©Cambridge University Press 2012
ISSN 1466-2523

Animal Health Research Reviews **13**(1); 40–53
doi:10.1017/S1466252312000126

High throughput sequencing methods for microbiome profiling: application to food animal systems

Sarah K. Highlander*

Department of Molecular Virology and Microbiology, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030 USA

Received 1 May 2012; Accepted 25 May 2012

Abstract

Analysis of microbial communities using high throughput sequencing methods began in the mid 2000s permitting the production of 1000s to 10,000s of sequence reads per sample and megabases of data per sequence run. This then unprecedented depth of sequencing allowed, for the first time, the discovery of the 'rare biosphere' in environmental samples. The technology was quickly applied to studies in several human subjects. Perhaps these early studies served as a reminder that though the microbes that inhabit mammals are known to outnumber host cells by an order of magnitude or more, most of these are unknown members of our second genome, or microbiome (as coined by Joshua Lederberg), because of our inability to culture them. High throughput methods for microbial 16S ribosomal RNA gene and whole genome shotgun (WGS) sequencing have now begun to reveal the composition and identity of archaeal, bacterial and viral communities at many sites, in and on the human body. Surveys of the microbiota of food production animals have been published in the past few years and future studies should benefit from protocols and tools developed from large-scale human microbiome studies. Nevertheless, production animal-related resources, such as improved host genome assemblies and increased numbers and diversity of host-specific microbial reference genome sequences, will be needed to permit meaningful and robust analysis of 16S rDNA and WGS sequence data.

Keywords: microbial community analysis, 16S rRNA gene sequencing, metagenomics, genome sequencing, whole genome shotgun sequencing, Roche 454, Illumina, microbial diversity

Introduction

Since the vast majority of microbes in any environment are uncultured, polymerase chain reaction (PCR) typing and sequencing methods targeting the 16S ribosomal RNA (rRNA) gene, and other housekeeping or organism-specific genes have been widely adopted to identify and quantitate members of microbial communities (Wooley *et al.*, 2010). For decades, 16S rDNA profiling has been the mainstay of studies of archaeal and bacterial communities in terrestrial and marine environments. In the mid 2000s, these technologies began to be applied to human and animal microbial communities.

Sequencing full-length 16S rRNA gene clones by the Sanger dideoxy chain termination method (Sanger *et al.*, 1977) is the gold standard because paired-end sequencing provides high-quality overlapping reads that can cover the entire gene. Low error rates (<0.1%) are particularly important when using 16S rDNA sequences for taxonomic assessment since a 3% difference across the *ca.* 1500 bp gene usually discriminates between species, phylotypes or operational taxonomic units (OTUs). The transition to 'next-generation' sequencing technologies such as 454 picotiter pyrosequencing (Margulies *et al.*, 2005; Sogin *et al.*, 2006) allowed generation of thousands of reads per

*Corresponding author. E-mail: sarahh@bcm.edu

sample, but the read length was much shorter (originally <100 bases) and the error rates were between 1 and 3%. Such metrics caused significant overestimates of community abundance and diversity and identification of 'new rare' phylotypes. It was soon recognized that poor read quality, including homopolymer tracts, contributed to many of these (Huse *et al.*, 2007).

In 2008, the NIH/NHGRI officially launched the Human Microbiome Project (HMP; www.hmpdacc.org); the MetaHIT (Metagenomics of the Human Intestinal Tract) European research project (www.metahit.eu) was initiated shortly thereafter, and both stimulated interest in all aspects of human metagenomics, from protocol design to data analysis and visualization, to disease correlation. One of the primary goals of the HMP was to collect clinical samples from 15 to 18 body sites (dependent on sex) from 300 healthy subjects then use the nucleic acid obtained from these samples for microbial community analysis by evolving methods of nucleic acid sequencing. In the USA, the involvement of four genome sequencing centers that would sequence thousands of samples from 300 healthy human subjects required development of consistent and reproducible protocols for sample handling, nucleic acid extraction, sequencing and read filtering, assembly and analysis. These protocols and recommendations, as well as analysis and interpretation of much of the subject sequence data, are described in key 'marker' papers (Huttenhower *et al.*, 2012; Methé *et al.*, 2012; Ward *et al.*, 2012) that form the basis for some of the following descriptions and recommendations for current microbial community sequencing and analysis approaches. These approaches are broadly applicable to microbiome projects in production animal species.

Until very recently, few studies focused on the microbiota of animals of agricultural significance. The exceptions were projects focused on the microbial composition and metagenomic function of the bovine rumen (Brulc *et al.*, 2009; Bretschger *et al.*, 2010). Studies of the rumen microbiome and its metabolome continue to be of broad interest to microbiologists, ecologists, nutritionists, and chemists and biologists interested in biomass processing. Early microbial community surveys in production animals were limited to semi-quantitative methods such as polymerase chain reaction (PCR)-denaturing gradient gel electrophoresis or terminal restriction fragment length polymorphism length determinations, or to clone-based Sanger sequencing of the 16S rDNA gene (Yu and Morrison, 2004a; Ozutsumi *et al.*, 2005; Scupham *et al.*, 2008; Durso *et al.*, 2011b; Lowe *et al.*, 2011). The first applications of 454 pyrosequencing to examine the microbiota of food production animals were published in 2008 (Dowd *et al.*, 2008; Qu *et al.*, 2008). High throughput sequencing studies of these microbial communities have not been numerous, and often involve only a few animals, perhaps because of lack of available funding and appropriate protocols, but also because of limited access to sequencing facilities that routinely process and sequence 16S rDNA amplicons at reasonable cost. Table 1 lists some high throughput microbial community profiling projects involving production animal species. Some of these were surveys, while others were hypothesis driven projects; only a small fraction of these addressed any aspect of disease in the host species. Note that some were virus discovery projects that benefited from sample enrichment methods and amplification techniques for high throughput sequencing.

Sequence-based 16S rDNA surveys of the fecal or rumen microbiota of many other non-rodent and non-primate mammals have also been published. These include the studies of the dog (Suchodolski *et al.*, 2008), the dromedary camel (Samsudin *et al.*, 2011), reindeer (Sundset *et al.*, 2009), the polar bear (Glad *et al.*, 2010), as well as a broad survey by Ley *et al.* (2008), in which the fecal diversity of 106 mammals including kangaroo, elephant, rhinoceros, giraffe, panda, zebra, bear, and wild pigs, was examined.

Although the field of microbial community profiling and metagenomics is under rapid development and change, this review will attempt to summarize current approaches, with the knowledge that new sequencing platforms are constantly entering the sequencing arena and being used to develop applications for 16S rDNA and whole genome shotgun (WGS) metagenomic sequencing. In the following sections, methods and considerations for sample handling and nucleic acid extraction for microbiome surveys and metagenomic sequencing will be discussed. Current 16S rDNA and WGS sequencing strategies will be described, followed by brief discussions of established methods for community analysis using the two data types.

Sample acquisition and nucleic acid preparation

Most animal studies utilize fecal samples as the source of microbial nucleic acid, though rumen contents and samples from other body sites, such as the tonsils (see Table 1) have been used. In each case, appropriate handling of the sample is the key as microbial nucleic acid is ubiquitous. An object (tube, tip, swab, etc.) or solution that is bacteriologically 'sterile' is not necessarily free of nucleic acid. This can be especially challenging in veterinary settings. The primary sample (feces, swab sample, milk, and tissue) should be placed in a tube containing extraction buffer that includes nuclease inhibitors, then, if possible, the sample should be appropriately diluted and immediately processed for nucleic acid purification. If prompt DNA extraction is not possible, then the sample should be frozen on dry ice and stored at −80°C for the shortest time possible. The effect of prolonged sample storage is still a subject of debate (Lauber *et al.*, 2010; Wu *et al.*, 2010; Zhao *et al.*, 2011; Bahl *et al.*, 2012), so in most cases, prudent sample handling and experimental 'convenience' are tolerated.

