The Pharma Innovation Journal 2022; SP-11(10): 310-315

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(10): 310-315

www.thepharmajournal.com
Received: 10-08-2022
Accepted: 13-09-2022

Amitoz Kaur
Department of Veterinary Microbiology, SKUAST Jammu RS Pura, Jammu & Kashmir, India

Gurdeep Singh
Department of Animal Genetics & Breeding, SKUAST Jammu RS Pura, Jammu & Kashmir, India

Anil Taku
Department of Veterinary Microbiology, SKUAST Jammu RS Pura, Jammu & Kashmir, India

Altaf Bhatt
Department of Veterinary Microbiology, SKUAST Jammu RS Pura, Jammu & Kashmir, India

Rajni Sharma
Department of Veterinary Microbiology, SKUAST Jammu RS Pura, Jammu & Kashmir, India

Corresponding Author:
Amitoz Kaur
Department of Veterinary Microbiology, SKUAST Jammu RS Pura, Jammu & Kashmir, India

Exploration of microbiome in the respiratory tract of sheep using whole metagenomic sequencing

Amitoz Kaur, Gurdeep Singh, Anil Taku, Altaf Bhatt and Rajni Sharma

Abstract
Sheep is an important species of livestock that has an immense potential to provide livelihood to a large proportion of small and marginal landless farmers, but respiratory diseases are one of the major hurdles in development of this sector. This study was conducted to explore the microbial diversity respiratory tract of apparently healthy sheep through metagenomic approach using Illumina Hiseq platform. The analysis revealed a total of 11, 142, 309 reads with sequence length of 250 bp having GC content of 46.21%. On the basis of comparative taxonomic abundance 37154 contigs belonged to Phylum *Proteobacteria* followed by *Chordata.* Analysis revealed majority sequences were belonging to bacteria and only 3% belonged to Viruses. The study revealed presence of a rich and very diverse microbial flora in the lower respiratory tract of sheep and is probably the first comprehensive study on sheep respiratory tract in India that can serve as a benchmark for further research in respiratory diseases of sheep.

Keywords: Metagenome, contigs, microbiota, phylum, reads

Introduction
Sheep is an important species of livestock that contributes to the agrarian economy of India by providing livelihood to a large proportion of small and marginal landless farmers. Respiratory diseases are one of the major constraints of sheep rearing. It represents almost 5.6% of all the diseases in Sheep (Hindson *et al*, 2002) [2] causing considerable financial losses to the farmers. Keeping in view of the multifactorial nature of respiratory diseases, survey of indigenous microbial populations (microbiota) is emerging as one of the most important aspect of animal health. For optimizing livestock productivity, it is mandatory that small ruminant rearers and health providers to realize the importance of identification and significance of indigenous respiratory flora in health as well as in diseased conditions [Maier *et al*, 2008] [4]. This will help further in reduction of current on-farm vulnerabilities, and increase in animal protein in the country.

Despite being accurate and gold standard procedure, only small number of organisms (< 1%) have been identified using prior knowledge of standard microbiological techniques and genome based identification procedures (Schuster, 2008) [5]. The advanced culture independent sequencing techniques such as pyrosequencing, has proved to be an effective tool for complete representation or identification of the resident microbiota. Furthermore, with the vast and ever-increasing amount of genomic information available in public databases, there is an immense potential for identifying novel pathogens or associating pathogens with clinical outcomes using these approaches. (Wylie *et al*., 2012) [6]. In case of sheep, recent studies have focused largely on the culture-independent identification of the bacterial community in the gastrointestinal tract. Nevertheless, there is a paucity of information regarding the culture-independent analysis of respiratory microbiota. Obtaining basic information on respiratory microbiota of sheep can ultimately be used to devise health and disease control strategies. This study is important with the point of view of minimizing the cost of healthy sheep production by reducing economic loses on respiratory diseases, resulting in establishment of more sheep farms that creates employment opportunity and will increase animal protein production.

Thus, the study is devised with an objective of performing whole metagenomic studies of respiratory tract of apparently healthy sheep using Ilumina Hiseq sequencing to reveal the micro-communities present.

