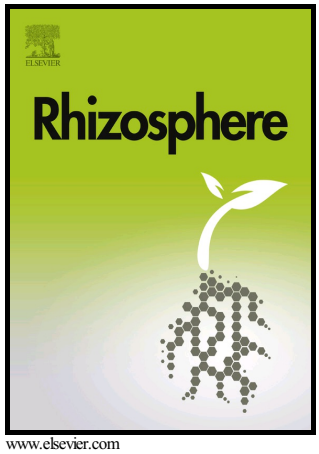

The state of rhizospheric science in the era of multi-omics: A practical guide to omics technologies

Richard Allen White, Albert Rivas-Ubach, Mark I. Borkum, Martina Köberl, Aivett Bilbao, Sean M. Colby, David W. Hoyt, Kerem Bingol, Young-Mo Kim, Jason P. Wendler, Kim K. Hixson, Christer Jansson

PII: S2452-2198(17)30066-6
DOI: http://dx.doi.org/10.1016/j.rhisph.2017.05.003
Reference: RHISPH64

To appear in: *Rhizosphere*

Received date: 25 March 2017
Revised date: 3 May 2017
Accepted date: 4 May 2017

Cite this article as: Richard Allen White, Albert Rivas-Ubach, Mark I. Borkum Martina Köberl, Aivett Bilbao, Sean M. Colby, David W. Hoyt, Kerem Bingol, Young-Mo Kim, Jason P. Wendler, Kim K. Hixson and Christer Jansson, The state of rhizospheric science in the era of multi-omics: A practical guide to omics technologies, *Rhizosphere,* http://dx.doi.org/10.1016/j.rhisph.2017.05.003

Available online at www.sciencedirect.com

ScienceDirect

Rhizosphere

Rhizosphere 00 (2017) 1–23

Review Article

The state of rhizospheric science in the era of multi-omics: A practical guide to omics technologies

Richard Allen White III[a], Albert Rivas-Ubach[a], Mark I. Borkum[a], Martina Köberl[b], Aivett Bilbao[a], Sean M. Colby[a], David W. Hoyt[a], Kerem Bingol[a], Young-Mo Kim[a], Jason P. Wendler[a], Kim K. Hixson[a,*], Christer Jansson[a,*]

[a]*Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99352, USA*
[b]*Institute of Environmental Biotechnology, Graz University of Technology, Rechbauerstraße 12, 8010 Graz, Austria*

Abstract

Over the past century, the significance of the rhizosphere has been increasingly recognized by the scientific community. This complex biological system is comprised of vast interconnected networks of microbial organisms that interact directly with their plant hosts, including archaea, bacteria, fungi, picoeukaryotes, and viruses. The rhizosphere provides a nutritional base to the terrestrial biosphere, and is integral to plant growth, crop production, and ecosystem health. There is little mechanistic understanding of the rhizosphere, however, and that constitutes a critical knowledge gap. It inhibits our ability to predict and control the terrestrial ecosystem to achieve desirable outcomes, such as bioenergy production, crop yield maximization, and soil-based carbon sequestration. Multi-omics have the potential to significantly advance our knowledge of rhizospheric science. This review covers multi-omic techniques and technologies; methods and protocols for specific rhizospheric science questions; and the challenges to be addressed during this century of rhizospheric science.

Keywords: rhizosphere, multi-omics, next-generation sequencing, metagenomics, metatranscriptomics, metabolomics, metaproteomics

*Corresponding author

Email addresses: richard.white@pnnl.gov (Richard Allen White III), albert.rivas.ubach@pnnl.gov (Albert Rivas-Ubach), mark.borkum@pnnl.gov (Mark I. Borkum), martina.koeberl@tugraz.at (Martina Köberl), aivett.bilbao@pnnl.gov (Aivett Bilbao), sean.colby@pnnl.gov (Sean M. Colby), david.hoyt@pnnl.gov (David W. Hoyt), kerem.bingol@pnnl.gov (Kerem Bingol), young-mo.kim@pnnl.gov (Young-Mo Kim), jason.wendler@pnnl.gov (Jason P. Wendler), kim.hixson@pnnl.gov (Kim K. Hixson), christer.jansson@pnnl.gov (Christer Jansson)

1. Introduction

1.1. Brief history of rhizospheric science

The rhizosphere is the interface between a plants roots and the surrounding soil (Figure 1). The rhizosphere microbiome is the specific microbial community within the first several millimeters away from the root. It consists of all domains of microbial life, including archaea, bacteria, fungi, picoeukaryotes, and viruses, but also constitutes a microbial community specific to plant species and genotype. The activities of a rhizosphere's microbiome are driven by the nature and characteristics of the plant root exudates [1, 2]. The rhizosphere is one of the most dynamic interfaces on Earth, containing up to 10^{11} microbial cells per gram of root [3], representing over ~30,000 bacterial species [4].

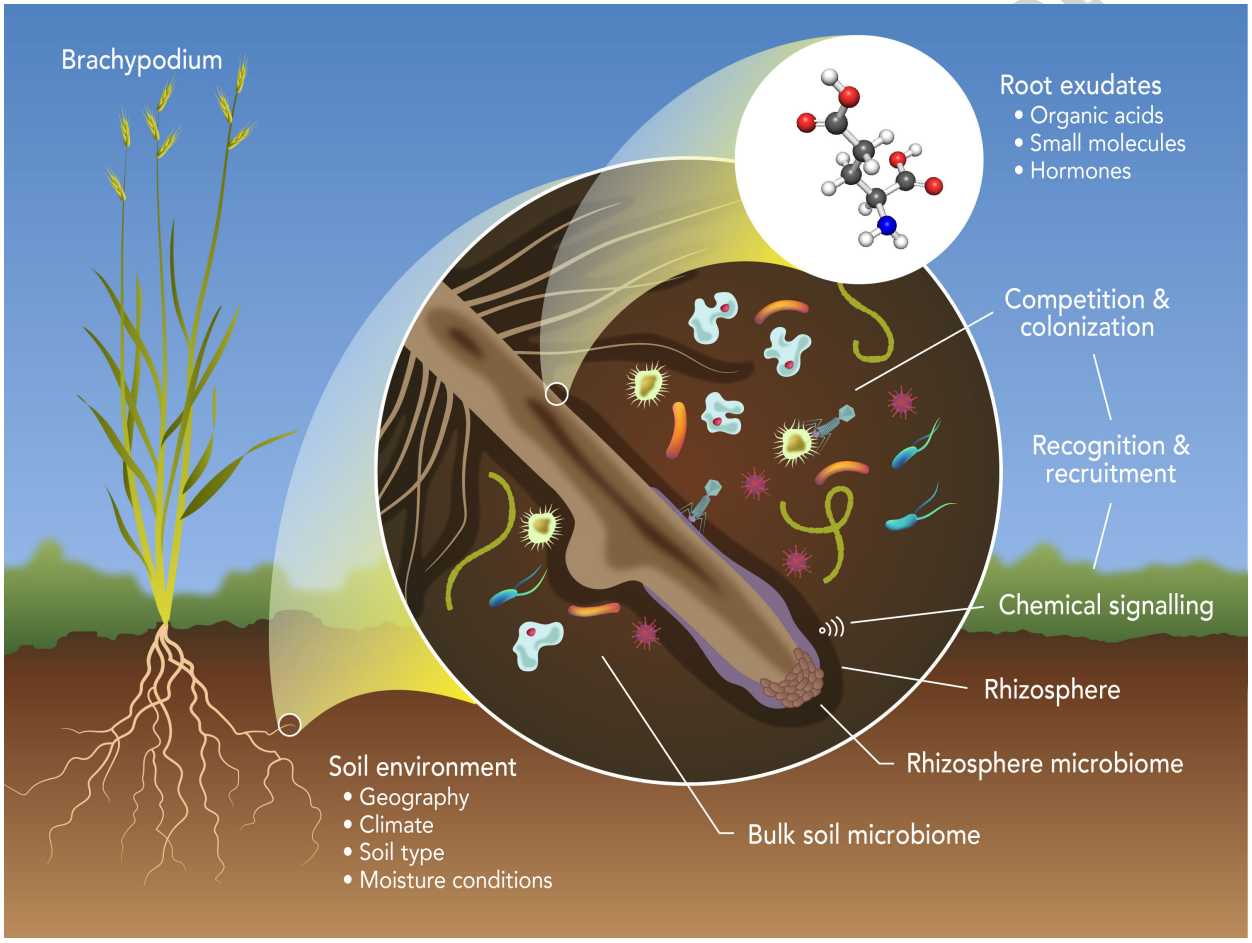

Figure 1. Depiction of the rhizosphere components and processes (not to scale).

The term was coined in 1904 by Lorenz Hiltner, who demonstrated that a plants nutrition is significantly influenced by the composition of its rhizosphere microbiome [5]. Hiltners research on germination and plant growth

demonstrated that rhizosphere-based microbial communities exhibit plant-specificity because of differences in the composition and concentration of plant root exudates [6]. Successive researchers built upon Hiltners original work. Most prominent among them was Albert Rovira, whose research on "the rhizosphere effect" [7], provided detailed views of plant-driven microbial colonization of the rhizosphere at the microscopic scale.

