Methods 124 (2017) 108–119

Contents lists available at ScienceDirect

Methods

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

Integrated metagenomic analysis of the rumen microbiome of cattle reveals key biological mechanisms associated with methane traits

Haiying Wang [a], Huiru Zheng [a,*], Fiona Browne [a], Rainer Roehe [b], Richard J. Dewhurst [b], Felix Engel [c], Matthias Hemmje [c], Xiangwu Lu [d], Paul Walsh [d]

[a] School of Computing and Mathematics, Computer Science Research Institute, Ulster University, United Kingdom
[b] Future Farming Systems, Scotland's Rural College, Edinburgh, United Kingdom
[c] Research Institute for Telecommunication and Cooperation, Germany
[d] NSilico Life Science Ltd., Cork, Ireland

ARTICLE INFO

Article history:
Received 1 April 2017
Received in revised form 29 May 2017
Accepted 31 May 2017
Available online 9 June 2017

Keywords:
Rumen microbial community
Metagenomics
Network-based approaches
Random matrix theory

ABSTRACT

Methane is one of the major contributors to global warming. The rumen microbiota is directly involved in methane production in cattle. The link between variation in rumen microbial communities and host genetics has important applications and implications in bioscience. Having the potential to reveal the full extent of microbial gene diversity and complex microbial interactions, integrated metagenomics and network analysis holds great promise in this endeavour. This study investigates the rumen microbial community in cattle through the integration of metagenomic and network-based approaches. Based on the relative abundance of 1570 microbial genes identified in a metagenomics analysis, the co-abundance network was constructed and functional modules of microbial genes were identified. One of the main contributions is to develop a random matrix theory-based approach to automatically determining the correlation threshold used to construct the co-abundance network. The resulting network, consisting of 549 microbial genes and 3349 connections, exhibits a clear modular structure with certain trait-specific genes highly over-represented in modules. More specifically, all the 20 genes previously identified to be associated with methane emissions are found in a module (hypergeometric test, $p < 10^{-11}$). One third of genes are involved in methane metabolism pathways. The further examination of abundance profiles across 8 samples of genes highlights that the revealed pattern of metagenomics abundance has a strong association with methane emissions. Furthermore, the module is significantly enriched with microbial genes encoding enzymes that are directly involved in methanogenesis (hypergeometric test, $p < 10^{-9}$).

1. Introduction

As one of the most complicated anaerobic microbial ecosystems in nature [1], the rumen provides an environment with stable and favorable physiological conditions for microbial growth and fermentation. Microbes in the rumen are a complex ecosystem predominantly consisting of bacteria, archaea, protozoa and fungi. These microorganisms confer the ability to break down complex polysaccharides and harvest energy from otherwise indigestible food components [2,3]. It has been shown that *Bos taurus* gut microbiota has a paramount role in cattle performance, productivity, health and immunity [4].

However, the rumen microbes are also responsible for the production of the highly potent greenhouse gas methane and nitrogen-rich wastes causing not only the loss of feed gross energy but also contributing to the greenhouse gas emissions and global warming [1,5,6]. Understanding the topological difference in gut microbial community composition is crucial to provide knowledge on the functions of each member of the microbiota to the physiological maintenance of the host. Thus a better understanding of the composition of rumen microbial communities and the association between host genetic and microbial activities has significant applications and implication in bioscience [6,7].

Early exploration of rumen microbiology was mainly dominated by culture-based approaches. Examples include the description of well characterized rumen bacteria based on the isolation of the functionally significant bacterial groups [9,10]. While successfully identifying more than 200 microbial species including bacteria and protozoa from the rumen [1,8], culture-dependent techniques requiring a careful design of protocol for growth of organisms exhibit several significant limitations [11]. They are

* Corresponding author.
E-mail address: h.zheng@ulster.ac.uk (H. Zheng).

http://dx.doi.org/10.1016/j.ymeth.2017.05.029
1046-2023/

not only time consuming and cumbersome [8] but more importantly, culture-based studies are usually unable to reveal the full extent of microbial diversity due to the nature of protocol design and constraints due to culture conditions [11,12].

Advances in next-generation sequencing (NGS) have opened up new avenues in microbial ecology studies. Metagenomics, defined as the direct genetic analysis of DNA from microbial communities sampled in their specific environment without prior need for culturing, is offering unparalleled coverage and depth in determining microbial gut dynamics as long as the analytic resources are available [13,14].

