Rumen Metagenomics

16

Robert W. Li

Abstract

The rumen microbiome plays a critical role in normal physiology and nutrition of ruminants. Alterations in the rumen microbiome have important physiological and pathological implications. The advent of next-generation sequencing technologies and rapid development of computational tools and reference databases provide powerful tools in rumen microbiome studies. Rumen metagenomics enables studies on the collective genetic structure and functional composition of the rumen microbial community in a culture-independent manner and can be simply divided into functional metagenomics and sequencing-based computational metagenomics. Recent progresses in mining the rumen microbial community for novel enzymes, such as fibrolytic enzymes, or other biomolecules for industry and biotechnology applications using functional screening are discussed. Rapid advances in computational metagenomic tools and methods are summarized. Metagenomics has provided novel insights into the structure and function of the rumen microbiome. Recent efforts suggest that the core rumen microbiome consists of 8 phyla and 15 families, which likely contribute to the basic function of the rumen. Systematic investigations of the rumen microbiome, including its viral (virome) and plasmid (plasmidome) fractions, have revealed previously unrecognized biodiversity in the rumen. Resistance and resilience of the rumen microbial community in response to perturbation is also discussed. Moreover, the need for mechanistic models and applications of general ecological theories and principles in rumen metagenomic studies is emphasized.

R.W. Li (✉)
United States Department of Agriculture, Agriculture Research Service, Animal Genomics and Improvement Laboratory, 10300 Baltimore Avenue, Beltsville, MD 20705, USA
e-mail: Robert.Li@ars.usda.gov

16.1 Introduction

The rumen microbiome represents the totality of rumen microorganisms, their genetic elements, and environmental interactions. The rumen microbiome plays an essential role in ruminant physiology and nutrition and pathology as well as host immunity. Rumen microorganisms convert plant fiber to short-chain fatty acids (SCFA), which contribute up to 75 % of the total metabolizable energy in ruminants. In addition to the fibrolytic capacity, rumen microorganisms also participate in ruminal nitrogen metabolism, including dietary protein degradation. However, nitrogen losses during protein degradation and methane produced during rumen fermentation are substantial contributors to water and air pollution as well as global warming. Rumen microorganisms produce a large amount of vitamins. As a result, ruminants generally do not need dietary supplement of water-soluble vitamins and vitamin K. Rumen microorganisms are able to modulate nutrient absorption and may be among the major determinants of nutrient utilization efficiency (Li et al. 2012a; Jami et al. 2014). Moreover, ruminal biohydrogenation, the saturation process of dietary unsaturated fatty acids controlled by rumen microorganisms, can be manipulated for healthier meat products (Jenkins et al. 2008). It is well known that rumen microbes play a key role in detoxifying plant secondary compounds (Wallace 2008). The involvement of the rumen microbiome in xenobiotic metabolism has been well documented (Li et al., 2014). Previous studies have identified rumen microbes responsible for the degradation of nitroaromatic explosive compounds, such as 2,4,6-trinitrotoluene (De Lorme and Craig 2009) and hexahydro-1,3,5-trinitro-1,3,5-triazine (Eaton et al. 2011).

The complexity of the rumen microbiome has long been appreciated, as evidenced by the presence of myriad microbial interactions (Li et al. 2012a). One of the major obstacles hindering our understanding of the structure and function of the rumen microbiome is that only approximately 11 % of rumen bacteria appear to be culturable (Edwards et al. 2004). DNA fingerprinting techniques widely used in earlier studies, such as terminal-restriction fragment length polymorphism (t-RFLP) and single-strand conformation polymorphism (SSCP), have limited throughput and low resolution and are therefore unable to provide a holistic view of the structure and function of the rumen microbiome. Furthermore, the rumen microbiome functions as a tightly integrated system in which all resident species interacts closely to contribute to its emergent properties. Predominant species perform all major microbial conversions in this ecosystem. Nevertheless, numerically minor species also play an important role in maximizing rumen ecosystem outputs. Disruption of one species could cause a chain reaction and result in undesired or unpredicted consequences. These properties call for a move from studies of individual rumen microorganisms in isolation or in pure culture to community-level studies, especially in their natural habitats.

Metagenomics has emerged in the past few years as a powerful tool for studying the rumen microbiome, thanks to the advent of next-generation sequencing (NGS) technologies and rapid progress in reference databases and bioinformatic tools. Metagenomics addresses the collective genetic structure and functional composition of a microbial community without the bias or necessity for culturing its individual inhabitants (Galbraith et al. 2004). Rumen metagenomics enables comprehensive studies of

the structure and function of the rumen microbiome using culture-independent approaches. Rumen metagenomics generally includes two major arenas: high-throughput screening of cloned expression libraries made from rumen metagenome DNA for gene products of interest (functional metagenomics) and sequencing-based characterization of the aggregate collection of genomes and genes present in rumen microbial communities, at both DNA (metagenomics) and RNA levels (metatranscriptomics). Functional screening technology was first applied to rumen materials to mine novel enzymes in 2005 (Ferrer et al. 2005), whereas the first publication using next-generation sequencing-based rumen computational metagenomics can be traced back to 2009 (Brulc et al. 2009). Since then, metagenomic technologies have been extensively utilized to investigate rumen microbial communities. The rumen of an individual animal is believed to harbor hundreds and up to 1,000 microbial species. Therefore, microarrays (such as PhyloChips), DNA fingerprinting techniques, or traditional Sanger sequencing-based methods that are unable to provide a sequencing depth of greater than 1,000 sequence reads will be excluded for discussion in this chapter. We will summarize recent advances in metagenomic technologies and novel metagenomic insights into the structure and function of the rumen microbiome.

16.2   Functional Metagenomics

Functional metagenomics is the study of the collective genome of a microbial community by expressing it in a foreign host (Ekkers et al. 2012). The vast majority of enzymes that catalyze biochemical reactions are encoded by genes present in microbial communities under various environmental conditions. For example, a recently developed database lists 510 commercially useful enzymes used in various sectors, including agriculture, energy, and biomedicine (Sharma et al. 2010). Therefore, functional screening has become an increasingly important

field for discovering novel biomolecules for applications in biotechnology and medicine. This approach relies on cloning of vast genetic diversity from a target habitat in various vectors (e.g., plasmids, cosmids, fosmids, or bacterial artificial chromosomes) and then expressing cloned metagenome libraries in foreign host systems (e.g., *E. coli*) followed by detection and characterization of desired functional activities in the expression libraries using various strategies (Simon and Daniel 2009). Functional screening provides direct access to largely unexploited microbial genetic diversity in the environment. Lignocellulose biomass, including cellulose, hemicellulose, pectin, and lignin, is the most abundant source of organic carbon on the planet. Efficient enzymatic conversion of biomass into biofuel has been of great interest recently. The complete degradation of lignocelluloses requires a concerted action of dozens of enzymes from various families, such as endo-β-1,4-glucanases, cellobiohydrolases, β-glucosidases, endoxylanases, β-xylosidases, α-l-arabinofuranosidases, acetyl xylan esterases, feruloyl esterases, and α-glucuronidases. Previous studies suggest that the rumen microbial ecosystem harbors a dazzling array of microbial diversity (Li et al. 2012a, b, c; Sparks et al. 2012) and is a rich source of efficient fibrolytic enzymes. A relatively small fraction of rumen microorganisms have been successfully cultured to date. The largely unexplored ruminal microbial diversity represents an untapped source of unique lignocellulose-digesting enzymes, especially those with multiple functions. Numerous efforts have been made to isolate fiber-digesting enzymes from the rumen, including various hydrolases from at least 8 glycosyl hydrolase families, such as GH3, GH5, GH8, GH9, GH10, GH13, GH26, GH43, GH48, and GH57. Morgavi et al. (2013) summarized the screening results prior to 2012. Results from functional screening of the rumen microbiome since 2012 are listed in Table 16.1.

Despite the fact that the huge potential of functional screening in mining genetic diversity for biotechnology applications has been demonstrated by the abovementioned case studies,

Table 16.1 Lignocellulose-digesting enzymes mined from the rumen using functional metagenomic approaches since 2012

Enzyme (family)	Rumen	Reference
Cellulases (GH5)	Buffalo	Nguyen et al. (2012)
Cellulases (GH5)	Buffalo	Cheema et al. (2012)
Cellulases (GH5, GH9, GH45, GH48)	Yak	Dai et al. (2012)
Endocellulase/xylanase (GH5)	Bovine	Rashamuse et al. (2013)
Endoglucanase cellulases	Cow	Gong et al. (2012)
Endohemicellulases (GH8, GH10, GH11, GH26, GH28, GH53)	Yak	Dai et al. (2012)
Esterase	Cow	Kim et al. (2012a)
Exocellulase (GH48)	Cow	Ko et al. (2013)
Feruloyl esterase	Calf	Ferrer et al. (2012)
Feruloyl esterase	Cow	Cheng et al. (2012a)
Glycosyl hydrolases (GH43)	Calf	Ferrer et al. (2012)
Glycosyl hydrolases (GH5)	Reindeer	Pope et al. (2012)
Xylanase (GH10)	Bovine	Cheng et al. (2012b)
Xylanase (GH10)	Bovine	Gong et al. (2012)
α-Glucuronidase (GH67)	Cow	Lee et al. (2012a)
β-Glucosidase (GH3)	Bovine	Gruninger et al. (2014)

numerous challenges remain. First, only a small fraction of functional diversity is captured in expression libraries, partially due to the difficulty in expressing target genes in a foreign host. Moreover, current methods to detect desired function or enzyme activities are less sensitive; and the throughput of screening methods is relatively low. Novel strategies, such as fractionation of the microbial community using habitat biasing methods to reduce the complexity of the microbiome or to enrich desired activities, have been developed to overcome these limitations (Ekkers et al. 2012). Furthermore, the potential power of novel technology using in vitro compartmentalization (IVC) in combination with fluorescent-activated cell sorting (FACS) in aiding functional screening of complex microbial ecosystems has been recognized (Ferrer et al. 2009). It is foreseeable that in combination with rapid advances in directed evolution techniques and methods (Dalby 2011), more enzymes and biomolecules with improved activities will be isolated using functional screening from the rumen microbiome for a wide range of applications.

