Master's Thesis

Association between microbiota community diversity and metabolite diversity: bank voles (*Clethrionomys glareolus*) inhabiting the Chornobyl Exclusion Zone as a model

Hanna Sipola

University of Jyväskylä

Department of Biological and Environmental Science

19 May 2024

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science
Department of Biological and Environmental Science
Master's Degree Programme in Cell and Molecular Biology

Sipola, Hanna

Association between microbiota community diversity and metabolite diversity: bank voles (*Clethrionomys glareolus*) inhabiting the Chornobyl Exclusion Zone as a model

MSci Thesis: 55 p., 5 appendices (18 p.)
Supervisors: Prof Phillip Watts and DVM Ilze Brila
Inspectors: Saana Sipari and Sameli Piirto
May 2024

Keywords: Earth Microbiome Project, functional redundancy, metabolomics

Exposure to excess ionising radiation can cause cellular damage and thus affect organism health. Bank voles (*Clethrionomys glareolus*) inhabiting the Chornobyl Exclusion Zone (CEZ) may be exposed to elevated levels of radiation due to the presence of radionuclides in the environment following the accident in the nuclear power plant. One of the reported impacts of exposure to high dose of radiation is a change in the composition of the gut microbiota. Whether this change alters the metabolites is not known. For example, many gut microbes could perform similar functions, in which case the alteration of the microbiota may not have a notable effect on the services provided to the host, e.g. metabolite production. Addressing this knowledge gap, bank vole faecal samples were collected from contaminated and uncontaminated areas within the CEZ and uncontaminated areas outside the CEZ and near Kyiv. The microbiota and metabolite composition of all samples was collected as part of the Earth Microbiome 500 project. Briefly, the microbiota were characterised using 16S rRNA v4 amplicon sequencing, with data processed in QIIME2 and analysed using packages within PHYLOSEQ and VEGAN. Metabolite determined using untargeted metabolomics (ultra-high performance liquid chromatography coupled to a mass spectrometer), with peak calling in MZmine and subsequent data analyses in PHYLOSEQ/VEGAN. I found that (1) radiation significantly affected the alpha and beta diversities of gut microbiota, but not the metabolites and (2) the changes in gut bacteria diversity were not readily associated with the changes in metabolite diversity. These results strengthened an earlier understanding that exposure to an increased level of radiation alters the gut microbiota of bank voles. Since exposure radiation affected the gut microbiota diversity, but no association was found between changes in gut microbiota and metabolites, this reinforces a complexity of understanding whether changes in microbiota communities inferred from amplicon sequencing data can be interpreted in term of alterations to key services such as the provision of metabolites.

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta
Bio- ja ympäristötieteiden laitos
Solu- ja molekyylibiologian maisteriohjelma

Sipola, Hanna	Mikrobiyhteisön monimuotoisuuden ja metaboliittien monimuotoisuuden välinen yhteys: Chornobyl Exclusion Zone -aluetta asuttavat metsämyyrät (*Clethrionomys glareolus*) mallina
Pro gradu tutkielma:	55 s., 5 liitettä (18 s.)
Työn ohjaajat:	Prof Phillip Watts ja DVM Ilze Brila
Tarkastajat:	Saana Sipari ja Sameli Piirto
Toukokuu 2024	

Hakusanat: Earth Microbiome Projekti, metabolomiikka, toiminnallinen redundanssi

Ionisoivalla säteilylle altistuminen voi aiheuttaa soluvaurioita ja siten vaikuttaa isännän terveyteen, kuten muuttamalla suolistomikrobistoa. Metsämyyrä (*Clethrionomys glareolus*) on uudelleen asuttanut Chornobyl Exclusion Zone (CEZ) -aluetta tapahtuneen ydinvoimalaonnettomuuden jälkeen, mikä tarjoaa ainutlaatuisen mahdollisuuden tutkia kroonisen pieniannoksisen säteilyn pitkäaikaisia biologisia vaikutuksia. On olemassa tutkimuksia, jotka osoittavat, että suurille säteilyannoksille altistuneiden metsämyyrien suolistomikrobisto on muuttunut. Ei kuitenkaan tiedetä, vaikuttaako nämä suolistomikrobiston muutokset isännän terveyteen muuttamalla suolistomikrobistojen tuottamia metaboliitteja. Esimerkiksi, mitä läheisempää sukua mikrobit ovat toisilleen, sitä samankaltaisemmat ovat niiden isännälle tarjoamat hyödyt, kuten metaboliittien tuotanto. Asian tutkimiseksi metsämyyrien ulostenäytteitä kerättiin saastuneilta alueilta CEZ:ssä sekä saastumattomilta alueilta CEZ:n ja Kiovan läheltä. Näytteet sekä niiden mikrobi että metaboliitti koostumukset kerättiin osana Earth Microbiome 500 -projektia. Lyhyesti, mikrobiota tunnistettiin käyttämällä 16S rRNA v4 amplikoni sekvensointia, saatu data käsiteltiin QIIME2:lla ja analysointi suoritettiin PHYLOSEQ ja VEGAN ohjelmapaketeilla. Metaboliitit määriteltiin kohdistamattomalla metabolomiikalla (ultra-korkean suorituskyvyn nestekromatografia yhdistettynä massaspektrometriin), saatu data käsiteltiin MZmine ohjelmalla ja analysointi suoritettiin PHYLOSEQ ja VEGAN ohjelmapaketeilla. Havaitsin, että (1) säteily vaikutti suolistomikrobiston alfa ja beeta diversiteetteihin, mutta ei metaboliittien ja (2) näillä muutoksilla ei ollut assosiaatiota keskenään. Nämä tulokset vahvistivat aikaisempien tutkimusten tuloksia sekä havainnollistivat suolistomikrobiston ja metaboliittien välistä kompleksia suhdetta.

TABLE OF CONTENTS

TERMS AND ABBREVIATIONS

Terms

ALLAS	The object storage service provided by CSC.
Feature	A unique amplicon sequence variant or mass-to-charge profile of metabolite.
Functional redundancy	An ecological phenomenon where various species are filling similar, or even identical, roles in the same ecosystem functionality.
Taxa	Groups of one or more organism populations are classified together by their features.

Abbreviations

CEZ	Chornobyl Exclusion Zone
CH	contaminated sampling area in CEZ
CL	uncontaminated sampling area in CEZ
EIC	extracted ion chromatogram
EMP500	the Earth Microbiome Project 500
LC	liquid chromatography
KL	uncontaminated area near Kyiv
MC	mass spectrometry
ROS	reactive oxygen species
SV	sequence variant

1 INTRODUCTION

Exposure to an increased level of ionising radiation is often damaging to living organisms. Radiation can cause damage to the cell membrane and single- and double-stranded breaks in DNA, which both are regarded as lethal or leading to the events causing apoptosis or necrosis, with mitosis-linked death being the most numerous ones causing apoptosis (Cohen–Jonathan et al. 1999). Radiation also causes increased activity of macrophages soon after a single dose of 30 Gy, which can increase the amount of reactive oxygen species (ROS), which can lead to considerable damage to the structure of the cell (Kim et al. 2014). Additionally to direct injuries in cells after ionising radiation causing oxidative damages by reactive oxygen species (ROS), induced gene expression can be thought to be an indirect injury in the cells, which further can increase the amount of proinflammatory cytokines, also affecting the amount of reactive oxygen species (ROS), and inflammations leading to the death of the cell (Kim et al. 2014).

The universe has natural radiation, but the Earth's ozone layer and magnetosphere protect Earth's organisms from cosmic radiation (Mousseau 2021). However, there are places in the world where radioactive nuclides originate from anthropogenic events such as nuclear bombs, their testing, and nuclear power plant accidents (Mousseau 2021). The immediate effects of atomic bombs on humans have been relatively well studied (Mousseau 2021). However, the long-term consequences of a constant low level of radiation still require further research, as does its impact on the biosphere (Mousseau 2021).

On April 26 1986, an accident at reactor 4 of the former nuclear power plant at Chornobyl, Ukraine, led to the release of about 9 million terabecquerels of radionuclides into the atmosphere, with the fallout largely deposited in Eastern Europe, Western Russia, and Fenno-Scandinavia (Møller and Mousseau 2006). To limit human exposure to high concentrations of persistent environmental radionuclides (notably Strontium-90 (half-life of 29 years), Cesium-137 (half-life of 30 years), and Plutonium-239 (half-life of 24,000 years) (Møller and Mousseau 2015)), the Chornobyl Exclusion Zone (CEZ) was established at an approximately 30 km radius around the site of the accident (Møller and Mousseau 2006). Wildlife inhabiting the CEZ provide the best-studied natural models of the biological impacts of exposure to elevated levels of ionising radiation derived from environmental radionuclides (Mousseau 2021).

A hazardous environment can affect the functional pathways of the microbes, and the radiation can be associated with changes in the host's gut microbe composition (Lavrinienko et al. 2018a). A several studies study shows that bank voles' gut microbiome is affected by environmental radiation by changing composition (Lavrinienko et al. 2018a, 2018b, 2021), but there is also a study where the effect was not discovered when using small sample size and no replication (Antwis et al. 2021). However, it is unknown if this apparent impact of radiation exposure has any health consequences, for example, if the change in

the bacterial community affects the functional roles of the gut microbiome but not the functional outcome, the metabolites. This phenomenon, where the changes in microbiota composition do not affect the function as taxa closely related to each other usually have similar functions, i.e. similar metabolic e.g. many bacteria belonging in phylum of Firmicutes participates in dietary fibre fermentation (Sun et al. 2023), is known as functional redundancy.

Functional diversity of microbial communities is complex, and it can be divided into two perspectives: measuring the traits at taxa level or community level (Escalas et al. 2019). The measurements that bases on the traits studied at taxa-level resembles the approach used in macrobial functional studies. Unravelling the functional niches of taxa enables a better understanding of community dynamics when the link between the phylogeny and function, and the identification of functional redundancy is untangled. Since the introduction of next-generation sequencing methods, the genotypic trait are widely microbial trats (Escalas et al. 2019). However, functional pathways identified from metagenomes may be more relevant as functional traits (Escalas et al. 2019). To use genotypic traits as a standard for microbial functional ecology, the validation of the link between the genotypic traits and phenotypic traits need to be done. Phenotypic traits include i.e. metabolic capabilities (Barberán et al. 2017) when considering the taxa-centered perspective. The lack of the culturable representatives for most microbial groups can create its own challenges (Barberán et al. 2017). However, the current available database of taxon-traits, helps to fill the gap between the functional ecology of micro- and microorganisms (Escalas et al. 2019). Previous studies presents an example of how to reckon the functional diversity using the taxa-traits approach; the distribution of abundance across the trait and its effects on the distribution of functionality in communities can be used to cluster species according to their biological characteristics (Escalas et al. 2019).

The gut microbiota can provide many functions for its host, one of the most important of which is the production of metabolites, for example, as part of the nutrient metabolism of the host (Jandhyala et al. 2015, Andoh 2016, Lukáčová et al. 2023, Sun et al. 2023). A few studies of bank voles' gut microbiota have been revealed to have a high diversity of taxa, and environmental factors, such as pollution and radiation, can affect the gut microbiota composition (Lavrinienko et al. 2018a, Brila et al. 2021), yet it is still unknown if a change in microbiota community associates with change in function (provision of metabolites) that could impact the health of the host. However, recent advances in untargeted metabolomics using standardised protocols and analytical methods enable quantifying bacteria diversity association with metabolic diversity, i.e. metabolite production (Shaffer et al. 2022).

The effect of environmental changes on the phylogenetic and functional alpha diversity differs depending on the studied species and composition of the gut microbiota when mammals are kept in captivity (Koziol et al. 2023). Metabolic diversity studies have been done in humans and models, but in wildlife there are very few studies. For wildlife, the contribution of phylogenetic

diversity to metabolic diversity still needs to be studied more, as well as the contribution of rare taxa to metabolic diversity. Taxa from the same family have similar or identical functions, and as taxa get less related to each, i.e. are phylogenetically farther apart other, the function differs more. Because environmental radiation alters the microbiota composition (Lavrinienko et al. 2018a, b, 2021), the differences should be seen when considering the effect of the contaminated environment on the phylogenetic and functional diversities compared to the uncontaminated environment. Whether the amount of rare taxa changes and how this affects the metabolic composition (e.g. changes in the outliers of metabolic data) needs to be studied.

In this thesis, the bank vole (*Clethrionomys glareolus*) is used as a model species for studying the effects of radiation to the gut microbiota and metabolism. The bank vole is a small rodent that inhabits the mixed woodlands of much of northern Europe and Asia (Amori 2008). While bank voles inhabiting contaminated areas within the CEZ show no clear evidence of acute health impacts (such as an increase in DNA damage or mutation rate (Kesäniemi et al. 2018, 2019)), exposure to radionuclides tends to associate with poor health. For example, bank vole populations in contaminated areas within the CEZ exhibit an increase in frequency of cataracts (Lehmann et al. 2016). Also, bank voles exposed to radionuclides show signs of metabolic remodelling (Kesäniemi et al. 2019), an increase in damage to telomeres and mitochondrial DNA (Kesäniemi et al. 2019b) and changes in genome content of satellite DNA and the ribosomal RNA cassette (Jernfors et al. 2021). An impact of radionuclide exposure on the condition of animal digestive tracts also seems possible because the gut microbiota of bank voles inhabiting areas contaminated by radionuclides differs to the gut microbiota of animals that are not exposed to radionuclides (Lavrinienko et al. 2018a, 2020). This association between exposure to radionuclides and a change in gut microbiota occurs in other rodents (such as species of *Apodemus* mice) inhabiting the CEZ (the external dose rate 0.313 - 0.498 mGy day^{-1} in contaminated area) and at the site of the Fukushima nuclear accident in Japan (the external dose rate 0.174 - 0.233 mGy day^{-1} in contaminated area) (Lavrinienko et al. 2021).

The thesis has two general aims: (1) to determine whether a change in gut microbiota associated with exposure to elevated levels of radiation (via exposure to environmental radionuclides) elicits a corresponding change in the diversity of faecal metabolites, and (2) to quantify what aspects of gut microbiota (e.g. species diversity or phylogenetic diversity), if any, can best explain variation in metabolite diversity.

As the functional redundancy of the gut microbiota supports the host to tolerate to changes in the environment, for example, by securing meaningful metabolic interactions (Koziol et al. 2023), I hypothesise that differences in bank vole gut microbiota are not associated with the change in function (production of metabolites). The less related taxa are to each other, the more their functions differ, which is why phylogenetic diversity should have more effect on metabolite diversity than species diversity. Since previous studies have shown

that radiation alters the composition of the gut microbiota, and since the production of metabolites is one of its primary functions, I hypothesized that bank voles' exposure to elevated environmental values also affects the diversity of metabolites.

