Analysis of community structure and species richness of protozoa-enriched rumen metagenome from Indian Surti by shotgun sequencing

Ravi K. Shah[1,2], Amrutlal K. Patel[1], Tejas M. Shah[1], Krishna M. Singh[1], Neelam M. Nathani[1] and Chaitanya G. Joshi[1,*]

[1]Department of Animal Biotechnology, College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Anand 388 001, India
[2]BRD School of Biosciences, Sardar Patel University, V. V. Nagar 388 120, India

The present study applies metagenomics to characterize the diversity and relative abundance of protozoa residing in the rumen of Indian Surti buffalo (*Bubalus bubalis*). To increase protozoa representation in the metagenome, protozoa enrichment was performed by density sedimentation and confirmed by quantitative real time PCR. The highly enriched metagenome sample was subjected to shotgun sequencing by Ion Torrent PGM which resulted in 10,303,375 reads totalling 1.6 gigabases. The taxonomic profile obtained by comparison with SILVA SSU database showed predominance of the class Litostomatea (99.78%) followed by Coccidia (0.10%) and Aconoidasida (0.06%). At the genus level *Isotricha* (48.06%) followed by *Polyplastron* (9.90%), *Dasytricha* (9.87%) and *Eudiplodinium* (7.47%) were predominant. The taxonomic assignment based on protein coding regions showed discrepancy with the SSU-based assignments, possibly due to the absence of most eukaryotic genomes in public databases. According to the SEED subsystems annotation database, genes for protein metabolism were the most abundant followed by genes for RNA metabolism, regulation and cell signalling. The present study offers a preliminary snapshot of diversity, functional potential and relative abundance of protozoa within the Indian Surti buffalo rumen and also expands our knowledge of these unicellular eukaryotes present in the rumen ecosystem.

Keywords: Buffalo rumen, metagenome, protozoa enrichment, quantitative real time PCR, shotgun sequencing.

RUMEN, one of the most diverse ecosystems in nature, harbours a complex consortium of interdependent anaerobic prokaryotes, protozoa, fungi, bacteriophages and methanogens. These act synergistically to convert lignocellulosic feeds into volatile fatty acids serving as an energy supplement for host ruminants[1]. Owing to the complexity of rumen ecosystem and the intricate interactions among various microbial groups within the rumen, it is difficult to decipher a precise role for each microorganism. Despite the fact that rumen microbiomes have been studied extensively, a majority of them have focussed on domain bacteria, because of their numerical abundance and metabolic diversity[2], thus leaving rumen fungi, archaea, bacteriophages and protozoa underexplored.

Rumen protozoa, besides contributing nutrients[3] to the host animal, help in digesting carbohydrate and protein containing feedstuff by secretion of various saccharolytic and proteolytic enzymes[4,5] in microbial protein turnover[6], modulation of bacterial populations[7–9] and association with methanogens[8,10]. Numerous species of ciliates occupy the rumen, where they constitute about half of the total microbial biomass[11]. Although bacterial species play a vital role in digestion, protozoa are reported to digest 25–50% of the total fibre[12]. The polymorphic nature and complexity in cultivation of protozoa, has hampered the assessment of their taxonomy within the rumen ecosystem.

Traditionally, microscopy has been the preferred method for identifying and enumerating protozoal populations in rumen samples[13]; however, it has several inherent limitations like misidentification and low sensitivity. Therefore, molecular techniques have recently been used in ecological studies of ruminal protozoa[14–19]. Advances in metagenomics and direct analysis of DNA fragments from environmental samples, have greatly increased our ability to understand the metabolic potential, gene content, functional capacity, diversity and relative abundance of indigenous microbial communities of a given habitat.

The exploration of less dominant species within a habitat requires enormous sequence coverage, which necessitates higher computational resources. In response, keeping in mind the significance of protozoa in rumen ecosystem and the practical power of a metagenomic approach, a sample, intentionally enriched for rumen

*For correspondence. (e-mail: cgjoshi@rediffmail.com)

protozoa, was used to study the diversity and relative abundance of protozoa within Surti buffalo rumen. The sequences obtained by high throughput sequencing were analysed using bioinformatics tools and comparative analysis was performed between reference ruminal liquor and an enriched metagenome sample.

Materials and methods

Sampling and protozoa enrichment

The experiment was performed using an individual Surti buffalo, reared at Department of Animal Nutrition, College of Veterinary Science and Animal Husbandry, Anand, Gujurat, India. Permission was obtained from the Committee for Purpose of Control and Supervision of Experiments on Animals (CPCSEA) before initiating the study. Approximately 300 ml of rumen fluid was collected using a flexible stomach tube and vacuum pump, three hours post feeding. The diet consisted of mature pasture grass (*Dicanthium annulatum*), green fodder Napier Bajra 21 (*Pennisetum purpureum*) and compound concentrate mixture (20% crude protein along with 65% total digestible nutrients). Ruminal contents were distributed into 50 ml falcon tubes (each tube subjected to specific number of washes of saline) after appropriate mixing and enrichment of protozoa was obtained using the modified procedure of William and Strachan[20]. The term enrichment, in our study, refers to the increase in protozoa : bacteria ratio. Briefly, the ruminal content was allowed to stand still for 20 min and pellets obtained from all the tubes were pooled. Glucose solution (0.1%) was added to the pooled sample in equal volume and allowed to stand again for 15 min. The supernatant was cautiously removed, to not to disturb the sediment. The pellet was expected to contain protozoa in addition to some bacteria and other high specific gravity rumen contents, and was given multiple washes with normal saline. NaCl (5 ml of 0.8%) was added to the pellet, vortexed and centrifuged at 400 *g* for 3 min. The supernatant was removed carefully and pellet was resuspended in 8 ml of 0.8% NaCl solution. This washing step was repeated 5 to 9 times (in individual tubes), and the resulting protozoa enriched sample (PES) was stored at −80°C until DNA extraction. Although there was a single animal, the sample was divided into different tubes for giving a different number of washes of saline (i.e. 5, 7 and 9 washes of saline). Reference rumen liquor was similarly collected as mentioned above and processed further for DNA extraction without subjecting for enrichment.

