

Contents lists available at ScienceDirect

Data in Brief

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

## Data Article

# ITS and 16S rDNA metagenomic dataset of different soils from flax fields



Daiana A. Zhernova<sup>a,#,\*</sup>, Elena N. Pushkova<sup>a,#</sup>, Tatiana A. Rozhmina<sup>b</sup>, Liubov V. Povkhova<sup>a</sup>, Roman O. Novakovskiy<sup>a</sup>, Anastasia A. Turba<sup>a</sup>, Elena V. Borkhert<sup>a</sup>, Elizaveta A. Sigova<sup>a</sup>, Ekaterina M. Dvorianinova<sup>a</sup>, George S. Krasnov<sup>a</sup>, Nataliya V. Melnikova<sup>a</sup>, Alexey A. Dmitriev<sup>a,\*</sup>

<sup>a</sup> Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilova, Moscow 119991, Russia
<sup>b</sup> Federal Research Center for Bast Fiber Crops, 35 Lunacharskogo, Torzhok 172002, Russia

## ARTICLE INFO

Article history: Received 24 April 2023 Revised 28 October 2023 Accepted 14 November 2023 Available online 18 November 2023

Dataset link: Microbial community of flax fields (Original data)

Keywords: Flax Linum usitatissimum Soil metagenomics Microbial diversity Targeted deep sequencing Amplicon sequence data ITS 165 rDNA

## ABSTRACT

Flax (Linum usitatissimum L.), one of the important and versatile crops, is used for the production of oil and fiber. To obtain high and stable yields of flax products, L. usitatissimum varieties should be cultivated under optimal conditions, including the composition of the soil microbiome. We evaluated the diversity of microorganisms in soils under conditions unfavorable for flax cultivation (suboptimal acidity or herbicide treatment) or infected with causative agents of harmful flax diseases (Septoria linicola, Colletotrichum lini, Melampsora lini, or Fusarium oxysporum f. sp. lini). For this purpose, twenty-two sod-podzolic soil samples were collected from flax fields and their metagenomes were analyzed using the regions of 16S ribosomal RNA gene (16S rDNA) and internal transcribed spacers (ITS) of the ribosomal RNA genes, which are used in phylogenetic studies of bacteria and fungi. Amplicons were sequenced on the Illumina MiSeq platform (reads of 300 + 300 bp). On average, we obtained 8,400 reads for ITS and 43,300 reads for 16S rDNA per sample. For identification of microorganisms in the soil samples, the Illumina reads were processed using DADA2. The raw

\* Corresponding authors.

E-mail addresses: zhernova.d@yandex.ru (D.A. Zhernova), alex\_245@mail.ru (A.A. Dmitriev).

<sup>#</sup> These authors contributed equally to this work.

https://doi.org/10.1016/j.dib.2023.109827

<sup>2352-3409/© 2023</sup> The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

data are deposited in the Sequence Read Archive under the BioProject accession number PRJNA956957. Tables listing the microorganisms identified in the soil samples are available in this article. The obtained dataset can be used to analyze the fungal and bacterial composition of flax field soils and their relationship to environmental conditions, including suboptimal soil acidity and infection with fungal pathogens. In addition, it can help to understand the influence of herbicide treatment on the microbial diversity of flax fields. Another useful application of our data is the ability to assess the suitability of the soil microbiome for flax cultivation.

© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

## Specifications Table

Subject	Agricultural Microbiology, Microbiology: Microbiome.
Specific subject area	Soil Metagenomics.
Data format	Raw (fastq files) and Analyzed (lists of microorganisms identified in the soil samples).
Type of data	Amplicon sequence data and Tables.
Data collection	Twenty-two sod-podzolic soil samples with different characteristics were
	collected from flax fields at a depth of about 20 cm. Total DNA was extracted from each sample. To assess the fungal and bacterial composition of the soil
	(paired-end reads, $300 + 300$ bp). Lists of microorganisms were generated using DADA2.
Data source location	Soil samples were collected from flax fields of the Institute for Flax (Torzhok, Russia). Details are given in Table 1 of this article.
Data accessibility	Raw amplicon sequences are available at the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA956957
	(https://www.ncbi.nlm.nih.gov/sra/PRJNA956957). Lists of microorganisms
	identified in the soil samples are available in this article.