I will address my aims and test my hypothesis using samples of bank vole faeces collected from animals inhabiting contaminated and uncontaminated areas as a model. The samples were collected and processed according to the protocol of Earth Microbiome Project 500 (EMP500) (Shaffer et al. 2022, https://earthmicrobiome.org/emp500/). These samples provided data for 16S amplicon sequencing and untargeted metabolomics. I conduct Shannon diversity, Jaccard distance, and Bray-Curtis dissimilarity tests to inspect microbiota and metabolite data changes in environment with elevated levels of ionising radiation. I found that increased levels of environmental radiation increased the variation in the diversity of taxa, which strengthened earlier knowledge of radiation's effect on the gut microbiota (Lavrinienko et al. 2018a, b, 2021, Jernfors et al. 2024) as well as contradicted some other results (Antwis et al. 2021). The increased levels of environmental radiation did not significantly decrease the variation in metabolite diversity. Additionally, the Jaccard distance and the Bray-Curtis dissimilarity were used to inspect the correlation of metabolite and bacteria data. These led to the discoveries that pairwise dissimilarities among the microbiota and metabolite samples did not correlate. These results support the earlier knowledge of effect of radiation on gut microbiota. However, clear evidence of functional redundancy was not discovered.

2 METHODS AND MATERIALS

2.1 Sample collection and preparation

Faecal samples of bank voles (*Clethrionomys glareolus*) were collected from contaminated areas (CH1 and CH2) in the Chornobyl Exclusion Zone (CEZ) with higher radiation dose rates (mean of 30.1 μSv/h) and uncontaminated areas (CL1 and CL2, a mean dose of 0.25 μSv/h) within the CEZ (Figure 1A) and near Kyiv (KL, a mean does of 0.33 μSv/h) with lower radiation dose rates by Lavrinienko et al. during May to July 2016 (Lavrinienko et al. 2018a). In context, safety limits for workers who expose to radiation is 20 mSv/year and 1 mSv/year for general public (Niu 2011). Measured ambient radiation levels at each trapping location were measured with a hand-held Geiger-Mueller dosimeter (Inspector, International Medcom Inc., CA, USA), placed one cm above the ground (Lavrinienko et al. 2018a). Ambient dose rate measurements give a reasonable approximation of the external absorbed dose rate for bank voles (Beresford et al. 2008).

Bank voles were live-captured using Ugglan Special2 live traps from mixed forest habitats from three study areas in Ukraine: the (1) east and (2) west side of

Dnieper river near Kyiv, and from (3) Gluboke lake and (4) Vesnyane within the CEZ, henceforth referred as uncontaminated (areas 1 and 2) and contaminated (areas 3 and 4) groups (Lavrinienko et al. 2018a). At each location, 16 traps were placed in a grid with an inter-trap distance of 20 m up to three trapping nights per trapping site. Traps were checked the following morning, and caught animals were transferred to the laboratory for internal absorbed dose rate estimation, body size measurement (body weight and head width), and faecal sampling. The sampling was performed within three weeks to reduce the effect of seasonal variation (Maurice et al. 2015). For this study, faecal metabolite data from sampling sites CH3 and KL2 from Lavrinienko et al. (Lavrinienko et al. 2018a) were not included as they were not included in archived metadata (Supplementary 1, available in the ALLAS, https://docs.csc.fi/data/Allas/) (see Figure 1A for the sampling sites).

Faecal samples for metabolomics and microbiota analysis were collected from live animals immediately after capture (see (Lavrinienko et al. 2018a)). 112 samples, including males and females, were collected, with 88 of these samples frozen (-80°C) and the remaining 24 samples stored in ethanol (Lavrinienko et al. 2018a). All faecal samples were divided into two aliquots, stored on dry ice and then archived at -80°C. One aliquot of each sample was sent (on dry ice) to laboratory in University of California SD, US, for faecal metagenomics and untargeted metabolomics (see Thompson et al. 2018 and Shaffer et al. 2022 for methods).

2.2 Sample analysis

Bank vole faecal microbiota were characterised using 16S amplicon sequencing that targeted the variable region four (v4) of the bacterial 16S rRNA locus. Briefly, faecal samples were sent to the EMP500 (https://earthmicrobiome.org/emp500/) following the standard submission protocol (Thompson et al. 2018) where they were processed using standardised, published laboratory methods (Thompson et al. 2018, Shaffer et al. 2022) (that included two rounds of DNA extraction and sequencing per sample to ensure sufficient coverage for all submitted samples). The primers 515F (Parada et al. 2016) (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (Apprill et al. 2015) (5'-GGACTACNVGGGTWTCTAAT-3') were used for PCR (that included Illumina adapters, barcodes as described in (Caporaso et al. 2011, 2012). Each 25 µl PCR contained Platinum Hot Start PCR master mix (0.8X final reaction concentration), 0.2 µM each primer, and 1 µl of template DNA. Thermal cycling conditions (on 384-well plates) were 94 °C, 3 min, followed by 35 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 105 s, and then 72 °C for 10 min. All PCRs were completed in triplicate 25 µl reactions that were then pooled, cleaned using an UltraClean PCR Clean-Up kit (QIAGEN) following the manufacturer's instructions, quantified to check DNA quality and quantity, and then sequenced using Illumina chemistry.

The samples for untargeted metabolomics were submitted by Lavrinienko et al. (Lavrinienko et al. 2018a) to the EMP500, and the metabolite data were

obtained using liquid chromatography with tandem mass spectrometer (LC-MS/MS) (Liu and Locasale 2017) as described in the EMP500 (Shaffer et al. 2022, https://earthmicrobiome.org/emp500/) (Supplementary 1, available in the ALLAS, https://docs.csc.fi/data/Allas/). Briefly, the samples were extracted, subjected to solid phase extraction (SPE) and resuspended for mass spectrometry. Ultra-high performance liquid chromatography (UHPLC) (Vanquish, Thermo Fisher) coupled to a quadrupole-Orbitrap mass spectrometer (Q Exactive, Thermo Fisher) was used to obtain mass spectra (Shaffer et al. 2022). The metabolite data for bank voles were archived (Supplementary 1, available in the ALLAS, https://docs.csc.fi/data/Allas/) and available for analysis.

2.3 Sample design

To determine the effect of radiation dose on microbiota and metabolite diversity, samples were divided into three categories based on the radiation levels and sampling location (Figures 1A and 1B). These three categories were named as a treatment_1 and treatment_2 in metadata (Appendix 1 and 2). Treatment_1 consisted of three treatment groups based on those three categories. Treatment group "hot" consisted of samples from contaminated sampling sites CH1 and CH2, "clean" of samples from uncontaminated sampling sites CL1 and CL2, and "control" of samples from uncontaminated sampling site KL. Treatment_2 consisted of uncontaminated (uncont.) samples (from sampling sites CL1, CL2, and KL) and contaminated (cont.) samples (from sampling sites CH1 and CH2). (1) contaminated samples (from sampling sites CH1 and CH2) had elevated dose of ionising radiation (mean of 30.1 μSv/h) and were sourced from the CEZ, while the uncontaminated samples were derived from (2) the CEZ (sample sites CL1 and CL2, with a mean dose of 0.25 μSv/h) or (3) near Kyiv (sample site KL, with a mean dose of 0.33 μSV/h) (Lavrinienko et al. 2018a). The estimated radiation doses of the uncontaminated samples were close to natural background levels of <0.1 μSv/h (Mousseau 2021). Variation among (1) CH and other sites can be attributed to effects of radiation dose, whereas variation between CH/CL and KL are likely due to a marked difference in geographic location.

2.4 Methods

2.4.1 Detection of metabolomic features

To identify the metabolites from the LC-MS/MS data, I processed the mass-spectra using MZmine (3.9.0) (Schmid et al. 2023), using best practice to remove features associated with laboratory reagents (i.e. metabolites that may be derived from the plastic/glassware, or the extraction process). See Appendix 3 and 4 for parameters used and full details of used modules of each step-in sample processing, respectfully. Briefly, mass detection was performed to define the threshold value to filter out the "noise", i.e. the impurities from the raw data. The noise level was visually set to filter the background noise from the MS level 1

scans. The second mass detection was performed to filter out the background

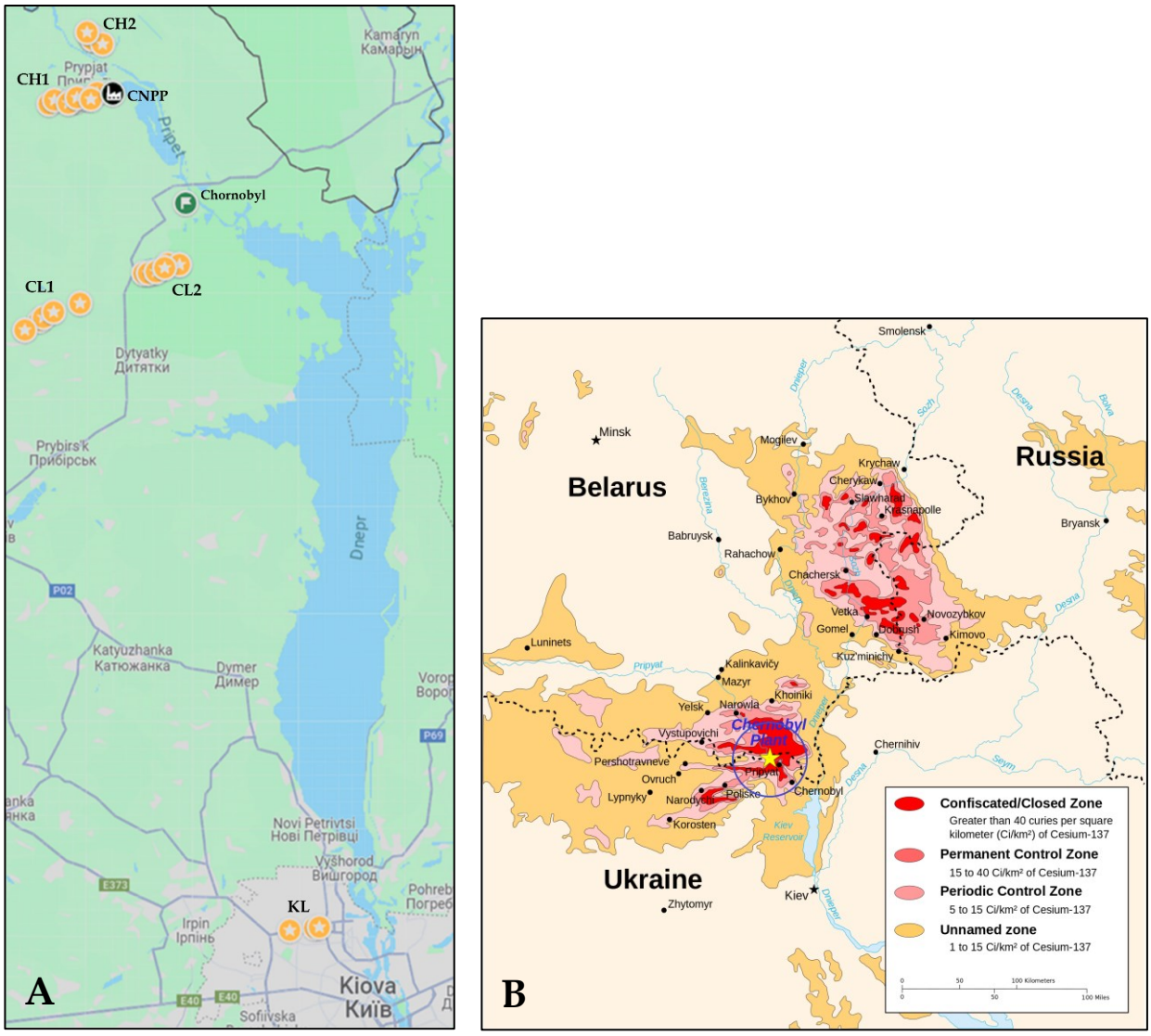

Figure 1. Geographical maps of sampling locations within and outside Chornobyl Exclusion Zone (CEZ). A. The bank voles' (*Clethrionomys glareolus*) faecal samples' collection location around the Chornobyl Nuclear Power Plant (CNPP) within the Chornobyl Exclusion Zone. Faecal samples were collected from trapped bank voles from contaminated areas (CH1 and CH2) with elevated radiation levels and from uncontaminated areas (CL1 and CL2) within the Chornobyl Exclusion Zone. Faecal samples was also collected from uncontaminated areas (KL) outside the Chornobyl Exclusion Zone and near Kyiv (Kiova). Map data ©2024 Google. The image was modified by adding dark black texts and editing the CNPP icon afterwards. B. Chornobyl radiation map. Chornobyl Exclusion Zone is marked as a blue circle. Chornobyl power plant is marked as a star. Red to darker orange areas indicates the amount of Cesium-137 radiation. "File: Chernobyl radiation map 1996 30km zone.png" by CIA Factbook, Sting (vectorisation), MTruch (English translation), Makeemlighter (English translation), 四葉亭四迷 is licensed under CC BY-SA 2.5. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/2.5/?ref=openverse.

noise from the MS level 2 scans. The second mass detection's noise level was also defined visually. The first and the second mass detection was done for each raw data file.

The mass-detected data was used to build an extracted ion chromatogram (EIC) for each mass-to-charge (m/z) value for each raw data file using the ADAP chromatogram builder (Myers et al. 2017) module. The parameter for minimum consecutive scans was determined by inspecting the raw data's usual minimum number of data points of the LC peaks.

The local minimum resolver module was used to split the "shoulders" of LC peaks into individual features using appropriate parameters determined by inspecting apparently noisy EICs and good EICs in a few feature lists. The parameters were set to obtain the feature list of LC peaks with the minimum amount of noise but with as many separated peaks as possible (Appendix 3).

Next, a 13C isotope filter (isotopic grouping) module was performed to filter out the features corresponding to 13C isotopes of the same analyte. Then, an isotope pattern finder (isotopic peak finder) module was performed to search for isotope patterns in the feature list. Following this, the join aligner module was used to align the detected peak in the samples through a match score (based on the mass and retention time of each peak and the mass spectrometer's tolerance range). These modules produced a feature list, from which apparent contaminants could be subtracted (based on their appearance in blank samples), and also fill the gaps in the aligned feature table, remove the misaligned feature list rows and finally, to obtain the feature table.

The features were grouped based on retention time, feature shape, and feature height correlation using the metaCorrelate feature grouping module (Schmid et al. 2021, 2023). Metabolomics data processing resulted in a total of 15,778 features in all 91 metabolomics samples which included the LC-MS/MS information (Supplementary 2, available in the ALLAS https://docs.csc.fi/data/Allas/)

2.4.2 16S microbiota feature detection

Reads for the EMP500 data were demultiplexed, trimmed to 150 bp and denoised using the default parameters in DEBLUR (Amir et al. 2017) to generate feature tables. Feature tables are archived in QIITA (study: 13114) (https://qiita.ucsd.edu/) (Gonzalez et al. 2018).

The feature tables that contained data from bank voles were obtained from QIITA (https://qiita.ucsd.edu/) and merged using QIIME2's (2023.5) (Estaki et al. 2020) '*feature-table*' plugin, with the '*filter-features*' option used to retain data for bank voles, as well as the negative (extraction and run) controls used for each flow cell (n=185 samples in total). These data provided between 9,240 and 97,927 reads per sample. Taxonomy was assigned to features using the naive Bayes classifier (Bokulich et al. 2018) implemented by the QIIME2's *feature-classifier* (*classify-sklearn*) plugin trained against the SILVA (139_99) database of 16S sequences (Quast et al. 2013, https://www.arb-silva.de/) that had been aligned

and trimmed to the v4 region of the rRNA locus (i.e. to match the 515 and 806 primer pair). The feature table was then filtered (*taxa filter-table* plugin) to remove all features that were not assigned a bacterial taxonomic classification (i.e. as either *Archaea*, *Eukaryota*, mitochondria, chloroplast, or unassigned). After this, a midpoint rooted phylogenetic tree was constructed using FASTREE (*phylogeny align-to-tree-mafft-fasttree* plugin) and the (i) feature table, (ii) phylogenetic tree, (iii) taxonomic classification, and (iv) metadata were exported to PHYLOSEQ (McMurdie and Holmes 2013). Next, possible contaminants were removed using DECONTAM (Davis et al. 2018). From all the microbiota samples 27,201 features were detected (Supplementary 3, available in the ALLAS https://docs.csc.fi/data/Allas/).