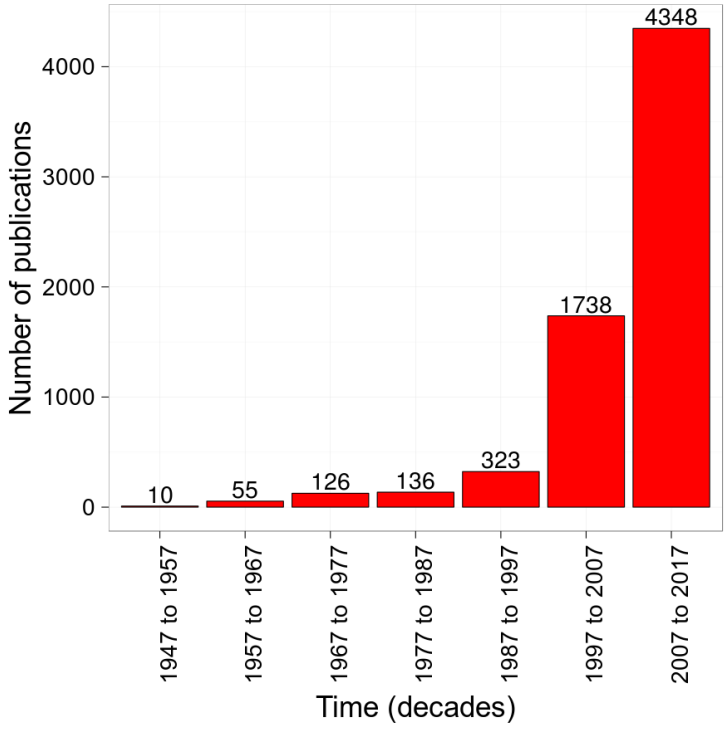

Figure 2. Number of manuscripts found on PubMed featuring the keyword "rhizosphere" published from 1947 to 2017 (to February 2017).

1.2. Current state of rhizospheric science

A PubMed search from 1947 to 2017 revealed a total of 6,344 manuscripts published featuring the keyword "rhizosphere," of which almost 70% were published in the last decade (Figure 2). Early publications, describing research conducted without the use of computers, emphasized the need for prefatory understanding of the nature and characteristics of the rhizosphere at microscopic scales. Over the last decade, the surge in rhizosphere publications can be linked to several developments in high-throughput technologies, to improvements in computational capabilities, to the maturity of the Web as a data dissemination platform, and to the development of novel algorithms for the analysis of complex, multi-omic datasets. These new methodologies target the microbial composition, specificity, and functionality of the rhizosphere microbiome at the community scale [8, 9, 10, 11], as well as the complex network plant-microbe interactions at the molecular scale [12, 13, 14, 15, 16].

1.3. What is multi-omics?

Multi-omics is a modern, holistic approach to systems biology that leverages the datasets of multiple -omes (the genome, for instance, as well as the proteome, transcriptome, and metabolome). These multiple -omes are from individual organisms (such as archaea, bacteria, fungi, picoeukaryotes, plants, animals, and viruses) as well as from multi-organism communities and consortia. Directly measured data is gathered with a variety of instrumentation pertaining to, at minimum, the "central dogma" of molecular biology [17]: that is, DNA, mRNA, and proteins are measured, analyzed, and integrated using multivariate statistics to be considered multi-omics. Examples of such microbiome-based multi-omics studies (which integrate data from DNA, mRNA, and protein) are described in Hultman *et al.,* [18] and Heintz-Buschart *et al.,* [19].

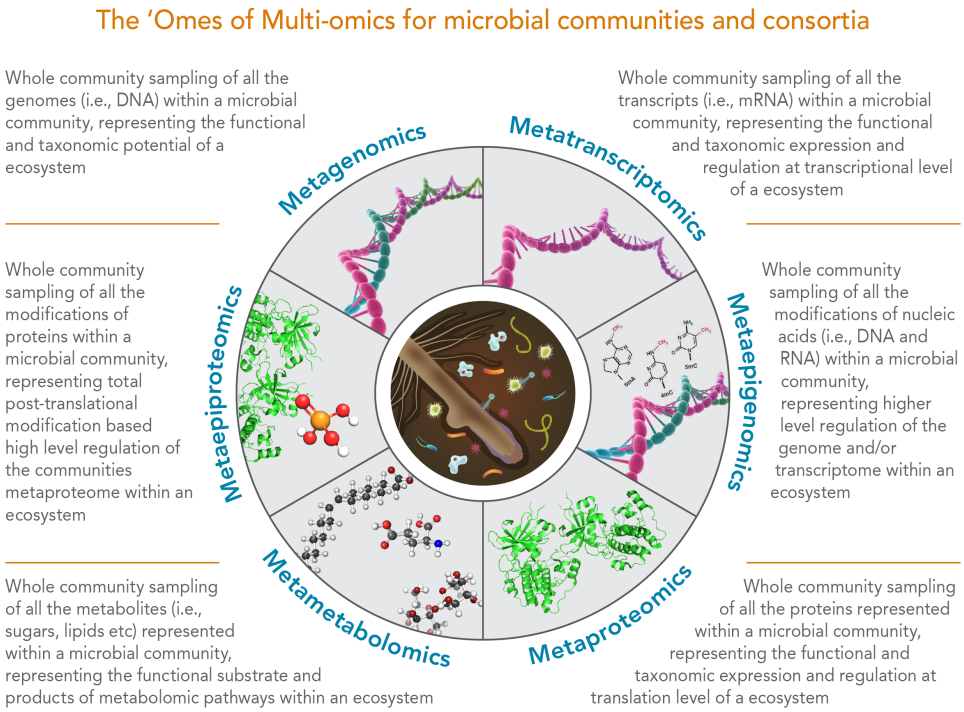

Figure 3. The -omes of multi-omics for microbial communities and consortia.

1.4. Multi-omics terminology

In this section, we introduce and discuss the terminology used in subsequent sections of this review (Figure 3). The word "-ome," used in a biological context, is the totality of a class of biological molecules. Current high-throughput technologies discussed later provide global measurements of the totality of an entire -ome. For example, the "genome" is the total genetic information encoded by the DNA of an entire organism. The term "sampling," as

used in this review, refers to the process of obtaining a point-in-time profile of a single sample for a given -ome. The suffix "meta" has been adopted in this biological context to mean a collection of many individuals that refers to whole microbial communities. The term "metagenomics" refers to a whole-community sampling of the totality of all genetic information encoded by the DNA of all individuals found within that sample. Metagenomics was derived from the original definition of the term "metagenome" by Handelsman *et al.,* [20], referred to the "cloning and functional analysis of the collective genomes." Metagenomics should only refer to whole-community genomics, where DNA is randomly fragmented and directly sequenced, and where both function and taxonomy are predicted from direct measurements. Metagenomics should not be used to describe whole-community polymerase chain reaction (PCR) amplification of individual markers (mostly commonly 16S ribosomal RNAs (16S rRNA)), which uses the form of sequencing known as PCR amplicon sequencing. PCR amplicon sequencing is limited to direct measurement of taxonomy and diversity measurements.

1.5. General challenges of obtaining -omics from the rhizosphere

Rhizosphere and soil science are rapidly undergoing a transition to interdisciplinary approaches that require collaboration in many areas of science, including geology, pedology, mineralogy, physics, biology, bioinformatics, mathematics, statistics, and computer science. Successful holistic studies of the rhizosphere will in the future depend on developments of both techniques and technologies, along with robust data analytics and software. Our review discusses developments in the techniques and technologies of -omics as they are applied to the rhizosphere or to surrogate environmental ecosystems. Another short review contained in this special edition issue will tackle the largest challenge of the obtaining multi-omic models of the rhizosphere: the data analysis.

As mentioned above, rhizosphere soil represents one of the most exceptionally complex ecosystems because of the numerical diversity of microbes alone. A current rhizosphere study using 16S rRNA PCR amplicon sequencing of *Arabidopsis thaliana* across 613 samples illustrates the high diversity of microbes contained within the rhizosphere. It found >2,000 species (as operational taxonomic units − OTUs) per gram of rhizosphere soil [10]. If we assume that the average microbial genome has ~3,000 protein-coding genes with ~2,000 species, that represents $6x10^6$ bacterial proteins at the low end. At the high end, an estimate of ~30,000 species per gram posits $9x10^7$ proteins [10].

Rhizosphere soil contains very high levels of interfering humic acids from the soil matrix, as well as plant polyphenols and other degraded macromolecules, all of which makes extracting biological molecules such as DNA, RNA, proteins, and metabolites difficult. These compounds interfere due to co-extraction and they inhibit PCR and the ionization required for analyzing proteins and metabolites. Separating these biological molecules from interfering compounds is critical for high resolution -omics results.

