BiomeSeq: A Tool for the Characterization of Animal Microbiomes from Metagenomic Data

Kelly A. Mulholland[1] and Calvin L. Keeler, Jr.[1]*

[1] Department of Animal and Food Sciences, University of Delaware, Newark, Delaware, United States of America

*Corresponding author: Department of Animal and Food Sciences
University of Delaware
Newark, DE 19716
ckeeler@udel.edu

Abstract

The complete characterization of a microbiome is critical in elucidating the complex ecology of the microbial composition within healthy and diseased animals. Many microbiome studies characterize only the bacterial component, for which there are several well-developed sequencing methods, bioinformatics tools and databases available. The lack of comprehensive bioinformatics workflows and databases have limited efforts to characterize the other components existing in a microbiome. BiomeSeq is a tool for the analysis of the complete animal microbiome using metagenomic sequencing data. With its comprehensive workflow, customizable parameters and microbial databases, BiomeSeq can rapidly quantify the viral, fungal, bacteriophage and bacterial components of a sample and produce informative tables for analysis. BiomeSeq was employed in detecting and quantifying the respiratory microbiome of a commercial poultry broiler flock throughout its grow-out cycle from hatching to processing. It successfully processed 780 million reads, of which 5,163 aligned to avian DNA viral genomes, 71,936 aligned to avian RNA viral genomes, 469,937 aligned to bacterial genomes, 504,682 aligned to bacteriophage genomes and 1,964 aligned to fungal genomes. For each microbial species detected, BiomeSeq calculated the normalized abundance, percent relative abundance, and coverage as well as the diversity for each sample. BiomeSeq provides for the detection and quantification of the microbiome from next-generation metagenomic sequencing data. This tool is implemented into a user-friendly container that requires one command and generates a table consisting of taxonomical information for each microbe detected as well as normalized abundance, percent relative abundance, coverage and diversity calculations.

Background

Specific and unique animal microbiomes contribute to the biological function of various locations on the body including the gut, skin, vagina, oral cavity, and respiratory tract (Cui *et al*., 2013). Disturbances of these environments by colonization of a new bacteria, eukaryotic virus, or fungi can lead to competition, invasion and replacement. Under appropriate conditions this may result in disease. Advancements in next-generation sequencing technology enable investigations into individual components of the microbiome, thereby gaining insight into the dynamic interactions taking place (Barzon *et al*., 2011). Identification of microbial communities within these environments can aid in elucidating the role they play in both healthy and diseased animals.

Recent studies attempting to characterize the microbiomes of mammals have focused primarily on their bacterial composition, as there are well established and rapid methods of sequencing and analyzing this component (Bond *et al*., 2017; De Boeck *et al*., 2015; Gaeta *et al*., 2017; Glendinning *et al*., 2017; Johnson *et al*., 2018; Shabbir *et al*., 2015). The 16S rRNA gene is commonly used to identify and compare the bacterial genera present in a given sample (Clarridge *et al*., 2004). Accessible bacterial databases, such as Greengenes (DeSantis *et al*., 2006) and Silva (Quast *et al*., 2013), in addition to well-developed bioinformatics workflows are available to facilitate these analyses (Meyer *et al*., 2008; Caporaso *et al*., 2010; Schloss *et al*., 2009). Internal Transcribed Spacer, or ITS, is a widely used fungal genetic marker gene. Similar to 16S rRNA, accessible fungi databases (Kõljalg *et al*., 2013) and bioinformatics workflows for fungal analysis exist (Caporaso *et al*., 2010).

Characterizing the viral component of the microbiome presents unique difficulties. Unlike the ribosomal genes of bacteria and fungi, viruses are heterogeneous in their genetic content and

therefore do not have a conserved genomic region that can be easily sequenced and employed for taxonomic classification (Zou *et al*., 2016). In addition, host DNA contamination has been found to negatively impact the interpretation of results (Daly *et al*., 2015). As a result, there have been fewer efforts to develop comprehensive viral genome databases similar to those available for bacteria (DeSantis *et al*., 2006; Quast *et al*., 2013; Kõljalg *et al*., 2013). Quantification of viral abundance is another limitation with characterizing the virome. Due to the lack of eukaryotic viral genome databases, a sequence-similarity independent approach is often employed to detect eukaryotic viruses, but this approach does not allow for accurate abundance calculations. In addition, many of the available virome bioinformatics tools require the user to possess extensive command-line knowledge and computational resources to successfully install and run the necessary programs and their dependencies on the command line. A user-friendly tool for the analysis of the viral, fungal, bacterial and bacteriophage components is essential to elucidating the complete ecology of a microbiome.