Materials and Methods
Sample Collection and Processing
For studying the respiratory tract microbiome Broncheoalvoelar lavage fluid (BAL) and deep

nasal swabs were collected from two apparently healthy Sheep kept under good husbandry practices at sheep farm of Faculty of veterinary sciences and Animal husbandry, SKUAST-Jammu, R.S. Pura Jammu. The collection of Bronchoalveolar lavage fluid was done as per the protocol prescribed by (Katsoulos DP *et al*, 2009) [3] with few modifications. Sheep were anesthetized by intramuscular administration of Xylazine Hydochloride (1mg/kg bd wt) and an endotracheal tube was inserted through a tracheostomy incision, and a flexible Polyvinyl tubing was inserted through it and gently wedged into a lower lobe bronchus. Three 50-ml aliquots of sterile Normal saline (0.9% NaCl) were slowly infused and aspirated back into the injecting syringe after a dwell time of less than 5 seconds. The recovered lavage fluid was placed at 4º C until further processing. Deep nasal samples were collected by inserting sterile cotton-tipped applicator sticks into the nasal passage after proper cleaning and disinfecting the external nares and transported in a cool box to the laboratory for analysis. These swabs tips were carefully cut and put into normal saline solution for DNA extraction.

DNA was isolated using MOBIO power soil Kit (Qiagen) as per the protocol described by (Glendinning *et al*, 2016) [1] with only initial step modification. The bead tubes were not heated with Fastprep FP120 cell disrupter due to unavailability of the solution. Both lavage samples and nasal swabs were processed in duplicates and both samples were pooled as a single sample to increase the overall yield of the genomic DNA (gDNA).

Metagenomic sequencing

Whole metagenome sequencing of the DNA sample was performed using the Illumina HiSeq 2500 sequencer (Illumina, USA). For sequencing, a dual-indexed Paired-End sequencing (2 × 251 base pairs) strategy with a total of 250 cycles, six bp index sequence was used as shown in Table 1. The entire sequencing run was completed in 39 hours. The sequencing steps that were performed for analysis of the data are shown in Figure 1. From the fastq file, the read quality check included Base quality score distribution, sequence quality score distribution, average base content per read, GC distribution, PCR amplification, adapter contamination. Sequence reads were trimmed to retain high quality sequence for further analysis and low-quality sequence reads were excluded from the analysis. The adapter trimming was performed using Cut adapt (version - 1.8.1) software.

Table 1: Whole metagenomic analysis summary

Sample	Pooled BAL (Broncho alveolar lavage fluid) and nasal swabs of apparently healthy Sheep
Sequencing Platform	Illumina HiSeq
Library type	Paired End (250 bp x 2)
Type	*De novo* whole genome metagenomics

The *de novo* assembly of the adapter trimmed fastq files was carried out using MetaSPAdes (v 3.10.1). Further, contigs were linked by the assembly algorithm to create scaffolds and bad or mis assemblies were removed from the result. Assembly was performed with default k-mer length using de-bruijn graph method and In-house PERL (Practical Extraction and Report Language) codes were used to parse the fasta files for the downstream analysis.

Functional Annotation

As a result of *de novo* assembly, contigs in FASTA format were generated for the sample, while the contigs obtained from the assembly were used as input to Meta Gene Annotator (MGA) for the prediction of open reading frames (ORFs). The gene-finding program MGA was used to identify the coding regions and distinguish them from noncoding DNA. The ORFs were then filtered and taken for the functional annotation and taxonomic classification. For Protein Homology Search, the predicted ORFs were searched against the non-redundant (NCBI-nr) database using DIAMOND (v0.7.9.58) programme which is high-throughput program for aligning a file of short reads against a protein reference database such as NR, at 20,000 times the speed of BLASTX, with high sensitivity.

Analysis of dominant population

Rarefaction curve was generated by comparing the species abundance between the samples based on number of leaves in the taxonomy and number of sequence occurred. The curve is made for all taxa including Bacteria, Archaea, Eukaryote, Viruses, unclassified and other sequences.