2.5 Statistical analysis and data visualization

Both microbiota (bacteria) and metabolomics data were analysed in R (4.2.1) statistical software (R Core Team 2021) using PHYLOSEQ (McMurdie and Holmes 2013) as a part of the vegan package (2.6.2) (Oksanen et al. 2022). See all used packages and version from Appendix 5. R was used via CSC – IT center for Science (https://csc.fi/en/).

2.5.1 Statistical analysis and data visualization of metabolomics data

First, the metabolite data (91 samples) was normalized for statistical analysis and data visualization using log-transformation (FENG et al. 2014). The Shannon entropy was used as an unweighted method to calculate the differences in the composition of metabolites between different contaminations (dose of radiation) within individual bank vole within three samples groups (clean, control and hot) (see Supplementary 4, available in the ALLAS, https://docs.csc.fi/data/Allas/). Statistical Bartlett's test was used to determine homogeneity of variances. Observed features were used as a weighted method to calculate the number of different metabolites present. The pairwise dissimilarities among the samples in each feature table were calculated using the binary Jaccard metric and weighted Bray-Curtis metric to examine the importance of feature (i.e. metabolite) abundance. The effect of dose on gut metabolites was determined using permutation analysis of variance (PERMANOVA) (Stevens 2019).

2.5.2 Statistical analysis and data visualization of bacteria data

First, the bacteria data (111 samples) were normalized for statistical analysis and data visualization by rarefying the data using an even depth of 10,000 sequences without replacement. The Shannon diversity index was used as an unweighted method to calculate the differences in the composition of bacteria between different treatments (dose of radiation) within individual bank vole (see Supplementary 4 and Supplementary 5, both available in the ALLAS, https://docs.csc.fi/data/Allas/). Statistical Bartlett's test was used to determine homogeneity of variances. Observed features were used as a weighted method

to calculate the number of different bacteria present. Statistical tests Kruskal-Wallis and Dunn's test were used for the non-parametric analyses to assess the difference between the treatments. The pairwise dissimilarities among the samples in each feature table were calculated using an unweighted Jaccard metric to examine the importance of bacteria abundance. The unweighted UniFrac metric (Lozupone et al. 2011) was used to account for phylogenetic differences among communities among pairs of samples for gut bacteria informs about the relative importance of rare (versus abundant) taxa and phylogenetic differences in gut bacteria community composition. The effect of dose on gut microbiota was determined using permutation analysis of variance (PERMANOVA) (Stevens 2019).

2.5.3 Identifying best predictors of metabolic diversity

First, the bacterial data samples without the sample equivalent of the metabolite data were filtered out. This produced 91 samples for both bacteria and metabolite data. Two outliers, sample_040 and sample_043, were removed from the data to obtain the easier to read figure, so analysed data composed of 89 samples in both data. Microbiome (11.8.0), vegan (2.6.2) and ggplot2 (3.3.6) were used as the generalised linear models (GLMs) in R (4.2.1) to quantify the strength of the relationship (e.g. significance and effect size) between corresponding pairwise dissimilarities for bacteria and metabolites (see Supplementary 4, available in the ALLAS, https://docs.csc.fi/data/Allas/). The metabolites and bacteria data were not normalized for the comparing data using beta diversity. The correlation plot of unweighted Jaccard metric and weighted Bray-Curtis metric was used to examine if the changes in gut bacteria diversity could explain the changes in metabolite diversity. The statistical significances were examined using Mantel test.

3 RESULTS

3.1 Radiation's effect to the microbiota and metabolites alpha diversity

3.1.1 Metabolites alpha diversity

Radiation did not significantly impact on metabolite alpha diversity (Bartlett's test, $p = 0.806$) (Table 1). Contaminated samples had higher mean alpha diversity of metabolites than uncontaminated samples (Figure 2A). However, contaminated samples had the slightest variation in alpha diversity. Uncontaminated samples within CEZ had more variation in alpha diversity than contaminated samples but smaller mean. Uncontaminated samples within CEZ samples also had less variation in alpha diversity than uncontaminated samples

outside of CEZ. Uncontaminated samples outside of CEZ had the most comprehensive variation in diversity and the most minor mean diversity. The variety and abundance of metabolites within the samples was negatively affected by the radiation. However, radiation had no statistically significant impact samples grouped by radiation levels ($p = 0.806$) (Table 1) when calculated using Bartlett's test, suggesting that the variances are likely to be homogeneous across the treatment groups (clean, control, and hot).

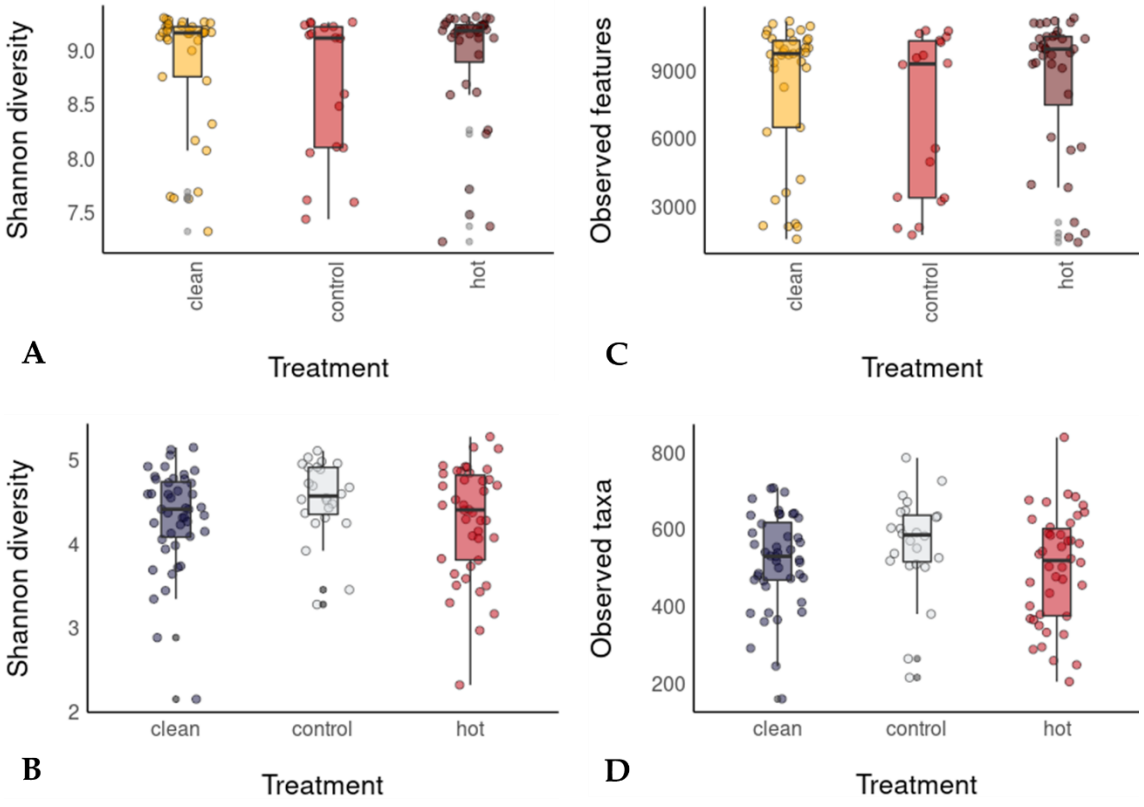

Figure 2. Statistical tests' results from metabolite and bacteria data. The clean group presents bank voles' (*Clethrionomys glareolus*) faecal samples collected from sampling locations from uncontaminated area within the Chornobyl Exclusion Zone (CL1 and CL2). The control group presents bank voles' faecal samples collected from sampling locations from uncontaminated areas outside of the Chornobyl Exclusion Zone and near Kyiv (KL). The hot group presents bank voles' faecal samples collected from contaminated area with elevated radiation levels within the Chornobyl Exclusion Zone (CH1 and CH2). A. The Shannon diversity of metabolites within individual bank voles in three different treatment group. B. The Shannon diversity of bacteria taxa within individual bank voles in three different treatment group. C. The number of different metabolites present within the bank voles' faecal samples in three treatment group using Observed method. D. The number of different bacteria taxa present within the bank voles' faecal samples in three treatment group using Observed method.

Radiation decreased the number of observed features (metabolites). Contaminated samples had the slightest variation in observed features within the samples but a higher amount than in uncontaminated samples (Figure 2C). Contaminated samples also had the highest mean of observed features. Uncontaminated samples within CEZ had higher variation in observed features but fewer than contaminated samples. Uncontaminated samples outside of CEZ have the highest variation in observed features but the most minor mean diversity. It can also be noticed that contaminated and uncontaminated samples within CEZ did have more metabolites within the samples (often more than 9000) than the uncontaminated samples outside of CEZ, where the number of metabolites within the samples was not weighted to high amounts but distributed more evenly, i.e. there was more variation in the number of features.

3.1.2 Bacteria alpha diversity

Radiation did not significantly affect bacteria alpha diversity when using the Shannon diversity index (Figure 2B) (Bartlett's test, $p = 0.184$) (Table 1). The differences within the contaminated and uncontaminated samples were notable. The contaminated samples had the most notable variation, and the uncontaminated samples outside of CEZ had the lowest. The mean alpha diversity of all samples groups were quite similar to each other. The variation in alpha diversity within the uncontaminated samples within CEZ and contaminated samples differs more than in the uncontaminated samples outside of CEZ. The variety and abundance of bacteria taxa within the contaminated samples was found to be positively affected by the radiation (Figure 2B).

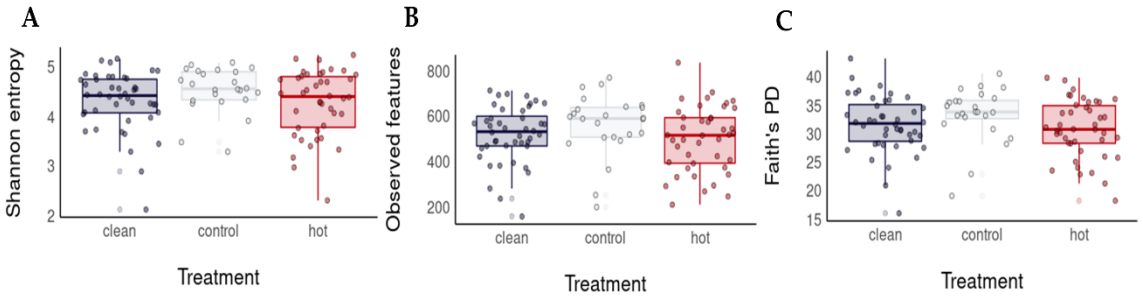

Figure 3. Alpha diversity metrics of bacteria data in phylogenetic level. The clean group presents bank voles' (*Clethrionomys glareolus*) faecal samples collected from sampling locations from uncontaminated area within the Chornobyl Exclusion Zone (CL1 and CL2). The control group presents bank voles' faecal samples collected from sampling locations from uncontaminated areas outside of the Chornobyl Exclusion Zone and near Kyiv (KL). The hot group presents bank voles' faecal samples collected from contaminated area with elevated radiation levels within the Chornobyl Exclusion Zone (CH1 and CH2). A. Shannon diversity of bacteria data in phylogenetic level. B. Observed features of bacteria data in phylogenetic level. C. Faith's phylogenetic diversity (PD) of bacteria data in phylogenetic level.

However, radiation had no statistically significant impact samples grouped by radiation levels (p = 0.184) (Table 1) when calculated using Bartlett's test, suggesting that the variances are likely to be homogeneous across the treatment groups (clean, control, and hot).

The radiation significantly increased the number of observed bacteria taxa between the uncontaminated samples outside the CEZ and contaminated samples (Dunn's test, p = 0.0491) (Table 1). The contaminated samples had the most considerable variation of observed taxa (Figure 2D); however, they also had the lowest mean of observed taxa. The uncontaminated samples outside of CEZ had the highest number of observed taxa of all the treatment groups. They also had the highest mean of observed taxa but the lowest variation. The uncontaminated samples had more outliers than the contaminated samples. The pairwise comparison test (Dunn's test) of control and hot treatment groups revealed a significant difference between the groups (p = 0.049), when significant differences between the groups clean – hot and clean – control were not observed (p = 0.566, p = 0.134, respectively).

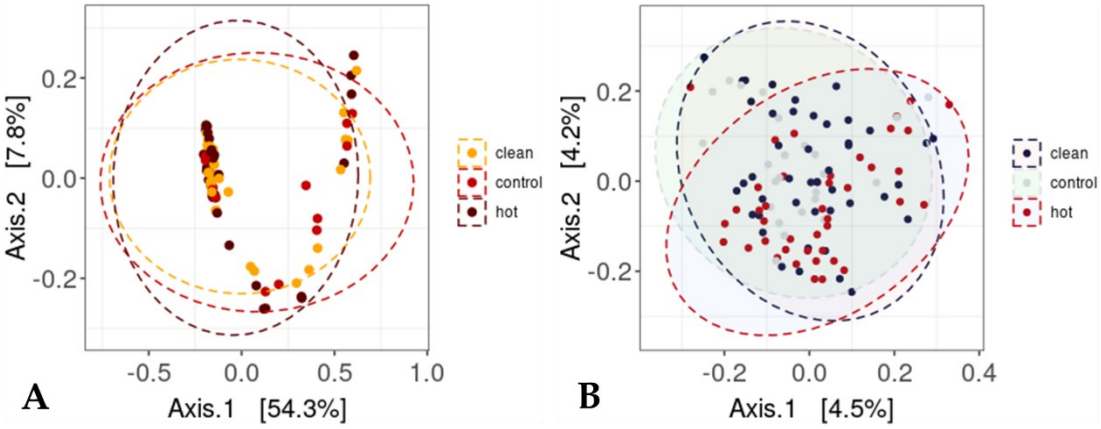

Figure 4. The pairwise dissimilarities among the samples using the Jaccard metric for metabolites (A) and bacteria taxa (B) combined with an ordination plot. The clean group presents bank voles' (*Clethrionomys glareolus*) faecal samples collected from sampling locations from uncontaminated area within the Chornobyl Exclusion Zone (CL1 and CL2). The control group presents bank voles' faecal samples collected from sampling locations from uncontaminated areas outside of the Chornobyl Exclusion Zone and near Kyiv (KL). The hot group presents bank voles' faecal samples collected from contaminated area with elevated radiation levels within the Chornobyl Exclusion Zone (CH1 and CH2).

Radiation did not have a clear effect on bacteria data when samples were examined in phylogenetic level. When comparing alpha diversity metrics Shannon diversity (Figure 3A), Observed features (Figure 3B) and Faith's phylogenetic diversity (Figure 3C), no clear differences could not be found. In all the metrics, the uncontaminated samples outside of CEZ had the highest

diversity. Uncontaminated samples outside the CEZ also differed the most from the samples collected within the CEZ when Faith's phylogenetic diversity was inspected (Figure 3C).