A number of metagenomics studies have investigated rumen microbial populations. These include research by Henderson et al. [7], which investigated whether the microbial community composition was influenced by diet, host species, or geography. It has been found that the composition of rumen microbial community varies with diet and host, but similar bacteria and archaea dominated in nearly all samples. Based on the simultaneous exploration of rumen microbiota and the metabolic phenotype, the study carried out by Morgavi et al. [5] brought new insights on the interactions between microbial populations and the association with the host. By varying a host's diet, Faith et al. [15] examined a human gut microbiota's response to diet in gnotobiotic mice, in which 60% of the variation in species abundance was predicted due to the differences in diet.

More recently, based on the relative abundance of 1570 microbial genes identified in a metagenomics analysis, Roehe and his colleagues [6] developed new selection criteria to be used for predicting methane emissions and other traits such as feed conversion efficiency. Using the partial least squares analysis, 20 and 49 microbial genes were found to be associated with methane emissions and feed conversion efficiency in cattle respectively. Furthermore, functional clusters of microbial genes were identified based on the analysis of the co-abundance network in which the correlation threshold was manually set to 0.9.

By extending our preliminary analysis [16], this study aims to further examine the rumen microbial community in cattle through the integration of metagenomic and network-based approaches. The main objectives include

- to develop an automatic computational technique to objectively determine the correlation threshold used to construct a condition-specific co-abundance network.
- to adopt network systems biology approaches for the identification of key biological mechanisms associated the methane traits.

The rest of the paper is organized as follows. Section 2 briefly describes the methodology and datasets under study. The detailed description of automatic determination and its implementation is provided. The results and discussion are presented in Section 3. The conclusions, together with future research, are given in Section 4.

2. Methodologies

The framework for integrated metagenomic analyses adopted in this study is illustrated in Fig. 1. Based on the relative abundance of

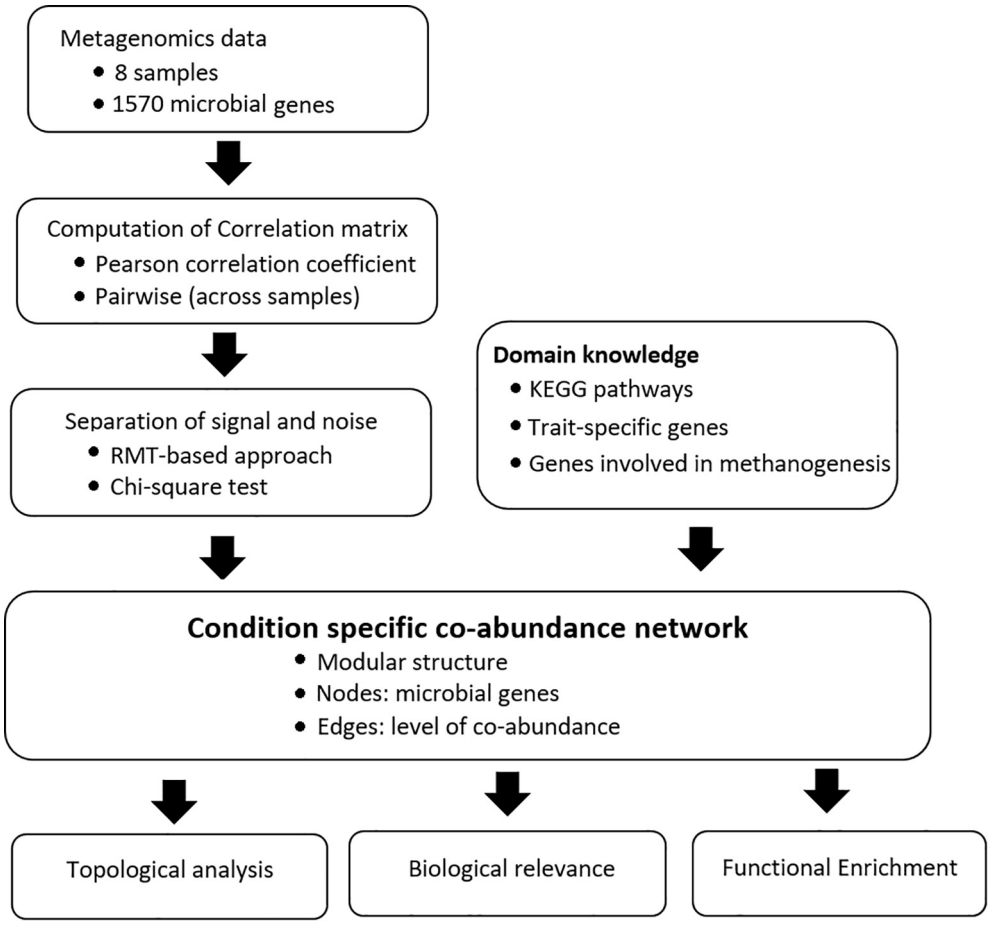

Fig. 1. The framework for integrated metagenomic analyses of the rumen micobiome.