# Molecular characterization of faecal RNA virome of healthy chickens using next-generation sequencing

**Vivian Chiamaka Nwokorogu**

**(22176492)**

**Submitted in complete fulfilment of the requirements for the degree of**

**Master of Applied Science**

**in**

**Biotechnology**

**Department of Biotechnology and Food Science**

**Faculty of Applied Sciences**

**Durban University of Technology**

**Durban, South Africa**

**Supervisor: Prof. Saheed Sabiu**

**Co-supervisor: Prof. Santhosh Kumar Kuttan Pillai**

**Co-supervisor: Prof. Martin Munene Nyaga**

**August 2023**

# REFERENCE DECLARATION

I, Ms Nwokorogu Vivian Chiamaka – Student number 22176492, Prof Sabiu Saheed, Prof Santosh Pillai, and Prof Martin Nyaga do hereby declare that in respect of the following dissertation:

Title**: Molecular characterization of faecal RNA virome of healthy chickens using next-generation sequencing**

1. To the best of our knowledge no other dissertation of a similar nature exists

2. All references mentioned in the dissertation are complete in terms of all personal communications engaged in and published works consulted.

| | |
|---|---|
| | **12/08/2023** |
| **Signature of student** | **Date** |
| | 13/08/2023 |
| **Signature of Supervisor** | **Date** |
| | 13 . 08. 2023 |
| **Signature of co-supervisor** | **Date** |
| | 12-August-2023 |
| **Signature of co-supervisor** | **Date** |

# AUTHORS DECLARATION

The present study presents the original work of the author. This dissertation has not been previously submitted to any other academic institution in any form or for any degree. whenever the work of others was utilized, it was duly acknowledged within the text as citations. The research presented in this dissertation was carried out at Durban University of Technology's Department of Biotechnology and Food Science, Faculty of Applied Sciences, under the esteemed supervision of Prof. Sabiu Saheed, Prof. Santosh Pillai and Prof. Martin Nyaga.

**12/08/2023**

**Signature of student** **Date**

# DEDICATION

This work is dedicated to my dearest son, Godwin Clinton Munachimso, my lovely husband, Onwuka Godwin Chukwudi and my super Dad, Prince Tony Okezie Nwokorogu for their love and immeasurable support.

# ACKNOWLEDGEMENTS

I am extremely grateful to my esteemed supervisor, Prof. Saheed Sabiu, for his unwavering support, scholarly mentorship, invaluable counsel and for availing all necessary resources for this virological project. I express my profound gratitude to you. I would like to specially extend my sincere gratitude to my esteemed co-supervisors, Prof. Santosh Pillai and Prof. Martin Nyaga for your relentless support, academic guidance, and the privilege of an opportunity to do this project. The knowledge and skills I have gained on the journey towards this completing this project has positive reshaped my understanding of a variety of concept. Thank you so much for your patience, guidance, tolerance, constructive criticism and words of encouragement.

My sincere gratitude to Dr. San Emmanual James from Krips unit, Nelson Mandela University, for his technical assistance.

I would like to thank the entire members of Prof. Santhosh's enzyme research group and Prof. Saheed's Computational and Systems Biology (D3VM) research group, in the department of Biotechnology and Food science, Durban University of Technology, for being friendly and supportive. My sincere gratitude to Dr. Charlene Pillay for your immense assistance with sample collection. I am indeed grateful to Dr. Prashant for the numerous times you have allowed steal into your time with my unresolved questions.

I would also like to thank the entire members of University of the Free State next-generation sequencing unit (UFS-NGS-Unit) for being so supportive and for training me in different capacities during my stay with you all. Dr. Emmanuel, Dr. Peter, Mr. Milton, Ms. Hlengiwe and Ms. Robyn I am so grateful for your assistance, encouragement and reassurances.

A special thank you to all the friends that I have made in the Biotechnology and Food Science department, for all the friendliness, motivation and support. I am extremely grateful to all my friends and colleagues, both near and far. Ms. Grace, Mr. Dayo, Dr. Deji, Mrs. Betty, Mrs. Nonye, Mr. Jamiu, Dr. Kabange, and Dr. Stanley, there are way too many names to mention. It is indeed a pleasure to get to know everyone.

Also, I would like to express my heartfelt gratitude to my dearest husband Godwin, my lovely parents, my son Clinton, Uncle Tony, Blessing, Sandra and Promise for your unconditional love, reassurances, immense support and for always believing in me and praying for me.

I would like to acknowledge the financial support of National Research Foundation (NRF) Research Development Grant for rated researchers (Grant number 120433) and the supportive grant by the Directorate of Research and Postgraduate Support, Durban University of Technology awarded to Prof. Saheed Sabiu.

I would also like to thank and acknowledge the Directorate of Research and Postgraduate Support, Durban University of Technology Scholarship Scheme, for providing me scholarship/bursary throughout my study.

# RESEARCH OUTPUTS

**Manuscript under review**

Vivian C. Nwokorogu, Santhosh Pillai, James E. San, Charlene Pillay, Martin M. Nyaga, Saheed Sabiu. A metagenomic investigation of the gut RNA virome structure of asymptomatic chickens obtained from Durban, KwaZulu-Natal province, South Africa. *Virus Research*, 2023.

**Manuscripts under processing for submission**

i. Vivian C. Nwokorogu, Santhosh Pillai, Martin M. Nyaga, and Saheed Sabiu. A snapshot of mNGS based analysis RNA viral diversity in bird species: A systematic review. *Viruses*, 2023
ii. Vivian C. Nwokorogu, Santhosh Pillai, Martin M. Nyaga, and Saheed Sabiu. Discovery of a novel astrovirus identified from broiler chicken in southern Africa. *Plos Pathogens*, 2023
iii. Vivian C. Nwokorogu, Santhosh Pillai, Martin M. Nyaga, and Saheed Sabiu. Emergence and molecular characterization of an *Avihepevirus magniiceur* virus isolated from chickens in the eastern coast of South Africa, a zoonotic threat. *Journal of Viral Hepatitis*, 2023.

**Conference presentation**

Vivian C. Nwokorogu, Santhosh Pillai, Martin M. Nyaga, and Saheed Sabiu. Metagenomic detection and characterization of RNA viruses in faecal samples from South African chickens. Faculty Research Day, Faculty of Applied Sciences, Durban University of Technology, Durban, South African, 17 November 2022.

# ABSTRACT

The incidence of emerging and re-emerging diseases has been on the rise, affecting both wild and domestic animals. Globally, it is noteworthy that major disease outbreaks that have caused significant morbidity and mortality in poultry systems, other animal species and human populations, have been attributed to viruses originating from animals including birds. Some of these viral outbreaks, especially those characterized by highly unstable RNA genomes have escalated into epidemics or even pandemics. Instances of RNA viral outbreaks, notably associated with animal origins, include severe acute respiratory syndrome coronavirus 2 (SARS-CoV 2), Ebola, Swine flu, and Middle East respiratory syndrome (MERS), Spanish flu, Asian flu, and Hong Kong flu. Globally, substantial losses in poultry, have been attributed to RNA viral-linked infections including Newcastle disease, avian influenza, avian leukosis, Gumboro disease, bronchitis, and acute enteritis. In South Africa, the poultry industry has remained the largest agricultural sector, with significant contribution to the nation's gross domestic product from proceeds of poultry meat and eggs. Interestingly, chicken is the most farmed poultry bird in South Africa and a major source of protein consumed across all income classes. As a result of the increasing demand for chicken in South Africa, its consumption has outweighed its local production, leading to importations. Though the nation's chicken production strives to remain competitive for its growing demand, however, this goal has been threatened by the rising production cost and infectious disease outbreaks including those of viral origin in flocks. Therefore, it has remained imperative to carry out an in-depth evaluation of these factors, particularly infectious diseases that are associated with suboptimal performance, lowered productivity, and chicken mortality in poultry production.

The productive performance and feed utilization rate of chickens are significantly impacted and reliant upon the state of health and proper functioning of their gastrointestinal tract (GIT). Chicken's GIT is the site of metabolism and may contain diverse microorganisms including fungi, bacteria and viruses whose composition and abundance vary remarkably across its growth stages. Among these organisms, viruses have been implicated in major infectious diseases leading to seasonal culling of poultry birds. These viral diseases cause low productivity in chickens due to immune suppression, subclinical growth impedance, and malabsorption. Chicken flocks are homogenous, often crowded and possess similar genetic features, leaving them vulnerable to viral infections. Thus, with infected birds being initially asymptomatic and the viruses unidentified, these viruses spread rapidly causing outbreaks leading to substantial colossal losses. Poor GIT health, even in the absence of a recognized disease state, can affect poultry performance and result in low productivity. Unfortunately, studies on

the GIT of farm animals and birds are relatively scarce, from the African continent, though there are a few studies from other continents available with information on the prokaryotic microbiome of birds GIT, with chickens being more studied because of its economic importance. Nevertheless, virome studies on birds including chickens are relatively few, despite being implicated in major outbreaks.

Viruses, unlike bacteria lack a universal gene marker for identification and the low amount of their nucleic acid in biological samples makes their identification difficult. Importantly, studies have characterized one or a few of these known viruses using non-NGS molecular methods. However, these approaches do not represent the occurrence of these viruses in natural settings and ignore certain factors such as virus-virus interactions, bird age, host taxonomy and community structure dynamics which have been shown to influence the emergence and abundance of viruses in birds. In addition, while non-NGS methods effectively study each viral species or fewer viruses under experimental settings, they are flawed by the limitation of characterizing only known viruses. Hence, characterizing the viruses present in the GIT of chickens using high-throughput technologies, has remained important to determine the key viral agents associated with poultry infectious outbreaks.