3.2 Radiation's effect to the microbiota and metabolites beta diversity

3.2.1 Metabolites beta diversity

There were no statistically significant differences in metabolite composition between the samples when Jaccard distance were used (PERMANOVA, p = 0.22) (Tabel 1). The uncontaminated samples were quite similar, and contaminated samples differed from uncontaminated samples when metabolites' differences between the treatment groups were compared (Figure 4A). The analysis revealed a modest effect size, with the treatment explaining approximately 2.835% of the variation in composition (R2 = 0.028), however, radiation had no statistically significant impact on Jaccard distance, with no clear clusters of samples grouped by radiation levels (p = 0.22).

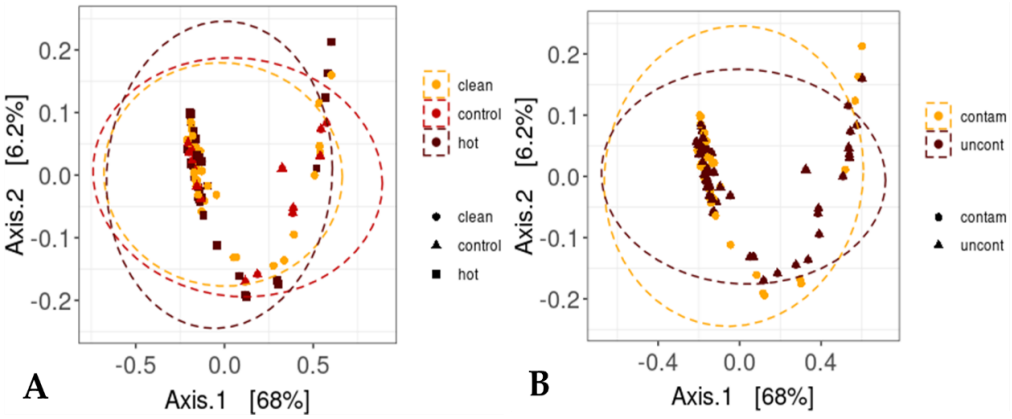

Figure 5. The pairwise dissimilarities among the samples using Bray-Curtis metric for metabolites within treatment groups clean, control and hot (A) and treatment groups contam and uncontam (B) combined with an ordination plot. The clean group presents bank voles' (*Clethrionomys glareolus*) faecal samples collected from sampling locations from uncontaminated area within the Chornobyl Exclusion Zone (CL1 and CL2). The control group presents bank voles' faecal samples collected from sampling locations from uncontaminated areas outside of the Chornobyl Exclusion Zone and near Kyiv (KL). The hot group presents bank voles' faecal samples collected from contaminated area with elevated radiation levels within the Chornobyl Exclusion Zone (CH1 and CH2). The contam presents the samples from contaminated areas (CH1 and CH2) within the Chornobyl Exclusion Zone and the uncont presents the samples from uncontaminated areas (CL1 and CL2) within the Chornobyl Exclusion Zone and outside of the Chornobyl Exclusion Zone (KL).

The radiation did not significantly affect the metabolite composition by increasing the dissimilarities among the samples when using Bray-Curtis metric (PERMANOVA, for treatment_1 group hot, clean and control, p = 0.306, for treatment_2 groups contaminated and uncontaminated, p = 0.371) (Tabel 1). The differences between the uncontaminated samples were similar, while the contaminated samples differed from the above (Figure 5A). Yet, the differences were not statistically significant, and the effect size was low (R2 = 0.026, p = 0.306). Additionally, treatment_2 group (contaminated and uncontaminated) was also examined resulting findings confirming the treatment_1 group's (Figure 5B) also with no statistically significant results (R2 = 0.010, p = 0.371).

3.2.2 Bacteria beta-diversity

The radiation had a statistically significant impact on bacteria taxa composition when dissimilarities between the samples groups hot, clean and control were calculated using Jaccard distance (PERMANOVA, R2 = 0.02715, p = 0.001) (Table 1). The contaminated samples differ greatly from uncontaminated samples

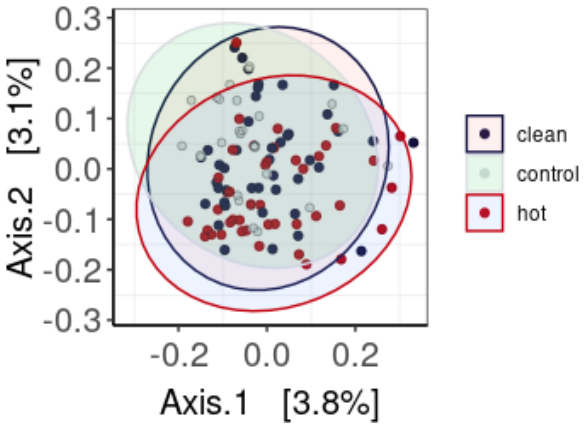

Figure 6. The UniFrac metric used to account for phylogenetic differences among communities among pairs of samples for gut bacteria in treatment groups clean, control and hot. The clean group presents bank voles' (*Clethrionomys glareolus*) faecal samples collected from sampling locations from uncontaminated area within the Chornobyl Exclusion Zone (CL1 and CL2). The control group presents bank voles' faecal samples collected from sampling locations from uncontaminated areas outside of the Chornobyl Exclusion Zone and near Kyiv (KL). The hot group presents bank voles' faecal samples collected from contaminated area with elevated radiation levels within the Chornobyl Exclusion Zone (CH1 and CH2).

(Figure 4B), and samples from uncontaminated areas within and outside the CEZ are similar to each other.

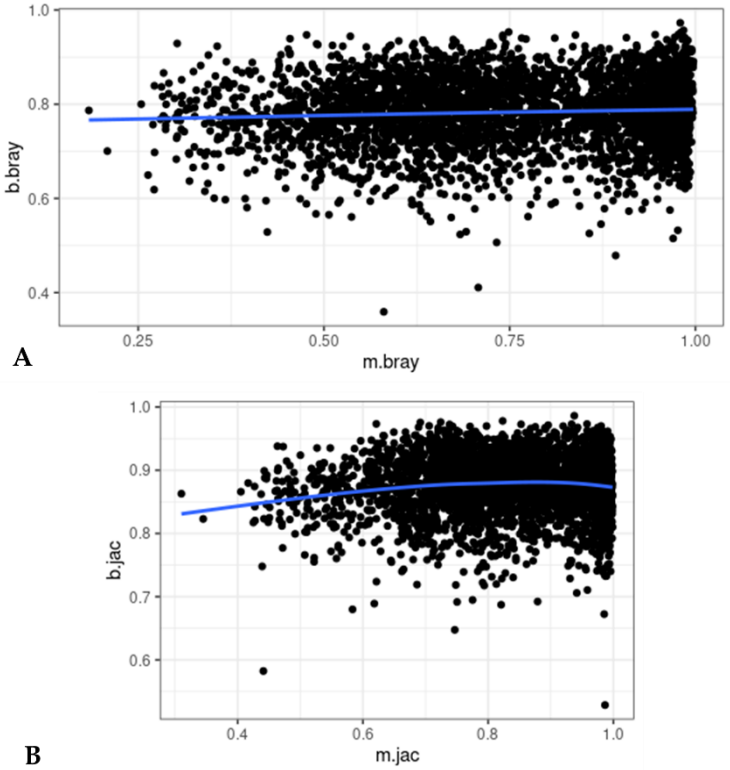

Figure 7. Correlation plot of bacteria taxa and metabolite data using Bray-Curtis metric (A) and Jaccard metric (B). Faecal samples of bank voles' (*Clethrionomys glareolus*) were collected from sampling locations from uncontaminated area within the Chornobyl Exclusion Zone and outside Chornobyl Exclusion Zone near Kyiv and from contaminated area with elevated radiation levels within the Chornobyl Exclusion Zone. Samples were used for both 16S rRNA v4 amplicon sequence data and untargeted metabolomics data. A. On x-axis m.bray present the Bray-Curtis metric of metabolite samples and on y-axis b.bray present the Bray-Curtis metric of bacteria samples. B. On x-axis m.jac present the Jaccard metric of metabolite samples and on y-axis b.jac present the Jaccard metric of bacteria samples.

The radiation had a significant effect on phylogenetic differences among communities among pairs of samples for gut bacteria in treatment groups clean, control and hot (UniFrac metric, R2 = 0.026, p = 0.001) (Figure 6). Samples within treatment groups clean and control were more phylogenetically alike than in treatment group hot, similar to the results obtained from Jaccard distance.

3.3 The features of gut bacteria best explaining the changes in metabolites

There was no significant correlation between the bacteria taxa and metabolites, when using the Bray-Curtis metric (Mantel test, p = 0.068931) (Figure 7A) or the Jaccard metric (Mantel test, p = 0.078921) (Figure 7B) (Table 1). Additionally, the effect size of Bray-Curtis (0.065) and Jaccard metrics (0.065) were low (Table 1).

Table 1. Results of statistical analysis of bank voles' faecal samples from contaminated area and uncontaminated area within the Chornobyl Exclusion Zone and from uncontaminated area outside the Chornobyl Exclusion Zone. See Appendix 1 and 2 for treatment groups explanation.

Data	Diversity	Metric	Test	Tested treatment group		Effect size	p-value
Metabolites							
	alpha diversity	Shannon	Bartlett's test	treatment_1			0.806
	beta diversity	Bray-Curtis	PERMANOVA	treatment_1		0.02624	0.306
		Bray-Curtis	PERMANOVA	treatment_2		0.00954	0.371
		Jaccard	PERMANOVA	treatment_1		0.02835	0.22
Bacteria							
	alpha diversity	Shannon	Bartlett's test	treatment_1			0.1839
		Observed	Dunn's test	treatment_1	control - hot		0.0491
					clean - control		0.134
					clean - hot		0.566
		Observed	Kruskal-Wallis test	treatment_1		0.0184	0.138
	beta diversity	Jaccard	PERMANOVA	treatment_1		0.02715	0.001
		UniFrac	PERMANOVA	treatment_1		0.02626	0.001
Bacteria compared to metabolites	beta diversity	Bray-Curtis	Mantel test			0.0649	0.068931
		Jaccard	Mantel test			0.0649	0.078921

4 DISCUSSION

Radiation can cause various issues to the hosts living in contaminated area, e.g. change the composition on microbiota community (Lavrinienko et al. 2018a, 2021) and affect metabolism pathways (Kesäniemi et al. 2019a), but the effect of these changes to the host's health is unknown. To address this knowledge gap, the bank voles (*Clethrionomys glareolus*) inhabiting the Chornobyl Exclusion Zone (CEZ) were used as a model to determine if the changes in gut microbiota diversity influences the services, e.g. the production of metabolites. The bank voles' faecal samples were collected from contaminated and uncontaminated areas within the CEZ and from uncontaminated areas outside the CEZ and near Kyiv to examine the effect of ionising radiation on metabolite and bacteria

diversities as well as if changes in bacteria diversity associate the changes in metabolite diversity. Regarding functional redundancy, if bacteria composition does not change too much at the phylogenetic level, metabolite diversity should not be affected, as bacteria taxa closely related to each other have similar functions. To see if radiation has any effect on bacteria or metabolite diversity and, more broadly, the host's health, both 16S amplicon sequencing and untargeted metabolomics (processed via LC-MS/MS) data were obtained from the faecal samples and data were analysed using QIIME2 and MZmine, both combined to R, respectively. The analysis discovered that radiation significantly affected the bacteria alpha and beta diversities, but not the metabolite alpha or beta diversities. Additionally, the association between the gut microbiota diversity and metabolite diversity was not found. These findings highlights the complexity of gut microbiota metabolite production and all the elements affecting that. Clear evidence of functional redundancy was also not discovered.

4.1 The effect of radiation on alpha and beta diversities

As exposure to ionising radiation significantly affected bacteria diversity when using alpha and beta diversity methods (Figures 2 and 4) is similar to discoveries of previous studies (Lavrinienko et al. 2018a, 2021, Zhu et al. 2021, Jernfors et al. 2024), this reinforces the knowledge provided by previous studies, that radiation does affect the microbiota composition. The effect of radiation on alpha diversity has been found in gut microbes (Lavrinienko et al. 2018a, 2021) and microbes living in the soil (Cheng et al. 2023), but there is also contradiction (Antwis et al. 2021). Here, the radiation positively correlated with the variety and abundance of bacteria within the samples (Figure 2D), which aligns with some of the results of previous studies (Lavrinienko et al. 2018b, 2021). As the uncontaminated samples were more similar to each other than contaminated samples, this suggest that radiation has more impact on the bacteria composition than the geographical location. If geographical location would affect more, the samples from CEZ would differ the samples collected outside the CEZ and near Kyiv. Radiation also impacted the similarities between the bacteria samples when considering the taxa, as the samples from contaminated areas differed from those from uncontaminated areas (Figure 4B). The results suggested, that radiation did have a greater impact on bacteria taxa composition than the geographical location. If geographical location would be more important factor, the samples collected within CEZ should differ from samples collected outside the CEZ as the sampling location are more geographically closer to each other. The impact of radiation on beta diversity was also discovered in earlier studies (Lavrinienko et al. 2018a, 2021, Zhu et al. 2021).

The effect of ionising radiation on the variety and abundance of metabolites was not discovered to be significant (Figure 2A and 2C) (Table 1). Additionally, beta diversity (Figure 4A and 5) showed no statistically significant effect of radiation. This suggests that radiation may have some impact on metabolite

diversity, but whether the metabolite production is too complex completeness or if the effect of functional redundancy is not enough to affect the metabolite production needs to be studied in the future. Faecal metabolites seem not be affected by radiation, which is consistent with the study done by Jernfors et al. (Jernfors et al. 2024). They found that metabolites (small subset of short chain fatty acids (SCFAs)) isolated from plasma were affected by radiation, which could suggest that the host itself could be more affected by the radiation than the gut microbiota. However, the host is a multicellular and more complex organism than a single-celled and rapidly dividing bacterium, so the effects of radiation may be more significant in more complex cellular processes. Tintori et al. found out that radiation did not impact on the genomes of *Oschiesus tipulae* (Tintori et al. 2023), a nematode, but mammals are more sensitive. Additionally, some bacteria taxa may be more resistant against the radiation (Confalonieri and Sommer 2011). Although my study did not identify metabolic pathways that resulted in a change in the diversity of metabolites, previous studies have shown that, the radiation affected the primary energy metabolism pathways in the livers of bank voles, e.g. via upregulation of genes participating in mitochondrial fatty acid oxidation and gluconeogenesis (Kesäniemi et al. 2019a) and altering the production of reactive oxygen species (ROS) (Kim et al. 2014). Radiation can also participate to development of metabolic syndrome (Huang et al. 2023). Studies with different results underscore the need for standardized protocols related to the study of metabolites.

