Methods 186 (2021) 42–51

Contents lists available at ScienceDirect

Methods

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

Rumen metaproteomics: Closer to linking rumen microbial function to animal productivity traits

Thea Os Andersen[a], Benoit J. Kunath[b,c], Live H. Hagen[c], Magnus Ø. Arntzen[c], Phillip B. Pope[a,c,*]

[a] Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway
[b] Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg
[c] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

ARTICLE INFO

Keywords:
Rumen
Rumen microbiome
Metaproteomics
Methane
Metagenome-assembled genomes
Microbial diversity

ABSTRACT

The rumen microbiome constitutes a dense and complex mixture of anaerobic bacteria, archaea, protozoa, virus and fungi. Collectively, rumen microbial populations interact closely in order to degrade and ferment complex plant material into nutrients for host metabolism, a process which also produces other by-products, such as methane gas. Our understanding of the rumen microbiome and its functions are of both scientific and industrial interest, as the metabolic functions are connected to animal health and nutrition, but at the same time contribute significantly to global greenhouse gas emissions. While many of the major microbial members of the rumen microbiome are acknowledged, advances in modern culture-independent meta-omic techniques, such as metaproteomics, enable deep exploration into active microbial populations involved in essential rumen metabolic functions. Meaningful and accurate metaproteomic analyses are highly dependent on representative samples, precise protein extraction and fractionation, as well as a comprehensive and high-quality protein sequence database that enables precise protein identification and quantification. This review focuses on the application of rumen metaproteomics, and its potential towards understanding the complex rumen microbiome and its metabolic functions. We present and discuss current methods in sample handling, protein extraction and data analysis for rumen metaproteomics, and finally emphasize the potential of (meta)genome-integrated metaproteomics for accurate reconstruction of active microbial populations in the rumen.

1. Theoretical basis and framework of rumen metaproteomics

With the expected growth of the global human population from 7.8 billion in 2020 to 10 billion in 2050 [1], there is pressure to increase sustainability within agricultural industries in order to secure both animal welfare and human nutritional needs for the future. Additionally, the expansion of livestock production systems necessitates options that ensure both animal health and productivity as well as mitigate negative impacts such as greenhouse gas (GHG) emissions (e.g. CH_4) [2]. While intensive livestock industries involve different animals such as poultry, pigs and ruminants (e.g. sheep, goats and cattle), all these systems rely on close interactions between the host animal and its inherent microbiomes. Moreover, it is widely accepted that a critical means to address sustainability challenges in ruminant livestock systems is to optimize the intimate relationship between the environment (e.g. feed), the animal host and their gut microbiota, which collectively play an integral role in digesting feedstuffs into nutrients whilst producing GHG as a natural by-product [3].

Using both traditional culturing and molecular omics-based approaches, it has been inferred that differences in rumen microbiota are associated with cattle production and health phenotypes, such as feed conversion ratio (FCR) [4], methane production [5], milk composition [6], and ruminal acidosis [7]. In particular, modern meta-omic techniques can be used to accommodate the complexity of the rumen microbiome and to study microorganisms and microbial populations in their natural ecosystems without the limitations of standard cultivation methods [8,9]. The term “metaproteomics” was first coined by Wilmes and Bond in 2004 as “the large scale characterization of the entire protein complement of environmental microbiota at a given point in time” [10]. Since then, advances in sensitivity and accuracy in current mass spectrometry analysis, and development of proteomics software has made it possible to identify and quantify thousands of proteins from environmental samples and provide information on expressed function of proteins in a microbial community [11,12]. While metaproteomics hold great potential for microbial systems ecology, there are challenges in studying complex microbial communities regarding protein

* Corresponding author.
E-mail address: phil.pope@nmbu.no (P.B. Pope).

https://doi.org/10.1016/j.ymeth.2020.07.011
Received 1 April 2020; Received in revised form 12 June 2020; Accepted 27 July 2020
Available online 03 August 2020

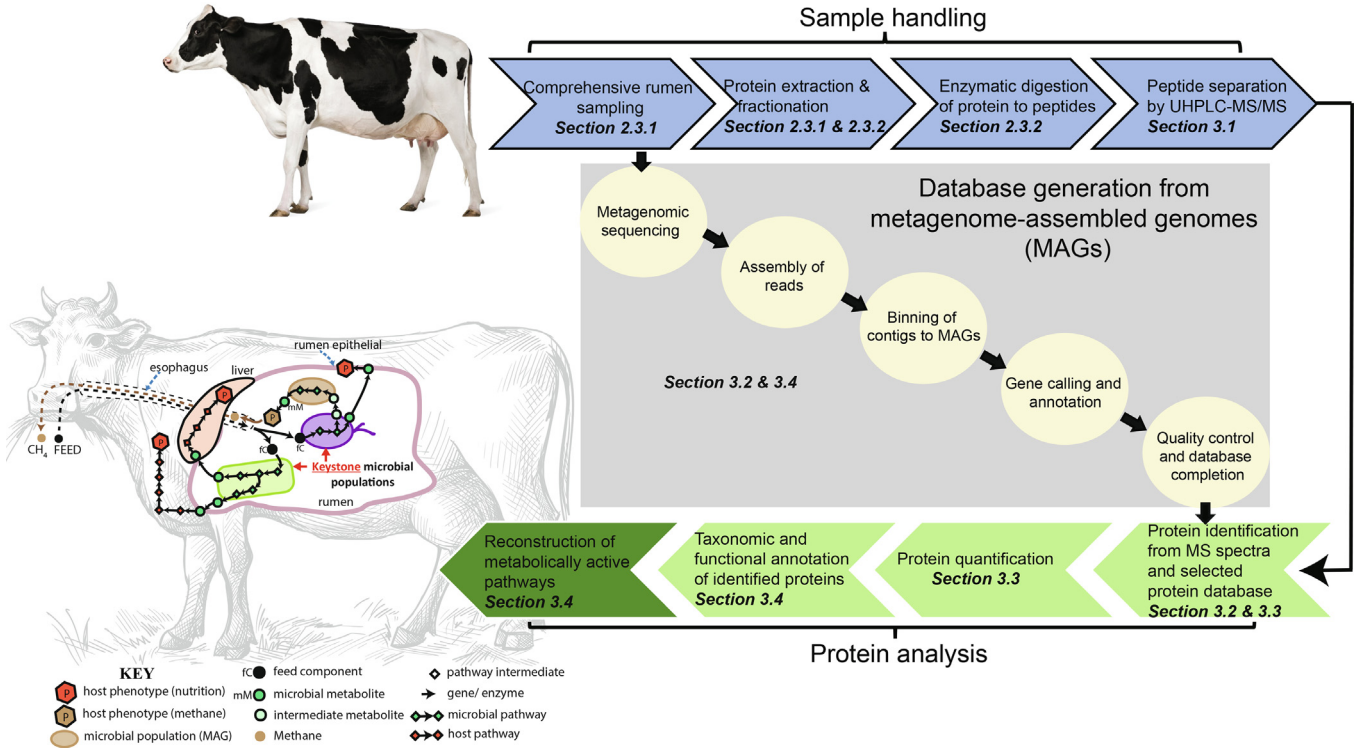


Fig. 1. Workflow for rumen metaproteomics. The ultimate goal of rumen metaproteomics is to visualize and quantify the active microbial populations, their metabolic pathways and their interactions within the microbiome as well as with the host. High-quality proteomic data depends on optimization of each step in the workflow. The workflow consists of two main phases; 1) sample handling in the lab, consisting of sampling, protein extraction and protein and peptide fractionation, and 2) protein analysis with proteomic tools, consisting of protein identification and quantification, annotation of identified proteins and finally, reconstruction of metabolically active populations in the microbiome using annotated recovered metagenome-assembled genomes (MAGs). Database design and generation is essential for accurate protein identification and quantification. The integration of sample-specific metagenomic data can provide information on novel proteins otherwise lost and allow for metabolic reconstruction of the MAGs in the rumen microbiome. The section where each step in the workflow is described is noted in the figure.

identification rates (compared to single-organism proteomics), protein extraction from environmental samples, large sequence databases, sequence unavailability, and search engine sensitivity [12].