One main challenge that is rarely discussed is extraction bias, based on the fact that many microbes are resistant to lysis; thus, we are getting a bias view point of microbial communities. Extraction bias is an important problem in rhizosphere studies since a single commercial kit or method is commonly used for isolating biological molecules such as DNA, RNA, proteins, and metabolites from rhizosphere soil microbial communities. All -omics require robust extraction and lysis in order to get the most comprehensive downstream comparisons and analysis with lowest bias extraction. We discuss this challenge in greater detail later in the sections on the challenges of next-generation sequencing (NGS) and metaproteomics.

2. Next-generation sequencing (NGS) of rhizospheric microbial communities

2.1. Introduction to next generation sequencing (NGS)

NGS platforms are used for direct measurements of the genomes and transcriptomes of individuals, and for the metagenomes and metatranscriptomes of whole communities. In recent years, NGS has become vastly more cost- and time-efficient compared to other technologies for multi-omic analysis, including mass spectrometry (MS) based proteomics and metabolomics [21]. NGS platforms are characterized by two parameters: mean read length and throughput (Table 1). The preeminent manufacturers of NGS platforms are Illumina, Pacific Biosciences of California, Inc. (i.e., PacBio), and Oxford nanopore technologies (i.e., Oxford) (Table 1).

Table 1. Summary of Next Generation Sequencing (NGS) platforms. 'Per unit' refers to flow cell lane (Illumina), MinION cell (Oxford), or SMRT cell (PacBio) [22].

Vendor	Platform	Instrument	Date	Read length (bp)	Accuracy	Data (per unit)	Data (total)
Illumina	Illumina	HiSeq/MiSeq	2007	100–300	99.9%	1.8 Tb	18 Tb
Moleculo	Illumina	HiSeq	2012	6,000–16,000	99.9%	1 Gb	16 Gb
PacBio	SMRT	PacBio RS II (P6)	2011	10,000–25,000	89.0%	300 Mb	10 Gb
Oxford	Nanopore	MinION	2012	1,000–90,000	90.0%	1 Gb	1 Gb

Illumina is the most mature maker of NGS platforms, and aims to provide great utility in sequencing capabilities. Illumina's NGS platforms use oligonucleotide array flow cells, reversible chain terminators, and bridge PCR amplification in order to directly sequence nucleic acids by synthesis [23]. Nucleic acids are fragmented based on obtaining wanted insert size (usually ~350 bp) then oligonucleotide adapters/barcodes are added resulting in library for sequencing. The library adapter/barcode sequences are complementary to array flow cells, which are then annealed and amplified by a special PCR on the flow-cell known as bridge PCR [23]. Illumina provides overlapping insert libraries with paired end reads (medium throughput); varying read lengths (high throughput); and short reads (ultra-high throughput) [21]. Illumina offers a novel longer-read technology that provides read lengths of approximately 8-15 kbp. The technology, formally called Moleculo, is now referred to as the TruSeq synthetic long-read

DNA library prep kit [24, 25]. In recent studies, this approach has been demonstrated [26, 25, 27] to improve *de novo* metagenomic assembly and genomic binning of microbial communities, including prairie soils [26].

NGS platforms made by PacBio and Oxford nanopore technologies offer generally low-to-medium throughput when compared to Illumina platforms. However, they yield significantly longer read lengths by two orders of magnitude (Table 1) and have been reviewed further elsewhere [21]. PacBio uses tethered DNA polymerases and zero-mode waveguides to detect and sequence nucleic acids [28, 29]. Oxford Nanopore Technologies NGS platforms thread individual nucleic acid molecules through nano-size pores, and then detect the individual nucleotide bases through changes in electrical conductivity [30]. It should be noted that, to date, Oxford nanopore technologies NGS platforms have demonstrated the longest read lengths (up to 90 kbp) [30]; however, these long reads were rare and prone to error but can be corrected by novel software [31].

Both PacBio and Oxford nanopore technologies NGS platforms offer direct detection of modifications to nucleic acids (e.g., methylation and/or hydroxymethylation of cytosine). While they have not been applied to microbial communities as they have been in metaepigenomics (Figure 3), both PacBio and Oxford nanopore could be the technologies that access whole community modification profiles of nucleic acids. Microbial genomes appear to have modification of their DNA; the study by Murray *et al.,* [32] is an excellent example of epigenomics on single microbes. PacBio's NGS platforms also detect modifications to nucleic acids by measuring the "wobble" of unnatural bases during polymerase threading [32]. Oxford nanopore technologies can detect modified bases (e.g., methylated bases), by a decrease in voltage as the modified bases are threaded through the nanopore [33].

2.2. Usage and applications of next-generation sequencing

NGS has been used to infer rhizospheric microbial community structure [10]; the core microbial community [11]; microbial community structure and metabolic potential by metagenomics [34]; microbial community expression under pesticide treatment (i.e., glyphosate) [35]; and plant-microbe interactions [36]. These and other studies have provided a wealth of information pertaining to responses to perturbation, as well as to the abundance, structure, diversity, spatial distribution, and core members of rhizospheric communities. Studies of the rhizosphere employing NGS platforms have mainly focused on model plants such as *Arabidopsis thaliana* [10, 11], legumes (e.g., peas and soybeans) [34, 35, 36], and cereals (e.g., wheat, oats, and corn) [35, 36].

After posing a research question, the first step in a successful NGS workflow is the selection of an appropriate experimental technique for the downstream application to the rhizosphere. Two techniques are quite useful for inferring core- and pan-microbial community structure [10, 11] and spatial and temporal distributions [37]: amplifying 16S rRNA genes (e.g., bacteria/archaea-specific markers) or using internal transcribed spacer (ITS) (e.g., fungal specific markers) regions for PCR amplicon sequencing as phylogenetic markers [38, 39].

Both DNA- and cDNA-based templates can be used for PCR analysis. They provide the most inexpensive NGS experiments, high replication, and limited sample input (approximately 1 ng nucleic acid template). These cDNA-based PCR amplicon templates can also measure the metabolically active members of a microbial community. However, PCR amplification creates bias towards certain microbial groups [38, 40]. Illumina HiSeq and MiSeq represent the best options for PCR amplicon studies. They are cost-effective and robust, obtain deep coverage (often to level rarefaction even for diverse communities), and they scale to many samples [38]. Recently, Illumina Moleculo [26], Illumina MiSeq [41] and PacBio sequencing [42] have been used to obtain long full-length 16S rRNA genes for high-resolution microbial taxonomy and phylogenetic placement. Still, these technologies cannot provide the sequencing depth offered by shorter PCR amplicons. In the near future, these long-read technologies may provide higher numbers of sequences.

Metagenomics provides taxonomy in a less biased manner than 16S rRNA PCR amplicons and it allows direct inference of the metabolic potential of a microbial community. *De novo* assembly of metagenomics data provides error correction of reads, connects protein coding gene fragments, and it provides more robust assignment of function and taxonomy due to added length during alignment to reference databases [43]. Metagenomics by using *de novo* assembly allows for reconstruction of complete to near-complete genomes from unknown and uncultivated phyla [26, 44, 45].

For the reconstruction of genomes, long reads are superior to short reads for *de novo* metagenomic assembly due to increased contig length [43, 46]. A length of reads or contigs of ~7,000 bp or greater has been called the "golden threshold," of *de novo* assembly of microbial genomes [46]. Long reads or contigs ~7,000 bp can resolve the most common repeats in bacterial genomes, which allows for the reconstruction of complete microbial genomes [46]. Longer reads offer more overlapping data for higher accuracy annotation, and potentially allow for higher genome completeness (that is, the percentage of how complete a genome is, based on universal marker genes found across all clades of bacteria) for reconstructing genomes from metagenomes [25, 26, 27]. Illumina Moleculo, PacBio, and Oxford nanopore are the best sequencing platforms for providing long reads (Table 1). Note that short reads can be used for genome reconstruction for low-complexity microbial communities [44, 45, 47]. However, long reads provide better protection from assembly errors caused by genome repeats [46] and often contain gene operons on single reads [48]. Metagenomics alone cannot be used to deduce whether any of the metabolic genes are expressed at the time of sampling. For this information, metatranscriptomics is needed.

Metatranscriptomics is a powerful, robust, and cost-efficient use of NGS because it allows for a variety of questions to be answered with one experiment. If rRNAs are not depleted, they can provide indicators of protein synthesis; however, due to lifestyle strategies, they cannot be used as general indicators of activity or growth [49]. If meta-

transcriptomics are combined with temporal and/or spatial measurements, it can answer questions relating to when functions are transcriptionally present and active in relation to time and space. Expressed functions obtained from metatranscriptomics facilitates the inference of biogeochemical cycles [34], models of plant-microbe interactions [50], and functional diversity measurements [51].

For metagenomic and metatranscriptomic studies, Illumina HiSeq provides many counts at relatively low cost. The reads provided by Illumina HiSeq are very short, however, from 100 to 250 bp. These shorter reads are useful if the microbial community in question is well characterized with known reference genomes; if so, then high-quality alignment/mapping to these reference genomes can provide insights into functional and taxonomic distributions. However, if the microbial community is not well characterized, which is very common in soil/rhizosphere studies, then very short reads often lead to lower gene prediction, and to less usable data [52]. If there are no reference genomes and/or high quality metagenomes, then *de novo* assembly of the metatranscriptome should be completed in order to allow for higher quality gene prediction. Shorter-read technologies, such as those offered by Illumina, are highly useful for counting mRNA transcripts (with high quality reference genomes/metagenomes), and should be used to provide further providence in gene predictions within genomes obtained from metagenomes (genomic bins) [26]. Metatranscriptomics can also be done on longer-read technologies, and sequencing can be done directly on Oxford Nanopore [53].