16.3 Computational Metagenomics: Methods and Approaches

The advent of ultrahigh-throughput next-generation sequencing technologies and rapid development of computational tools and resources have stimulated computational metagenomic studies. As a result, computational metagenomics provides novel insights into the structure and function of microbial communities of host-associated habitats or from environmental samples at unprecedented resolution. The approach targeting small subunits (SSU) of rRNA genes (16S or 18S) allows us to interrogate the microbial composition and structure of the rumen microbiome. The whole-genome shotgun (WGS) approach provides unique opportunities to gain novel insights into the protein repertoire and metabolic potential of the microbiome, which lead to biological pathway reconstruction. Moreover, WGS approach enables taxonomical assignment to understand the microbial composition and structure of the rumen microbiota.

16.3.1 Ribosomal RNA Gene-Based Analysis

SSU ribosomal RNA genes, such as 16S rRNA for prokaryotes and 18S rRNA genes for eukaryotes, can be amplified from metagenomic DNA of various fractions of rumen materials. These two genes are most frequently used for phylogenetic analysis and microbial diversity studies in the rumen. Taxonomic informativeness of 9 well-defined hypervariable regions (V1 to V9) of the 16S rRNA gene varies tremendously (Chakravorty et al. 2007). As a result, effects of various primer combinations on classification accuracies have been designed and compared (Nossa et al. 2010; Soergel et al. 2012). The position of primers and amplicon length are major determinants of taxonomic precision. Most importantly, taxonomic informativeness of primers is habitat dependent. No primers are truly universal and work best in all environments (Soergel et al. 2012). For example, primer pairs 343 F and 798R, targeting on hypervariable regions V3 to V4, produce maximal classification accuracy under the current limitation of NGS platforms and may be the most suitable for human foregut microbiome studies (Nossa et al. 2010). Primer pairs targeting on V1 to V3, V3 to V5, and V6 to V9 generally result in overall similar and yet accurate classification with minor bias (Vilo and Dong 2012). Indeed, primers targeting V1 to V3 and V3 to V5 regions are commonly used in rumen microbiome studies (Table 16.2).

The amplicons from target regions of 16S (18S) rRNA genes are sequenced using next-generation sequencers, such as 454 FLX or Illumina sequencers. While barcoded pyrosequencing has been the mainstay in sequencing the 16S amplicons of the rumen samples (Jami et al. 2013; Li et al. 2012c; Wu et al. 2012b), Illumina-based sequencing technology is increasingly gaining attention. The newly launched Illumina MiSeq sequencer with version 3 reagent kits enables generation of up to 25 million sequences with a length up to 2×300 bp (pair end). The reagent cost for such a run is approximately $1400.

Raw sequence quality needs to be checked and then filtered and trimmed. Sequencing error and PCR single-base substitutions from the 454 platform can be removed using AmpliconNoise, a development of the PyroNoise algorithm that is capable of separately removing 454 sequencing errors and PCR single-base errors (Quince et al. 2011). The Perseus program can be used to remove chimeras. Processed 16S sequence reads are then analyzed using taxonomy-dependent and taxonomy-independent approaches. The taxonomy-dependent approach generally assigns 16S sequences to various levels of taxa based on sequence similarities to annotated sequences deposited in public databases. The commonly used SSU databases include EzTaxon-e (Kim et al. 2012b), Greengenes (DeSantis et al. 2006), RDP (Cole et al. 2009), and SILVA (Quast et al. 2013). The algorithm RDP Classifier (Wang et al. 2007) is among the most frequently used programs for taxonomic classification and has resulted in the publication of more than 400 articles since its launch. However, inherent limitations of this approach are obvious: (1) inability to assign novel sequences from previously undescribed species that have no matches in existing reference databases, (2) accuracy and robustness of taxonomic classification that is dependent on the coverage and quality of the database used, and (3) low resolution. This approach is often unable to assign input query sequences to species or strain levels. These limitations become more serious for rumen microbiome studies because the SSU sequences of rumen origin are particularly underrepresented in public databases. To overcome these limitations, the taxonomy-independent clustering approach has been developed. This approach uses various clustering algorithms to assign query sequences into operational taxonomic units (OTUs) based on a distance matrix at a specified threshold (Chen et al. 2013). Its independence from existing databases allows the analysis of novel sequences. More than 15 taxonomy-independent algorithms, such as CD-HIT-OTU (Fu et al. 2012; Li and Godzik 2006), CROP (Hao et al. 2011), ESPRIT (Sun et al. 2009) and ESPRIT-tree (Cai and Sun 2011),

Table 16.2 A summary of metagenomic studies in ruminants using next-generation sequencing technologies

Sequencing	Species	Reference
16S hypervariable regions targeted		
V1–V3 (and V4–V5; V6–V8)	Buffalo	Pitta et al. (2014)
V1–V3	Cattle (Angus heifers)	Petri et al. (2013a, b)
V1–V3	Cattle (cows)	Mao et al. (2013)
V1–V2	Cattle (cows)	de Menezes et al. (2011)
V6–V8	Cattle (cows)	Hess et al. (2011)
V1–V3	Cattle (Friesian cows)	Sandri et al. (2014)
V3–V5	Cattle (Holstein calves)	Li et al. (2012a)
V1–V3	Cattle (Holstein calves)	Malmuthuge et al. (2014)
V3–V5	Cattle (Holstein cows and Angus beef steers)	Wu et al. (2012a)
V2–V3	Cattle (Holstein cows)	Jami et al. (2013)
V2–V3	Cattle (Holstein cows)	Jami and Mizrahi (2012)
V2–V3	Cattle (Holstein cows)	Jami et al. (2014)
V3–V5	Cattle (Holstein cows)	Li et al. (2012b)
V1–V2	Cattle (Holstein cows)	Pinloche et al. (2013)
V3–V4	Cattle (Holstein cows)	Thoetkiattikul et al. (2013)
V6–V8	Cattle (Nellore steer)	de Oliveira et al. (2013)
V1–V3, V4 (archaeal V3, V6–V8), 18S, ITS1	*Cervus*, red deer, sheep, cattle	Kittelmann et al. (2013)
V1–V3	Goats	Lee et al. (2012b)
V1–V3	Goats	Huo et al. (2014)
V1–V3	Reindeer	Pope et al. (2012)
V1–V2	Sheep	Castro-Carrera et al. (2014)
V6–V8	Sheep (West African Dwarf)	Omoniyi et al. (2014)
Whole-genome shotgun (WGS)		
WGS	Buffalo	Singh et al. (2012a, b)
WGS	Camels	Bhatt et al. (2013)
WGS	Cattle (Angus steers)	Brulc et al. (2009)
WGS	Cattle (cows)	Ferrer et al. (2012)
WGS	Cattle (cows)	Hess et al. (2011)
WGS	Cattle (Holstein calves)	Li et al. (2012b)
WGS	Cattle (Jersey cows)	Wang et al. (2013)
WGS	Goats	Lim et al. (2013)
WGS	Reindeer	Pope et al. (2012)
WGS	Sheep	Ellison et al. (2014); Li et al. (2014)

MOTHUR (Schloss et al. 2009), UClust (Edgar 2010), and UPARSE (Edgar 2013), have been published to date. A novel algorithm, TBC, incorporating the basic concept of taxonomy into clustering has been published (Lee et al. 2012c). Recently, the relative performance and parameters of some of these algorithms have been compared (Chen et al. 2013; Wu et al. 2012a). For the rumen data, CD-HIT-OTU performs well in our hands (Wu et al. 2012a; Li et al. 2012d). This algorithm, which uses a greedy incremental clustering process to identify OTUs from 16S rRNA gene sequences, is able to assign millions of reads in a relatively short time. Most

importantly, the program avoids overestimation of OTUs, a common problem associated with many existing programs, and results in accurate estimation of microbial diversity. In addition to these algorithms, publicly accessible pipelines, such as MOTHUR and QIIME (www.qiime.org), are very popular in analyzing SSU sequences and have been widely used to analyze the rumen datasets (Castro-Carrera et al. 2014; Lee et al. 2012b; Omoniyi et al. 2014; Pitta et al. 2014; Pope et al. 2012). QIIME also wraps other applications, such as FastTree, PyNAST, RDP Classifier, and UClust. The microbial community structure between different samples can then be compared and visualized using UniFrac (Lozupone and Knight 2005) and Fast UniFrac (Hamady et al. 2010).

16.3.2 Whole-Genome Shotgun Approach

WGS sequencing provides an opportunity to analyze both microbial diversity and functionality encoded in the genomes of rumen microbial communities. Driven by its application potential in metagenomics, numerous tools have been developed to analyze WGS data (Fig. 16.1). NGS technologies, such as those from Roche 454 pyrosequencing, Illumina, Ion Torrent, and PacBio platforms, have significantly reduced the time and cost of metagenome sequencing, which are revolutionizing metagenomic studies. Unique features and advantages of various NGS platforms, including future DNA sequencing technologies, have been extensively reviewed (Zhang
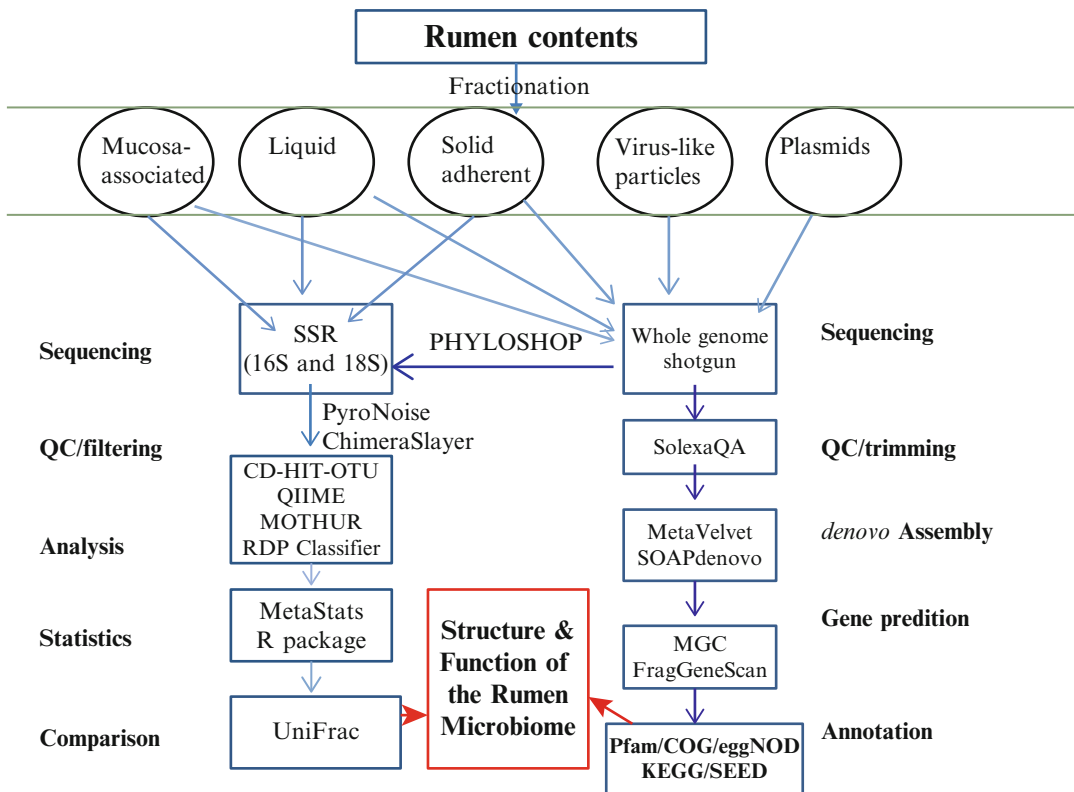

