The Indian Journal of Veterinary Sciences & Biotechnology ***(2018) Volume 13, Issue 4, 74-77***
ISSN (Print) : 2394-0247 : ISSN (Print and online) : 2395-1176, abbreviated as IJVSBT
10.21887/ijvsbt.v13i4.11564

Submitted : 20-01-2018 ***Accepted : 22-02-2018*** ***Published : 08-04-2018***

Viral Metagenomics of chickens with respiratory infection using MG-RAST

Manisha R. Sajnani[1], D. Sudarsanam[2], Subhash J. Jakhesara, Siddhardha Solosanc[3], Chaitanya G. Joshi and Vaibhav D. Bhatt[4*]

Department of Animal Biotechnology, Anand Agricultural University, Anand, Gujarat

[1]Bharathiar University, Coimbatore, Tamilnadu, [2]Loyola College, Chennai,

[3]Madras Medical College, Chennai, [4]Saurashtra University, Rajkot, Gujarat, India

Corresponding Author: bhatt_vbhv@yahoo.co.in

Abstract

Respiratory tract infections in poultry flocks lead to increased rate of mortality and thereby increase economic burden to the poultry industry. The aim of the study was to identify variety of known and unknown microbial organisms participating in the respiratory tract infections in poultry. Infected samples from 8 different poultry farms located in the Anand district of Gujarat, India were collected. Using Next Generation Sequencing (NGS) platform Ion Torrent PGM, approximately 1.1 giga-bases of sequences by metagenomics pipeline were obtained. To understand the community of bacterial as well as viral sequences involved in the infection, MG-RAST, an automated online based approach was used. The taxonomical classfication resulted in sequences with 78.91% of bacteria, 9.35% of eukaryota, 8.17% of viruses, 0.10% of archaea, 0.02% of unclassified sequences, 0.45% of other sequences and 2.99% were unannotated. While functional classification of sequences resulted in 22.15% of predicted proteins with known functions, 77.74% of predicted proteins with unknown function, and very negligible 0.01% of ribosomal RNA genes. The most abundant virus families observed were Podoviridae, Myoviridae and Siphoviridae belonging to the Caudovirales order. This study is the first approach for annotating metagenomics of poultry having respiratory infection using MG-RAST approach.

Key words: Respiratory tract infection, Poultry, Metagenomics, NGS, MG-RAST

Introduction

Diseases of the respiratory tract are a significant component of the overall disease incidence in poultry. Respiratory tract infections are leading causes of morbidity in poultry farms all over the world that is the reason for the economic loss in the poultry industry (Sid *et al*., 2015). In many cases, respiratory disease observed in a flock may be a component of a multi-systemic and multi-etiology disease. It may be the predominant disease with lesser involvement of other organ systems and many a times caused by a group of pathogens (co-infection), rather than a single causative pathogen species. Various pathogens may initiate respiratory disease in poultry, including a variety of viruses, bacteria, and fungi. Environmental factors may provide suitable environment with genetics

of host to result in effective virulence, clinical signs and lesions (Garcia, 2017). To understand the diversity of microbial communities present in respiratory tract in poultry as well as their functions associated with the disease, this metagenome sequencing of DNA viruses from eight infected broilers farms of Gujarat state was attempted for the first time using MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST, Meyer *et al.,* 2008) in assigning sequence reads to microbes at different phylogenetic levels.

Materials and Methods

Eight tissue samples of trachea and nasal swab of the poultry affected with respiratory infections from eight different poultry farms in Gujarat were collected. A library preparation as well as Next Generation Sequencing using Ion Torrent PGM platform (318 chip with 400bp chemistry) on the DNA Virome sequences was performed to study the diversity and function of microbial organisms causing infections. The raw reads were checked and filtered for its quality using PRINSEQ (Schmieder and Edwards, 2011) and host specific sequences were screened using Bowtie2 tool (Langmead and Salzberg, 2012). Further the screened reads were assembled using SPAdes Assembler (Bankevich *et al*., 2012) showing optimum of 55 Kmer size. High quality contigs of DNA Virome datasets were used for our studies and uploaded them to the MG RAST server for automated analysis. The methodology is described in detail earlier (Sajnani *et al*., 2018).

Default parameters for the analysis were applied. The MG-RAST server followed following steps of analysis: 1. Initial sequence statistics calculation, 2. Adapter Trimming, 3. Denoising and normalization, 4. Removal of sequencing artifacts, 5. Host DNA contamination removal, 6. RNA feature identification, 7. RNA clustering, 8. RNA similarity search, 9. Identifying putative protein coding features, 10. Filter putative protein features overlapping rRNA features, 11. Amino acid sequence clustering, 12. Protein similarity search, 13. Protein similarity annotation, 14. RNA similarity annotation, 15. Merge and index similarities, 16. Annotate and index similarities, 17. Feature abundance profile, 18. LCA abundance profile, 19. Data source abundance profile, 20. Extract features with no similarity hits, 21. Abundance profile load, 22. Abundance profile build and load, and 23. Summary statistics.

The raw sequencing data is available in the Sequence Read Archive (SRA) of NCBI under the BioProject No. PRJNA322592 and Accession No. MAUZ00000000, MAVA00000000, MAVB00000000, MAVC00000000, MAVD00000000, MAVE00000000, MAVF00000000, MAVG00000000 (https://www.ncbi.nlm.nih.gov/bioproject/?term= PRJNA322592).

Results and Discussion

MG-RAST is freely available tools for metagenomic analyses that have been used in a wide range of studies. In the present study, MG-RAST server assisted in the analysis of virome datasets at the microbial level predicting bacterial and viruses associated with respiratory disease. It not only helps in analyzing contigs but also raw reads with .fna or .fasta files allowing quality files for better analysis.