Herein, we present BiomeSeq, a tool for the analysis of complete animal microbiomes from metagenomic data. The BiomeSeq workflow and databases address the challenges of characterizing the eukaryotic virome by including quality filtering and host decontamination, sequence-similarity dependent alignment to microbial reference genome databases and accurate quantification of microbial abundance. It also analyzes the fungal, bacteriophage and bacterial components using the same sequencing data to produce a complete analysis of the microbiome without requiring additional sequencing of the 16S rRNA and ITS genes. Additionally, utilizing shotgun metagenomics to analyze the bacterial and fungal components can increase taxonomic resolution, permit analysis of complete genomes instead of a conserved genomic region, and allow for a comparison of bacteria and fungi to the viral and bacteriophage components (Jovel *et*

al., 2016). BiomeSeq is available as a user-friendly docker container. This versatility allows BiomeSeq to be accessible to users with varied degrees of command-line knowledge and computational resources. While BiomeSeq has been developed and tested on avian species, it can be used to characterize microbiomes of a variety of species.

Implementation

BiomeSeq is currently available as an open-access and user-friendly tool on Docker Hub. As the docker container is self-contained, it simplifies installation and execution by eliminating the need for downloading and installing dependent software and requires only one command. Additionally, BiomeSeq is customizable and allows the user to adjust parameters similar to a command-line tool. Table 1 includes all software and parameters used in BiomeSeq.

BiomeSeq accepts both single- and paired-end reads in fastq format generated by DNA-seq or RNA-seq methods. Along with the fastq file, the user may customize a number of parameters including: the host genome that the sample was derived from, a host-specific viral database, mapping quality threshold, output file name and an output directory. **Figure 1** shows an overview of the BiomeSeq workflow. BiomeSeq generates a table consisting of NCBI RefSeq accession number, microbe name, taxonomy, number of mapped reads in the file, normalized abundance, percent relative abundance, genome coverage for each eukaryotic virus, bacteria, bacteriophage and fungi detected, as well as an alpha diversity calculation for the sample. **Table 2** is an example of an output table for the viral component. Similar tables are generated for bacteria, bacteriophage and fungal data. Visualizations of this data can be easily generated using several different packages in R.

Quality and Decontamination

Individual fastq sequence files are first analyzed for per-base sequence quality, per-sequence quality, sequence length distribution, duplicate sequences, and overrepresented Kmers. Reads with a quality phred score below 30, reads under 100 base pairs and adapter sequences are removed. Quality steps are processed using Trim-Galore (Martin *et al*., 2011). The remaining

reads are then aligned to the user-specified host reference genome using Bowtie2 and only unmapped reads are extracted and analyzed further (Langmead *et al*., 2012). This step removes host genome contamination from the data, increasing analytical efficiency and mapping accuracy (Daly *et al*., 2015).

Databases

The remaining sequencing reads are aligned to an host-specific viral genome database, a bacterial database, a fungal database and a bacteriophage database using the Bowtie 2 alignment algorithm (Langmead *et al*., 2012). Mapping quality threshold default is 20, however this parameter may be customized by the user. The eukaryotic viral genome database currently includes avian-specific viral genomes and was constructed using full genome reference sequences of both DNA and RNA avian viruses obtained from the National Center for Biotechnology Information (NCBI) Virus Database (O'Leary, 2016). The avian DNA viral genomes include 48 viral elements from 9 unique families and the avian RNA viral genomes include 63 viral elements from 13 families. The avian DNA and RNA viral database is organized by the classification of their viral structure and genome organization. DNA viruses are organized hierarchically by whether the virus is double- or single-stranded and whether the virus is enveloped or non-enveloped. RNA viruses are organized hierarchically by whether the virus is double- or single-stranded, negative or positive sense, segmented or non-segmented and whether the virus is enveloped or non-enveloped. The eukaryotic viral genome database will include additional host-specific viruses from a variety of species.

Custom bacterial, fungal and bacteriophage databases were constructed using complete and representative genomes obtained from the NCBI Reference Sequence Database and contain

3,623, 1,281 and 2,212 genomes, respectively (O'Leary, 2016). Each microbial database and corresponding aligner index files can be downloaded from CyVerse. As an additional feature, BiomeSeq also accepts custom microbial databases provided by the user.

