



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Comparison of tissue sample processing methods for harvesting the viral metagenome and a snapshot of the RNA viral community in a turkey gut



Jigna D. Shah^a, Joshua Baller^b, Ying Zhang^b, Kevin Silverstein^b,
Zheng Xing^a, Carol J. Cardona^{a,*}

^a Veterinary and Biomedical Sciences, University of Minnesota, MN, USA

^b Minnesota Supercomputing Institute, University of Minnesota, MN, USA

ABSTRACT

Article history:

Received 14 April 2014

Received in revised form 18 July 2014

Accepted 15 August 2014

Available online 1 September 2014

Keywords:

RNA viral metagenomics

Turkey

Gastrointestinal virome

RNA viruses have been associated with enteritis in poultry and have been isolated from diseased birds. The same viral agents have also been detected in healthy flocks bringing into question their role in health and disease. In order to understand better eukaryotic viruses in the gut, this project focused on evaluating alternative methods to purify and concentrate viral particles, which do not involve the use of density gradients, for generating viral metagenome data. In this study, the sequence outcomes of three tissue processing methods have been evaluated and a data analysis pipeline has been established for RNA viruses from the gastrointestinal tract. In addition, with the use of the best method and increased sequencing depth, a glimpse of the RNA viral community in the gastrointestinal tract of a clinically normal 5-week old turkey is presented. The viruses from the Reoviridae and Astroviridae families together accounted for 76.3% of total viruses identified. The rarefaction curve at the species level further indicated that majority of the species diversity was included with the increased sequencing depth, implying that viruses from other viral families were present in very low abundance.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Recent studies have demonstrated the diversity of the gut bacterial microbiome in many species including poultry species. The microbiome of the gut plays a significant role in performing numerous biochemical functions for the host (Beckmann et al., 2006; Wong et al., 2006; Dunkley et al., 2007; Shakouri et al., 2009; Rodriguez et al., 2012; Stanley et al., 2012, 2013), and alterations of the microbiome have been associated with diverse effects on the physiology and the health of the host (Van Immerseel et al., 2002; Turnbaugh et al., 2008; Yang et al., 2009; Gaggia et al., 2010; Bull-Otterson et al., 2013).

Several studies have characterized the bacterial component of the microbiome associated with specific host phenotypes or status of the bird (Dumoncaux et al., 2006; Danzeisen et al., 2011; Torok et al., 2011; Corrigan et al., 2012; Oakley et al., 2013; Wei

et al., 2013). Despite the popularity and economic importance of poultry as a food source, relatively little is known about the viral component of their gut microbiome and its source. Because the bacterial microbiota and the phage/DNA viral component of the microbiome are interdependent (Qu et al., 2008), the application of 'viral metagenomics' approaches that address DNA viruses has been the next step in assessing gut microbial communities (Breitbart et al., 2003; Cheval et al., 2011). But, in addition to DNA viruses, the poultry gut also harbors a complex consortium of eukaryotic or RNA viruses (Pantin-Jackwood et al., 2008) which have been associated with a host of enteric disease syndromes (Jindal et al., 2010). Enteric disease in poultry species often involves diarrhea and more progressive forms, such as poult enteritis mortality syndrome (PEMS), poult enteritis complex (PEC), runting-stunting syndrome of broilers (RSS) and several unclassified enteric diseases (Guy, 1998; Day and Zsak, 2013) result in high mortality. Not only do severe forms result in economic loss, but mild cases of enteritis and those that only affect a few birds in a flock do as well (Skinner et al., 2010). Although a number of different RNA viruses such as avian reoviruses, rotaviruses, astroviruses and coronaviruses, have been isolated from the intestinal contents of clinically ill birds (Reynolds et al., 1987b), the precise causative agent(s) of enteric disease are

* Corresponding author at: University of Minnesota, 301C Veterinary Science Building, 1971 Commonwealth Avenue, St. Paul, MN 55108, USA.
Tel.: +1 612 625 5276; fax: +1 612 625 5203.

E-mail address: ccardona@umn.edu (C.J. Cardona).

not known because knowledge of the composition of the eukaryotic or RNA viral community of the healthy poultry gut is rather limited making it difficult to distinguish between the 'core' and the 'transient' members and to thus identify causes of enteritis.

Recent advances in next-generation sequencing technology have the potential to circumvent the limitations of conventional detection methods and enable insights into gut community microbiota at a high resolution. In this study, our strategy was to examine the intestinal RNA viral flora of a clinically normal turkey, in order to aid in identifying RNA viral enteric pathogens accurately. However, eukaryotic viruses are a small component of the whole gut microbiome and sequencing RNA viruses from any niche has its own set of hurdles to overcome including the overabundance of the contaminating non-viral nucleic acids, lack of a universal phylogenetic marker and labor intensive viral purification methods involving 'density gradients'. There have been several viral metagenomic studies reported to date, characterizing DNA (Breitbart et al., 2003; Angly et al., 2006; Thurber et al., 2009; Minot et al., 2013) or RNA viruses (Culley et al., 2006; Zhang et al., 2006; Day et al., 2010) from various parts of the gastrointestinal system and environmental niches. Based on the sample type, these studies have used different protocols to recover viral particles and different sequencing platforms. Variations in the (a) methods to concentrate and purify viral particles, (b) library preparation protocols, and (c) sequencing platforms, have been reported to impact the DNA viral metagenome results (Hurwitz et al., 2013; Solonenko and Sullivan, 2013). We used the conceptual frameworks of these studies to design and test three different modified protocols with the potential to be used with large sample numbers for processing intestinal tissue and its contents to recover viral particles, while keeping the other parameters such as RNA extraction, library preparation and the sequencing platform exactly the same. In the first part of this study, the same sample was used to compare three tissue processing protocols for viral recovery and viral species richness as determined with an Illumina Miseq sequencing platform. The second part of this study explores modified bioinformatics approaches to analyze the resultant sequence data and gives a snapshot of the viral species richness and diversity present in a clinically normal turkey gut.

2. Materials and methods

2.1. Sample collection and processing

A 5-week old turkey with no clinical signs of disease was humanely euthanized following guidelines of the University of Minnesota Institutional Animal Care and Use Committee. The intestinal tract, (from duodenum to colon) along with contents, was harvested followed by removal of mesenteric fat and pancreas. The intestinal tissue was re-suspended in an approximately equal volume of 0.9% sterile saline and spiked with 1 ml of Influenza A virus suspension (A/mallard/Minnesota/Sg-00177/2007) at the concentration of 128 HAU/50 μ l, followed by thorough homogenization using the PRO250 homogenizer (PRO Scientific, Oxford, CT). The homogenate (~250 ml) was centrifuged at $4355 \times g/4^\circ\text{C}$ for 20 min and again at $7741 \times g/4^\circ\text{C}$ for 20 min to remove large tissue debris. The supernatant was filtered sequentially through 0.8 μ m, 0.45 μ m and 0.2 μ m filters (Millipore, Billerica, MA) to separate virus particles (VPs) less than 0.2 μ m in size, from bacteria and larger particles (Day et al., 2010). Filters with a PES membrane, were specifically used throughout this study for their low-protein binding characteristics in order to reduce the loss of the VPs (Mocé-Llivina et al., 2003). The filtrate was divided into six equal parts for further processing and two parts were processed in duplicate for each of the subsequent methods. A schematic of processing methods is presented in Fig. 1.