This review will provide insight into how metaproteomics and in particular, metagenome-resolved metaproteomics can be used to visualize changes in rumen microbiome functions in larger animal experiments that aim to modify performance metrics such as feed conversion ratios and methane production or improve animal health and welfare. We will also introduce and discuss current metaproteomic methodological workflows (Fig. 1) that consider metaproteomic sampling in the rumen, identifying and quantifying proteins and visualizing active microbial populations. Finally, this review will highlight the potential of (meta)genome-resolved metaproteomics and how this added level of resolution can be applied to reconstruct active metabolic pathways, and how they function together in the rumen microbiome (Fig. 1).

2. Rumen sampling, sample preparation and protein extraction

Capturing the heterogeneity and complexity of ecological niches is a substantial challenge for meaningful metaproteomic analysis, as it is crucial that the proteins that are extracted from environmental samples reflect the original microbiome and its functional properties [12]. Given the highly dynamic and ever-changing ecosystem of the rumen, comprehensive and standardized sampling is especially important to ensure the detailed and complete “microbiome” portrait at a specific point in time. The following section will introduce and discuss some important considerations in sampling, sample handling and processing for rumen metaproteomics.

2.1. Influences on composition of the rumen microbiome

In an adult ruminant, the rumen is the biggest of the four compartments that makes up the stomach. The rumen is specialized in the degradation and fermentation of complex carbohydrates from plant biomass, made possible by carbohydrate active enzymes (CAZymes) produced predominantly by endogenous microorganisms in the rumen microbiome [13]. The rumen maintains a pH value of about 6.0–6.8 depending on the fermentation stage and a temperature at around 39 °C [15], which makes the rumen optimal for microbial growth and activity. The major microbial populations of the rumen ecosystem are well known and understood, but the relationships between these individual populations and the collective rumen function remains poorly characterized. Large and continuing efforts have been made to identify a core rumen microbiome [8,16] such as the Global Rumen Census, which represents a collaborative research effort to catalogue rumen microorganisms that are available for in depth culture analysis [17]. In contrast, other studies have focused on cataloguing uncultured microbial populations [8,9,18] or statistically correlating core rumen microbiome to host animal breeds (i.e. genotype) and production efficiency (i.e. phenotype) [16].

Henderson et al. examined rumen microbial communities across ruminant and camelid species, diets and geographical regions, and found that similar bacteria and archaea dominated in nearly all samples, while protozoal communities were more variable [8]. *Prevotella, Butyrivibrio, Ruminococcus, Lachnospiraceae, Ruminococcaceae, Bacteroidales* and *Clostridiales,* were dominant bacterial groups in over 90% of the studied samples and accordingly represent the core bacterial microbiome, yet the metabolic functions of some of these bacterial groups are not well characterized [8]. The role of eukaryotic microorganisms

in the rumen microbial ecosystem, such as protozoa and fungi, is not fully understood. Even though protozoal species can make up a substantial part, up to 50%, of biomass of the microorganisms present in the rumen [15,19], they are currently poorly characterized. The lack of appreciation is predominately due to a lack in genome reference sequences for rumen protozoal species, and the fact that laborious microscopic identification and counting remains the standard method for analysing the protozoal contribution of the rumen microbiome [19]. Moreover, although the role of rumen fungal species as fibre-degraders is acknowledged, only 11 anaerobic gut fungi from mammalian herbivores were described until a recent study expanded the taxonomic diversity with seven new genera [20]. The shortcomings are mostly due to technical difficulties associated with their cultivation as well as sequencing and analysing their eukaryotic genomes [21,22].

The products from microbial fermentation of carbohydrate biomass in the rumen are volatile fatty acids (VFAs), such as acetate, butyrate and propionate, lactate and succinate, in addition to CO_2 and H_2 [15]. VFAs are taken up by the host animal across the rumen epithelial wall and are important energy and carbon sources [15]. Hydrogen can be converted to CH_4 by methanogenic archaeal populations, which in turn is emitted during eructation and represents up to 12% of dietary gross energy loss for the ruminant [15,23,24]. The rate of passage of plant material depends on feed content, particle size and how efficiently feed is digested. Digestion efficiency and production rate of fermentation products are dependent of microbial composition in the rumen. Factors that impact microbial composition and function in the rumen are pH, temperature, the host species, age and geographic location [25], in addition to diet and dietary interventions, which have been suggested as the most influential factor for altered rumen microbial composition (and function) [26–29].

2.2. Rumen sampling

Given the scale and diversity of the rumen ecosystem, correct sampling methods that generate an accurate and reproducible reflection of the microbiome structure and function are of upmost importance. Microorganisms are associated with the liquid phase of the rumen content, attached to feed particles, and to a lesser extent also to the rumen wall [30,31]. Because of differences in density and the presence of ruminal pillars and their movements during fermentation, well-digested particles with high density sink towards the ventral parts of the rumen. In contrast gas (CO_2 and CH_4) and low-density particles that are less digested are found in the more dorsal part of the rumen, and contribute to the stratification of the rumen content [15,30,31]. Subsequently, there are higher rates of fermentation in regions where feed particles are not yet digested, i.e. the intermediate zone of the rumen content [15].

Spatial differences in rumen content are caused by both stratification and variations in microbial profiles that are associated with different fractions of rumen digesta, which can ultimately affect comprehensive and representable rumen sample collection. Studies have suggested that composition of rumen microbiota can be affected by different sampling technique, rumen fractions sampled and DNA extraction methods. For example, Henderson et al. found that community composition was generally similar irrespective of sampling via cannula or oesophageal tubing [32]. However in the same study, community composition structure differed significantly between rumen sample fractions (solid vs liquid) possibly reflecting niche-specific microbes that inhabit these varying regions [32]. While oesophageal tubing is considered a less invasive, cheaper and more accessible sampling method compared to rumen cannulation, cannula sampling allows for collection of representative and repeated samples of both liquid and solid content of the rumen [32,33]. Canula sampling additionally allows for consistent collection of representative and repeated samples for similar sites in the rumen. Henderson et al. noted that while they observed generally similar community structure for both sampling methods, the possible effects should not be disregarded [32]. Differences in relative abundance for certain taxa can be explained by tube size (when using oesophageal tubing) as it only allows for sampling of small and highly degraded fibre and plant material [32] and consequently microbial populations that dominate those particular rumen fractions.

Previous research has also demonstrated significant differences in microbial community structure between solid and liquid fractions of rumen content at broad taxonomic levels (i.e. phylum), especially in bacterial and archaeal groups [32,34–36]. However, Ji et al. found that although the rumen bacterial diversity was biased in different fractions, it was predominately affected by individual cow and diet rather than rumen fraction [37]. Correspondingly, Vaidya et al. observed differences in microbial composition in solid and liquid fractions of the rumen content yet questioned if these differences were the result of the physical nature of the fractions or due to differences in microbes present in the fractions [34]. Irrespective of these findings, it must be considered that microbiome structure and function are two separate entities and that structural measurements (such as 16S rRNA gene analysis) may not necessarily match functional activities. Variation in community function will likely lead to subsequent variation in VFA production levels and pH levels. The considerations for sampling mentioned in this review illustrate the complex and dynamic rumen microbiome, and we further recognise that analyses of rumen microbial community structure and function are still required to develop a standardization of protocols for rumen sampling and sample handling for comparable results.

2.3. Recovering the rumen metaproteome

As mentioned, the rumen microbiome contains prokaryotes, eukaryotes and viruses, and collected rumen samples for metaproteomic analysis should reflect this heterogeneity as accurately as possible. Once samples are collected optimal protocols for protein extraction must be used that enable unbiased and complete portrayal of the rumen metaproteome. If storing is necessary, we recommend flash freezing samples and storing at −80 °C in order to minimize the activity of proteases on protein abundance in environmental samples [38,39]. Protein extraction from environmental samples requires multiple steps, including protein clean-up and protein separation/fractionation [39,40], for which standardized methods will be discussed below [31,41–43].

