechetov:** Conceptualization, Field work, Investigation, Methodology, Project administration. **Tatiana V. Koroleva:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing-Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data on the temporal changes in soil properties and microbiome composition after a jet-fuel contamination during the pot and field experiments (Original data) (figshare).

Acknowledgments

Sequencing service was provided by the Genomics Core Facility of the Skolkovo Institute of Science and Technology. Soil physico-chemical properties were controlled in the analytical center of the Faculty of Chemistry of the M.V. Lomonosov Moscow State University.

Funding

The work was supported by the RFBR project No. 19–29–05206 (field work and data description) and the M.V. Lomonosov Moscow State University (the Interdisciplinary Scientific and Educational School of “Future Planet and Global Environmental Change” and project I.4. “Anthropogenic geochemical transformation of the components of landscapes”) (chemical-analytical work).

References

[1] A.V. Sharapova, I.N. Semenkov, P.P. Krechetov, S.A. Lednev, T.V. Koroleva, The effect of kerosene pollution on the cellulolytic activity of Albic Retisols and Protic Arenosols: a laboratory experiment, Eurasian Soil Sci. Sci. 55 (2022) 235–241, doi:10.1134/S1064229322020119.

[2] T.V. Koroleva, P.P. Krechetov, I.N. Semenkov, A.V. Sharapova, S.A. Lednev, A.M. Karpachevskiy, A.D. Kondratyev, N.S. Kasimov, The environmental impact of space transport, Transp. Res. Part D Transp. Environ. 58 (2018) 54–69, doi:10.1016/j.trd.2017.10.013.

[3] T.V. Koroleva, I.N. Semenkov, A.V. Sharapova, P.P. Krechetov, S.A. Lednev, Ecological consequences of space rocket accidents in Kazakhstan between 1999 and 2018, Environ. Pollut. 268 (2021) 115711, doi:10.1016/j.envpol.2020.115711.

[4] P.V. Shelyakin, I.N. Semenkov, M.N. Tutukina, D.D. Nikolaeva, A.V. Sharapova, Y.V. Sarana, S.A. Lednev, A.D. Smolenkov, M.S. Gelfand, P.P. Krechetov, T.V. Koroleva, The influence of kerosene on microbiomes of diverse soils, Life 12 (2022) 221, doi:10.3390/life12020221.

[5] S.A. Lednev, I.N. Semenkov, G.V. Klink, P.P. Krechetov, A.V. Sharapova, T.V. Koroleva, Impact of kerosene pollution on ground vegetation of southern taiga in the Amur Region, Russia, Sci. Total Environ. 772 (2021) 144965, doi:10.1016/j.scitotenv.2021.144965.

[6] M.F. Dorokhova, P.P. Krechetov, T.V. Koroleva, A.V. Sharapova, Algo-cyanobacterial communities as indicators of soil pollution with jet-fuel, in: II Int. Sci. Pract. Conf. Dedic. to 105th Anniv. Birth Profr. Emilia Adrianovna Shtina "Algae Cyanobacteria Nat. Agric. Ecosyst., Vyatskaya GSHA, Kirov, 2015, pp. 118–122. https://istina.msu.ru/publications/article/11446550/.

[7] S.A. Lednev, T.V. Koroleva, I.N. Semenkov, G.V. Klink, P.P. Krechetov, A.V. Sharapova, A.M. Karpachevskiy, G.V. Klink, The natural regeneration of desert ecosystem vegetation at the 2013 crash site of a Proton-M launch vehicle, Republic of Kazakhstan, Ecol. Indic. 101 (2019) 603–613, doi:10.1016/j.ecolind.2019.01.045.

[8] P.P. Krechetov, A.V. Sharapova, I.N. Semenkov, T.V. Koroleva, Protocol of conjugate evaluation of the biological activity of soils in terms of cellulolytic activity and biological consumption of oxygen, MethodsX 9 (2022) 101841, doi:10.1016/j.mex.2022.101841.

[9] T.A. Bolotnik, Y.V. Timchenko, I.V. Plyushchenko, V.V. Levkina, A.V. Pirogov, A.D. Smolenkov, M.V. Popik, O.A. Shpigun, Use of Chemometric Methods of Data Analysis for the Identification and Typification of Petroleum and Petroleum Products, J. Anal. Chem. 74 (2019) 1336–1340, doi:10.1134/S1061934819130045.

[10] M. Pansu, J. Gautheyrou, Handbook of Soil Analysis, 2006. https://doi.org/10.1007/978-3-540-31211-6.

[11] SS 54650-2011, Soils. Determination of mobile phosphorus and potassium compounds by Kirsanov method modified by CINAO, (2011) 11. https://docs.cntd.ru/document/1200094361.

[12] SS-26205-91, Soils. Determination of mobile compounds of phosphorus and potassium by Machigin method modified by CINAO, (1991) 10. https://docs.cntd.ru/document/1200023449.

[13] SS-17.4.4.01-84, Nature protection. Soils. Methods for determining the capacity of cation exchange, (1984). https://docs.cntd.ru/document/1200012802.

[14] S.R. Andrews, A Quality Control tool for High Throughput Sequence Data, FastQC. (2019). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/(accessed December 17, 2021).

[15] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: high-resolution sample inference from Illumina amplicon data, Nat. Methods. 13 (2016) 581–583, doi:10.1038/nmeth.3869.

[16] P.J. McMurdie, S. Holmes, Phyloseq: an R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data, PLoS ONE 8 (2013) e61217, doi:10.1371/journal.pone.0061217.

[17] N.M. Davis, Di.M. Proctor, S.P. Holmes, D.A. Relman, B.J. Callahan, Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data, Microbiome. 6 (2018) 226. https://doi.org/10.1186/s40168-018-0605-2.

[18] A. Murali, A. Bhargava, E.S. Wright, IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences, Microbiome. 6 (2018) 140. https://doi.org/10.1186/s40168-018-0521-5.

[19] E.S. Wright, Using DECIPHER v2.0 to analyze big biological sequence data in R, R J 8 (2016) 352–359, doi:10.32614/rj-2016-025.

[20] G. Kim, J. Bae, M.J. Kim, H. Kwon, G. Park, S.J. Kim, Y.H. Choe, J. Kim, S.H. Park, B.H. Choe, H. Shin, B. Kang, T. Magoč, S.L. Salzberg, C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glöckner, S. Trachtenberg, M. Reyman, M.A. van Houten, D. van Baarle, A.A.T.M. Bosch, W.H. Man, M.L.J.N. Chu, K. Arp, R.L. Watson, E.A.M. Sanders, S. Fuentes, D. Bogaert, A. Ramírez-Guzmán, Y. Taran, M.A. Armienta, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucl. Acids Res 41 (2019) D590–D596.

[21] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2 - Approximately maximum-likelihood trees for large alignments, PLoS ONE 5 (2010) e9490, doi:10.1371/journal.pone.0009490.

[22] G.M. Douglas, V.J. Maffei, J.R. Zaneveld, S.N. Yurgel, J.R. Brown, C.M. Taylor, C. Huttenhower, M.G.I. Langille, PICRUSt2 for prediction of metagenome functions, Nat. Biotechnol. 38 (2020) 685–688, doi:10.1038/s41587-020-0548-6.