2.1.1. First processing method

Polyethylene glycol (PEG) – 6000 (Sigma–Aldrich, St. Louis, MO) was added at 8% (w/v) to the filtrate and stirred overnight at 4°C to concentrate the VPs. The mixture was centrifuged at $9798 \times g/4^\circ\text{C}$ for 30 min, the pellet saved at 4°C and the process was repeated on the supernatant after which the pellets from the first and second centrifugations were combined. The resultant pellet was re-suspended in an equal volume of 50 mM Tris–HCl, pH 7.4 (Fisher Scientific, Pittsburgh, PA).

2.1.2. Second processing method

The filtrate was treated as described for the first method except that the combined pellet from the two centrifugation steps was re-suspended in 15 ml of cold 0.9% saline. The suspension was sonicated with 6–10 cycles of 20 s runs/10 s pause to separate VPs from PEG particles. PEG particles were pelleted by centrifugation at $7741 \times g/4^\circ\text{C}$ for 5 min. The supernatant containing the VPs was transferred to a clean tube and ultra-centrifuged at $156,530 \times g/4^\circ\text{C}$ for 6 h. The supernatant was aspirated, the pellet harvested and re-suspended in an equal volume of 50 mM Tris–HCl, pH 7.4 (Fisher Scientific).

2.1.3. Third processing method

The filtrate was directly ultra-centrifuged at $156,530 \times g/4^\circ\text{C}$ for 6 h. The supernatant was aspirated, pellet harvested and re-suspended in an equal volume of 50 mM Tris–HCl, pH 7.4 (Fisher Scientific).

2.2. RNA extraction

The re-suspended pellets from each of the three processing methods i.e. six samples were individually mixed in 50 mM Tris–HCl, pH 7.4 and treated with 5 μ l of RNaseA (1 mg/ml) (ThermoScientific, Lafayette, CO). The reactions were incubated for 30 min at room temperature, to degrade any non-viral RNA. After RNaseA treatment, Trizol LS reagent (Invitrogen, Grand Island, NY) was added to each suspension at a ratio of 3:1 and was mixed. For each reaction, total RNA from the intact VPs was extracted according to the manufacturer's instructions. The quality of the total RNA was determined with a Nanodrop 2000 (ThermoScientific, Wilmington, DE).

2.3. Sequencing

2.3.1. First sequencing

Total RNA from each of the processing method was converted to Illumina sequencing libraries using Illumina's TruSeq RNA Sample Prep Kit v2 according to manufacturer's instructions. Final library size distribution was validated and indexed, and libraries from the three processing methods were then normalized, pooled and size selected at 270 bp \pm 5% using Caliper's XT instrument giving an average insert size of approximately 150 bp. The pooled libraries i.e. six samples were sequenced on a MiSeq 151 PE run. De-multiplexed FASTQ files were used for subsequent analysis.

2.3.2. Second sequencing

To achieve greater sequencing depth and coverage, the library for the processing methods which resulted in the highest species richness was re-sequenced. For re-sequencing, the library size was selected to be 320 bp \pm 10% giving an average insert size of approximately 200 bp. This library was sequenced on a MiSeq 151 PE run. The FASTQ files were used for subsequent analysis.

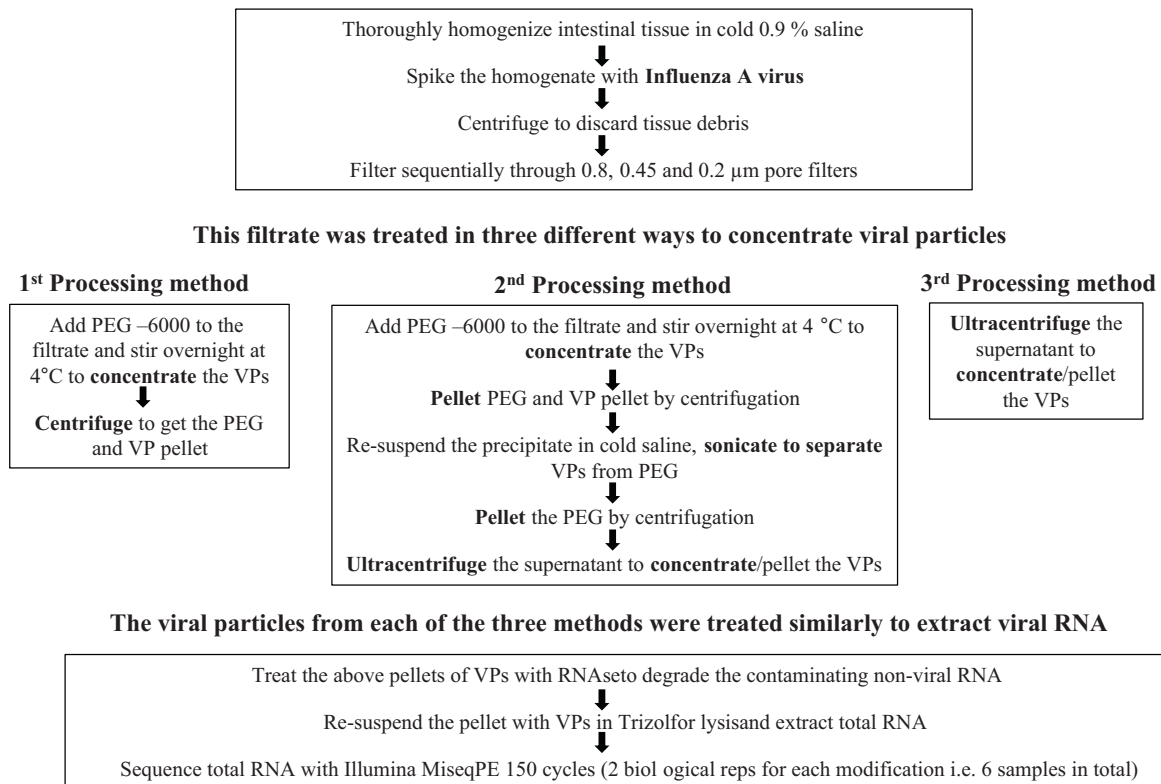


Fig. 1. Schematic representation of three processing methods used to concentrate viral particles. VPs: viral particles and PEG: polyethylene glycol – 6000.

2.4. Data analysis

2.4.1. Sequencing data from all processing methods

Sequence data was analyzed to identify rRNA-like sequences using the riboPicker tool, version 0.4.3 (Schmieder et al., 2012) which uses Burrows–Wheeler Aligner program. All datasets available as of August 2012, on the web-version were selected as reference databases with the thresholds set at 50% coverage, 90% identity and 100 bp alignment length. The quality check for the sequences from each of the processing methods was done using FastQC. AbokiaBLAST, a commercial parallel implementation of NCBI BLAST was used for all blast searches against the NT and NR databases downloaded from NCBI in August 2013. The blastn against the NT database, for the sequences from a single end i.e. only R1 reads from each of the processing methods, was done using default parameters. Blast hits with bit scores below 35 were removed. For the remaining BLAST hits, the NCBI gi ID was used to identify the associated taxon ID from the NCBI taxonomy database. Visualization of the weights of taxonomic nodes was done using MEGAN 4 software.

2.4.2. Analysis of second sequencing

The quality of sequences from the re-sequencing run was evaluated using FastQC. The sequences with bases having a minimum *phred* quality score <20 were trimmed from the 3' end using FastQ Quality Trimmer (Blankenberg et al., 2010).