2.3.1. Sample handling and protein extraction

Flash freezing of whole rumen samples and storing at −80° until further processing is a common storage method, and is used to minimize the effect of natural proteases that can have a detrimental impact on microbial protein abundance [38,39]. However, a consideration to this method can be drawn via a study by Martinez-Fernandez et al., which discovered that centrifuging fresh rumen fluid, removing the supernatant and freezing the cell pellet increased the abundance of readily lysed gram-negative bacterial species in a DNA sequencing effort, compared to the standard immediate flash freezing for whole samples [44]. Martinez-Fernandez et al. concluded that while further analyses are needed to confirm their results, their findings indicate that flash freezing rumen samples and using the cell pellet for analysis can alter the abundance of genetic material from species that were easily lysed [44]. In order to detach microbial cells from undigested plant matter and disrupt all aggregated cell types in the rumen (e.g. biofilms), common extraction methods often combine chemical and mechanical cell lysis to maximize protein recovery from the sample(s) [39,45]. Such protocols combine a chemical lysis method using a lysis buffer with moderate to high concentrations of detergent with a mechanical disruption of cells, such as bead-beating or sonication with heat [31]. To enhance denaturation of proteins, sodium dodecyl sulfate (SDS) can be added in moderate to high concentrations (0.1–5%) [12,46]. Natural

protease activity in the samples may be a problem if no detergents or denaturing agents are used, which is typical for so called gel-free approaches. This may result in non-tryptic or semi-tryptic peptides and cause low identification yields during database searches. Natural protease activity can be restricted with the use of protease inhibitors, such as phenylmethylsulfonyl fluoride (PMSF) or by complex protease inhibitor cocktails. However, addition of protease inhibitors that are peptides may increase complexity of downstream analysis [47]. From experience, SDS will also increase the solubility of proteins and typically increase protein recovery rather than lysis buffer without detergents.

2.3.2. Sample processing and protein fractionation

While humic substances has been proven to be beneficial for rumen microbial growth, they are also known to possibly decrease the number of identified proteins and the protein coverage percentage in proteomic analysis and should be removed from the protein extract before further proteomic analysis [48,49]. Trichloroacetic acid (TCA) is known to be an effective precipitation agent in order to remove humic or other interfering substances, yet TCA precipitation is generally connected with potential loss of big proteins and difficulty in re-solubilizing proteins [12,39,49]. Other commonly used precipitation agents are phenol, acetone, acetone/deoxycholate, methanol/ammonium acetate and methanol/chloroform [48]; however, some affiliated with poor protein recovery (reviewed in [12]). For meta-omic analysis of rumen samples, the effect of humic interference on further proteomic analysis needs to be weighed against the effect of potential poor protein recovery due to precipitation. As meta-omic techniques aim to reflect the unaffected, unbiased and complete composition and/or function of the environmental microbiomes, any (potentially biased) loss of proteins can have a detrimental impact on our understanding of the functions of microbial communities.

The metaproteome of the rumen microbial community consists of a diverse mixture of proteins due to the complexity of the rumen microbial community. Efficient separation of the large number of proteins present is crucial for accurate protein identification and quantification, and studies have indicated that pre-fractionation can lead to increased protein identification [50]. To ensure accurate and high-resolution protein quantification, a typical bottom-up proteomic workflow is often employed, including one or more fractionating steps to separate proteins and/ or peptides [40,51]. While modern proteomics software often have imbedded normalization algorithms to ensure unbiased quantification between samples (e.g. the delayed normalization algorithm MaxLFQ in MaxQuant [52]), evaluation of protein concentration can be advantageous prior to sample fractionating to ensure equal amounts of protein from each sample is analysed downstream [39]. However, when using a lysis buffer with high levels of detergent it is essential to use detergent-compatible protein concentration assays to ensure accurate estimates, as detergents can bind to proteins and compete with the dye reagent (e.g. in the classic Bradford protein assay method), causing underestimations of protein concentration [53]. Since the development of metaproteomics, two-dimensional (2D) gel-based protein separation has been the most widespread method for protein separation. But in recent years, 2D-gels have been replaced by gel-free or one dimensional (1D) gel separations methods, such as SDS – polyacrylamide gel electrophoresis (SDS-PAGE). This is mainly due to drawbacks of 2D-gel analysis regarding recovery of large and hydrophobic proteins and because of recent developments in protein/peptide labelling techniques, liquid chromatography and high resolution-mass spectrometry [39,54]. In addition, modern proteomic experiments use an increasing number of samples, rendering (tedious) 2D-gel analysis impractical. Thus, after protein extraction, current protocols include fractionation of proteins by either one dimensional gel electrophoresis, e.g. SDS-PAGE, followed by *in-gel* digestion, or digested directly *(in-solution*) into peptides commonly by trypsin, a specific and non-selective serine protease [55]. SDS-PAGE can also be used as a powerful clean-up technique for removing interfering substances. SDS-PAGE is an inexpensive and easy method for separating proteins from interfering substances, which does not enter the gel or, if charged, passes through the gel very fast, while at the same time fractionating proteins based on protein size [12]. Further, cysteine residues in peptides are usually reduced with a strong reducing agent, such as dithiothreitol (DTT), and alkylated, with e.g. iodoacetamide (IAA), prior to peptide clean up and mass spectrometry (MS) analysis. Studies have reported issues in protein extraction in SDS-PAGE gel possibly due to co-precipitation of interfering humic compounds deriving from plant material [56,57]. As mentioned, sample preparation protocols often include a phenol extraction or other precipitation agents such as TCA or acetone in order to remove the majority of humic compounds known to interfere with downstream analysis. However, precipitation is also connected with potential protein loss and has to be weighed against the effect of humic interference. A recently published study by Honan and Greenwood recommend the use of non-gel-based fractionation methods to characterize and identify the rumen metaproteome in order to overcome the effect of poor protein recovery or protein loss when using SDS-PAGE for the purpose of fractionation [58].

3. Protein identification and quantification

3.1. Peptide separation and UHPLC-MS/MS

Today, MS-based proteomics is the most commonly applied approach in “shotgun” metaproteomics. Electrospray ionization (ESI) MS enables high-resolution analysis of peptides due to its convenient coupling with liquid peptide-separation techniques such as ultra-high-performance liquid chromatography (UHPLC). Furthermore, as the mass spectrometer is more sensitive to low molecular-weight molecules, separation of peptides aid in achieve deep proteomic coverage and high analytical resolution [59]. Prior to UHPLC-MS analysis, peptides can be concentrated, cleaned, and further fractionated though binding, e.g. to a reverse-phase material in microcolumns, often C^{18} [60,61]. Typically, StageTips or ZipTips® (Merck-Millipore, cat. no. Z720070) are used for this purpose and resulting peptides are subsequently eluted with organic solvents. The hydrophobic stationary phase in C^{18} columns ensures easy purification and concentration of peptides for MS analysis, and can increase stability of UHPLC-MS/MS systems by removing impurities, such as gel pieces, hence preventing clogging of the column used for UHPLC [60,61]. The duration of the UHPLC gradient is selected based on sample complexity and extent of protein fractionation; for metaproteomics, where sample complexity usually is high, long gradients of 90–240 min are often used. Peptides are then analysed by the mass spectrometer, where their mass-to-charge (m/z) ratios can be analysed, generating mass spectra or fragmented tandem spectra (MS/MS) [59]. For metaproteomics, due to sample complexity and large protein databases (explained in Section 3.2), it is of immense importance that the mass spectrometer used is of high-resolution and high accuracy, to limit the list of potential sequences matching to one MS/MS spectrum.

Protein identification in proteomics can be achieved through: 1) *de novo* sequencing, by interpreting the amino acid sequence directly from MS/MS spectra and following identification of the protein using sequencing-similarity search engine such as BLAST, 2) peptide sequence identification using spectral libraries, or 3) theoretical matching - the most common strategy in metaproteomics. Theoretical matching identifies detected proteins by matching experimental MS/MS spectra to theoretical peptide fragmentation patterns from *in silico* peptide digestion of a user specific protein sequence database. This approach has proven to be successful in protein identification and quantification even in large-scale soil metaproteomics studies, a microbial niche even more complex than that of the rumen microbiome [39,62].