4.2 Association between bacteria diversity and metabolite diversity

The changes in bacteria taxa diversity did not significantly associate with the changes in metabolite diversity. There are relatively few studies where the changes in microbiota community are compared to changes in metabolite, and even less when considering the diversity of gut bacteria and metabolite. Previous studies showed a positive correlation between the metabolite and microbe for many alpha diversity (Shaffer et al. 2022), however, the obtained results were not aligned with previous research. According to results of this thesis, the changes in gut microbiota diversity was not associated with the changes in metabolite diversity. The differences in the results were expected as the samples in Shaffer et al. study, were so different from each other and collected around the world, whereas here the comparable analysis was made with samples from a very small geographic scale. Also, Shaffer et al. separated the primary and secondary metabolites, which was not done for this thesis. In plants, primary metabolites participate to the function related to growth and secondary metabolites participate functions related to environmental interaction, such as regulating on plants defence (Erb and Kliebenstein 2020). The possible association between, for example, differences in gut microbiota phylogenetic level and changes in metabolites would be a fascinating perspective to examine. When the relationship between the gut microbiota and the host's health, possibly via

metabolites, is uncovered, or at least better understood, could information obtained from these be used, e.g. to study the health of human patients via faecal samples or to study the environment effect to various kind of species which could help to maintain and improve the global diversity of the species.

The phenomenon of functional redundancy did not appear between the bacteria and metabolites samples. As the functional diversity of microbial communities is a complex wholeness, the link between phylogeny and function and the identification of functional redundancy can be complex (Escalas et al. 2019). Also, in this study, the primary and secondary metabolites were not separated. A taxa-centered approach can help identify those phylogeny traits, including e.g. metabolism (Barberán et al. 2017). The functional diversity of a bacteria community can be reckoned using taxa-traits. The distribution of abundance across the trait values seems to affect the distribution of functionality in the community; the more uneven the distribution of abundance, the more uneven the distribution of functionality (Escalas et al. 2019). As the number of observed taxa (Figure 2D) does not correlate to the number of observed features (metabolites) (Figure 2C), i.e. as the number of taxa is increasing between the samples from uncontaminated areas (clean and control samples) and samples from contaminated areas (hot samples), but the number of features decreases between the samples from uncontaminated areas (clean and control samples) and samples from contaminated areas (hot samples), the phenomenon between the distribution of abundance and the distribution of functionality as described above seems not to occur. This highlights the perspective that both functional redundancy and metabolite production are very complex.

4.3 Limitations and future aspects

Even though the results supported other studies when considering the effect of radiation to the bacteria taxa composition, the effect of environmental changes, such as the alternation of seasons, can also affect gut microbiota rodents (Maurice et al. 2015). Furthermore, since metabolites are metabolic products formed when nutrients are broken down, the food consumed will likely also affect the composition of the metabolites (Jandhyala et al. 2015) and the diet changes during the seasons. In a contaminated area, diet sources of bank voles may also have received their share of the effects of radiation as diet sources may also have been exposed to radiation, as a result of which they may also have radiation-induced effects that may still affect the bank voles' gut microbiota, metabolism and its products, metabolites. Also, the deeper insight of host's metabolites and microbiota metabolites is needed to clear conclusions; is there difference between the metabolites produced by the host and metabolites produced by the microbiota under radioactive conditions? Additionally, as bank voles are mobile animals and are capable of covering the distance of about 1 km (Kozakiewicz et al. 2007), the change of environment can also affect the results. Due to this possible immigration between the sampling sites, there was lot of noise in the samples. Rational conclusions on the effects of environmental radiation, in

particular, would require a study to distinguish between the effects of internal and external radiation on the subjects. Obtaining even more accurate and precise results on how environmental conditions, such as radiation, affect the gut microbiota requires a more significant and more geographically extensive sampling of samples from areas with lower and higher radiation levels to reveal effective doses of radiation. It is still unclear which radiation dose does impact the gut microbiota and which dose the gut microbiota can still recover from. Additionally, the longer or shorter exposure time, the dose, and early life of bank voles should also be inspected, for example. Typically, empirical data from wild animals are noisy and various study design should be used to obtain indisputable results. One of the most significant issues affecting the results was the lack of a uniform protocol or a proven course of action for handling metabolite LC-MS/MS data so far. However, as shown in this study, the data from LC-MS/MS can be obtained and used for metabolomic studies, which sets the scene for more studies. Standards and protocol are absolute for the results of the various studies to be truly comparable. Standardised methods will hopefully be available shortly as interest in gut microbiota, the factors influencing it, and its effects on the host are on the rise.

If an association is found between the production of metabolites and the gut microbiota, this can be used, for example, for diet metabarcoding, determining various intestinal microbiota imbalances, and separating genes and functions from bacteria (metagenomics). The filling of this gap in knowledge could be utilized in both ecology and evolution studies. Additionally, these finding could later be expanded to include medicine as, e.g. treating cancers (Plaza-Diaz and Álvarez-Mercado 2023) and revealing other the unknown of microbe-metabolite-disease associations (Feng et al. 2022).

5 CONCLUSIONS

The association of changes in gut microbial diversity and the changes in metabolite diversity was studied using bank voles inhabiting Chornobyl as a model. The effect of radiation on gut microbiota and metabolite diversity was studied by analysing 16S amplicon sequence and non-targeted metabolomics data from faecal samples collected from previous studies. The results of the analysis found that radiation affected the composition of the gut microbiota, but not the metabolite diversity. In addition, it was found that radiation had more effect on changes than geographical location when considering the changes in gut microbiota diversity. In order to study the effective dose of radiation, samples should be collected from a wider geographical area from both higher and lower radiation levels to determine the effective dose of radiation. In addition, based on the samples of this study, changes in the composition of gut microbiota diversity did not associate with the changes in metabolite diversity, i.e. the gut microbiota's effect on the host's health still remains unclear. Further studies could investigate which taxa change affected the changes in which metabolites in order

to exploit the results on an even broader scale, such as the effects of environmental change on living organisms and the treatment of those with chronic intestinal diseases.

ACKNOWLEDGEMENTS

The author wishes to acknowledge CSC – IT Center for Science, Finland, for generous computational resources.

REFERENCES

Amir A., McDonald D., Navas-Molina J.A., Kopylova E., Morton J.T., Zech Xu Z., Kightley E.P., Thompson L.R., Hyde E.R., Gonzalez A. & Knight R. 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* 2: e00191-16.

Amori G. 2008. IUCN Red List of Threatened Species: Clethrionomys glareolus. *IUCN Red List Threat. Species*.

Andoh A. 2016. Physiological Role of Gut Microbiota for Maintaining Human Health. *Digestion* 93: 176–181.

Antwis R.E., Beresford N.A., Jackson J.A., Fawkes R., Barnett C.L., Potter E., Walker L., Gaschak S. & Wood M.D. 2021. Impacts of radiation exposure on the bacterial and fungal microbiome of small mammals in the Chernobyl Exclusion Zone. *J. Anim. Ecol.* 90: 2172–2187.

Apprill A., McNally S., Parsons R. & Weber L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75: 129–137.

Barberán A., Caceres Velazquez H., Jones S. & Fierer N. 2017. Hiding in Plain Sight: Mining Bacterial Species Records for Phenotypic Trait Information. *mSphere* 2: 10.1128/msphere.00237-17.

Beresford N.A., Gaschak S., Barnett C.L., Howard B.J., Chizhevsky I., Strømman G., Oughton D.H., Wright S.M., Maksimenko A. & Copplestone D. 2008. Estimating the exposure of small mammals at three sites within the Chernobyl exclusion zone--a test application of the ERICA Tool. *J. Environ. Radioact.* 99: 1496–1502.

Bokulich N.A., Kaehler B.D., Rideout J.R., Dillon M., Bolyen E., Knight R., Huttley G.A. & Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6: 90.

Brila I., Lavrinienko A., Tukalenko E., Ecke F., Rodushkin I., Kallio E.R., Mappes T. & Watts P.C. 2021. Low-level environmental metal pollution is associated with altered gut microbiota of a wild rodent, the bank vole (Myodes glareolus). *Sci. Total Environ.* 790: 148224.

Caporaso J.G., Lauber C.L., Walters W.A., Berg-Lyons D., Huntley J., Fierer N., Owens S.M., Betley J., Fraser L., Bauer M., Gormley N., Gilbert J.A., Smith G. & Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6: 1621–1624.

Caporaso J.G., Lauber C.L., Walters W.A., Berg-Lyons D., Lozupone C.A., Turnbaugh P.J., Fierer N. & Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* 108: 4516–4522.

Cheng F., Huang X., Qin Q., Chen Z., Li F. & Song W. 2023. The effect of aboveground long-term low-dose ionizing radiation on soil microbial diversity and structure. *Front. Ecol. Evol.* 11.

Cohen–Jonathan E., Bernhard E.J. & McKenna W.G. 1999. How does radiation kill cells? *Curr. Opin. Chem. Biol.* 3: 77–83.

Confalonieri F. & Sommer S. 2011. Bacterial and archaeal resistance to ionizing radiation. *J. Phys. Conf. Ser.* 261: 012005.

Davis N.M., Proctor D.M., Holmes S.P., Relman D.A. & Callahan B.J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6: 226.

Erb M. & Kliebenstein D.J. 2020. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* 184: 39–52.

Escalas A., Hale L., Voordeckers J.W., Yang Y., Firestone M.K., Alvarez-Cohen L. & Zhou J. 2019. Microbial functional diversity: From concepts to applications. *Ecol. Evol.* 9: 12000–12016.

Estaki M., Jiang L., Bokulich N.A., McDonald D., González A., Kosciolek T., Martino C., Zhu Q., Birmingham A., Vázquez-Baeza Y., Dillon M.R., Bolyen E., Caporaso J.G. & Knight R. 2020. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. *Curr. Protoc. Bioinforma.* 70: e100.

FENG C., WANG H., LU N., CHEN T., HE H., LU Y. & TU X.M. 2014. Log-transformation and its implications for data analysis. *Shanghai Arch. Psychiatry* 26: 105–109.

Feng J., Wu S., Yang H., Ai C., Qiao J., Xu J. & Guo F. 2022. Microbe-bridged disease-metabolite associations identification by heterogeneous graph fusion. *Brief. Bioinform.* 23: bbac423.

Gonzalez A., Navas-Molina J.A., Kosciolek T., McDonald D., Vázquez-Baeza Y., Ackermann G., DeReus J., Janssen S., Swafford A.D., Orchanian S.B., Sanders J.G., Shorenstein J., Holste H., Petrus S., Robbins-Pianka A., Brislawn C.J., Wang M., Rideout J.R., Bolyen E., Dillon M., Caporaso J.G.,

Dorrestein P.C. & Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* 15: 796.

Huang R., Miszczyk J. & Zhou P.-K. 2023. Risk and mechanism of metabolic syndrome associated with radiation exposure. *Radiat. Med. Prot.* 4: 65–69.

Jandhyala S.M., Talukdar R., Subramanyam C., Vuyyuru H., Sasikala M. & Reddy D.N. 2015. Role of the normal gut microbiota. *World J. Gastroenterol. WJG* 21: 8787–8803.

Jernfors T., Danforth J., Kesäniemi J., Lavrinienko A., Tukalenko E., Fajkus J., Dvořáčková M., Mappes T. & Watts P.C. 2021. Expansion of rDNA and pericentromere satellite repeats in the genomes of bank voles Myodes glareolus exposed to environmental radionuclides. *Ecol. Evol.* 11: 8754–8767.

Jernfors T., Lavrinienko A., Vareniuk I., Landberg R., Fristedt R., Tkachenko O., Taskinen S., Tukalenko E., Mappes T. & Watts P.C. 2024. Association between gut health and gut microbiota in a polluted environment. *Sci. Total Environ.* 914: 169804.

Kesäniemi J., Boratyński Z., Danforth J., Itam P., Jernfors T., Lavrinienko A., Mappes T., Møller A.P., Mousseau T.A. & Watts P.C. 2018. Analysis of heteroplasmy in bank voles inhabiting the Chernobyl exclusion zone: A commentary on Baker et al. (2017) "Elevated mitochondrial genome variation after 50 generations of radiation exposure in a wild rodent." *Evol. Appl.* 11: 820–826.

Kesäniemi J., Jernfors T., Lavrinienko A., Kivisaari K., Kiljunen M., Mappes T. & Watts P.C. 2019a. Exposure to environmental radionuclides is associated with altered metabolic and immunity pathways in a wild rodent. *Mol. Ecol.* 28: 4620–4635.

Kesäniemi J., Lavrinienko A., Tukalenko E., Boratyński Z., Kivisaari K., Mappes T., Milinevsky G., Møller A.P., Mousseau T.A. & Watts P.C. 2019b. Exposure to environmental radionuclides associates with tissue-specific impacts on telomerase expression and telomere length. *Sci. Rep.* 9: 850.

Kim J.H., Jenrow K.A. & Brown S.L. 2014. Mechanisms of radiation-induced normal tissue toxicity and implications for future clinical trials. *Radiat. Oncol. J.* 32: 103–115.

Kozakiewicz M., Chołuj A. & Kozakiewicz A. 2007. Long-distance movements of individuals in a free-living bank vole population: an important element of male breeding strategy. *Acta Theriol. (Warsz.)* 52: 339–348.

Koziol A., Odriozola I., Leonard A., Eisenhofer R., San José C., Aizpurua O. & Alberdi A. 2023. Mammals show distinct functional gut microbiome dynamics to identical series of environmental stressors. *mBio* 14: e01606-23.

Lavrinienko A., Hämäläinen A., Hindström R., Tukalenko E., Boratyński Z., Kivisaari K., Mousseau T.A., Watts P.C. & Mappes T. 2021. Comparable response of wild rodent gut microbiome to anthropogenic habitat contamination. *Mol. Ecol.* 30: 3485–3499.

Lavrinienko A., Mappes T., Tukalenko E., Mousseau T.A., Møller A.P., Knight R., Morton J.T., Thompson L.R. & Watts P.C. 2018a. Environmental radiation alters the gut microbiome of the bank vole Myodes glareolus. *ISME J.* 12: 2801–2806.

Lavrinienko A., Tukalenko E., Mappes T. & Watts P.C. 2018b. Skin and gut microbiomes of a wild mammal respond to different environmental cues. *Microbiome* 6: 209.

Lavrinienko A., Tukalenko E., Mousseau T.A., Thompson L.R., Knight R., Mappes T. & Watts P.C. 2020. Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. *Sci. Data* 7: 312.

Lehmann P., Boratyński Z., Mappes T., Mousseau T.A. & Møller A.P. 2016. Fitness costs of increased cataract frequency and cumulative radiation dose in natural mammalian populations from Chernobyl. *Sci. Rep.* 6: 19974.

Liu X. & Locasale J.W. 2017. Metabolomics: A Primer. *Trends Biochem. Sci.* 42: 274–284.

Lozupone C., Lladser M.E., Knights D., Stombaugh J. & Knight R. 2011. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5: 169–172.

Lukáčová I., Ambro Ľ., Dubayová K. & Mareková M. 2023. The gut microbiota, its relationship to the immune system, and possibilities of its modulation. *Epidemiol. Mikrobiol. Imunol. Cas. Spolecnosti Epidemiol. Mikrobiol. Ceske Lek. Spolecnosti JE Purkyne* 72: 40–53.

Maurice C.F., CL Knowles S., Ladau J., Pollard K.S., Fenton A., Pedersen A.B. & Turnbaugh P.J. 2015. Marked seasonal variation in the wild mouse gut microbiota. *ISME J.* 9: 2423–2434.

McMurdie P.J. & Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8: e61217.

Mousseau T.A. 2021. The Biology of Chernobyl. *Annu. Rev. Ecol. Evol. Syst.* 52: 87–109.

Myers O.D., Sumner S.J., Li S., Barnes S. & Du X. 2017. One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Anal. Chem.* 89: 8696–8703.