A modified approach was used to blast the sequences and reconcile the outputs from the paired ends for the re-sequenced dataset (Turkey_reseq) including the removal of duplicate fragments. To do this, paired reads were concatenated and one copy of each fragment was retained. The joined fragments were split to regenerate the forward (split R1) and reverse (split R2) reads. The paired reads were blasted independently against the NT database using the program's default parameters. The blast hits with bit scores below 35

were removed. For the remaining blast hits, the NCBI taxonomic ID associated with the sequence was determined and hits without taxonomy IDs were excluded. In order to keep the most robust hits, the blast results were independently grouped by their query IDs for each paired end and the hits with bit scores less than 90% of the highest bit score for the group were discarded. Lastly, all non-viral hits were filtered from the blast results. To reconcile the information from the paired ends F was defined as a sequenced DNA fragment, A as the non-unique set of taxon IDs associated with F's R1 read and B as the non-unique set of taxon IDs associated with F's R2 read. Using these definitions, the taxon IDs for the paired ends were reconciled into a single non-unique set of taxon IDs, O, for the F fragment according to the following rules: (a) if $A = \emptyset$ or $B = \emptyset$ then $O = A \cup B$ and (b) $A \neq \emptyset$ and $B \neq \emptyset$ then $O = A \cap B$. Our set notation is non-standard in that all sets are non-unique, allowing multiple occurrences of the same taxon ID. As a result, the union and intersection operators are additive with respect to the number of occurrences of a given taxon ID. This reconciled output was used for additional processing i.e. distribution of weight to leaf nodes and pruning the leaf nodes with weight lower than cutoff as described in Section 2.5.1 before MEGAN was used to visualize the taxonomic weights.

2.5. Diversity measurements

2.5.1. Exploration of various parameters on the stability of diversity indices using the Turkey_reseq dataset

The taxon IDs from the reconciled output indicated associations with nodes in the taxonomic tree. However, most sequences had multiple taxon IDs that mapped to different levels of the tree. To impose a degree of order to the tree weightings, all sequences were forced to map to leaf taxons rather than nodes farther up the tree. If each sequence fragment is considered to provide evidence of a weight of 1, the weight of non-leaf nodes was divided

among multiple leaf nodes in a two-step process. First, the weight was divided evenly to all taxonomic nodes associated with a fragment. Secondly, that weight was distributed evenly among all leaf nodes below it. The taxonomic diversity and distinctness indices were calculated as described by Clarke and Warwick (Pienkowski et al., 1998; Warwick and Clarke, 1995) with the modification that 'N', the number of nodes, was fixed to the number of leaf nodes in the viral domain (213,392) rather than the number of leaf nodes with weight. To provide a consistent sequencing depth for comparison between samples, five subsets of 100,000 reconciled fragments were selected by random sampling. These subsets (roughly 10% of the full dataset) were used to evaluate the effect of even weight distribution on diversity.

The effect of leaf node pruning on diversity scores was evaluated with whole dataset. In order to ensure that a given weight cutoff value had a similar effect regardless of sequencing depth or taxonomy region studied, the leaf node weights were rescaled. In order to rescale, the weight of the leaf nodes was divided by the total number of mapped fragments and multiplied by the total number of leaf nodes. This had the effect of making a cutoff value of 1 equivalent to the weight expected per node if the total weight was distributed uniformly. Actual cutoff values used ranged from 0 to 16 with a cutoff of 1 used for further analysis.

The relationship between sequencing depth and saturation of the taxonomic tree with respect to the calculation of diversity scores was evaluated. The analysis was run with various sized sub-samples of the dataset ranging from 0.625% to 100% of the data.

2.5.2. Evaluation of the stability of diversity indices using a randomly generated dataset as negative control

A control dataset of one million fragments was created by randomly sampling from the NT database using a custom script. Paired end reads were modeled by selecting two 100 bp reads from single selected fragments such that the overall fragment size distribution was centered at 220 bp with standard deviation of 20 bp. This dataset was then 'mutated' by randomly altering 20% of the fragment's nucleotides. Five different datasets of 0.1 million reads each were created by randomly sampling from this larger dataset. These randomly generated datasets were processed similarly to the Turkey_reseq-derived sequences for even weight distribution and checking the stability of diversity indices, and sub-sampling for calculation of diversity indices. The whole dataset was used for checking the effect of leaf node pruning on diversity scores as described above in Section 2.5.1.

3. Results

3.1. Comparison of processing methods

The RNA concentrations and the number of sequences obtained with the first, second and third processing methods are summarized in Table 1. The sequence quality in the differentially processed material was checked with FastQC. For all three methods, the majority of the reads maintained *phred* quality scores above 20, and as a result, no quality trimming was performed. Since there

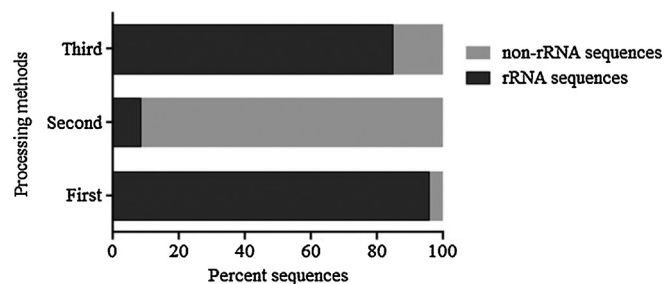


Fig. 2. Ratio of ribosomal RNA to non-ribosomal RNA sequences in the material from three processing methods. The sequencing data from all the three processing methods was analyzed using similar parameters with riboPicker tool version 0.4.3. The methods involved (1) concentrating the VPs using PEG, (2) concentrating the VPs using PEG followed by sonication to separate VPs from PEG particles and ultracentrifugation to concentrate the VPs again, and (3) ultracentrifugation to concentrate the VPs.

was no significant variation in the replicates for following analysis, only the results of one replicate from each processing method are discussed below.

The sequencing data from a single end (read 1) was first analyzed to identify rRNA-like sequences using the riboPicker tool. Since viral genomes do not contain 16s or 18s ribosomal genes, the presence of these sequences provided a semi-quantitative measure of eukaryotic or prokaryotic contamination. The analysis revealed that the highest and lowest percentages of ribosomal RNA was detected in the total RNA from the first (95.61%) and second (8.47%) processing methods respectively (Fig. 2).

The blast search against the NT database demonstrated that the total RNA from the second processing method resulted in the largest proportion of reads with hits to the viral domain (23.36% of reads) (Table 2). Similarly, this also had the highest viral richness with 15 different viral genera detected. The most abundant viral sequences detected were from the genera Rotavirus (81.4% of the reads assigned) and Avastrovirus (18.4% of the reads assigned). The Influenza A virus, which was added to the tissue before processing, was detected in the material from the second processing method with 14 reads assigned to the node (Fig. 3b). Influenza A virus was also detected in the total RNA from the third processing method with 469 reads assigned to the node, but viruses belonging to only three other viral genera including Rotavirus and Avastrovirus were detected in this material (Fig. 3c). Influenza virus was not detected in the material from first processing method and only three viral genera including Rotavirus and Avastrovirus were detected in this RNA (Fig. 3a). Based on these results, the RNA obtained with the second processing method was re-sequenced for greater depth and analyzed further.

3.2. Deep sequencing of the RNA from second processing method

The RNA from the second processing method was re-sequenced for increased coverage resulting in 8,111,627 sequences with an average of 46% GC content for each of the paired end reads for this dataset. Paired reads with perfect fragment duplicates (R1 and

Table 1
RNA yields and the number of sequences for each of the six samples obtained from the first sequencing run.