3.2. Database selection and assembly

Compared to proteomics of single organisms, metaproteomics faces several challenges because of increased complexity and heterogeneity of samples. Rumen microbial communities are estimated to consist of hundreds to thousands of different species [8]. This inherent diversity means the rumen microbiome likely encodes up to several million proteins, even before further estimations are made due to alternative gene splicing and post translational modifications (PTMs), which can cause the number of expressed proteins to exceed the number of genes in an organism [63]. In addition, many species consist of closely related proteins, due to e.g. strain variations or horizontal gene transfer [42,64]. Therefore, accurate quantification and identification of proteins from the rumen microbiome is highly determined by the database choice [65]. An optimal protein sequence database should be comprehensive, of high-quality and should theoretically include protein sequences for all proteins expected to be expressed in the microbial community and detectable via MS, as well as potential contaminants such as MS standards and human keratins [42].

Creation of rumen-specific protein sequence databases can be based on the use of complete public protein sequence repositories that can be further refined by 16s rRNA analysis of the sample at hand (referred to as *pseudo*-meta*genomic* databases [42,66]) or metagenome assembled genome (MAG) inventories [21,67]. While it is easy to think that an increase in search-space (i.e. database) subsequently will increase the likelihood of identifying detected proteins, protein identification as such is not trivial. As large, non-sample specific database can represent only a fraction of the microbial species present in a given microbial sample, protein identification with aforesaid databases may lead to an increased fraction of false positive hits, low identification rates and few significant hits [12]. The generation of rumen-specific MAG catalogues is expanding rapidly [9,18,68], however it has been highlighted that individual variation of rumen microbiota exist in both beef [69] and dairy cattle [70], even when animals are fed the same diet and managed under the same environment. Therefore, even with steps taken to customize a rumen database from sequence/MAG repositories, it is likely that protein identification will still be sub-optimal if the protein sequence database is missing protein entries for present proteins or even missing species [42].

Sample-specific databases, generated using shotgun metagenomics data from the same sample from which MS raw data are produced, are considered far superior to any of the above mentioned options as they will encompass individual variation that potentially exists in a given sample [12,42]. While the integration of sample-specific metagenomic sequence data with metaproteomic analysis requires increased financial and processing efforts compared to the use of publicly available sequence data, there are clear advantages. Not only do metagenome-integrated protein databases include protein sequences for (nearly all) expected expressed proteins by the target microbial community, they also restrict the size of the database, making it both more complete yet concise, which minimizes issues associated with false positive identifications. Moreover, the integration of sample-specific metagenomic data can provide information on expressed proteins from novel and uncultured microorganisms that are not present in public sequence repositories, and with this contribute to our increased understanding of their metabolic functions in their innate microbial ecosystem. Lastly, and importantly, the integration of metagenomics and metaproteomics does not only enhance metaproteomic detection, but also allows for metabolic reconstruction and functional assessment of the individual MAGs (see Sections 3.4.1 and 3.4.2).

Database generation from metagenomic data generally consists of five main steps; 1) metagenomic sequencing of the community, 2) assembly of reads into longer, continuous and overlapping segments of DNA (contigs), 3) binning of contigs into MAGs, 4) gene calling, and 5) functional annotation [12]. A detailed description of the different steps of metagenomic database generation is provided in [12]. Briefly, the sequencing should aim to comprise both dominating and rare microbial populations in the samples, thus reflecting the microbial composition and its complexity as accurately as possible. Short read techniques, creating reads of about 150–300 base pairs (bp), can generate a tremendous amount of data with high sequencing depth and low error rates. While long read techniques (PacBio and Oxford Nanopore) have a reduced sequencing depth and increased error rates, there are also several advantages of incorporating long reads into high-quality assemblies for the purpose of metagenomic database creation [71].

Once assembled, larger contigs are binned into MAGs, where each assembled contig is assigned to (ideally) one population-level genome. Moreover, to avoid miss-assemblies and inaccurate genomic assignments, results from binning and assembly should always be inspected. After assembly and binning of the metagenomic dataset, protein coding regions in the different MAGs are identified through prediction of open reading frames (ORF) in a process called gene calling. Functional annotation of ORFs can be performed using a multitude of different computational resources and is described in more detail in Section 3.4 below. In most cases, it will also be useful to perform a taxonomical assignment of the contigs or the MAGs. As there is currently no standard workflow for *de novo* genome assembly from complex microbial communities, choice of sequencing technique, assembler and binning software should be determined based on target community and research goals. In this context, the Critical Assessment of Metagenome Interpretation (CAMI) is an example of an initiative that aims to benchmark software selection to answer specific research questions [72]. Today as standard practice, our lab routinely uses MAG-centric databases for the purpose of protein identification and quantification, as it provides targeted knowledge into individual population activity, and at the same time reduces database size and complexity.

3.3. Protein identification and quantification

As mentioned, protein identification can be achieved in different ways, yet the most common method in metaproteomics is the use of a database and match experimental MS/MS spectra to theoretical fragment patterns from *in silico* digestion of the database. MS raw data contain information on peptide *m/z* ratio and intensity and can be used to identify and quantify peptides in proteomics software through search engines, e.g. Andromeda [73] and Mascot [74]. There are also search engines and workflows designed to overcome challenges specific for metaproteomics, e.g. database size; these include among others ComPIL workflow [75] and ProteoStorm [76]. For reviews on proteomic search engines, see [77,78].

In order to discriminate correct protein identifications from incorrect identifications, strict control of the false discovery rate (FDR) is necessary. The most common approach to control the FDR is the target-decoy approach, where reversed or randomly scrambled (decoy) sequences are included in the target database and peptide-to-spectrum matches (PSMs) to the decoys are considered false positive identifications. FDR can be estimated as the ratio of the number of decoy matches above a threshold to the number of database hits above the same given threshold, to give a FDR of e.g. 1% [12]. The target-decoy approach has been shown to be less sensitive for FDR estimation in metaproteomic analysis as increased number of analytes and high sequence similarity in metaproteomics hampers the differentiation between correct and false identifications [64]. To regain some sensitivity, alternative options have been suggested, such as the use of multiple search engines to increase protein identifications for metaproteomics [64].

Label-free quantification (LFQ) as a measure of protein abundance has become the method-of-choice in metaproteomics, as it can be applied directly to protein identification data and does not suffer from potentially error-prone labelling of proteins or peptides, which could lead to further challenges regarding analytical reproducibility for vastly complex samples [79]. Yet, recent studies suggest that accurate and precise LFQ still suffers from heterogeneity and complexity of microbial

samples and lack of standardized protein sample protocols, complicating the quantification of subtle changes or rare identifications in microbial samples [79]. In recent years, a variety of bioinformatic approaches have been developed for acquisition, quantification and processing methods, such as transformation, normalization, missing value filtering, imputations and match-transfer between samples [52], to increase precision and accuracy, and enhance reproducibility of LFQ in metaproteomic studies [77,78]. One such tool is ANPELA, an open access server constructed to enable performance assessment of quantification workflows for the facilitation of accurate proteome quantification for metaproteomic research [78]. Further, the MetaProteomeAnalyzer is a protein quantification tool that includes multiple search engines and groups similar proteins into meta-proteins to overcome some of the challenges [80]. Other robust and commonly used quantification software include MaxQuant [81], the MetaPro-IQ workflow [82], as well as new developments designed for the Galaxy framework, such as metaQuantome [83].