Fig. 16.1 A typical workflow and computational tools for rumen metagenomics

et al. 2012). NGS generally relies on either synthesis or hybridization at a massively parallel scale. For example, Illumina HiSeq 2500, in conjunction with improved version 4 chemistry, generates up to 1 terabase (1 trillion base) of sequence data in a single run (~167 Gb per day). The ultralow cost, enormous throughput, and extreme convenience of these NGS technologies have directly contributed to their instant acceptance and utility in the metagenomic community. Sequences generated by NGS technologies have some unique characteristics, such as short read lengths, platform-specific biases, and relatively high error rates, which could have a significant impact on downstream analyses. The computational challenges in handling short sequence reads have been extensively discussed (Pop 2009; Pop and Salzberg 2008). The challenges generally include difficulties in dealing with repetitive sequences as well as the need to modify existing algorithms to solve platform-specific errors and high error rates (Pop 2009). Furthermore, the production of billions of reads in a single sequencing run by such as Illumina HiSeq 2500 sequencers poses a tremendous challenge on computational resources.

Bioinformatic pipelines for NGS shotgun sequences generally include six steps: raw read quality control (QC) and trimming, assembly, functional annotation and metagenomic pathway reconstruction, taxonomy assignment, statistical analysis, and global network inference (Fig. 16.1). These processes have been extensively reviewed (Kim et al. 2013; Luo et al. 2013). The first step in dealing with WGS sequences involves QC, filtering, and trimming processes. Host sequence contamination can be removed using DeconSeq (Schmieder and Edwards 2011) or BLAST/Blat. Sequencing errors and PCR single-base substitutions as well as chimeras from the 454 platform can be removed using AmpliconNoise. Raw sequences generated by the Illumina platform can be trimmed using SolexaQA (Cox et al. 2010).

WGS sequences after these QC steps will generally need to be assembled into longer contigs for downstream applications. Assembly improves functional annotation. The basic framework of NGS assembly includes 4 stages: a preprocessing filtering, a graph construction process, a graph simplification process, and a post-processing filtering (El-Metwally et al. 2013). More than a dozen short-read assemblers have been developed to facilitate the analysis of short WGS sequences (Huang et al. 2012; Miller et al. 2010). The de Bruijn graph-based approach is among the most commonly used in short-read de novo assemblers, such as ABySS (Simpson et al. 2009), EULER-USR (Chaisson et al. 2009), SOAP*denovo* and its memory-efficient version, SOAP*denovo*2 (http://soap.genomics.org.cn/soapdenovo.html), and Velvet (Zerbino and Birney 2008). Recently, efforts have been made to understand the unique features and limitations of these short-read assemblers (Zhang et al. 2011; Huang et al. 2012; Mende et al. 2012; Vazquez-Castellanos et al. 2014). For 454 pyrosequencing data, Newbler is among the widely used assemblers and has been used in the analysis of rumen WGS sequences (Li et al. 2012b). Genovo, a de novo assembler specifically designed for 454-based metagenomic sequences using a generative probabilistic model (Laserson et al. 2011), and its extended version, Xgenovo (Afiahayati et al. 2013), perform well and are able to generate long contigs (Vazquez-Castellanos et al. 2014). Our results using simulated metagenomic datasets show that ABySS and SOAP*denovo* produce higher N50 and require relatively low memory usage, while Velvet and SOAP*denovo* produce higher genome coverage (Huang et al. 2012). Both Velvet and SOAP*denovo* result in a lower percentage of contig chimerism; while not specifically designed for metagenomic datasets, de Bruijn graph-based assemblers have been proven appropriate for large datasets with hundreds of millions of short reads (Zhang et al. 2011) and have been nevertheless extensively used in metagenomic studies. Recently, by making use of abundance differences and graph connectivity for the decomposition of the de Bruijn graph, an extended version of Velvet, MetaVelvet, has been developed to handle metagenomic data (Namiki et al. 2012). MetaVelvet is able to generate significantly higher N50 scores than other short-read assemblers, leading to an increased number

of predicted genes in our hands. Similarly, a de novo metagenomic assembler, Meta-IDBA (Peng et al. 2011), and its revised version, IDBA-UD (Peng et al. 2012), have been shown to generate longer contigs with high accuracy.

Genes or open reading frames (ORF) are then predicted from assembled contigs using a variety of gene prediction (gene-calling) algorithms. A dozen gene-calling programs have been developed for the metagenomic datasets, such as FragGeneScan (Rho et al. 2010), Glimmer-MG (Kelley et al. 2012), MetaGene (Noguchi et al. 2006) and MetaGeneAnnotator (Noguchi et al. 2008), MetaGeneTack (Tang et al. 2013), and Orphelia (Hoff et al. 2009). Gene prediction is an essential step for WGS metagenome data analysis for two reasons: (1) it is necessary for functional annotation and pathway reconstruction, and (2) gene prediction reduces the computational burden of protein similarity searches by nearly a factor of 6, compared to BLASTX (Trimble et al. 2012). In a direct comparison of 5 commonly used gene-calling algorithms, it is found that FragGeneScan performs better than MetaGeneAnnotator, MetaGeneMark, or Orphelia, especially for short reads (<1,000 bp) with sequence errors (Trimble et al. 2012). FragGeneScan combines sequencing error models and codon usages in a hidden Markov model to predict ORF in short reads (Rho et al. 2010) and has been used in the publically available MG-RAST pipeline (Wilke et al. 2013). Recently, a newly improved algorithm, MGC, has been published (El Allali and Rose 2013). This program relies on different models for different regions with various GC-contents and includes amino acid usage features to improve overall accuracy. It performs better in terms of sensitivity and specificity than both FragGeneScan and Orphelia in simulated metagenomic data (El Allali and Rose 2013).

To gain insights into functional potentials, predicted genes are further annotated against various public databases using a homology-based approach. For example, COG (Tatusov et al. 2000) and eggNOG (Powell et al. 2014) databases can be used to classify functional categories of predicted genes. Pfam (Punta et al. 2012), TIGRfam (Haft et al. 2003), and FIGfam (Meyer et al. 2009) databases can be used for protein family analysis. For rumen metagenomic data, the Carbohydrate-Active enZYmes database (CAZy), a database collecting and annotating the families of catalytic and carbohydrate-binding modules of enzymes that degrade, modify, or create glycosidic bonds, has been frequently used to mine fibrolytic enzymes in the rumen (Brulc et al. 2009; Hess et al. 2011). Pfam is a widely used database for protein family analysis and includes more than 14,800 annotated protein families in its latest release (v27.0). These families are also organized into groupings of related families (clans) based on similarity of sequences and structures. Furthermore, Gene Ontology (GO) can be extracted from these protein families using the Pfam2GO program (Hayete and Bienkowska 2005). The metagenomic features, such as COG functional profiles and metabolic subsystem data, between samples from two treatment groups can be analyzed using statistical packages, such as MetaStats (White et al. 2009). Metabolic pathways can then be reconstructed using databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa 2002) and BRENDA, the enzyme database (Schomburg et al. 2004). Metabolic pathways that differ between samples from two treatment groups can then be analyzed for statistical significance using MetaPath (Liu and Pop 2011) and Metagenomic Annotation Networks (Vey and Moreno-Hagelsieb 2012). Furthermore, network analysis tools and algorithms can be used to infer global co-occurrence patterns for the microbiome-wide microbial interactions (Faust and Raes 2012; Faust et al. 2012). The tools available for co-occurrence and association network analysis in metagenomic databases include CoNet (http://psbweb05.psb.ugent.be/conet/download.php), extended Local Similarity Analysis (eLSA) (Xia et al. 2011), QIIME, and a metagenome-wide association study (Qin et al. 2012).

Recently, using statistical distribution methods that ignore known biological processes to handle metagenomic data has been questioned (Liberles et al. 2013). These authors suggest that mechanism models based on ecological relationships, such as predator–prey dynamics and competitive relationships that widely exist in the

rumen microbial community, should be incorporated in metagenomic data analysis for ecological inference.

In addition to the individual bioinformatic tools and resources described above, several publicly available Web-based pipelines have been developed for metagenomic data analysis. These online platforms, such as CAMERA (http://camera.calit2.net/), IMG/M (http://img.jgi.doe.gov/), METAREP (http://jcvi.org/metarep/), and MG-RAST (http://metagenomics.anl.gov/), have integrated various tools for gene prediction, functional or protein family assignment, and protein interaction and metabolic pathway inference in a user-friendly format. For example, the latest version of the MG-RAST server integrates data uploading, QC, and annotation and analysis of various datasets, such as 16S or amplicon sequences and WGS metagenome and metatranscriptome sequences (Wilke et al. 2013). MG-RAST relies on FragGeneScan as a gene-calling program and BLAT for homology-based similarity search. Since its launch in 2007, more than 108,500 datasets have been analyzed.