Processing method	Replicate	RNA (ng/μl)	260/280 ratio	Number of sequences	% GC content
Processing method 1	Replicate 1	3039.1	2.07	568,199	54
	Replicate 2	2913.3	2.06	543,682	54
Processing method 2	Replicate 1	24.0	1.71	563,877	48
	Replicate 2	24.9	1.74	547,587	47
Processing method 3	Replicate 1	330.7	1.87	837,655	57
	Replicate 2	331.2	1.87	886,840	57

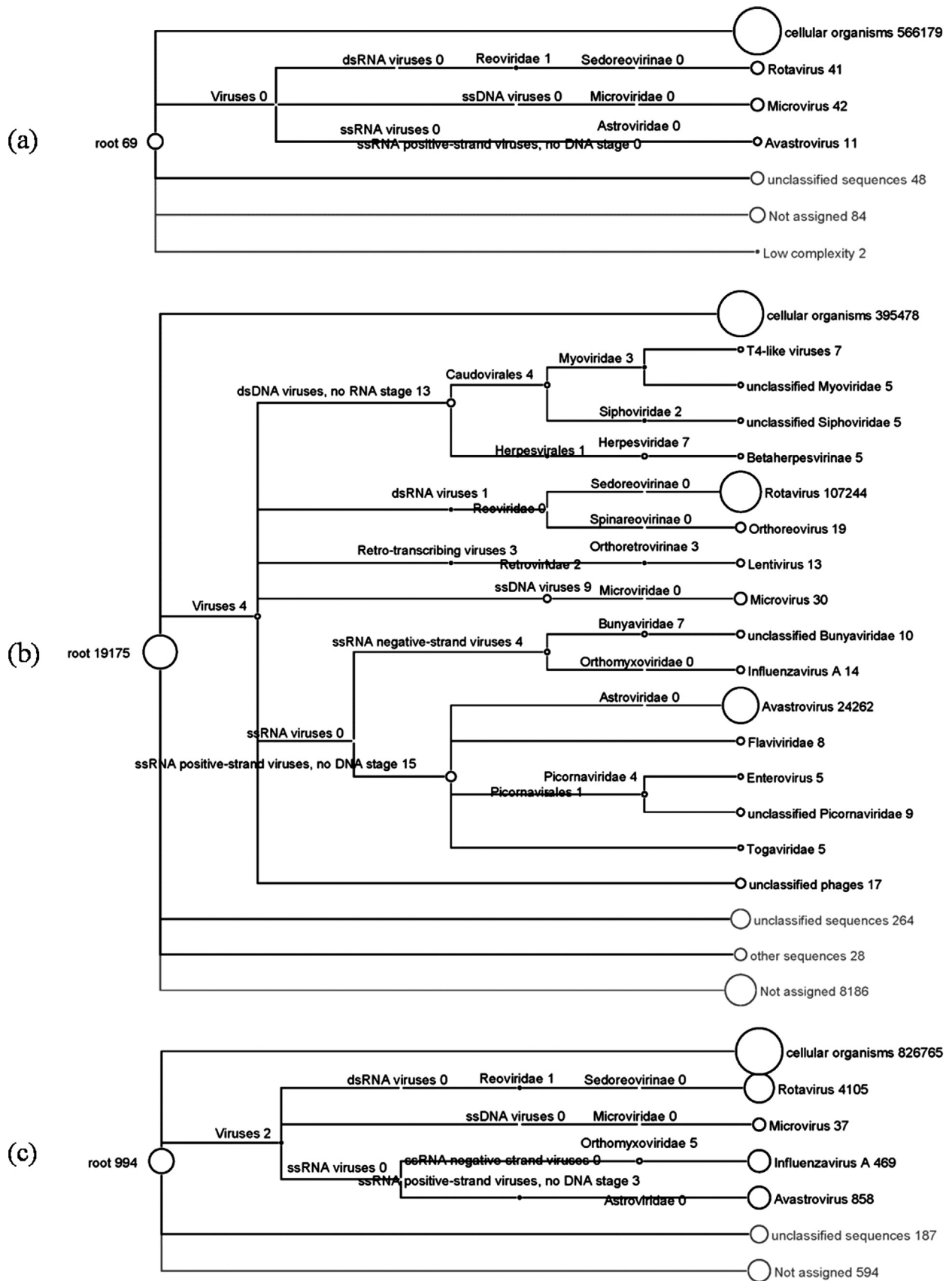


Fig. 3. Results of blastn mapped onto a NCBI taxonomic tree using MEGAN. The three panels show the results for RNA derived from three processing methods (a) first, (b) second, and (c) third processing methods. The sizes of the circles are proportional to the number of reads assigned to the particular node and the numbers of reads assigned are shown beside each taxon node.

Table 2
Results of blastn mapped to NCBI taxonomic tree for the data from three different processing methods. The numbers in the table represent the number of sequences with hits to a particular node.

Node	Processing method 1	Processing method 2	Processing method 3
Total sequences	568,199	563,877	837,655
Root	69	19,175	994
Cellular organisms	566,179	395,478	826,765
Viruses	95	131,741	5480
Unclassified sequences	48	264	187
Other sequences	–	28	–
Not assigned	84	8186	594
Low complexity	2	–	–
% of viruses/total	0.02%	23.36%	0.65%

Table 3
Results of blastn mapped to NCBI taxonomic tree for re-sequenced data from processing method 2.

	Split R1	Split R2
Total reads		8,111,627
Unique reads		5,872,593
Reads with hits	5,663,193 (96%)	5,557,275 (94%)
Total hits	599,555,638	581,830,758
Hits lacking a taxonomy match	21,928,511	21,292,938
Hits mapping to non-viral taxonomy	111,695,287	109,848,537
Reads with hits to viral taxonomy	1,202,396	1,181,025

R2 identical) were assumed to be PCR artifacts and only one fragment was retained resulting in 5,872,593 unique fragments. The R1 and R2 reads were blasted independently resulting in 1,202,396 (20.47%) and 1,181,025 (20.11%) reads with hits to viral sequences (Table 3). In reconciling the output of hits to virus sequences, we found that (a) 120,895 fragments had hits to R1 but not R2, (b) 99,524 fragments had hits to R2 but not R1, (c) 67,262 fragments had no agreement between the hits to R1 and R2, and (d) 1,014,239 fragments had some agreement between hits to R1 and R2. This resulted in a total of 1,234,658 (calculated from (a) + (b) + (d)) fragments (21.0%) from the Turkey_reseq dataset which had hits to viral sequences.

3.3. Effect of different parameters on the stability of the diversity indices

The reconciled data of fragments with hits to viral sequences was used to visualize the taxonomic tree with MEGAN. The data in Table 4 summarizes the number of leaf nodes before and after the weight distribution to the leaf nodes of a taxonomic tree. For the randomly sampled datasets of the Turkey_reseq dataset and the Random_NT_1M dataset, the taxonomic diversity and distinctness indices were calculated after weight was distributed evenly (Fig. 4a)

Table 4
The number of leaf nodes before and after modification for five randomly sampled datasets each from the Turkey_reseq dataset and the Random_NT_1M dataset.

Datasource	Seed number	Unique taxa (pre-distribution)	Weighted leaf nodes (even distribution)
Random_NT_1M	1	44,008	65,041
	2	43,355	66,280
	3	43,971	65,847
	4	44,735	65,749
	5	44,383	66,216
Turkey_reseq	1	13,756	38,718
	2	11,801	40,144
	3	14,110	39,338
	4	14,110	39,074
	5	13,230	38,444

and the diversity indices were stable for both the experimental and control datasets.