3.4. Biological interpretation of identified and quantified proteins

The final step in the metaproteomic analysis is the biological interpretation of the acquired protein identifications and abundances through taxonomic and functional annotation. Meaningful metaproteomics rely heavily on accurate functional assignment of identified proteins in order to reconstruct active microbial populations and their metabolic pathways. A considerable challenge in metaproteomics, with direct implications for taxonomic and functional annotation, is protein inference [42,84]. Peptide sequences can be unique to a single protein, but often, and more so for shorter peptides, the peptide sequence can match to several proteins with similar sequences. The mass spectrometer can be tuned in order to disregard peptides under a specific length in order to reduce the number of interfering peptides in protein identification and taxonomic evaluation. Peptide-to-protein-to-species inference is more difficult for large databases, and especially for conserved proteins; however, when using a sample-specific database constructed from MAGs, this process will not be precluded by taxa that are not present in the sample. This is important, and one of the reasons why sample-specific databases are most favourable in metaproteomics. However, in the absence of such databases, options exist. Muth et al. suggested a workflow for taxonomic evaluation where identified proteins are submitted to protein BLAST [85] as a pre-processing step, and subsequent results are analysed with MEGAN [86] to compute a phylogenetic tree [64]. An alternative strategy is to estimate the taxonomy of identified proteins using the lowest common ancestor (LCA) of all the peptides matching to a protein. This method is used in the proteomic software MetaProteomeAnalyzer, as described above [80]. UniPept is another tool for taxonomic annotation of peptide/protein sequences, which also include visualization feature [87].

Functional annotation of proteins makes it possible to reconstruct active metabolic pathways from environmental samples and contributes to our understanding of functions of active populations in complex microbial communities. Several publicly available databases can be used for functional annotation of identified proteins, e.g. UniProtKB. The Gene Ontology (GO) Consortium aims to be the largest source of functional gene information, ranging from molecular functions and biological processes to organism level [88]. The InterPro database represents protein domains, families and functional sites from multiple other protein databases [89], such as Pfam [90], and can therefore reveal functions or domains on otherwise uncharacterized protein and contribute to an expanded functional understanding [64]. There are also specialized functional databases, such as the CAZy database [13], consisting of around 300 families with carbohydrate active enzyme modules, contributing to the understanding of carbohydrate degrading systems, such as the rumen.

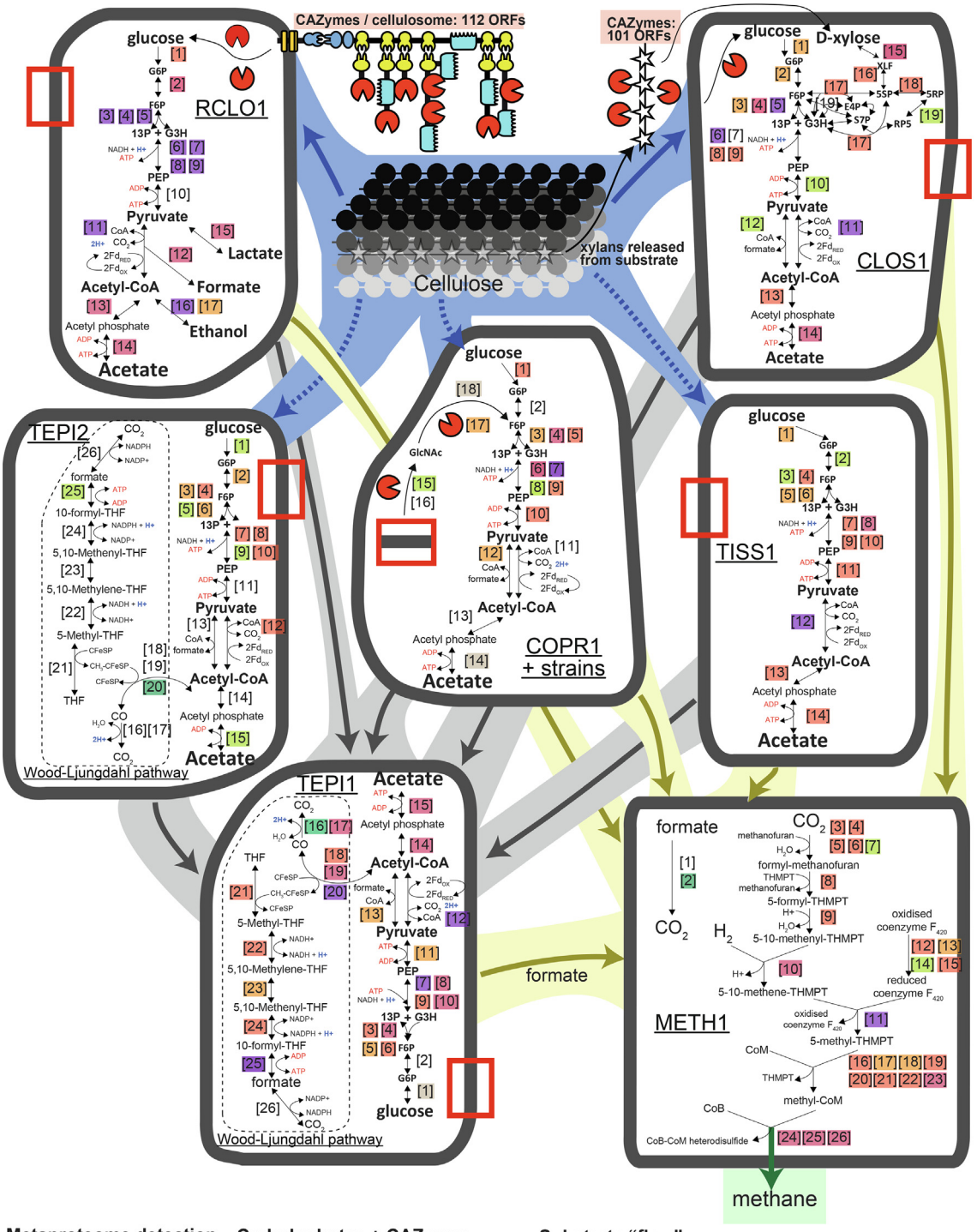
For metabolic pathway analysis, the Kyoto Encyclopedia for Gene and Genomes (KEGG) integrates genomic, biochemical and functional information into a visualizing feature called PATHWAY to facilitate the reconstruction of metabolically active pathways [91]. This can be achieved through their metagenomic annotation tool, GoastKOALA [92], with annotation of MAGs and their ORFs using Enzyme Commission (EC) numbers, that represents catalytic activities, which are then mapped on KEGG pathways to highlight abundant and active pathways in a microbial community. Quantitative expression patterns can be complemented with functional information in software tools such as Perseus [93], to visualize expression data and generate heatmaps. A survey of metaproteomic software tools for functional microbiome analysis is described in [94]. Finally, over the past decades several initiatives have been started in order to make proteomics data publicly available, such as the Proteomics Identifications database (PRIDE) [95], PeptideAtlas [96], the Global Proteome Machine Database (GPMDB) [97] and MassIVE [98]. Many of these repositories are now part of the ProteomeXchange Consortium [99] and collectively contains approximately 157 metaproteomic studies (March 2020).

3.4.1. Rumen metaproteomic studies: the current state of the art

In many ways, rumen metaproteomics is still in its infancy and these limited studies include rumen metaproteomes from adult sheep [56], pre-weaned lambs [100], cows [21,31,101,102] and moose [67]. Metagenome-centric studies have been used to highlight the active proteins and saccharolytic machineries that are used by different rumen microbiota, in particular polysaccharide utilization loci by gram negative Bacteroidetes and cellulosomes by anaerobic fungi, which were surprisingly detected at higher detection levels than their bacterial counterparts [21]. Moreover, we recently combined both metagenome-centric metaproteomics and biochemistry to identify and describe a novel Bacteroidetes family (“Candidatus MH11”) composed entirely of uncultivated strains that are predominant in ruminants [103]. While these aforementioned examples have focused on selected populations and their activity, broader community-wide metaproteomic efforts have generated metabolic networks that reveal highly connected “hub” populations hypothesized to be of central importance to the greater rumen microbiome function [67]. While all metaproteomic studies to date have detected important microbial functions, they lack both deep functional resolution at a population level *as well as* broader systems-wide metabolic networks that are required to ultimately connect rumen microbiome function to phenotypic traits in the host animal (Fig. 1). We believe that in order to reach this level of understanding, steps must be taken to connect active metabolic functions, i.e. genes and pathways that are “switched-on” in a host and its microbiome alike. Such “holoomic” approaches that integrate metadata (i.e. feed, host traits) and different levels of molecular data (DNA, RNA and protein) from both host and microbes are hypothesized to reveal functional interactions that would otherwise remain undetected.