## 1. Value of the Data

- The dataset is a valuable source of information on fungal and bacterial biodiversity in soil ecosystems. They are particularly important in the context of the cultivation of agricultural plants, which are influenced by various external abiotic factors and microorganisms.
- The data would be useful for ecological and microbiological analyses of the diversity of microorganisms in soil ecosystems under different conditions. The dataset could also help to reveal a possible relationship between flax fungal infection and the composition of soil microorganisms.
- The dataset allows to assess the level of pressure caused by human agricultural activities (herbicide treatment, artificial changes in soil acidity) on soil ecosystems. These data could be compared with others obtained from wild communities.
- The data could be used to assess the suitability of the soil microbiome for flax cultivation.
- The dataset provides an opportunity to study evolution and relationships between taxa using ITS sequences for fungi and 16S rDNA sequences for bacteria.

## 2. Data Description

Such a valuable agricultural crop as flax (*Linum usitatissimum* L.) is used for its fiber and seeds, which are the richest source of lignans and are high in omega-3, easily digestible pro-

tein, dietary fiber, vitamins, and minerals. Linseed is also used in the production of environmentally friendly paints, varnishes, composites, and animal feed [1–3]. To produce high-quality flax products, *L. usitatissimum* varieties should be grown under specific conditions in soils with an optimal microbiome composition. The majority of known causative agents of flax diseases are soil-borne and seed-borne [4]. Research by Hartman and Tringle suggested that the composition and function of the root microbiome were closely related to abiotic stress and the ultimate consequences for plant health [5]. In the present study, we collected samples of soils infected with *Septoria linicola, Colletotrichum lini, Melampsora lini,* or *Fusarium oxysporum* f. sp. *lini,* which cause the most harmful flax diseases [6,7], soils with suboptimal acidity for flax cultivation [8,9], and soils treated with herbicides. Understanding the diversity of the microbiome in relation to soil conditions and its association with flax growing plays an important role in the successful cultivation of this crop. In the future, it may be possible to assess the suitability of soils for flax growing based on the composition of the microbiome and improve it, for example, through microbiological treatment.

To evaluate the diversity of microorganisms under the studied conditions (different soil acidity, herbicide treatment, and infection with pathogens), we collected 22 sod-podzolic soil samples in flax fields divided into experimental sections. Soil samples infected with pasmo, anthracnose, or rust causative agents were obtained from the first field. Soil samples from the second field had suboptimal soil acidity (pH=6.0) or were treated with about 10 mg/ha or 50 mg/ha of herbicide. We also obtained samples of soils with pH levels between 4.5 and 6.0 or infected with *Fusarium oxysporum* f. sp. *lini* (Table 1). Each group of experimental samples had a matched control sample.

For the 22 soil samples, we deep-sequenced the fragments of the regions commonly used in phylogenetic studies of bacteria and fungi: 16S ribosomal RNA gene (16S rDNA) and internal transcribed spacers (ITS) of the ribosomal RNA genes [10,11].

Amplicon libraries were prepared by two-step PCR. The following primers were used in the first step: ITS1\_fu\_Illu\_F and ITS4\_Illu\_R for ITS, and 16S\_F and 16S\_R for 16S rDNA (Table 2). In the second step, Nextera XT Index primers were applied. Amplicons were sequenced on the Illumina MiSeq platform (reads – 300+300 bp). On average, we obtained 8,400 (range – 1,700-17,500) reads for ITS and 43,300 (range – 31,000-58,800) reads for 16S rDNA. Raw data (fastq format) were deposited in the Sequence Read Archive (SRA) under the BioProject accession number PRJNA956957. These data play a key role in assessing the genetic diversity of soil microorganisms, including pathogenic fungi and bacteria, and in developing time-saving and accurate test systems for assessing soil quality. Based on this dataset, future research will provide further insight into the question of soil microbiome composition and its effect on flax cultivation.