Møller A.P. & Mousseau T.A. 2006. Biological consequences of Chernobyl: 20 years on. *Trends Ecol. Evol.* 21: 200–207.

Møller A.P. & Mousseau T.A. 2015. Strong effects of ionizing radiation from Chernobyl on mutation rates. *Sci. Rep.* 5: 8363.

Niu S. 2011. Radiation protection of workers. *SafeWork Inf. Note Ser. Int. Labour Organ.*

Oksanen J., Simpson G.L., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Solymos P., Stevens M.H.H., Szoecs E., Wagner H., Barbour M., Bedward M., Bolker B., Borcard D., Carvalho G., Chirico M., De Caceres M., Durand S., Evangelista H.B.A., FitzJohn R., Friendly M., Furneaux B., Hannigan G., Hill M.O., Lahti L., McGlinn D., Ouellette M.-H., Cunha E.R., Smith T., Stier A., Ter Braak C.J.F. & Weedon J. 2022. Community Ecology Package. *2022-10-11*.

Parada A.E., Needham D.M. & Fuhrman J.A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18: 1403–1414.

Plaza-Diaz J. & Álvarez-Mercado A.I. 2023. The Interplay between Microbiota and Chemotherapy-Derived Metabolites in Breast Cancer. *Metabolites* 13: 703.

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J. & Glöckner F.O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41: D590–D596.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schmid R., Heuckeroth S., Korf A., Smirnov A., Myers O., Dyrlund T.S., Bushuiev R., Murray K.J., Hoffmann N., Lu M., Sarvepalli A., Zhang Z., Fleischauer M., Dührkop K., Wesner M., Hoogstra S.J., Rudt E., Mokshyna O., Brungs

C., Ponomarov K., Mutabdžija L., Damiani T., Pudney C.J., Earll M., Helmer P.O., Fallon T.R., Schulze T., Rivas-Ubach A., Bilbao A., Richter H., Nothias L.-F., Wang M., Orešič M., Weng J.-K., Böcker S., Jeibmann A., Hayen H., Karst U., Dorrestein P.C., Petras D., Du X. & Pluskal T. 2023. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* 41: 447–449.

Schmid R., Petras D., Nothias L.-F., Wang M., Aron A.T., Jagels A., Tsugawa H., Rainer J., Garcia-Aloy M., Dührkop K., Korf A., Pluskal T., Kameník Z., Jarmusch A.K., Caraballo-Rodríguez A.M., Weldon K.C., Nothias-Esposito M., Aksenov A.A., Bauermeister A., Albarracin Orio A., Grundmann C.O., Vargas F., Koester I., Gauglitz J.M., Gentry E.C., Hövelmann Y., Kalinina S.A., Pendergraft M.A., Panitchpakdi M., Tehan R., Le Gouellec A., Aleti G., Mannochio Russo H., Arndt B., Hübner F., Hayen H., Zhi H., Raffatellu M., Prather K.A., Aluwihare L.I., Böcker S., McPhail K.L., Humpf H.-U., Karst U. & Dorrestein P.C. 2021. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat. Commun.* 12: 3832.

Shaffer J.P., Nothias L.-F., Thompson L.R., Sanders J.G., Salido R.A., Couvillion S.P., Brejnrod A.D., Lejzerowicz F., Haiminen N., Huang S., Lutz H.L., Zhu Q., Martino C., Morton J.T., Karthikeyan S., Nothias-Esposito M., Dührkop K., Böcker S., Kim H.W., Aksenov A.A., Bittremieux W., Minich J.J., Marotz C., Bryant M.M., Sanders K., Schwartz T., Humphrey G., Vásquez-Baeza Y., Tripathi A., Parida L., Carrieri A.P., Beck K.L., Das P., González A., McDonald D., Ladau J., Karst S.M., Albertsen M., Ackermann G., DeReus J., Thomas T., Petras D., Shade A., Stegen J., Song S.J., Metz T.O., Swafford A.D., Dorrestein P.C., Jansson J.K., Gilbert J.A., Knight R., & Earth Microbiome Project 500 (EMP500) Consortium. 2022. Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. *Nat. Microbiol.* 7: 2128–2150.

Stevens M.H.H. 2019. adonis: Permutational Multivariate Analysis of Variance Using... in vegan: Community Ecology Package. *https://rdrr.io/rforge/vegan/man/adonis.html*.

Sun Y., Zhang S., Nie Q., He H., Tan H., Geng F., Ji H., Hu J. & Nie S. 2023. Gut firmicutes: Relationship with dietary fiber and role in host homeostasis. *Crit. Rev. Food Sci. Nutr.* 63: 12073–12088.

Thompson L., Ackermann G., Humphrey G., Gilbert J., Jansson J. & Knight R. 2018. EMP Sample Submission Guide.

Tintori S.C., Çağlar D., Ortiz P., Chyzhevskyi I., Mousseau T.A. & Rockman M.V. 2023. Environmental radiation exposure at Chornobyl has not systematically affected the genomes or mutagen tolerance phenotypes of local worms. *bioRxiv*: 2023.05.28.542665.

Zhu J., Sun X., Zhang Z.-D., Tang Q.-Y., Gu M.-Y., Zhang L.-J., Hou M., Sharon A. & Yuan H.-L. 2021. Effect of Ionizing Radiation on the Bacterial and Fungal Endophytes of the Halophytic Plant Kalidium schrenkianum. *Microorganisms* 9: 1050.

APPENDIX 1. THE BACTERIA METADATA

Table 2. The bacteria metadata. Sample-id is the sample identification number given for this thesis. Emp500 is the sample name for Earth Microbiome Project. Metabolite column indicates if there is a metabolomics data from particular sample. Treatment_1 indicates if the sample belongs to treatment group hot, clean or control. Treatment_2 indicates is samples was collected from contaminated or uncontaminated area. Storage indicates the storage form of the sample.

sample-id	emp500	metabolite	treatment_1	treatment_2	Storage
sample_900	13114.mousseau.88.s001	no	hot	contam	frozen
sample_040	13114.mousseau.88.s002	yes	hot	contam	frozen
sample_901	13114.mousseau.88.s003	no	hot	contam	frozen
sample_043	13114.mousseau.88.s004	yes	hot	contam	frozen
sample_051	13114.mousseau.88.s005	yes	hot	contam	frozen
sample_052	13114.mousseau.88.s006	yes	hot	contam	frozen
sample_042	13114.mousseau.88.s007	yes	hot	contam	frozen
sample_902	13114.mousseau.88.s008	no	hot	contam	frozen
sample_012	13114.mousseau.88.s009	yes	clean	uncont	frozen
sample_005	13114.mousseau.88.s010	yes	clean	uncont	frozen
sample_006	13114.mousseau.88.s011	yes	clean	uncont	frozen
sample_903	13114.mousseau.88.s012	no	clean	uncont	frozen
sample_010	13114.mousseau.88.s013	yes	clean	uncont	frozen
sample_904	13114.mousseau.88.s014	no	clean	uncont	frozen
sample_007	13114.mousseau.88.s015	yes	clean	uncont	frozen
sample_009	13114.mousseau.88.s016	yes	clean	uncont	frozen
sample_039	13114.mousseau.88.s017	yes	hot	contam	frozen
sample_003	13114.mousseau.88.s018	yes	clean	uncont	frozen
sample_905	13114.mousseau.88.s019	no	clean	uncont	frozen
sample_906	13114.mousseau.88.s020	no	clean	uncont	frozen
sample_004	13114.mousseau.88.s021	yes	clean	uncont	frozen
sample_907	13114.mousseau.88.s022	no	clean	uncont	frozen
sample_008	13114.mousseau.88.s023	yes	clean	uncont	frozen
sample_002	13114.mousseau.88.s024	yes	clean	uncont	frozen
sample_011	13114.mousseau.88.s025	yes	clean	uncont	frozen
sample_908	13114.mousseau.88.s026	no	clean	uncont	frozen
sample_909	13114.mousseau.88.s027	no	hot	contam	frozen
sample_047	13114.mousseau.88.s028	yes	hot	contam	frozen
sample_910	13114.mousseau.88.s029	no	hot	contam	frozen
sample_056	13114.mousseau.88.s030	yes	hot	contam	frozen
sample_911	13114.mousseau.88.s031	no	hot	contam	frozen
sample_060	13114.mousseau.88.s032	yes	hot	contam	frozen
sample_038	13114.mousseau.88.s033	yes	hot	contam	frozen
sample_041	13114.mousseau.88.s034	yes	hot	contam	frozen

sample_059	13114.mousseau.88.s035	yes	hot	contam	frozen
sample_044	13114.mousseau.88.s036	yes	hot	contam	frozen
sample_061	13114.mousseau.88.s037	yes	hot	contam	frozen
sample_912	13114.mousseau.88.s038	no	hot	contam	frozen
sample_054	13114.mousseau.88.s039	yes	hot	contam	frozen
sample_050	13114.mousseau.88.s040	yes	hot	contam	frozen
sample_045	13114.mousseau.88.s041	yes	hot	contam	frozen
sample_049	13114.mousseau.88.s042	yes	hot	contam	frozen
sample_055	13114.mousseau.88.s043	yes	hot	contam	frozen
sample_046	13114.mousseau.88.s044	yes	hot	contam	frozen
sample_057	13114.mousseau.88.s045	yes	hot	contam	frozen
sample_048	13114.mousseau.88.s046	yes	hot	contam	frozen
sample_053	13114.mousseau.88.s047	yes	hot	contam	frozen
sample_023	13114.mousseau.88.s048	yes	clean	uncont	frozen
sample_025	13114.mousseau.88.s049	yes	clean	uncont	frozen
sample_015	13114.mousseau.88.s050	yes	clean	uncont	frozen
sample_017	13114.mousseau.88.s051	yes	clean	uncont	frozen
sample_001	13114.mousseau.88.s052	yes	clean	uncont	frozen
sample_028	13114.mousseau.88.s053	yes	clean	uncont	frozen
sample_024	13114.mousseau.88.s054	yes	clean	uncont	frozen
sample_019	13114.mousseau.88.s055	yes	clean	uncont	frozen
sample_022	13114.mousseau.88.s056	yes	clean	uncont	frozen
sample_014	13114.mousseau.88.s057	yes	clean	uncont	frozen
sample_013	13114.mousseau.88.s058	yes	clean	uncont	frozen
sample_016	13114.mousseau.88.s059	yes	clean	uncont	frozen
sample_913	13114.mousseau.88.s060	no	hot	contam	frozen
sample_020	13114.mousseau.88.s061	yes	clean	uncont	frozen
sample_058	13114.mousseau.88.s062	yes	hot	contam	frozen
sample_018	13114.mousseau.88.s063	yes	clean	uncont	frozen
sample_021	13114.mousseau.88.s064	yes	clean	uncont	frozen
sample_508	13114.mousseau.88.s065	yes	control	uncont	frozen
sample_505	13114.mousseau.88.s066	yes	control	uncont	frozen
sample_504	13114.mousseau.88.s067	yes	control	uncont	frozen
sample_509	13114.mousseau.88.s068	yes	control	uncont	frozen
sample_502	13114.mousseau.88.s069	yes	control	uncont	frozen
sample_510	13114.mousseau.88.s070	yes	control	uncont	frozen
sample_507	13114.mousseau.88.s071	yes	control	uncont	frozen
sample_914	13114.mousseau.88.s072	no	control	uncont	frozen
sample_503	13114.mousseau.88.s073	yes	control	uncont	frozen
sample_506	13114.mousseau.88.s074	yes	control	uncont	frozen
sample_915	13114.mousseau.88.s075	no	control	uncont	frozen
sample_916	13114.mousseau.88.s076	no	control	uncont	frozen
sample_917	13114.mousseau.88.s077	no	control	uncont	frozen
sample_918	13114.mousseau.88.s078	no	control	uncont	frozen
sample_919	13114.mousseau.88.s079	no	control	uncont	frozen

sample_501	13114.mousseau.88.s080	yes	control	uncont	frozen
sample_511	13114.mousseau.88.s081	yes	control	uncont	frozen
sample_512	13114.mousseau.88.s082	yes	control	uncont	frozen
sample_514	13114.mousseau.88.s083	yes	control	uncont	frozen
sample_513	13114.mousseau.88.s084	yes	control	uncont	frozen
sample_516	13114.mousseau.88.s085	yes	control	uncont	frozen
sample_517	13114.mousseau.88.s086	yes	control	uncont	frozen
sample_515	13114.mousseau.88.s087	yes	control	uncont	frozen
sample_518	13114.mousseau.88.s088	yes	control	uncont	frozen
sample_032	13114.mousseau.88.s089	yes	clean	uncont	EtOH
sample_031	13114.mousseau.88.s090	yes	clean	uncont	EtOH
sample_029	13114.mousseau.88.s091	yes	clean	uncont	EtOH
sample_026	13114.mousseau.88.s092	yes	clean	uncont	EtOH
sample_033	13114.mousseau.88.s093	yes	clean	uncont	EtOH
sample_027	13114.mousseau.88.s094	yes	clean	uncont	EtOH
sample_920	13114.mousseau.88.s095	no	clean	uncont	EtOH
sample_030	13114.mousseau.88.s096	yes	clean	uncont	EtOH
sample_064	13114.mousseau.88.s097	yes	hot	contam	EtOH
sample_063	13114.mousseau.88.s098	yes	hot	contam	EtOH
sample_034	13114.mousseau.88.s099	yes	clean	uncont	EtOH
sample_035	13114.mousseau.88.s100	yes	clean	uncont	EtOH
sample_069	13114.mousseau.88.s101	yes	hot	contam	EtOH
sample_067	13114.mousseau.88.s102	yes	hot	contam	EtOH
sample_036	13114.mousseau.88.s103	yes	clean	uncont	EtOH
sample_037	13114.mousseau.88.s104	yes	clean	uncont	EtOH
sample_068	13114.mousseau.88.s105	yes	hot	contam	EtOH
sample_066	13114.mousseau.88.s106	yes	hot	contam	EtOH
sample_073	13114.mousseau.88.s107	yes	hot	contam	EtOH
sample_072	13114.mousseau.88.s108	yes	hot	contam	EtOH
sample_070	13114.mousseau.88.s109	yes	hot	contam	EtOH
sample_062	13114.mousseau.88.s110	yes	hot	contam	EtOH
sample_065	13114.mousseau.88.s111	yes	hot	contam	EtOH
sample_071	13114.mousseau.88.s112	yes	hot	contam	EtOH

APPENDIX 2. THE METABOLITE METADATA

Table 3. The metabolite metadata. Sample-id is the sample identification number given for this thesis. Emp500 is the sample name for Earth Microbiome Project. Metabolite column indicates if there is a metabolomics data from particular sample. Treatment_1 indicates if the sample belongs to treatment group hot, clean or control. Treatment_2 indicates is samples was collected from contaminated or uncontaminated area. Storage indicates the storage form of the sample.