The use of viral metagenomics through the NGS approach has allowed the investigation of viruses including novel viruses in animal samples and birds, regardless of the sample type. This approach offers a combined advantage of speed and high-throughput recovery of viruses. While the information on the virome composition of African birds is scarce, the data on their RNA virome including chickens are even scantier despite the continuous evolution of RNA viruses and their associated disease outbreaks. Therefore, it has become paramount to characterize RNA viruses in chicken's GIT using metagenomic NGS (mNGS). South Africa being the highest poultry producing country in the African continent has been plagued by many seasonal outbreaks of RNA viral diseases in flocks. Hence, determining the complex RNA viral constituents present in the GIT of South African chickens is imperative. In this study, the diversity and abundance of the total RNA viruses found in healthy South African chickens was studied using the mNGS technique.This was achieved through optimized enrichment strategies for better virus recovery using the Illumina Miseq sequencing. The use of Novel Enrichment Technique of VIRomes (NetoVIR) standardized sample preparation protocol, whole transcriptome amplification (WTA) and QIAseq FX library preparation method while using the non-invasive faecal sampling method. The effect of age (2, 4 and 7 weeks) and seasons (winter and summer) were studied as factors that may modulate the abundance and/or diversity of viruses in the GIT of chickens. This was achieved using established ecological metrics of alpha and beta

diversities and their result was statistically evaluated. In addition, the evolutionary relatedness of some of the identified viruses were explored using phylogenetic analysis.

The results obtained from the RNA virome investigation of 10 asymptomatic, commercially bred South African chickens revealed a total of 48 RNA viral species. The identified viruses spanned across 11 orders, 15 families and 21 genera. The viral families such as *Coronaviridae, Picornaviridae, Reoviridae, Astroviridae, Caliciviridae, Picorbirnaviridae* and *Retroviridae* were the most abundant. Among these families, picornaviruses, reoviruses, astroviruses, picobirnaviruses and coronaviruses were most prevalent at 100%, 88.9%, 81.5%, 81.5% and 74% occurrence across the 27 samples, respectively. Specifically, virus genera such as *Rotavirus, Orthoreovirus, Gammacoronavirus, Sicinivirus* and *Megrivirus* relatively prevailed in the 2 weeks faecal samples regardless of season. Significantly, *Rotavirus G* and *Avian Orthoreovirus* with high abundance observed at 2 weeks, drastically decreased by the 7th week of development and this may be attributed to their stable, fully developed immune system compared to their juvenile stages. Furthermore, the complete genome of novel chicken astroviruses (CAstV) and genomes of many previously known viruses, including pathogenic avian viruses, mammalian, fungal and plant viruses were identified in this study. Additionally, results from the investigated factors (age and season), showed that there was no effect on viral shedding within samples in a group (alpha diversity) for age ($P = 0.146$) and season ($P = 0.242$), which was contrasting to beta diversity (between groups) metrics that indicated that viral diversity and abundance was significantly influenced by age ($P = 0.01099$) and season ($P = 0.00099$). More viruses were abundant in the 2 weeks and 4 weeks samples, while for the two seasons, the winter samples had more viruses. Interestingly, for age, this outcome could be attributed to the higher viral susceptibility of chickens at juvenile and intermediate ages as a result of their weaker, still developing immune system while for season, it could be deduced that due to temperature differences of the two seasons, more viruses thrive at winter compared to summer season. Furthermore, the outcomes of the viral evolutionary relatedness demonstrated global distribution and distinctiveness in terms of some specific genotypes or virus lineages for identified viruses.

Taken together, the results obtained from this study show that viral structure in the GIT of South African chickens are diverse. It was noted that chickens might carry pathogenic viruses even in the absence of an observable disease condition where pathogenesis may be triggered under certain conditions. Furthermore, the relative abundance profiles of specific avian viruses may be dependent on the age of the bird investigated. Based on the samples analysed, the overall GIT viral abundance in chickens within the same group may be homogenous. However, the viral diversity and abundance

of chickens GIT may vary between different chicken groups characterised by distinct features, for instance, age and season, provided other underlining nutritional and environmental factors are considered. Undoubtedly, based on the chicken faecal samples studied and the diverse viruses recovered/characterized, mNGS has proven to be a valuable tool for effectively studying the virome in the GIT of avian chickens.

Overall, this viral metagenomic study offers some insights into the diversity and composition of RNA viruses circulating in commercially bred chickens in South Africa and this information would be helpful towards understanding the key RNA viruses present in chicken's GIT at early, intermediate and mature stages of growth. In addition, this study has provided baseline data that will be handy for research endeavours aiming to compare RNA virome structure between healthy and diseased chickens. The identification of some pathogenic viruses in apparently healthy/asymptomatic chickens provides information that may be beneficial for further epidemiological studies looking to decipher the transition dynamics of gut viruses in chicken host from being asymptomatic carriers to diseased condition, aimed at averting illnesses and improving chicken gut health. This is a significant stride towards better preparedness for emerging or reoccurring viral infections from chickens in South Africa and beyond.

**Keywords:** Viral metagenomics, Faecal virome, Gastrointestinal tract, RNA viruses, Next-generation sequencing, Poultry, Chicken, Zoonosis, Viral diversity and relative abundance.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

µg: Microgram

µl: Microlitres

16S: 16 Svedberg ribosomal ribonucleic acid

aa: Amino acid

AREC: Animal Research Ethics Committee

ATP: Adenosine triphosphate

BLAST: Basic local alignment search tool

BP: Base pair

cDNA: Complementary DNA

CDS: Coding sequence

DdDp: DNA-dependent DNA polymerase

DNA: Deoxyribonucleic acid

dNTP: Deoxynucleoside triphosphate

dsDNA: Double-stranded deoxyribonucleic acid

dsRNA: Double-stranded ribonucleic acid

EB: Elution buffer

EDTA: Ethylenediamine tetra acetic acid

Gb: Giga base

gDNA: Genomic DNA

GDP: Gross domestic product

GIT: Gastrointestinal tract

ICTV: International Committee on Taxonomy of Viruses

ITS: Internal transcribed spacer

Kb: Kilobase

MEGA: Molecular Evolutionary Genetics Analysis

MERS: Middle East respiratory syndrome

$MgCl_{2:}$ Magnesium chloride

mNGS: Metagenomic next-generation sequencing

MUSCLE: MUltiple Sequence Comparison by Log- Expectation

NaOH: Sodium hydroxide

NCBI: National Center for Biotechnology Information

NetoVIR: Novel techniques of enrichment techniques of viromes

ng: Nanogram

NGS: Next-generation sequencing

Nr: Non-redundant

nt: Nucleotide

OIE: Office International des Epizooties

ORF: Open read frame.

PBS: Phosphate buffer saline

PCR: Polymerase chain reaction

pH: Potential of hydrogen

QC: Quality control

qPCR: quantitative polymerase chain reaction

RdRp: RNA-dependent RNA polymerase

RefSeq: Reference Sequence data

RNA: Ribonucleic acid

RNase: Ribonuclease

Rpm: Revolutions per minute

rRNA: Ribosomal RNA

RSB: Resuspension buffer

RT-PCR: Reverse transcriptase polymerase chain reaction

SAPA: South African poultry association

SARS-CoV: severe acute respiratory syndrome coronavirus 2

ssRNA: Single-stranded ribonucleic acid

SOLiD-ABI: Supported Oligonucleotide Ligation and Detection - Applied Biosystems

WHO: World health organization

WTA: Whole transcriptome amplification

β-ME: Beta mercaptoethanol

# CHAPTER 1. Introduction

## 1.1 Background information

Emerging and re-emerging diseases continue to take the upward trend with both wild and domestic animals implicated. It is noteworthy that over 60% of identified infectious agents are of zoonotic nature and comprise a diverse range of pathogenic microorganisms, including bacteria, parasites, protozoans, fungi, and viruses, among other types (Rahman *et al.,* 2020). Domestic animals such as livestock, pets, and poultry may frequently transmit pathogens to humans because of their close interactions (Delahoy *et al.,* 2018; Maurin *et al.,* 2021; Shi *et al.,* 2021b). Globally, most of the emerging viruses, including highly pathogenic avian influenza virus (H5N1 and pandemic H1N1), Ebola virus, Nipah virus (Hauser *et al.,* 2021), Zika virus, Marburg virus (Sah *et al.,* 2022), SARS virus and SARS-Cov-2 that have caused severe infections in humans are from animals, including chickens (Uyeki *et al.,* 2022; Saxena *et al.,* 2023). Chicken is a major source of animal protein for human consumption, with an approximate annual production of 122 million tonnes globally (Sood *et al.,* 2020; Whitton *et al.,* 2021). The increased demand for chicken meat and egg products has resulted in its early marketable age. In addition, due to its quick financial returns to poultry farmers, with high-profit margins have made chicken the most preferred bird for poultry farming (Long *et al.,* 2017).

According to the South African Poultry Association (SAPA), the poultry industry is the largest agricultural sector in South Africa, contributing about 16.60% of all agricultural products' gross domestic product (GDP), 39.90% of animal product GDP, and an annual income estimate of about 50.96 and 11.44 billion Rands for meat and eggs, respectively (SAPA 2021). In addition, poultry products are the most consumed source of animal protein in South Africa. However, the consumption of poultry products is far greater than its local production (Nkukwana 2018; SAPA 2019). As a result, the consumption demand for poultry meat in South Africa has increased from 1.88 million tonnes in 2020 to 1.90 million tonnes in 2021 (SAPA 2021). Despite this increased demand, the South African poultry industry still fights to remain competitive as its profit margins are pressured by rising production costs (Lubinga *et al.,* 2018; Nkukwana 2018). These costs are primarily attributed to the increased cost of grains (Nkukwana 2018), importations, additional expenses on electricity due to surging loadshedding levels, and the emergence of infectious disease outbreaks in flocks. Hence, it has remained crucial to critically examine these factors, particularly infectious diseases that are implicated in underperformance, low productivity, and mortality in poultry production.

Globally, chicken's productive performance and feed conversion rate are greatly influenced and dependent on the state of health and proper functioning of its gastrointestinal tract (GIT), which is the site of nutrition, metabolism, and build-up of diverse microbes (Shah *et al.,* 2016). Chicken's GIT is naturally colonized by microorganisms like bacteria, fungi, and viruses upon hatch which could be harmless, symbiotic, or pathogenic (Ding *et al.,* 2020). Therefore, poor GIT health, even in the absence of a recognized disease state, can affect poultry performance and result in low productivity (Day *et al.,* 2015). Chicken flocks are frequently dense and possess homogeneous gene traits, leaving them vulnerable to breakouts of viral infections. Hence, even on high-quality poultry farms, a wide range of viruses can accumulate, especially if they are multi-age and since they are often asymptomatic and unidentified, they spread quickly causing substantial economic losses (Akinyemi *et al.,* 2020).