The list of microorganism taxa in each sample was generated by automated comparison of the obtained sequence variants with DNA sequences deposited in GenBank using DADA2 [12]. The ratio of bacterial and fungal taxa in each soil sample is available in Supplementary Table 1 (ITS) and Supplementary Table 2 (16S rDNA). Figures representing the ratio of fungal and bacterial Classes (as an example) in soil samples from flax fields with different characteristics are available in Supplementary Figure 1 (ITS) and Supplementary Figure 2 (16S rDNA).

## 3. Experimental Design, Materials and Methods

## 3.1. Material

Twenty-two soil samples from experimental flax fields and soil boxes were provided by the Institute for Flax (Torzhok, Russia). The studied soil was of the sod-podzolic type and was taken at a depth of 20 cm. The studied material included soil samples with suboptimal acidity, treated with herbicides, and infected with pathogens leading to pasmo (*Septoria linicola*), anthracnose (*Colletotrichum lini*), rust (*Melampsora lini*), or fusarium wilt (*Fusarium oxysporum* f. sp. *lini*) (Table 1). The infected areas of flax fields were created according to the methods described in

#### Table 1

List	of	soil	sam	oles	under	study,	their	characteristics	and	sampling	locations.	
											· · · · · · · · · · · · · · · · · · ·	

Sample name	Soil characteristics	Sampling location*
S1 S2	Infected with a pasmo causative agent ( <i>Septoria linicola</i> )	Flax field, 57.053056 N, 35.045556 E Flax field, 57.053056 N, 35.045556 E
52	( <i>Colletotrichum lini</i> ). No plants. Biological replicate for S3	Tax field, 57.055050 N, 55.045550 E
S3	Infected with an anthracnose causative agent (Colletotrichum lini). No plants. Biological replicate for S2	Flax field, 57.053056 N, 35.045556 E
S4	Infected with an anthracnose causative agent (Colletotrichum lini)	Flax field, 57.053056 N, 35.045556 E
S5	Infected with a rust causative agent ( <i>Melampsora lini</i> ). Biological replicate for S6	Flax field, 57.053056 N, 35.045556 E
S6	Infected with a rust causative agent ( <i>Melampsora lini</i> ). Biological replicate for S5	Flax field, 57.053056 N, 35.045556 E
K_S1-S6	Control for S1-S6	Flax field, 57.053611 N, 35.045833 E
S7	рН=6.0	Flax field, 57.051667 N, 35.052222 E
S8	Herbicide «Magnum», 10 mg/ha	Flax field, 57.050556 N, 35.052222 E
S9	Herbicide «Magnum», 50 mg/ha. Biological replicate for S10	Flax field, 57.050556 N, 35.052222 E
S10	Herbicide «Magnum», 50 mg/ha. Biological replicate for S9	Flax field, 57.050556 N, 35.052222 E
K_S7-S10	Control for S7-S10 (pH=5.2). Biological replicate for K2_S7-S10	Flax field, 57.050556 N, 35.052222 E
K2_S7-S10	Control for S7-S10 (pH=5.2). Biological replicate for K_S7-S10	Flax field, 57.050556 N, 35.052222 E
S11	Infected with a fusarium wilt causative agent ( <i>Fusarium</i> oxysporum f. sp. lini). Biological replicate for S12	Experimental soil box, Institute for Flax, Torzhok, Russia
S12	Infected with a fusarium wilt causative agent ( <i>Fusarium</i> oxysporum f. sp. lini). Biological replicate for S11	Experimental soil box, Institute for Flax, Torzhok, Russia
K_S11-S12	Control for S11-S12	Experimental soil box, Institute for Flax, Torzhok, Russia
S13	pH=6.0. Biological replicate for S14	Experimental soil box, Institute for Flax, Torzhok, Russia
S14	pH=6.0. Biological replicate for S13	Experimental soil box, Institute for Flax, Torzhok Russia
K_\$13-\$14	Control for \$13-\$14 (pH=5.2)	Experimental soil box, Institute for Flax, Torzhok, Russia
S15	Liming to pH=5.2. Biological replicate for S16	Experimental soil box, Institute for Flax, Torzhok Russia
S16	Liming to pH=5.2. Biological replicate for S15	Experimental soil box, Institute for Flax, Torzhok, Russia
K_S15-S16	Control for S15-S16 (pH=4.5)	Experimental soil box, Institute for Flax, Torzhok, Russia