sample-id	emp500	metabolite	treatment_1	treatment_2	storage
sample_001	13114.mousseau.88.s052	yes	clean	uncont	frozen
sample_002	13114.mousseau.88.s024	yes	clean	uncont	frozen
sample_003	13114.mousseau.88.s018	yes	clean	uncont	frozen
sample_004	13114.mousseau.88.s021	yes	clean	uncont	frozen
sample_005	13114.mousseau.88.s010	yes	clean	uncont	frozen
sample_006	13114.mousseau.88.s011	yes	clean	uncont	frozen
sample_007	13114.mousseau.88.s015	yes	clean	uncont	frozen
sample_008	13114.mousseau.88.s023	yes	clean	uncont	frozen
sample_009	13114.mousseau.88.s016	yes	clean	uncont	frozen
sample_010	13114.mousseau.88.s013	yes	clean	uncont	frozen
sample_011	13114.mousseau.88.s025	yes	clean	uncont	frozen
sample_012	13114.mousseau.88.s009	yes	clean	uncont	frozen
sample_013	13114.mousseau.88.s058	yes	clean	uncont	frozen
sample_014	13114.mousseau.88.s057	yes	clean	uncont	frozen
sample_015	13114.mousseau.88.s050	yes	clean	uncont	frozen
sample_016	13114.mousseau.88.s059	yes	clean	uncont	frozen
sample_017	13114.mousseau.88.s051	yes	clean	uncont	frozen
sample_018	13114.mousseau.88.s063	yes	clean	uncont	frozen
sample_019	13114.mousseau.88.s055	yes	clean	uncont	frozen
sample_020	13114.mousseau.88.s061	yes	clean	uncont	frozen
sample_021	13114.mousseau.88.s064	yes	clean	uncont	frozen
sample_022	13114.mousseau.88.s056	yes	clean	uncont	frozen
sample_023	13114.mousseau.88.s048	yes	clean	uncont	frozen
sample_024	13114.mousseau.88.s054	yes	clean	uncont	frozen
sample_025	13114.mousseau.88.s049	yes	clean	uncont	frozen
sample_026	13114.mousseau.88.s092	yes	clean	uncont	EtOH
sample_027	13114.mousseau.88.s094	yes	clean	uncont	EtOH
sample_028	13114.mousseau.88.s053	yes	clean	uncont	frozen
sample_029	13114.mousseau.88.s091	yes	clean	uncont	EtOH
sample_030	13114.mousseau.88.s096	yes	clean	uncont	EtOH
sample_031	13114.mousseau.88.s090	yes	clean	uncont	EtOH
sample_032	13114.mousseau.88.s089	yes	clean	uncont	EtOH
sample_033	13114.mousseau.88.s093	yes	clean	uncont	EtOH
sample_034	13114.mousseau.88.s099	yes	clean	uncont	EtOH
sample_035	13114.mousseau.88.s100	yes	clean	uncont	EtOH
sample_036	13114.mousseau.88.s103	yes	clean	uncont	EtOH

sample_037	13114.mousseau.88.s104	yes	clean	uncont	EtOH
sample_038	13114.mousseau.88.s033	yes	hot	contam	frozen
sample_039	13114.mousseau.88.s017	yes	hot	contam	frozen
sample_040	13114.mousseau.88.s002	yes	hot	contam	frozen
sample_041	13114.mousseau.88.s034	yes	hot	contam	frozen
sample_042	13114.mousseau.88.s007	yes	hot	contam	frozen
sample_043	13114.mousseau.88.s004	yes	hot	contam	frozen
sample_044	13114.mousseau.88.s036	yes	hot	contam	frozen
sample_045	13114.mousseau.88.s041	yes	hot	contam	frozen
sample_046	13114.mousseau.88.s044	yes	hot	contam	frozen
sample_047	13114.mousseau.88.s028	yes	hot	contam	frozen
sample_048	13114.mousseau.88.s046	yes	hot	contam	frozen
sample_049	13114.mousseau.88.s042	yes	hot	contam	frozen
sample_050	13114.mousseau.88.s040	yes	hot	contam	frozen
sample_051	13114.mousseau.88.s005	yes	hot	contam	frozen
sample_052	13114.mousseau.88.s006	yes	hot	contam	frozen
sample_053	13114.mousseau.88.s047	yes	hot	contam	frozen
sample_054	13114.mousseau.88.s039	yes	hot	contam	frozen
sample_055	13114.mousseau.88.s043	yes	hot	contam	frozen
sample_056	13114.mousseau.88.s030	yes	hot	contam	frozen
sample_057	13114.mousseau.88.s045	yes	hot	contam	frozen
sample_058	13114.mousseau.88.s062	yes	hot	contam	frozen
sample_059	13114.mousseau.88.s035	yes	hot	contam	frozen
sample_060	13114.mousseau.88.s032	yes	hot	contam	frozen
sample_061	13114.mousseau.88.s037	yes	hot	contam	frozen
sample_062	13114.mousseau.88.s110	yes	hot	contam	EtOH
sample_063	13114.mousseau.88.s098	yes	hot	contam	EtOH
sample_064	13114.mousseau.88.s097	yes	hot	contam	EtOH
sample_065	13114.mousseau.88.s111	yes	hot	contam	EtOH
sample_066	13114.mousseau.88.s106	yes	hot	contam	EtOH
sample_067	13114.mousseau.88.s102	yes	hot	contam	EtOH
sample_068	13114.mousseau.88.s105	yes	hot	contam	EtOH
sample_069	13114.mousseau.88.s101	yes	hot	contam	EtOH
sample_070	13114.mousseau.88.s109	yes	hot	contam	EtOH
sample_071	13114.mousseau.88.s112	yes	hot	contam	EtOH
sample_072	13114.mousseau.88.s108	yes	hot	contam	EtOH
sample_073	13114.mousseau.88.s107	yes	hot	contam	EtOH
sample_501	13114.mousseau.88.s080	yes	control	uncont	frozen
sample_502	13114.mousseau.88.s069	yes	control	uncont	frozen
sample_503	13114.mousseau.88.s073	yes	control	uncont	frozen
sample_504	13114.mousseau.88.s067	yes	control	uncont	frozen
sample_505	13114.mousseau.88.s066	yes	control	uncont	frozen
sample_506	13114.mousseau.88.s074	yes	control	uncont	frozen
sample_507	13114.mousseau.88.s071	yes	control	uncont	frozen
sample_508	13114.mousseau.88.s065	yes	control	uncont	frozen

sample_509	13114.mousseau.88.s068	yes	control	uncont	frozen
sample_510	13114.mousseau.88.s070	yes	control	uncont	frozen
sample_511	13114.mousseau.88.s081	yes	control	uncont	frozen
sample_512	13114.mousseau.88.s082	yes	control	uncont	frozen
sample_513	13114.mousseau.88.s084	yes	control	uncont	frozen
sample_514	13114.mousseau.88.s083	yes	control	uncont	frozen
sample_515	13114.mousseau.88.s087	yes	control	uncont	frozen
sample_516	13114.mousseau.88.s085	yes	control	uncont	frozen
sample_517	13114.mousseau.88.s086	yes	control	uncont	frozen
sample_518	13114.mousseau.88.s088	yes	control	uncont	frozen

APPENDIX 3. USED MODULES AND PARAMETERS IN MZMINE

Table 4. The used modules and parameters for detecting the features in MZmine.

Import MS data		
Advanced import: false		
MS1 detector (Advanced): false (Factor of lowest signal)		
MS2 detector (Advanced): false (Factor of lowest signal)		
Denormalize fragment scans (traps): false		
Spectral library files: []		
Mass detection		
Raw data files: 5E4_4_16_mousseau-88-s021-a02.mzML		
5D12_4_5_mousseau-88-s017-a02.mzML		
5E2_4_12_mousseau-88-s018-a02.mzML		
5A12_SPE_Blank.mzML		
5E5_4_17_mousseau-88-s010-a02.mzML		
5D11_4_4_mousseau-88-s052-a02.mzML		
5E3_4_14_mousseau-88-s007-a02.mzML		
5B10_Blank.mzML		
5D10_4_1_mousseau-88-s033-a02.mzML		
5A5_Blank.mzML		
5D5_Blank.mzML		
5E8_4_22_mousseau-88-s023-a02.mzML		
5E1_4_11_mousseau-88-s002-a02.mzML		
5E6_4_18_mousseau-88-s011-a02.mzML		
5E9_4_23_mousseau-88-s016-a02.mzML		
5E7_4_20_mousseau-88-s015-a02.mzML		
5E10_SPE_Blank.mzML		
5F2_4_27_mousseau-88-s028-a02.mzML		
5F1_4_26_mousseau-88-s004-a02.mzML		
5E11_4_24_mousseau-88-s024-a02.mzML		
5E12_4_25_mousseau-88-s034-a02.mzML		
5F4_4_30_mousseau-88-s025-a02.mzML		
5F8_4_36_mousseau-88-s040-a02.mzML		
5F3_4_28_mousseau-88-s013-a02.mzML		
5F9_4_37_mousseau-88-s005-a02.mzML		
5F6_4_32_mousseau-88-s046-a02.mzML		
5F5_4_31_mousseau-88-s009-a02.mzML		
5F7_4_35_mousseau-88-s042-a02.mzML		
5F10_4_38_mousseau-88-s036-a02.mzML		
5F11_4_39_mousseau-88-s041-a02.mzML		
5F12_4_40_mousseau-88-s044-a02.mzML		
5G1_4_41_mousseau-88-s006-a02.mzML		

5G2_4_42_mousseau-88-s047-a02.mzML		
5G3_4_43_mousseau-88-s073-a02.mzML		
5G6_4_48_mousseau-88-s066-a02.mzML		
5G7_4_49_mousseau-88-s057-a02.mzML		
5G10_4_52_mousseau-88-s080-a02.mzML		
5G5_4_46_mousseau-88-s039-a02.mzML		
5G9_4_51_mousseau-88-s074-a02.mzML		
5G8_4_50_mousseau-88-s043-a02.mzML		
5G11_4_53_mousseau-88-s069-a02.mzML		
5G4_4_45_mousseau-88-s067-a02.mzML		
5H1_4_55_mousseau-88-s030-a02.mzML		
5G12_4_54_mousseau-88-s058-a02.mzML		
5H2_4_56_mousseau-88-s059-a02.mzML		
5H3_4_57_mousseau-88-s065-a02.mzML		
5H4_4_58_mousseau-88-s062-a02.mzML		
5H5_Blank.mzML		
5H10_4_64_mousseau-88-s050-a02.mzML		
5H12_4_67_mousseau-88-s045-a02.mzML		
5H7_4_60_mousseau-88-s035-a02.mzML		
5H6_4_59_mousseau-88-s068-a02.mzML		
5H9_4_63_mousseau-88-s051-a02.mzML		
6A1_4_68_mousseau-88-s063-a02.mzML		
5H11_4_66_mousseau-88-s071-a02.mzML		
5H8_4_61_mousseau-88-s070-a02.mzML		
6A2_4_69_mousseau-88-s037-a02.mzML		
6A4_4_71_mousseau-88-s064-a02.mzML		
6A3_4_70_mousseau-88-s061-a02.mzML		
6A6_4_72_mousseau-88-s081-a02.mzML		
6A5_Blank.mzML		
6A7_4_74_mousseau-88-s056-a02.mzML		
6A10_4_77_mousseau-88-s032-a02.mzML		
6A9_4_76_mousseau-88-s054-a02.mzML		
6B1_4_80_mousseau-88-s049-a02.mzML		
6A11_4_78_mousseau-88-s055-a02.mzML		
6B4_5_2_mousseau-88-s091-a02.mzML		
6B2_4_81_mousseau-88-s053-a02.mzML		
6A8_4_75_mousseau-88-s048-a02.mzML		
6A12_SPE_Blank.mzML		
6B5_5_5_mousseau-88-s096-a02.mzML		
6B3_5_1_mousseau-88-s082-a02.mzML		
6B7_5_7_mousseau-88-s090-a02.mzML		
6B6_5_6_mousseau-88-s110-a02.mzML		
6B9_5_9_mousseau-88-s089-a02.mzML		
6B8_5_8_mousseau-88-s084-a02.mzML		
6B11_5_10_mousseau-88-s092-a02.mzML		

6B12_5_11_mousseau-88-s094-a02.mzML		
6C1_5_13_mousseau-88-s083-a02.mzML		
6C3_5_15_mousseau-88-s111-a02.mzML		
6C2_5_14_mousseau-88-s097-a02.mzML		
6B10_Blank.mzML		
6C4_5_16_mousseau-88-s099-a02.mzML		
6C6_5_18_mousseau-88-s106-a02.mzML		
6C9_5_21_mousseau-88-s103-a02.mzML		
6C10_5_23_mousseau-88-s087-a02.mzML		
6C5_5_17_mousseau-88-s100-a02.mzML		
6C12_5_26_mousseau-88-s098-a02.mzML		
6C11_5_24_mousseau-88-s093-a02.mzML		
6C8_5_20_mousseau-88-s102-a02.mzML		
6C7_5_19_mousseau-88-s085-a02.mzML		
6D2_5_28_mousseau-88-s086-a02.mzML		
6D3_5_29_mousseau-88-s109-a02.mzML		
6D1_5_27_mousseau-88-s105-a02.mzML		
6D11_5_39_mousseau-88-s101-a02.mzML		
6D4_5_31_mousseau-88-s112-a02.mzML		
6D7_5_34_mousseau-88-s088-a02.mzML		
6D9_5_36_mousseau-88-s107-a02.mzML		
6D6_5_33_mousseau-88-s104-a02.mzML		
6D8_5_35_mousseau-88-s108-a02.mzML		
6D5_Blank.mzML		
6E10_SPE_Blank.mzML		
Scan filters: MS1	level = 1	Polarity (Any)
Scan number: <not set>		
Base Filtering Integer: <not set>		
Retention time: <not set>		
Mobility: <not set>		
MS level filter: MS1	level = 1	
Scan definition:		
Polarity: Any		
Spectrum type: ANY		
Scan types (IMS): All scan types		
Mass detector: Centroid		
Noise level: 20000.0		
Detect isotope signals below noise level: false		
Chemical elements: [Element(1464767129	S:H	AN:1)
m/z tolerance: 5.0E-4 m/z or 10.0 ppm		
Maximum charge of isotope m/z: 1		
Denormalize fragment scans (traps): false		
Output netCDF filename (optional): false (null)		
Mass detection		

Raw data files: same as for the first Mass detection		
Scan filters: MS2	level = 2	Polarity (Any)
Scan number: <not set>		
Base Filtering Integer: <not set>		
Retention time: <not set>		
Mobility: <not set>		
MS level filter: MS2	level = 2	
Scan definition:		
Polarity: Any		
Spectrum type: ANY		
Scan types (IMS): All scan types		
Mass detector: Centroid		
Noise level: 1500.0		
Detect isotope signals below noise level: false		
Chemical elements: [Element(1955285241	S:H	AN:1)
m/z tolerance: 5.0E-4 m/z or 10.0 ppm		
Maximum charge of isotope m/z: 1		
Denormalize fragment scans (traps): false		
Output netCDF filename (optional): false (null)		
ADAP Chromatogram Builder		
Raw data files: same as for the first Mass detection		
Scan filters: MS1	level = 1	Polarity (Any)
Scan number: <not set>		
Base Filtering Integer: <not set>		
Retention time: <not set>		
Mobility: <not set>		
MS level filter: MS1	level = 1	
Scan definition:		
Polarity: Any		
Spectrum type: ANY		
Minimum consecutive scans: 5		
Minimum intensity for consecutive scans: 60000.0		
Minimum absolute height: 140000.0		
m/z tolerance (scan-to-scan): 0.002 m/z or 5.0 ppm		
Suffix: chromatograms		
Allow single scan chromatograms: {}		
Local minimum feature resolver		
Feature lists:		
Suffix: resolved		
Original feature list: KEEP		
MS/MS scan pairing: true		
MS1 to MS2 precursor tolerance (m/z): 0.01 m/z or 10.0 ppm		
Retention time filter: RtLimitsFilter[filter=Use tolerance	rtTolerance=0.15 minutes]	

