J. Dairy Sci. 101:1–14
https://doi.org/10.3168/jds.2017-13356

Mining metagenomic and metatranscriptomic data for clues about microbial metabolic functions in ruminants[1]

Fuyong Li, Andre L. A. Neves, Bibaswan Ghoshal, and Le Luo Guan[2]
Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada T6G 2P5

ABSTRACT

Metagenomics and metatranscriptomics can capture the whole genome and transcriptome repertoire of microorganisms through sequencing total DNA/RNA from various environmental samples, providing both taxonomic and functional information with high resolution. The unique and complex rumen microbial ecosystem is receiving great research attention because the rumen microbiota coevolves with the host and equips ruminants with the ability to convert cellulosic plant materials to high-protein products for human consumption. To date, hundreds to thousands of microbial phylotypes have been identified in the rumen using culture-independent molecular-based approaches, and genomic information of rumen microorganisms is rapidly accumulating through the single genome sequencing. However, functional characteristics of the rumen microbiome have not been well described because there are numerous uncultivable microorganisms in the rumen. The advent of metagenomics and metatranscriptomics along with advanced bioinformatics methods can help us better understand mechanisms of the rumen fermentation, which is vital for improving nutrient utilization and animal productivity. Therefore, in this review, we summarize a general workflow to conduct rumen metagenomics and metatranscriptomics and discuss how the data can be interpreted to be useful information. Moreover, we review recent literatures studying associations between the rumen microbiome and host phenotypes (e.g., feed efficiency and methane emissions) using these approaches, aiming to provide a useful guide to include studying the rumen microbiome as one of the research objectives using these 2 approaches.

Key words: rumen, microbiome, microbiota, metagenomics, metatranscriptomics

INTRODUCTION

The rumen provides anaerobic conditions and redox potentials that favor microbial growth and expression of fiber-degrading enzymes, which allows rumen microorganisms to break down cellulosic plant materials and to meet ruminants' daily energy requirement through producing VFA. Meanwhile, rumen microorganisms are also responsible for producing greenhouse gases [e.g., methane (CH_4), CO_2]. It has been suggested that the rumen microbiome is associated with host phenotypes, such as feed efficiency (Hernandez-Sanabria et al., 2010; Shabat et al., 2016), CH_4 emissions (Shi et al., 2014; Wallace et al., 2015), milk production (Jami et al., 2014), and ruminal acidosis (Khafipour et al., 2009). Due to their importance, rumen microorganisms have attracted increased attention, and rumen microbiology research has rapidly evolved in the last decade to examine their microbial phylogenetic diversities and functional characteristics.

To date, hundreds to thousands of microbial phylotypes have been identified from various rumen samples using sequencing of marker gene PCR amplicons (amplicon sequencing). In this review, the term *microbiota* refers to the assemblage of all microorganisms within a microbial community, and the term *microbiome* refers to the entirety of microbial genomes and transcriptomes from a microbiota, being consistent with the recent definitions from Claesson et al. (2017). It is reported that the rumen microbiota consists of bacteria, archaea, fungi, ciliated protozoa, and phages (with the concentration up to 10^{11}, 10^9, 10^6, 10^6, and 10^{10} per gram of digesta or per milliliter of fluid, respectively; Morgavi et al., 2013). At the same time, genomic information of rumen microorganisms has rapidly accumulated through single genome sequencing, particularly via the recent Hungate 1000 project (http://www.rmgnetwork.org/hungate1000.html). Based on the genomic information, underlying functions of each sequenced phylotype can be predicted. However, due

Received June 19, 2017.
Accepted October 27, 2017.
[1]Presented as part of the Growth and Development Symposium: Microbial Endocrinology in Ruminant Growth and Development at the ADSA Annual Meeting, Pittsburgh, Pennsylvania, June 2017.
[2]Corresponding author: lguan@ualberta.ca

to the lack of available genomes for many important but uncultivable rumen microorganisms (e.g., members belonging to Bacteroidetes), there is still a large gap in the accurate prediction of rumen microbial functional profiles based on the existing genomic information. Therefore, metagenomics and metatranscriptomics represent useful tools for globally cataloging microbial gene and transcript profiles and reflecting overall metabolic functions of rumen microorganisms.