DNA extraction from the protozoa enriched sample

Approximately, 100 mg of sample was mixed with 1 ml of lysis buffer (10 mM *Tris*Cl, 10 mM KCl, 10 mM $MgCl_2$, 0.5 M NaCl, 2 mM EDTA, 10% SDS (freshly added)) and homogenized by bead beating with 25 mg of acid washed glass beads at 30 Hz per second for 3 min in TissueLyser (Qiagen). The homogenized sample was then incubated at 70°C for 15 min with periodic inversion. The tube was centrifuged at 16,000 *g* for 15 min at 4°C and the supernatant was recovered into a fresh tube and stored at 4°C. The pellet was resuspended in 800 µl of lysis buffer followed by homogenization and centrifugation. The supernatants obtained were pooled and used for DNA extraction via phenol–chloroform method. The quality of the extracted DNA was evaluated by running DNA on 0.8% agarose gel and the quantity was measured using a Nanodrop ND-1000 spectrophotometer.

Evaluation of protozoa enrichment by real time PCR

To quantify the enrichment of protozoa in PES, the total protozoa rDNA copies were quantitated relative to rDNA copies of total bacteria by qPCR using the ABI 7500 FAST real time PCR system (Applied Biosystems, USA), with QuantiFast SYBR green PCR master mix (Qiagen, USA). The 18S (specific for protozoa) and 16S (specific for bacteria) rRNA gene-targeted primer sets used in the study are listed in Table 1. The amplification reactions were performed in triplicates with a total volume of 15 µl, containing 10 ng of template DNA, 7.5 µl of 2× QuantiFast SYBR Green PCR master mix (Qiagen, USA), 0.5 µl of each primer (10 pmol/µl) and 5.5 µl of nuclease free water. The cycling conditions comprised of an initial denaturation step of 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. After each extension step, fluorescent readings were taken, and a final melting analysis obtained by slow heating with 0.1°C/s increment from 60°C to 95°C, and fluorescence was collected at 0.1°C intervals. The threshold cycle was determined during the exponential phase of amplification. The relative enrichment of protozoa over bacteria in protozoa enriched sample was deduced from the difference in Ct value obtained for 18S (protozoa) and 16S (bacterial) specific primers.

Metagenome sequencing and data analysis

Fragment DNA library construction and shotgun sequencing was carried out on an Ion Torrent PGM platform (Life Technologies), using Ion Torrent 316 chip following the manufacturer's instructions. In brief, the metagenomic DNA extracted from the PES sample was subjected to enzymatic fragmentation to obtain fragments in the range of 280 to 300 bp. The desired size fragments were ligated to library adaptors followed by emulsion PCR and subsequet recovery and loading of beads on the Ion Torrent 316 chip. All resultant sequencing data sets were uploaded to the Metagenomics – Rapid Annotation using

Table 1. PCR primer specific for 18S and 16S rDNA copies of rRNA genes to detect protozoa and total bacteria in rumen fluid

Target group	Sequence (5′–3′)	Product size (bp)
Protozoa (18S)	Forward – GCTTTCGWTGGTAGTGTATT Reverse – CTTGCCCTCYAATCGTWCT	223
Total Bacteria (16S)	Forward – CCTACGGGAGGCAGCAG Reverse – ATTACCGCCGCTGTTGG	194

Table 2. C_T values for protozoa and total bacteria obtained from 18S and 16S gene specific primers

Sample	C_T value protozoa	Bacteria	Difference (ΔC_T)	$\Delta\Delta C_T$
P1 (5 washes)	14.569	26.811	–12.242	–14.306
P2 (7 washes)	14.751	27.401	–12.65	–14.714
P3 (9 washes)	15.253	27.207	–11.954	–14.018
Ruminal fluid (reference)	15.901	13.837	2.064	0

Subsystem Technology (MG-RAST) server (http://metagenomics.anl.gov/) checked for low quality reads prior to dereplication, annotation and phylogenetic identification. Taxonomic analysis in MG-RAST compared the metagenomic sequences with non-redundant multi-source protein annotation database (M5NR) and SILVA small subunit rRNA database (SSU) on MG-RAST server. Functional analysis was carried out by comparing selected sequence features against the SEED subsystems database, on the MG-RAST server.

Results

To study the protozoa community within rumen ecosystem and to overcome the problem of low representation of protozoa within the metagenome, rumen protozoa were enriched from a ruminal liquor sample by density sedimentation. The quality of DNA recovered from PES has been shown in Figure S1 (see Supplementary Information online). As expected, mechanical homogenization by bead beating caused some shearing of metagenomic DNA. A DNA yield of around 300 µg was obtained from the protozoa fractionated from approximately 300 ml of ruminal fluid.