The taxonomic tree was pruned to eliminate bias after even weight distribution and to reduce the effect of spurious mappings. The effect of various pruning values, on diversity measures was investigated (Fig. 4b). As expected, based on the observed reduction in the number of species present, more aggressive pruning reduced diversity. Given a similar response to pruning from both the 'Random_NT_1M' and 'Turkey_reseq' datasets, and a lack of evidence for a noise floor in either the diversity vs. cutoff curve or in a histogram of leaf node weights (not shown), the cutoff was selected to be equal to the weight expected by uniform mapping (a cutoff of 1).

3.4. Snapshot of the composition of the RNA viral domain of a turkey gut

The data used to calculate diversity scores, with even weight distribution and the pruning of leaf nodes with weight lower than the cutoff, was used as input for MEGAN to visualize the taxonomic tree. Results revealed that approximately 85% of the sequences had hits to RNA viruses and approximately 15% of the sequences had hits to DNA and other viruses (Fig. 5). Among the RNA viruses, the dsRNA viruses and the ssRNA positive-strand viruses were found in abundance (81% together) compared with ssRNA negative-strand viruses and Retro-transcribing viruses (4% together). When summarized to the family level, the dsRNA viruses, ssRNA positive strand viruses and Retroviruses had a dominant single family, Reoviridae (99.6%), Astroviridae (90.9%) and Retroviridae family (92.7%) respectively. Additional families detected in low abundance from the dsRNA viruses, Retroviruses and ssRNA positive strand viruses are shown in Fig. 6a, b and Table 5 respectively. Hits to ssRNA negative-strand viruses showed greater richness with viruses from the Bunyaviridae (34.0%), Paramyxoviridae (26.0%), Rhabdoviridae (17.3%) and Arenaviridae (10.3%) families being well represented. The additional families detected in low abundance are shown in Fig. 6c. The viruses from the Reoviridae and Astroviridae families together accounted for 76.3% of total viruses identified. These results were in agreement with the preliminary results of the second processing method in which viruses of the genera Astrovirus (Astroviridae family) and Rotavirus (Reoviridae family) were detected in abundance (Fig. 3b).

The increase in the resolution of the taxonomic tree to the lowest species level possible further revealed that there were numerous other viruses present in low abundance, with sequences that mapped specifically to leaf nodes. These RNA viruses or variants of these known pathogenic viruses or closely related viruses can be presumed to be present in the turkey gut (supplementary data Fig. S1).

Supplementary Fig. S1 related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jviromet.2014.08.011>.

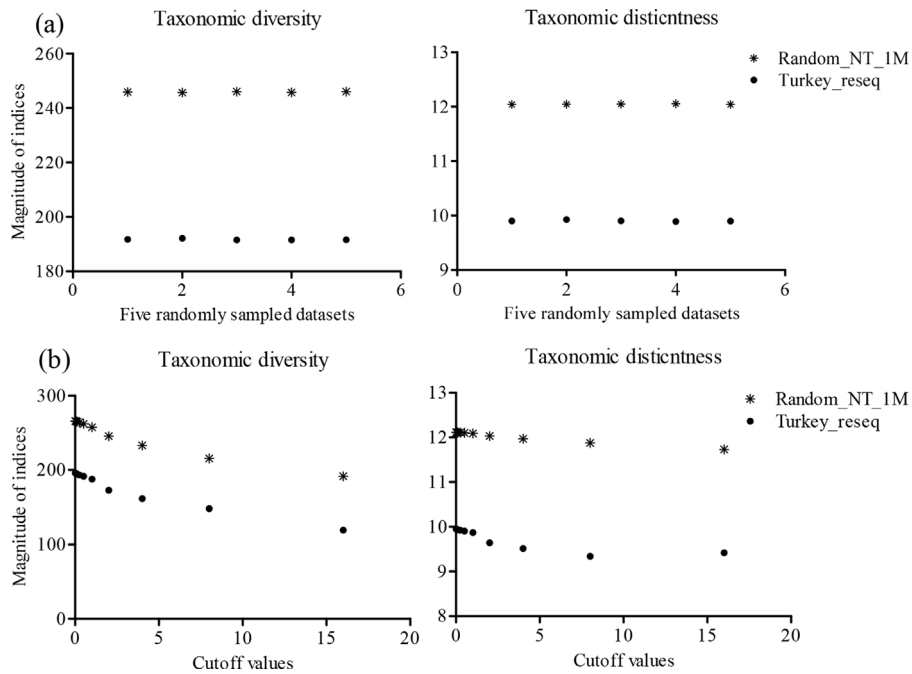


Fig. 4. Effect of different parameters on the taxonomic diversity and distinctness indices. (a) Diversity indices after even weight distribution to the leaf nodes for five randomly sampled datasets of 0.1 million reads each from each of the Turkey_reseq and the Random_NT_1M datasets. The X-axis is the dataset and Y-axis is the magnitude of the indices. (b) Diversity indices after leaf node pruning with different cutoff levels for each of the Turkey_reseq (●) and the Random_NT_1M (*) datasets. The X-axis is the cutoffs and Y-axis is the magnitude of the indices.

3.5. Dataset size and viral diversity

To assess the measurement of diversity as an effect of dataset size, the stability of diversity indices was explored with both the Turkey_reseq and the Random_NT_1M datasets. For both, repeated sub-sampling of increasing portions of the data demonstrated that both indices were relatively stable after application of the even weight distribution method and pruning of leaves with weights

lower than the cutoff (Fig. 7). Since the diversity measures for the Turkey_reseq dataset were stable even when a small portion of the sequence was sub-sampled, we concluded that most of the viral diversity for the tissue was detected and sequenced at a sequencing depth lower than what was obtained in the re-sequencing.

To account for species richness, rarefaction curves were computed with increasing proportions of the data (MEGAN with default parameters). At the genus level, the slope increased sharply until

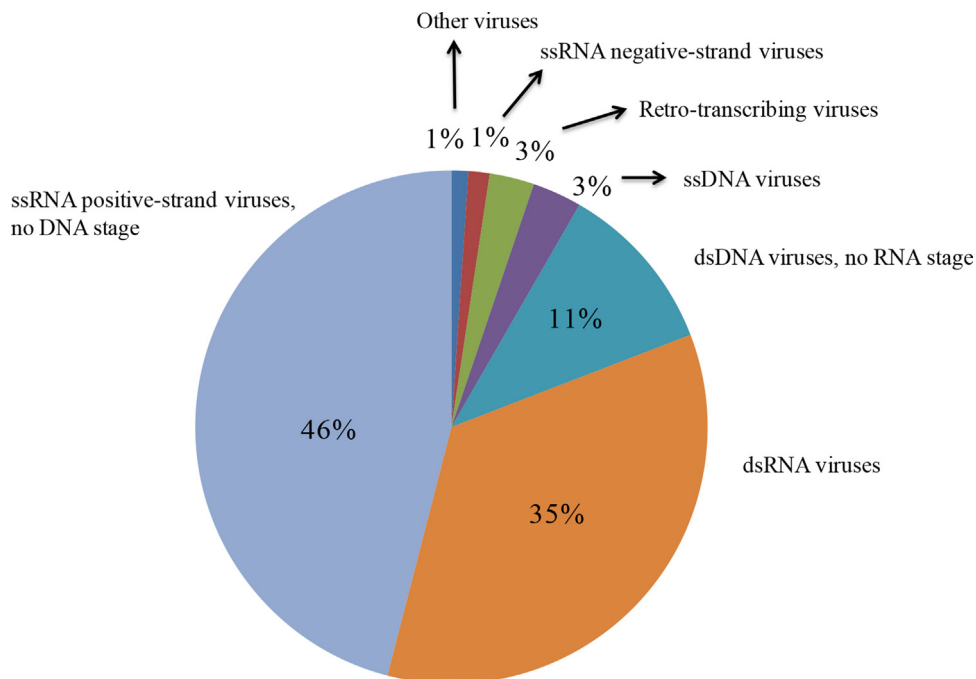


Fig. 5. Results of blastn as the percentage of reads by taxon nodes in the viral domain. The 'other viruses' group is comprised of Deltavirus, Satellites, Emaravirus, environmental viruses, unassigned and unclassified viruses, unclassified virophages, ssRNA viruses, archaeal viruses, and phages.