The development of next generation sequencing (**NGS**) technologies has accelerated research of the microbiome using metagenomics and metatranscriptomics. Metagenomics and metatranscriptomics are approaches to studying the entirety of genomes (genes) and transcriptomes (transcripts) of a microbial community (Aguiar-Pulido et al., 2016). Initially, these methodologies started with the cloning of environmental DNA and RNA into vectors (e.g., fosmids, bacterial artificial chromosome vectors), followed by sequence-based or function-driven analysis. Nevertheless, the rapid reduction of costs for NGS accelerated the development of those methodologies, and currently they are usually defined as the direct high-throughput shotgun sequencing of total DNA and RNA in environmental samples (also known as shotgun metagenomics and metatranscriptomics). In this review, we discuss a general workflow to conduct rumen metagenomics and metatranscriptomics with the focus on the direct high-throughput shotgun sequencing and summarize implications of these 2 approaches, aiming to provide a useful guide for the efficient application of those 2 approaches to studying the rumen microbiome in the future.

UNDERSTANDING THE MICROBIOME USING SEQUENCING-BASED APPROACHES

As mentioned above, metagenomics and metatranscriptomics have become powerful and feasible tools for exploring the microbiome of an ecosystem. By using metagenomics and metatranscriptomics, functional potentials (DNA based) and functional activities (RNA based) can be estimated, respectively. The advantage of these approaches is that they can better answer the 2 basic questions "Which members are there?" and "What are they doing?"

"Which Members Are There?"

To date, culture-independent molecular-based taxonomic analysis of the microbiota highly relies on the sequencing of PCR amplicons of targeted marker genes and transcripts [e.g., 16S rDNA/rRNA for bacteria and archaea, 18S rDNA/rRNA for protozoa, and internal transcribed spacer (*ITS*) gene/transcript for fungi], also known as amplicon sequencing. This approach is rapid and low cost; however, the taxonomic assessment of the microbiota can be misleading due to the PCR biases from the primer selection (Hong et al., 2009) and amplification cycling conditions (Huber et al., 2009). In addition, amplicon sequencing has limited potential for discovering new phylotypes because PCR primers are usually designed according to known sequences (Urich et al., 2008; Ross et al., 2012). Recently, it has been reported that metagenomics and metatranscriptomics can be used for the taxonomic assessment of the microbiota, which is less biased and more quantitative than amplicon sequencing (Logares et al., 2014; Li et al., 2016).

"What Are They Doing?"

For a long time, the study of microbial functions highly relied on pure culture-based approaches, which allow researchers to investigate a specific isolate's metabolic functions and to obtain its single genome or transcriptome. However, culture-based methods are not able to characterize functions of uncultured microorganisms or to elucidate overall functions of all microorganisms within a complex microbiota. Because metagenomics and metatranscriptomics can capture the whole genomic and transcriptomic repertoire for both cultivable and uncultivable microorganisms, they are immensely helpful in the functional prediction.

Until now, the functional repertoires of microbiomes from various ecosystems have been explored using metagenomics and metatranscriptomics, including soil (Urich et al., 2008; Tveit et al., 2014), sea water (Baker et al., 2013; Martínez et al., 2013), the animal gastrointestinal tract (Qin et al., 2010; Franzosa et al., 2014), and the plant rhizosphere (Mendes et al., 2014). However, the number of rumen-related metagenomic and metatranscriptomic studies is still low. The understanding of microbial taxonomic and functional characteristics is vital to link the rumen microbiome to host phenotypes, which can help us develop strategies to optimize rumen microbial fermentation for higher productivity.

A GENERAL WORKFLOW TO CONDUCT RUMEN METAGENOMICS AND METATRANSCRIPTOMICS

As more and more researchers are interested in applying metagenomics and metatranscriptomics to answer research questions in various ruminant trials, it is necessary to generate high-quality data and have reliable bioinformatics pipelines for these 2 approaches. In this section, we provide an overview of the major