Quantification and Output

A sequence similarity-dependent approach for detecting viruses contributes to the rapid detection of known viruses while also allowing for the quantification of biodiversity, which similarity-independent approaches lack (Herath *et al*., 2017). This approach can be applied to bacteria, fungi and bacteriophage as well. For each individual sample, the reads that map to each microbe are normalized based on both microbial and reference genome length per 100,000 host cells using an adaptation to the equation presented by Moustafa and his colleagues in 2017 to quantify viral abundance (Moustafa *et al*., 2017):

$$Microbial\ Abundance = \frac{2\ x\ \frac{number\ of\ reads\ mapped\ to\ microbe\ genome}{microbe\ genome\ size}}{\frac{number\ of\ reads\ mapped\ to\ host\ genome}{host\ genome\ size}}\ x\ 10^5$$

Percent relative abundance is quantified using the following equation:

$$Percent\ Relative\ Abundance = \frac{microbial\ abundance}{total\ microbial\ abundance}\ x\ 100$$

Genome coverage is approximated using the following equation:

$$Genome\ Coverage = \frac{(\#\ mapped\ to\ microbe\ \ x\ \ read\ length)}{microbe\ reference\ genome\ size}$$

149 Alpha diversity for each sample is calculated using the Shannon Diversity Index, a commonly
150 used equation for calculating species diversity in a microbiome as it accounts for both abundance
151 and evenness of the species.

Results and Discussion

A Longitudinal Study of the Microbial Ecology of a Healthy Broiler Flock

Recent studies have identified specific bacterial and viral agents within the respiratory microbiome of both humans and animals that are associated with the severity and spread of disease (Bakaletz *et al*., 1995; Pettigrew *et al*., 2008; De Steenhuijsen Piters *et al*., 2016; Teo *et al*., 2015). In order to understand the complex etiology of a disease and the changes in the microbial ecology of a diseased microbiome, a comprehensive analysis of a healthy microbiome is first required. BiomeSeq was employed to detect and quantify eukaryotic viruses, bacteria, bacteriophage, and fungi in a healthy commercial broiler flock during the grow-out cycle from hatching to processing.

Tracheal swabs were collected at hatching and at weekly intervals through processing at day 50 (8 samples) from an antibiotic-free commercial broiler flock. Both DNA and RNA were isolated and sequencing was performed for each of the eight time points using the Illumina HiSeq platform producing 1 X 100 single-end reads. Each of the resulting 16 samples were processed using BiomeSeq with the following parameters: -g chicken -d avianALL_db -q 20.

In total, BiomeSeq detected 5,163 reads aligned to avian DNA viruses and 71,936 reads aligned to avian RNA viruses. A total of 11 viral species, representing 9 genera and 8 families, were identified from the avian respiratory tract during the grow-out period. This data is represented in a heatmap (**Figure 2**). A total of 469,937 reads were aligned to the bacterial genome database. A total of 533 unique bacterial species were identified, of which 45 had a calculated relative abundance greater than 0.5%. The 45 most abundant species detected extend from 4 phyla, 7 classes, 13 orders, 26 families and 45 genera. This data is represented in a phylogenetic tree

174 generated using the Phytools package in R (**Figure 3**; Revell, 2012). A total of 504,682 reads
175 aligned to the bacteriophage genome database. A total of 30 unique bacteriophage species
176 extended from 1 classified and 1 unclassified order, 4 classified and 1 unclassified families, and
177 5 classified and 4 unclassified genera were identified. This data is represented in a Venn diagram
178 of the common bacteriophage species detected at Week 0, Week 3 and Week 7, generated using
179 the VennDiagram package in R (**Figure 4**; Chen *et al*., 2011). A total of 1,964 reads aligned to
180 the fungal genome database. Sixty-one unique fungal species were identified which extended
181 from 2 phyla, 9 classes, 20 orders, 37 families and 50 genera. This data is represented in a
182 microbial network generated with Cytoscape (**Figure 5**; Shannon *et al*., 2003).

183

184

Conclusions

The complete characterization of a microbiome is critical in elucidating the complex ecology of the microbial composition within healthy and diseased animals. Recent studies have focused on the bacterial component, as there are several well-developed sequencing methods, bioinformatics tools and databases available. The lack of comprehensive bioinformatics workflows and databases have limited efforts to characterize the other components. BiomeSeq is a tool for the analysis of the animal microbiome using metagenomic data. With its comprehensive workflow and custom databases, this tool can rapidly quantify the eukaryotic viral, fungal, bacteriophage and bacterial components of a sample and produces informative tables for analysis. The sequence-dependent approach that BiomeSeq utilizes provides the necessary information required to accurately quantify microbial abundance, genome coverage and diversity. Conversely, this method limits BiomeSeq's ability to perform in de novo microbe discovery. Moreover, a sequence-dependent approach including only representative microbial genomes may underrepresent the abundance of specific microbial strains. To resolve these limitations, BiomeSeq accepts custom microbial databases provided by users which may include microbial genomes derived by other host species and novel microbial sequences.