3.4.2. Case examples: metabolic visualization of anaerobic digestion

Genome-wide association studies with cows has identified heritable rumen bacteria [16], and it has also been demonstrated that genetic variation in cows can lead to differences in microbial gene/taxa abundance and methane production [104,105]. However, these studies have all relied on the relative abundance of microbial DNA only (singular genes such as 16S rRNA) and are yet to elucidate how the expressed metabolic pathways at a profound functional level within (multiple) microbial populations, are linked to host phenotypes. While there are only a few rumen metaproteomic datasets that follow the metagenome-centric workflow described in this review, we believe such approaches will enable a deeper mechanistic level of understanding into rumen microbiota that are correlated to host genetic traits and/or desirable phenotypes such as high feed efficiency and/or low methane emissions. While to the best of our knowledge protein-mediated pathway analysis of methanogenesis have not yet been reconstructed from the rumen, metagenome-centric metaproteomics approaches have been used at an enrichment scale to visualize microbial interactions



(caption on next page)

Fig. 2. Selected metabolic features of a cellulose degrading, methane-producing enrichment as inferred from genome and proteome comparisons. The different metabolic pathways are displayed for each of the seven recovered population (MAG). The SEM1b consortium is composed of seven populations, including two saccharolytic bacteria (RCLO1 and CLOS1), one sugar fermenter (TISS1), two syntrophic acetate-oxidizing bacteria (TEPI1 and TEPI2), one hydrogenotrophic methanogen (METH1) and at least three strains of a generalist bacteria Coprothermobacter proteolyticus, which is represented in this figure by one MAG (COPR1).Graphical representation of pathways (inferred from EC and KEGG annotation), enzymes, CAZymes, and cellular features are based on functional annotations and metaproteomic data. LFQ values for detected proteins from one time point are indicated as numbered boxes. Main transfers of key metabolites (carbohydrates, acetate, formate, hydrogen and methane) are represented by colour-highlighted arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

involved in digesting plant fibre to VFAs and methane (Fig. 2) [106]. In this study, metagenomic DNA and temporal protein datasets from a highly efficient fibre-degrading and methane-producing consortium (SEM1b) were combined, leading to the identification of about 7500 proteins from seven populations (MAGs). Using these proteins, we highlighted the importance of database selection and construction for accurate and precise protein identification, thus allowing a greater understanding of microbial community function, as well as accurate monitoring of community members over time. Moreover, by using comparative metaproteomics and EC/KEGG annotations, this study highlighted the different metabolic pathways expressed by SEM1b populations which was used to reconstruct the “carbon flow” of the community from hydrolysis of polysaccharides to production of methane (Fig. 2).

In this particular example of SEM1b [106], we generated metagenomic data from replicate enrichment samples using the Illumina HiSeq3000 platform (Section 3.2). Reads were further co-assembled and subsequent contigs binned, before the quality of recovered MAGs were assessed. After genomic feature prediction and KEGG annotation, the annotated ORFs were used as a reference database for metaproteomic analysis (Section 3.4). Proteins were extracted from four time points and centrifuged, prior to cell lysis with buffer and mechanical disruption (Section 2.3.1). Extracted proteins were quantified using the Bradford method and separated by SDS-PAGE (Section 2.3.2). Further, each gel lane was cut into slides and reduction, alkylation and tryptic digestion was performed in gel, as described above (Section 2.3.2). Tryptic peptides were extracted from the gel and desalted, prior to UHPLC-MS/MS analysis, and eluted using 90-minute gradients (Section 3.1). The total 192 MS raw files were analysed using MaxQuant [81], where common contaminants were removed and reversed sequences of protein entries was used for FDR estimation (Section 3.3). Identifications were filtered to achieve a FDR 1%. Taxonomy was assigned to protein groups in instances where all proteins within the group originated from the same species, otherwise only protein function were recovered. Identified proteins were quantified using the log of their LFQ intensities and expression values were analysed using hierarchical clustering (Section 3.3). The KEGG annotations and the expression profiles of the proteins were retrieved and the main metabolic pathways for each community members were identified, as shown in Fig. 2 (Section 3.4).

Due to the high number of proteins identified, complete pathways involving various stages of carbon metabolism (hydrolysis, fermentation, VFA oxidation and methanogenesis) could be detected for most of the different populations identified in this study. Collectively, four of the MAGs were predicted to generate common fermentation products such as hydrogen, carbon dioxide and acetate. While issues such as incomplete genomic information or difficulties to distinguish closely related strains still made this task difficult, this high level of protein mapping enabled the visualization of metabolic activities of every member of the community over time.

4. Future perspectives and conclusions

Today, it has become commonplace for researchers to apply meta-omic techniques in order to recover and reconstruct composition and functions of complex microbiomes. For the rumen ecosystems, such tools can be used to enhance the understanding of how the rumen microbiome is linked to methane production and/or performance measures such as VFA production and feed conversion ratios. Integrated meta-omic techniques are used in current rumen studies to show and understand how the active rumen microbiome relate to low or high methane yield [107,108], and feed efficiency [69], while both Kamke et al. and Li & Guan performed integrated metatranscriptomics studies where they concluded that compositional and functional characterization of the rumen microbiome can serve as foundation to understand rumen functions and be used as screening tools for methane yield and feed efficiency. The power of multi-omic approaches enhances our ability to better understand how manipulation of the rumen (e.g. with dietary interventions) affects methane production and performance metrics via our ability to visualize what microbial populations and metabolic pathways are active. Despite these progressions, the rumen research community has been slow to adopt metaproteomic analyses, however this is beginning to change as methodology rapidly improves. Finally, an additional hurdle is how rumen metaproteomic data can be integrated with host multi-omic or meta-data to better understand the host-microbiome-environment axis (i.e. the holobiont) and how it affects animal health and productivity. While this element of “holo-omics” is still in its infancy, several approaches are worthy of consideration including constructing of co-expression networks [109] as well as constraint-based models [110] that contain both host and microbiome multi-omics/metadata. We believe both methods show promise to enable researchers to follow the flow of feed components through multiple keystone microbial populations and into host tissue where it is metabolized.

In conclusion, this review has introduced and discussed current methods and considerations for accurate and meaningful metaproteomic analysis of the complex rumen microbiome. Metaproteomics can be added as a functional layer to metagenomic data displaying microbial composition for increased understanding of symbiotics relationships and metabolically active populations in the rumen. In this method-based review we have provided a “up-to-date” workflow of metagenome-integrated metaproteomics and exemplified how this can increase our protein identification. While there exists a multitude of protocols, software and tools in metaproteomic analysis, rumen metaproteomics is still dependent on standardized methods regarding sampling, protein extraction and protein identification and quantification for comprehensive metaproteomic analysis. We also show how metagenomic integration of metaproteomics can serve as an added level of resolution and how this can be utilized to reconstruct active metabolic pathways and visualize the “flow” of specific microbial activities and metabolites, such as hydrogen, methane or carbon. We envisage that coupling high-resolution metaproteomic approaches to broader genetic and phenotypic association-based analyses, will create a deeper “systems-wide” level of understanding into the interactions between the animal host and its microbiota (i.e. the holobiont) and their effects on the production efficiency.

Acknowledgements

We are grateful for support from The Research Council of Norway (Project no. 250479), the Novo Nordisk Foundation (Project no. 0054575 - SuPAcow), Norwegian Centennial Seed Grant (Project no.

720500) and the European Research Commission Starting Grant Fellowship (Project no. 336355 - MicroDE).

