The Indian Journal of Veterinary Sciences & Biotechnology ***(2018) Volume 13, Issue 3, 60-66***

ISSN (Print) : 2394-0247 : ISSN (Print and online) : 2395-1176, abbreviated as IJVSBT

http://dx.doi.org/10.21887/ijvsbt.v13i3.10612

Submitted : 24-11-2017 ***Accepted : 22-12-2017*** ***Published : 09-01-2018***

DNA Virome: Sequencing and Data Analysis of Viral Metagenome of Poultry Suffering from Respiratory Diseases

Manisha R. Sajnani, D.Sudarsanam, Subhash J. Jakhesara, Siddhardha Solosanc,

Chaitanya G. Joshi and Vaibhav D. Bhatt*

Institute of Medical Microbiology, Madras Medical College, Chennai, India

Corresponding Author: bhatt_vbhv@yahoo.co.in

Abstract

Respiratory diseases are the most common causes of death in a poultry flock. Knowledge of the avian respiratory system is essential for developing a health monitoring plan for a poultry flock, recognizing problems that may occur, and taking action to correct them. Recently, respiratory diseases in commercial broiler chicken flocks have increased in India and no significant cause has been identified till date. Viral populations are predominant in respiratory tract infections and they cause severe economic loss through decreased productivity. We performed shotgun sequencing approach to understand viruses associated with or causing respiratory infections (RI) in broilers. We report high quality sequencing reads, de novo assembled sequences, taxonomical as well as functional classifications of virome of the infected broiler birds.

Key words: Respiratory diseases, poultry, de novo assembly shotgun sequencing

Introduction

Respiratory disease complex has emerged as a great challenge to poultry industry. The DNA viruses are Marek's disease (a gallid herpes virus 2, GaHV-2, MDV) (Wozniakowski *et al.,* 2015), Chicken Anaemia (CAV)(Sommer and Cardona, 2003) and Fowl pox virus (FPV)(Miller and Torchetti, 2014). These viruses spread and infect flock very rapidly and are relatively highly resistant to chemicals and environmental conditions. Retroviruses replication within chicken cells employs a rapid transition to DNA by a reverse transcriptase step, following which the viral genome is incorporated into the cellular genome and transmitted as DNA viruses. As these agents are predominant in respiratory tract infections, they continued to create economic burden to poultry industry by causing severe loss in production. Morbidity is typically 10-20%, whereas the mortality is 5-10% in poultry. It is presumed to be due to involvement of many unknown viruses and genetically novel strain of viruses.

The traditional culture methods are highly laborious and time consuming and fail to produce desirable results (multiplex PCR, DNA microarray, etc.) to identify avian viruses. It has many more limitations as it is not possible to detect multiple organisms at once in short time. With the advent of molecular biotechnology, we applied metagenomic sequencing of the next-generation sequencing technology to study avian viral microbiome. Metagenomics approach is a power full tool to study avian viral microbiome diversity in infectious respiratory tract. Moreover, functional metagenomics

can identify novel functional genes, microbial pathways, antibiotic resistance genes, functional infectious respiratory tract microbiome, and determine interactions and co-evolution between microbiota and host; there is no need to cultivate individual organisms, as in mixed communities, genomic analysis can be done to discover novel members of community. It is worthwhile to note that not a single survey is available to determine the whole viral population associated with respiratory infections. Therefore the present study was made as an attempt to study the viral metagenome population present in respiratory infections (RI) of poultry in India.

Materials and Methods

Sample Collection: In the present work, tracheal and nasal swabs of 34 broilers affected with respiratory diseases were collected in sterile tubes from 8 different poultry farms located in Gujarat, India (Table 1). The collected samples were then filtered through 0.2 ìm filter and filtrates were immediately stored at -80ºC. For purifying viruses associated with tracheal tissue, the tissue was processed via blending into a ~20% homogenate in sterile phosphate buffered saline (PBS) followed by centrifugation at 7500 rpm (5500 × G) for 15 min at 4°C. A stepwise filtration process involving 0.8 ìm and 0.45 ìm was used to remove eukaryotic and bacterial cells and nuclei. Virus-sized particles were pelleted by ultracentrifugation for 5 hr at 4°C (113,000 × G) using CsCl density gradient centrifugation.