BiomeSeq was employed in detecting and quantifying the respiratory microbiome of a commercial poultry broiler flock throughout its grow-out cycle from hatching to processing. This study provides the first comprehensive analysis of the ecology of the avian respiratory microbiome and will facilitate future investigations of animal diseases. BiomeSeq is accessible as a container, available as a user-friendly container on Docker Hub.

Abbreviations

NGS: Next generation sequencing

16S rRNA: 16s ribosomal RNA

NCBI: National Center for Biotechnology Information

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

Declarations

Ethics Statement

Not applicable.

Consent for publication

All authors have consented to publication

Availability of data and material

The BiomeSeq Docker container is available at http://dockerhub.com.

BiomeSeq custom databases are available at https://de.cyverse.org.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was supported by Agriculture and Food Research Initiative Competitive Grant #2015-68004-23131 from the USDA National Institute of Food and Agriculture. Computational infrastructure support by the University of Delaware Center for Bioinformatics and Computational Biology Core Facility was made possible through funding from Delaware INBRE (NIH P20 GM103446) and the Delaware Biotechnology Institute.

Authors' contributions

KAM is the primary author of this manuscript. KAM developed the bioinformatics workflow and constructed the animal-specific viral genome database, the fungal database, the bacteriophage database and the bacterial database. KAM wrote all programs for microbial calculations and

233 programs to generate visual representations of microbial data. CLK is the corresponding author
234 of this work. CLK contributed to the design of the work, the acquisition of samples, the analysis
235 and interpretation of the data, and revised and edited the manuscript.

236 **Acknowledgements**

237 We thank Monique Robinson, Sharon Keeler, Hong Li and Daniel Bautista for their
238 contributions in collecting and processing experimental samples and Shawn Polson for his
239 insightful comments and suggestions for this manuscript.

240

241 **References**

242 Bakaletz L. 1995. Viral potentiation of bacterial superinfection of the respiratory tract.
243 Trends Microbiol. 3:110-114.

244 Barzon, L, Lavezzo, E, Militello, V, Toppo, S, Palù, G. 2011. Applications of next-
245 generation sequencing technologies to diagnostic virology. Int J Mol Sci. 12;11: 7861–7884.

246 Bond S, Timsit E, Workentine M, Alexander T, Léguillette R. 2017. Upper and lower
247 respiratory tract microbiota in horses: bacterial communities associated with health and mild
248 asthma (inflammatory airway disease) and effects of dexamethasone. BMC Microbiol.
249 17:184.

250 Caporaso J, Kuczynski J, Stombaugh J, Bittinger K, Bushman F, Costello E, Fierer N, Peña
251 A, Goodrich J, Gordon J, Huttley G, Kelley ST, Knights D, Koenig JE, Ley R, Lozupone C,
252 McDonald D, Muegge B, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh P, Walters W,
253 Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. Qiime allows analysis of high-
254 throughout community sequencing data. Nat Methods. 7:335-336.

255 Chen H and Boutros P. 2011. VennDiagram: a package for the generation of highly-
256 customizable Venn and Euler diagrams in R. Software. 12:35.

257 Clarridge J. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria
258 on clinical microbiology and infectious diseases. Clin Microbiol Rev. 17:840–862.
259 Cui L, Morris A, Ghedin E. 2013. The human mycobiome in health and disease. Genome
260 Med. *5*:63.

261 Daly G, Leggett R, Rowe W, Stubbs S, Wilkinson M, Ramirez-Gonzalez R, Mario C, Bernal
262 W, Heeney J. 2015. Host subtraction, filtering and assembly validations for novel viral
263 discovery using next generation sequencing data. PLoS One. 10:6.

264 De Boeck C, Kalmar I, Dumont A, Vanrompay D. 2015. Longitudinal monitoring for
265 respiratory pathogens in broiler chickens reveals co-infection of Chlamydia psittaci and
266 Ornithobacterium rhinotracheale. J Med Microbio. 64:565-574.

267 De Santis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, Huber T, Dalevi D, Hu
268 P, Andersen G. 2016. Greengenes, a chimera-checked 16S rRNA gene.

269 De Steenhuijsen Piters W, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-Arrabal M,
270 Damien C, Cohen D, Sanders EA, Ramilo O, Bogaert D, Mejias A. 2016. Nasopharyngeal
271 microbiota, host transcriptome, and disease severity in children with respiratory syncytial
272 virus infection. Am J Respir Crit Care Med. 194:1104-1115.