References

[1] 2019 Revision of World Population Prospects https://population.un.org/wpp/ (accessed 06 August 2020).
[2] C. Mbow, et al., Food Security, in: P.R. Shukla, et al. (Ed.), Climate Change and Land, Intergovernmental Panel on Climate Change, 2019.
[3] S.A. Huws, et al., Addressing global ruminant agricultural challenges through understanding the rumen microbiome: past, present, and future, Front. Microbiol. 9 (2018) 2161.
[4] S.K.B. Shabat, et al., Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants, ISME J. 10 (2016) 2958–2972.
[5] R.J. Wallace, et al., The rumen microbial metagenome associated with high methane production in cattle, BMC Genomics 16 (2015) 839.
[6] E. Jami, B.A. White, I. Mizrahi, Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency, PLoS One 9 (2014).
[7] J.C. McCann, et al., Induction of subacute ruminal acidosis affects the ruminal microbiome and epithelium, Front. Microbiol. 7 (2016) 701.
[8] G. Henderson, et al., Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range, Sci. Rep. 5 (2015) 14567.
[9] R.D. Stewart, et al., Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery, Nat. Biotechnol. 37 (2019) 953–961.
[10] P. Wilmes, P.L. Bond, The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms, Environ. Microbiol. 6 (2004) 911–920.
[11] P. Wilmes, A. Heintz-Buschart, P.L. Bond, A decade of metaproteomics: where we stand and what the future holds, Proteomics 15 (2015) 3409–3417.
[12] B.J. Kunath, et al., Metaproteomics: sample preparation and methodological considerations, Adv. Exp. Med. Biol. 1073 (2019) 187–215.
[13] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucl. Acids Res. 42 (2014) D490–D495.
[15] O.V. Sjaastad, K. Hove, O. Sand, Physiology of Domestic Animals, Scan. Vet. Press, 2016, pp. 629–724.
[16] R.J. Wallace, et al., A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions, Sci. Adv. 5 (2019) eaav8391.
[17] R. Seshadri, et al., Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection, Nat. Biotechnol. 36 (2018) 359–367.
[18] R.D. Stewart, et al., Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen, Nat. Commun. 9 (2018) 870.
[19] C.J. Newbold, G. De La Fuente, A. Belanche, E. Ramos-Morales, N.R. McEwan, The role of ciliate protozoa in the rumen, Front. Microbiol. 6 (2015) 1313.
[20] R.A. Hanafy, et al., Seven new Neocallimastigomycota genera from wild, zoo-housed, and domesticated herbivores greatly expand the taxonomic diversity of the phylum, Mycologia (2020) 1–28.
[21] L.H. Hagen, et al., Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber, bioRxiv (2020), https://doi.org/10.1101/2020.01.16.907998.
[22] R.J. Gruninger, et al., Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential, FEMS Microbiol. Ecol. 90 (2014) 1–17.
[23] D.E. Johnson, G.M. Ward, Estimates of animal methane emissions, Environ. Monit. Assess. 42 (1996) 133–141.
[24] P.H. Janssen, M. Kirs, Structure of the archaeal community of the rumen, J Appl. Environ. Microbiol. 74 (2008) 3619–3625.
[25] N. Malmuthuge, Understanding host-microbial interactions in rumen: searching the best opportunity for microbiota manipulation, J. Anim. Sci. Biotechnol. 8 (2017) 8.
[26] S.J. Noel, et al., Rumen and fecal microbial community structure of holstein and jersey dairy cows as affected by breed, diet, and residual feed intake, Animals 9 (2019) 498.
[27] J. Song, et al., Effects of sampling techniques and sites on rumen microbiome and fermentation parameters in hanwoo steers, Microbiol. Biotechnol 28 (2018) 1700–1705.
[28] H.A. Paz, C.L. Anderson, M.J. Muller, P.J. Kononoff, S.C. Fernando, Rumen bacterial community composition in Holstein and Jersey cows is different under same dietary condition and is not affected by sampling method, Front. Microbiol. 7 (2016) 1206.
[29] F. Li, et al., Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle, Microbiome 7 (2019) 92.
[30] J. Miron, D. Ben-Ghedalia, M. Morrison, Invited review: adhesion mechanisms of rumen cellulolytic bacteria, J. Dairy Sci. 84 (2001) 1294–1309.
[31] S. Deusch, J. Seifert, Catching the tip of the iceberg–evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota, Proteomics 15 (2015) 3590–3595.
[32] G. Henderson, et al., Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities, PLoS One 8 (2013) e0074787.
[33] T. Geishauser, A. Gitzel, A comparison of rumen fluid sampled by oro-ruminal probe versus rumen fistula, Small Ruminant Res. 21 (1996) 63–69.
[34] J.D. Vaidya, et al., The effect of DNA extraction methods on observed microbial communities from fibrous and liquid rumen fractions of dairy cows, Front. Microbiol. 9 (2018) 92.
[35] A.B. De Menezes et al. Microbiome analysis of dairy cows fed pasture or total mixed ration diets 78, 256-265 (2011).
[36] K.M. Singh, et al., Microbial profiles of liquid and solid fraction associated biomaterial in buffalo rumen fed green and dry roughage diets by tagged 16S rRNA gene pyrosequencing, Mol. Biol. Rep. 42 (2015) 95–103.
[37] S. Ji et al. Comparison of rumen bacteria distribution in original rumen digesta, rumen liquid and solid fractions in lactating Holstein cows. 8, 16 (2017).
[38] J. Hultman, et al., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes, Nature 521 (2015) 208–212.
[39] K.M. Keiblinger, S. Fuchs, S. Zechmeister-Boltenstern, K. Riedel, Soil and leaf litter metaproteomics—a brief guideline from sampling to understanding, FEMS Microbiol. Ecol. 92 (2016).
[40] A. Tanca, et al., A straightforward and efficient analytical pipeline for metaproteome characterization, Microbiome 2 (2014) 49.
[41] R. Starke, N. Jehmlich, F. Bastida, Using proteins to study how microbes contribute to soil ecosystem services: the current state and future perspectives of soil metaproteomics, J. Proteomics 198 (2019) 50–58.
[42] R. Heyer, et al., Challenges and perspectives of metaproteomic data analysis, J. Biotechnol. 261 (2017) 24–36.
[43] P. Wilmes, P.L. Bond, Metaproteomics: studying functional gene expression in microbial ecosystems, Trends Microbiol. 14 (2006) 92–97.
[44] G. Martinez-Fernandez, S.E. Denman, C.S. McSweeney, Sample processing methods impacts on rumen microbiome, Front. Microbiol. 10 (2019) 861.
[45] B.J. Kunath, A. Bremges, A. Weimann, A.C. McHardy, P.B. Pope, Metagenomics and CAZyme Discovery, Methods Mol. Biol. 1588 (2017) 255–277.
[46] A. Zougman, P.J. Selby, R.E. Banks, Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis, Proteomics 14 (2014) 1006–1009.
[47] F. Amado, M.J. Calheiros-Lobo, R. Ferreira, R. Vitorino, Sample Treatment for Saliva Proteomics, Adv. Exp. Med. Biol. 1073 (2019) 23–56.
[48] J. Speda, M.A. Johansson, U. Carlsson, M. Karlsson, Assessment of sample preparation methods for metaproteomics of extracellular proteins, Anal. Biochem. 516 (2017) 23–36.
[49] S.A. Terry, et al., Effect of humic substances on rumen fermentation, nutrient digestibility, methane emissions, and rumen microbiota in beef heifers, J. Anim. Sci. 96 (2018) 3863–3877.
[50] F. Kohrs, et al., Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and thermophilic biogas plants, Anaerobe 29 (2014) 59–67.
[51] H.J. Issaq, T.P. Conrads, G.M. Janini, T.D. Veenstra, Methods for fractionation, separation and profiling of proteins and peptides, Electrophoresis 23 (2002) 3048–3061.
[52] J. Cox, et al., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, Mol. Cell. Proteomics: MCP 13 (2014) 2513–2526.
[53] J.M. Walker, The Bicinchoninic Acid (BCA) Assay for Protein Quantitation, in: J.M. Walker (Ed.), The Protein Protocols Handbook, Springer Protocols Handbooks, Humana Press, Totowa, NJ, 2009, , https://doi.org/10.1007/978-1-59745-198-7_3.
[54] B. Thiede, et al., High resolution quantitative proteomics of HeLa cells protein species using stable isotope labeling with amino acids in cell culture (SILAC), two-dimensional gel electrophoresis (2DE) and nano-liquid chromatograpohy coupled to an LTQ-OrbitrapMass spectrometer, Mol. Cell. Proteomics 12 (2013) 529–538.
[55] J. Cox, M. Mann, Quantitative, high-resolution proteomics for data-driven systems biology, Annu. Rev. Biochem. 80 (2011) 273–299.
[56] T.J. Snelling, R.J. Wallace, The rumen microbial metaproteome as revealed by SDS-PAGE, BMC Microbiol. 17 (2017).
[57] E. Hart, C. Creevey, T. Hitch, A. Kingston-Smith, Meta-proteomics of rumen microbiota indicates niche compartmentalisation and functional dominance in a limited number of metabolic pathways between abundant bacteria, Sci. Rep. 8 (2018) 1–11.
[58] M.C. Honan, S.L. Greenwood, Characterization of variations within the rumen metaproteome of Holstein dairy cattle relative to morning feed offering, Sci. Rep. 10 (2020) 1–8.
[59] R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function, Nature 537 (2016) 347–355.
[60] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2 (2007) 1896.
[61] J. Rappsilber, Y. Ishihama, M. Mann, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, Anal. Chem. 75 (2003) 663–670.
[62] D. Becher, J. Bernhardt, S. Fuchs, K. Riedel, Metaproteomics to unravel major microbial players in leaf litter and soil environments: C hallenges and perspectives, Proteomics 13 (2013) 2895–2909.
[63] M.R. Wilkins, et al., Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it, Biotechnol. Genet. Eng. Rev. 13 (1996) 19–50.
[64] T. Muth, D. Benndorf, U. Reichl, E. Rapp, L. Martens, Searching for a needle in a stack of needles: challenges in metaproteomics data analysis, Mol. BioSyst. 9 (2013) 578–585.
[65] A. Tanca, et al., The impact of sequence database choice on metaproteomic results