DNA Isolation: DNA was extracted from each sample separately using extraction kit from Roche. Before processing DNA samples were pooled in equimolar concentration for each farm. The samples were amplified with whole genome amplification kit from Qiagen for DNA. The qualitative and quantitative evaluation of extracted DNA was done using Nanodrop 1000 UV-Vis Spectrophotomer as well as 0.7% agarose gel electrophoresis. For isolating DNA viruses, after homogenization samples were treated by DNase I for removal of host background DNA. Viral DNA was extracted using standard phenol:chloroform extraction method and amplified to increase quantity.

Library Preparation and Next Generation Sequencing: Amplified products were used further for library preparation as manufacturer's protocol. In brief, the samples were sonicated to generate fragments of 400–500 bp size range followed by end polishing and adaptor ligation to the free ends. The quality and average size of the library were accessed on the Agilent 2100 bioanalyzer with the DNA high sensitivity kit (Agilent Technologies, USA). Genomic libraries were clonally amplified, enriched and subjected to sequencing run using Ion Torrent PGM 316 Chip with 300 bp chemistry following the manufacturer's protocol. Individual farms samples were separated using molecular barcoding. Data were transferred to high end cluster having 2 TB RAM and 100 nodes for further analysis.

Quality Filtering of Data and Host Specific Screening: Raw reads were scanned with pred score >20 and read length >50 bp for quality filtering using PrinSeqtool (http://prinseq.sourceforge.net/) (Table 2). Host specific screening was performed by mapping against host genome of *Gallus gallus* from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/ Gallus_gallus/ and Assembly name: Gallus_gallus-5.0) using Bowtie2.2.8 (https:// sourceforge.net/projects/bowtie-bio/files/bowtie2/ 2.2.8/) with default parameter (Table 3). The same procedure was followed for each sample separately. Reads that mapped partially or completely with host genomic sequences were removed from further analysis. Unmapped reads (unaligned to host sequences) were considered as clean reads and used for assembly and downstream analysis. The raw sequencing data used to perform metagenomic analysis is available in the Sequence Read Archive (SRA) 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).

De novo Assembly of DNA Viromes: Assembly for each sample was performed separately using high quality reads using SPAdes assembler (http://bioinf.spbau.ru/en/spades3.7) based on multiple

k-mer (k-mer length 21, 33, and 55). The best assembly was obtained at 55 k-mer size that was used for downstream analysis. The detailed assembly statistics having number of contigs, total contig bases, N50 size, and GC % is provided in the Table 4.

Virome Analysis: Virome analysis was carried out by uploading contigs to the online web applications: Virome (http://virome.dbi.udel.edu/) and METAVIR (http://metavir-meb.univ-bpclermont.fr/index.php?page=Welcome) for taxonomical and functional classification as well as performing BLASTx using complete virus genome database from NCBI (ftp://ftp.ncbi.nlm.nih.gov/ genomes/ Viruses/) with query coverage >= 50 and e-value 0.0003.

Predominant Viruses: Highly abundant viruses were also predicted and validated for above analysis by using reference mapping approach on the high quality reads using virus genome database from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/viruses/). We used "map to read" CLC genomics workbench version 7.0.4.1 for predicting predominantly occurring viruses associated with the infection. The mapped reads with >90% coverage i.e. total 13686 blast yielded 547 virome and 558 metavir.

Results and Discussion

Ion Torrent Sequencing and Data Analysis

Respiratory disease in the poultry is known to be the reason for the heavy loss in the poultry industry. Various studies have been carried out worldwide to understand the cause of infections in the poultry, however there are more than one pathogens involved in causing respiratory diseases which makes it a complex to understand its etiology. These pathogens can cause disease individually or in combination with other pathogens. Mixed infections of different respiratory agents may occur because of extensive use of multiple live vaccines, high geographic population, and housing densities (Gowthaman *et al.,* 2014). Thus present work was carried out to investigate the

Table 1: Data showing number of samples collected from different farms of Gujarat State having severe infection