273 Gaeta N, Lima S, Teixeira A, Ganda E, Oikonomou G, Gregory L, Bichalho R. 2017.
274 Deciphering upper respiratory tract microbiota complexity in healthy calves and calves that
275 develop respiratory disease using shotgun metagenomics. J Dairy Sci. 100:1445-1458.
276 Glendinning L, McLachlan G, Vervelde L. 2017. Age-related differences in the respiratory
277 microbiota of chickens. PloS One. 12:11.

278 Herath D, Jayasundara D, Ackland D, Saeed I, Tang S, Halgamuge S. 2017. Assessing
279 species diversity using metavirome data: methods and challenges. Comp Struc Biotechnol. J
280 15:447–455.

281 Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL,
282 Madsen K, Wong G. 2016. Characterization of the Gut Microbiome Using 16S or Shotgun
283 Metagenomics. Front Microbiol. 7:459.
284 Kõljalg U, Nilsson R, Abarenkov K, Tedersoo L, Taylor A, Bahram M, Bates S, Bruns T,
285 Bengtsson-Palme J, Callaghan T, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc
286 T, Griffith G, Hartmann M, Kirk P, Kohout P, Larsson E, Lindahl B, Lücking R, Martín M,
287 Matheny P, Nguyen N, Niskanen T, Oja J, Peay K, Peintner U, Peterson M, Põldmaa K, Saag
288 L, Saar I, Schüßler A, Scott J, Senés C, Smith M, Suija A, Taylor D, Telleria M, Weiss M,
289 Larsson K. 2013. Towards a unified paradigm for sequence-based identification of fungi.
290 Mol Ecol. 22:5271-5277.

291 Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods.
292 9:357-359.

293 Li H and Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler
294 Transform. Bioinformatics. Epub.

295 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
296 reads. EMBnet. J. 17:10–12.

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

301 Moustafa A, Xie C, Kirkness E, Biggs W, Wong E, Turpaz Y, Bloom K, Delwart E, Nelson
302 K, Venter J, Telenti A. 2017. The blood DNA virome in 8,000 humans. PLoS Pathog.
303 13:e1006292.

304 O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse
305 B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V,
306 Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher
307 E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy
308 MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS,
309 Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W,
310 Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016.
311 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and
312 functional annotation. Nucleic Acids Res. 4:D733-45.

313 Pettigrew M, Gent JF, Revai K, Patel J, Chonmaitree T. 2008. Microbial interactions during
314 upper respiratory tract infections. Emerg Infect Dis. 14:1584-1591.

315 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner F. 2013.
316 The SILVA ribosomal RNA gene database project: improved data processing and web-based
317 tools. Nucl Acids Res. 41:590-596.

318 Revell L. 2012. Phytools: An R package for phylogenetic comparative biology (and other
319 things). Methods Ecol. Evol. 3:217-223.

320 Shabbir M, Malys T, Ivanov Y, Park J, Shabbir M, Rabbani M, Yaqub T, Harvill E. 2015.
321 Microbial communities present in the lower respiratory tract of clinically healthy birds in
322 Pakistan. Poultry Sci. 94:612-620.

323 Shannon P, Markiel A, Ozier O. 2003. Cytoscape: a software environment for integrated
324 models of biomolecular interaction networks. Genome Res.13:2498–2504.

325 Schloss P, Westcott S, Ryabin T, Hall J, Hartman M, Hollister E, Lesniewski R, Oakley B,
326 Parks D, Robinson C, Sahl J, Stres B, Thallinger G, Van Horn D, Weber C. 2009.
327 Introducing mothur: Open-source, platform-independent, community-supported software for
328 describing and comparing microbial communities. Appl Enviro Microbiol. 75:7537-7541.

329 Teo D, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt B, Hales B, Walker M, Hollams
330 E, Bochkov Y, Grindle K, Johnston S, Gern J, Sly P, Holt P, Holt K, Inouye M. 2015. The
331 infant airway microbiome in health and disease impacts later asthma development. Cell Host
332 Microbe. 17:704-715.

333 Zou S, Caler L, Colombini-Hatch S, Glynn S, Srinivas P. 2016. Research on the human
334 virome: where are we and what is next. Microbiome. 4:32.

335

Table 1. Software tool and parameters used in BiomeSeq

Process	Tool Name	Parameters
Quality	Trim Galore	
Decontamination	Bowtie 2	-x -S
	Samtools	view -bS
Microbial Database Alignment	Bowtie 2	-x -S
	Samtools	view -bSq [user input]

Table 2. Example table generated by BiomeSeq of the viral component of a commercial poultry flock at Week 6.