It is noteworthy that both RNA and DNA viruses cause infections in poultry birds. However, while DNA viruses such as adenoviruses (Ishag *et al.,* 2022; Lebdah *et al.,* 2022) and parvoviruses (Pradeep *et al.,* 2020; Cui *et al.,* 2022) identified in chickens have been associated with enteritis, studies have shown that RNA viruses constitute a greater proportion of all infections caused by viruses (King *et al.,* 2018; Zhang *et al.,* 2018; Zhang *et al.,* 2019; Wells *et al.,* 2020). This has been attributed to their highly unstable nature, mutations, and errors involved in their replication leading to their re-assortments, formation of diverse variants, and subtypes (Naguib *et al.,* 2017). Studies have also shown that RNA viruses are implicated in most mild and/or severe infectious disease outbreaks in humans and animals including chickens (Chowdhury *et al.,* 2021; Kang *et al.,* 2021; OIE 2021; Shi *et al.,* 2021b; Kwok *et al.,* 2022; WHO 2023). Some of these RNA-linked viral infections in chickens include; Newcastle disease (NCD) (Ogali *et al.,* 2018), avian influenza (Kariithi *et al.,* 2020; Sakuma *et al.,* 2021), infectious laryngotracheitis (Carnaccini *et al.,* 2022), infectious bronchitis (IB) (Wang *et al.,* 2020), infectious bursal disease (IBD) (Fan *et al.,* 2019), and enteritis (Xia *et al.,* 2023). These infectious viral outbreaks may often result from, low biosecurity measures in live bird markets, interactions between infected birds, and intensified poultry systems. Consequently, these infections have resulted in low productivity, malabsorption, enteritis, immune suppression, and other infections in poultry (Kock and Heuer 2019; Johnson *et al.,* 2020) leading to colossal economic losses as well as substantial morbidity and mortality globally (Fasanmi *et al.,* 2017). In addition, avian RNA viral outbreaks asides from causing significant financial losses in poultry are detrimental to human health; for instance, from January 2003 to January 2023, about 868 human infections of HPAIV (H5N1) resulting in 457 deaths, have been reported globally (WHO 2023). Hence, characterizing the RNA

viruses in chickens will lead to a greater knowledge of key viral agents associated with poultry infectious outbreaks.

To date, the use of faecal samples has been reported as an ideal specimen in animal studies as it allows continuity, temporal analysis, and tracking of the digestive tract microbiome at different stages without killing the animal (Stanley *et al.,* 2015; Videvall *et al.,* 2018). Previous investigations through next-generation sequencing (NGS) have reported faecal virome compositions of both healthy and diseased animals. For example, in a comparative study on diarrheic and non-diarrheic piglets, an average of 4.2 and 5.2 mammalian viruses was shed in the faeces of non-diarrheic and diarrheic piglets, respectively (Theuns *et al.,* 2016). A more specific study, which compared breeder source and age, in the development of intestinal RNA viruses in healthy chickens, using stool samples observed that *Astroviridae, Reoviridae* were most prevalent at 2 weeks whereas *Picornaviridae, Flaviviridae,* and other unclassified viruses showed increased abundance with age (Shah *et al.,* 2016). In another related study, the *Caliciviridae, Adenoviridae, Parvoviridae, Picornaviridae, Picobirnaviridae, and Reoviridae* were the major viral families identified from faecal samples of Brazilian chickens (Lima *et al.,* 2017). In addition, a more recent study in the Netherlands demonstrated that close interaction of humans with chickens and its farm dust facilitated zoonosis and also identified picornaviruses, parvoviruses, caliciviruses, and novel chicken astrovirus using NGS (Kwok *et al.,* 2022).

Worldwide, NGS, a high throughput sequencing method has allowed unprecedented advances in the characterization of complex microbial communities including viruses (Shah *et al.,* 2016; Kang *et al.,* 2021). The NGS approach offers the combined advantages of speed, sensitivity, automation, and high-throughput deep sequencing and has successfully been used to characterize faecal microbiota of animals such as bats (Phan *et al.,* 2018), cats (Yamanaka *et al.,* 2020), dog (Shi *et al.,* 2021b), pigs (Theuns *et al.,* 2016), wild birds (Chang *et al.,* 2020; Vibin *et al.,* 2020a) and chickens (Patzina-Mehling *et al.,* 2020; Kwok *et al.,* 2022). In addition, the revolutionization of viral metagenomics concerning epidemiological studies has allowed credible, faster, better detection and surveillance of multifunctional viruses in poultry, which is critical for the sustainability of healthy chicken production globally (Malik *et al.,* 2018; Yadav *et al.,* 2019). Decades of avian research have focused on zoonotic viral pathogens, or viruses causing significant economic losses in poultry while overlooking other RNA viruses which also constitute the avian virome. With the continuous evolution of highly unstable RNA viruses and quasi species emergence, these undetected, unidentified RNA viruses might become the basis for a potential outbreak in the nearest future (Fitzpatrick *et al.,* 2021).

Also, despite the complexity, evolution, and high mutation rates of RNA viruses, the information on the total RNA virome of birds including chickens, particularly in apparently healthy ones is still patchy. Although it would be crucial to simultaneously explore both healthy and diseased chickens despite the difficulties in their sample availability as well as ethical concerns regarding sampling diseased birds. Therefore, monitoring the total RNA viral diversities in healthy chickens is a crucial step towards understanding the chicken RNA virome. Also, seemingly healthy chickens are more readily available and in high demand, since they are reared in large commercial, medium-size, and backyard farming systems and may carry viral pathogens even in the absence of any observable disease condition or growth impediment (healthy). Therefore, this study was underscored as a crucial step toward future epidemiological investigations of avian enteric viruses in diseased birds. Hence, this pilot study characterized and obtained baseline data on the RNA faecal virome of healthy South African chickens with the possibility of considering diseased birds in future studies.

Overall, this study provides information on the characterization of gut RNA viral communities present in the faecal samples of healthy chickens from a commercial farm in Durban, KwaZulu-Natal, South Africa using the NGS method. In addition, this study examined the effect of seasonal variations and chicken's developmental age on viral gut shedding as well as the evolutionary relatedness and diversity among the identified viruses. Furthermore, information obtained from this study serves as baseline data for future studies comparing viruses in healthy and diseased chickens. It is also hoped that the data from this study will be helpful for studies looking to compare RNA virome structure in healthy and diseased chickens and will serve as an effective step towards better preparedness for emerging or reoccurring infections from chickens.

## 1.2 Statement of the research problem

Infectious RNA viral outbreaks of animal origin have been and still is a global issue, leaving several countries in a quandary; for instance, the Coronavirus pandemic (Peacock *et al.,* 2021). Continuous viral surveillance of livestock species is crucial globally because of their economic importance. It could be noted that more than 60% of emerging infectious viruses are of animal origin (Lloyd-Smith *et al.,* 2009; Brucker 2020; Rahman *et al.,* 2020), though recently, there has been considerable attention on zoonosis devoted to bats and other mammals; nevertheless, birds are also important natural hosts for quite a number of zoonotic infections/transmissions globally. These infections include the *Crimean-Congo haemorrhagic fever virus* (Lindeborg et al. 2012), *Japanese encephalitis virus* (Hameed *et al.,* 2021), Louping-ill virus (Jeffries *et al.,* 2014), *Newcastle disease virus* (Alexander 2000), *St. Louis encephalitis virus* (Gruwell *et al.,* 2000), *Tick-borne encephalitis virus*

(Deviatkin *et al.*, 2020), *West Nile virus* (Vidaña *et al.,* 2020), and the *Usutu virus* (Kuchinsky *et al.,* 2021) among others. In addition, Influenza A viruses, particularly the H5N1, H5N6, and the latest H7N9 have been implicated with potential pandemic concern. Globally, from January 2003 to January 2023, about 868 human infections of H5N1 avian influenza (H5N1) including 457 deaths were recorded while since early 2013, a total of 1,568 confirmed human infections with H7N9 avian influenza A virus and 616 deaths, have been reported by world health organisation (WHO) (WHO 2023). The burden of avian infectious RNA viruses is still high causing substantial economic losses, morbidity and mortality, globally. Hence it appears that the vulnerability of both novel and re-emerging viruses may have been seriously underestimated in the light of poultry farming and cross-species transmission occurrences of viral pathogens (Hassan *et al.,* 2016; Gardy and Loman 2018; Kalkowska *et al.,* 2018). These viruses have been implicated with enteritis, malabsorption, immune suppression, and other infections resulting in low productivity of chickens and its allied products, as well as high costs of management and treatment in poultry systems. South Africa has recorded about 68 bird flu outbreaks in poultry as of 2017 with more than half occurring in chickens, costing the nation about 900 million Rands as well as export restrictions (Abolnik *et al.,* 2021; Marimwe *et al.,* 2021). Recently, there have been about 145 avian-flu outbreaks from the first outbreak in April 2021 to March 2022 with more than 4 million birds culled in South Africa (Uwishema *et al.,* 2021; SAPA, 2022b). In addition, an outbreak from Influenza A virus H5N2 led to the stamping out of about 10% of South Africa's local ostrich population and approximately 7.4 million USD in compensation to owners (Abolnik *et al.,* 2019; Marimwe *et al.,* 2021). Consequently, the loss of substantial revenue resulting from mobility restrictions of poultry products exports, the medical cost of surveillance systems during outbreaks in birds, vaccine procurements, morbidity, and fatal human avian cases are major impacts of avian viral diseases (Newmana and Abrahamsb 2018; Arndt *et al.,* 2020). Nevertheless, decades of avian research have focused on zoonotic viral pathogens, or viruses causing significant economic losses in poultry. However, there is a paucity of information on the overall RNA viruses as well as factors such as age, seasonality, latitude, and sex which may impact the presence or prevalence of RNA viruses in chickens. Also, with the continuous evolution of highly unstable RNA viruses and quasi-species emergence (Carrasco-Hernandez *et al.,* 2017; Fitzpatrick *et al.,* 2021), these undetected, unidentified RNA viruses might become the basis for the emergence of novel viruses, with a potential outbreak in the nearest future. Hence, this study will, for the first time, determine and analyse the overall RNA virome in the GIT of healthy South African chickens using next-generation sequencing (NGS). Also, it will seek to determine the RNA viruses that might be causative agents for diseases affecting chickens and those with the potential to cross species barriers.