WGS metagenomic sequence data not only enable analysis of functionality and metabolic potential of the microbial community but also provide a means for taxonomical assignment (binning). Numerous tools, such as metaBEETL (Ander et al. 2013), have been developed for taxonomic classification of WGS data from microbial communities. These tools are generally divided into 3 major categories: homology or similarity-based, composition-based, and hybrid approach. The lowest common ancestor (LCA) algorithm has formed a base for many of the similarity-based classification methods, such as in WebCARMA (Gerlach et al. 2009), CloudLCA (Zhao et al. 2012), MEGAN (Huson et al. 2007), and DiScRIBinATE (Ghosh et al. 2010). It is shown that the latter significantly reduces binning time with superior assignment accuracy. Composition-based methods, which exploit the uniqueness of DNA base composition in genomes of different taxonomic entities, have been implemented in programs such as PhyloPythia (McHardy et al. 2007), Phymm (Brady and Salzberg 2009), TACOA (Diaz et al. 2009), and

TaxSOM (Weber et al. 2011). The hybrid approach for binning generally combines both similarity- and composition-based methods, such as PhymmBL (Brady and Salzberg 2009) and RITA (MacDonald et al. 2012), for better accuracy. However, these two methods tend to be computationally time-consuming. To overcome this problem, a new algorithm, MetaPhlAn, has been developed (Segata et al. 2012). MetaPhlAn estimates the relative abundance of microbial cells by mapping short reads against a set of 400,141 clade-specific marker genes and allows for more accurate taxonomical assignment down to a species level in minutes of computational times for millions of WGS reads (Segata et al. 2012).

16.3.3 Stable Isotope Labeling

Stable or "heavy" isotopes, such as ^{13}C and ^{15}N, can be used to label various substrates, either small molecules (e.g., glucose) and polysaccharides (such as inulin) or even whole plants. Microorganisms that utilize these labeled substrates will most likely incorporate the heavy or stable isotope more efficiently into their biomolecules, including DNA or RNA. The labeled DNA or RNA can be readily separated from unlabeled, normal "light" DNA or RNA by isopycnic density-gradient ultracentrifugation, in combination with magnetic-bead capture techniques and isotope ratio mass spectrometry. The enriched SIP-RNA/DNA can be then studied using standard molecular technologies. For example, SIP has been widely used to investigate community function in microbial ecosystems or genes responsible for bioremediation (Uhlik et al. 2013). The potential of SIP technology in metagenomic studies is immediately recognized. The combined approaches not only permit the detection of low-abundance species in a complex microbial community but also facilitate the discovery of novel enzymes and bioactive compounds (Chen and Murrell 2010). Moreover, SIP-metagenomic technologies enable the enrichment of metabolically active fractions of microorganisms from environmental samples or

the gut microbiome, which can provide a powerful link between microbial phylogeny and metabolic functionality and activity. Such a link is important in understanding the role of rumen microbiota in normal physiology and pathogenesis. SIP-RNA technology has demonstrated that changes in the functional activity of the human gut microbiota are associated with nutrient sources and medium types (Reichardt et al. 2011). In ruminants, compounds labeled by SIP have been widely used in nutrition studies. For example, ^{13}C-labeled n-alkanes of plant origin are used as an internal tracer to assess digesta passage kinetics through the gastrointestinal tract (Warner et al. 2013). In addition, a stable isotope tracer, ^{13}C-linolenic acid, has been used to investigate the biohydrogenation process of linolenic acid in a bovine rumen microbial community (Lee and Jenkins 2011). Using a small-scale repeated batch culture model of cattle fecal microbial communities, ^{13}C-labeled fructose in combination with a modified t-RFLP molecular fingerprinting identifies *Streptococcus bovis* as the most dominant and *Lactobacillus vitulinus* and *Megasphaera elsdenii* as minor fructose fermenter, while several species of *Clostridium* cluster IV are non-fermenters of fructose (Michinaka and Fujii 2012). It is conceivable that the importance of SIP-RNA/DNA technology in rumen metagenomic studies will soon be recognized.

16.3.4 Gnotobiotic Rumen Models

Gnotobiotic, including germ-free, animals, which have well-defined microbial composition, provide an elegant model system to study myriad interactions between individual microbes and between microbes and the host. The microbial communities of varying complexity and origin can be then sequentially introduced to gnotobiotic animals to examine the effects of genetic background, dietary conditions, and physiological stages on the microbial community structure and dynamics. Synthetic gut microbiota with known microbial composition and abundance can be created in germ-free animals. When the

complete genome and transcriptome of these introduced microbial species become known, these systems can then be used to measure perturbation dynamics of the entire microbial community and to refine tools and algorithms using computational metagenomics.

Gnotobiotic lambs have been used to study rumen microbial establishment sequences and interactions of microbes of different functional groups for more than a decade (Fonty et al. 1989). The rumen of these animals harbors a defined microbial community. Rumen microbial species with known function can be sequentially added to the defined community. Therefore, gnotobiotic lambs are an ideal model to study the role of specific microorganisms and their interactions with other species in the rumen, such as the relationship between H_2-producing and H_2-consuming communities. Early results show that lambs lacking methanogens can be raised to adulthood, although their feed intake is lower compared to conventional lambs with functional methanogens (Fonty et al. 2007). A concomitant reduction in SCFA production as well as overall microbial complexity in these lambs is also observed. Moreover, acetogens can colonize and become rapidly established in the rumen of methanogen-free lambs, suggesting their establishment is independent of other microbes, unlike cellulolytic bacteria that generally require the presence of a diverse microbiota for establishment. On the other hand, methanogen colonization in the rumen does not substantially affect acetogen diversity (Gagen et al. 2012). Recently, interactions between fibrolytic species and methanogens have been examined using the gnotobiotic model (Chaucheyras-Durand et al. 2010). Methane emission is reduced when the dominant fibrolytic species in the rumen is a non-H_2 producer, such as *Fibrobacter succinogenes*, compared to the rumen with H_2-producing fibrolytic species, such as ruminococci and anaerobic fungi. These results suggest that dietary intervention strategies to promote non-H_2-producing fibrolytic species may represent a novel approach to mitigate methane production in farm animals. Utilization of metagenomic tools in the gnotobiotic

rumen model will facilitate our understanding of microbial establishment sequence and succession of the rumen microbiome.

16.3.5 Metatranscriptomics, Metaproteomics, and Metabolomics

Sequencing-based metagenomics addresses the collective genetic structure and functional composition at the DNA level of a microbial community in a culture-independent manner. While vitally important, metagenome characterization using the DNA-based approach does not itself reveal how the genetic information of a given microbiota is actually expressed. To characterize how genes in the metagenome are expressed and regulated, metatranscriptomics, a comprehensive measure of mRNA transcript abundance, dynamics, and regulation under a variety of environmental conditions or developmental and physiological/pathological stages, is developed (Lim et al. 2013; Qi et al. 2011). While metatranscriptomic analysis provides insights into how the metagenome is expressed and regulated, metaproteomics allows comprehensive characterization of the gene products (proteins) encoded in the metagenome and their posttranslational modifications and turnover. Metaproteomics has recently been applied to analyze the human salivary supernatant (Jagtap et al. 2012). However, due to limitations in accurate detection and mass measurement of peptides and their annotation, metaproteomics currently allow characterization of only a relatively small fraction of the gene products in a complex gut microbiota (Wilmes and Bond 2006). Similarly, a comprehensive survey of metabolites in the host, diet (forage), and its rumen microbiome, metabolomics, has been conducted to provide information on key players responsible for the microbiome function (Lee et al. 2012b; Kingston-Smith et al. 2013). Metabolomic profiling using nuclear magnetic resonance spectroscopy, in combination with 454 pyrosequencing, demonstrates uniqueness of the microbial composition and metabolites in the rumen of Korean native goats (Lee et al. 2012b).

Metabolic data facilitate the study of interactions between bacteria-specific metabolites and host proteins (Jacobsen et al. 2013). Together, rapid integrations of these OMIC technologies, including metagenomics for metagenomic DNA, metatranscriptomics for RNA, metaproteomics for proteins and peptides, and meta-metabolomics for metabolites, will provide a holistic insight into the structure and function of the rumen microbiota.

16.4 Metagenomic Insights into the Structure and Function of the Rumen Microbiome

16.4.1 Microbial Establishment and Succession During Rumen Development

Microbial products, such as SCFAs, play a critical role in rumen development. As a result, numerous attempts have been made to understand microbial establishment sequences and ecological succession of the developing rumen microbiome. It is generally accepted that the rumen is sterile at birth. Earlier studies demonstrate that bacteria start colonization in the rumen within the first 24 h of life and strictly anaerobic bacteria become predominant by the second day after birth (Fonty et al. 1989). Major functional groups of microorganisms, including fibrolytic bacteria and methanogens, become established in the rumen within the first week of life, followed by protozoa (Morvan et al. 1994; Quigley et al. 1985). Our knowledge of the microbial establishment and succession in the developing rumen and during the transition to mature rumen has been significantly expanded (Li et al. 2012b; Jami et al. 2013; Malmuthuge et al. 2014), largely due to the advent of metagenomic technologies. For example, it is generally accepted that methanogens start colonizing the rumen 3–4 days after birth (Fonty et al. 1989). A recent study shows that methanogens can be detected in the ovine rumen 17 h after birth (Gagen et al. 2012). A systematic cataloging of microbial diversity and functionality in the developing rumen using both

16S rRNA gene-based and WGS approaches has been attempted (Li et al. 2012a, b, c). Sequences from more than 24 prokaryotic phyla and 22 eukaryotic phyla are identified in rumen microbial communities of preruminant (14-day-old and 42-day-old) calves fed the same milk diet. Furthermore, the rumen microbiome of preruminant calves harbors considerable functional diversity, as evidenced by the existence of 8,298 Pfam protein families. A total of 156 and 120 genera are identified in the rumen microbiota of 14-day-old and 42-day-old calves, respectively. Fibrolytic bacteria and glycoside hydrolase protein families are abundant in the developing rumen, before the calves are fed a solid diet. Moreover, the fibrolytic capability of the developing rumen increases as calves' age. Interestingly, rumen development has a significant impact on microbial diversity. Genus-level richness indices ACE and Chao1 are significantly higher in the rumen of 14-day-old calves than that of 42-day-old calves. The rumen microbiome of younger calves displays a more heterogeneous microbial composition and harbors a greater number of bacterial genera (many of them may be transient) than the older calves fed the same milk replacer diet (Li et al. 2012b). This observation is in agreement with a general ecological theory that biodiversity tends to increase during early succession as new species arrive but may decline in later succession as competition eliminates opportunistic species. Rumen microbial composition changes from birth to adulthood have been monitored using pyrosequencing (Jami et al. 2013). This study shows that the fibrolytic species, such as *Ruminococcus albus*, is detectable in the developing rumen as early as one day of age while another major fibrolytic species, *Fibrobacter succinogenes*, appears much later. The observation that the presence of fibrolytic capacity in the developing rumen prior to exposure of solid diet is in agreement with previous studies (Li et al. 2012b) and in human infants (Koenig et al. 2011). Developmental stages appear to be one of the major determinants of the rumen microbial establishment as evidenced by the significant differences observed in rumen microbial composition between 14-day-old and 42-day-old calves (Li et al. 2012b) and between 6-month-old and 2-year-old cattle (Jami et al. 2013) that are fed the same diet. The direct comparison of microbial establishment sequences in the bovine rumen (Jami et al. 2013) and the hindgut of the human infants (Koenig et al. 2011) identifies similar wavelike patterns, coincident with critical life events such as diet, development, and health status, suggesting that similar forces may drive the establishment of microbial communities in two different habitats (Jami et al. 2013).