2.3. NGS challenges

Many challenges remain in using NGS for the rhizosphere, including: data analysis, data storage, extraction of nucleic acids, free nucleic acids, and nucleic acid stability. The greatest challenge in rhizosphere science is data analysis. For data analysis and data storage there are excellent reviews on NGS microbial community studies [54, 55, 56, 57].

Extraction bias and the presence of free and/or high molecular weight nucleic acids remain the most stubborn obstacles to using NGS for rhizospheric studies. Many microbes within soil and rhizosphere samples have robust cell walls (e.g., gram positives), which makes them difficult to lyse [58]. Soil and rhizosphere samples also have dormant microbes and spores that are difficult to extract nucleic acids [58]. Since no single extraction method is capable of extracting all members of a microbial community, more than one extraction technique, as well as different methods of lysis, have to be used in order to get more complete coverage. If enough material is available, to minimize extraction bias we suggest using multiple-extraction replicates from the same extraction technique and using different methods of extracting nucleic acids.

Free nucleic acids from dead microbial cells are also problematic. The DNA from these cells can result in both abundances and metabolic potential being misrepresented. Using metatranscriptomes to validate DNA based

metagenomes or 16S PCR amplicon studies is beneficial because they are more reliable than only DNA in predicting functions and microbial communities that are transcriptionally active. However, transcriptional activity doesnt always mean an organism is growing or active, but it provides greater evidence than DNA alone [59]. Nucleases, DNA-binding chemicals, and stable-isotope probing (SIP) may also help enrich metabolically active members [60, 61].

The stability of nucleic acid in field-collected samples is an ongoing problem in soil and rhizosphere science. Using liquid nitrogen for the rapid freezing of samples is the most common way to preserve samples, but it may not be possible in all field conditions. RNAlater (ThermoFisher) and Lifeguard (Qiagen) may be alternatives to protecting nucleic acids in field conditions, but their chemistry may not be compatible with all extraction kits [62]. Ethanol and RNAlater showed bias in the amplification of certain microbial groups [62]. Phenol-chloroform may be an alternative when freezing is not possible in the field and has been shown to stabilize nucleic acids [62].

Few technologies are available to obtain samples of high quality in high-molecular nucleic acids from soil and rhizosphere samples. This is a major challenge in moving to long-read technologies since such methods require samples of high purity and of the highest molecular weight in order to obtain the longest possible sequence reads.

3. Mass-spectrometry-based omics: metaproteomics and metabolomics

Metaproteomics and metabolomics provide deep characterization of, respectively, all proteins and metabolites that are present in a given sample [63, 64]. A variety of analytical techniques have been used to characterize proteins and metabolites from samples. However, because of significant advances in techniques and technologies, mass spectrometry (MS) is currently the predominant method for both metaproteomics and metametabolomics analysis. Global protein and metabolite characterization of complex samples is typically achieved by combining chromatography with mass analysis. MS-based metaproteomics typically employs reverse phase liquid chromatography (LC), while MS-based metabolomics can be also performed using gas chromatography (GC) through the derivatization of molecular compounds [65, 66].

MS-based proteomics is completed by the extraction of proteins via solvents (e.g., chloroform/methanol), detergents (e.g., sodium dodecyl sulphate (SDS), or by physical means (e.g., sonication). Once a protein is extracted, it is cleaved into peptides (most commonly trypsin) for downstream analysis. Metabolites are extracted using similar methods in proteomics; however, detergents are rarely used due to their lack of compatibility and the potential of downstream loss. For proteins, there are methods for quantification (e.g., bicinchoninic acid assay, or BCA assay) and detergent removal; but methods for global metabolite quantification and detergent removal without loss of material have yet to be established.

Electrospray ionization (ESI) is currently the most typical ionization method for liquid chromatography mass

spectrometry (LC-MS), proteomics [65], and metabolomics [66]. LC-MS uses ESI, yielding a softer ionization of compounds and facilitating measurements of intact species [67]. ESI technique consists in by applying a high voltage electric field to a liquid sample passing through a capillary tube forming millions of nanometric charged droplets. Gas chromatography mass spectrometry (GC-MS), in metabolomics only, uses electron impact (EI), which often results in extensive fragmentation. A typical EI induces ionization by applying a high potential electron beam under vacuum to ionize from the analytes. After molecules (either peptides or metabolites) are ionized, a mass analyzer measures its mass-to-charge ratio (m/z) and a detector accounts for the number of ions at each m/z value.

4. Metaproteomics of rhizospheric microbial communities

4.1. Introduction to metaproteomics

In recent decades, metaproteomics has evolved into a large-scale assessment of produced and/or modified proteins (e.g., post-translational modifications, or PTMs) in microbial communities [68]. Metaproteomics methodologies are summarized in Table 2.

Table 2. Summary of metaproteomics methodologies, where direct measurements are obtained using LC-MS.

Name	Proteome Coverage	Detectable Moiety Size	Ease of Detection
Bottom-up	High	Small	High
Hybrid	Medium	Medium	Medium
Top-down	Low	Large	Low

Bottom-up proteomics is the most frequently applied strategy for metaproteomics. The bottom-up approach requires that proteins within the sample be digested into smaller peptides using a protease (e.g., trypsin); then further separated by various multidimensional LC strategies (e.g., strong cation exchange and/or reversed phase); and then ultimately analyzed using MS/MS [69, 70, 71]. Bottom-up metaproteomics provides high proteome coverage, making it one of the best approaches currently employed for microbial community analysis. Metaproteomics is mainly accomplished using a label-free strategy via spectral counts (i.e., the total number of spectra identified) of detected proteins. The samples are run individually, at the same total peptide mass, to avoid bias in analysis.

To obtain more specificity in proteomic community assessments, an alternative strategy is top-down proteomics [72], which analyzes intact proteins in their original form by using high resolution MS. It answers targeted questions for specific proteins within a complex sample. Currently, top-down proteomics are not high-throughput enough to measure whole communities. However, as the technology advances these techniques will become more commonplace.

4.2. Metaproteomics usage and applications

Metaproteomics has been used to characterize rhizospheres associated with leaf litter decomposition [73], methanotrophs in rhizosphere/root tissues of rice plants [74], crop rhizospheres [75, 76], and black truffles [77]. The permafrost metaproteome from Hultman *et al.,* [18]; which thus far has identified 7,000 proteins, has been estimated to contain billions of unique proteins [68]. Both bulk and rhizosphere soil have much higher microbial diversity than permafrost. It is likely that the number of unique proteins would be higher in bulk and rhizosphere soils because of the higher growth rates of microbes in those soil types when compared to low-growth frozen permafrost. As a result, these soils might have a greater number of unique proteins than permafrost by an order of magnitude [68].

With this predicted vast number of unique proteins in the rhizosphere, which technique is the best? LC-MS/MS bottom-up is the current standard for shotgun metaproteomics measurements of protein abundances in microbial communities. For quantitative and qualitative metaproteomics, spectral counts (i.e., the total number of spectra identified for a protein) are the standard in the field. But stable-isotope labeling (SIL), if deconvoluted correctly, could prove helpful in quantifying proteins as that area of proteomics continues to develop [78]. Using SIL tagging for relative abundance determination is common in proteomics, but is rarely utilized in metaproteomics. That presents an immediate opportunity for rhizosphere science. SIL is recommended if an analysis requires a combination of growth rates, substrate utilization, post-translational modifications of proteins (PTMs), translational rates, time series, and quantitative data.

4.3. Metaproteomics challenges

Metaproteomics in the rhizosphere and surrounding soil is very challenging for several reasons: the high diversity of organisms present, protein redundancy, and the general difficulty in protein extraction[79]. By digesting proteins into peptides (typically with trypsin), protein specificity is lost; many organisms, after all, contain redundant peptides that are similar to many other proteins. This challenge reduces bottom-down proteomics to a technique primarily used to assess protein changes in related groups of organisms. Another caution regarding bottom-up proteomics is its bias for detecting peptides coming from highly translated proteins such as ribosomal proteins, chaperones, and others involved in cellular maintenance. On the other hand, top-down proteomics measures intact proteins. It could be used to mitigate the problems of protein redundancy in protein digestion associated with bottom-up analysis. Top-down proteomics presents another immediate opportunity in studies of the rhizosphere.