In addition, the impact of bird age on viral composition, and the effects of seasonal variations during summer and winter periods on viral diversities and abundance was ascertained. Hence, characterizing the overall RNA viruses in healthy chickens provides insights into the knowledge of key viral agents associated with poultry infectious diseases which is an effective step towards better preparedness and management of future emerging and recurring diseases in chickens.

## 1.3 Justification/Significance of the study

This study was viewed as an important step towards field epidemiology investigations of enteric avian RNA viruses. In South Africa, a dearth of information exists about the total gut virome of animals including chickens. Therefore, this study seeks to determine and analyse the genomes of faecal RNA viruses present in healthy chickens that could contribute to underlying disease agents affecting normal host immunity and/or alter the severity of diseases such as diarrhoea or act as catalysts for more serious and/or opportunistic diseases that result to high morbidity or mortality of avian species. Hence, rather than studying one or fewer viruses in isolation, the Illumina next-generation sequencing technology and modern bioinformatics tools in viral metagenomics were adopted, which hold the potential to provide a clearer understanding of the total viral RNA composition of chickens. This study seeks to find information on the diversity and composition of the total RNA viruses in healthy chicken's gut. The tracking of chicken gut viruses at early (2 weeks), intermediate (4 weeks) and mature stages (7 weeks) provide insights into the viral colonization of the GIT of avian species. In addition, this study provides a baseline of reference sequences and information for future studies looking to compare the virome of healthy and diseased African chickens. Furthermore, the data obtained from this study could be vital in improving poultry health management systems and the development of prevention strategies for RNA viral infections. This will, in-turn provide a baseline knowledge of viruses in asymptomatic chickens for better poultry production strategies in South Africa and result in improved and efficient production of healthier chickens and their derived products.

## 1.4 Hypothesis

i. Would enteric RNA viruses be detected from Apparently healthy commercially bred South African chicken?

ii. What impact would chicken age have on viral abundance and diversity?

iii Would more viruses be detected during winter collection time points than in summer?

## 1.5 Aim

The overall aim of this study is to determine the composition, diversity and abundance of the total RNA viruses present in healthy South African chickens using the Illumina NGS technique.

## 1.6 Objectives

The specific objectives are:

i. To profile the total RNA viruses, present in the faecal samples of healthy chickens in Durban, KwaZulu-Natal province, South Africa using the NGS method.
ii. To determine the effect of developmental ages (2, 4, and 7 weeks) and seasonal variations [summer (December to March) and winter (June to August)] on the chicken's GIT RNA viral abundance and diversities.
iii. To describe the evolutionary relationship of the identified chicken gut RNA viruses using phylogenetic analysis and to establish a database of reference sequences.

## 1.7 Dissertation layout/organization

The present dissertation has been organized into six distinct chapters.

The first chapter (introduction) provides the broad context of the study, its problem description, goals, specific objectives, and the significance of the study.

The second chapter gives an in-depth review of the published literature pertaining to the composition of avian gut virome, its relation to diseases, and the next-generation sequencing characterization technologies employed in recent times, while chapter three presents a detailed methodology on targeted enrichment methods and optimization undertaken for characterization and better recovery of mRNA viral transcripts employing metagenomic NGS.

Chapter four presents the overall results of different RNA optimization steps used, enteric RNA viruses identified in chickens, the differences in abundance and diversity across sample groups, the ecological diversity measures of the impact of age and seasons on avian gut virome diversities and phylogenies of the identified RNA viruses, all based on the three study objectives.

The fifth chapter provides a discussion/comparison of the results obtained, along with their potential implications, while the last chapter (chapter six), presents the concluding findings of the study, as well as the limitations and potential avenues for future research.

# CHAPTER 2. Literature review

## 2.1 Chicken gut virome

Avian GIT extends from the mouth through the pharynx, oesophagus, crop, proventriculus, small intestines, caeca, rectum and finally the cloaca. Avian gut, unlike the human GIT is relatively short, with lesser transition time, and structurally comprises of diverse microbiota (Grond *et al.*, 2018; Abd El-Hack *et al.*, 2022). Chicken's productive performance and feed conversion rate are greatly influenced and dependent on the state of health and proper functioning of its GIT (Shah *et al.*, 2016), which is the site of nutrition, metabolism, and build-up of diverse microbes. In addition, the chickens GIT has a significant impact on immune cell development and system regulation, which in turn controls its intestinal flora and maintains infection resistance (Shah *et al.*, 2016; Abd El-Hack *et al.*, 2022; Yin *et al.*, 2022a). Research on the chicken microbiome has been significant, mostly concentrating on the gut bacterial composition, whereas studies on avian viral constituents of the microbiome are quite sparse with the information on asymptomatic birds even scantier. This may be attributed to the study of avian viruses as pathogens, hence has led to the obvious differences in research efforts between avian virome and bacteriome studies, as most virome studies are merely descriptive (François and Pybus 2020). In addition, it appears that most chicken viral studies are surveillance studies geared towards tracking specific pathogenic viruses. Furthermore, the lack of a common identifier for viruses (like the bacterial 16S gene and ITS for fungal), limited genomic database for comparisons and sparse analytical techniques have contributed to the challenges of virome investigations (François and Pybus 2020; Mogotsi *et al.*, 2020). Although the gut microbiome has been shown to influence the body's natural immunity to microbes, little is known about the role of the microbiome in determining the susceptibility of chickens to viral infections (Abd El-Hack *et al.*, 2022; Yin *et al.*, 2022a). The assemblage of viruses that populate or infiltrates chicken’s digestive tract is referred to as its gut virome. These chicken gut viruses may be categorized into enteric viruses, bacteriophages, mycoviruses, archaeal viruses, endogenous and diet associated viruses which contribute to the entire make up of its gut virome.

### 2.1.1 Enteric viruses

Avian enteric viruses encompass diverse viruses that trigger diseases and villus withering by replicating in the mucosa and cytoplasmic epithelial cells that surround the GIT. Inflammatory damages caused by these viruses’ changes the integrity of the gut, which compromises the mucosal barrier that maintains intestinal homeostasis and innate immunity of the avian host. Although younger

birds are more susceptible to enteric viruses owing to their still-developing body defence mechanisms, nevertheless these viruses may affect birds of all ages, consequently increasing the vulnerability of subsequent infections and lowering feed conversion efficiency. An investigation of the gut RNA virome of American chickens revealed that RNA viral abundance declined with age, with the exception of the picornaviruses (Shah *et al.,* 2016). This exception may not be surprising as different avian studies globally have identified members of the *Picornaviridae* from asymptomatic birds.

Virologists have employed recent sequencing technologies to characterize known and novel enteric viruses in human and animal hosts including chickens. Some of these enteric gut viruses have well-established roles and their associated disease conditions, while others still need to be further investigated. For instance, previous studies have associated caliciviruses, coronaviruses, reoviruses and parvoviruses to diarrheal related infections in chickens (Devaney *et al.,* 2016; Lima *et al.,* 2019; Kubacki *et al.,* 2022). Significant studies have tried to uncover the overall gut virome of chickens. Table 2.1 gives an overview of all different research efforts in the last decade to uncover the virome of chickens currently available in literature databases. The information depicted in Table 2.1 were extracted from chicken studies that exclusively investigated their virome through the metagenomics approach, and not studies that employed other molecular methods and/ or with interest in studying specific virus(es). A total of 11 studies were recovered at the time of this study from seven countries (Switzerland, Netherlands, Brazil, U.S.A, Hungary, China, and United Kingdom) spanning across only 4 continents: Europe, South America, North America, and Asia. Majority of the 11 studies used specimens from chicken such as faeces, jejunum, cloaca, intestines, ileum, duodenum, bursa, lungs, liver, heart, kidney, spleen, and pancreases. However, Kwok *et al.,* (2022) and Kubacki *et al.,* (2022) utilized chicken farm dust and farm cloth in their studies respectively. Within, these studies, for the physiological state of chicken used, half of these studies (5) used healthy chickens, 3 compared healthy and diseased chickens, 2 exclusively considered diseased birds and only study utilized dead chickens. It is important to state that within the available studies on chicken virome studies, the information on the African and Australian continents were underreported. Hence, to our knowledge, this is the first study on chicken virome from the African continent. With continuous emphasis made on the importance of characterizing the overall viral constituents of the poultry GIT's. Thus, it is crucial to also determine the roles which these viruses play in enteric health disorders or in the overall physiology of the chicken's GIT (Day *et al.,* 2015; Aruwa *et al.,* 2021).