Evaluation of protozoa enrichment by real time PCR

C_T (threshold cycle) values for 16S and 18S rDNA copies, demonstrating relative abundance of bacteria and protozoa measured after washes of saline, in PES and reference ruminal liquor, are presented in Table 2. Comparative threshold ($\Delta\Delta C_T$) value was calculated by the formula: $\Delta\Delta C_T = (C_{T(\text{protozoa})} - C_{T(\text{Bacteria})})_{Pn} - (C_{T(\text{protozoa})} - C_{T(\text{Bacteria})})_{\text{Ruminal fluid}}$; where P_n is the number of washes of saline given to protozoa-enriched sample. About 2^{12} fold increase in protozoa rDNA concentration was detected over bacterial rDNA after 7 washes of saline, compared to the four-fold higher rDNA concentration of bacteria than protozoa in reference ruminal liquor, thus indicative of significant enrichment of protozoa.

Metagenome analysis

Metagenome sequencing performed using Ion Torrent PGM resulted in 10,303,375 reads with an average read length of 155 ± 85 bp length totalling 1,602,222,749 bases. Out of the entire sequences uploaded to MG-RAST, 4,533,325 sequences (44%) failed to pass the QC pipeline. Among the sequences that passed QC, 241,433 sequences (2.3%) contained ribosomal RNA genes. Of the remainder, 1,105,531 sequences (10.7%) contained predicted proteins with known functions and 2,712,312 sequences (26.3%) contained predicted proteins with unknown function. Sequences numbering 1,710,774 (16.6%) did not match with any rRNA genes or predicted proteins. Taxonomic information available from all annotation source database used by MG-RAST is presented in Figure 1.

Community structure of protozoa within enriched-metagenome based on 18S rRNA genes

According to SILVA SSU database, at maximum *e*-value of $1e^{-20}$, a minimum identity of 90%, and a minimum alignment length of 50 bp, a total of 24,246 sequence features were identified in the given metagenome dataset, having putative rRNA regions. Of these 14,443 (59.57%) sequences affiliated to eukaryota, 9433 (38.91%) to

bacteria, 129 (0.5%) to archaea, 6 (0.02%) to viruses and 235 (0.97%) sequences were unable to be assigned at domain level, whereas in reference ruminal liquor, out of the total 462 rRNA sequence features identified, 422 (91.34%) were assigned to domain bacteria, 33 (7.14%) to eukaryota and 7 (1.52%) sequences remained unassigned, thus clearly demonstrating enrichment of protozoa in PES. Within PES, 13,364 (92.52%) rRNA sequences from eukaryota, belonged to the protozoa, all being ciliates and no hits for flagellate protozoa were found. At the class level, rRNA genes for Litostomatea (99.78%, 13,335 sequences) were the most abundant among ciliates, followed by Coccidia (0.10%, 14 sequences) and Aconoidasida (0.06%, 9 sequences). At the genus level, Litostomatea comprised of 11 different genera of ciliate rumen protozoa, viz. *Dasytricha*, *Isotricha*, *Eudiplodinium*, *Didinium*, *Polyplastron*, *Ophyroscolex*, *Entodinium*, *Epidinium*, *Diploplastron*, *Metadinium* and *Spathidium* with rRNA genes of genera *Isotricha* (48.06%, 6423 sequences) dominating the class, followed by *Polyplastron* (9.90%, 1324 sequence), *Dasytricha* (9.87%, 1320 sequences) and *Eudiplodinium* (7.47%, 999 sequences). The pathogenic protozoa of the class Coccidia and Aconoidasida (phylum Apicomplexa) were represented by *Toxoplasma*, *Eimeria* and *Plasmodium*. Relative abundance of rRNA genes of all genera detected in the protozoa enriched metagenome is shown in Figure 2. The rarefaction curve obtained for the enriched metagenome dataset, based on SILVA SSU rRNA database of MG-RAST is shown in Figure 3. Adequate coverage of rumen microbiome was illustrated in the rarefaction curve, depicting that it already passed the steep region and levelled off to where fewer new species could be found in enlarged sequencing depth.

Phylogenetic analysis based on protein coding regions

To obtain information on taxonomic diversity, based on protein encoding genes as taxonomic markers using M5NR database of MG-RAST, at the minimum identity cut-off of 70% and maximum *e*-value cut-off of $1e^{-5}$, we found that protozoa enriched metagenome consists of 20,964 sequences (64.77%) of protozoa out of the total eukaryotic content as compared to the 1141 sequences (59.67%) in reference ruminal liquor. Moreover, the total eukaryotic content was quite low in reference ruminal fluid (3.77%, 1912 sequences) compared to protozoa enriched metagenome (28.83%, 32,365 sequences), demonstrating effective enrichment of rumen protozoa. Ciliate protozoa consisted of 61.95% (19,961 sequences) of the total eukaryotic content in PES as compared to the 46.33% (886 sequences) in reference liquor. Regarding diversity, hits were obtained for 77 genus and 124 species of protozoa in PES as compared to 21 genus and 29 species in reference ruminal fluid. *Entodinium*, *Epidinium*, *Polyplastron*, *Eudiplodinium*, *Isotricha* and *Dasytricha* genus observed in the PES were among previously reported ciliates found in the ruminant animals. *Entodinium* (28.22%, 5917 sequences) was found to be most abundant among protozoa, followed by *Toxoplasma* (14.94%, 3133 sequences) and *Epidinium* (7.2%, 1511 sequences). The five most abundant pathogenic organisms to the host animals in the PES belonged to genus *Toxoplasma*, *Neospora*, *Eimeria*, *Plasmodium* and *Babesia*. Relative abundance of all the protozoa genera found within the given metagenome is shown in Table S1 (see Supplementary Information online). Out of total bacterial population, Firmicutes (35.9%, 354,542 sequences) represented the most abundant phyla in protozoa enriched sample, followed by Proteobacteria (31.3%, 309,786 sequences) and Bacteroidetes (27.3%, 270,009 sequences); whereas in case of reference ruminal liquor Bacteroidetes (68.25%, 127,128 sequences) dominated with substantially lower