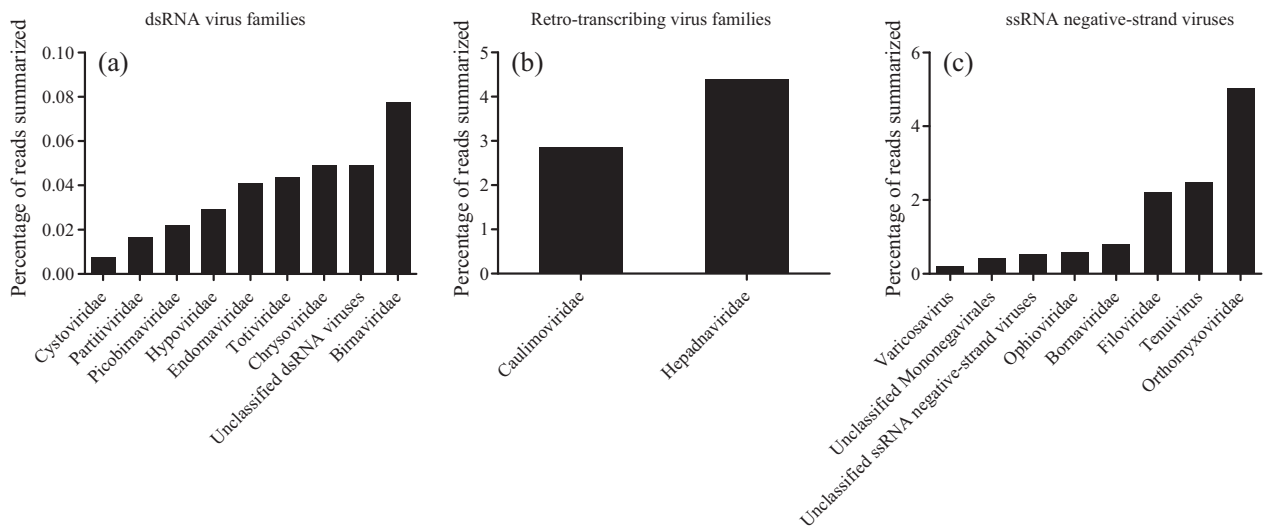


Fig. 6. Results of blastn with the percentage of reads summarized at each taxon node by the respective RNA viral families. Panels are (a) dsRNA viruses, (b) Retro-transcribing viruses, and (c) ssRNA negative-strand viruses. The X-axis lists the virus families and the Y-axis is the percentage of the reads summarized within a particular group.

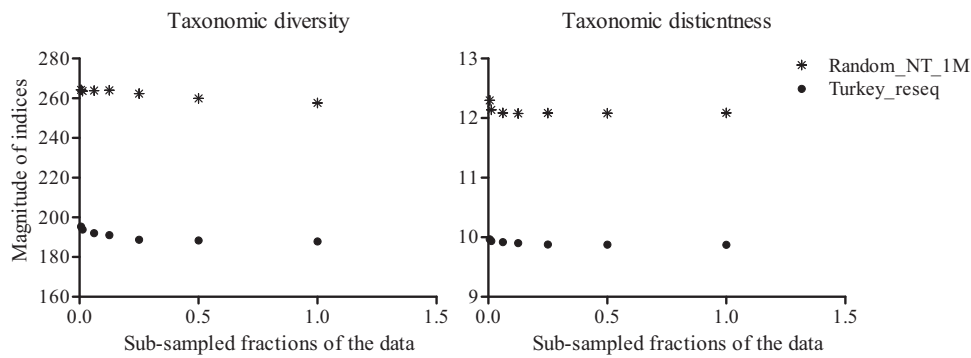


Fig. 7. Effect of the sub-sampling of different fractions of data (dataset size) on the taxonomic diversity and distinctness indices for each of the Turkey_reseq (●) and the Random_NT_1M (*) datasets. The average of multiple replicates is plotted for each data point. The X-axis is the sub-sampled fractions of the dataset and Y-axis is the magnitude of the indices.

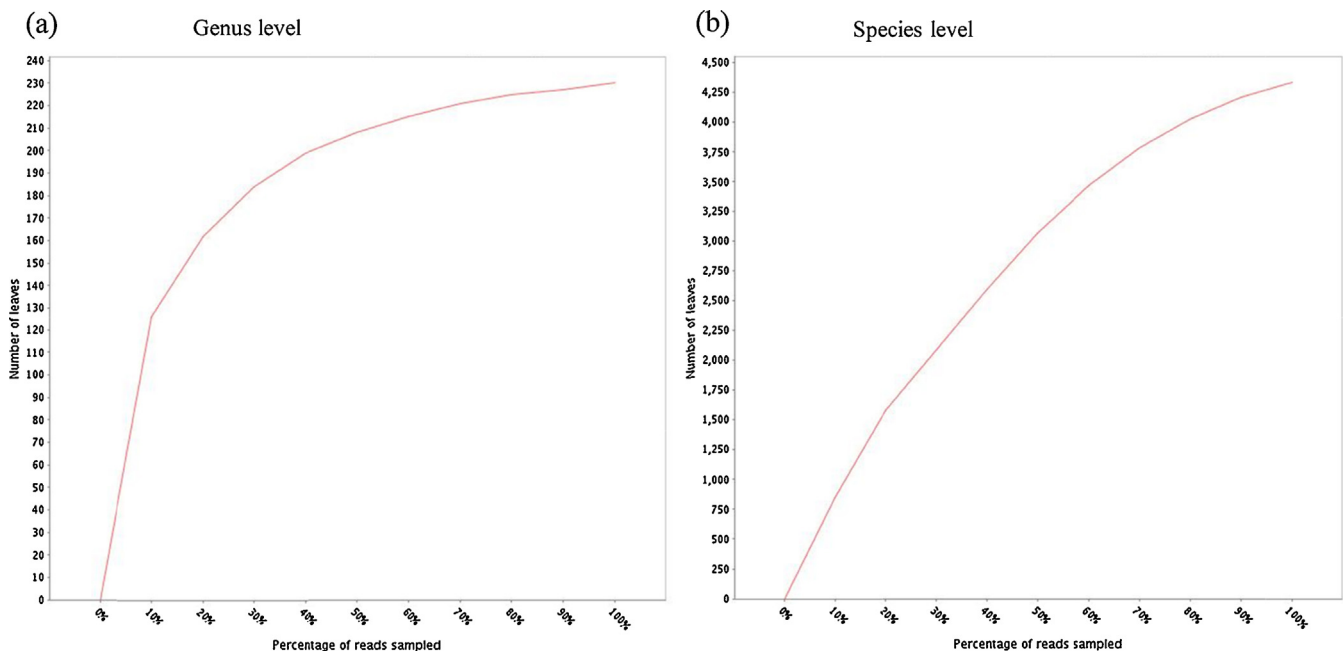


Fig. 8. Rarefaction curves for the Turkey_reseq dataset at the (a) genus level and (b) species level, of the viral domain. The X-axis is the percentage of the data sampled and the Y-axis is the number of species. The curves were created with default parameters using MEGAN.