Table 2.1 A synopsis of chicken virome metagenomic investigations over the past decade (2013-2023)

| Country | Published Year | Specimen | Microbes studied | Sample size | Age | Health state | Diseased condition | Next generation sequencing platform | Data availability status | ID/Accession numbers | Sampling date | References |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U.S.A | 2015 | Intestines | Bacteria and RNA viruses | 5 | Unspecified | Healthy | Healthy | Ion torrent | Yes | MG-RAST (4509873.3 - 4509879.3) | Unspecified | Day *et al.,* (2015) |
| China | 2023 | Liver, spleen, kidney, and recta | Bacteria, DNA and RNA viruses | 20 | Unspecified | Dead | Hepatomegaly rupture syndrome | Illumina | Yes | Genebank (OQ749505 - OQ749511); SRA (PRJNA923241) | 2021 | Wang *et al.,* (2023) |
| Brazil | 2021 | Duodenum, ileum, bursa, lungs, liver, heart, kidney, pancreas | RNA viruses only | 243 | 1-11 days | Diseased | Runting stunting syndrome (RSS), apathy and growth retardation | Illumina | Yes | Unspecified | 2016 to 2017 | de Oliveira *et al.,* (2021) |
| Brazil | 2018 | Faeces | RNA viruses | 1 pooled sample | 41 days | Healthy | Healthy | Illumina sequencing | Yes | Genebank (MG708279) | July, 2009 | Castro *et al.,* (2018) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| United Kingdom | 2016 | Intestines | DNA and RNA viruses | 2-7 birds from each group (RSS and normal) | 13-21 days | Healthy and diseased | RSS and normal | Illumina and Pyrosequencing | Yes | MG-RAST 4689857 | Unspecified | Devaney *et al.*, (2016) |
| Hungary | 2016 | Cloaca and intestines | Picornaviruses | Unspecified (Pooled samples) | 4 weeks | Diseased | Diarrheic | Illumina | Yes | Genebank (KT880665-KT880677) | April, 2013 | Boros *et al.*, (2016) |
| U.SA | 2016 | Jejunum and ileum | RNA viruses | 36 | 2- 6 weeks | asymptomatic | Healthy | Illumina | Unspecified | unspecified | Unspecified | Shah *et al.*, (2016) |
| Brazil | 2017 | Faeces | DNA and RNA viruses | 5 per house | 3-5 weeks | Asymptomatic | Healthy | | Yes | Genebank (KY086292-99) (KY69111-13) (KY56250-51) (KY120883) | | Lima *et al.*, (2017) |
| Brazil | 2019 | Faeces | DNA and RNA viruses | 70 | 3- 5 weeks | Healthy and diseased | Malabsorption syndrome and healthy | Illumina | Yes | Genebank (MG846351-MG846490) | May to December, 2015 | Lima *et al.*, (2019) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Netherland | 2022 | Faeces and farm dust | DNA and RNA viruses | 102 | 2-5 weeks | Apparently healthy | Asymptomatic | Illumina | Unspecified | unspecified | May to August 2019 | Kwok *et al.*, 2022 |
| Switzerland | 2022 | Poultry farm cloth | DNA and RNA viruses | 28 | 2-11 days and 30 days | Healthy and diseased | RSS and normal | Illumina | Yes, SRA and Genebank | Genebank (OM469021-OM469308); SRA (PRJNA802076) | Unspecified | Kubacki *et al.*, (2022) |

### 2.1.2 Bacteriophages

Bacteriophages or phages are bacterial-replicating viruses which predominate the gut virome. These immunogenic bacteriophages on the mucosal lining may aid in the building of an innate immune barrier by establishing an antibacterial first line of defence against luminal bacterial infections (Popescu *et al.,* 2021). Research on viruses present in faecal samples has demonstrated that bacterial viruses are the most dominant viruses in the gut (Breitbart *et al.,* 2008). The term "avian phageome" could be referred to as the sum of all bacteriophages found in a bird host. In addition to *Microviridae*, research has shown that gut phages are predominated by phages with dsDNA viral genomes within the families, *Siphoviridae* (61% of tailed phages), *Myoviridae* (25%), and *Podoviridae* (14%) within the Caudovirales order (Ackermann 2003; Leung *et al.,* 2019). However, recent research has shown that the distribution of RNA phages is more widespread than initially anticipated in a few environmental niches, including animal faeces (Krishnamurthy *et al.,* 2016; Callanan *et al.,* 2020). Antibiotic resistance genes are known to be highly prevalent in faecal samples of broiler chickens (Yan *et al.,* 2019).

Over the years, we have witnessed a surge in the prevalence of multidrug-resistant bacteria, attributed to food-borne pathogens. Unfortunately, the dearth of new and effective treatments has left us with limited options for combating these recalcitrant bacteria. Various studies have suggested bacteriophages (phages), as a potential substitute for antibiotics to ensure food safety and public health. Bacteriophages have been used for a variety of purposes, including disinfection (Lee *et al.,* 2016), biocontrol of food (Nabil *et al.,* 2018), and even in the production of chickens (Żbikowska *et al.,* 2020). However, some threats were raised by the persistence of these bacteriophages on poultry meat surfaces (EFSA, 2009) and these phages may become fully adapted super phages and develop resistance in the nearest future (Principi *et al.,* 2019). It is also believed that phage treatment in people and animals, including poultry, may be affected by the immune responses elicited by the bacteriophages (Cisek *et al.,* 2017; Majewska *et al.,* 2019). Overall, though many promising results have emerged from studies of bacteriophage applications, there are still several drawbacks and uncertainties that should be resolved (Loc-Carrillo and Abedon 2011). In addition, owing to the discovery of 122 partial genomes of new RNA phages from a metagenomic dataset, it is clear that the "real" diversity and abundance of RNA have not been fully explored (Krishnamurthy *et al.,* 2016; Callanan *et al.,* 2018). Presently, little is known about the healthy chicken gut phageome (Walk *et al.,* 2016), and far less is understood about the interactions between phage and the chicken immune system (Żbikowska *et al.,* 2020).

### 2.1.3 Endogenous retroviruses

Retroviruses are RNA viruses that replicates through a DNA pro-viral stage linearly existent in the host genome, because of the availability of a *pol gene* encoding the reverse transcriptase required for the transcribing of RNA to DNA in the viral genome. Endogenous retroviruses (ERVs) often occur in a wide range of vertebrate genomics, including avian species. ERVs make up around 3% of the avian genome and act as the evolutionary history of past avian retroviral infections (Mason *et al.*, 2016; Kapusta and Suh 2017). Within the avian genomics, chicken has been the model organism for the study of ERVs and with only about 1.3% of its genome made up of ERVs compared to a much higher 5% that was observed in humans (Lander *et al.,* 2001). When retroviruses incorporate itself into the germline, they adapt and become ERVs, which have a lasting effect on genome modification of the avian genome. ERVs are known to be vertically and horizontally transmissible in the host genome and can be further propagated through reinfection and retro-transposition, particularly in chickens and turkeys (Lee *et al.*, 2017).

The three major coding genes in ERVs are called *gag, pol*, and *env*. These genes are surrounded by long terminal repeats (LTRs), which are regulatory areas that include promoters, enhancers, and polyadenylation signals (Grifford and trestem, 2003). Additionally, ERV-derived non-coding RNA has been shown to have biological activity (Chen *et al.,* 2019). However, most ERVs lack the env protein, due to mutations and/or negative selection, and can be complemented by another retrovirus with a functional envelope (Feschotte and Gilbert 2012; Magiorkinis *et al.*, 2012). To keep up with the rapid pace of evolution in their host organisms, exogenous retroviruses often acquire new accessory genes through horizontal gene transfer from other vertebrates, their pathogens, or other viruses (Kanda *et al.*, 2013). Novel, structurally complete ERVs may be produced by reinfection and intracellular retro transposition, and these ERVs have the capacity to influence gene expression, enable chromosomal rearrangements (Kanda *et al.*, 2013) and modify the innate immunity to retroviral infections.

Endogenous retroviruses of avian origin are classified into three major exogenous retroviral classes (class I to III), according to *pol* amino acid sequences (Jern *et al.*, 2005), and consist of four internal coding regions: group-specific antigen (*gag*), protease gene (*pro*), RNA-dependent DNA polymerase gene (*pol*), and envelope gene (*env*), which are flanked by LTRs. In chickens and turkeys, the reticuloendotheliosis virus spreads vertically and horizontally. Researchers have shown that one of the 49 full-length *Gallus gallus* endogenous retroviruses (GGERV10) has the potential to be used as a diagnostic biomarker for chicken breed determination (Lee *et al.*, 2016). In contrast, commercially

significant characteristics in chicken production may be negatively impacted by these exogenous and, in some instances, endogenous viruses. Even though retroviruses are often classified as endogenous sequences integrated into a host genome, they can also be exogenous viral particles that infiltrate the avian species. For instance, Shan *et al.,* (2022) identified a novel retrovirus from cloacal samples of wild birds, whose mean mapping coverage (74.4) was significantly higher than that of the host genome (0.006), indicating that the retrovirus genome obtained from this library is most likely from a virus particle and not of the host. Three species of avian retrovirus cause disease in poultry: the avian leukosis/sarcoma virus (ALSV), reticuloendotheliosis virus (REV), and lymphoproliferative disease virus (LPDV) of turkeys (Payne 1998). It is widely established that ERVs may influence host responses on different levels. ERVs have been linked to a variety of phenotypes, including blue eggshell (Wang *et al.,* 2013; Wragg *et al.,* 2013), recessive white (Chang *et al.,* 2006), late feathering (Elferink *et al.,* 2008), and henny feathering (Li *et al.,* 2019). In addition, strain T, an abruptly changing type of reticuloendotheliosis virus, contains a viral oncogene and can produce reticuloendotheliosis in birds within a few days (Payne, 1998). Overall, chicken endogenous retroviruses (ERVs) have been studied for their effects on a variety of diseases, not only those linked to tumours (Li *et al.,* 2019; Dai *et al.,* 2022). Nevertheless, it appears that our knowledge of chicken ERVs and their roles in chicken illnesses is limited.

## 2.2 The interplay between the gut virome and chicken diseases

Studies have shown that the gut microbiome has a role in the development of illnesses both within and outside the intestines (Cao *et al.,* 2022). One of the cutting-edge areas of gut microbiome research is the gut virome, which is being shown to have an essential role in the pathogenesis of diverse viral infections. The avian gut is colonized by microorganisms shortly after hatching (Oakley *et al.,* 2014). In a balanced chicken gut, the gut mucosal barrier provides both physical and chemical protection and helps maintain host−microbial homeostasis. The disruption of the gut integrity by viruses induces inflammatory damage in the intestinal mucosa which in turn compromises the innate immunity of the host defences thus permitting the appearance and adaptation of pathogenic RNA viruses. Studies have revealed that the microbiome structure of avian species may be affected by a few environmental variables, such as age, breed, sanitation, nutrition, temperature, and housing (Clavijo and Flórez 2018; Ocejo *et al.,* 2019). For instance, a longitudinal study which compared breeder source and age, in the development of intestinal RNA viruses in healthy American chickens, using stool samples revealed that gut RNA viruses were more abundant in the earlier stages of the chicken's life (Shah *et al.,* 2016). Unlike the DNA virome, which has been extensively investigated, the RNA virome in the avian gut

has been examined only to a limited extent since RNA viruses tend to be less durable in samples than DNA viruses, making their detection by metagenomic sequencing more challenging (Turner *et al.,* 2021). The gut microbiota functions of chicken have been extensively studied in relation to resistance to chicken viral diseases (Clavijo and Flórez 2018; Abaidullah *et al.,* 2019; Alqazlan *et al.,* 2020; Shi *et al.,* 2021c). However, there is currently a lack of information on the impacts of gut virome on avian infections and requires clarity.