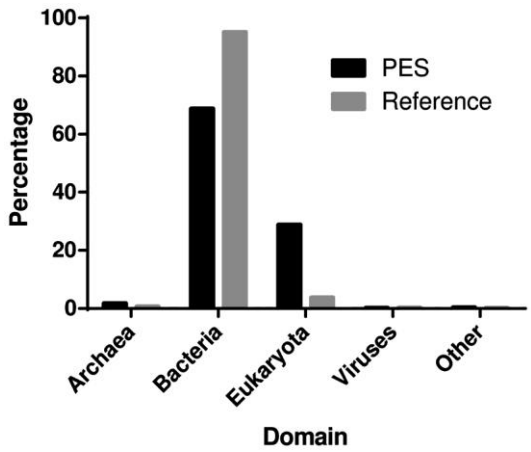

Figure 1. Taxonomic composition of protozoa enriched sample and reference ruminal liquor. The information is based on all annotation source database used by MG-RAST server. PES: Protozoa enriched sample, Reference: Sample without any pre-treatment.

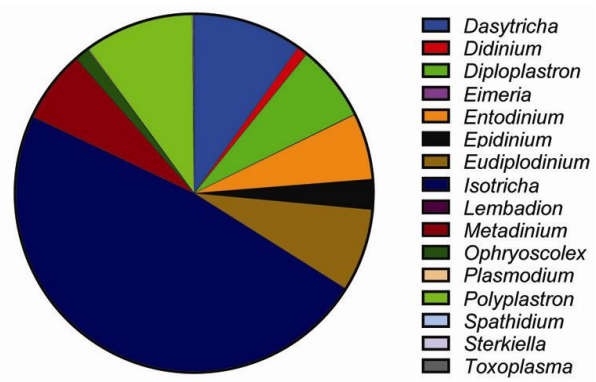

Figure 2. Relative percentage of protozoa genera (in percentage of total annotated reads) in the protozoa enriched metagenome as deduced from SILVA SSU database.

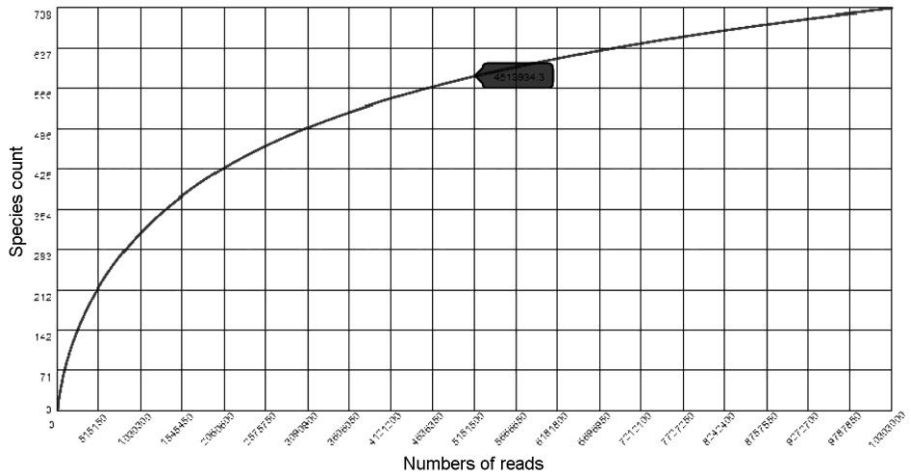

Figure 3. Rarefaction curve derived from 16S/18S reads from the metagenome based on SILVA SSU database on MG-RAST server.

proportions of Firmicutes (16.95%, 31,572 sequences) and Proteobacteria (8.0%, 15,032 sequences).

On changing the minimum identity cut-off to more stringent 90% criteria, the number of protozoa genus reduced to 24 in protozoa enriched sample, including *Dasytricha* and *Isotricha* which are commonly found rumen ciliates along with pathogenic *Neospora*. However, in case of reference ruminal liquor, protozoa genus decreased to 6 including common rumen ciliates of *Epidinium* genus and no hits for pathogenic protozoa. Among bacterial population, the results showed at these identity cut-off were quite distinctive, Proteobacteria (99.35%, 158,897 sequences) being the most abundant phyla followed by Actinobacteria (0.59%, 957 sequences). The phyla Proteobacteria was mainly dominated by the class Gammaproteobacteria (91.30%, 155,629), followed by Betaproteobacteria (8.06%, 13,744 sequences) and Alphaproteobacteria (0.015%, 27 sequences).

The proportion of archaea observed in protozoa enriched metagenome (1.83%) was significantly higher as compared to reference rumen liquor (0.68%) as seen from all annotation source database used by MG-RAST.