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

Table 1. Examples of high throughput sequence-based microbiome studies in food animal species

Host	Body site/sample	Project goal(s)	Platform template(s)	Reference
Holstein cow	Feces	Survey of commensals and potential food-borne pathogens in 20 animals	454 FLX V4–V5 rDNA amplicons	(Dowd *et al.*, 2008)
Chicken	Cecum	Identify host-specific metavirulomes/ horizontal gene transfer elements	454 GS20 WGS	(Qu *et al.*, 2008)
Beef steer	Rumen fistula	Compare metagenomes of fiber-adherent and liquid fractions	454 GS WGS	(Brulc *et al.*, 2009)
Piglets	Cecum	Compare gut gene expression in neonatal piglets fed sow's milk or formula	454 Titanium RNA-Seq cDNA libraries V1–V4 rDNA	(Poroyko *et al.*, 2010)
Turkey	Intestinal tract	RNA virus discovery	454 FLX cDNA	(Day *et al.*, 2010)
Beef heifer	Feces	Identify virulence-associated and antibiotic-resistance genes	454 FLX WGS	(Durso *et al.*, 2011a)
Pig	Feces	Virus discovery	454 Titanium RT–PCR amplicons	(Shan *et al.*, 2011)
Beef cattle	Feces	Compare effect of feeding practice on the bovine fecal microbiome	454 FLX V6 rDNA amplicons	(Shanks *et al.*, 2011)
Dairy cattle	Rumen fistula	Examine rumen microbial metabolic community activities predicted to degrade cellulosic plant material	Illumina GAIIX and HiSeq2000 WGS libraries	(Hess *et al.*, 2011)
Holstein bull calf	Abomasum	Examine the effect of nematode infection on the abomasmal microbiome	454 Titanium V3–V5 rDNA amplicons, WGS	(Li *et al.*, 2011)
Chicken	Cecum	Examine the effects of antibiotics on the cecal microbiome and metagenome	454 FLX V3 rDNA amplicons, WGS	(Danzeisen *et al.*, 2011)
Pig	Tonsils	Define core microbiome of healthy tonsil	454 FLX V4 16S rDNA amplicon	(Lowe *et al.*, 2012)
Holstein cow	Rumen	Bacteriophage and CRISPR* associations in the bovine rumen	454 FLX Purified phage DNA	(Berg Miller *et al.*, 2012)
Pig	Feces	Survey and comparison of swine microbiome with that from other species	454 G20 and FLX WGS	(Lamendella *et al.*, 2011)
Cross-bred pig	Feces	Examine the effect of bacteriocin-producing *Lactobaccillus salivarius* on the GI microbiota	454 Titanium V4–V5 rDNA amplicons	(Riboulet-Bisson *et al.*, 2012)
Holstein–Friesian cow	Rumen	Survey of taxa in 16 animals	454 FLX V2–V3 rDNA amplicons	(Jami and Mizrahi, 2012)
Holstein cow	Rumen	Identify the rumen 'plasmidome'	Illumina GAIIX phi29 amplified plasmid DNA	(Kav *et al.*, 2012)
Pre-ruminant Holstein bull calf	Rumen	Characterize microbiota of calves fed milk replacement	454 Titanium V3–V5 rDNA amplicons, WGS	(Li *et al.*, 2012)

*CRISPR: 'clustered regularly interspaced short palindromic repeats' implicated in bacterial resistance to bacteriophage (Horvath and Barrangou, 2010).

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

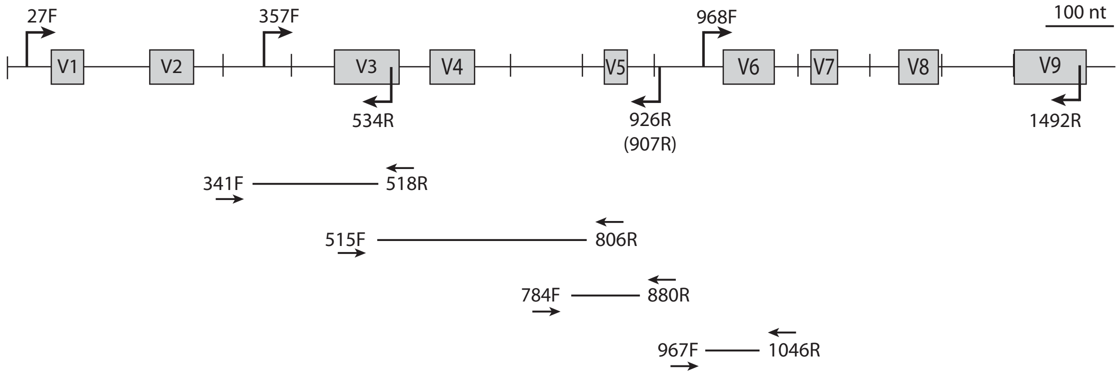

Fig. 1. Linear map of a 'consensus' eubacterial 16S rRNA gene with coordinates, variable regions, and some commonly used 'universal' primer locations shown. The variable regions are drawn to scale using the coordinates derived from Lane (1991) and Woese *et al.* (1990). Common primers used for full-length Sanger PCR amplification and sequencing (such as 27F and 1492R) and those used to create multiple region amplicons (V1–V3, V3–V5 and V6–V9) for HMP 454 sequencing are shown on the coordinate map. Primer pairs that have been used for single variable region amplification for 454 and/or Illumina sequencing are illustrated as horizontal bars with corresponding primers (arrows) and locations are shown below the coordinate map.

A variety of commercial kits and associated techniques have been used for isolation of DNA samples from vertebrates. For the HMP, a single kit, the MO BIO Laboratories, Inc. (Carlsbad, CA) PowerSoil® DNA Isolation Kit, was selected for all body site samples (http://www.hmpdacc.org/tools_protocols/tools_protocols.php). This kit utilizes SDS and bead beating to lyse the cells and includes a proprietary PCR inhibitor removal solution. The DNA is purified from the lysate using a silica spin column. Another kit that is commonly used is the Qiagen (Valencia, CA) QIAamp DNA Stool Kit (Dowd *et al.*, 2008; Riboulet-Bisson *et al.*, 2012). In this protocol, microbes are lysed in SDS at high temperature (e.g. 70°C), the sample is then incubated with a proprietary PCR inhibitor, followed by proteinase K treatment and then the DNA is purified on a spin column. Often, a bead-beating lysis step is used prior to purification using a Qiagen DNA purification kit (Yu and Morrison, 2004b; Qu *et al.*, 2008; Danzeisen *et al.*, 2011). The importance of protocols that include both chemical and bead-beating lysis of the cells has been discussed and examined on many occasions. Recently, Yuan *et al.* (2012) published a systematic evaluation of six common DNA extraction methods that utilized combinations of chemical, enzyme-based and bead-beating-based lysis techniques; five of these used silica columns for final DNA purification. An important aspect of their approach was the application of these methods to a mock community composed of a mixture of equivalent numbers of 11 bacterial species. 16S rDNA sequencing using the 454 FLX platform was performed on technical replicates of DNA isolated by each lysis method. The authors advised that DNA extraction methods for bacterial communities should employ a chemical lysis component plus bead beating and/or addition of mutanolysin to yield a reasonable recovery and representation of the species within a metagenomic sample. The current strategy in our laboratory is similar: we use the MO BIO PowerSoil® DNA Isolation Kit, beginning with a 1 h incubation at 37°C using a mixture of lysozyme, lysostaphin, and mutanolysin.