16.4.2 Rumen Microbial Diversity and the Core Rumen Microbiome

The collective microbial diversity in the rumen has been illustrated by a meta-analysis of all curated 16S rRNA gene sequences of rumen origin (13,478 bacterial and 3,516 archaeal sequences) deposited into the RDP database (Kim et al. 2011). This analysis has identified a total of 19 phyla and 5,271 and 942 OTUs for rumen bacteria and archaea, respectively. Approximately 1,000 OTUs are likely present in the rumen of fistulated cows and 587 OTUs are detected in all 4 samples from these two cows (Hess et al. 2011). In our studies, a total of 21 phyla are collectively identified from the rumen microbiome of dairy cows (Li et al. 2012c), with the mean of 16 phyla for the mature rumen of dairy cattle (Wu et al. 2012a). The mean number of genera in the rumen of individual cattle is 79.9 ± 14.0 ($\pm$ sd). In the mature rumen of dairy and beef cattle, the mean numbers of OTUs identified are 512 and 343, respectively. Together, these results suggest that the number of microbial species in a typical rumen of individual animals will likely be in the range of several hundreds.

The rumen microbiome is highly responsive to diet (Ellison et al. 2014), developmental stage (Li et al. 2012b; Jami et al. 2013), genetics, and numerous environmental factors. Substantial variations exist in microbial compositions and functionality among individual rumen samples within a species and between ruminant species (Jami and Mizrahi 2012). It is probable that a set

of core taxa or OTUs (species) are shared by the rumen microbiome of individual animals in all ruminant species within the large context of variation. The core rumen microbiome, consisting of a common set of microbial taxa that are shared by all individual samples, may contribute to basic rumen function. Defining the core rumen microbiome is of significance in understanding basic structure and function of the rumen microbiome and has been recently attempted (Li et al. 2012b, c; Wu et al. 2012a; Petri et al. 2013b). The core rumen microbiome of the bovine rumen microbiome, both the developing and mature rumen of dairy and beef cattle, consists of 8 phyla, 11 classes, and 15 families (Wu et al. 2012a). These 8 phyla, accounting for 99.5 % of input 16S sequences, are *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Spirochaetes*, *Fibrobacteres*, *Verrucomicrobia*, *Synergistetes*, and *Actinobacteria*, in descending order of relative abundance. The core bovine rumen microbiome likely represents minimal components of the rumen microbial community. As Table 16.2 shows, only a small number of approximately 150 ruminant species have been systematically studied for rumen microbial diversity using metagenomic tools. The rumen microbiome composition between ruminant species has been

investigated (Kittelmann et al. 2013). Recently, we have compared microbial community compositions of the bovine ($N=8$), caprine ($N=10$), and ovine ($N=10$) rumen in order to define the core rumen microbiome using deep 16S sequencing. The mean number of 16S rRNA gene sequences per sample is $79,213.0 \pm 11,682.2$ (mean $\pm$ SD; $N=28$; the mean read length $=300$ bp), a sequencing depth estimated to reach 99.9 % coverage (Kim et al. 2011). Our preliminary results show that collectively, 22 phyla and 94 families are detected in the rumen microbiome of cattle (cows), goats, and sheep. The mean number of the phyla per animal in the rumen microbiome of cattle, goats, and sheep is 16.8, 11.8, and 16.9, respectively. The caprine rumen has a significantly smaller number of phyla than the bovine and ovine rumen ($P < 10^{-5}$) in this study. The family-level composition follows the similar trend. The core rumen microbiome consists of 8 phyla, *Actinobacteria*, *Bacteroidetes*, Euryarchaeota, *Firmicutes*, *Proteobacteria*, *Spirochaetes*, *Synergistetes*, and *Verrucomicrobia*, which is consistent with our previous study of the core bovine microbiome (Wu et al. 2012a).

The 15 families consisting of the core rumen microbiome (Table 16.3), representing >95 % of assigned 16S sequences in each sample, will

Table 16.3 The relative abundance of the 15 families consisting of the core rumen microbiome

Family	Cattle ($N=8$)	Goats ($N=10$)	Sheep ($N=10$)
Prevotellaceae	51.12 (±5.99)	24.36 (±16.68)	35.16 (±11.56)
Lachnospiraceae	22.01 (±3.56)	24.35 (±13.25)	23.97 (±11.87)
Ruminococcaceae	8.18 (±3.18)	4.43 (±4.83)	19.74 (±9.60)
Veillonellaceae	4.33 (±3.90)	13.51 (±14.96)	10.09 (±6.78)
Acidaminococcaceae	4.21 (±1.26)	0.73 (±0.64)	2.39 (±1.37)
Porphyromonadaceae	1.76 (±1.28)	0.90 (±1.33)	1.21 (±0.99)
Spirochaetaceae	1.40 (±0.44)	0.45 (±0.65)	0.73 (±0.65)
Succinivibrionaceae	1.26 (±0.68)	29.84 (±20.46)	0.36 (±0.30)
Erysipelotrichaceae	0.88 (±0.32)	0.24 (±0.32)	1.73 (±1.05)
Synergistaceae	0.07 (±0.04)	0.03 (±0.02)	0.09 (±0.06)
Coriobacteriaceae	0.06 (±0.02)	0.05 (±0.06)	0.24 (±0.15)
Desulfobulbaceae	0.06 (±0.04)	0.00 (±0.00)	0.21 (±0.12)
Moraxellaceae	0.03 (±0.03)	0.02 (±0.01)	0.02 (±0.06)
Clostridiales incertae sedis XIII	0.03 (±0.02)	0.05 (±0.04)	0.16 (±0.11)
Campylobacteraceae	0.01 (±0.01)	0.05 (±0.04)	0.12 (±0.08)

The number represents the percentage of the abundance (± SD). The abundance = the number of 16S sequences assigned to this particular family/the total number of 16S sequences assigned to all families at 80 % confidence cutoff using RDP Classifier

likely contribute to the basic function of the rumen microbial ecosystem. However, the relative abundance of 15 families consisting of the core rumen microbiome varies significantly among the 3 host species, despite their high prevalence. For example, the abundance of the family *Prevotellaceae*, the most abundant family in the rumen of all 3 species, in the caprine rumen (24.26 %) is significantly lower than in the bovine (51.12 %) and ovine (35.16 %) rumen, while the abundance of 4 families, *Acidaminococcaceae*, *Desulfobulbaceae*, *Campylobacteraceae*, and *Succinivibrionaceae*, is significantly different between species. The abundance of *Ruminococcaceae* is significantly higher in the ovine rumen (19.74 %) than in the bovine (8.18 %) and the caprine (4.43 %) rumen. On the other hand, the abundance of *Lachnospiraceae* is relatively stable in the rumen of the 3 species (between 22.00 and 24.35 %).

16.4.3 Rumen Virome and Plasmidome

Previous studies have focused on bacterial and archaeal diversity in the rumen. Lack of conserved proteins and genes, such as 16S rRNA genes, among viruses or phages has hindered their discovery. The advent of computational metagenomics has made possible systematically cataloging viruses in the rumen and surveying the phage diversity. A recent study shows that more than 20,000 viral genotypes exist in the rumen of 2 of the 3 cattle analyzed (Berg Miller et al. 2012). While over 70 % of viral sequences have no significant matches to sequences in public databases, sequences associated with prophages outnumber those lytic phages approximately 2:1. Moreover, rumen viral sequences carry functional genes; and the majority of these genes belong to phages, prophages, transportable elements, and plasmid subsystem according to the SEED database, as expected. In another study, 14 putative viral sequences over 30 kb are identified (Ross et al. 2013). Cows housed together and fed the same diet display similar taxonomical virome profiles than those housed separately. Intriguingly,

these two cohorts have similar functional characterizations, suggesting the rumen virome appears to be functionally conserved (Ross et al., 2013). Together, these results provide further evidence that viruses play an important role in horizontal gene transfer between various microorganisms, spreading antibiotic resistance genes, controlling bacterial population dynamics, and affecting animal nutrition and protein metabolism in the rumen.

Due to a relatively low copy number per bacterial cell and difficulty in distinguishing them from chromosomal DNA, rumen plasmids have not been systematically studied until recently (Brown Kav et al. 2012; Mizrahi 2012). Brown Kav et al. (2013) recently reported a method to enrich plasmid DNA from the rumen. The method takes advantage of an exonuclease that is able to digest chromosomal DNA that is sheared and linearized during extraction procedures. The resultant circular plasmid DNA is amplified using a phi29 DNA polymerase and further sequenced using an Illumina sequencer. This study provides a first glimpse of the diversity and function of the rumen plasmidome (i.e., the collective plasmid population of rumen origin). Notably, while the rumen microbial hosts can be assigned to the three major phyla, *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*, using rumen plasmid sequences, a significant difference in the relative abundance is evident compared to the phylogenetic assignment based on the 16S sequences from the very same rumen source. For example, *Proteobacteria* account for approximately 20 % of rumen plasmid sequences, whereas its abundance appears to be significantly lower (~5 %) according to the 16S approach (Brown Kav et al. 2012). Functional analysis using the SEED database suggests that, in addition to the intrinsic plasmid-coding functions, the rumen plasmidome shares similarities with the rumen metagenomes and displays a significantly higher representation of the functional categories, such as "amino acids," "cell wall and capsule," "cofactors, vitamins, etc.," and "protein metabolism." These results demonstrate that rumen plasmids may play an important role in lateral gene transfer between rumen microorganisms.