Functional analysis

Prior to functional analysis, the sequencing reads from bacterial contaminants were removed by subjecting selected sequence features (20,964 sequences assigned to protozoa by MG-RAST using M5NR database), from the metagenomic dataset, to BLASTN analysis using locally generated bacterial database. Metagenomic reads having identity greater than 80% with any sequence in the database were not considered for further analysis. The preliminary assessment of these filtered reads for the metabolic potential of protozoa within the rumen ecosystem, was performed using the SEED subsystem[21] database of MG-RAST. At maximum *e*-value cut-off of $1e^{-20}$, a minimum identity of 90%, and a minimum alignment length of 50 bp, the predominant functional category included carbohydrate metabolism (31.42% of assigned sequence features), clustering-based subsystems (12.76%), cell wall and capsule (8.59%), fatty acids, lipids, and isoprenoids (7.99%), membrane transport (4.36%), cofactors, vitamins, prosthetic groups, pigments (4.19%), RNA metabolism (3.48%), DNA metabolism (1.49%), dormancy and sporulation (0.64%), amino acids and derivatives (0.29%) and protein metabolism (0.18%). Interestingly, the second largest percentage of sequence features was assigned to the functional category iron metabolism and acquisition (24.55%). The functional classification of selected sequence features from protozoa enriched metagenome, to the subsystem level 1 of SEED database is shown in Figure 4. Furthermore, within the carbohydrate metabolism gene category, central carbohydrate metabolism (30.65%) was the most abundant functional category, followed by one carbon metabolism (23.62%) and monosaccharide metabolism (13.17%). Heme and hemin uptake and utilization systems in Gram-negatives, were most abundant functional categories (40.98%), followed by ferrous iron transporter (23.28%) and siderophores (17.53%) within the iron acquisition and metabolism category.

Discussion

According to our knowledge, this is the first report describing the protozoa community of rumen, through high throughput shotgun sequencing. Although the role of protozoa in rumen cellulose digestion is widely recognized[22], compared to bacteria and fungi relatively less information is available on rumen protozoa which can be attributed to difficulties with isolation, culture or maintenance of protozoa and lack of available genomes within the database. Moreover, there is a frequent loss in viability of ruminal isolates during purification or sub-culturing[23].

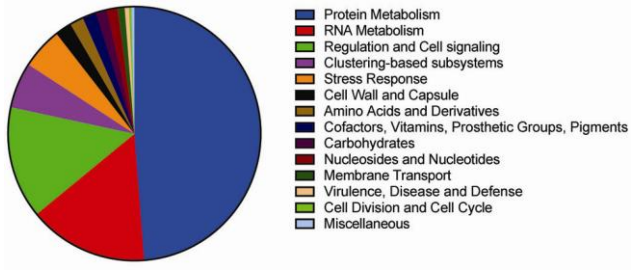

Figure 4. Functional assignment of protozoa specific sequence features identified from M5NR database using the SEED subsystem database of MG-RAST server. Percentage of reads assigned to each hierarchical level-1 subsystem of SEED database.

Traditionally, protozoa is identified by classical culture-based techniques and microscopic examination of protozoal morphological features. However, it is often hampered by the polymorphic nature of protozoa, difficulty in cultivation, time consuming and laborious techniques involved and the technical expertise required for identifying each species morphologically. In recent times, culture-independent techniques, for instance, denaturing gradient gel electrophoresis (DGGE), thermal gradient gel electrophoresis, single-strand conformation polymorphism and terminal restriction fragment length polymorphism have been used to study protozoal communities in various ecosystems[24–26]. Although these methods are useful to study less diverse communities, they may not reflect the integral diversity of rumen protozoa due to their low throughput. High throughput sequencing method was recently applied as promising method to identify the microbial community diversity within a given habitat[27–32]. Although metagenomics has aided to attain a comprehensive view of genetic diversity of hard to exploit natural ecosystems, such as rumen, allowing better assessment of the microbial diversity present within the niche, storage and analysis of the tremendous data thus generated has been a primary bottleneck for the scientific progress. MG-RAST[33], a freely available public domain metagenome annotation server, running at Argonne National Laboratory (ANL) has served primarily to overcome such bottleneck and has hugely impacted metagenome data analysis.

To assess the true diversity of protozoa overcoming the abundance of bacteria present in the rumen, an intentional enrichment of rumen protozoa was performed from ruminal liquor by a modified procedure of Williams and Strachan[20]. The scum layer of flocculent plant debris is formed by microbial fermentation when ruminal liquor is allowed to stand. The protozoa pellet thus obtained when washed with 0.1% glucose solution optimizes flocculation, removing protozoa from other ruminal contents. The dislodgement of protozoa from feed particles reduces particle specific gravity, causing them to float, thus facilitating separation. The protozoa pellet obtained when washed with normal saline helps to curtail contamination by bacteria and plant debris.

Previously, rRNA genes were used for quantification of protozoa biomass[34,35], but rRNA levels may vary dramatically during various life stages although the protozoa population does not undergo any change. Thus rDNA has been proposed as a specific and precise marker that can be accurately quantified by quantitative real time PCR[17–19]. The primer set used for amplification of protozoa rDNA in the present study adapted from Sylvester *et al.*[17] has been designed from the internal region of 18S rDNA gene and thus any non-specific amplification would result in erroneous dissociation curves. Dissociation curve profiles for both 18S and 16S rDNA genes were compact with absence of any other peaks, suggesting high level of specificity and absence of primer dimeriziation.