Table 5

Low abundance families of positive strand ssRNA viruses detected. The percentage corresponds to the number of reads summarized at each taxon node of the respective families.

ssRNA positive strand viruses	Percentage
Narnaviridae	0.0022
Tetraviridae	0.0026
Unclassified Picornavirales	0.0048
Bacillariornaviridae	0.0067
Unclassified Nidovirales	0.0082
Unclassified Flexiviridae	0.0082
Umbravirus	0.0108
Sobemovirus	0.0112
Roniviridae	0.0123
Benyvirus	0.0134
Iflaviridae	0.0295
Unclassified ssRNA positive-strand viruses	0.0407
Leviridae	0.0414
Bromoviridae	0.0455
Dicistroviridae	0.0563
Tombusviridae	0.0578
Tymoviridae	0.0914
Luteoviridae	0.0970
Virgaviridae	0.0988
Alphaflexiviridae	0.1294
Togaviridae	0.2130
Closteroviridae	0.2137
Hepeviridae	0.2159
Caliciviridae	0.2667
Arteriviridae	0.3633
Betaflexiviridae	0.3886
Secoviridae	0.4006
Coronaviridae	0.5374
Potyviridae	0.9880
Picornaviridae	1.8748
Flaviviridae	2.8695

approximately 50% of the total data was sampled, after which the curve reached a plateau (Fig. 8a). At the species level, the curve reached a plateau only after approximately 90% of the data was sampled (Fig. 8b). The rarefaction curve indicated that not only the majority of the genera, but also the majority of the species diversity as well was sampled in the Turkey_reseq dataset.

4. Discussion

Conventionally, most methods for viral particle purification have involved the use of density gradients which are laborious. The results of this study provide a reasonably efficient protocol for recovery of viruses smaller than 0.2 μm , and which can be used to process multiple samples together, as would be necessary for epidemiologic and surveillance studies.

An initial comparison of the blast results from the three tissue processing methods demonstrated that the majority of the sequences recovered mapped to cellular organisms at varying levels likely because whole intestinal tissue homogenate was used to ensure the recovery of cell-associated RNA viruses. Simultaneously, a comparison of the ribosomal RNA vs. non-ribosomal RNA ratios revealed that the sample from the second processing method had the lowest number of sequences that mapped to ribosomal RNA and the most sequences mapping to the viral domain. Based on these measures, this processing method was selected for the remaining data analysis.

Of all the hits to the viral domain from the Turkey_reseq dataset, approximately 15% of the reads had hits to DNA viruses. This was likely due to the fact that DNase treatment was not incorporated before RNA extraction. Of the 85% of the reads with hits to RNA viruses, approximately 76% had hits to viruses from the Reoviridae and Astroviridae. Of these families, viruses of genera Rotavirus and Astrovirus were present in abundance. Viruses

belonging to these genera have been associated with enteric disease syndromes in many host species including turkeys (Koo et al., 2013; Reynolds et al., 1987a). In addition, recent studies have shown the widespread prevalence of these viruses in poultry flocks across various geographical locations with or without the clinical manifestations of disease (Pantin-Jackwood et al., 2008; Roussan et al., 2012). It is very likely that specific viruses from these genera are very closely related and minute differences in the genetic content alter their virulence or changes their ability to infect the host. It is also plausible that some of the species detected in this study are opportunistic pathogens and their population increases significantly in the gut in response to some perturbation, and for this reason, they are routinely isolated from the intestinal contents of birds with enteric disease. It is also possible that the turkey from which the tissue was collected would have developed enteric disease or had recently recovered from enteritis and because data from this study is a snapshot from one bird, it cannot be established if the viruses detected were core or transient members of the gut microbiota. Finally, one major limitation in assigning roles to these viruses is that the identification of the viruses from enteric disease samples has been traditionally done with molecular methods which do not have the species specific resolution. So, although the results from this study, are consistent with previous studies (Day et al., 2010) which showed the presence of viruses from Reoviridae, Caliciviridae and Picornaviridae in the turkey gut, for all of the stated reasons, the results do not clarify what these viruses are doing in the turkey gut and what role, if any, they play in causing disease.

The roles of the other viruses, present in the gut, in the health or disease of the turkey host are not known either. A large number of plant viruses were detected, which are likely consumed in the diet or from the environment. A study that investigated the RNA viruses from the feces of healthy individuals detected an abundance of plant pathogenic viruses in human gut (Zhang et al., 2006). Although plant viruses have emerged in animal hosts (Meehan et al., 1997; Gibbs and Weiller, 1999) they are not generally considered a risk to their animal hosts, a conclusion that is not based on any substantive data. To draw any further biological conclusions and to determine the core vs. the transient RNA viral community, data from multiple birds in various flocks across different geographical locations are needed which can be generated using the methods established with this study.

Acknowledgements

We thank Ms. Sarah Scharber for her tireless technical assistance with this project. We are grateful for the funding we received from the University of Minnesota, College of Veterinary Medicine Emerging and Zoonotic Disease signature program and the Pomeroy Chair for Avian Health endowment.

References

- Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J.M., Mueller, J.E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C.A., Rohwer, F., 2006. The marine viromes of four oceanic regions. *PLoS Biol.* 4, e368.
- Beckmann, L., Simon, O., Vahjen, W., 2006. Isolation and identification of mixed linked beta-glucan degrading bacteria in the intestine of broiler chickens and partial characterization of respective 1,3-1,4-beta-glucanase activities. *J. Basic Microbiol.* 46, 175–185.
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26, 1783–1785.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185, 6220–6223.
- Bull-Otterson, L., Feng, W., Kirpich, I., Wang, Y., Qin, X., Liu, Y., Gobejishvili, L., Joshi-Barve, S., Ayyaz, T., Petrosino, J., Kong, M., Barker, D., McClain, C., Barve, S., 2013. Metagenomic analyses of alcohol induced pathogenic alterations in the