High throughput sequencing has also permitted identification of viruses and bacteriophages in production animal species (Table 1), including at least two new animal viruses (Hoffmann *et al.*, 2012; Reuter *et al.*, 2012) and numerous bacteriophages (Qu *et al.*, 2008; Berg Miller *et al.*, 2012). For descriptions of techniques to enrich and isolate both RNA and DNA virus sequences from metagenomic samples, see Edwards and Rohwer (2005) and Allander *et al.* (2001). In brief, these protocols involve selective filtration and fractionation steps of the primary sample to isolate viral particles away from host and microbial cells, followed by nuclease treatment, recovery of viral particles, capsid lysis, and random PCR amplification, followed by adapter-specific PCR amplification to create fragments for high throughput sequencing, as detailed below.

16S rDNA gene sequencing methods and read processing

The bacterial and archaeal 16S ribosomal small subunit RNA is approximately 1500 nucleotides (nt) long and can be encoded by one to as many as 14 rRNA operons within an organism. The 16S rRNA molecule has a highly conserved and functionally constrained tertiary structure mainly composed of highly conserved domains (Gutell *et al.*, 1994). These conserved domains are punctuated by nine 'variable' regions (Fig. 1), which in general, map to open loop structures within the molecule (Lane, 1991). Although near full-length sequences of cloned 16S rRNA genes are considered the gold standard for archaeal and bacterial classification, usually to the species level, full-length clone-based approaches have been replaced

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

V1–V3	
Forward	5′-454 B Adapter---27F (AGAGGTTTGATCCTGGCTCAG)-3′
Reverse	5′-454 A Adapter---<u>Bar Code</u>---534R (ATTACCGCGGCTCTGG)-3′
V3–V5	
Forward:	5′-454 B Adapter---357F (CCTACGGGAGGCAGCAG)-3′
Reverse:	5′-454 A Adapter---<u>Bar Code</u>---926 (907) R-MP (CCGTCAATTCMTTTRAGT)-3′
V6–V9	
Forward:	5′-454 B Adapter---U968F (AACGCGAGAACCTTAC)-3′
Reverse:	5′-454 A Adapter---<u>Bar Code</u>---1492R-MP (TACGGYTACCTTGTTAYGACTT)-3′

Fig. 2. Summarized features of primers used to create V1–V3, V3–V5, and V6–V9 16S rDNA amplicons for 454 Titanium sequencing, respectively, where M=A or C; R=A or G; Y=C or T. B adapter=5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3′; A adapter=5′-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3′; Bar code=one of 96 known sequences of 5–10 bp in length, see Ward *et al.*, (2012). Note that primer 926R is equivalent to 907R-MP (Lane, 1991).

by high-throughput methods that directly sequence PCR products (amplicons). In most cases, amplicons that span two or more variable regions serve as reasonable surrogates for classification of organisms within a microbial community at the genus level or higher (Lan *et al.*, 2012), and the lengths of these variable region amplicons (*ca.* 550–600 bp) are appropriate for the current 454 Titanium chemistry.

An important caveat, however, is that 'universal' primers do not amplify all 16S rRNA genes within a sample at equal efficiency. For this reason, some primers contain degenerate bases and in some cases, mixtures of primers are used to capture the expected diversity within a sample. Although many publications describe the deficiencies of particular universal 16S rDNA primer sets, especially for specific microbial communities (Matsuki *et al.*, 2002; Frank *et al.*, 2008; Sim *et al.*, 2012), it is surprising how rarely the choice of variable region and primers for a particular experiment or community is discussed or justified.

454 16S rDNA variable region amplicon sequencing

Individual 16S rRNA gene variable regions, or combinations thereof, have been used for amplicon sequencing on the 454 platform, beginning with the first application of the technology by Sogin *et al.* using the V6 region (967F and 1046R, Fig. 1) and 100 nt reads (Sogin *et al.*, 2006). The current standard 454 Titanium chemistry produces more than 400 nt per amplicon read. A standardized protocol for 16S rDNA amplicon sequencing from human metagenomic DNA samples from all human body sites was developed based on extensive benchmarking by the HMP Consortium (including the use of standard DNA mock communities as sequencing controls) (Methé *et al.*, 2012; Ward *et al.*, 2012). The HMP protocol involved creation of PCR-derived amplicons for each DNA sample that encompass V1–V3, V3–V5 or V6–V9 regions using primers that begin with a 5′ 454 Titanium A or B adapter sequence followed by a 5–10 base sequence, called a 'barcode', and ending with a sequence complementary to the desired variable region (see Fig. 1 for a map of primer locations and Fig. 2 for primer sequence features). A set of 96 bar-coded primers was designed for each of the three 16S rRNA gene regions (http://www.hmpdacc.org) to permit sample multiplexing and inclusion of controls and replicates on a single picotiter sequencing plate. Following PCR amplification, the amplicons are purified, quantified, normalized and pooled. The pools are then used for emulsion PCR (emPCR) and 454 'One-way Read' sequencing. For the HMP, the goal was to achieve 3000 high-quality reads per sample; this metric was based on a number of factors including estimates of coverage, sequencing capacity, and budget (Methé *et al.*, 2012). For comparative analyses, where one expects to examine abundances of 1% or more, then 1000 high-quality reads should be sufficient. This level of coverage will not, however, permit sampling and identification of minor species or complete community analysis where diversity (number of OTUs) is expected to be high. All of these factors should be considered during the design phase of any microbiome study.

Illumina 16S rDNA variable region amplicon sequencing

The high throughput and coverage provided by the Illumina reversible dye terminator sequencing technology

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

(Bentley *et al.*, 2008) has led to the development of protocols for 16S rDNA profiling using shorter reads (75–100 nt). 16S rDNA amplicons have been sequenced on Illumina GAII instruments to examine human and environmental samples. The benefit is a 1000-fold increase in the number of reads per run when compared with 454 sequencing. The short read lengths are problematic, however, for 16S rRNA gene classification, especially at the level of OTU discrimination. This also is due, in part, to increased error rates.

The general scheme for Illumina amplicon preparation is similar to that for 454. Short bar-coded amplicons, corresponding to a single variable region or subregion (usually *ca.* 100–300 bp in length), are generated by bar-coded adapter-based PCR; amplicons are then normalized, pooled, and sequenced. The V3 (Bartram *et al.*, 2011), V4 (Goll *et al.*, 2010; Caporaso *et al.*, 2011), V5 (Lazarevic *et al.*, 2009), and V6 (Bateman *et al.*, 2004) regions have been used with some success (Fig. 1). In most cases, amplicons are sequenced from both ends (i.e. paired-end reads) to obtain the maximal information from the amplicon. As proof of the utility of paired-end sequencing, amplicons derived from Illumina tagged versions of 515F and 806R (Fig. 1), sequenced on a MiSeq System instrument (2×150 nt reads), were able to span the 254 bp V4 region and provide community discrimination at the phylum level (Illumina Application Note 770-2011-013).

A major bottleneck for both the 454 and Illumina-based 16S rDNA sequencing strategies is the creation, purification, and quantification of each individual PCR amplicon library prior to sequencing. Although the on-instrument sequencing costs are continuing to decline, individual library construction remains labor intensive. Methods that automate this aspect of the process will ultimately reduce costs and process time.