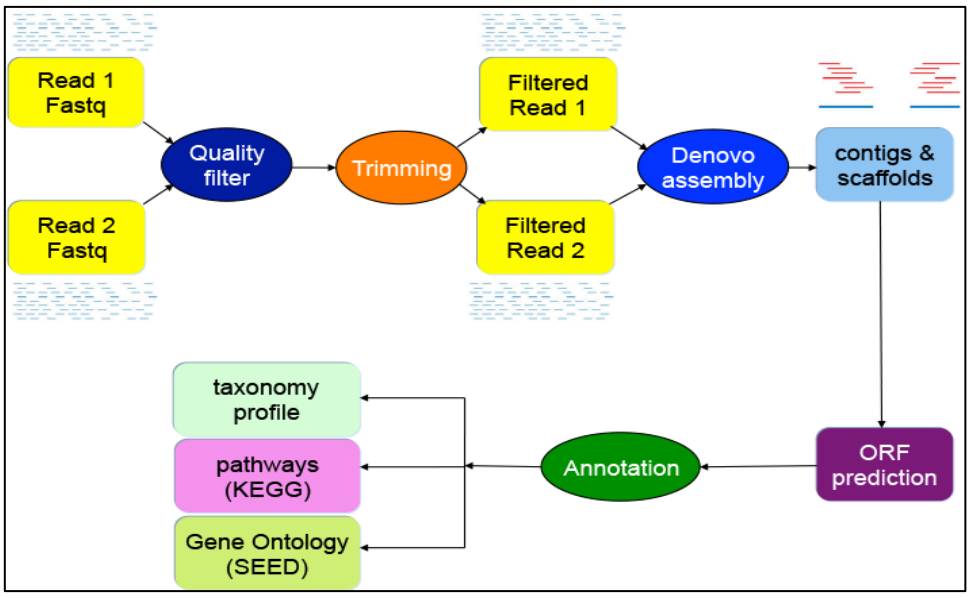

Fig 1: De novo metagenomics pipeline

Results & Discussions

Metagenomic analysis of respiratory tract of sheep revealed a total reads of 11,142,309 from sequencer with the sequence length of 250 bp as shown in *Table 2*. 84.4% reads were having phred score over Q30.

Table 2: Raw read summary

Sample ID	Total Reads	Sequence Length(bp)	Total Data (GB)	% of GC	Average Base Quality	Q0-Q10	Q10-Q20	Q20-Q30	>=Q30
BAL-sample	11,142,309	250	5.57	46.21	34.95	.01	9.55	5.84	84.64

The base quality of left and right end of the paired-end read sequence of the sample is shown in the bar-plot (Figure 2 - 3) respectively. Regarding the base composition of nucleotides in the sequence read for the sample is calculated from the quality checking. The x-axis represents sequencing cycle and y-axis represents nucleotide percentage. The base composition of left and right end of the paired-end read sequence is shown in Figure 4-5. On an average, the majority of the reads in sample have GC content is 40-70% (Figure 6).