Accession No.	Poultry Farm (PF)	Village/ Town	Age of Poultry Bird	No. of Samples collected	Latitude and Longitude
MAUZ00000000	Unique PF	Mogri	40 days	4	22° 31' 39.1584" N 72° 55' 57.9684" E
MAVA00000000	Ketul PF	Bansal	20 days	5	30° 52' 35.6628" N 75° 51' 17.5284" E
MAVB00000000	Nasib PF	Godhra	19 days	2	22° 46' 43.6944" N 73° 36' 51.408" E
MAVC00000000	Panchmahal PF	Godhra	22 days	4	22° 46' 43.6944" N 73° 36' 51.408" E
MAVD00000000	Jay PF	Vaghasi	30 days	5	22° 32' 46.5756" N 72° 58' 58.8468" E
MAVE00000000	Faizan PF	Anklav	13 days	5	22° 22' 38.9136" N 72° 59' 57.1236" E
MAVF00000000	S. K. PF	Petlad	26 days	5	22° 29' 1.014" N 72° 48' 4.8636" E
MAVG00000000	Shahin PF	Anand	22 days	4	22° 33' 52.2648" N 72° 55' 43.9356" E

involvement of these pathogens in the complex respiratory infections using Next Generation Sequencing approach. We performed metagenome sequencing of DNA viruses in infected broilers from eight different poultry farms of Gujarat State in India. Details of each poultry farm, its location, number of samples are given in Table 1.

Next Generation Sequencing with Ion Torrent PGM of pooled eight DNA virome samples resulted in 6,888,239 reads and 1,627,633,377 (1.6 GB) bases. Quality filtration resulted in 6,174,274 reads and 15,588,747 bases showing 89.63% of reads passed quality check (Table 2). The host specific screening resulted in 4,314,327 reads and 1,052,038,233 bases (1.05 GB) which means 62.63% passed screening and were subjected for downstream analysis (Table 3). The average mean read length obtained before and after quality check was 229 bp and 253 bp, respectively. The reads were assembled using De novo assembly approach using SPAdes tool which produced 15,746 contigs (7,676,901 bps) at k-mer 55. The majority of contig were found to be ranging between 56 - 1000 bp. The assembly statistics report shows largest contig size at 21535bp, N50 size at 475 bp and GC content 44.27% (Table 4). Contigs less than 100 bases were discarded from further analysis.

Virome analysis using Blastx, Virome and MetaVir (BVM) showed 15,765 contigs, which were assigned to 103 classified and unclassified viral families. Across all samples the most abundant viral families and top 20 identified were Siphoviridae (632), Myoviridae (342), Podoviridae (232), Potyviridae (103), Papillomaviridae (102), Polydnaviridae (88), Rhabdoviridae (79), Flaviviridae (71), Picornaviridae (68), Baculoviridae (59), Herpesviridae (56), Betaflexiviridae (56), Paramyxoviridae (52), Bunyaviridae (52), Retroviridae (48), Adenoviridae (48), Caulimoviridae (46), Secoviridae (43), Coronaviridae (41) and Closteroviridae (41) as classified, and unclassified (224) families. Those contigs which did not produce hits against virus database were either belonging to cellular organisms (bacteria, fungi, and avian species) or had no hits against the BLAST nucleotide database. Total of 2945 viruses were predicted by Blastx having 13,686 contigs hits, by Virome having 547 contigs hits and from MetaVir having 558 contigs hits which includes phages. For analysis we focussed on mainly viral sequences and removed phage hits. This showed total 1663 viruses predicted by BVM, where Blastx resulting in 6383 contigs hits, Virome showing 109 hits and MetaVir showing 558 contig hits. MetaVir and Virome produced less hits compared to Blastx, may be because NCBI database is updated and populated with approximately 7463 complete viral genome sequences till date which can be aligned to the metagenomic sequences and produce hits. The most common viruses having highest count and ranking top 10 were *Pandoravirus salinus, Pandoravirus dulcis, Acanthamoeba polyphagamimivirus, Acanthamoeba polyphaga moumouvirus, Megaviruslba, Acanthocystis turfacea Chlorella virus* 1, Cafeteria roenbergensis virus BV-PW1, Microviridae phi-CA82*, Pithovirus sibericum, Lausanne virus*, Marine gokushovirus (Table 4).

Predominant Viruses

We carried out reference guided mapping approach to also map high quality and host specific screened reads against the virus database and predict the abundance of viral populations. Our aim was to align maximum sequences with the viruses and in a way to validate the analysis done using Blastx, Virome and MetaVir by *insilico* approach. We predicted presence of 3126 viruses showing 1810 viral hits and 1316 phages by mapping approach. We filtered out mapping hits having <10 reads aligned to the references. Most predominantly existing viruses with >90% genome coverage were Avian Gyrovirus 2, Gy-rovirus 4, complete genome, Chicken anemia virus, complete genome, Gyrovirus Tu789, complete genome, Human Gyrovirus type 1, Mongoose feces-associated gemycircular virus b strain 160b.