Processing raw 16S rDNA sequence data

A critical aspect of interpretation of high throughput 16S rDNA sequence data is read processing and error correction. Without proper read processing, the diversity of a sample can be overestimated and reads can be misclassified. The same is true for data generated by Sanger sequencing, but the general requirement for high quality bases imposed by the Human Genome Sequencing Project, the relatively low number of sequences generated per community (usually many <1000 per sample) and the lack of appreciation of the formation of chimeric 16S rDNA molecules (caused by template switching during PCR amplification), made these issues seem insignificant compared with application to tens of thousands of reads that covered less than a third of the 16S rRNA gene. These factors came into clear view when sequencing amplicons created from a mixed DNA community containing an equal mixture of DNAs from 20 bacterial strains representing 18 OTUs (mock community) were examined as part of protocol development for the HMP (Ward *et al.*, 2012). Using raw reads generated from 454 sequencing of variable region amplicons, the 18 OTU community was estimated to contain about 300 members. Following stringent quality filtering and trimming, the predicted community size fell to about 200 OTUs. Removal of chimeric 16S rDNA sequences reduced the predicted community size to *ca.* 50–100, dependent on the variable region amplicon sequenced. The application of these same processing steps to full-length Sanger reads predicted exactly 18 OTUs. Thus, quality filtering, trimming, and chimera removal are essential. We follow a 454 16S rDNA read processing pipeline that is similar to that described by Ward *et al.*, (2012), though chimeric sequences are detected and removed using UCHIME (Edgar *et al.*, 2011), which is more sensitive and is 1000 times faster than the method used by them.

Overview of 16S rDNA analysis methods

The analysis of 16S rDNA data can be targeted toward broad questions concerning the microbial community, including: (1) How diverse is the community? (2) How do two or more communities differ? (3) Who are the members of the community? (4) How are members of communities interrelated? The answers to these types of queries can be qualitative (is a member present or absent?) or quantitative (how abundant are members?) and numerous statistical and bioinformatic tools are available to explore these questions. Following is a brief overview of some common 16S rDNA analysis approaches.

Question one addresses the issue of alpha diversity, which is a measure of how many different members are within a community. A common illustration of alpha diversity is a collector's curve (Fig. 3a), which plots the number of OTUs or taxa discovered in a sample as a function of the number of sequences generated. These plots can reveal the richness of a community and provide clues as to whether the sample has been sufficiently sampled (by sequencing) to capture all of its diversity (a curve approaching a plateau) or not (a climbing curve). Other common measures of alpha diversity are the Chao1 estimator (Chao, 1984) that calculates species richness by extrapolation, and Simpson's Reciprocal Diversity Index (1/D) (Magurran, 1988), which reflects both the richness (number of taxa) and the evenness (relative taxon abundance) within a sample (Fig. 3b).

The differences (and similarities) between communities is called beta diversity. This can be calculated using one of a number of qualitative or quantitative similarity indices to create distance matrices that can be used to visualize community clusters. The Unique Fraction metric, or UniFrac (Lozupone and Knight, 2005), has been used by many groups to examine beta diversity in microbial

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

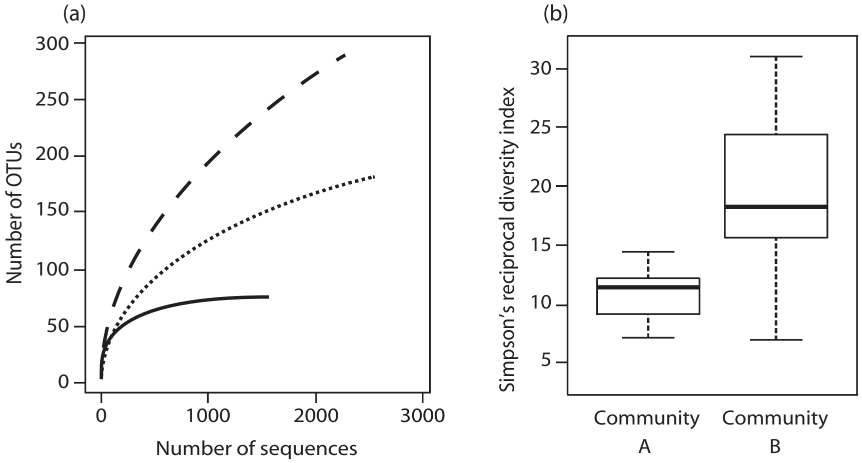

Fig. 3. (a) Example of collector's curves showing three different hypothetical microbial communities. Accumulating numbers of OTUs are plotted as a function of number of sequences determined per community. The community indicated by the filled line is estimated to have been sequenced to saturation between 800 and 1000 sequences and is predicted to contain about 60 OTUs. The community represented by the dotted line has not reached saturation, but is predicted to be composed of not more than 200 OTUs after about 3000 sequences have been collected. The third community, shown by the hashed line, is continuing to accumulate OTUs beyond the level of sequencing depth suggesting a high level of community richness. (b) Box and whisker plot of the Simpson's Reciprocal Diversity Index for two hypothetical microbial communities, where community B is predicted to be more diverse than community A.

communities. This is a metric that uses the evolutionary relatedness of sequences and communities (see below) to calculate distance matrices. In contrast, beta diversity coefficients that do not depend on pre-determined phylogenies are, for example: Bray-Curtis, Morisita-Horn, and the Sörensen indices (Magurran, 1988). A common method used to visualize beta diversity between communities is principal coordinate analysis (PCoA). In simple terms, PCoA processes the calculated distance matrix into a smaller number of variables, called principal components, then reorients the matrix so that the first component accounts for the majority of the variability in the dataset, the second component accounts for the second tier of variability and so on. One can then plot the first two (*x* and *y*), or more coordinates of the analysis, to visualize community clusters (Fig. 4). Other methods for visualization of beta diversity include hierarchical clustering and generation of trees.

To identify and quantitate the membership of a community to answer, 'Who is there?', one can simply submit raw (trimmed and chimera-checked) sequences to a web-based rRNA classifier. The most commonly used classifier is the naïve Bayesian rRNA Classifier (Wang *et al.*, 2007), maintained as part of the Ribosomal Database Project (RDP) (Cole *et al.*, 2009). Although RDP and its associated taxonomy (rdp.cme.msu.edu) are those most commonly used and cited, one should be aware that other 16S rRNA gene taxonomies and associated databases are available, including those as part of Greengenes (McDonald *et al.*, 2012), SILVA (Pruesse *et al.*, 2007), and the NCBI taxonomy (Federhen, 2012). No two taxonomies are equivalent and some are more inclusive than others, some are more frequently updated and curated, and some place special emphasis on particular groups of organisms. About 90% of high quality, chimera checked 454 16S rDNA amplicon sequences from human body sites, other than those from the stool, could be classified to the genus level with greater than 80% confidence using the RDP Classifier with RDP Release 10 (Ward *et al.*, 2012). The most recent update of Greengenes reports improvements for under classified environmental sequences; this update has not yet been tested with HMP data.