## 2.3 Pathology of RNA viruses

The virome represents one of the most varied biomes on our planet, exhibiting an astounding array of species and genetic variation. The gut flora promotes the progress of immune cells and boosts the innate antiviral immunity (Alqazlan *et al.,* 2020; Shi *et al.,* 2021c). Nonetheless it has been observed that viruses with pathogenic properties have the potential to manifest within the GIT of avian hosts during periods of stress that are induced by infection. Yet another effect to be examined is that of competitive exclusion for viruses. For bacteria, a gut microbial community that is balanced serves as a protective mechanism that can potentially prevent the overgrowth of pathogenic bacteria through a process known as competitive exclusion. This process involves the gut microbiota that settle first in the epithelium competing for space and nutrients with potential pathogens. As a result, an unsuitable environmental atmosphere for potential pathogens may be created, or their growth may even be inhibited through the production of antimicrobial compounds such as bacteriocins (Garcia-Gutierrez *et al.,* 2019). Research findings have indicated that competitive exclusion is an effective measure in providing defence against enteric bacterial pathogens in juvenile poultry (Methner and Rösler 2020; Lee *et al.,* 2023). This competitive exclusion effect conferred is yet to be examined for opportunistic viral pathogens. According to Virgin *et al.,* (2009), it has been predicted that a healthy individual may harbour more than 10 chronic viral infections, and in some cases, even more. The microbiome is significantly impacted by viruses due to their parasitic nature. It has been observed that viruses possess the capability to modulate and exert control over a significant portion of the prokaryotic microbiome, consequently impacting both the composition of the microbiome and the host. The phenomenon of trans-domain relationships, which refers to the relationships that exist between organisms belonging to different domains of life, is frequently observed in the context of phage/bacteria as well as virus/host/bacteria connections (Virgin 2014). The investigation of the virome constitutes a pivotal constituent in the pursuit of comprehending the function of the microbiome in animals including birds.

## 2.4 Sequencing technologies

### 2.4.1 First generation sequencing

The decryption of genomic sequences and the knowledge of DNA composition of organisms were made possible by the 1953 crystallographic discovery of the double helix structure of DNA, which is comprised of the four bases A, T, C, and G. To better comprehend and cure genetic illnesses, the investigation of genomic sequences of organisms has benefited greatly from the emergence of sequencing technologies (Tourneau and Kamal *et al.*, 2015). A sequencing device generates files containing DNA sequences in the form of strings known as reads on an alphabet composed of the five letters A, T, C, and G, with the letter N representing uncertainty (Shendure and Ji, 2008; Kchouk *et al.*, 2017). The advancement of sequencing technologies over time has allowed unprecedented knowledge of the genomic information of all life forms.

#### 2.4.1.1 Sanger sequencing

Frederick Sanger and colleagues in 1975, and Allan Maxam and Walter Gilbert in 1977, were the first to discover methods for sequencing DNA. According to the di-deoxy nucleotide chain termination principle (Sanger, 1988), the template is split into four aliquots and then all the substrates needed for sequence synthesis are added. This is the basis for Sanger sequencing. Radiolabelled chain termination di-deoxynucleotides (ddATP, ddGTP, ddCTP, ddTTP) are originally utilized throughout this sequencing procedure and are subsequently added to each reaction to halt DNA strand synthesis. Separating the template's terminated lengths derived from the platform using gels was the final step (Sanger, 1988). Chemical degrading and base-specific cleaving of DNA was at the heart of Maxam and Gilbert's approach (Maxam and Gilbert, 1977) with each of the four reactions (C, T+C, G, A+G) and chemical treatment causing breakage at a tiny percentage of one or two of the four nucleotide bases. This reaction generates a sequence of fluorescently labelled fragments that may be electrophoretically resolved based on their relative sizes. However, for Sanger sequencing which did not involve DNA cloning, or the use of potentially dangerous reagents became the method of choice. Following this, automated laser-based fluorescence dye detection and capillary electrophoresis were then later introduced (Smith *et al.*, 1986). Later, Sanger sequencing technology was improved to the currently used capillary sequencing where fluorescently labelled chain terminating di-deoxy-nucleotides are used as chain terminators. The sequenced templates obtained from the platform are separated using capillary electrophoresis (Liu *et al.*, 2012; Swerdlow & Gesteland, 1990). Although capillary sequencing yields extremely precise reads, it is a time-consuming, costly, low-throughput, and a labour-intensive process.

### 2.4.2 Second generation sequencing

The development of second-generation sequencing technology was spurred by the need for rapid, low-cost, high-throughput, and accurate sequencing methods. This new generation of sequencers emerged in 2005, expanding beyond the capabilities of its predecessors. Second-generation sequencers are intriguing development that was made possible by fusing Sanger technology with fluorescence detection, which ultimately enabled for many sequence processes to occur in parallel. There are two broad types of second-generation sequencing technologies, distinguished by means breakdown, amplification genomic segments and identification of bases. Those based on the principle of sequencing by ligation such as the SOLiD-ABI (Applied Biosystems). The second group rely on sequencing by synthesis such as 454-pyrosequencing, Illumina sequencing and Ion torrent sequencing (Tipu & Shabbir, 2015). Among this generation sequencers, researchers appear to have a preference for the Illumina sequencer. The advantages of these modern technologies include faster processing times, more precision, wider coverage, and higher throughput, all while requiring less time investment. Most importantly, their capacity to sequence samples without culture or previous knowledge of the sample makeup has made viral metagenomics research a reality (Mogotsi, *et al.*, 2019). Table 2.2 compares the different features of major NGS platforms.

#### 2.4.2.1 SOLiD-ABI sequencing

The Supported Oligonucleotide Ligation and Detection (SOLiD) is an NGS sequencer marketed by Life Technologies which was bought by Applied Biosystems (ABI) in 2006, and a merger sequencer named ABI/SOLiD sequencing. This sequencing approach involves the use of a DNA ligase in ligating four fluorescently labelled bases competing for binding to an oligonucleotide complementing the template being sequenced. Multiple cycles of ligation, detection and cleavage are performed with the number of cycles determining the eventual read length. The strength of ABI/SOLiD platform is high accuracy because each base is read twice while the drawback is the relatively short reads and long run times (Kchouk *et al.,* 2017). However, it is not able to produce good read lengths and is faulted with base substitution errors.

#### 2.4.2.2 454 pyrosequencing

The luminous approach is used by the 2005-introduced 454-pyrosequencing technique to measure the release of pyrophosphate after the incorporation of each nucleotide into the synthetic DNA strand. An enzyme called adenosine triphosphate sulfurase converts pyrophosphate, which is produced during sequencing by synthesis, into a more usable form of energy called adenosine triphosphate

(ATP). This ATP serves as a substrate for the enzyme luciferase, which emits light in direct proportion to the quantity of pyrophosphate it liberates (Tedersoo *et al.,* 2010). However, counting consecutive nucleotides was a challenge while using this method. Roche/454's newest instrument, the GS FLX+, can output 1 million reads each run with read lengths of up to 1000 bp (454.com GS FLX+Systems http://454.com/products/gs-flx+system/index.asp). As a result of homopolymer regions, insertion and deletion mistakes predominate in sequencing data (Huse et al. 2007). Noisy signals were often produced if there were four or more identical nucleotides, even though the amount of light emitted was proportional to the homopolymer's length (Froehlich & Heindl, 2010). Additionally, mistakes in identifying nucleotides might be brought on by signals of insufficient or excessive strength, respectively.

#### 2.4.2.3 Ion torrent sequencing

In ion torrent sequencing method, a chip containing a series of micro wells, each of which carries a bead with multiple identical fragments, is employed. Its principle typically, involves picking up pH variations brought on by the emission of hydrogen ions during the sequencing procedure. Given the evidence of the process's extreme specificity, no ionic alterations take place if a complimentary base cannot be added to the expanding template being sequenced. A sensor at the micro well's base detects this shift, translating it into a voltage signal that is directly proportionate to the quantity of nucleotides integrated. This ion proton sequencer output may exceed 10 Giga bases, and the sequencer can generate reads of 200 bp, 400 bp, and 600 bp in length. Importantly, the main benefits of this sequencing technique are its quick sequencing time (within 2 and 8 h) and greater read lengths (when compared to other second-generation sequencers) (Kchouk *et al.,* 2017). Nevertheless, due to signal degradation when several matching dNTPs are included, this method of sequencing is not ideal for deciphering homopolymer sequences (greater than 6 bp) (Rothberg *et al.,* 2011; Reuter *et al.,* 2015).

#### 2.4.2.4 Illumina sequencing

Illumina sequencing was introduced in 2006 and remains the most popular second-generation sequencing platform. It features more stable chemistry and an error rate of 0.1%. In this approach, adapters are ligated to the beginning and end of each fragmented DNA sequence extracted from a sample. Following this, the adapters are affixed to their corresponding complements, which are in turn hooked on a flow cell containing a wide variety of adapters. These sequences that have been connected to a solid plate are then amplified by bridge PCR amplification which results in many copies of each sequence called cluster, consisting of several duplicates of the same sequence. Four modified nucleotide sequencing primers, and DNA polymerases are combined, and the primers are

hybridized to the sequences to identify each individual nucleotide. Then, polymerases are employed to extend the primers using the modified bases. To ensure the individuality of each nucleotide type, they are all fluorescently tagged. An inert 3'-hydroxyl group in the nucleotides prevents the addition of any further phosphate groups, guaranteeing that just a single nucleotide will be used. Clusters are stimulated by lasers, which in turn release a light signal unique to each nucleotide. This excited signal is captured by a coupled-charge device (CCD) camera, and the data is then interpreted by a computer to reveal the nucleotide sequence (Tipu & Shabbir, 2015). However, the use of this sequencing platform has a limitation of generating short read sequences, which makes genome assembly difficult, particularly in places with repetitive sequences (Hengyun, & Francesca, 2016). In addition, sequencing control is often a necessity as overloading of samples might cause them to merge into one another, leading to low-quality sequencing output. Furthermore, the most prevalent kind of mistakes associated with this technique is base substitutions owing to incorrect identification of the integrated nucleotide (Dohm *et al.,* 2008).