- intestinal microbiome and the effect of *Lactobacillus rhamnosus* GG treatment. *PLoS ONE* 8, e53028.
- Cheval, J., Sauvage, V., Frangeul, L., Dacheux, L., Guigon, G., Dumey, N., Pariente, K., Rousseaux, C., Dorange, F., Berthet, N., Brisse, S., Moszer, I., Bourhy, H., Manuguerra, C.J., Lecuit, M., Burguiere, A., Caro, V., Eloit, M., 2011. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J. Clin. Microbiol.* 49, 3268–3275.
- Corrigan, A., Horgan, K., Clipson, N., Murphy, R.A., 2012. Effect of dietary prebiotic (mannan oligosaccharide) supplementation on the caecal bacterial community structure of turkeys. *Microb. Ecol.* 64, 826–836.
- Culley, A.I., Lang, A.S., Suttle, C.A., 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312, 1795–1798.
- Danzeisen, J.L., Kim, H.B., Isaacson, R.E., Tu, Z.J., Johnson, T.J., 2011. Modulations of the chicken cecal microbiome and metagenome in response to anticoccidial and growth promoter treatment. *PLoS ONE* 6, e27949.
- Day, J.M., Ballard, L.L., Duke, M.V., Scheffler, B.E., Zsak, L., 2010. Metagenomic analysis of the turkey gut RNA virus community. *Virology* 40, 313–319.
- Day, J.M., Zsak, L., 2013. Recent progress in the characterization of avian enteric viruses. *Avian Dis.* 57, 573–580.
- Dumoncaux, T.J., Hill, J.E., Hemmingsen, S.M., Van Kessel, A.G., 2006. Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. *Appl. Environ. Microbiol.* 72, 2815–2823.
- Dunkley, K.D., Dunkley, C.S., Njongmeta, N.L., Callaway, T.R., Hume, M.E., Kubena, L.F., Nisbet, D.J., Ricke, S.C., 2007. Comparison of in vitro fermentation and molecular microbial profiles of high-fiber feed substrates incubated with chicken cecal inocula. *Poult. Sci.* 86, 801–810.
- Gaggi, F., Mattarelli, P., Biavati, B., 2010. Probiotics and prebiotics in animal feeding for safe food production. *Int. J. Food Microbiol.* 141 (Suppl.), S15–S28.
- Gibbs, M.J., Weiller, G.F., 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8022–8027.
- Guy, J., 1998. Virus infections of the gastrointestinal tract of poultry. *Poult. Sci.* 77, 1166–1175.
- Hurwitz, B.L., Deng, L., Poulos, B.T., Sullivan, M.B., 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* 15, 1428–1440.
- Jindal, N., Patnayak, D.P., Chander, Y., Ziegler, A.F., Goyal, S.M., 2010. Detection and molecular characterization of enteric viruses from poult enteritis syndrome in turkeys. *Poult. Sci.* 89, 217–226.
- Koo, B.S., Lee, H.R., Jeon, E.O., Jang, H.S., Han, M.S., Mo, I.P., 2013. An unusual case of concomitant infection with chicken astrovirus and group A avian rotavirus in broilers with a history of severe clinical signs. *J. Vet. Sci.* 14, 231–233.
- Meehan, B.M., Creelan, J.L., McNulty, M.S., Todd, D., 1997. Sequence of porcine circovirus DNA: affinities with plant circoviruses. *J. Gen. Virol.* 78 (Pt 1), 221–227.
- Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., Bushman, F.D., 2013. Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12450–12455.
- Mocé-Llivina, L., Jofre, J., Muniesa, M., 2003. Comparison of polyvinylidene fluoride and polyether sulfone membranes in filtering viral suspensions. *J. Virol. Methods* 109, 99–101.
- Oakley, B.B., Morales, C.A., Line, J., Berrang, M.E., Meinersmann, R.J., Tillman, G.E., Wise, M.G., Siragusa, G.R., Hiatt, K.L., Seal, B.S., 2013. The poultry-associated microbiome: network analysis and farm-to-fork characterizations. *PLoS ONE* 8, e57190.
- Pantin-Jackwood, M.J., Day, J.M., Jackwood, M.W., Spackman, E., 2008. Enteric viruses detected by molecular methods in commercial chicken and turkey flocks in the United States between 2005 and 2006. *Avian Dis.* 52, 235–244.
- Pienkowski, M.W., Watkinson, A.R., Kerby, G., Warwick, R.M., Clarke, K.R., 1998. Taxonomic distinctness and environmental assessment. *J. Appl. Ecol.* 35, 532–543.
- Qu, A., Brulc, J.M., Wilson, M.K., Law, B.F., Theoret, J.R., Joens, L.A., Konkel, M.E., Angly, F., Dinsdale, E.A., Edwards, R.A., Nelson, K.E., White, B.A., 2008. Comparative metagenomics reveals host specific metaviromes and horizontal gene transfer elements in the chicken cecum microbiome. *PLoS ONE* 3, e2945.
- Reynolds, D.L., Saif, Y.M., Theil, K.W., 1987a. Enteric viral infections of turkey poults: incidence of infection. *Avian Dis.* 31, 272–276.
- Reynolds, D.L., Saif, Y.M., Theil, K.W., 1987b. A survey of enteric viruses of turkey poults. *Avian Dis.* 31, 89–98.
- Rodriguez, M.L., Rebole, A., Velasco, S., Ortiz, L.T., Trevino, J., Alzueta, C., 2012. Wheat- and barley-based diets with or without additives influence broiler chicken performance, nutrient digestibility and intestinal microflora. *J. Sci. Food Agric.* 92, 184–190.
- Roussan, D.A., Shaheen, I.A., Khawaldeh, G.Y., Totanji, W.S., Al-Rifai, R.H., 2012. Simultaneous detection of astrovirus, rotavirus, reovirus and adenovirus type 1 in broiler chicken flocks. *Pol. J. Vet. Sci.* 15, 337–344.
- Schmieder, R., Lim, Y.W., Edwards, R., 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* 28, 433–435.
- Shakouri, M.D., Iji, P.A., Mikkelsen, L.L., Cowieson, A.J., 2009. Intestinal function and gut microflora of broiler chickens as influenced by cereal grains and microbial enzyme supplementation. *J. Anim. Physiol. Anim. Nutr. (Berl.)* 93, 647–658.
- Skinner, J.T., Bauer, S., Young, V., Pauling, G., Wilson, J., 2010. An economic analysis of the impact of subclinical (mild) necrotic enteritis in broiler chickens. *Avian Dis.* 54, 1237–1240.
- Solomenko, S.A., Sullivan, M.B., 2013. Preparation of metagenomic libraries from naturally occurring marine viruses. *Methods Enzymol.* 531, 143–165.
- Stanley, D., Denman, S.E., Hughes, R.J., Geier, M.S., Crowley, T.M., Chen, H., Haring, V.R., Moore, R.J., 2012. Intestinal microbiota associated with differential feed conversion efficiency in chickens. *Appl. Microbiol. Biotechnol.* 96, 1361–1369.
- Stanley, D., Geier, M.S., Denman, S.E., Haring, V.R., Crowley, T.M., Hughes, R.J., Moore, R.J., 2013. Identification of chicken intestinal microbiota correlated with the efficiency of energy extraction from feed. *Vet. Microbiol.* 164, 85–92.
- Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., Rohwer, F., 2009. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470–483.
- Torok, V.A., Hughes, R.J., Mikkelsen, L.L., Perez-Maldonado, R., Balding, K., MacAlpine, R., Percy, N.J., Ophel-Keller, K., 2011. Identification and characterization of potential performance-related gut microbiotas in broiler chickens across various feeding trials. *Appl. Environ. Microbiol.* 77, 5868–5878.
- Turnbaugh, P.J., Bäckhed, F., Fulton, L., Gordon, J.I., 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* 3, 213–223.
- Van Immerseel, F., Cauwerts, K., Devriese, L.A., Haesebrouck, F., Ducatelle, R., 2002. Feed additives to control *Salmonella* in poultry. *Worlds Poult. Sci. J.* 58, 501–513.
- Warwick, R.M., Clarke, K.R., 1995. New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Mar. Ecol. Prog. Ser.* 129, 301–305.
- Wei, S., Morrison, M., Yu, Z., 2013. Bacterial census of poultry intestinal microbiome. *Poult. Sci.* 92, 671–683.
- Wong, J.M., de Souza, R., Kendall, C.W., Emam, A., Jenkins, D.J., 2006. Colonic health: fermentation and short chain fatty acids. *J. Clin. Gastroenterol.* 40, 235–243.
- Yang, Y., Iji, P.A., Choct, M., 2009. Dietary modulation of gut microflora in broiler chickens: a review of the role of six kinds of alternatives to in-feed antibiotics. *Worlds Poult. Sci. J.* 65, 97–114.
- Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W., Hibberd, M.L., Liu, E.T., Rohwer, F., Ruan, Y., 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 4, e3.