in gut microbiota studies, Microbiome 4 (2016) 51.
[66] A. Géron, J. Werner, R. Wattiez, P. Lebaron, S. Matallana Surget, Deciphering the functioning of microbial communities: shedding light on the critical steps in metaproteomics, Front. Microbiol. 10 (2019) 2395.
[67] L.M. Solden, et al., Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem, Nat. Microbiol. 3 (2018) 1274–1284.
[68] M. Hess, et al., Metagenomic discovery of biomass-degrading genes and genomes from cow rumen, Science 331 (2011) 463–467.
[69] F. Li, L.L. Guan, Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle, Appl. Environ. Microbiol. 83 (2017).
[70] E. Jami, I. Mizrahi, Composition and similarity of bovine rumen microbiota across individual animals, PLoS One 7 (2012) e33306.
[71] S.L. Amarasinghe, et al., Opportunities and challenges in long-read sequencing data analysis, Genome Biol. 21 (2020) 30.
[72] A. Sczyrba, et al., Critical assessment of metagenome interpretation—a benchmark of metagenomics software, Nat. Methods 14 (2017) 1063–1071.
[73] J. Cox, et al., Andromeda: a peptide search engine integrated into the MaxQuant environment, J. Proteome Res. 10 (2011) 1794–1805.
[74] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis 20 (1999) 3551–3567.
[75] S.K.R. Park, et al., ComPIL 2.0: an updated comprehensive metaproteomics database, J. Proteome Res. 18 (2018) 616–622.
[76] D. Beyter, M.S. Lin, Y. Yu, R. Pieper, V. Bafna, Proteostorm: an ultrafast metaproteomics database search framework, Cell Syst. 7 (2018) 463–467.
[77] J.K. Eng, B.C. Searle, K.R. Clauser, D.L. Tabb, A face in the crowd: recognizing peptides through database search, Mol. Cell. Proteomics 10 (2011) R111.009522.
[78] T. Välikangas, T. Suomi, L.L. Elo, A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation, Briefings Bioinf. 19 (2018) 1344–1355.
[79] J. Tang, et al., ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies, Briefings Bioinf. 21 (2020) 621–636.
[80] T. Muth, et al., The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation, J. Proteome Res. 14 (2015) 1557–1565.
[81] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372.
[82] X. Zhang, et al., MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota, Microbiome 4 (2016) 31.
[83] C.W. Easterly, et al., metaQuantome: an integrated, quantitative metaproteomics approach reveals connections between taxonomy and protein function in complex microbiomes, Mol. Cell. Proteomics 18 (2019) S82–S91.
[84] A.I. Nesvizhskii, R. Aebersold, Interpretation of shotgun proteomic data: the protein inference problem, Mol. Cell. Proteomics 4 (2005) 1419–1440.
[85] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[86] D.H. Huson, A.F. Auch, J. Qi, S.C. Schuster, MEGAN analysis of metagenomic data, Genome Res. 17 (2007) 377–386.
[87] R. Gurdeep Singh, et al., Unipept 4.0: functional analysis of metaproteome data, J. Proteome Res. 18 (2018) 606–615.
[88] M. Ashburner, et al., Gene ontology: tool for the unification of biology, Nat. Genet. 25 (2000) 25–29.
[89] S. Hunter, et al., InterPro: the integrative protein signature database, Nucl. Acids Res. 37 (2009) D211–D215.
[90] A. Bateman, et al., The Pfam protein families database, Nucl. Acids Res. 32 (2004) D138–D141.
[91] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucl. Acids Res. 28 (2000) 27–30.
[92] M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences, J. Mol. Biol. 428 (2016) 726–731.
[93] S. Tyanova, et al., The Perseus computational platform for comprehensive analysis of (prote)omics data, Nat. Methods 13 (2016) 731–740.
[94] R. Sajulga, et al., Survey of metaproteomics software tools for functional microbiome analysis, bioRxiv (2020), https://doi.org/10.1101/2020.01.07.897561.
[95] L. Martens, et al., PRIDE: the proteomics identifications database, Proteomics 5 (2005) 3537–3545.
[96] F. Desiere, et al., The peptideatlas project, Nucl. Acids Res. 34 (2006) D655–D658.
[97] R. Craig, J.P. Cortens, R.C. Beavis, Open source system for analyzing, validating, and storing protein identification data, J. Proteome Res. 3 (2004) 1234–1242.
[98] M. Wang, et al., Assembling the community-scale discoverable human proteome, Cell Syst. 7 (2018) 412–421.e415.
[99] E.W. Deutsch, et al., The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics, Nucl. Acids Res. 48 (2019) D1145–D1152.
[100] A. Palomba, et al., Multi-omic biogeography of the gastrointestinal microbiota of a pre-weaned lamb, Proteomes 5 (2017) 36.
[101] E.H. Hart, C.J. Creevey, T. Hitch, A.H. Kingston-Smith, Meta-proteomics of rumen microbiota indicates niche compartmentalisation and functional dominance in a limited number of metabolic pathways between abundant bacteria, Sci. Rep. 8 (2018) e10504.
[102] S. Deusch, et al., A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments, Front. Microbiol. 8 (2017) 1605.
[103] A.E. Naas, et al., "Candidatus Paraporphyromonas polyenzymogenes" encodes multi-modular cellulases linked to the type IX secretion system, Microbiome 6 (2018) 44.
[104] R. Roehe, et al., Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance, PLoS Genet. 12 (2016) e1005846.
[105] G.F. Difford, et al., Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows, PLoS Genet. 14 (2018) e1007580.
[106] B.J. Kunath, Interpreting the Irrecoverable Microbiota in Digestive Ecosystems, Doctoral thesis Norwegian University of Life Sciences, 2018.
[107] J. Kamke, et al., Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation, Microbiome 4 (2016) 56.
[108] W. Shi, et al., Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome, Genome Res. 24 (2014) 1517–1525.
[109] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[110] L. Heirendt, et al., Creation and analysis of biochemical constraint-based models using the COBRA Toolbox vol 3.0, Nat. Protoc. 14 (2019) 639–702.