Important Viruses

The homology searches using BVM and CLC mapping predicted various known and unknown viruses which could be independently or jointly involved in the respiratory infections. Viruses namely

Table 2: Quality and length filtering of DNA virus reads

Accession No.	Total Reads	Total Bases	Mean Read Length	Reads <50 bp	Reads <20 quality score	High Quality bases	High Quality Mean Length	Total High Quality Reads used for Screening	% High Quality Reads
MAUZ00000000	8,84,693	23,48,05,603	265.41	55341	13834	22,23,92,266	287.33	8,15,518	92
MAVA00000000	10,15,266	28,96,26,968	285.27	49946	16164	28,20,33,137	301.95	9,49,156	93
MAVB00000000	9,04,801	22,07,15,149	243.94	58637	15331	21,21,34,615	263.8	8,30,833	92
MAVC00000000	2,65,217	4,67,13,857	176.13	46282	5798	4,41,18,624	209.19	2,13,137	80
MAVD00000000	10,62,736	21,88,81,880	205.96	95737	21813	20,66,36,396	230.38	9,45,186	89
MAVE00000000	7,83,283	14,14,44,442	180.58	104810	17928	13,32,73,138	207.2	6,60,545	84
MAVF00000000	9,06,034	20,29,75,398	224.03	91369	18471	19,48,44,941	248.98	7,96,194	88
MAVG00000000	10,66,209	27,24,70,080	255.55	84806	17698	26,34,41,632	277.91	9,63,705	90

Table 3: Host specific screening of high quality reads using *Gallus gallus* genome

Accession No.	Total High Quality Reads used for Screening	Number of Reads aligned to Host	Percentage of Reads aligned to Host	No. of Reads unaligned to host	Percentage of Reads unaligned to Host	No. of Remaining bases	No. of Remaining bases in MB
MAUZ00000000	8,15,518	283000	34.70%	532518	65.30%	150412375	150
MAVA00000000	9,49,156	87392	9.21%	861764	90.79%	272863210	272
MAVB00000000	8,30,833	111855	13.46%	718978	86.54%	199324024	199
MAVC00000000	2,13,137	92406	43.35%	120731	56.64%	25500359	25
MAVD00000000	9,45,186	302838	32.04%	642348	67.96%	162062566	162
MAVE00000000	6,60,545	255463	38.67%	405082	61.33%	91345731	91
MAVF00000000	7,96,194	319368	40.11%	476826	59.89%	91640245	91
MAVG00000000	9,63,705	407625	42.30%	556080	57.70%	58889723	58

Pandoravirus salinus and *Pandoravirus dulcis* belonging to the Pandoraviridae family produced 210 contigs hits by Blastx and 11004 reads mapping whereas 130 contig hits and 5542 reads mapping, respectively. Both these viruses belonging to the host amoeba Acanthamoeba *castellani* have been recently discovered and are the largest viruses known having 2 megabase genomes exceeding in size the genomes of numerous bacteria and archaea. Pandoraviruses show a distant relationship with other nucleocytoplasmic large DNA viruses (NCLDV) of eukaryotes (Yutin and Koonin, 2013). Though both the viruses showed highest contig hits in the present analysis, but there are no significant reports available which shows its association in the poultry RI. Another largest virus showed third highest contig hit (91) and 2599 reads mapped to Acanthamoeba polyphagamimivirus, whereas contig hit (87) and 1917 reads mapped to Acanthamoeba polyphagamoumouvirus belonging to the Mimiviradae family and residing inside amoeba. The Acanthamoeba polyphagamimivirus has been isolated from the patients suffering from pneumonia and shows its involvement in the respiratory diseases but no report has been found in the chickens (Saadi *et al.,* 2013). The Acanthamoeba polyphagamoumouvirus represents a third lineage beside mimivirus and megavirus. The moumouvirus and megavirus genomes share near perfect orthologous gene collinearity in the central part of the genome, with the variations concentrated in the terminal regions. In addition, genomic comparisons of the Mimiviridae reveal substantial gene loss in the moumouvirus lineage. The majority of the remaining moumouvirus proteins are most similar to homologs from other Mimiviridae members (Yoosuf *et al.*, 2012). Its presence in the DNA virome is significantly high but has no information available describing its role in causing RI in poultry. No major report is available for the Megaviruslba which shows 73 contig hits and 1423 mapped reads in our infected samples. We found 68 contigs hits and 652 reads mapping to the Acanthocystis turfacea chlorella virus 1 (ATCV-1) which is a member of the genus Chlorovirus (family Phycodnaviridae). This family of algae-infecting viruses have been found in human oropharyngeal samples with changes in cognitive functions in humans and mice (Yolken *et al.*, 2014). We found the presence of previously known viruses like Fowl pox virus (24 hits), Newcastle disease virus B1 (1 hit), Infectious bronchitis virus (2 hits), Gallid herpes virus 2 (14 hits), Chicken anemia virus (17 hits) reported earlier as causing respiratory infection in poultry.