Note: Flax plants were growing in soil at the time of collection, unless otherwise noted. \* - north latitude and east longitude.

Loshakova et al. [6]. The acidity of pH=5.2 (optimal for flax cultivation) for samples S15 and S16 was achieved by liming. All other samples were taken from fields and had a natural pH level (see Table 1). The treatment with the herbicide "Magnum" was carried out in June, about 3 weeks after the germination of flax. The herbicide was applied at concentrations of about 10 mg/ha and 50 mg/ha. The collected soil samples were stored at -75°C until DNA extraction.

## 3.2. DNA extraction and quality control

DNA was extracted from 22 soil samples (average mass was about 250-300 mg) according to the standard protocol of Syntol "MetaGen" kit (Russia). Agarose gel electrophoresis (2% agarose) and Qubit fluorometer (Thermo Fisher Scientific, USA) were used to control DNA quality and evaluate DNA quantity.

Primer name	Primer sequence	Target locus
ITS1_fu_IIIu_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTGGTCATTTAGAGGAAGTAA	ITS [16]
ITS4_IIIu_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTCCGCTTATTGATATGC	ITS [16]
16S_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG	16S rDNA [15]
16S_R	GTCTCGTGGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC	16S rDNA [15]

#### Table 2

Primer sequences used for the first step of library preparation.

Note: The Illumina overhang adapter sequences (bold font) were added to the locus-specific primers (normal font).

#### 3.3. DNA library preparation and sequencing

Amplification of ITS and 16S rDNA regions was performed for 22 soil samples. Amplicon libraries were prepared according to the protocol with two-step PCR, as described in our previous articles [13,14]. Briefly, in the first step, target sequences were amplified using primers containing locus-specific sequences for ITS and 16S rDNA amplification [15,16] and overhang adapters (Table 2). For each sample, amplicons were pooled equimolarly and the second PCR was performed with Nextera XT Index primers consisting of dual-index barcodes and sequencing adapters. All PCR products were then pooled equimolarly and the library was assessed for quality using the 2100 Bioanalyzer (Agilent Technologies, USA) and for quantity using the Qubit fluorometer (Thermo Fisher Scientific). The library was sequenced on the MiSeq platform (Illumina, USA) with the Illumina MiSeq Reagent Kit v3 (600 cycles).

#### 3.4. Preliminary data analysis

The pairs of Illumina reads were demultiplexed by forward primers using cutadapt (-noindels, -action=none) [17]. Then, reverse reads from the demultiplexed pairs were treated as forward reads, and the pairs were demultiplexed again by reverse primers using the same software (-no-indels, -action=none). In the next step, the reads were processed with DADA2 [12], including error correction, ribosomal sequence variants (RSV) inference, and chimeric RSV removal. Forward and reverse reads were merged using MeFiT [18] prior to DADA2 processing, as the overlap region of the reads is quite small, with a general tendency to drop in quality towards the 3'-end. Annotation of the obtained RSV sequences was performed using the Silva 138.1 database [19] with the built-in DADA2 engine (namely, the RDP naive Bayesian classifier [20]).