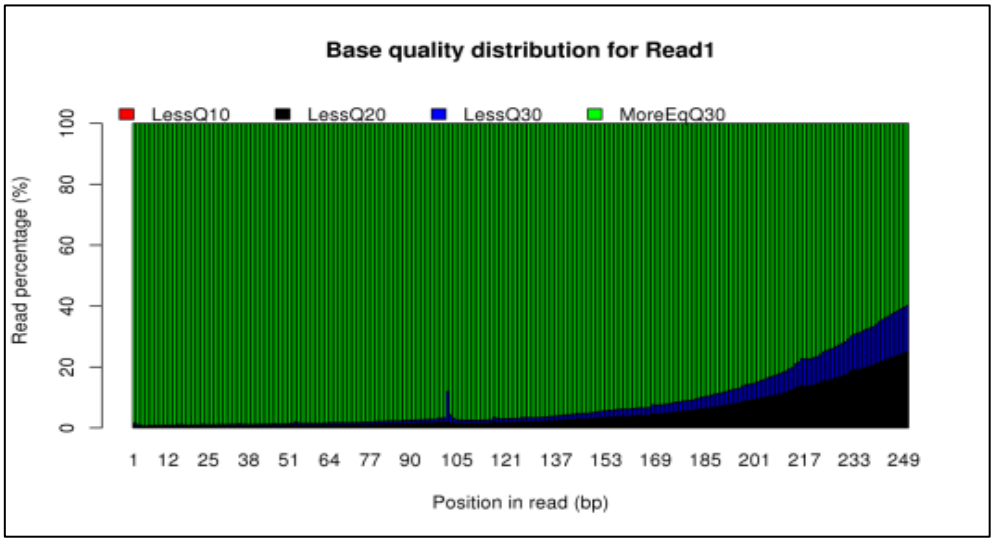

Fig 2: Base quality distribution of left end of paired-end read BAL-Sample (R1). The x-axis represents sequencing cycle and y-axis represents percentage of total reads

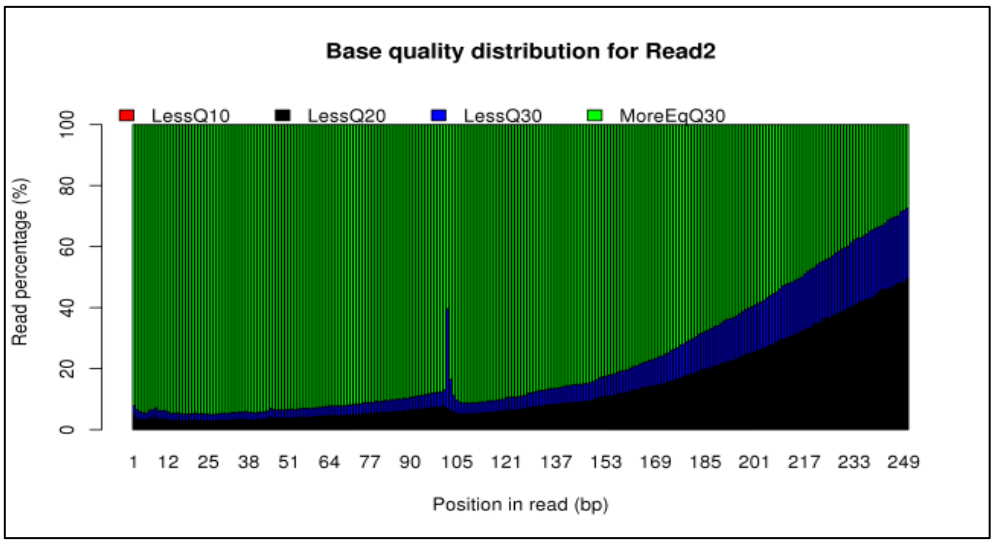

Fig 3: Base quality distribution of right end of paired-end read BAL-Sample (R2)