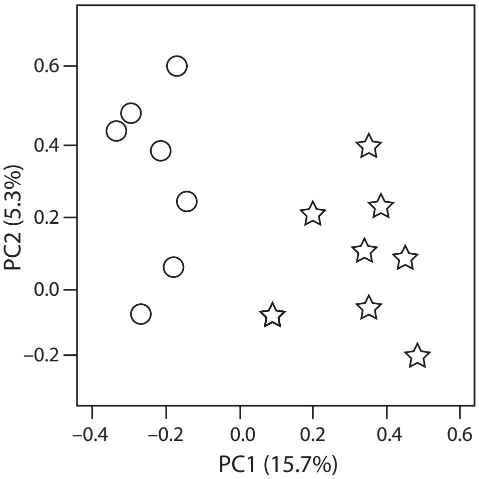

Fig. 4. Example of a scatterplot of the first two components from a PCoA analysis of two microbial communities (indicated by circles or stars) where 15.7% of the difference between the communities is described within the first component (PC1) and 5.3% is reflected within the second component (PC2).

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

Although classification using the RDP Classifier alone can provide rank abundance data, high throughput methods for large datasets usually divide sequences into OTUs, or taxa, before proceeding with downstream analyses. Here is a point of divergence in approach. OTUs can be assigned based on sequence similarity (or distance), or they can be assigned by relating them to a phylogenetic tree. These are considered: (1) phylogeny-independent, or OTU-based methods (Schloss *et al.*, 2009), versus (2) phylogeny-dependent, or phylogenetic methods (Lozupone and Knight, 2005; Lozupone *et al.*, 2011). An advantage of the OTU-based approach is that it has no taxonomic bias and readily accepts novel sequences from unknown taxonomies. A disadvantage is that *de novo* OTU assignment is computationally intensive. The phylogenetic method, such as that used in UniFrac, is considered to be more useful for examining similarities and differences among species in communities (Lozupone and Knight, 2008). Considerable controversy surrounded the rationale for choosing one method or another, especially for beta diversity analyses (Schloss, 2008; Lozupone *et al.*, 2011), though most agree that there are benefits and pitfalls associated with each scheme.

Two freely available 16S rRNA gene analysis packages, QIIME (qiime.org) (Kuczynski *et al.*, 2011) and mothur (http://www.mothur.org) (Schloss *et al.*, 2009), can perform all of the processes described above, plus many more, while a wrapper, called CloVR (clovr.org), that incorporates most of the features of QIIME and mothur has also been developed (Angiuoli *et al.*, 2011). The QIIME package (Kuczynski *et al.*, 2011) uses UniFrac-based phylogenetic beta diversity metrics (Lozupone and Knight, 2005) to create distance matrices that are either unweighted or weighted, for qualitative or quantitative assessments, respectively. The input required is an OTU abundance table and a phylogenetic tree, which can be generated by QIIME. Other more complex types of community interactions, including hierarchical clustering, and networks can also be visualized using the distance matrix file as input. The platform will also perform OTU-based analyses. The mothur package is more focused on OTU-based approaches to evaluate community diversity but also includes the UniFrac algorithms. It includes a useful read processing pipeline that is used by the HMP. Both QIIME and mothur include clustering and network analysis tools that can be used to begin to address the fourth and most complex question that was posed at the beginning of this section, 'How are members of communities interrelated?'

Whole genome shotgun sequencing and analysis methods

Although 16S rRNA gene surveys provide immense information about microbial communities, the problems associated with the use of 16S rDNA data alone, especially in high throughput contexts where only a portion of the 16S rRNA gene is amplified and analyzed, have been previously discussed. As sequencing technologies have evolved, throughput has increased and costs have decreased, so more studies are including or are exclusively composed of WGS sequencing data. This method also has benefits and drawbacks. WGS sequencing is capable of providing true metagenomic data though analysis is complex and processor intensive. Indeed, the first application of 454 sequencing to examine a microbial community was a metagenomic sequencing project of a deep mine environment (Edwards *et al.*, 2006); in food production animals the first use was a study of the chicken cecum (Qu *et al.*, 2008; Table 1). WGS sequencing permits annotation of most bacterial genes within a sample and is not restricted to issues associated with 16S rRNA gene phylogenies. Gene annotations can be used to predict function of a metagenome and thus be used to construct predicted 'metabolomes' (Turnbaugh and Gordon, 2008).

Most recent WGS metagenomic sequencing has been performed on 454 or Illumina platforms, though early studies used Sanger sequencing to obtain data from shotgun libraries of metagenomic DNA constructed in bacterial artificial chromosome (BAC) (Beja *et al.*, 2000a, b) or ColE1-type vectors (Venter *et al.*, 2004; Tringe and Rubin, 2005). A limitation that persists is the quantity (100 ng) of high quality, high molecular weight DNA needed to create a WGS library. In theory, this could be overcome by multi-displacement amplification using an enzyme such as phi29 (Binga *et al.*, 2008) if the metagenomic DNA contains little host contamination. This is rarely the case. In HMP samples, the percent of human contamination ranged from about 1% in stool to more than 80% in samples obtained from the saliva, nares, and vagina (Methé *et al.*, 2012).

WGS read processing

The first step for processing metagenomic data generated by WGS sequencing is to mask reads corresponding to host sequence. This can be accomplished using a tool such as Best Match Tagger (BMTagger) using a reference genome as input. Fortunately, genomes of many important food production animals are available (*Genus species* assembly number): *Bos taurus* UMD_3.1; *Bos indicus* 1.0; *Bubalus bubalis* ASM18099v; *Sus scrofa* 9.2; *Ovis aries* 1.0; and *Gallus gallus* 4.0. Unfortunately, however, most of these are low-quality draft sequences, so host masking is likely to be very inefficient. Once host sequences are removed from the WGS data, duplicate reads, which are sequencing artifacts, are removed, low-quality bases are trimmed, then low-quality reads are identified using an aligner such as the Burrows-Wheeler Aligner (Li and Durbin, 2009), and are discarded.

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

Analysis of WGS data

WGS data can be analyzed as individual processed reads that are mapped to microbial reference genome databases using tools such as MetaPhyler (Liu *et al*., 2011) or WebCARMA (Gerlach *et al*., 2009). Individual read mapping can also be useful for assessment of abundance of taxa. In addition, reads can be assembled into contigs then annotated. Both the HMP and MetaHIT Consortia used SOAPdenovo (Li *et al*., 2010) for assembly of Illumina microbiome data. Hybrid 454/Illumina assemblies were also generated from HMP sequences using the Newbler assembler (Margulies *et al*., 2005; Methé *et al*., 2012).

The J. Craig Venter Institute (JCVI) developed a metagenomics analysis pipeline that can be used for annotation of WGS reads or assemblies (Tanenbaum *et al*., 2010). The pipeline annotates RNAs (rRNAs, tRNAs, and ncRNAs) and performs *ab initio* open reading frame (ORF) calling using MetaGeneAnnotator (Noguchi *et al*., 2008). Predicted protein sequences are examined for functional motifs and cellular localization signals using a variety of tools such as PRIAM (Claudel-Renard *et al*., 2003) for enzyme classification, HMM-Pfam (Bateman *et al*., 2004) and TIGRFAM (Haft *et al*., 2003) for functional motifs, TMHMM (Sonnhammer *et al*., 1998) to identify transmembrane potential domains, and BLAST against JCVI's internal non-redundant nucleotide and protein database, PANDA. The BLAST results from this pipeline can be directly imported into the MEGAN MEtaGenome Analyzer (Mitra *et al*., 2011), where reads are assigned to the NCBI taxonomy and functional analysis is performed using the SEED (http://www.theseed.org) classification system (Overbeek *et al*., 2005). MEGAN also uses the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.kegg.jp) (Kanehisa and Goto, 2000) to construct metabolic pathways from metagenome sequences.