The reliability of MG-RAST for profiling of shotgun sequenced metagenome data has been reported previously[36]. Taxonomic assignment of the given metagenomic dataset when analysed with the SSU rRNA database of SILVA, rRNA genes for the genus, *Isotricha* was most dominant within the domain eukaryote, followed by *Polyplastron*, *Dasytricha* and *Eudiplodinium*. Previous studies report dominance of *Dasytricha* sp. in the Surti buffalo rumen[37]. This discrepancy in results may be attributed to variation in diet, primers used, samples analyzed, sampling procedures, DNA extraction methods and host animals. Pathogenic species of the phylum Apicomplexa were detected in the metagenome with a percentage identity of 100% and average alignment length of 60. The taxonomic assignment of these organisms may be considered true up to the genera level, but taxonomic assignment at species level should be construed with caution because of a high level of conservation within the 18S rRNA gene sequences between closely related species and absence of 18S rRNA gene sequences for some organisms within the database.

Further taxonomic assignment was carried out based on protein coding regions by M5NR database of MG-RAST. At minimum identity cut-off of 70%, abundance was observed for previously reported common rumen ciliate genera like *Entodinium*, *Epidinium*, *Polyplastron*, *Eudiplodinium*, *Isotricha* and *Dasytricha* alongwith pathogenic genus *Toxoplasma* and *Neopsora*. Dehority[13] and Skillman *et al.*[19] reported *Entodinium* sp. to be major rumen ciliate in sheep rumen with hay-fed diet. Dehority *et al.*[13] reported that rumen consists of many protozoa out of which *Entodinium* sp. are dominant. According to previous reports, *Entodinium* can constitute up to 90% of the total ciliate protozoa[38]. However the abundance of *Entodinium* in our study was quite low (28.22%), which might be due to the fact that all previous studies were based on construction of clone libraries, which might have overrepresented the predominant genera and neglected rare organisms. High throughput sequencing approach used in this study may help to represent rare genera present in the community which may account for decrease in the percentage of *Entodinium* species. The incongruity between

results based on M5NR database and SSU rRNA database may be due to the presence of a few complete genome sequences of rumen protozoa, within the database, perhaps leading to incorrect assignment of the reads. Moreover there is variation between the rDNA copy numbers per cell/genome and considerable difference in numbers between taxa[39]. It is to be noted that less than 40% of the total sequences were annotated by MG-RAST. Bacterial genomes dominate these databases and thus it is a possibility that a fraction of these unclassified sequences is associated with genes of yet-to-be annotated eukaryotic genomes. Therefore, forthcoming studies should concentrate on sequencing genomes of these yet-to-be sequenced rumen protozoa to provide valuable insights into the protozoa community.

Archaea population was another major difference between protozoa enriched metagenome sample (1.83%) and reference ruminal liquor (0.68%). Increase in methanogens (archaea) population may be due to increase in protozoa number which is known to have frequent and intimate association with methanogens[40]. The association between the hydrogenotrophic methanogens and bovine ciliates[41] is a known characteristic of several anaerobic systems rich in hydrogen, such as rumen. The association is important to both participants, as rumen protozoa depend on hydrogen evolving fermentation, that provides substrate to the methanogens, which in turn is beneficial to the protozoa, as hydrogen is inhibitory to protozoa metabolism if not removed[42].

Anaerobic protozoa often harbours synergistic relationship with the bacteria in rumen. Santos *et al.*[43] described a few species in the class Proteobacteria having intracellular life-cycle. According to Brigge and Gortz[44], most of the intracellular bacteria in protozoa are Proteobacteria. Abundance of Proteobacteria (99.35%) at minimum identity cut-off of 90% in our protozoa-enriched sample, may be linked to higher number of protozoa in the metagenome, but due to lack of sequence availability of the protozoa in the database, reads may be wrongly assigned to bacteria instead of protozoa.

While assessing functional potential of protozoa community in the rumen, carbohydrate metabolism was found to be the most abundant functional category. Genes coding for central carbohydrate metabolism and one-carbohydrate metabolism were most dominant within this category and it is unsurprising that protozoa play a critical role in all the subsystems of carbohydrate metabolism. Previous reports have suggested that about 62% of the total cellulolytic activity associated with plant materials may be attributed to protozoa[45–47]. A noteworthy feature of the functional potential of rumen protozoa was the abundance of genes for iron acquisition and metabolism, suggesting a high potential of protozoa to compete for iron with other members of microbial community. The prominence of iron for microbial viability and its undoubtedly complex role in all aspects of life cycle of protozoa needs further understanding in future studies.

In conclusion, this article reported diversity, species richness and functional architecture of the protozoa species harboured in Surti buffalo rumen, by metagenomic analysis of protozoa enriched metagenome. The enrichment method resulted in a significantly higher representation of protozoa species compared to metagenomic analysis of whole rumen sample, thereby enabling us to gain a preliminary snapshot of community structure, metabolic potential and relative abundance of protozoa within rumen ecosystem and has helped to expand our current knowledge of these unique eukaryotic species present within the rumen. Rumen protozoa were found to play a significant role in protein and RNA metabolism. As reflected from the results, previously reported ruminal protozoa species are dominant within the metagenome. Pathogenic protozoa identified in the rumen metagenome might be due to their sequence similarity with the rumen protozoa, or else they may represent potential source of reservoir for transmission of diseases. If the assignment of these pathogenic genera is true, their association with non-pathogens and importance in maintenance of rumen dynamics, need to be considered in further studies.