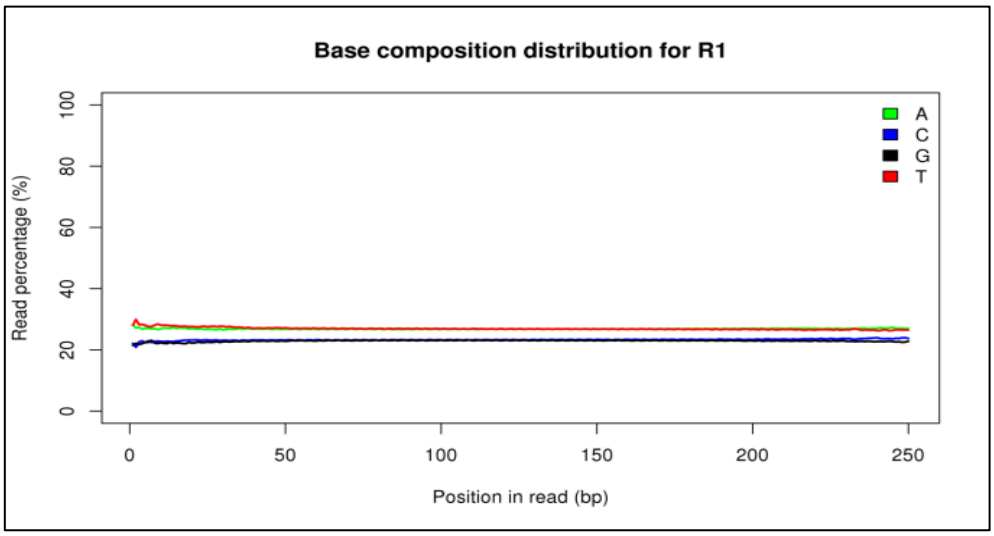

Fig 4: Base composition in left end of paired-end read BAL-Sample (R1)

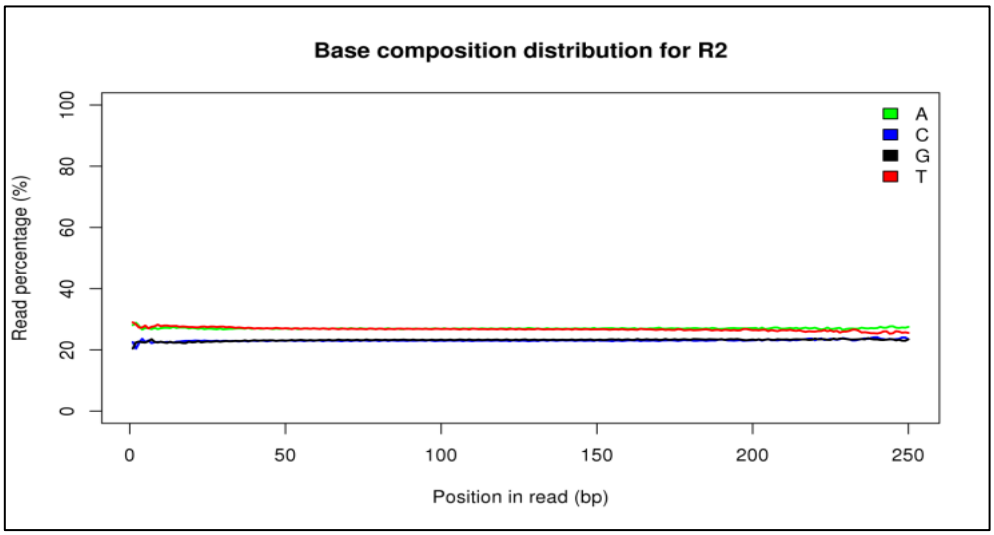

Fig 5: Base composition in right end of paired-end read BAL-Sample (R2)

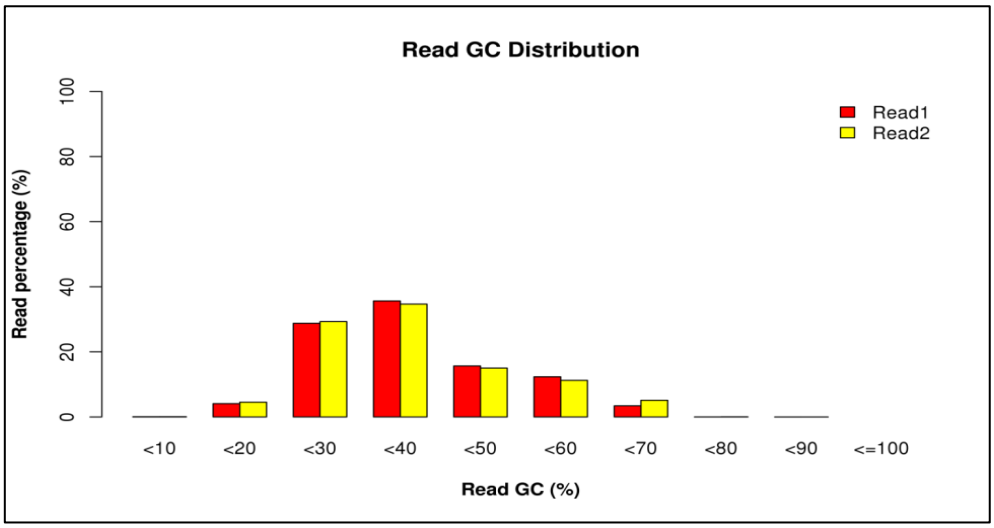

Fig 6: GC distribution of read sequence for BAL-Sample (The x-axis represents average GC content in the sequence and y-axis represents percentage of sequences)