#### Limitations

Absence of biological replicates for some samples.

### **Ethics Statement**

Authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

## **CRediT** author statement

Daiana A. Zhernova: Data analysis, Writing. Elena N. Pushkova: Performing experiments, Writing. Tatiana A. Rozhmina: Conceptualization, Performing experiments. Liubov V. Povkhova: Performing experiments. Roman O. Novakovskiy: Performing experiments. Anastasia A. Turba: Performing experiments. Elena V. Borkhert: Performing experiments. Elizaveta A. Sigova: Data analysis. Ekaterina M. Dvorianinova: Performing experiments, Data analysis. George S. Krasnov: Data analysis, Writing. Nataliya V. Melnikova: Conceptualization, Data analysis, Writing. Alexey A. Dmitriev: Conceptualization, Data analysis, Writing.

## **Data Availability**

Microbial community of flax fields (Original data) (Sequence Read Archive (SRA))

## Acknowledgments

This work was financially supported by the Russian Science Foundation [grant number 22-16-00169]. This work was performed using the equipment of EIMB RAS "Genome" center (http://www.eimb.ru/ru1/ckp/ccu\_genome\_ce.php).

## **Declaration of Competing Interests**

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 article.

## Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2023.109827.

#### References

- A. Mueed, S. Shibli, M. Jahangir, S. Jabbar, Z. Deng, A comprehensive review of flaxseed (*Linum usitatissimum L.*): health-affecting compounds, mechanism of toxicity, detoxification, anticancer and potential risk, Crit. Rev. Food Sci. Nutr. (2022) 1–24, doi:10.1080/10408398.2022.2092718.
- [2] A.P. More, Flax fiber-based polymer composites: a review, Adv. Compos. Hybrid Mater. 5 (1) (2022) 1–20, doi:10. 1007/s42114-021-00246-9.
- [3] P. Kezimana, A.A. Dmitriev, A.V. Kudryavtseva, E.V. Romanova, N.V. Melnikova, Secoisolariciresinol diglucoside of flaxseed and its metabolites: biosynthesis and potential for nutraceuticals, Front. Genet. 9 (2018) 641, doi:10.3389/ fgene.2018.00641.
- [4] E. Gruzdevienė, I. Brazauskienė, J. Repečkienė, A. Lugauskas, The occurrence of pathogenic fungi during flax growing season in Central Lithuania, J. Plant Protect. Res. 48 (2) (2008) 255–265, doi:10.2478/v10045-008-0029-2.
- [5] K. Hartman, S.G. Tringe, Interactions between plants and soil shaping the root microbiome under abiotic stress, Biochem. J. 476 (19) (2019) 2705–2724, doi:10.1042/BCJ20180615.
- [6] N.I. Loshakova, T.V. Krylova, L.P. Kudryavtseva, Methodological guidelines on the phytopathological assessment of the resistance of fiber flax to diseases, RAAS, Moscow, 2000, In Russian.
- [7] R.O. Novakovskiy, E.M. Dvorianinova, T.A. Rozhmina, L.P. Kudryavtseva, A.A. Gryzunov, E.N. Pushkova, L.V. Povkhova, A.V. Snezhkina, G.S. Krasnov, A.V. Kudryavtseva, N.V. Melnikova, A.A. Dmitriev, Data on genetic polymorphism of flax (*Linum usitatissimum L.*) pathogenic fungi of *Fusarium, Colletotrichum, Aureobasidium, Septoria*, and *Melampsora* genera, Data Brief 31 (2020) 105710, doi:10.1016/j.dib.2020.105710.
- [8] T.A. Rozhmina, A.A. Zhuchenko Jr, N.V. Melnikova, A.D. Smirnova, Resistance of flax gene pool samples to edaphic stress caused by low acidity, Agric. Sci. Euro-North-East 21 (2) (2020) 133–140 In Russian, doi:10.30766/2072-9081. 2020.21.2.133-140.
- [9] A.A. Dmitriev, G.S. Krasnov, T.A. Rozhmina, A.V. Zyablitsin, A.V. Snezhkina, M.S. Fedorova, E.N. Pushkova, P. Kezimana, R.O. Novakovskiy, L.V. Povkhova, M.I. Smirnova, O.V. Muravenko, N.L. Bolsheva, A.V. Kudryavtseva, N.V. Melnikova, Flax (*Linum usitatissimum* L.) response to non-optimal soil acidity and zinc deficiency, BMC Plant Biol. 19 (Suppl 1) (2019) 54, doi:10.1186/s12870-019-1641-1.
- [10] D.M. Ward, R. Weller, M.M. Bateson, 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community, Nature 345 (6270) (1990) 63-65, doi:10.1038/345063a0.