Minimum relative feature height: true (0.25)		
Minimum required signals: true (1)		
Limit by ion mobility edges: false		
Merge MS/MS spectra (TIMS): false		
Minimum signal intensity (absolute	TIMS): false (250.0)	
Minimum signal intensity (relative	TIMS): true (0.01)	
Dimension: Retention time		
Chromatographic threshold: 0.95		
Minimum search range RT/Mobility (absolute): 0.1		
Minimum relative height: 0.0		
Minimum absolute height: 140000.0		
Min ratio of peak top/edge: 2.5		
Peak duration range (min/mobility): [0.0..3.0]		
Minimum scans (data points): 5		
13C isotope filter (formerly: isotope grouper)		
Feature lists:		
Name suffix: deisotoped		
m/z tolerance (intra-sample): 5.0E-4 m/z or 2.5 ppm		
Retention time tolerance: 0.075 minutes		
Mobility tolerance: false (null)		
Monotonic shape: false		
Maximum charge: 3		
Representative isotope: Most intense		
Never remove feature with MS2: true		
Original feature list: KEEP		
Isotope finder module		
Feature lists:		
Chemical elements: [Element(874926039	S:H	AN:1)
m/z tolerance (feature-to-scan): 0.001 m/z or 5.0 ppm		
Maximum charge of isotope m/z: 3		
Search in scans: SINGLE MOST INTENSE		
Join aligner		
Feature lists:		
Feature list name: Aligned feature list		
m/z tolerance (sample-to-sample): 0.001 m/z or 5.0 ppm		
Weight for m/z: 3.0		
Retention time tolerance: 0.2 minutes		
Weight for RT: 1.0		
Mobility tolerance: false (null)		
Mobility weight: 1.0		
Require same charge state: false		
Require same ID: false		

Compare isotope pattern: false		
Isotope m/z tolerance: 0.001 m/z or 5.0 ppm		
Minimum absolute intensity: <not set>		
Minimum score: <not set>		
Compare spectra similarity: false		
Spectral m/z tolerance: 0.001 m/z or 10.0 ppm		
MS level: 2		
Compare spectra similarity: Weighted cosine similarity		
Weights: MassBank (mz^2 * I^0.5)		
Minimum cos similarity: 0.7		
Handle unmatched signals: KEEP ALL AND MATCH TO ZERO		
Original feature list: KEEP		
Filtering feature list rows		
Feature lists:		
Name suffix: rows		
Minimum aligned features (samples): true (abs=1 and rel=0.1)		
Minimum features in an isotope pattern: false (2)		
Validate 13C isotope pattern: false		
m/z tolerance: 5.0E-4 m/z or 5.0 ppm		
Max charge: 3		
Estimate minimum carbon: true		
Remove if 13C: true		
Exclude isotopes: [Element(823043095	S:O	AN:8)]
Remove redundant isotope rows: false		
m/z: false ([100.0244..1347.8822])		
Retention time: false (null)		
features duration range: false ([0.0..3.0])		
Chromatographic FWHM: false ([0.0..1.0])		
Charge: false ([1..2])		
Kendrick mass defect: false		
Kendrick mass defect: [0.0..1.0]		
Kendrick mass base:		
Shift: 0.0		
Charge: 1		
Divisor: 1		
Use Remainder of Kendrick mass: false		
Parameter: No parameters defined		
Only identified?: false		
Text in identity: false ()		
Text in comment: false ()		
Keep or remove rows: Keep rows that match all criteria		
Feature with MS2 scan: false		
Never remove feature with MS2: true		
Reset the feature number ID: false		

Mass defect: false (null)		
Original feature list: KEEP		
Feature filter		
Feature lists:		
Name suffix: feature		
Duration: true ([0.0..3.0])		
Area: false ([0.0..1.0E7])		
Height: false ([0.0..1.0E7])		
# data points: true ([3..10000])		
FWHM: false ([0.0..2.0])		
Tailing factor: false ([0.5..2.0])		
Asymmetry factor: false ([0.5..2.0])		
Keep only features with MS/MS scan: false		
Original feature list: KEEP		
Feature list blank subtraction		
Aligned feature list:		
Blank/Control raw data files: 5A12_SPE_Blank.mzML		
5B10_Blank.mzML		
5A5_Blank.mzML		
5D5_Blank.mzML		
5E10_SPE_Blank.mzML		
5H5_Blank.mzML		
6A5_Blank.mzML		
6A12_SPE_Blank.mzML		
6B10_Blank.mzML		
6D5_Blank.mzML		
6E10_SPE_Blank.mzML		
Minimum # of detection in blanks: 11		
Quantification: Height		
Ratio type: MAXIMUM		
Fold change increase: true (3.0)		
Keep or remove features (of rows) below fold change:	REMOVE - Only keep features above fold change	
Create secondary list of subtracted features: false		
Suffix: subtracted		
Gap filling		
Feature lists:		
Name suffix: gap-filled		
Intensity tolerance: 0.05		
m/z tolerance (sample-to-sample): 0.001 m/z or 5.0 ppm		
Retention time tolerance: 0.12 minutes		
Minimum scans (data points): 1		

Original feature list: KEEP		
Duplicate feature list rows filter		
Feature lists:		
Name suffix: dup		
Filter mode: NEW AVERAGE		
m/z tolerance: 0.001 m/z or 5.0 ppm		
RT tolerance: 0.1 minutes		
Mobility tolerance: false (Mobility tolerance: 0.008)		
Require same identification: false		
Original feature list: KEEP		
Correlation grouping (metaCorrelate)		
Feature lists:		
RT tolerance: 0.08 minutes		
Minimum feature height: 50000.0		
Intensity threshold for correlation: 30000.0		
Min samples filter: Min samples in all: abs=2 and rel=0	Min samples in group: abs=0 and rel=0	Min %-intensity overlap: 0.5
Min samples in all: abs=2 and rel=0		
Min samples in group: abs=0 and rel=0		
Min %-intensity overlap: 0.5		
Exclude gap-filled features: true		
Feature shape correlation: true		
Min data points: 5		
Min data points on edge: 2		
Measure: PEARSON		
Min feature shape correlation: 0.65		
Min total correlation: false (0.5)		
Feature height correlation: true		
Minimum samples: 3		
Measure: PEARSON		
Min correlation: 0.65		
Suffix (or auto): true (cor)		

APPENDIX 4. USED MZMINE MODULES AND EXPLANATIONS

The following modules of MZmine were used in order of shown to analyse the LC-MS/MS spectra. First *Mass detection* with higher noise level (Table 1) was used to define the threshold value to filter out the noise from the raw data. Then, second *Mass detection* with lower noise level (Table 1) was to define the threshold value to filter out the noise from the raw data. *ADAP chromatogram builder* was used to build an extracted ion chromatogram (EIC) for each mass-to-charge (m/z) value according to set parameters (Table 1). *Local minimum resolver* module was used to split "shoulders" in LC peaks into individual features according to set parameters. Parameters were mainly determined visually by inspecting a few noisy EIC and good EIC in a few feature lists by using *Show preview* box. *13C isotope filter (isotopic grouping)* module was used to filter out the features corresponding to 13C isotopes of the same analyte. *Isotope pattern finder (isotopic peak finder)* module was used to search isotope patterns in selected feature list. *Join aligner* module was used to align detected peak in the samples through a match score, which is based on the mass and retention time of each peak and ranges of tolerance specified in the parameters (Table 1). *Feature list rows filter* module was used to filter out the rows that do not match to the set parameters (Table 1). *Feature filter* module was used to delete features which does not match with the set parameters (Table 1). *Feature list blank subtraction* module was used to subtract the features appearing in blanks. *Peak finder (gap filling)* module tried to fill the gaps in the aligned feature table. *Duplicate peak filter* module was used to help to remove misaligned feature list rows. *Feature grouping (metaCorrelate)* module was used to group features based on set parameters (Table 1).

Table 5. The modules used in MZmine for analysing the LC-MS/MS spectra of bank voles' faecal samples from in- and outside of Chornobyl Exclusion zone, modules parameters and their explanations.

Module	Parameter	Explanation
1. Mass detection	Raw data files	All raw data files were selected as the mass detection was to be performed on all raw data files.
	Scan filters -> Show -> MS level filter	Select the desired MS level. As the data is LC-MS/MS data, first the MS level 1 mass detection was performed.
	Mass detector	Select suitable algorithm for mass detection. As data was already centroided, the *centroid* algorithm was used.
	Mass detector Setup -> Noise level	The *noise level* was set to 2.0E4 for MS level 1 scans to filter out any signals which have intensity lower than 2.0E4. The noise level was determined by using *Show preview* box and the threshold value was visually set to filter the background noise out from MS level 1 scans.

2. Mass detection	Raw data files	All raw data files were selected as the mass detection was to be performed on all raw data files.
	Scan filters -> Show -> MS level filter	Select the desired MS level. As the data is LC-MS/MS data and the MS level 1 mass detection was already performed, next the MS level 2 was performed.
	Mass detector	Select suitable algorithm for mass detection. As data was already centroided, the *centroid* algorithm was used.
	Mass detector Setup -> Noise level	The *noise level* was set to 1.5E3 for MS level 2 scans to filter out any signals which have intensity lower than 1.5E3. The noise level was determined by using *Show preview* box and the threshold value was visually set to filter the background noise out MS level 2 scans.
ADAP chromatogram builder	Raw data files	All raw data files were selected as the ADAP chromatogram builder was to be performed on all raw data files. This module builds an extracted ion chromatogram (EIC) for each mass-to-charge (m/z) value according to set parameters.
	Scan filters -> Show -> MS level filter	The module processes only MS level 1 scans.
	Minimum consecutive scans	The minimum number of scans, where a m/z must be detected to build the EIC and retained in the feature list. The parameter was determined by inspecting the raw data's usual minimum number of data points of the LC peak.
	Minimum intensity for consecutive scans	Determines the minimum intensity of the consecutive scans to build the EIC and retained in the feature list.
	Minimum absolute height	Determines the minimum height of the EIC, so it will be built and retained in the feature list.
	m/z tolerance (scan-to-scan)	Determines maximum difference between m/z within the EIC.
Local minimum resolver	Feature lists	Select feature lists created in previous step.
	MS/MS scan pairing -> Show -> MS1 to MS2 precursor tolerance (m/z)	Pairs MS/MS fragmentation spectra to resolved features.
	MS/MS scan pairing -> Show -> Retention time filter -> Use tolerance	Parameter was kept as default.
	Dimension	Dimension to be resolved, here used dimension was *retention time*.

	Chromatographic threshold	Determines the percentage of data points in EIC to be filtered out before local minimum search.
	Minimum search range RT/Mobility (absolute)	Determines the size of retention time window for local minimum research. Too low value can cut off peak edges and too high value can lead to incomplete peak separation.
	Minimum absolute height	Determines the peak's heigh (intensity) to be retained as a feature.
	Min ratio of peak top/edge	Determines the minimum difference between peaks top and sides to be retained as a feature.
	Peak duration range	Determines the acceptable range of the peak length.
	Minimum scans	The minimum number of scans (data points) that resolved peak needs to have to retain as a feature. The parameter was determined by inspecting the raw data's usual minimum number of data points of the LC peaks.
13C isotope filter (isotopic grouping)	Feature lists	Select feature lists created in previous step.
	m/s tolerance (intra-sample)	The maximum difference between the measured and predicted m/z of the potential 13C isotope to be grouped as isotopologues. Can be strict to avoid discarding false 13C.
	Retention time tolerance	The maximum retention time between the feature and its potential 13C isotope to be grouped as isotopologues. Can be strict to avoid discarding false 13C.
	Maximum charge	The maximum charge state considered to predict 13C isotopes' m/z.
Isotope pattern finder (isotopic peak finder)	Feature lists	Select feature lists created in previous step.
	m/z tolerance (feature-to-scan)	Parameter was kept as default.
	Maximum charge of isotope m/z	Parameter was kept as default.
Join aligner	Feature lists	Select feature lists created in previous step.
	m/z tolerance (sample-to-sample)	Maximum difference between two m/z values to be considered the same.
	Weight for m/z	Weight for m/z difference for the match score calculation between peak rows.
	Retention time tolerance	Maximum difference between two retention times in order to be considered the same.
	Weight for RT	Weight for retention time difference for the match score calculation between peak rows.
Feature list rows filter	Feature lists	Select feature list created in previous step.
	Minimum aligned features (samples)	The minimum number of features detected in the row. If less, the row will be removed.
	Keep or remove rows	Keep rows that match all criteria

Feature filter	Duration	Determines the peak duration
	Data points	Determines the minimum data points.
Feature list blank subtraction	Feature lists	Select feature list created in previous step.
	Blank raw data files	Blank raw data files
	Minimum # of detection in blanks	The amount of the blanks
	Quantification	Height
	Ratio type	Maximum
	Fold change increase	If intensity in the filtered list increases more than given percentage, it will not be removed.
	Keep or remove features (of rows) below fold change	Remove
Peak finder (gap filling)	Feature lists	Aligned feature list filtered
	Intensity tolerance	Maximum allowed deviation from the expected peak shape in chromatographic direction.
	m/z tolerance (sample-to-sample)	The m/z range for feature search in the raw data.
	Retention time tolerance	The retention time range for feature search in the raw data.
	Minimum scans (data points)	1
Duplicate peak filter	Feature lists	Select feature list created in previous step.
	m/z tolerance	The maximum difference between duplicate peaks.
	RT tolerance	The maximum difference between duplicate peaks.
Feature grouping (metaCorrelate)	Feature lists	Select feature list created in previous step.
	RT tolerance	0.080 absolute (min)
	Min height	Determinates the minimum feature high.
	Intensity threshold for correlation	Determinates the threshold for the feature.
	Min sample filter -> Min sample in all	Determinates the minimum number of samples.
	Min sample filter -> Min %-intensity overlap	Determinates the minimum percent of overlap intensity.
	Feature shape correlation -> Min feature shape correlation	Determinates the feature shape correlation.
	Feature height correlation -> Minimum sample	Determinates the feature height correlation.
	Feature height correlation -> Min correlation	Determinates the correlation value.

APPENDIX 5. USED R PACKAGED AND VERSION

Table 6. The used R packages and versions for performing the statistical analysis of metabolite and bacteria data.

Package name	Version
phyloseq	1.40.0
ggplot2	3.3.6
ggpubr	0.4.0
vegan	2.6.2
GenomeInfoDb	1.32.3
GenomeInfoDbData	1.2.8
bitops	1.0.7
dplyr	1.0.9
hrbrthemes	0.8.0
gcookbook	2.0
tidyverse	1.3.1
devtools	2.4.3
knitr	1.39
rstatix	0.7.0
ggsci	2.9
plyr	1.8.7
ComplexHeatmap	2.12.1
RColorBrewer	1.1.3
breakaway	4.7.9
microbiome	1.18.0
decontam	1.16.0
gameofthrones	1.0.2
R	4.2.1
readxl	1.4.0
DESeq2	1.36.0
eulerr	6.1.1
MicEco	0.9.19
reshape2	1.4.4
MuMIn	1.46.0
SIBER	2.1.6
rjags	4.13
picante	1.8.2
magrittr	2.0.3