16.4.4 Resistance and Resilience of the Rumen Microbiome

One of the primary functions of the rumen is to degrade lignocellulosic fiber to produce short-chain fatty acids for energy. The relative stability of structure and composition of the rumen microbiome becomes a prerequisite for such functions. The stability can be defined as (1) the ability to return to an equilibrium state following perturbation and (2) the ability to resist changes (resistance) or the rate of return to an equilibrium following perturbation or overall system variability (Robinson et al. 2010). Therefore, the stability of the rumen microbiome imparts resilience to perturbation, ensuring continued rumen function.

The rumen microbiome is susceptible to both natural and anthropogenic stresses and is highly responsive to changes in environmental conditions and host factors, such as critical physiological or pathological events, resulting in the creation of novel niche for other microbial species. The microbial diversity of the rumen is a reflection of the coevolution between microbial communities and their host and represents equilibrium between functional redundancy of a stable community and niche specialization. Although a few dominant microorganisms are likely to be responsible for the majority of the metabolic activity and energy influx in the rumen, it is well known that uncommon species often serve as a reservoir of genetic and functional diversity, playing key roles in microbial ecosystems. These species can become numerically important if environmental conditions change. Recently, scientific communities have begun to study the extent of temporal and spatial shifts in functionality and phylogenic composition of the rumen microbiome in response to various stresses, such as diet, critical life events (such as weaning and acidosis), and antibiotic usage, as well as their ecological and physiological implications (Li et al. 2012b, c; Jami et al. 2013; Petri et al. 2013a, b).

We have characterized temporal changes of the rumen microbiome of dairy cows in response to an exogenous butyrate disturbance (Li et al. 2012c). We reanalyzed the raw data using improved bioinformatic tools for this chapter. Our results indicate that in the rumen microbiome of dairy cows in their mid-lactation, the five most abundant phyla, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Fibrobacteres*, and *Spirochaetes* in this order, account for >99 % of observed 16S sequences. A 168-h exogenous butyrate perturbation results in significant changes in abundance of 4 of the 5 most abundant phyla. The relative abundance of *Bacteroidetes* and *Fibrobacteres* is significantly decreased, from 68.20 % to 56.74 % (at the basal level to after perturbation) and 1.43 % to 0.82 %, respectively. On the other hand, the abundance of *Firmicutes* and *Spirochaetes* is significantly increased, from 23.34 % to 30.32 % and 0.98 % to 2.17 %, respectively. The phylum *Firmicutes* includes a majority of butyrate-producing bacteria. The observation of exogenous butyrate increasing the relative abundance of *Firmicutes* suggests that butyrate itself may be butyrogenic. Indeed, a readily available energy source (exogenous butyrate) reduces the need of fibrolytic capacity of the rumen, resulting in a decreased abundance of *Fibrobacteres*. The abundance of these 5 phyla returns to a pre-disturbance level 168 h after perturbation, suggesting the resilient nature of the rumen microbiome. The rumen microbiome is also recovered 1 week after diet-induced acidosis challenge (Petri et al. 2013a). The analysis at family level demonstrates the same trend. The perturbation with exogenous butyrate significantly impacts 7 of the 20 most abundant families (accounting for 99.8 % of the sequences), including the 3 most abundant families, *Prevotellaceae*, *Lachnospiraceae*, and *Ruminococcaceae*. After the perturbation withdrawal, the abundance of these families returns to their pre-disturbed levels. Among the 20 top families, the long-lasting impact of the disturbance is observed for only 2 minor families: *Anaeroplasmataceae* remains elevated, while *Acetobacteraceae* is further repressed, 168 h after perturbation withdrawal. Our data demonstrate that the rumen microbial ecosystem displays substantial resilience to short-term disturbances. Furthermore, considerable hysteresis of the rumen microbiome is

observed. The ecological role and consequences of the two families, *Anaeroplasmataceae* and *Acetobacteraceae*, in the new established rumen microbial community are worthy of further scientific attention.

16.5 Conclusions

Metagenomics has significantly expanded our knowledge of the rumen microbial diversity and the structure and function of the rumen microbiome, thanks to rapid advances in next-generation sequencing technologies and computational tools and resources. The rumen microbiome, consisting of hundreds of microbial species and myriad microbial interactions, plays a critical role in nutrition and normal physiology of ruminants. It is highly responsive to changes in diet, development, environmental factors, and host genetics. Alterations in the rumen microbiome have important implications for animal well-being and production efficiency. While metagenomics has proven to be a powerful tool in rumen microbiome studies, numerous challenges remain. Temporal and spatial fluctuations as well as intra- and interindividual variations of ruminal microbial composition have yet to be assessed. Current development of bioinformatic tools and resources that cope with fragmental and voluminous next-generation sequencing data is still in its infancy. Notably, the lack of fully assembled and annotated reference genomes of rumen origin in public databases has hindered functional annotation of metagenomic data. The lack of commonly accepted data analysis pipelines and standardized report formats makes direct comparisons of various metagenomic studies difficult, if not impossible. Most importantly, general theories and principles of microbial ecology have yet to be fully applied to rumen metagenomic studies. Mechanistic models are still needed to aid the interpretation of biological relevance of metagenomic data. The recent launch of the Hungate1000 community sequencing project, which aims to sequence up to 1,000 microbial genomes of rumen origin (http://www.hungate1000.org.nz/), representing a broad spectrum of rumen micro-

bial taxa, will undoubtedly facilitate the assembly and annotation of WGS metagenomic data. It is conceivable that comprehensive studies of the rumen microbiome using metagenomic tools will broaden our understanding of the structure and function of the rumen microbiome and its role in normal physiology and pathology of ruminants, which in turn should guide our efforts to develop optimal rumen manipulation strategies for more efficient fiber digestion as well as mitigation of environmental footprints of animal farming.

References

Afiahayati, Sato K, Sakakibara Y (2013) An extended genovo metagenomic assembler by incorporating paired-end information. Peer J 1:e196

Ander C, Schulz-Trieglaff OB, Stoye J et al (2013) MetaBEETL: high-throughput analysis of heterogeneous microbial populations from shotgun DNA sequences. BMC Bioinfo 14(Suppl 5) :S2

Berg Miller ME, Yeoman CJ, Chia N et al (2012) Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. Environ Microbiol 14:207–227

Bhatt VD, Dande SS, Patil NV et al (2013) Molecular analysis of the bacterial microbiome in the forestomach fluid from the dromedary camel (*Camelus dromedarius*). Mol Biol Rep 40:3363–3371

Brady A, Salzberg SL (2009) Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nat Methods 6:673–676

Brown Kav A, Sasson G, Jami E et al (2012) Insights into the bovine rumen plasmidome. Proc Natl Acad Sci U S A 109:5452–5457

Brown Kav A, Benhar I, Mizrahi I (2013) A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. J Microbiol Methods 95:272–279

Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC et al (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proc Natl Acad Sci U S A 106:1948–1953

Cai Y, Sun Y (2011) ESPRIT-tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. Nucleic Acids Res 39:e95

Castro-Carrera T, Toral PG, Frutos P et al (2014) Rumen bacterial community evaluated by 454 pyrosequencing and terminal restriction fragment length polymorphism analyses in dairy sheep fed marine algae. J Dairy Sci 97:1661–1669

Chaisson MJ, Brinza D, Pevzner PA (2009) De novo fragment assembly with short mate-paired reads: does the read length matter? Genet Res 19:336–346

Chakravorty S, Helb D, Burday M et al (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods 69:330–339

Chaucheyras-Durand F, Masseglia S, Fonty G et al (2010) Influence of the composition of the cellulolytic flora on the development of hydrogenotrophic microorganisms hydrogen utilization and methane production in the rumens of gnotobiotically reared lambs. Appl Environ Microbiol 76:7931–7937

Cheema TA, Jirajaroenrat K, Sirinarumitr T (2012) Isolation of a gene encoding a cellulolytic enzyme from swamp buffalo rumen metagenomes and its cloning and expression in Escherichia coli. Anim Biotechnol 23:261–277

Chen Y, Murrell JC (2010) When metagenomics meets stable-isotope probing: progress and perspectives. Trends Microbiol 18:157–163

Chen W, Zhang CK, Cheng Y et al (2013) A comparison of methods for clustering 16S rRNA sequences into OTUs. PLoS One 8:e70837

Cheng F, Sheng J, Cai T et al (2012a) A protease-insensitive feruloyl esterase from China Holstein cow rumen metagenomic library: expression characterization and utilization in ferulic acid release from wheat straw. J Agric Food Chem 60:2546–2553

Cheng F, Sheng J, Dong R, Men Y, Gan L, Shen L (2012b) Novel xylanase from a holstein cattle rumen metagenomic library and its application in xylooligosaccharide and ferulic Acid production from wheat straw. J Agric Food Chem 60:12516–12524

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141–D145

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinf 11:485

Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L, Chen F, Wang M, Li J, Zeng X, Dong Z, Hu S, Li L, Xu J, Huang L, Dong X (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. PLoS One 7:e40430

Dalby PA (2011) Strategy and success for the directed evolution of enzymes. Curr Opin Struct Biol 21:473–480

De Eaton HL, Lorme M, Chaney RL, Craig AM (2011) Ovine ruminal microbes are capable of biotransforming hexahydro-135-trinitro-135-triazine (RDX). Microb Ecol 62:274–286

De Lorme M, Craig M (2009) Biotransformation of 246-trinitrotoluene by pure culture ruminal bacteria. Curr Microbiol 58:81–86

de Menezes AB, Lewis E, O'Donovan M, O'Neill BF, Clipson N, Doyle EM (2011) Microbiome analysis of dairy cows fed pasture or total mixed ration diets. FEMS Microbiol Ecol 78:256–265

de Oliveira MN, Jewell KA, Freitas FS, Benjamin LA, Totola MR, Borges AC, Moraes CA, Suen G (2013) Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Vet Microbiol 164:307–314

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072

Diaz NN, Krause L, Goesmann A, Niehaus K, Nattkemper TW (2009) TACOA: taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. BMC Bioinf 10:56

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998

Edwards JE, McEwan NR, Travis AJ, Wallace RJ (2004) 16S rDNA library-based analysis of ruminal bacterial diversity. Antonie Van Leeuwenhoek 86:263–281

Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD (2012) The great screen anomaly–a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol 93:1005–1020

El Allali A, Rose JR (2013) MGC: a metagenomic gene caller. BMC bioinformatics 14(Suppl 9):S6

Ellison MJ, Conant GC, Cockrum RR, Austin KJ, Truong H, Becchi M, Lamberson WR, Cammack KM (2014) Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. DNA Res 21(5):115–125

El-Metwally S, Hamza T, Zakaria M, Helmy M (2013) Next-generation sequence assembly: four stages of data processing and computational challenges. PLoS Comput Biol 9:e1003345

Faust K, Raes J (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C (2012) Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol 8:e1002606

Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VA, Strompl C, Elborough K, Jarvis G, Neef A, Yakimov MM, Timmis KN, Golyshin PN (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ Microbiol 7:1996–2010

Ferrer M, Beloqui A, Vieites JM, Guazzaroni ME, Berger I, Aharoni A (2009) Interplay of metagenomics and

in vitro compartmentalization. J Microbial Biotechnol 2:31–39

Ferrer M, Ghazi A, Beloqui A, Vieites JM, Lopez-Cortes N, Marin-Navarro J, Nechitaylo TY, Guazzaroni ME, Polaina J, Waliczek A, Chernikova TN, Reva ON, Golyshina OV, Golyshin PN (2012) Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. PLoS One 7:e38134

Fonty G, Gouet P, Nebout JM (1989) Development of the cellulolytic microflora in the rumen of lambs transferred into sterile isolators a few days after birth. Can J Microbiol 35:416–422

Fonty G, Joblin K, Chavarot M, Roux R, Naylor G, Michallon F (2007) Establishment and development of ruminal hydrogenotrophs in methanogen-free lambs. Appl Environ Microbiol 73:6391–6403

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152

Gagen EJ, Mosoni P, Denman SE, Al Jassim R, McSweeney CS, Forano E (2012) Methanogen colonisation does not significantly alter acetogen diversity in lambs isolated 17 h after birth and raised aseptically. Microb Ecol 64:628–640

Galbraith EA, Antonopoulos DA, White BA (2004) Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. Environ Microbiol 6:928–937

Gerlach W, Junemann S, Tille F, Goesmann A, Stoye J (2009) WebCARMA: a web application for the functional and taxonomic classification of unassembled metagenomic reads. BMC Bioinf 10:430

Ghosh TS, Monzoorul Haque M, Mande SS (2010) DiScRIBinATE: a rapid method for accurate taxonomic classification of metagenomic sequences. BMC Bioinf 11(Suppl 7):S14

Gong X, Gruninger RJ, Qi M, Paterson L, Forster RJ, Teather RM, McAllister TA (2012) Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. BMC Res Notes 5:566

Gruninger RJ, Gong X, Forster RJ, McAllister TA (2014) Biochemical and kinetic characterization of the multifunctional beta-glucosidase/beta-xylosidase/alpha-arabinosidase Bgxa1. Appl Microbiol Biotechnol 98(7):3003–3012

Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373

Hamady M, Lozupone C, Knight R (2010) Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J 4:17–27

Hao X, Jiang R, Chen T (2011) Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. Bioinformatics 27:611–618

Hayete B, Bienkowska JR (2005) Gotrees: predicting go associations from protein domain composition using decision trees. Pac Symp Biocomput 10:127–138

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331:463–467

Hoff KJ, Lingner T, Meinicke P, Tech M (2009) Orphelia: predicting genes in metagenomic sequencing reads. Nucleic Acids Res 37:W101–W105

Huang Y, Li W, Perkins D, Li RW (2012) Comparison of de novo short read assemblers on simulated metagenomic data (chapter 5). In: Li RW (ed) Metagenomics and its applications in agriculture biomedicine and environmental studies. Nova Science, New York

Huo W, Zhu W, Mao S (2014) Impact of subacute ruminal acidosis on the diversity of liquid and solid-associated bacteria in the rumen of goats. World J Microbiol Biotechnol 30:669–680

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386

Jacobsen UP, Nielsen HB, Hildebrand F, Raes J, Sicheritz-Ponten T, Kouskoumvekaki I, Panagiotou G (2013) The chemical interactome space between the human host and the genetically defined gut metabotypes. ISME J 7:730–742

Jagtap P, McGowan T, Bandhakavi S, Tu ZJ, Seymour S, Griffin TJ, Rudney JD (2012) Deep metaproteomic analysis of human salivary supernatant. Proteomics 12:992–1001

Jami E, Mizrahi I (2012) Composition and similarity of bovine rumen microbiota across individual animals. PLoS One 7:e33306

Jami E, Israel A, Kotser A, Mizrahi I (2013) Exploring the bovine rumen bacterial community from birth to adulthood. ISME J 7:1069–1079

Jami E, White BA, Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PLoS One 9:e85423

Jenkins TC, Wallace RJ, Moate PJ, Mosley EE (2008) Board-invited review: recent advances in biohydrogenation of unsaturated fatty acids within the rumen microbial ecosystem. J Anim Sci 86:397–412

Kanehisa M (2002) The KEGG database. Novartis Found Symp 247:91–101; discussion 101–103 119–128 244–152

Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL (2012) Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. Nucleic Acids Res 40:e9

Kim M, Morrison M, Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. FEMS Microbiol Ecol 76:49–63

Kim MK, Kang TH, Kim J, Kim H, Yun HD (2012a) Cloning and identification of a new group esterase

(Est5S) from noncultured rumen bacterium. J Microbiol Biotechnol 22:1044–1053

Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won S, Chun J (2012b) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol 62:716–721

Kim M, Lee KH, Yoon SW, Kim BS, Chun J, Yi H (2013) Analytical tools and databases for metagenomics in the next-generation sequencing era. Genomics Inform 11:102–113

Kingston-Smith AH, Davies TE, Rees Stevens P, Mur LA (2013) Comparative metabolite fingerprinting of the rumen system during colonisation of three forage grass (*Lolium perenne* L) varieties. PLoS One 8:e82801

Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI, Janssen PH (2013) Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial archaeal and eukaryotic microorganisms in rumen microbial communities. PLoS One 8:e47879

Ko KC, Han Y, Cheong DE, Choi JH, Song JJ (2013) Strategy for screening metagenomic resources for exocellulase activity using a robotic high-throughput screening system. J Microbiol Methods 94:311–316

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE (2011) Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108(Suppl 1):4578–4585

Laserson J, Jojic V, Koller D (2011) Genovo: de novo assembly for metagenomes. J Comput Biol 18:429–443

Lee YJ, Jenkins TC (2011) Biohydrogenation of linolenic acid to stearic acid by the rumen microbial population yields multiple intermediate conjugated diene isomers. J Nutr 141:1445–1450

Lee CC, Kibblewhite RE, Wagschal K, Li R, Orts WJ (2012a) Isolation of alpha-glucuronidase enzyme from a rumen metagenomic library. Protein J 31:206–211

Lee HJ, Jung JY, Oh YK, Lee SS, Madsen EL, Jeon CO (2012b) Comparative survey of rumen microbial communities and metabolites across one caprine and three bovine groups using bar-coded pyrosequencing and (1)H nuclear magnetic resonance spectroscopy. Appl Environ Microbiol 78:5983–5993

Lee JH, Yi H, Jeon YS, Won S, Chun J (2012c) TBC: a clustering algorithm based on prokaryotic taxonomy. J Microbiol 50:181–185

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Li RW, Sparks ME, Connor EE (2012a) Dynamics of the rumen microbiota. In: Li RW (ed) Metagenomics and its applications in agriculture biomedicine and environmental studies. Nova Science, New York

Li RW, Connor EE, Li C, Baldwin Vi RL, Sparks ME (2012b) Characterization of the rumen microbiota of

pre-ruminant calves using metagenomic tools. Environ Microbiol 14:129–139

Li RW, Wu S, Baldwin RLVI, Li W, Li C (2012c) Perturbation dynamics of the rumen microbiota in response to exogenous butyrate. PLoS One 7:e29392

Li RW, Wu S, Li W, Navarro K, Couch RD, Hill D, Urban JF Jr (2012d) Alterations in the porcine colon microbiota induced by the gastrointestinal nematode Trichuris suis. Infect Immun 80:2150–2157

Li RW, Giarrizzo JG, WU S, Li W, Duringer JM, Craig AM (2014) Metagenomic insights into the RDX-degrading potential of the ovine rumen microbiome. PLoS One 9(11), e110505

Liberles DA, Teufel AI, Liu L, Stadler T (2013) On the need for mechanistic models in computational genomics and metagenomics. Genome Biol Evol 5:2008–2018

Lim S, Seo J, Choi H, Yoon D, Nam J, Kim H, Cho S, Chang J (2013) Metagenome analysis of protein domain colection within cellulase genes of goat rumen microbes. Asian-Aust J Anim Sci 26(8):1144–1151

Lim YW, Schmieder R, Haynes M, Willner D, Furlan M, Youle M, Abbott K, Edwards R, Evangelista J, Conrad D, Rohwer F (2013) Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. J Cyst Fibros 12(2):154–164

Liu B, Pop M (2011) MetaPath: identifying differentially abundant metabolic pathways in metagenomic datasets. BMC Proc 5(Suppl 2):S9

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235

Luo C, Rodriguez RL, Konstantinidis KT (2013) A user's guide to quantitative and comparative analysis of metagenomic datasets. Methods Enzymol 531:525–547

MacDonald NJ, Parks DH, Beiko RG (2012) Rapid identification of high-confidence taxonomic assignments for metagenomic data. Nucleic Acids Res 40:e111

Malmuthuge N, Griebel PJ, Guan le L (2014) Taxonomic identification of commensal bacteria associated with the mucosa and digesta throughout the gastrointestinal tracts of preweaned calves. Appl Environ Microbiol 80:2021–2028

Mao S, Huo W, Zhu W (2013) Use of pyrosequencing to characterize the microbiota in the ileum of goats fed with increasing proportion of dietary grain. Curr Microbiol 67:341–350

McHardy AC, Martín HG, Tsirigos A, Hugenholtz P, Rigoutsos I (2007) Accurate phylogenetic classification of variable-length DNA fragments. Nat Methods 4:63–72

Mende DR, Waller AS, Sunagawa S, Jarvelin AI, Chan MM, Arumugam M, Raes J, Bork P (2012) Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS One 7:e31386