Another established online resource is the Metagenomics Analysis Server (MG-RAST; metagenomics.anl.gov) (Meyer *et al*., 2008), which is also associated with the SEED. MG-RAST annotates, calculates taxonomic distribution of species, rank abundances, and alpha diversity and bins ORFs into functional categories and subsystems. It includes tools to compare and visualize data from preloaded or user uploaded data. MG-RAST can export species and functional category abundance profiles that are compatible as input for QIIME.

Finally, the Joint Genome Institute (JGI) Integrated Microbial Genomes with Metagenome Samples system (IMG/M; img.jgi.doe.gov) (Markowitz *et al*., 2012) also provides tools for functional analysis of microbial communities. A key feature of IMG/M is integration with microbial genomes, finished and draft, in the JGI integrated microbial genome (IMG) system (Markowitz *et al*., 2010). One of the early and continuing goals of the HMP has been population of the NCBI genome database (and thus the IMG) with human microbial reference genome sequences to serve as a 'catalog' for classification of human microbial metagenomic sequences (Nelson *et al*., 2010). To date, about 800 HMP genomes have been deposited and are available from NCBI. The HMP goal is to complete 3000 draft microbial genomes by the end of 2013. Only 26% of the total of 46% of mappable WGS reads from 681 human subjects aligned to these reference genomes (Methé *et al*., 2012), underscoring the need for a broad and relevant database of microbial reference genomes for metagenomics studies. For animals of agricultural interest, a microbial reference genome resource is lacking. As such, data interpretation is limited to what can be gleaned from human microbial genomes. Many of these will be useful proxies, but many will not.

Conclusion/Perspective

Based on the extensive analysis performed by the HMP community, the recommendation in 2012 for microbial community profiling would be a multiplexed 454 amplicon sequencing approach targeting either the V1–V3 or V3–V5 16S rDNA regions followed by robust read trimming and chimera detection and removal. If sample DNA quantity is not limiting and host contamination is not expected to be substantial, then paired-end WGS sequencing on the Illumina platform is an attractive strategy that can provide additional taxonomic and functional predictions. For example, if stool is the sample of choice, then host contamination should be of minimal concern. On the other hand, if other body sites or tissues are to be studied, then the availability and quality of the host genomic sequence will be of paramount importance since host contamination of metagenomic reads will be significant. Another encumbrance is the lack of suitable microbial reference genomes for WGS read mapping. These problems may restrict food animal microbiome studies to 16S rDNA or other PCR-based gene surveys.

Although not discussed here, a critical component of all microbiome/metagenomic projects is proper attention to study design. Power calculations, using effect size to predict sample size, consideration of case-matched controls, appropriate sample handling, and care to prevent environmental DNA contamination are some of the lessons learned from the HMP. The HMP results also indicate that sample sequencing depth, both for 16S rDNA and WGS methods was more than adequate for the community comparisons made, but despite the exponential expansion of the human-related microbial reference genome sequences at NCBI, the reference catalog is far from saturated. Future projects, especially those involving food production animals and other lesser studied vertebrates will benefit from improved host genome sequences, new technologies to separate host from microbial DNA, and efforts to include commensal

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

microbes of agriculturally important animals in reference genome sequencing initiatives.

References

Allander T, Emerson SU, Engle RE, Purcell RH and Bukh J (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 11609–11614.

Angiuoli SV, Matalka M, Gussman A, Galens K, Vangala M, Riley DR, Arze C, White JR, White O and Fricke WF (2011). CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics* **12**: 356.

Bahl MI, Bergstrom A and Licht TR (2012). Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiology Letters* **329**: 193–197.

Bartram AK, Lynch MD, Stearns JC, Moreno-Hagelsieb G and Neufeld JD (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Applied and Environmental Microbiology* **77**: 3846–3852.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C and Eddy SR (2004). The Pfam protein families database. *Nucleic Acids Research* **32**: D138–D141.

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and Delong EF (2000a). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB, Feldman RA and Delong EF (2000b). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology* **2**: 516–529.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R and Smith AJ (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.

Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA, Flint HJ, Lamed R, Bayer EA and White BA (2012). Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environmental Microbiology* **14**: 207–227.

Binga EK, Lasken RS and Neufeld JD (2008). Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME Journal* **2**: 233–241.

Bretschger O, Osterstock JB, Pinchak WE, Ishii S and Nelson KE (2010). Microbial fuel cells and microbial ecology: applications in ruminant health and production research. *Microbial Ecology* **59**: 415–427.

Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM, Henrissat B, Nelson KE and White BA (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 1948–1953.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N and Knight R (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America* **108** (suppl. 1): 4516–4522.

Chao A (1984). Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**: 265–270.

Claudel-Renard C, Chevalet C, Faraut T and Kahn D (2003). Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research* **31**: 6633–6639.

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

Danzeisen JL, Kim HB, Isaacson RE, Tu ZJ and Johnson TJ (2011). Modulations of the chicken cecal microbiome and metagenome in response to anticoccidial and growth promoter treatment. *Public Library of Science ONE* **6**: e27949.

Day JM, Ballard LL, Duke MV, Scheffler BE and Zsak L (2010). Metagenomic analysis of the turkey gut RNA virus community. *Virology Journal* **7**: 313.

Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG and Edrington TS (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiology* **8**: 125.

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

Durso LM, Harhay GP, Bono JL and Smith TP (2011a). Virulence-associated and antibiotic resistance genes of microbial populations in cattle feces analyzed using a metagenomic approach. *Journal of Microbiological Methods* **84**: 278–282.

Durso LM, Harhay GP, Smith TP, Bono JL, Desantis TZ and Clawson ML (2011b). Bacterial community analysis of beef cattle feedlots reveals that pen surface is distinct from feces. *Foodborne Pathogens and Disease* **8**: 647–649.

Edgar RC, Haas BJ, Clemente JC, Quince C and Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Edwards RA and Rohwer F (2005). Viral metagenomics. *Nature Reviews Microbiology* **3**: 504–510.

Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander Jr EC and Rohwer F (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7**: 57.

Federhen S (2012). The NCBI Taxonomy database. *Nucleic Acids Research* **40**: D136–143.

Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA and Olsen GJ (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Applied and Environmental Microbiology* **74**: 2461–2470.

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

Glad T, Bernhardsen P, Nielsen KM, Brusetti L, Andersen M, Aars J and Sundset MA (2010). Bacterial diversity in faeces from polar bear (*Ursus maritimus*) in Arctic Svalbard. *BMC Microbiology* **10**: 10.

Goll J, Rusch DB, Tanenbaum DM, Thiagarajan M, Li K, Methe BA and Yooseph S (2010). METAREP: JCVI metagenomics reports – an open source tool for high-performance comparative metagenomics. *Bioinformatics* **26**: 2631–2632.

Gutell RR, Larsen N and Woese CR (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiological Reviews* **58**: 10–26.

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

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 and Rubin EM (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**: 463–467.

Hoffmann B, Scheuch M, Hoper D, Jungblut R, Holsteg M, Schirrmeier H, Eschbaumer M, Goller KV, Wernike K, Fischer M, Breithaupt A, Mettenleiter TC and Beer M (2012). Novel orthobunyavirus in cattle, Europe, 2011. *Emerging Infectious Diseases* **18**: 469–472.

Horvath P and Barrangou R (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**: 167–170.

Huse SM, Huber JA, Morrison HG, Sogin ML and Welch DM (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**: R143.