Metaproteomic coverage can be improved with better proteomic extraction methods and fractionation that decrease peptide complexity. Direct lysis (e.g., with SDS) is the least biased method compared to indirect methods since it captures all intracellular and extracellular proteins [79]. However, SDS is incompatible with downstream metabolomic

analysis. A novel method, known as metabolite, protein and lipid extraction (MPLEx), presents a potentially improved means for simultaneously extracting proteins, polar metabolites, and lipids from rhizospheric soil [80]. Metaproteomic coverage can be improved by online 2D separation, which is comprised of on-line strong cation exchange (SCX) or high-pH reversed phase (i.e., C-18), followed by a secondary separation of low-pH reversed phase separations typically electrosprayed directly into a mass spectrometer [81, 82]. These added separation techniques decrease complexity for the mass analyzer allowing more detection on coverage of individual proteins.

Sampling multiplexing (that is, running multiple samples in a single analysis) is not possible with standard peptide labeling approaches such as iTRAQ [83] and Tandem Mass Tag (TMT) labels [84]. In rhizosphere soils samples there is an abundance of humic acids in the soil matrix, which inhibits efficient labeling. Label-free samples cannot be multiplexed. The preferred method of multiplexing samples for metaproteomics is to feed the microbial community a isotopically labeled substrates (e.g., ^{13}C, ^{15}N and $H_2^{18}O$) [85]. This labels proteins directly for quantitative metaproteomics in rhizosphere and/or soil samples.

5. Metabolomics of rhizospheric microbial communities

5.1. Introduction to metabolomics

Metabolomics aims to analyze the metabolomes of organisms at a specific time under specific conditions [64]. An organism's metabolome includes thousands of cellular substrates and products (metabolites) from primary metabolism (e.g., amino acids and sugars), and from secondary metabolism, (e.g., flavonoids and terpenoids), all of which are involved in many specific functions, including signaling and stress responses. The metabolome is considered the chemical phenotype of an organism [64], and is typically much more susceptible to environmental fluctuations than either the proteome or the transcriptome [86]. That makes metabolomics a useful method of understanding the specific metabolic pathways involved in the phenotypic responses of organisms [87].

Metabolomics can be targeted (i.e., highly specific) and untargeted (i.e., global and unspecific) [88]. Targeted metabolomics focuses on the detection and quantification of metabolites selected *a priori* given the scientific question [88, 89]. Specific metabolites are detected and quantified through optimization and by user-generated standard curves (injecting internal standards) [90]. Untargeted global metabolomics aims to simultaneously analyze as many metabolites as possible in a single analysis, yielding hundreds of metabolites across many samples [90, 91]. This can be subsequently followed by a targeted approach for a selected subset of metabolites of biological interest.

5.2. Metabolomics usage and applications

Metabolomics has been widely used to follow metabolic changes in samples from many disciplines, including plant physiology [92, 93], ecology [87, 94, 95], and microbiology [96, 97]. However, metabolomics is still an

emerging field, with much room for further advancement in understanding many complex systems, including the rhizosphere, which has been significantly underrepresented. That underrepresentation is likely due to several factors: limitations in sampling specific exudates; the sensitivity of current metabolomics platforms; and the difficulties of tracking the spatial and temporal dynamics of exudates at a relevant scale [98].

How can metabolomics help analyze a system as complex as the rhizosphere? LC-MS and GC-MS, along with nuclear magnetic resonance (NMR) spectroscopy, are the most widely used technologies for analyzing metabolomes in biological samples because of their capacity to detect a wide range of metabolites in a single run [87, 99]. Consider what NMR can do: it can detect the most abundant metabolites within samples with high accuracy, precision, and reproducibility without the need for extensive sample preparation or fractionation. The detection limit of NMR is at single-digit micromolar concentration. NMR intensities correlate linearly with the relative concentrations of the mixture's components, providing absolute concentration by adding a standard molecule with known concentration [100, 101]. NMR generates a quick overview of the most abundant metabolites in the sample. NMR can detect the metabolites that are difficult to ionize for MS, a common occurrence in salty samples. NMR can robustly quantify small molecular weight volatile organic compounds (SMWVOCs) like alcohols and short-chain fatty acids (SCFAs), such as acetate, propionate and butyrate. These molecules are often undetected or underrepresented when using MS sample preparation methods, such as drying down and derivatization [102]. LC-MS and GC-MS are more sensitive techniques than NMR spectroscopy [94].

Both LC-MS and GC-MS instruments provide a similar data format, resulting in two orthogonal and independent variables: retention time (RT), and m/z relative to each of the detected ions, which can be further used for metabolite assignment [103]. Recently, it has been shown that combining MS with NMR spectroscopy enables accurate structure elucidation of unknown and known metabolites in complex mixtures [104]. GC-MS is a high-resolving, chromatographic separation method and it has been widely used to analyze volatile and semi-volatile compounds such as biogenic volatile organic compound (BVOCs) [98]. GC-MS has proven to also be suitable for detecting compounds from primary metabolism such as fatty acids, carbohydrates, and organic acids, after derivatization of the extracts [105]. Due to sample derivatization and compound fragmentation during EI, GC-MS provides indirect detection of compounds, which complicates the quantification and elucidation of novel metabolites in samples. LC-MS has been the most common method to resolve a large variety of non-volatile compounds. It covers metabolites like flavonoids, alkaloids, phenolic acids, and saponins, as well as primary metabolites such as amino acids, carbohydrates, and organic acids [106]. However, using LC-MS is limited to solvent-soluble compounds. The lack of derivatization of the samples for LC-MS analyses, and the use of ESI, allow for direct detection of metabolites in samples. Other practical aspects of MS for analyzing the rhizosphere have been reviewed elsewhere [107]. Nevertheless, no single analytical

method or combination of instrumentations can cover the entire metabolome composition of a given sample [99]. In the future, new analytical techniques that extend the current metabolome coverage will be very important.

Mass spectrometers are quickly becoming more powerful and sensitive methods for measuring metabolites and proteins in the rhizosphere and in soil samples. A substantial number of published metabolomics studies demonstrates that the recent Orbitrap technology, coupled to LC or GC, offers the best current option for untargeted metabolomics of plant exudates and microbes. However, in the near future it is likely that IMS-MS (i.e., ion-mobility spectrometry mass spectrometry) will represent an inflection point for untargeted metabolomics analysis, allowing the detection and characterization of hundreds of metabolites in a single run [108, 109]. IMS-MS metabolomics will therefore permit rapid and accurate metabolomic profiles from very complex metabolite mixtures, such as the rhizosphere, in time frames never achieved before. IMS-based technology for metabolomics and lipidomics should be commercially available in the near future [110]. A single technique cannot recover the whole metabolome of a sample, so coupling different technologies such as high-resolution LC-MS with GC-MS and/or the ultra-high resolution FT-ICR-MS [111] would facilitate the acquisition of extensive metabolomic fingerprints. That would enable the elucidation of novel compounds present in very low concentration from complex mixtures.

5.3. Metabolomics challenges

Characterizing the metabolome of a complex microbial community in a robust and comprehensive way still represents a significant challenge [88]. Currently, extracting exudates is one of the major challenges in rhizospheric metabolomics. In order to accurately understand the complex interplay between the metabolic processes of plants and microbes, it is imperative to obtain separate data from the microbe and plant fractions, as well as from the final products of their interaction. Applying the most advanced mass spectrometers to gnotobiotic (i.e., germ-free or microbe-free) plants may offer a breakthrough in rhizospheric science.

Studies have characterized the chemistry of given soil samples [112], but those studies have characterized soils as a single complex sample that includes decomposing mixed organic matter blended with exudates and microbes. These organic compounds, and the organic matter in decomposition, are constantly exposed to several transformation processes (e.g., oxidation, methylation, and dehydration), which makes it crucial to consider the temporal scale of chemical transformations in the rhizosphere [63]. Several attempts at plant exudate collection have been addressed through different methods [[63], but the biggest challenge for rhizospheric metabolomics is differentiating between products from the microbe and plant fractions, as well as from the final product of interactions within their natural functional networks. Optimized protocols for accurately sampling and analyzing different fractions of the rhizosphere are necessary. Such optimized protocols would address the integrated nature of the metabolic processes of plants and microbes in the rhizosphere and in the larger terrestrial ecosystem.

6. Conclusions

The rhizosphere is a fascinating habitat for research for both basic and applied microbiology, plant biology, and ecology. Multi-omic approaches will lead to greater understanding of rhizosphere organisms and their roles in plant growth, crop production, and ecosystem health. Many of the novel -omics techniques were initially developed in other research fields but can be modified for use in the rhizosphere. Since there are numerous plant species (~374,000 are estimated to live on our planet), we have a long road ahead to obtain a holistic view of the numerous and highly plant-specific rhizospheric communities [113]. A big challenge in rhizosphere studies is to choose the technique and technology platforms that can best answer the most pressing scientific questions. Such considerations remain important in several areas: obtaining data that answers a specific scientific hypothesis; preventing data deluge and long delays from study to publication; and allowing for meaningful scientific advancements. Poor planning or limited knowledge of available technology often creates big problems downstream. Here we have mentioned and described selected techniques and technologies that may resolve such problems. Considerable challenges lie ahead, but we believe that increased applications of multi-omic approaches in rhizosphere science offers great potential. They can help harness the rhizosphere as a resource for improved plant growth and quality, for sustainable crop production, and for increased soil carbon storage under a variety of environmental stresses.