1. Mackie, R., Gut environment and evolution of mutualistic fermentative digestion. In *Gastrointestinal Microbiology* (eds Mackie, R. and White, B.), Springer US, 1997, vol. 1, pp. 13–35.
2. Bryant, M. P., Nutritional requirements of the predominant rumen cellulolytic bacteria. *Fedproc*, 1973, **32**, 1809–1813.
3. Imai, S., Katsuno, M. and Ogimoto, K., Distribution of rumen ciliate protozoa in cattle, sheep and goat and experimental transfaunation of them. *Jpn. J. Zootech. Sci.*, 1978, **49**, 494–505.
4. Nagasawa, T., Uchida, T. and Onodera, R., Exopeptidase activity of mixed rumen ciliate protozoa. *Anim. Sci. Technol.*, 1992, **63**, 481–487.
5. Shinchi, S. and Abe, M., Species of rumen ciliate protozoa which produce extra-cellular protease and influence of bacteria on its activity (in cattle). *Jpn. J. Zootech. Sci*., 1987, **58**, 72–79.
6. Jouany, J.-P., Effect of rumen protozoa on nitrogen utilization by ruminants: altering ruminal nitrogen metabolism to improve protein utilization. *J. Nutr.*, 1996, **126**, 1335S–1346S.
7. Rønn, R., McCaig, A. E., Griffiths, B. S. and Prosser, J. I., Impact of protozoan grazing on bacterial community structure in soil microcosms. *Appl. Environ. Microbiol*., 2002, **68**, 6094–6105.
8. Schönhusen, U., Zitnan, R., Kuhla, S., Jentsch, W., Derno, M. and Voigt, J., Effects of protozoa on methane production in rumen and hindgut of calves around time of weaning. *Arch. Anim. Nutr*., 2003, **57**, 279–295.
9. Ozutsumi, Y., Tajima, K., Takenaka, A. and Itabashi, H., The effect of protozoa on the composition of rumen bacteria in cattle using 16s rrna gene clone libraries. *Biosci. Biotech. Biochem.*, 2005, **69**, 499–506.
10. Mould, D. and Thomas, G. J., The enzymic degradation of starch by holotrich protozoa from sheep rumen. *Biochem. J*., 1958, **69**, 327.
11. Williams, A. G. and Coleman, G. S., *The Rumen Protozoa*, Springer-Verlag, 1992.
12. Lee, S., Ha, J. and Cheng, K.-J., Relative contributions of bacteria, protozoa, and fungi to *in vitro* degradation of orchard grass cell walls and their interactions. *Appl. Environ. Microbiol*., 2000, **66**, 3807–3813.