De novo Metagenome Analysis

De novo metagenome assembly was carried out for the sample by assembling contigs from the reads using MetaSPAdes program. A total of 2,025,661 contigs were obtained, out of which 1,052,220 had contig length less than 150. 757,345 and contigs were having contig length between 150-500 whereas 216,096 were having contig length greater than 500. The length-wise distribution of the total contigs obtained from assembly is illustrated in Figure. 7. For further downstream analysis, 2,025,661 contigs were selected.

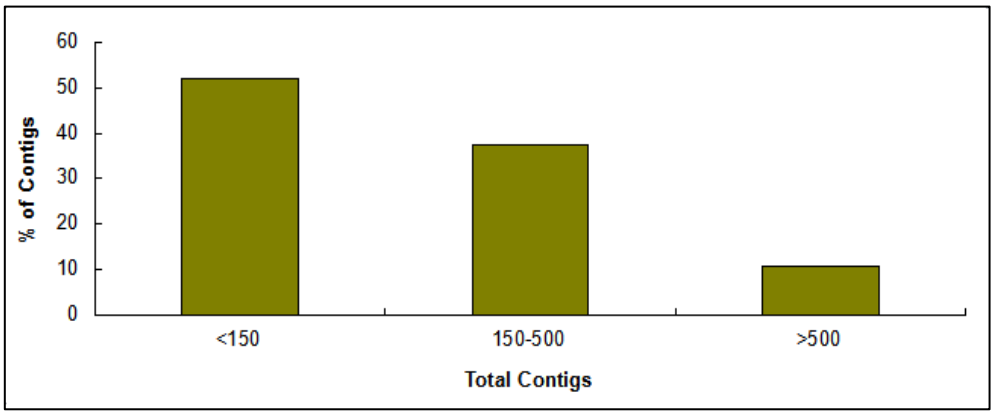

Fig 7: Assembled Contig Length Distribution of Sample BAL-Sample

Analysis of dominant population

The rarefaction curve was made for all taxa include *Bacteria*, *Archaea*, *Eukaryote*, *Viruses*, unclassified and other sequences. The plot (Figure 8) shows the taxanomic richness observed in which the curve length difference (shown in X-axis of plot) occurs between sample due to variation in the number of sequence between samples whereas Y-axis shows the number of leaves in taxon tree between samples.

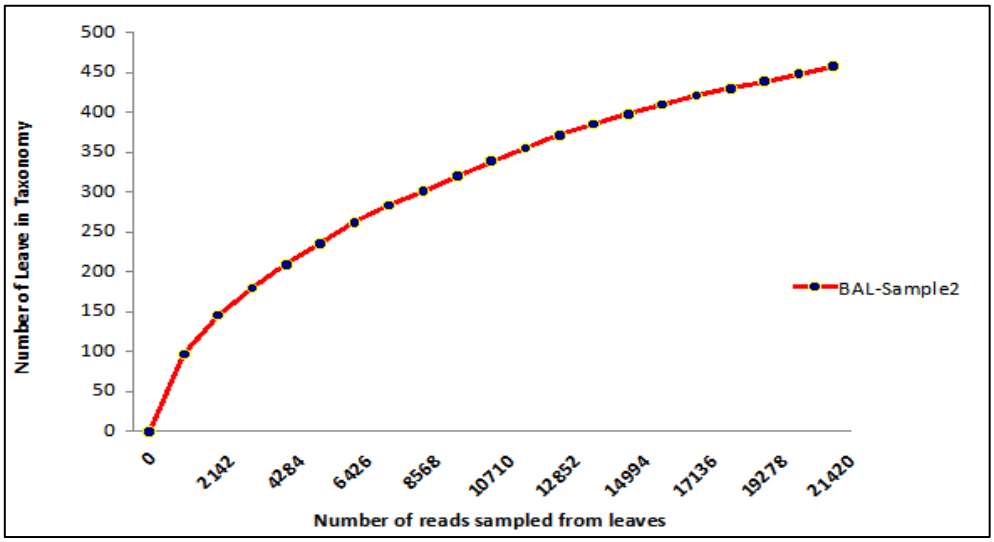

Fig 8: Rarefaction curves created in MEGAN. Rarefaction analysis was performed for all the samples at the most resolved taxonomic level of the NCBI taxonomy in MEGAN.