Acknowledgements

Funding was provided in part by the following PNNL initiatives: integrated Plant-Atmosphere-Soil System (iPASS; PNNL Project 204412) by DOE-Battelle Prime Contract for the Management and Operation of PNNL (DE-AC05-76RL01830). MK acknowledges support from the Austrian Science Fund (FWF; FWF Projects J-3638 and T-847). A portion of this research was conducted using the PNNL Institutional Computing (PIC) capability, EMSL, a DOE scientific user facility located at PNNL. We thank Nathan Johnson for his assistance in preparing excellent figures. We also thank Dr. Jennifer Kyle, Dr. Stephen Callister, and Dr. Ljiljana Paša-Tolić for there critical and helpful review of this manuscript. We thank Cory Ireland for copy-editing this manuscript.

Authorship statement

RAW and CJ conceptualized this article. RAW and MIB designed this article. All authors contributed to the drafting of this article. RAW, ARU, MIB and CJ made critical revision to this article. MIB copy-edited and typeset this article. All authors approved the final version of this article for publication.

References

[1] L. Philippot, J. M. Raaijmakers, P. Lemanceau, W. H. Van Der Putten, Going back to the roots: the microbial ecology of the rhizosphere, Nature Reviews Microbiology 11 (11) (2013) 789–799.

[2] R. F. Doornbos, L. C. van Loon, P. A. Bakker, Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. a review, Agronomy for Sustainable Development 32 (1) (2012) 227–243.

[3] D. Egamberdieva, F. Kamilova, S. Validov, L. Gafurova, Z. Kucharova, B. Lugtenberg, High incidence of plant growth-stimulating bacteria associated with the rhizosphere of wheat grown on salinated soil in Uzbekistan, Environmental Microbiology 10 (1) (2008) 1–9.

[4] R. L. Berendsen, C. M. Pieterse, P. A. Bakker, The rhizosphere microbiome and plant health, Trends in plant science 17 (8) (2012) 478–486.

[5] J. A. W. Morgan, J. M. Whipps, Methodological approaches to the study of rhizosphere carbon flow and microbial population dynamics, The Rhizosphere: Biochemistry and Organic Substance at the Soil-Plant Interface: Biochemistry and Organic Substance at the Soil-Plant Interface (2000) 373.

[6] A. Hartmann, M. Rothballer, M. Schmid, Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research, Plant and Soil 312 (1-2) (2008) 7–14.

[7] R. G. Burns, Albert Rovira and a half-century of rhizosphere research, in: The Rovira Rhizosphere Symposium, 2010, p. 1.

[8] G. Berg, K. Smalla, Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere, FEMS microbiology ecology 68 (1) (2009) 1–13.

[9] R. Mendes, M. Kruijt, I. de Bruijn, E. Dekkers, M. van der Voort, J. H. Schneider, Y. M. Piceno, T. Z. DeSantis, G. L. Andersen, P. A. Bakker, et al., Deciphering the rhizosphere microbiome for disease-suppressive bacteria, Science 332 (6033) (2011) 1097–1100.

[10] D. Bulgarelli, M. Rott, K. Schlaeppi, E. V. L. van Themaat, N. Ahmadinejad, F. Assenza, P. Rauf, B. Huettel, R. Reinhardt, E. Schmelzer, et al., Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota, Nature 488 (7409) (2012) 91–95.

[11] D. S. Lundberg, S. L. Lebeis, S. H. Paredes, S. Yourstone, J. Gehring, S. Malfatti, J. Tremblay, A. Engelbrektson, V. Kunin, T. G. Del Rio, et al., Defining the core Arabidopsis thaliana root microbiome, Nature 488 (7409) (2012) 86–90.

[12] C.-M. Ryu, M. A. Farag, C.-H. Hu, M. S. Reddy, H.-X. Wei, P. W. Paré, J. W. Kloepper, Bacterial volatiles promote growth in Arabidopsis, Proceedings of the National Academy of Sciences 100 (8) (2003) 4927–4932.

[13] D. V. Badri, J. M. Vivanco, Regulation and function of root exudates, Plant, Cell & Environment 32 (6) (2009) 666–681.

[14] A. Hartmann, A. Schikora, Quorum sensing of bacteria and trans-kingdom interactions of N-acyl homoserine lactones with eukaryotes, Journal of chemical ecology 38 (6) (2012) 704–713.

[15] L. A. Moe, Amino acids in the rhizosphere: from plants to microbes, American journal of botany 100 (9) (2013) 1692–1705.

[16] L. A. Weston, U. Mathesius, Flavonoids: their structure, biosynthesis and role in the rhizosphere, including allelopathy, Journal of chemical ecology 39 (2) (2013) 283–297.

[17] F. H. Crick, On protein synthesis, in: F. Sanders (Ed.), Symposia of the Society for Experimental Biology, Number XII: The Biological Replication of Macromolecules, Cambridge University Press, 1958, pp. 138—163.

[18] J. Hultman, M. P. Waldrop, R. Mackelprang, M. M. David, J. McFarland, S. J. Blazewicz, J. Harden, M. R. Turetsky, A. D. McGuire, M. B. Shah, et al., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes, Nature 521 (7551) (2015) 208–212.

[19] A. Heintz-Buschart, P. May, C. C. Laczny, L. Lebrun, C. Bellora, A. Krishna, L. Wampach, J. Schneider, A. Hogan, C. de Beaufort, et al., Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes., Nature microbiology 2 (2016) 16180.

[20] J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, R. M. Goodman, Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products, Chemistry & biology 5 (10) (1998) R245–R249.

[21] R. A. White III, S. J. Callister, R. J. Moore, E. S. Baker, J. K. Jansson, The past, present and future of microbiome analyses, Nature Protocols 11 (11) (2016) 2049–2053.

[22] V. C.-E. Jay S. Ghurye, M. Pop, Metagenomic assembly: Overview, challenges and applications, Yale J Biol Med 88 (3) (2016) 353–362. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5045144/

[23] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, et al., Accurate whole human genome sequencing using reversible terminator chemistry, Nature 456 (7218) (2008) 53–59.

[24] A. Voskoboynik, N. F. Neff, D. Sahoo, A. M. Newman, D. Pushkarev, W. Koh, B. Passarelli, H. C. Fan, G. L. Mantalas, K. J. Palmeri, et al., The genome sequence of the colonial chordate, Botryllus schlosseri, Elife 2 (2013) e00569.

[25] I. Sharon, M. Kertesz, L. A. Hug, D. Pushkarev, T. A. Blauwkamp, C. J. Castelle, M. Amirebrahimi, B. C. Thomas, D. Burstein, S. G. Tringe, et al., Accurate, multi-kb reads resolve complex populations and detect rare microorganisms, Genome research 25 (4) (2015) 534–543.

[26] R. A. White III, E. M. Bottos, T. R. Chowdhury, J. D. Zucker, C. J. Brislawn, C. D. Nicora, S. J. Fansler, K. R. Glaesemann, K. Glass, J. K. Jansson, Moleculo long-read sequencing facilitates assembly and genomic binning from complex soil metagenomes, mSystems 1 (3) (2016) e00045–16.

[27] V. Kuleshov, C. Jiang, W. Zhou, F. Jahanbani, S. Batzoglou, M. Snyder, Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome, Nature biotechnology 34 (1) (2016) 64–69.

[28] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, W. W. Webb, Zero-mode waveguides for single-molecule analysis at high concentrations, Science 299 (5607) (2003) 682–686.

[29] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, et al., Real-time DNA sequencing from single polymerase molecules, Science 323 (5910) (2009) 133–138.

[30] T. Laver, J. Harrison, P. O'neill, K. Moore, A. Farbos, K. Paszkiewicz, D. J. Studholme, Assessing the performance of the Oxford Nanopore Technologies MinION, Biomolecular detection and quantification 3 (2015) 1–8.

[31] M. Jain, H. E. Olsen, B. Paten, M. Akeson, The oxford nanopore MinION: delivery of nanopore sequencing to the genomics community, Genome Biology 17 (1). doi:10.1186/s13059-016-1103-0.
URL https://doi.org/10.1186%2Fs13059-016-1103-0

[32] I. A. Murray, T. A. Clark, R. D. Morgan, M. Boitano, B. P. Anton, K. Luong, A. Fomenkov, S. W. Turner, J. Korlach, R. J. Roberts, The methylomes of six bacteria, Nucleic acids research (2012) gks891.

[33] U. Mirsaidov, W. Timp, X. Zou, V. Dimitrov, K. Schulten, A. P. Feinberg, G. Timp, Nanoelectromechanics of methylated DNA in a synthetic nanopore, Biophysical journal 96 (4) (2009) L32–L34.

[34] L. W. Mendes, E. E. Kuramae, A. A. Navarrete, J. A. Van Veen, S. M. Tsai, Taxonomical and functional microbial community selection in soybean rhizosphere, The ISME journal 8 (8) (2014) 1577–1587.

[35] M. M. Newman, N. Lorenz, N. Hoilett, N. R. Lee, R. P. Dick, M. R. Liles, C. Ramsier, J. W. Kloepper, Changes in rhizosphere bacterial gene expression following glyphosate treatment, Science of the Total Environment 553 (2016) 32–41.