Huttenhower C, Gevers D, Knight R, Abubucker S, Badger J, Chinwalla A, Creasy H, Earl A, Fitzgerald M, Fulton R, Giglio M, Pepin K, Lobos E, Madupu R, Magrini V, Martin J, Mitreva M, Muzny D, Sodergren E, Versalovic J, Wollam A, Worley K, Wortman J, Young S, Zeng Q, Aagaard K, Allen-Vercoe E, Alm E, Alvarado L, Andersen G, Anderson S, Appelbaum E, Arachchi H, Armitage G, Arze C, Ayvaz T, Baker C, Begg L, Belachew T, Bihan M, Blaser M, Bloom T, Bonazzi V, Brooks J, Buck G, Buhay C, Busam D, Campbell J, Canon R, Cantarel B, Chen I-M, Chhibba S, Chu K, Ciulla D, Clemente J, Clifton S, Conlan S, Crabtree J, Cutting M, Davidovics N, Davis C, Desantis T, Deal C, Delehaunty K, Dewhirst F, Deych E, Dooling D, Dugan S, Dunne W, Durkin A, Edgar R, Erlich R, Farmer C, Farrell R, Faust K, Feldgarden M, Felix V, Fisher S, Fodor A, Forney L, Foster L, Di Francesco V, Friedman J, Friedrich D, Fronick C, Fulton L, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni M, Goldberg J, Goll J, Gonzalez A, Gujja S, Haake S, Haas B, Hamilton H, Harris E, Hepburn T, Herter B, Hoffmann D, Holder M, Huang K, Huse S, Izard J, Jansson J, Jiang H, Jordan C, Joshi V, Katancik J, Keitel W, Kelley S, Kells C, King N, Knights D, Kong H, Koren O, Kota K, Kovar C, Kyrpides N, Lee S, Lemon K, Lennon N, Lewis C, Lewis L, Ley R, Li K, Liolios K, Liu Y, Lozupone C, Lunsford R, Madden T, Mahurkar A, Mannon P, Mardis E, Markowitz V, McDonald D, McEwen J, McGuire A, McInnes P, Mehta T, Mihindukulasuriya K, Newsham I, Nusbaum C, O'laughlin M, Orvis J, Pagani I, Palaniappan K, Patel S, Pearson M, Peterson J, Podar M, Pohl C, Pollard K, Pop M, Priest M, Proctor L, Qin X, Raes J, Reid J, Rhodes R, Riehle K, Rivera M, Rodriguez-Mueller B, Rogers Y-H, La Rosa P, Ross M, Russ C, Sanka R, Sankar P, Sathirapongsasuti J, Schloss J, Schloss P, Schmidt T, Schriml L, Schubert A, Segata N, Segre J, Shannon W, Sharp R, Sharpton T, Shenoy N, Sheth N, Simone G, Singh I, Smillie C, Sobel J, Spicer P, Sutton G, Sykes S, Tabbaa D, Thiagarajan M, Tomlinson C, Torralba M, Truty R, Vishnivetskaya T, Walker J, Wang L, Wang Z, Ward D, Watson M, Wellington C, Wetterstrand K, White J, Wilczek-Boney K, Wu Y, Wylie K, Wylie T, Yandava C, Ye Y, Yooseph S, Youmans B, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker J, Birren B, Gibbs R, Highlander S, Methé B, Nelson K, Petrosino J, Weinstock G, Wilson R, White O, Griggs A, Liu B, Lo C-C, Howarth C, Sommer D, McCorrison J, Miller J, Mavromatis K, Chen L, Ye L, Scholz M, Rho M, Abolude O, Minx P, Chain P, Koren S, Treangen T, Bhonagiri V, Warren W, Ding Y and Ravel J (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 875–922.

Jami E and Mizrahi I (2012). Composition and similarity of bovine rumen microbiota across individual animals. *Public Library of Science ONE* **7**: e33306.

Kanehisa M and Goto S (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**: 27–30.

Kav AB, Sasson G, Jami E, Doron-Faigenboim A, Benhar I and Mizrahi I (2012). Insights into the bovine rumen plasmidome. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 5452–5457.

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG and Knight R (2011). Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current Protocols in Bioinformatics* Chapter 10: Unit 10 17. 934–935.

Lamendella R, Domingo JW, Ghosh S, Martinson J and Oerther DB (2011). Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiology* **11**: 103.

Lan Y, Wang Q, Cole JR and Rosen GL (2012). Using the RDP classifier to predict taxonomic novelty and reduce the search space for finding novel organisms. *Public Library of Science ONE* **7**: e32491.

Lane DJ (1991). 16S/23S rRNA sequencing. In: Stackebrandt E and Goodfellow M (eds) *Nucleic acid Techniques in Bacterial Systematics*, Chichester, England: John Wiley & Sons, pp. 115–175.

Lauber CL, Zhou N, Gordon JI, Knight R and Fierer N (2010). Effect of storage conditions on the assessment of bacterial

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

community structure in soil and human-associated samples. *FEMS Microbiology Letters* **307**: 80–86.

Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Osteras M, Schrenzel J and Francois P (2009). Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *Journal of Microbiological Methods* **79**: 266–271.

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R and Gordon JI (2008). Evolution of mammals and their gut microbes. *Science* **320**: 1647–1651.

Li H and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J and Wang J (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research* **20**: 265–272.

Li RW, Connor EE, Li C, Baldwin Vi RL and Sparks ME (2012). Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environmental Microbiology* **14**: 129–139.

Li RW, Wu S, Li W, Huang Y and Gasbarre LC (2011). Metagenome plasticity of the bovine abomasal microbiota in immune animals in response to *Ostertagia ostertagi* infection. *Public Library of Science ONE* **6**: e24417.

Liu B, Gibbons T, Ghodsi M, Treangen T and Pop M (2011). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* **12** (suppl. 2): S4.

Lowe BA, Marsh TL, Isaacs-Cosgrove N, Kirkwood RN, Kiupel M and Mulks MH (2011). Microbial communities in the tonsils of healthy pigs. *Veterinary Microbiology* **147**: 346–357.

Lowe BA, Marsh TL, Isaacs-Cosgrove N, Kirkwood RN, Kiupel M and Mulks MH (2012). Defining the "core microbiome" of the microbial communities in the tonsils of healthy pigs. *BMC Microbiology* **12**: 20.

Lozupone C and Knight R (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **71**: 8228–8235.

Lozupone C, Lladser ME, Knights D, Stombaugh J and Knight R (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME Journal* **5**: 169–172.

Lozupone CA and Knight R (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiology Reviews* **32**: 557–578.

Magurran AE (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF and Rothberg JM (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.

Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, Liolios K, Pagani I, Anderson I, Mavromatis K, Ivanova NN and Kyrpides NC (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Research* **40**: D123–129.

Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Anderson I, Lykidis A, Mavromatis K, Ivanova NN and Kyrpides NC (2010). The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Research* **38**: D382–390.

Matsuki T, Watanabe K, Fujimoto J, Miyamoto Y, Takada T, Matsumoto K, Oyaizu H and Tanaka R (2002). Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Applied and Environmental Microbiology* **68**: 5445–5451.

McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, Andersen GL, Knight R and Hugenholtz P (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME Journal* **6**: 610–618.