Taxonomy Analysis

Taxonomic profiling for all the metagenomics sample was performed using NCBI taxonomy data sets. The taxonomy tree was generated based on neighbour-joining method using MEGAN software.

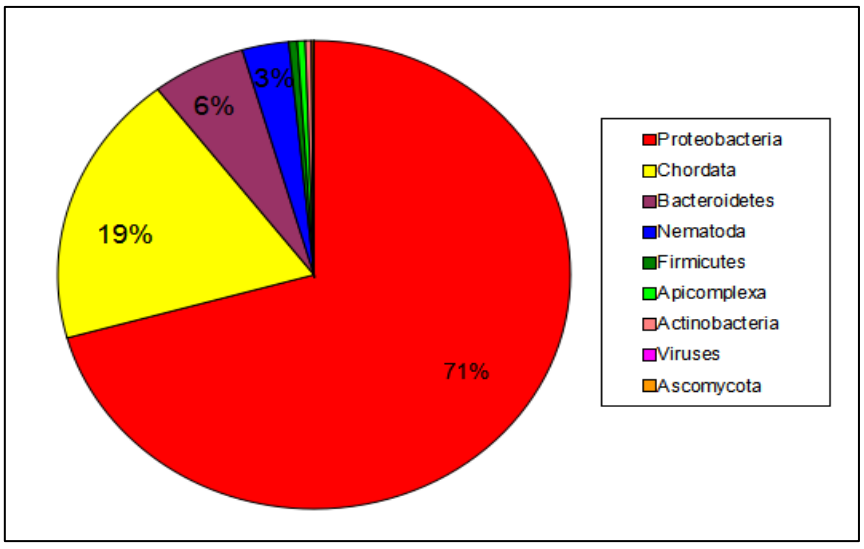

Fig 9: Phylum abundance of the BAL-Sample (only top phylum hits were highlighted)

Top phylum that were isolated included *Proteobacteria* constituting 71% of total reads while phylum Chordata constituted 19% of the reads. Phylum Bacteroidetes and Nematoda constituted about 6% and 3% respectively (Fig 9). Remaining Phylum that were detected by using metagenomics were Phylum *Firmicutes*, *Apicomplexa*, *Actinobacteria*, *Viruses*, *Ascomycota*, *Streptophyta*, *Verrucomicrobia*, *Cyanobacteria*, *Deinococcus-Thermus*, *Acidobacteria*, *Fusobacteria*, *Planctomycetes*, *Spirochaetes*,, *Basidiomycota*.. Analysis revealed that about 77.6% of total reads belonged to bacteria while viruses and fungus contributed .08% and .01% of total reads in phylum classification. Parasites and protozoa also contributed about 22.31% in phylum reads. Other sequences contributed about .02% to total phylum reads obtained.