- [11] C.L. Schoch, K.A. Seifert, S. Huhndorf, V. Robert, J.L. Spouge, C.A. Levesque, W. Chen, et al., Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi, Proc. Nat. Acad. Sci. U.S.A. 109 (16) (2012) 6241–6246, doi:10.1073/pnas.1117018109.
- [12] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J. Johnson, S.P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data, Nat. Methods 13 (2016) 581–583, doi:10.1038/nmeth.3869.
- [13] A.A. Dmitriev, P. Kezimana, T.A. Rozhmina, A.A. Zhuchenko, L.V. Povkhova, E.N. Pushkova, R.O. Novakovskiy, M. Pavelek, G.N. Vladimirov, E.N. Nikolaev, O.A. Kovaleva, Y.I. Kostyukevich, V.V. Chagovets, E.V. Romanova, A.V. Snezhkina, A.V. Kudryavtseva, G.S. Krasnov, N.V. Melnikova, Genetic diversity of SAD and FAD genes responsible for the fatty acid composition in flax cultivars and lines, BMC Plant Biol. 20 (Suppl 1) (2020) 301, doi:10.1186/s12870-020-02499-w.
- [14] L.V. Povkhova, E.N. Pushkova, T.A. Rozhmina, A.A. Zhuchenko, R.I. Frykin, R.O. Novakovskiy, E.M. Dvorianinova, A.A. Gryzunov, E.V. Borkhert, E.A. Sigova, G.N. Vladimirov, A.V. Snezhkina, A.V. Kudryavtseva, G.S. Krasnov, A.A. Dmitriev, N.V. Melnikova, Development and complex application of methods for the identification of mutations in the FAD3A and FAD3B genes resulting in the reduced content of linolenic acid in flax oil, Plants 12 (1) (2022) 95, doi:10.3390/plants12010095.
- [15] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F.O. Glockner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic. Acids. Res. 41 (1) (2013) e1, doi:10.1093/nar/gks808.
- [16] T.J. White, T. Bruns, S. Lee, J. Taylor, Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (Eds.), PCR Protocols: a Guide to Methods and Applications, Academic Press Inc., San Diego, 1990, pp. 315–322.
- [17] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet. J. 17 (1) (2011) 10–12, doi:10.14806/ej.17.1.200.
- [18] H.I. Parikh, V.N. Koparde, S.P. Bradley, G.A. Buck, N.U. Sheth, MeFiT: merging and filtering tool for illumina pairedend reads for 16S rRNA amplicon sequencing, BMC Bioinf. 17 (1) (2016) 491, doi:10.1186/s12859-016-1358-1.
- [19] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glockner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic. Acids. Res. 41 (D1) (2013) D590–D596, doi:10.1093/nar/gks1219.
- [20] Q. Wang, G.M. Garrity, J.M. Tiedje, J.R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Appl. Environ. Microbiol. 73 (16) (2007) 5261–5267, doi:10.1128/AEM.00062-07.