Table 4: De novo assembly by SPAdes

Assembly	Count
# contigs (>= 0 bp)	15746
# contigs (>= 1000 bp)	877
# contigs (>= 5000 bp)	68
# contigs (>= 10000 bp)	12
# contigs (>= 25000 bp)	0
# contigs (>= 50000 bp)	0
Total length (>= 0 bp)	7676901
Total length (>= 1000 bp)	1986782
Total length (>= 5000 bp)	499216
Total length (>= 10000 bp)	145663
Total length (>= 25000 bp)	0
Total length (>= 50000 bp)	0
Largest contig	21535
Total length	145663
GC (%)	44.22
N50	475
N75	409
L50	3616
L75	8050
# N's per 100 kbp	0

The present study was the first approach to study poultry virome having respiratory infections. We provide annotated virome sequences obtained from Blastx, Virome and MetaVir tools and applications. We have found novel viral sequences as well as previously reported viruses in our data sets which are probable to be involved in causing infections and can be used for designing effective and preventive measures.

Acknowledgement

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

Conflict of Interest: All authors express no conflict of interest.

References:

Gowthaman, V., Singh, S.D., Dhama, K., Barathidasan, R., Mathapati, B.S., Srinivasan, P., Saravanan, S. and Ramakrishnan, M.A. (2014). Molecular detection and characterization of infectious laryngotracheitis virus (Gallid herpesvirus-1) from clinical samples of commercial poultry flocks in India. Virus disease, **25**: 345-349.

Miller, P.J. and Torchetti, M.K. (2014). Newcastle disease virus detection and differentiation from avian influenza. Methods Mol. Biol., **1161**: 235-239.

Saadi, H., Pagnier, I., Colson, P., Cherif, J.K., Beji, M., Boughalmi, M., Azza, S., Armstrong, N., Robert, C., Fournous, G., La Scola, B. and Raoult, D. (2013). First isolation of Mimivirus in a patient with pneumonia. Clin. Infect. Dis., **57**: e127-134.

Sommer, F. and Cardona, C. (2003). Chicken anemia virus in broilers: dynamics of the infection in two commercial broiler flocks. Avian Dis., **47**: 1466-1473.

Wozniakowski, G., Mamczur, A. and Samorek-Salamonowicz, E. (2015). Common occurrence of Gallid herpesvirus-2 with reticuloendotheliosis virus in chickens caused by possible contamination of vaccine stocks. J. Appl. Microbiol., **118**: 803-808.

Yolken, R.H., Jones-Brando, L., Dunigan, D.D., Kannan, G., Dickerson, F., Severance, E., Sabunciyan, S., Talbot, C.C., Jr., Prandovszky, E., Gurnon, J.R., Agarkova, I.V., Leister, F., Gressitt, K.L., Chen, O., Deuber, B., Ma, F., Pletnikov, M.V. and Van Etten, J.L. (2014). Chlorovirus ATCV-1 is part of the human oropharyngeal virome and is associated with changes in cognitive functions in humans and mice. Proc. Natl. Acad. Sci. USA, **111**: 16106-16111.

Yoosuf, N., Yutin, N., Colson, P., Shabalina, S.A., Pagnier, I., Robert, C., Azza, S., Klose, T., Wong, J., Rossmann, M.G., La Scola, B., Raoult, D. and Koonin, E.V. (2012). Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol. Evol., **4**: 1324-1330.

Yutin, N. and Koonin, E.V. (2013). Pandoraviruses are highly derived phycodnaviruses. Biol. Direct, **8**: 25.

❑