13. Dehority, B. A., *Rumen Microbiology*, Nottingham University Press, Nottingham, 2003.
14. Karnati, S., Yu, Z., Sylvester, J., Dehority, B., Morrison, M. and Firkins, J., Technical note: specific PCR amplification of protozoal 18s RDNA sequences from DNA extracted from ruminal samples of cows. *J. Anim. Sci*., 2003, **81**, 812–815.
15. Regensbogenova, M. *et al.*, Assessment of ciliates in the sheep rumen by dgge. *Lett. Appl. Microbiol*., 2004, **39**, 144–147.
16. Shin, E. *et al.*, Phylogenetic analysis of protozoa in the rumen contents of cow based on the 18s rdna sequences. *J. Appl. Microbiol*., 2004, **97**, 378–383.
17. Sylvester, J. T., Karnati, S. K., Yu, Z., Morrison, M. and Firkins, J. L., Development of an assay to quantify rumen ciliate protozoal biomass in cows using real-time PCR. *J. Nutr*., 2004, **134**, 3378–3384.
18. Sylvester, J., Karnati, S., Yu, Z., Newbold, C. J. and Firkins, J., Evaluation of a real-time PCR assay quantifying the ruminal pool size and duodenal flow of protozoal nitrogen. *J. Dairy Sci*., 2005, **88**, 2083–2095.
19. Skillman, L. C., Toovey, A. F., Williams, A. J. and Wright, A.-D. G., Development and validation of a real-time PCR method to quantify rumen protozoa and examination of variability between entodinium populations in sheep offered a hay-based diet. *Appl. Environ. Microbiol*., 2006, **72**, 200–206.
20. Williams, A. G. and Strachan, N. H., The distribution of polysaccharide-degrading enzymes in the bovine rumen digesta ecosystem. *Curr. Microbiol*., 1984, **10**, 215–220.
21. Overbeek, R. *et al.*, The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucl. Acids Res*., 2005, **33**, 5691–5702.
22. Williams, A. G. and Coleman, G., The Rumen microbial ecosystem. In *The Rumen Protozoa*, Springer, 1997, pp. 73–139.
23. Singh, K., Pandya, P., Tripathi, A., Patel, G., Parnerkar, S., Kothari, R. and Joshi, C., Molecular diversity of protozoa in rumen of Indian buffalo (*Bubalus bubalis*). *Agric. Res*., 2013, **2**, 360–366.
24. Diez, B., Pedros-Alio, C., Marsh, T. L. and Massana, R., Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl. Environ. Microbiol*., 2001, **67**, 2942–2951.
25. Marie, D., Zhu, F., Balagué, V., Ras, J. and Vaulot, D., Eukaryotic picoplankton communities of the mediterranean sea in summer assessed by molecular approaches (DGGE, TTGE, QPCR). *FEMS Microbiol. Ecol*., 2006, **55**, 403–415.
26. Lepère, C., Boucher, D., Jardillier, L., Domaizon, I. and Debroas, D., Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (Lake Pavin). *Appl. Environ. Microbiol*., 2006, **72**, 2971–2981.
27. Martín, H. G. *et al.*, Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotech*., 2006, **24**, 1263–1269.
28. Mou, X., Sun, S., Edwards, R. A., Hodson, R. E. and Moran, M. A., Bacterial carbon processing by generalist species in the coastal ocean. *Nature*, 2008, **451**, 708–711.
29. Simon, C., Wiezer, A., Strittmatter, A. W. and Daniel, R., Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl. Environ. Microbiol*., 2009, **75**, 7519–7526.
30. Qin, J. *et al.*, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 2010, **464**, 59–65.
31. Singh, K. M. *et al.*, Metagenomic analysis of Surti buffalo (*Bubalus bubalis*) rumen: a preliminary study. *Mol. Biol. Rep*., 2012, **39**, 4841–4848.
32. Xie, G., Chain, P., Lo, C. C., Liu, K. L., Gans, J., Merritt, J. and Qi, F., Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing. *Mol. Oral. Microbiol*., 2010, **25**, 391–405.
33. Meyer, F. *et al.*, The metagenomics rast server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinf.*, 2008, **9**, 386.
34. Faichney, G., Poncet, C., Lassalas, B., Jouany, J., Millet, L., Doré, J. and Brownlee, A., Effect of concentrates in a hay diet on the contribution of anaerobic fungi, protozoa and bacteria to nitrogen in rumen and duodenal digesta in sheep. *Anim. Feed Sci. Technol*., 1997, **64**, 193–213.
35. Krause, D., Smith, W., Ryan, F., Mackie, R. and McSweeney, C., Use of 16S-rRNA based techniques to investigate the ecological succession of microbial populations in the immature lamb rumen: Tracking of a specific strain of inoculated Ruminococcus and interactions with other microbial populations *in vivo*. *Microb. Ecol*., 1999, **38**, 365–376.
36. Nathani, N. M., Patel, A. K., Dhamannapatil, P. S., Kothari, R. K., Singh, K. M. and Joshi, C. G., Comparative evaluation of rumen metagenome community using qPCR and MG-RAST. *AMB Exp*., 2013, **3**, 1–8.
37. Singh, K., Tripathi, A., Pandya, P., Rank, D., Kothari, R. and Joshi, C., Dasytricha dominance in Surti buffalo rumen revealed by 18S rRNA sequences and real-time PCR assay. *Curr. Microbiol*., 2011, **63**, 281–288.
38. Rossi, M. F., Martinele, I. and D'Agosto, M., Quantitative and differential analysis of ciliate protozoa in rumen content samples filtered before and after fixation. *Rev. Bras. Zooteen*, 2013, **42**, 831–834.
39. Gong, J., Dong, J., Liu, X. and Massana, R., Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist*, 2013, **164**, 369–379.
40. Sharp, R., Ziemer, C. J., Stern, M. D. and Stahl, D. A., Taxon-specific associations between protozoal and methanogen populations in the rumen and a model rumen system. *FEMS Microbiol. Ecol*., 1998, **26**, 71–78.
41. Vogels, G. D., Hoppe, W. F. and Stumm, C. K., Association of methanogenic bacteria with rumen ciliates. *Appl. Environ. Microbiol*., 1980, **40**, 608–612.
42. Wolin, M., Metabolic interactions among intestinal microorganisms. *Am. J. Clin. Nutr*., 1974, **27**, 1320–1328.
43. Santos, P. *et al.*, Gamma-proteobacteria *Aquicella lusitana* gen. Nov., sp. Nov., and *Aquicella siphonis* sp. Nov. infect protozoa and require activated charcoal for growth in laboratory media. *Appl. Environ. Microbiol*., 2003, **69**, 6533–6540.
44. Görtz, H.-D. and Brigge, T., Intracellular bacteria in protozoa. *Naturwissenschaften*, 1998, **85**, 359–368.
45. Coleman, G., The cellulase content of 15 species of entodiniomorphid protozoa, mixed bacteria and plant debris isolated from the ovine rumen. *J. Agric. Sci*., 1985, **104**, 349–360.
46. Newbold, C., Griffin, P. and Wallace, R., Interactions between rumen bacteria and ciliate protozoa in their attachment to barley straw. *Lett. Appl. Microbiol*., 1989, **8**, 63–66.
47. Williams, A. and Withers, S. E., Effect of ciliate protozoa on the activity of polysaccharide-degrading enzymes and fibre breakdown in the rumen ecosystem. *J. Appl. Microbiol*., 1991, **70**, 144–155.

ACKNOWLEDGEMENTS. We thank the supporting staff at Department of Animal Nutrition, Anand Agricultural University, Anand for their support in sample collection. This work was supported by Niche Area of Excellence project funded by Indian Council of Agricultural Research, New Delhi.

Received 6 October 2015; revised accepted 4 December 2015

doi: 10.18520/cs/v111/i1/184-191