[36] T. R. Turner, K. Ramakrishnan, J. Walshaw, D. Heavens, M. Alston, D. Swarbreck, A. Osbourn, A. Grant, P. S. Poole, Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants, The ISME journal 7 (12) (2013) 2248–2258.

[37] P. Marschner, G. Neumann, A. Kania, L. Weiskopf, R. Lieberei, Spatial and temporal dynamics of the microbial community structure in the rhizosphere of cluster roots of white lupin (Lupinus albus L.), Plant and Soil 246 (2) (2002) 167–174.

[38] J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, et al., Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, The ISME journal 6 (8) (2012) 1621–1624.

[39] C. L. Schoch, K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen, E. Bolchacova, K. Voigt, P. W. Crous, A. N.

Miller, M. J. Wingfield, M. C. Aime, K.-D. A. et al., Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi, Proceedings of the National Academy of Sciences 109 (16) (2012) 6241–6246. doi:10.1073/pnas.1117018109.
URL https://doi.org/10.1073%2Fpnas.1117018109

[40] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F. O. Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic acids research (2012) gks808.

[41] C. M. Burke, A. E. Darling, A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq, PeerJ 4 (2016) e2492.

[42] E. Singer, B. Bushnell, D. Coleman-Derr, B. Bowman, R. M. Bowers, A. Levy, E. A. Gies, J.-F. Cheng, A. Copeland, H.-P. Klenk, et al., High-resolution phylogenetic microbial community profiling, The ISME journal 10 (8) (2016) 2020–2032.

[43] A. C. Howe, J. K. Jansson, S. A. Malfatti, S. G. Tringe, J. M. Tiedje, C. T. Brown, Tackling soil diversity with the assembly of large, complex metagenomes, Proceedings of the National Academy of Sciences 111 (13) (2014) 4904–4909. doi:10.1073/pnas.1402564111.
URL https://doi.org/10.1073%2Fpnas.1402564111

[44] K. Anantharaman, C. T. Brown, L. A. Hug, I. Sharon, C. J. Castelle, A. J. Probst, B. C. Thomas, A. Singh, M. J. Wilkins, U. Karaoz, et al., Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system, Nature Communications 7.

[45] C. T. Brown, L. A. Hug, B. C. Thomas, I. Sharon, C. J. Castelle, A. Singh, M. J. Wilkins, K. C. Wrighton, K. H. Williams, J. F. Banfield, Unusual biology across a group comprising more than 15% of domain bacteria, Nature 523 (7559) (2015) 208–211.

[46] S. Koren, A. M. Phillippy, One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly, Current Opinion in Microbiology 23 (2015) 110–120. doi:10.1016/j.mib.2014.11.014.
URL https://doi.org/10.1016%2Fj.mib.2014.11.014

[47] G. W. Tyson, J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, J. F. Banfield, Community structure and metabolism through reconstruction of microbial genomes from the environment, Nature 428 (6978) (2004) 37–43. doi:10.1038/nature02340.
URL https://doi.org/10.1038%2Fnature02340

[48] C. B. Driscoll, T. G. Otten, N. M. Brown, T. W. Dreher, Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture, Standards in Genomic Sciences 12 (1). doi:10.1186/s40793-017-0224-8.
URL https://doi.org/10.1186%2Fs40793-017-0224-8

[49] S. J. Blazewicz, R. L. Barnard, R. A. Daly, M. K. Firestone, Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses, The ISME Journal 7 (11) (2013) 2061–2068. doi:10.1038/ismej.2013.102.
URL https://doi.org/10.1038%2Fismej.2013.102

[50] H.-L. Liao, Y. Chen, R. Vilgalys, Metatranscriptomic study of common and host-specific patterns of gene expression between pines and their symbiotic ectomycorrhizal fungi in the genus suillus, PLOS Genetics 12 (10) (2016) e1006348. doi:10.1371/journal.pgen.1006348.
URL https://doi.org/10.1371%2Fjournal.pgen.1006348

[51] Y. Jiang, X. Xiong, J. Danska, J. Parkinson, Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality, Microbiome 4 (1). doi:10.1186/s40168-015-0146-x.
URL https://doi.org/10.1186%2Fs40168-015-0146-x

[52] Q. Zhang, T. G. Doak, Y. Ye, Artificial functional difference between microbial communities caused by length difference of sequencing reads, in: Pacific Symposium on Biocomputing, 2012, pp. 259–270.

[53] D. R. Garalde, E. A. Snell, D. Jachimowicz, A. J. Heron, M. Bruce, J. Lloyd, A. Warland, N. Pantic, T. Admassu, J. Ciccone, et al., Highly

parallel direct RNA sequencing on an array of nanopores, bioRxiv (2016) 068809.

[54] A. Oulas, C. Pavloudi, P. Polymenakou, G. A. Pavlopoulos, N. Papanikolaou, G. Kotoulas, C. Arvanitidis, I. Iliopoulos, Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies, Bioinformatics and biology insights 9 (2015) 75.

[55] T. Thomas, J. Gilbert, F. Meyer, Metagenomics-a guide from sampling to data analysis, Microbial informatics and experimentation 2 (1) (2012) 3.

[56] V. Kunin, A. Copeland, A. Lapidus, K. Mavromatis, P. Hugenholtz, A bioinformatician's guide to metagenomics, Microbiology and molecular biology reviews 72 (4) (2008) 557–578.

[57] R. A. White III, M. I. Borkum, et al., Integrated multi-omics of the rhizosphere: from data to knowledge, Rhizosphere.

[58] L. N. Kajiura, S. D. Stewart, J. Dresios, C. F. T. Uyehara, Simultaneous extraction of viral and bacterial nucleic acids for molecular diagnostic applications, Journal of Biomolecular Techniques : JBT (2015) jbt.15–2604–002doi:10.7171/jbt.15-2604-002.
URL https://doi.org/10.7171%2Fjbt.15-2604-002

[59] S. J. Blazewicz, E. Schwartz, Dynamics of ^{18}O incorporation from $H_2^{18}O$ into soil microbial DNA, Microbial ecology 61 (4) (2011) 911–916.

[60] P. Carini, P. J. Marsden, J. W. Leff, E. E. Morgan, M. S. Strickland, N. Fierer, Relic DNA is abundant in soil and obscures estimates of soil microbial diversity, Nature Microbiology 2 (2016) 16242.

[61] E. Schwartz, Characterization of growing microorganisms in soil by stable isotope probing with $H_2^{18}O$, Applied and environmental microbiology 73 (8) (2007) 2541–2546.

[62] A. J. Rissanen, E. Kurhela, T. Aho, T. Oittinen, M. Tiirola, Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies, Applied Microbiology and Biotechnology 88 (4) (2010) 977–984. doi:10.1007/s00253-010-2838-2.
URL https://doi.org/10.1007%2Fs00253-010-2838-2

[63] A. Pandey, M. Mann, Proteomics to study genes and genomes, Nature 405 (6788) (2000) 837–846.

[64] O. Fiehn, Metabolomics–the link between genotypes and phenotypes, Plant molecular biology 48 (1-2) (2002) 155–171.

[65] R. Aebersold, D. R. Goodlett, Mass spectrometry in proteomics, Chemical reviews 101 (2) (2001) 269–296.

[66] M. R. Viant, U. Sommer, Mass spectrometry based environmental metabolomics: a primer and review, Metabolomics 9 (1) (2013) 144–158.

[67] J. A. Loo, H. R. Udseth, R. D. Smith, J. Futrell, Collisional effects on the charge distribution of ions from large molecules, formed by electrospray-ionization mass spectrometry, Rapid Communications in Mass Spectrometry 2 (10) (1988) 207–210.

[68] P. Wilmes, A. Heintz-Buschart, P. L. Bond, A decade of metaproteomics: Where we stand and what the future holds, Proteomics 15 (20) (2015) 3409–3417.

[69] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (6928) (2003) 198–207.

[70] A. I. Nesvizhskii, A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics, Journal of proteomics 73 (11) (2010) 2092–2123.

[71] Z. Szabo, T. Janaky, Challenges and developments in protein identification using mass spectrometry, TrAC Trends in Analytical Chemistry 69 (2015) 76–87.

[72] H. Zhang, Y. Ge, Comprehensive analysis of protein modifications by top-down mass spectrometry, Circulation: Cardiovascular Genetics 4 (6) (2011) 711–711.

[73] T. Schneider, K. M. Keiblinger, E. Schmid, K. Sterflinger-Gleixner, G. Ellersdorfer, B. Roschitzki, A. Richter, L. Eberl, S. Zechmeister-Boltenstern, K. Riedel, Who is who in litter decomposition? metaproteomics reveals major microbial players and their biogeochemical functions, The ISME journal 6 (9) (2012) 1749–1762.

[74] Z. Bao, T. Okubo, K. Kubota, Y. Kasahara, H. Tsurumaru, M. Anda, S. Ikeda, K. Minamisawa, Metaproteomic identification of diazotrophic

methanotrophs and their localization in root tissues of field-grown rice plants, Applied and environmental microbiology 80 (16) (2014) 5043–5052.