Methé B, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH, Chinwalla AT, Earl AM, Fitzgerald MG, Fulton RS, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bihan M, Blaser M, Bloom T, Bonazzi V, Brooks JP, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chen I-MA, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, Desantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Dooling DJ, Dugan SP, Dunne WM, Durkin AS, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney L, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Gujja S, Haake SK Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, King NB, Knight R, Knights D, Kong HH, Koren O, Kota KC, Kovar CL, Kyrpides NC, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu Y, Lozupone CA, Lunsford RD, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Priest ME, Proctor LM, Qin X, Raes J, Reid JG, Rhodes R, Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers Y-H, La Rosa PS, Ross MC, Russ C, Sanka RK, Sankar P, Sathirapongsasuti JF, Schloss JA, Schloss PD, Schmidt TM, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu YQ, Wylie KM, Wylie T, Yandava C, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Weinstock GM, and Wilson RK (2012). A framework for human microbiome research. *Nature* **486**: 1034–1080.

Meyer F, Paarmann D, D'souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J and Edwards RA (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Mitra S, Stark M and Huson DH (2011). Analysis of 16S rRNA environmental sequences using MEGAN. *BMC Genomics* **12** (suppl. 3): S17.

Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR, Rusch DB, Mitreva M, Sodergren E, Chinwalla AT, Feldgarden M, Gevers D, Haas BJ, Madupu R, Ward DV, Birren BW, Gibbs RA, Methe B, Petrosino JF, Strausberg RL, Sutton GG, White OR, Wilson RK, Durkin S, Giglio MG, Gujja S, Howarth C, Kodira CD, Kyrpides N, Mehta T, Muzny DM, Pearson M, Pepin K, Pati A, Qin X, Yandava C, Zeng Q, Zhang L, Berlin AM, Chen L, Hepburn TA, Johnson J, McCorrison J, Miller J, Minx P, Nusbaum C, Russ C, Sykes SM, Tomlinson CM, Young S, Warren WC, Badger J, Crabtree J, Markowitz VM, Orvis J, Cree A, Ferriera S, Fulton LL, Fulton RS, Gillis M, Hemphill LD, Joshi V, Kovar C, Torralba M, Wetterstrand KA, Abouellleil A, Wollam AM, Buhay CJ, Ding Y, Dugan S, FitzGerald MG, Holder M, Hostetler J, Clifton SW, Allen-Vercoe E, Earl AM, Farmer CN, Liolios K, Surette MG, Xu Q, Pohl C, Wilczek-Boney K and Zhu D (2010). A catalog of reference genomes from the human microbiome. *Science* **328**: 994–999.

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

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, De Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O and Vonstein V (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* **33**: 5691–5702.

Ozutsumi Y, Hayashi H, Sakamoto M, Itabashi H and Benno Y (2005). Culture-independent analysis of fecal microbiota in cattle. *Bioscience Biotechnology and Biochemistry* **69**: 1793–1797.

Poroyko V, White JR, Wang M, Donovan S, Alverdy J, Liu DC and Morowitz MJ (2010). Gut microbial gene expression in mother-fed and formula-fed piglets. *Public Library of Science ONE* **5**: e12459.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J and Glockner FO (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35**: 7188–7196.

Qu A, Brulc JM, Wilson MK, Law BF, Theoret JR, Joens LA, Konkel ME, Angly F, Dinsdale EA, Edwards RA, Nelson KE and White BA (2008). Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *Public Library of Science ONE* **3**: e2945.

Reuter G, Nemes C, Boros A, Kapusinszky B, Delwart E and Pankovics P (2012). Astrovirus in wild boars (*Sus scrofa*) in Hungary. *Archives of Virology* **157**: 1143–1147 [Epub ahead of print].

Riboulet-Bisson E, Sturme MH, Jeffery IB, O'donnell MM, Neville BA, Forde BM, Claesson MJ, Harris H, Gardiner GE, Casey PG, Lawlor PG, O'Toole PW and Ross RP (2012). Effect of *Lactobacillus salivarius* bacteriocin Abp118 on the mouse and pig intestinal microbiota. *Public Library of Science ONE* **7**: e31113.

Samsudin AA, Evans PN, Wright AD and Al Jassim R (2011). Molecular diversity of the foregut bacteria community in the dromedary camel (*Camelus dromedarius*). *Environmental Microbiology* **13**: 3024–3035.

Sanger F, Nicklen S and Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463–5467.

Schloss PD (2008). Evaluating different approaches that test whether microbial communities have the same structure. *ISME Journal* **2**: 265–275.

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 and Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**: 7537–7541.

Scupham AJ, Patton TG, Bent E and Bayles DO (2008). Comparison of the cecal microbiota of domestic and wild turkeys. *Microbial Ecology* **56**: 322–331.

Shan T, Li L, Simmonds P, Wang C, Moeser A and Delwart E (2011). The fecal virome of pigs on a high-density farm. *Journal of Virology* **85**: 11697–11708.

Shanks OC, Kelty CA, Archibeque S, Jenkins M, Newton RJ, McLellan SL, Huse SM and Sogin ML (2011). Community structures of fecal bacteria in cattle from different animal feeding operations. *Applied and Environmental Microbiology* **77**: 2992–3001.

Sim K, Cox MJ, Wopereis H, Martin R, Knol J, Li MS, Cookson WO, Moffatt MF and Kroll JS (2012). Improved detection of *Bifidobacteria* with optimised 16S rRNA-gene based pyrosequencing. *Public Library of Science ONE* **7**: e32543.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM and Herndl GJ (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**: 12115–12120.

Sonnhammer EL, Von Heijne G and Krogh A (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology* **6**: 175–182.

Suchodolski JS, Camacho J and Steiner JM (2008). Analysis of bacterial diversity in the canine duodenum, jejunum, ileum, and colon by comparative 16S rRNA gene analysis. *FEMS Microbiology Ecology* **66**: 567–578.

Sundset MA, Edwards JE, Cheng YF, Senosiain RS, Fraile MN, Northwood KS, Praesteng KE, Glad T, Mathiesen SD and Wright AD (2009). Rumen microbial diversity in Svalbard reindeer, with particular emphasis on methanogenic archaea. *FEMS Microbiology Ecology* **70**: 553–562.

Tanenbaum DM, Goll J, Murphy S, Kumar P, Zafar N, Thiagarajan M, Madupu R, Davidsen T, Kagan L, Kravitz S, Rusch DB and Yooseph S (2010). The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. *Standards in Genomic Sciences* **2**: 229–237.

Tringe SG and Rubin EM (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics* **6**: 805–814.

Turnbaugh PJ and Gordon JI (2008). An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**: 708–713.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C,

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press

Rogers YH and Smith HO (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**: 5261–5267.

Ward DV and Jumpstart Consortium Human Microbiome Project Data Generation Working Group (2012). Evaluation of 16S rDNA-based community profiling for human microbiome research. *Public Library of Science ONE*: 1229–1232.

Woese CR, Winker S and Gutell RR (1990). Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 8467–8471.

Wooley JC, Godzik A and Friedberg I (2010). A primer on metagenomics. *Public Library of Science Computational Biology* **6**: e1000667.

Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, Hwang J, Chen J, Berkowsky R, Nessel L, Li H and Bushman FD (2010). Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiology* **10**: 206.

Yu Z and Morrison M (2004a). Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology* **70**: 4800–4806.

Yu Z and Morrison M (2004b). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Bio Techniques* **36**: 808–812.

Yuan S, Cohen DB, Ravel J, Abdo Z and Forney LJ (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *Public Library of Science ONE* **7**: e33865.

Zhao J, Li J, Schloss PD, Kalikin LM, Raymond TA, Petrosino JF, Young VB and Lipuma JJ (2011). Effect of sample storage conditions on culture-independent bacterial community measures in cystic fibrosis sputum specimens. *Journal of Clinical Microbiology* **49**: 3717–3718.

https://doi.org/10.1017/S1466252312000126 Published online by Cambridge University Press