Ref Seq Number	Name	Taxonomy	Genome Size	Number Mapped	Norm. Abundance	Relative Abundance	Genome Coverage	Diversity
NC002229	Gallid Alphaherpesvirus 2	Double Stranded; Enveloped; Herpesviridae; Mardivirus	177874	1	27.38	0.10%	0	0.534
NC002577	Gallid Alphaherpesvirus 3	Double Stranded; Enveloped; Herpesviridae; Mardivirus	164270	1	29.65	0.10%	0	
NC015396	Avian Gyrovirus	Single Stranded; Non-Enveloped; Circoviridae; Gyrovirus	2383	72	147165.92	15.47%	3.05	
NC001720	Fowl Aviadenovirus	Double Stranded; Non-Enveloped; Adenoviridae; Aviadenovirus	43804	4560	507048.92	53.30%	10.51	

Figure 1. BiomeSeq Workflow.

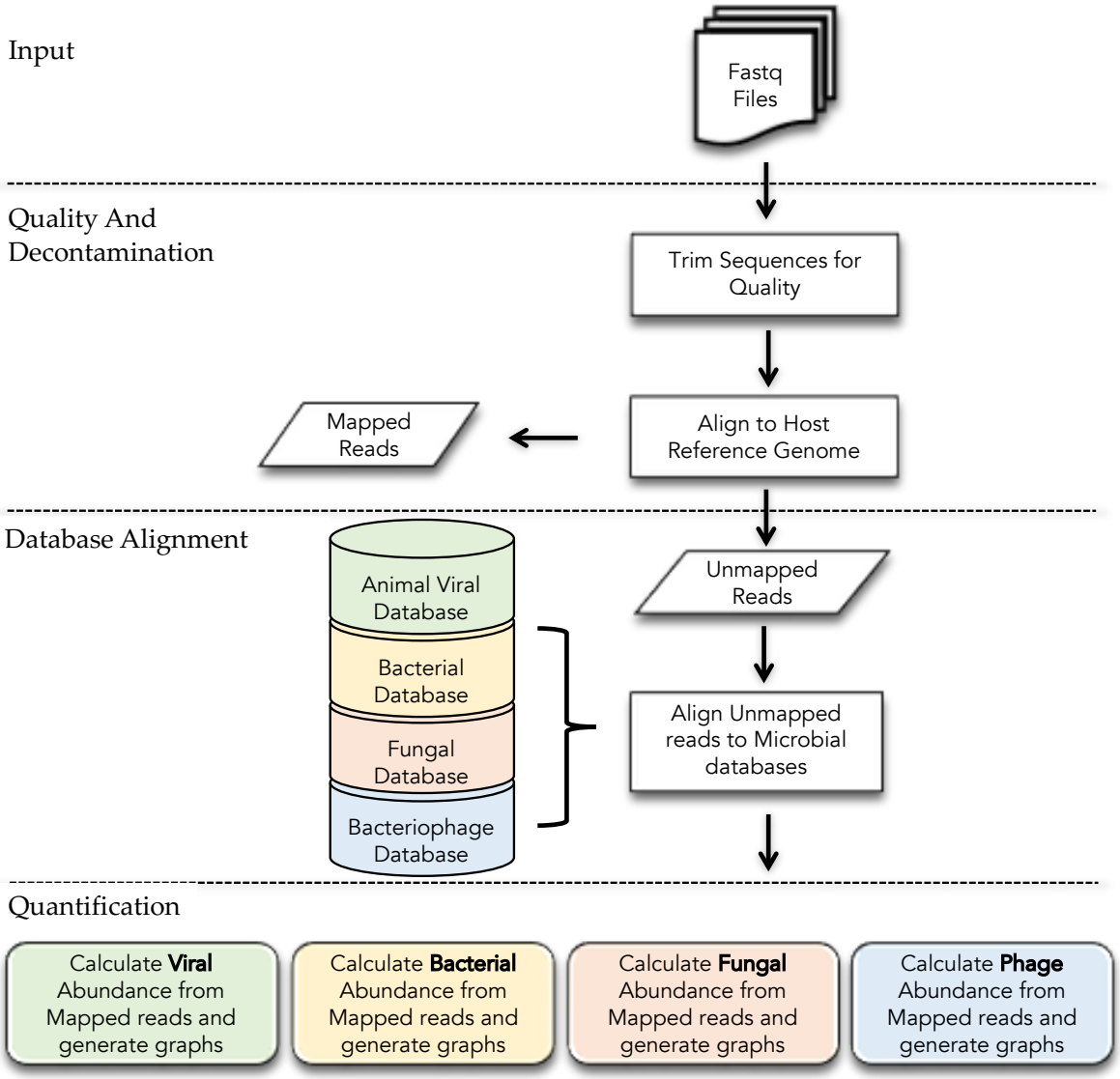

Figure 2. Heatmap consisting of each virus species identification and abundance in a commercial poultry flock from hatching to processing. Color corresponds to the range of relative abundance of each family from 0 to 100%. Green: 0-1%; yellow: 1-25%; orange: 25-75%; and red: 75-100%. The sum of each column, or week, is 100%.