[75] H.-B. Wang, Z.-X. Zhang, H. Li, H.-B. He, C.-X. Fang, A.-J. Zhang, Q.-S. Li, R.-S. Chen, X.-K. Guo, H.-F. Lin, et al., Characterization of metaproteomics in crop rhizospheric soil, Journal of proteome research 10 (3) (2010) 932–940.

[76] C. Knief, N. Delmotte, S. Chaffron, M. Stark, G. Innerebner, R. Wassmann, C. Von Mering, J. A. Vorholt, Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice, The ISME journal 6 (7) (2012) 1378–1390.

[77] E. Zampieri, M. Chiapello, S. Daghino, P. Bonfante, A. Mello, Soil metaproteomics reveals an inter-kingdom stress response to the presence of black truffles, Scientific reports 6.

[78] S. Bryson, Z. Li, J. Pett-Ridge, R. L. Hettich, X. Mayali, C. Pan, R. S. Mueller, Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton, mSystems 1 (2) (2016) e00027–15.

[79] K. Chourey, J. Jansson, N. VerBerkmoes, M. Shah, K. L. Chavarria, L. M. Tom, E. L. Brodie, R. L. Hettich, Direct cellular lysis/protein extraction protocol for soil metaproteomics, Journal of proteome research 9 (12) (2010) 6615–6622.

[80] E. S. Nakayasu, C. D. Nicora, A. C. Sims, K. E. Burnum-Johnson, Y.-M. Kim, J. E. Kyle, M. M. Matzke, A. K. Shukla, R. K. Chu, A. A. Schepmoes, et al., MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses, mSystems 1 (3) (2016) e00043–16.

[81] S. Elschenbroich, V. Ignatchenko, P. Sharma, G. Schmitt-Ulms, A. O. Gramolini, T. Kislinger, Peptide separations by on-line MudPIT compared to isoelectric focusing in an off-gel format: application to a membrane-enriched fraction from C2C12 mouse skeletal muscle cells, Journal of proteome research 8 (10) (2009) 4860–4869.

[82] F. Yang, Y. Shen, D. G. Camp, R. D. Smith, High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis, Expert review of proteomics 9 (2) (2012) 129–134.

[83] S. Wiese, K. A. Reidegeld, H. E. Meyer, B. Warscheid, Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research, Proteomics 7 (3) (2007) 340–350.

[84] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, Analytical chemistry 75 (8) (2003) 1895–1904.

[85] N. Jehmlich, F. Schmidt, M. Taubert, J. Seifert, F. Bastida, M. Von Bergen, H.-H. Richnow, C. Vogt, Protein-based stable isotope probing, nature protocols 5 (12) (2010) 1957–1966.

[86] J. Peñuelas, J. Sardans, Ecological metabolomics, Chemistry and Ecology 25 (4) (2009) 305–309.

[87] J. Sardans, J. Penuelas, A. Rivas-Ubach, Ecological metabolomics: overview of current developments and future challenges, Chemoecology 21 (4) (2011) 191–225.

[88] O. A. Jones, S. Sdepanian, S. Lofts, C. Svendsen, D. J. Spurgeon, M. L. Maguire, J. L. Griffin, Metabolomic analysis of soil communities can be used for pollution assessment, Environmental toxicology and chemistry 33 (1) (2014) 61–64.

[89] W. Lu, B. D. Bennett, J. D. Rabinowitz, Analytical strategies for LC–MS-based targeted metabolomics, Journal of Chromatography B 871 (2) (2008) 236–242.

[90] R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G. J. Patti, G. Siuzdak, An accelerated workflow for untargeted metabolomics using the METLIN database, Nature biotechnology 30 (9) (2012) 826–828.

[91] A. Rivas-Ubach, M. Pérez-Trujillo, J. Sardans, A. Gargallo-Garriga, T. Parella, J. Penuelas, Ecometabolomics: optimized NMR-based method, Methods in Ecology and Evolution 4 (5) (2013) 464–473.

[92] V. Shulaev, D. Cortes, G. Miller, R. Mittler, Metabolomics for plant stress response, Physiologia Plantarum 132 (2) (2008) 199–208.

[93] J. G. Bundy, M. P. Davey, M. R. Viant, Environmental metabolomics: a critical review and future perspectives, Metabolomics 5 (1) (2009)

3.

[94] A. Rivas-Ubach, J. Sardans, M. Pérez-Trujillo, M. Estiarte, J. Peñuelas, Strong relationship between elemental stoichiometry and metabolome in plants, Proceedings of the National Academy of Sciences 109 (11) (2012) 4181–4186.

[95] C. J. Bolten, P. Kiefer, F. Letisse, J.-C. Portais, C. Wittmann, Sampling for metabolome analysis of microorganisms, Analytical chemistry 79 (10) (2007) 3843–3849.

[96] M. M. Koek, B. Muilwijk, M. J. van der Werf, T. Hankemeier, Microbial metabolomics with gas chromatography/mass spectrometry, Analytical chemistry 78 (4) (2006) 1272–1281.

[97] A. K. Mathew, V. Padmanaban, Metabolomics: the apogee of the omics trilogy, Int J Pharm Pharm Sci 5 (2) (2013) 45–8.

[98] N. M. van Dam, H. J. Bouwmeester, Metabolomics in the rhizosphere: tapping into belowground chemical communication, Trends in plant science 21 (3) (2016) 256–265.

[99] W. B. Dunn, D. I. Ellis, Metabolomics: current analytical platforms and methodologies, TrAC Trends in Analytical Chemistry 24 (4) (2005) 285–294.

[100] I. A. Lewis, S. C. Schommer, B. Hodis, K. A. Robb, M. Tonelli, W. M. Westler, M. R. Sussman, J. L. Markley, Method for determining molar concentrations of metabolites in complex solutions from two-dimensional ^{1}H-^{13}C NMR spectra, Analytical Chemistry 79 (24) (2007) 9385–9390.

[101] A. M. Weljie, J. Newton, P. Mercier, E. Carlson, C. M. Slupsky, Targeted profiling: quantitative analysis of ^{1}H NMR metabolomics data, Analytical chemistry 78 (13) (2006) 4430–4442.

[102] P. Dalcin Martins, D. Hoyt, S. Bansal, C. Mills, M. Tfaily, B. Tangen, R. Finocchiaro, M. Johnston, B. McAdams, M. Solensky, G. Smith, Y. Chin, M. Wilkins, Abundant carbon substrates drive extremely high sulfate reduction rates and methane fluxes in Prairie Pothole Wetlands, Global Change Biology.

[103] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, et al., Proposed minimum reporting standards for chemical analysis, Metabolomics 3 (3) (2007) 211–221.

[104] L. R. Walker, D. W. Hoyt, S. M. Walker, J. K. Ward, C. D. Nicora, K. Bingol, Unambiguous metabolite identification in high-throughput metabolomics by hybrid 1D ^{1}H NMR/ESI MS1 approach, Magnetic Resonance in Chemistry 54 (12) (2016) 998–1003.

[105] J. Gullberg, P. Jonsson, A. Nordström, M. Sjöström, T. Moritz, Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry, Analytical biochemistry 331 (2) (2004) 283–295.

[106] R. C. De Vos, S. Moco, A. Lommen, J. J. Keurentjes, R. J. Bino, R. D. Hall, Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry, Nature protocols 2 (4) (2007) 778–791.

[107] A. Rugova, M. Puschenreiter, G. Koellensperger, S. Hann, Elucidating rhizosphere processes by mass spectrometry–a review, Analytica Chimica Acta.

[108] P. Dwivedi, P. Wu, S. J. Klopsch, G. J. Puzon, L. Xun, H. H. Hill, Metabolic profiling by ion mobility mass spectrometry (IMMS), Metabolomics 4 (1) (2008) 63–80.

[109] G. Astarita, G. Paglia, Ion-mobility mass spectrometry in metabolomics and lipidomics, LCGC North America 32 (9) (2015) 702—709.

[110] Y. M. Ibrahim, S. V. Garimella, S. A. Prost, R. Wojcik, R. V. Norheim, E. S. Baker, I. Rusyn, R. D. Smith, Development of an ion mobility spectrometry-orbitrap mass spectrometer platform, Analytical Chemistry 88 (24) (2016) 12152–12160.

[111] M. Ghaste, R. Mistrik, V. Shulaev, Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics, International journal of molecular sciences 17 (6) (2016) 816.

[112] M. M. Tfaily, R. Hamdan, J. E. Corbett, J. P. Chanton, P. H. Glaser, W. T. Cooper, Investigating dissolved organic matter decomposition in

northern peatlands using complimentary analytical techniques, Geochimica et Cosmochimica Acta 112 (2013) 116–129.

[113] M. J. CHRISTENHUSZ, J. W. BYNG, pstrongthe number of known plants species in the world and its annual increase/strong/p, Phytotaxa 261 (3) (2016) 201. doi:10.11646/phytotaxa.261.3.1.
URL https://doi.org/10.11646%2Fphytotaxa.261.3.1