The datasets uploaded at MG-RAST contained 33,652 sequences totaling 8,913,656 base pairs with an average length of 265 bps. The mean GC percent was 45 ± 10, Mean sequence length was 265 ± 487 bp, of the total sequences, 100% sequences passed the QC pipeline as well as artificial duplicate reads test. Amongst high quality sequences, 24 sequences (0.7%) contained ribosomal RNA genes, 4,944 sequences (22.15%) contained predicted proteins with known functions, and 17,352 sequences (77.74%) contain predicted proteins with unknown function.

The distribution of taxonomic categories at domain levels using protein similarities to entries in the 'refseq' protein databases were bacteria 49,881 (78.91%), eukaryota - 5,913 (9.35%), viruses - 5,167 (8.17%), other sequences - 286 (0.45%), archaea - 64 (0.10%), unclassified sequences - 10 (0.02%) and unannotated - 1,890 (2.99%). The first ten families showing highest abundance

were Enterobacteriaceae - 23,805, Staphylococcaceae - 7,497 (11.85%), Lactobacillaceae - 4,932 (7.80%), Phasianidae - 2,889 (4.57%), Pseudomonadaceae - 2,587 (4.09%), Moraxellaceae - 2,449 (3.87%), Siphoviridae - 2,046 (3.23%), Bacillaceae - 1,492 (2.36%) and Streptococcaceae - 1,169 (1.85%) whereas, 1934 (1.85%) of sequences remained unannotated.

At the species level, *Escherichia coli* with 689 hits, *Staphylococcus epidermidis* with 181 hits, *Staphylococcus aureus* with 147 hits, *Lactobacillus helveticus* with 109 hits, *Lactobacillu sacidophilus* with 108 hits, *Acinetobacter baumannii* with 77 hits, *Staphylococcus haemolyticus* with 71 hits, *Lactobacillus reuteri* with 63 hits, *Pseudomonas aeruginosa* with 61 hits and *Salmonella enterica* with 60 hits were observed in the bacterial classification.

Among them, *Escherichia coli* and *Staphylococcus aureus* has been previously identified for the cause of respiratory infection (Ammar *et al.,* 2016). For the virus classification most of the sequences were observed in the Podoviridae, Myoviridae, Siphoviridae family belonging to the Caudovirales order consisting of phages namely Enterobacteria phage Phieco 32, Enterobacteria phage RB49, Staphylococcus phage 37, Enterobacteria phage RTP, Staphylococcus phage CNPH82, Staphylococcus phage 2638A, Enterobacteria phage JSE, Staphylococcus phage G1, Staphylococcus phage EW, Staphylococcus phage PH15 and others.

The functional analysis was done using COGs, KOs, NOGs, and Subsystems to categorise sequences for metabolism, information storage processing, cellular processing and signalling. The subsystem technology showed percentage and number distribution of sequences for phages, prophages, transposable elements, plasmids 1,632 (51.21%), clustering-based subsystems - 206 (6.46%), carbohydrates - 197 (6.18%), cell wall and capsule - 159 (4.99%), miscellaneous - 102 (3.20%), nucleosides and nucleotides - 95 (2.98%), protein metabolism - 85 (2.67%), amino acids and derivatives - 82 (2.57%), DNA metabolism - 78 (2.45%), regulation and cell signaling - 74 (2.32%), cofactors, vitamins, prosthetic groups, pigments - 71 (2.23%), membrane transport - 65 (2.04%), RNA metabolism - 58 (1.82%) and fatty acids, lipids, and isoprenoids - 55 (1.73%). This shows microbial community composition which helps to evaluate the microbial diversity.

Conclusions

The metagenomic analysis of the poultry virome using MG-RAST was successfully carried out to identify microbial species involved in the respiratory disease. We reported presence of new species both known and unknown, which was not reported by the previously used BLAST, Virome and Metavir approach. The data analysis tool work with different capabilities and pipelines, as integrated databases used for annotations differs from one tool to another. The annotation provided by MG-RAST enhanced sequences analysis to further help in discovering preventive measures of respiratory infections in poultry flocks.

Acknowledgement

This work was financially supported by Gujarat State Biotechnology Mission Government of Gujarat, India. We are thankful to Pathology Department for providing sample and OME Research Facility for providing the facility of the next generation sequencing at College of Veterinary Sciences, AAU, Anand, India.

Conflict of Interest: All authors express no conflict of interest

References:

Ammar, A.M., Abd El-Aziz, N.K., Abd El Wanis, S. and Bakry, N.R. (2016). Molecular versus conventional culture for detection of respiratory bacterial pathogens in poultry. *Cell Mol. Biol. (Noisy-le-grand)*, 62: 52-56.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A.

and Pevzner, P.A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*., 19: 455-477.

Garcia, M. (2017). Current and future vaccines and vaccination strategies against infectious laryngotracheitis (ILT) respiratory disease of poultry. *Vet. Microbiol*., 206: 157-162

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357-359.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. and Edwards, R.A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9: 386.

Sajnani, M.R., Sudarsanam, D., Pandit, R.J., Oza, T., Hinsu, A.T., Jakhesara, S.J., Solosanc, S., Joshi, C.G. and Bhatt, V.D. (2018). Metagenomic data of DNA viruses of poultry affected with respiratory tract infection. *Data in Brief,* 16: 157-160.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863-864.

Sid, H., Benachour, K. and Rautenschlein, S. (2015). Co-infection with multiple respiratory pathogens contributes to increased mortality rates in Algerian poultry flocks. *Avian Dis*., 59: 440-446.

❑