Nucleic Acid Type	Strand	Sense	Enveloping	Family	Genus	Species	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
DNA	double stranded		enveloped	Herpesviridae	Iltovirus	Gallid alpha herpesvirus 1	0.171							
					Mardivirus	Gallid alphaherpesvirus 2&3				0.257	0.145	0.070	0.006	
						Meleagrid alphaherpesvirus 1			0.037		0.004			
			non-enveloped	Anelloviridae	Gyrovirus	Avian gyrovirus					88.664	12.054	15.469	36.773
	single stranded			Adenoviridae	Aviadenovirus	Fowl aviadenovirus							53.299	6.698
RNA	single stranded	negative	enveloped	Birnaviridae	Avibirnavirus	Infectious bursal disease virus					0.008		2.333	0.105
		positive	enveloped	Coronaviridae	Gammacoronavirus	Avian infectious bronchitis virus	0.382	54.762	58.947	16.278	1.884	23.602	21.786	19.319
				Retroviridae	Alpharetrovirus	Avian carcinoma virus						0.077		0.042
					Unclassified	Avian Endogenous Retrovirus	99.447	44.493	41.017	83.296	9.290	64.196	7.108	37.063
			non-enveloped	Astroviridae	Avastrovirus	Chicken astrovirus		0.744						
				Picornaviridae	Sicinivirus	Chicken sicinivirus JSY				0.169	0.005			

Figure 3. Phylogenetic tree of bacterial species detected in a commercial poultry flock.
Branches extend from phylum to species. Nodes indicate detected species and diameter indicates average abundance.

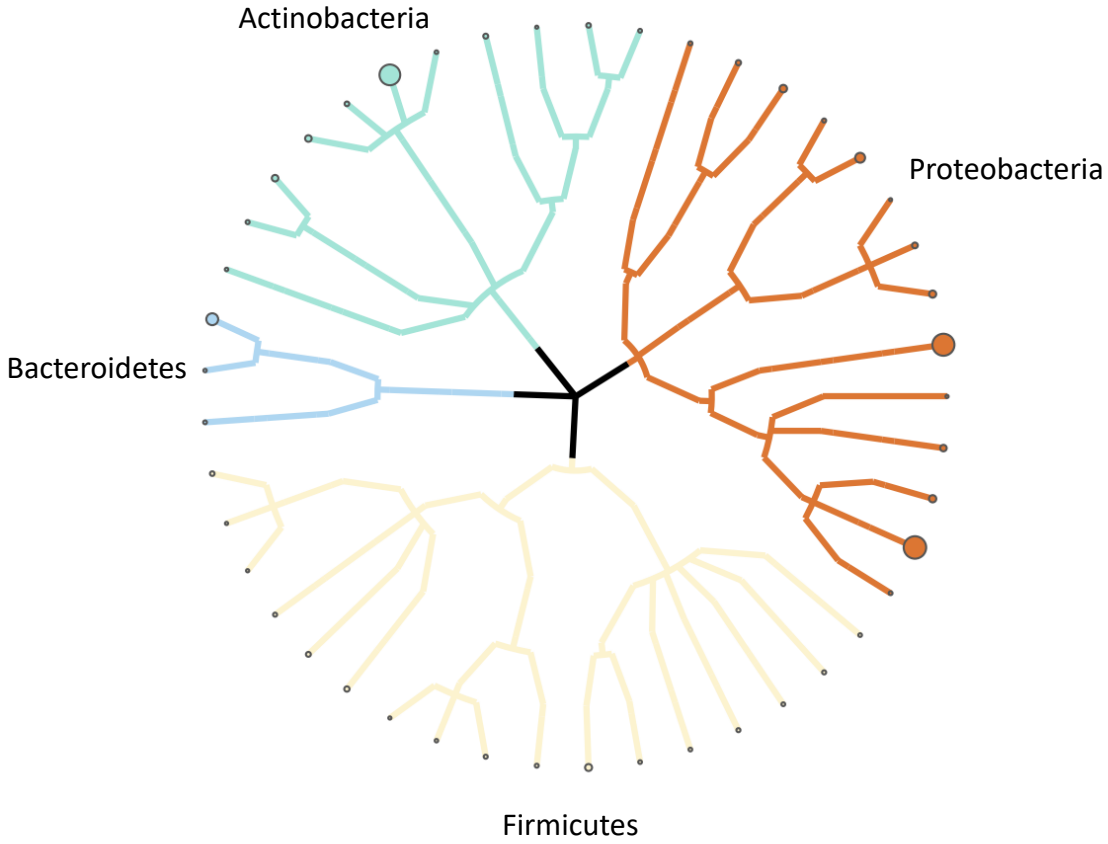

Figure 4. Venn Diagram of the detected bacteriophage species in a commercial poultry flock at Week 0, Week 1 and Week 7.