### 2.4.3 Third generation sequencing (TGS)

PacBio and Oxford Nanopore Technology are two examples of TGS platforms that can sequence long reads. The third gen-sequencers employs the principle of single molecule real time sequencing approach (Bentley *et al.,* 2008) and the synthetic method that builds on Illumina's current short-reads capabilities (Harris *et al.,* 2008). This generation sequencers have shown promising results in enhancing genome assembly, particularly in regions with high repetition (Kchouk *et al.,* 2017). These methods generate reads in real time, with a much shorter sequencing time compared to Illumina sequencing, which requires at least 48 h and allows access to the sequenced data only after it has been completed. This advantage makes them excellent sequencers for diagnostic testing offering real-time sequencing in a shorter time (Tyler *et al.,* 2018). Notwithstanding, the error rates for this generation sequencer are higher than those of alternative sequencing methods.

#### 2.4.3.1 Pacific Bio sequencing

Zero mode waveguides (ZMWs) are classified as wells of ten nanometres in diameter and are generated in a metal sheet. These ZWMs are used in single molecule real time (SMRT) cells in PacBio sequencing. To prohibit light from spreading, ZMWs allow light flow to through apertures with diameters smaller than its wavelength. Hence, with narrower diameters light diffuses along the wells, lighting the ground below. Embedded in the ZMWs' underside is a DNA polymerase and a nucleotide template which adds a fluorescently labelled nucleotide to the developing strand during sequencing (Eid *et al.,* 2009; Rhoads & Au, 2015; Travers *et al.,* 2010). The incorporation of nucleotides triggers

the production of a luminescent signal that is detected by the sensors, allowing the precise DNA sequence to be determined (Rhoads & Au, 2015). While the platform generates lengthy reads and sequences in real time, it is prohibitively costly and requires significant capital investment, making it impractical for use in most laboratories. In addition to this, the bulkiness of the sequencing equipment also makes its use awkward and impractical to be used in a real-world outdoor epidemiological investigation.

#### 2.4.3.2 Oxford Nanopore MinION sequencing

The limitation of Pac Bio sequencing was well compensated in the development of MinION sequencing. As a result of its mobility and compact size, the ONT MinION device can be used for real-time field epidemiological investigations (Laver *et al.*, 2015). The core concept of nanopore sequencing is the movement of nucleic materials through a nanopore. Nanopores come in two forms: those found in living organisms and those found in inorganic materials (Haque *et al.*, 2013; Liu *et al.*, 2012). Proteins in biological membranes create nanopores and once the desired sequence has been determined, the proteins are changed to alter their internal structure binding to molecular motors. *Staphylococcus aureus* α-haemolysin and *E. coli* Curlin sigma S dependent growth subunit G are two examples of proteins utilized to make nanopores (Goya *et al.*, 2018). In contrast, nanopores in the solid state are constructed openings in material-like graphene (Haque *et al.*, 2013; Magi *et al.*, 2017).

For sequencing purposes, the MinION device may be connected directly to a laptop equipped with 1 terabyte of storage, 16 gigabytes of random-access memory, and 4 cores of processing power (Loman *et al.*, 2015). The flow cell of a typical ONT MinION device has 512 paths, each of which contains four nanopores; however, only one nanopore per sensor is activated at any one moment (Stoddart *et al.*, 2009) and the MinKNOW program is used to manage the sequencing run procedure. The program facilitates setting up the experiment's run settings, gathering data, and receiving results and insights (Magi *et al.*, 2017; Rang *et al.*, 2018). By attaching a motor protein to the pore, the device may unzip double-stranded DNA and force it through a chemically or biologically created pore, allowing for sequencing to take place. Thus, variations in the nucleotides passing through the pore in motion generate an ionic current which is read as signals.

Table 2.2 A snapshot of the characteristic features of current next-generation sequencing (NGS) platforms

| **Present Company** | **NGS platform** | **Sequencing method** | **Maximum output** | **Maximum Read per run** | **Maximum read length (bp)** | **Run duration** | **Amplification method** |
|---|---|---|---|---|---|---|---|
| Illumina | Iseq 100 | Synthesis-based sequencing utilizing 3'-O-Azidomethyl reversible terminations | 1.2 Gb | 4 million | 2 X 150 | 9.5-19 h | Bridge PCR |
| | Miniseq | | 7.5 | 25 million | 2 X 150 | 4 to 24 h | |
| | Miseq | | 15 Gb | 25 to 30 million | 2 X 300 | 4 to 55 h | |
| | Nextseq 550 | | 120 Gb | 400 million | 2 X 150 | 12 to 30 h | |
| | Nextseq 2000 | | 330 Gb | 1.1 billion | 2 X 150 | 11 to 48 h | |
| | Novaseq 6000 | | 6 Tb | 40 billion (paired-end) | 2 X 250 | 13 to 44 h | |
| | Hiseq 3000/4000 | | 105 to 1500 Gb | 2.1 million to 5 billion | 2 X 150 | 84 h | |
| Oxford Nanopore Technology | Flongle | | 2 Gb | 0.1 million | Read length up to 4 Mb but is dependent | 72 h | PCR |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | MinION | Real time nanopore sequencing of relatively long fragments of DNA | 50 Gb | | on the length of molecule of interest | 72 h | |
| | GridION MK1 | | 250 Gb | | | 72 h | |
| | P2 and P2 Solo | | 580 Gb | | | 72 h | |
| | PromethION 24 | | 7 Tb | | | 72 h | |
| | PromethION 48 | | 14 Tb | | | 72 h | |
| Roche | 454 GS FLX + | Pyrosequencing | 700 Mb | 1 million | 700 bp | 23 h | Emulsion PCR |
| | Roche GS Junior | | 35 Mb | 0.1 million | 400 bp | 10 h | |
| Ion-Torrent | PGM 318 | Semiconductor synthesis | 10 Gb | - | 400 bp | 7 h | Emulsion PCR |
| | Ion-torrent Proton with Ion PI chip | | 15 Gb | 60 to 80 million | 200 bp | 2.5 h | |
| | Ion Torrent Genei studio S5 prime 540 chips | | 10 to 15 Gb | 60 to 80 million | 200 bp | 6.5 h | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Ion Torrent Genei studio S5 prime 520 chips | | 1.2 to 2 Gb | 4 to 6 million | 400 bp | 5.5 h | |
| | Ion Torrent Genei studio S5 530 chips | | 3 to 4 Gb | 15 to 20 million | 200 bp | 10.5 h | |
| ABI/SOLid Life technologies | | Oligonucleotide ligation | 3 Gb | 1.2 to 1.4 billion | 75 bp | 5 to 14 days | Emulsion PCR |
| Pacific Biosciences | PacBio RS C2 | Synthesis | 100 Mb | - | 2900 bp | 2 h | Real time single molecule |

The information depicted in Table 2.2 above was adapted from the official web pages of the respective sequencing platform manufacturers.

# CHAPTER 3. Methodology

## 3.1 Ethical approval and sampling

This study was carried out with the full approval of the Animal Research Ethics Committee (AREC) of the University of KwaZulu-Natal with the protocol Ethics Number, AREC 012/020 (Appendix 8.1.1). In addition, in compliance with section 20 of the Animal Disease Act (Act No 35 of 1984), a research permit was obtained from the South African Department of Agriculture, Land Reform and Rural Development, with the reference no 12/1/5/4 (1511AC) (Appendix 8.1.2). Furthermore, a provincial no restriction permit was obtained from the KwaZulu-Natal Department of Agricultural and Rural Development (DAR) for the collection of chicken faecal samples (Appendix 8.1.3). The faecal samples were obtained from a high-density, commercially driven poultry farm in Durban, KwaZulu-Natal, South Africa. The chickens were produced under standard rearing conditions in commercial flocks, and a non-invasive and non-intrusive faecal sampling method was adopted.

### 3.1.1 Sample collection and study design

A total of 10 chickens were used in all, five for summer and five for winter periods. This study, principally aimed to determine the RNA virome of healthy South African chickens was also underscored to track the viral colonization and composition of chickens' GIT. Hence, the early (2 weeks), intermediate (4 weeks), and mature (7 weeks) developmental stages of chickens were chosen. These ages were rationalized from the research design and the reality that under standard conditions, most commercially bred South African chickens, particularly broilers are off to the market by 7-8 weeks of age for consumption. Thus, five asymptomatic chickens showing proper development and normal behaviour with no observable disease condition (Veit 2021) were randomly pre-selected from flocks at 2 weeks of development during the summer period, July/August 2021. The selected chickens (n =5) at 2 weeks were marked and kept separately in medium-sized metal cages, layered with sterile plastic wrap and sawdust at each collection time point. The faecal contents from each cage were obtained from individual chicken immediately as they dropped using sterile plastic bags. The same collection procedure was repeated for the marked chickens (same summer chickens) at 4 and 7 weeks of age for summer samples. Similarly, for winter samples (December 2021 to January 2022), another five asymptomatic chickens were used, and the same selection, separation, and faecal sample time point collection procedure used for summer sampled chickens was followed resulting in a total of 15 winter samples. However, at age 7 weeks, only two samples were obtained from summer samples, while the remaining 3 chickens under study whose faecal samples were not obtained had been sold

off to the market for consumption, as at the time of sample collection. This reduced the expected sample number from 30 to 27 samples, all together. Hence, a total of 27 faecal samples (n=12 for summer) and (n= 15 for winter) were obtained, transported on dry ice, and were further stored at -80°C until further processing at the Department of Biotechnology and Food Science laboratory, Durban University of Technology. Figure 3.1 summarizes the faecal sample collection timeline of the 10 chickens used in this study. All wet lab analysis of the faecal samples, downstream processing and sequencing were carried out at the Next generation sequencing virological unit of the University of the Free State, Bloemfontein South Africa.