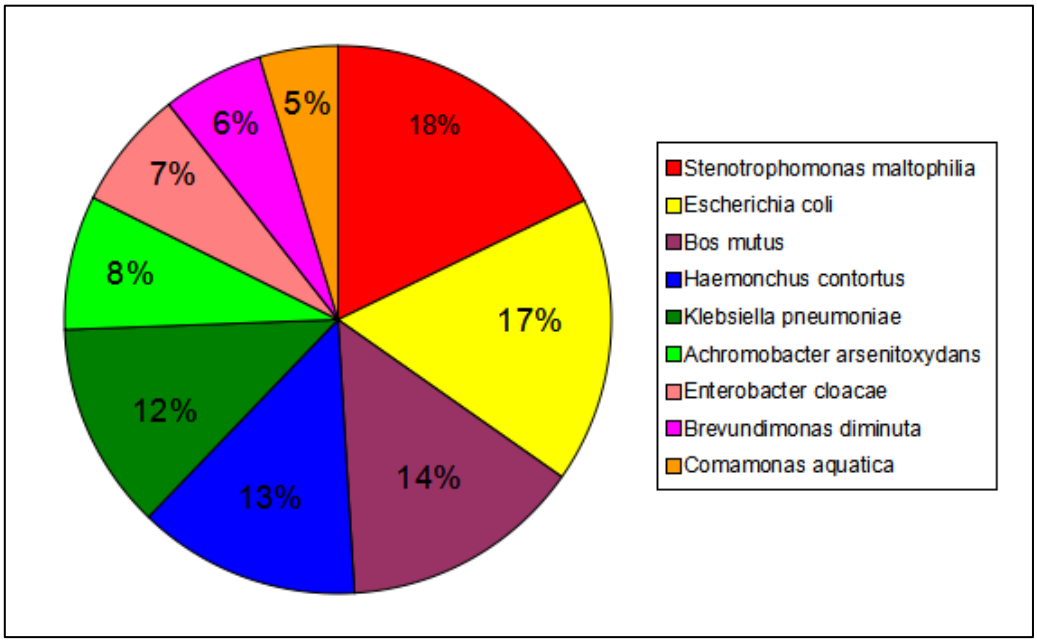

Fig 10: Species abundance of the sample BAL-Sample (only top species hits were highlighted)

A total of 599 species were obtained from the respiratory tract of sheep using whole metagenomics. Out of these top species hits are shown in Figure. 10. Taxonomic and functional diversity of a community quantified using whole metagenome shotgun sequencing revealed the dominance of bacterial population and their metabolisms. At the phylum level dominant bacterial phyla were *Actinobacteria*, *Bacteriodes*, *Deinococcus Thermus*, *Firmicutes*, and *Planctomycetes*. Bacterial genera like *Rhodococcus*, *Microbacterium*, *propionibacterium*, *Flavobacterium*, *Deinococcus*, *Caulobacter*, *Brevundimonas*, *Methylobacterium*, *Paracoccus*, *Roseomonas*, *Novosphingobium*, *Sphingomonas*, *Achromobacter*, *Acidovorax*, and *Aquabacterium* were also dominant. A number of DNA sequences remained unassigned with respect to taxonomic and functional coherence. To the best of our knowledge, this is the first study that deals with the description of complete profiling of microbial diversity from respiratory tract of sheep using next generation.

Conclusion

The study revealed very diverse microbial flora in the lower respiratory tract of sheep that were not previously identified by various cultivable methods while lungs were considered as sterile environment and thus this can add boon to further research in respiratory diseases of sheep and other species. Metagenome sequencing analysis may significantly provide important breakthroughs in depicting taxonomic structure and functional and/or metabolic pathways of sheep microbiome with the promise of novel genes and novel microbes for biotechnological applications.

References

1. Glendinning L, Steven W, Jolinda P, Peter T, David C, Gerry M. Variability of the Sheep Lung Microbiota. App & Env micro. 2016;82(11):3226-3237.
2. Hindson JC, Winter, AC. Respiratory disease. Manual of Sheep Disease. 2002;2:196-209.
3. Katsoulos P, Christodoulopoulos D, Georgios K, Georgios MA. Leukocyte counts in bronchoalveolar lavage fluid obtained from normal and Maedi-Visna-infected sheep. Vet Clinic. Path. 2009;6:0275-6382.
4. Maier RM, Pepper IL, Gerb CP. The Environmental Microbiology, Academic Press; c2009.
5. Schuster SC. Next-generation sequencing transforms today's biology. Nat. Meth. 2008;5:16-18.
6. Wylie KM, Truty RM, Sharpton TJ, Mihindukulasuriya JA, Zhou Y, Gao H, *et al*. Novel bacterial taxa in the human microbiome. Pub. Lib. of sci. one. 2012;7:e35294.