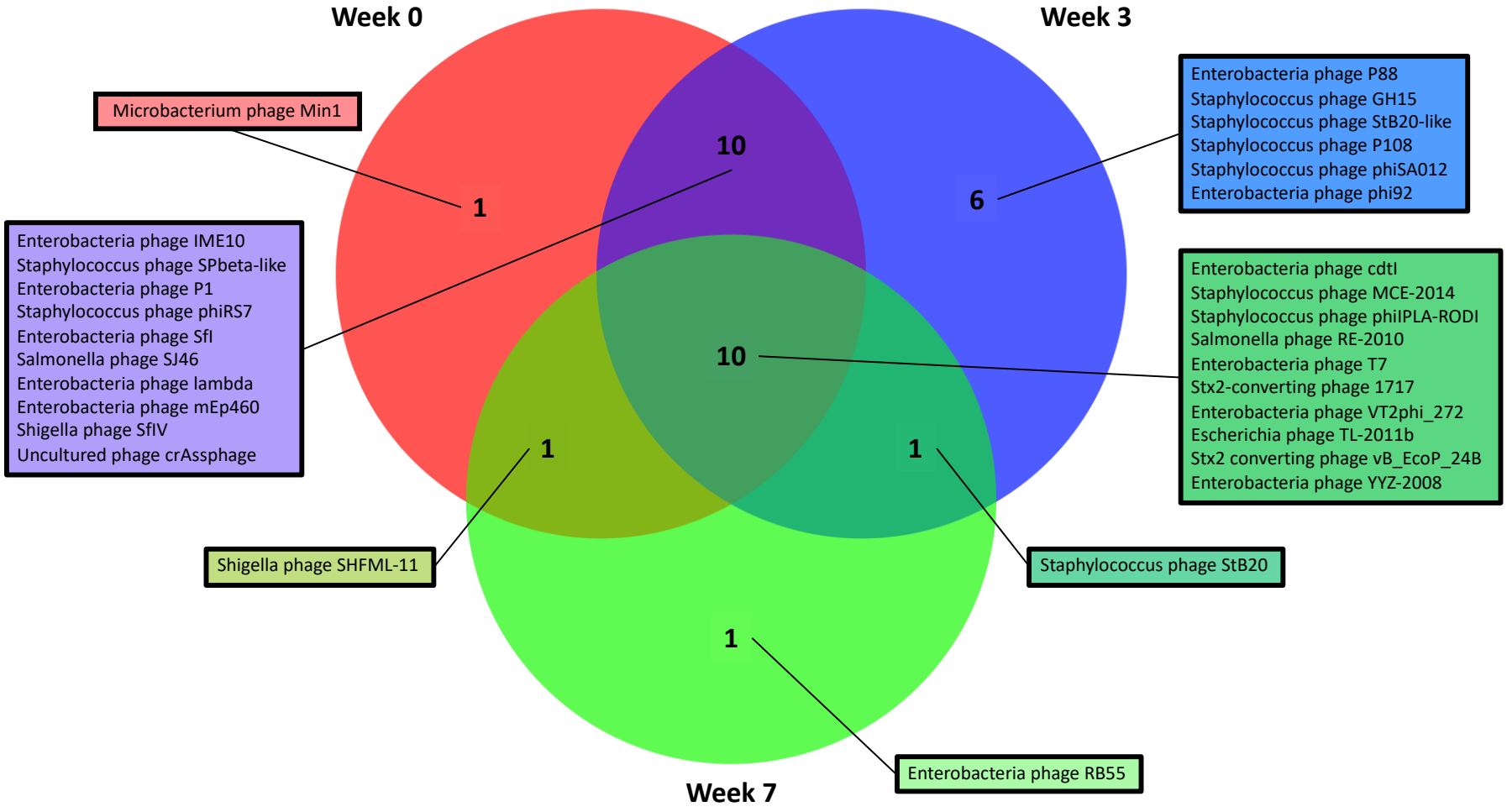

Figure 5. Network of fungal species detected in a commercial poultry flock. Outer nodes represent order level, while inner nodes represent species. Diameter of the inner nodes correlate to species frequency, or the number of weeks the species was detected.