Meyer F, Overbeek R, Rodriguez A (2009) FIGfams: yet another set of protein families. Nucleic Acids Res 37:6643–6654

Michinaka A, Fujii T (2012) Efficient and direct identification of fructose fermenting and non-fermenting bac-

teria from calf gut microbiota using stable isotope probing and modified t-RFLP. J Gen Appl Microbiol 58:297–307

Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95:315–327

Mizrahi I (2012) The rumen plasmidome: a genetic communication hub for the rumen microbiome. Mob Genet Elements 2:152–153

Morgavi DP, Kelly WJ, Janssen PH, Attwood GT (2013) Rumen microbial (meta)genomics and its application to ruminant production. Animal 7(Suppl 1):184–201

Morvan B, Dore J, Rieu-Lesme F, Foucat L, Fonty G, Gouet P (1994) Establishment of hydrogen-utilizing bacteria in the rumen of the newborn lamb. FEMS Microbiol Lett 117:249–256

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res 40:e155

Nguyen NH, Maruset L, Uengwetwanit T, Mhuantong W, Harnpicharnchai P, Champreda V, Tanapongpipat S, Jirajaroenrat K, Rakshit SK, Eurwilaichitr L, Pongpattanakitshote S (2012) Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. Biosci Biotechnol Biochem 76:1075–1084

Noguchi H, Park J, Takagi T (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res 34:5623–5630

Noguchi H, Taniguchi T, Itoh T (2008) MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387–396

Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, Desantis TZ, Brodie EL, Malamud D, Poles MA, Pei Z (2010) Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World J Gastroenterol 16:4135–4144

Omoniyi LA, Jewell KA, Isah OA, Neumann AP, Onwuka CF, Onagbesan OM, Suen G (2014) An analysis of the ruminal bacterial microbiota in West African Dwarf sheep fed grass- and tree-based diets. J Appl Microbiol 116(5):1094–1105

Peng Y, Leung HC, Yiu SM, Chin FY (2011) Meta-IDBA: a de Novo assembler for metagenomic data. Bioinformatics 27:i94–i101

Peng Y, Leung HC, Yiu SM, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ, McAllister TA (2013a) Changes in the rumen epimural bacterial diversity of beef cattle as affected by diet and induced ruminal acidosis. Appl Environ Microbiol 79:3744–3755

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ, McAllister TA (2013b) Characterization of the core rumen microbiome in cattle during transition from forage to concentrate as well as during and after an acidotic challenge. PLoS One 8:e83424

Pinloche E, McEwan N, Marden JP, Bayourthe C, Auclair E, Newbold CJ (2013) The effects of a probiotic yeast on the bacterial diversity and population structure in the rumen of cattle. PLoS One 8:e67824

Pitta DW, Kumar S, Veiccharelli B, Parmar N, Reddy B, Joshi CG (2014) Bacterial diversity associated with feeding dry forage at different dietary concentrations in the rumen contents of Mehshana buffalo (*Bubalus bubalis*) using 16S pyrotags. Anaerobe 25:31–41

Pop M (2009) Genome assembly reborn: recent computational challenges. Brief Bioinform 10:354–366

Pop M, Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. Trends Genet 24:142–149

Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M, Eijsink VG (2012) Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. PLoS One 7:e38571

Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldon T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P (2014) eggNOG v40: nested orthology inference across 3686 organisms. Nucleic Acids Res 42:D231–D239

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Res 40:D290–D301

Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB, Selinger LB, Butler G, Tsang A, McAllister TA, Forster RJ (2011) Snapshot of the eukaryotic gene expression in muskoxen rumen – a metatranscriptomic approach. PLoS One 6(5), e20521

Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55–60

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

Quigley JD 3rd, Schwab CG, Hylton WE (1985) Development of rumen function in calves: nature of protein reaching the abomasum. J Dairy Sci 68:694–702

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinf 12:38

Rashamuse KJ, Visser DF, Hennessy F, Kemp J, der Merwe MP R-v, Badenhorst J, Ronneburg T, Francis-Pope R, Brady D (2013) Characterisation of two bifunctional cellulase-xylanase enzymes isolated from a bovine rumen metagenome library. Curr Microbiol 66:145–151

Reichardt N, Barclay AR, Weaver LT, Morrison DJ (2011) Use of stable isotopes to measure the metabolic activity of the human intestinal microbiota. Appl Environ Microbiol 77:8009–8014

Rho M, Tang H, Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38:e191

Robinson CJ, Bohannan BJ, Young VB (2010) From structure to function: the ecology of host-associated microbial communities. Microbiol Mol Biol Rev 74:453–476

Ross EM, Petrovski S, Moate PJ, Hayes BJ (2013) Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. BMC Microbiol 13:242

Sandri M, Manfrin C, Pallavicini A, Stefanon B (2014) Microbial biodiversity of the liquid fraction of rumen content from lactating cows. Animal 8(4):572–579

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source platform-independent community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541

Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288

Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA the enzyme database: updates and major new developments. Nucleic Acids Res 32:D431–D433

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9:811–814

Sharma VK, Kumar N, Prakash T, Taylor TD (2010) MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. Nucleic Acids Res 38:D468–D472

Simon C, Daniel R (2009) Achievements and new knowledge unraveled by metagenomic approaches. Appl Microbiol Biotechnol 85:265–276

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123

Singh KM, Ahir VB, Tripathi AK, Ramani UV, Sajnani M, Koringa PG, Jakhesara S, Pandya PR, Rank DN, Murty DS, Kothari RK, Joshi CG (2012a) Metagenomic analysis of Surti buffalo (*Bubalus buba-lis*) rumen: a preliminary study. Mol Biol Rep 39:4841–4848

Singh KM, Jakhesara SJ, Koringa PG, Rank DN, Joshi CG (2012b) Metagenomic analysis of virulence-associated and antibiotic resistance genes of microbes in rumen of Indian buffalo (Bubalus bubalis). Gene 507:146–151

Soergel DA, Dey N, Knight R, Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J 6:1440–1444

Sparks ME, Huang Y, Baldwin RL VI, Li W, Connor EE, Li C, Sonstegard TS, Schroeder SG, Bequette BJ, Li RW (2012) Detection of functional shifts in the rumen microbiota in response to propionate intake in cattle. In: Li RW (ed) Metagenomics and its applications in agriculture biomedicine and environmental studies. Nova Science, New York

Sun Y, Cai Y, Liu L, Yu F, Farrell ML, McKendree W, Farmerie W (2009) ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. Nucleic Acids Res 37:e76

Tang S, Antonov I, Borodovsky M (2013) MetaGeneTack: ab initio detection of frameshifts in metagenomic sequences. Bioinformatics 29:114–116

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36

Thoetkiattikul H, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Pattarajinda V, Eurwilaichitr L, Champreda V (2013) Comparative analysis of microbial profiles in cow rumen fed with different dietary fiber by tagged 16S rRNA gene pyrosequencing. Curr Microbiol 67:130–137

Trimble WL, Keegan KP, D'Souza M, Wilke A, Wilkening J, Gilbert J, Meyer F (2012) Short-read reading-frame predictors are not created equal: sequence error causes loss of signal. BMC Bioinf 13:183

Uhlik O, Leewis MC, Strejcek M, Musilova L, Mackova M, Leigh MB, Macek T (2013) Stable isotope probing in the metagenomics era: a bridge towards improved bioremediation. Biotechnol Adv 31:154–165

Vazquez-Castellanos JF, Garcia-Lopez R, Perez-Brocal V, Pignatelli M, Moya A (2014) Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. BMC Genomics 15:37

Vey G, Moreno-Hagelsieb G (2012) Metagenomic annotation networks: construction and applications. PLoS One 7:e41283

Vilo C, Dong Q (2012) Evaluation of the RDP classifier accuracy using 16S rRNA gene variable regions. Metagenomics 1:e235551 doi: 10.4303/mg/235551

Wallace RJ (2008) Gut microbiology – broad genetic diversity yet specific metabolic niches. Animal 2:661–668

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA

sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267

Wang L, Hatem A, Catalyurek UV, Morrison M, Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. PLoS One 8:e78507

Warner D, Dijkstra J, Hendriks WH, Pellikaan WF (2013) Passage kinetics of 13C-labeled corn silage components through the gastrointestinal tract of dairy cows. J Dairy Sci 96:5844–5858

Weber M, Teeling H, Huang S, Waldmann J, Kassabgy M, Fuchs BM, Klindworth A, Klockow C, Wichels A, Gerdts G, Amann R, Glockner FO (2011) Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. ISME J 5:918–928

White JR, Nagarajan N, Pop M (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. PLoS Comput Biol 5:e1000352

Wilke A, Glass EM, Bartels D, Bischof J, Braithwaite D, D'Souza M, Gerlach W, Harrison T, Keegan K, Matthews H, Kottmann R, Paczian T, Tang W, Trimble WL, Yilmaz P, Wilkening J, Desai N, Meyer F (2013) A metagenomics portal for a democratized sequencing world. Methods Enzymol 531:487–523

Wilmes P, Bond PL (2006) Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol 14:92–97

Wu S, Baldwin RL VI, Li W, Li C, Connor EE, Li RW (2012a) The bacterial community composition of the bovine rumen detected using pyrosequencing of 16S rRNA genes. Metagenomics 1:e235571

Wu S, Li RW, Li W, Beshah E, Dawson HD, Urban JF Jr (2012b) Worm burden-dependent disruption of the porcine colon microbiota by Trichuris suis infection. PLoS One 7:e35470

Xia LC, Steele JA, Cram JA, Cardon ZG, Simmons SL, Vallino JJ, Fuhrman JA, Sun F (2011) Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. BMC Syst Biol 5(Suppl 2):S15

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhang W, Chen J, Yang Y, Tang Y, Shang J, Shen B (2011) A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. PLoS One 6:e17915

Zhang P, Gu J, He J, Gao W, Zhang W, Lindsay S, Meldrum DR (2012) Next-generation and future DNA sequencing technologies and metagenomics. In: Li RW (ed) Metagenomics and its applications in agriculture biomedicine and environmental studies. Nova Science, New York

Zhao G, Bu D, Liu C, Li J, Yang J, Liu Z, Zhao Y, Chen R (2012) CloudLCA: finding the lowest common ancestor in metagenome analysis using cloud computing. Protein Cell 3:148–152