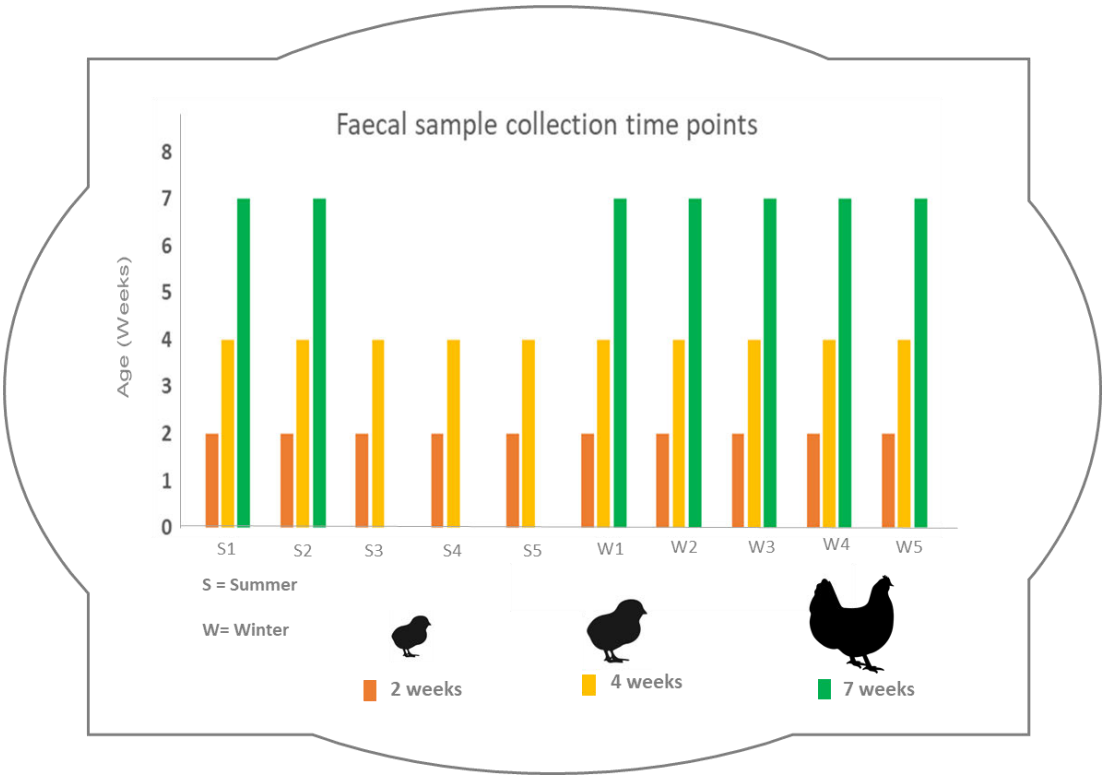



Figure 3.1. The time interval of chicken faecal samples collection. A total of 10 samples were collected from 2 weeks and 4 weeks developmental stages, whereas 7 samples were collected only at 7 weeks. On the X-axis, faecal samples S1 to S5 were collected during the summer, whereas samples W1 to W5 were obtained during the winter. In addition, the ages of the chickens used are indicated by the colors of the bars, and each bar represents only one sample.

### 3.1.2 Sample pre-treatment

Of the 27 samples, 5 each from each developmental stage of both winter and summer with the exception of the 7 weeks summer samples, all samples were subjected to pre-treatment in batches to

enrich for viral particles. The demographics of the individual samples and 10 chickens used in this study are depicted in Table 3.1. The pre-treatment of the faecal samples with antibiotics was done as described by Theuns *et al.,* (2014). The faecal samples were taken from a -80 °C Eppendorf CryoCube F570 bio freezer (Eppendorf, USA), placed on a pre-sterilized working bench, and allowed to defreeze to semi-solid composition by gradually normalizing to room temperature. Briefly, a 20% faecal suspension was constituted by making aliquots and weighing 200 mg using an analytical balance (Radwag, Radom, Poland) into a 2 ml Beadbug homogenization tube (Benchmark Scientific). This was further formulated by adding 1 ml of freshly prepared phosphate-buffered saline (PBS) (Sigma Aldrich, USA) at pH 7.5 containing 1000 U/ml penicillin, 1 mg/ml streptomycin, 1 mg/ml gentamicin and 500 U/ml amphotericin B (all from Sigma-Aldrich, USA). The faecal suspensions were homogenized in batches using an automated Beadbug homogenizer (Benchmark Scientific, USA) at 3000 rpm for 1 min.

Table 3.1 The demographics of the ten different gut RNA virome studied chickens

| **Sample tag** | **Chicken identity (ID)** | **Collection date** | **Collection age (weeks)** | **Sampling Season** |
|---|---|---|---|---|
| 2S1 | CH/2022/2S1 | 14/01/2022 | 2 | Summer |
| 2S2 | CH/2022/2S2 | 14/01/2022 | 2 | Summer |
| 2S3 | CH/2022/2S3 | 14/01/2022 | 2 | Summer |
| 2S4 | CH/2022/2S4 | 14/01/2022 | 2 | Summer |
| 2S5 | CH/2022/2S5 | 14/01/2022 | 2 | Summer |
| 4S1 | CH/2022/4S1 | 28/01/2022 | 4 | Summer |
| 4S2 | CH/2022/4S2 | 28/01/2022 | 4 | Summer |
| 4S3 | CH/2022/4S3 | 28/01/2022 | 4 | Summer |
| 4S4 | CH/2022/4S4 | 28/01/2022 | 4 | Summer |
| 4S5 | CH/2022/4S5 | 28/01/2022 | 4 | Summer |
| 7S1 | CH/2022/7S1 | 18/02/2022 | 7 | Summer |

| | | | | |
|---|---|---|---|---|
| 7S2 | CH/2022/7S2 | 18/02/2022 | 7 | Summer |
| 2W1 | CH/2022/2W1 | 07/07/2021 | 2 | Winter |
| 2W2 | CH/2022/2W2 | 07/07/2021 | 2 | Winter |
| 2W3 | CH/2022/2W3 | 07/07/2021 | 2 | Winter |
| 2W4 | CH/2022/2W4 | 07/07/2021 | 2 | Winter |
| 2W5 | CH/2022/2W5 | 07/07/2021 | 2 | Winter |
| 4W1 | CH/2022/4W1 | 21/07/2021 | 4 | Winter |
| 4W2 | CH/2022/4W2 | 21/07/2021 | 4 | Winter |
| 4W3 | CH/2022/4W3 | 21/07/2021 | 4 | Winter |
| 4W4 | CH/2022/4W4 | 21/07/2021 | 4 | Winter |
| 4W5 | CH/2022/4W5 | 21/07/2021 | 4 | Winter |
| 7W1 | CH/2022/7W1 | 11/08/2021 | 7 | Winter |
| 7W2 | CH/2022/7W2 | 11/08/2021 | 7 | Winter |
| 7W3 | CH/2022/7W3 | 11/08/2021 | 7 | Winter |
| 7W4 | CH/2022/7W4 | 11/08/2021 | 7 | Winter |
| 7W5 | CH/2022/7W5 | 11/08/2021 | 7 | Winter |

## 3.2 Viral enrichment

Enrichment of viral particles was carried out using the standard NetoVIR protocol for all the faecal samples (Conceição-Neto *et al.,* 2015) with some modifications. Following homogenization, the cell debris needs to be pelleted for the subsequent filtration step, hence the resultant homogenate in each case was subjected to a centrifugation step using a Labnet Prism™ R Refrigerated bench top micro-centrifuge (Labnet International, USA) for 5 min at 15000 x *g*. About 500 µl of the resulting supernatant was carefully filtered into a 1.5 ml qsp tube (Thermo Scientific, USA). The filtration process was carried out in two stages, initially using a 0.8 µm syringe filter (Sigma-Aldrich, USA)

and thereafter using a 0.45 µm filter (Merck, Millipore). In summary, the filtrate obtained from each sample ID was subjected to nuclease treatment using a cocktail of degrading enzymes which includes 2 µl of 25 U/µl benzonase nuclease (Sigma-Aldrich, USA) and 100 U/µl Micrococcal nucleases (Thermo Scientific, Massachusetts, USA). The resulting mixture was combined with 7 µl of freshly prepared buffer containing 1 M Tris buffer (Merck, Germany), and 30 mM $MgCl_2$ (Sigma Aldrich, USA) and 100 mM $CaCl_2$ (Sigma Aldrich, Missouri, USA) at pH 8.0. This mix was then incubated for 2 h at 37°C using the AccuBlock Digital Dry Bath (Labnet International, Inc., USA) to destroy the naked free-floating nucleic acids. Subsequently, after 2 h of incubation, 7 µl of 0.5 M EDTA was added to the samples for nuclease enzyme inactivation and all virally enriched specimens were kept at -20°C for 24 h, before extraction.

## 3.3 Viral RNA extraction

Total viral RNA from enriched samples was extracted using QIAamp viral RNA mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions with some modifications. First, the reaction buffers were modified to exclude carrier RNA, as in metagenomics studies unlike whole genome sequencing, random primers are used, resulting in amplified and sequenced carrier RNA more abundant than sample RNA targets. The buffer AVL contains chaotropic salts and detergents for cell wall degradation. Wash buffers 1 and 2 were reconstituted using ethanol instead of isopropanol, reducing yield. The final volume of AW1 and AW2 was 228 ml and 226 ml, achieved by adding 130 ml and 160 ml of 100% ethanol to 98 and 66 ml of buffer AW1 and AW2 concentrate, respectively.

The RNA extractions were performed in batches because the sample number exceeded the manufacturer's recommended 24 samples. Each enriched viral sample was equilibrated and 140 µl added to sterile 1.5 ml qsp tubes (Thermo Fisher Scientific USA) in duplicates. A volume of 560 µl of AVL buffer was added, homogenized using the pulse vortex method, and incubated at room temperature for 10 min. The tubes were placed with extra rack spaces, aerosol-barrier tips, and gloves changed after each centrifugation step to avoid contamination. After a 10 min incubation, the samples were briefly spun down, and 560 µl of absolute ethanol was added to each tube. Samples were briefly centrifuged using a microcentrifuge (Lasec Labs, Cape Town), and 630 µl of the lysate mixture was added to the pre-labelled spin column. The samples were spun down using a Labnet Prism™ R Refrigerated Microcentrifuge (Labnet, International, USA) at 6000 x *g* for 1 min, and the same steps were repeated until all liquid was spun down for each sample lysate.

The viral-bound spin columns were first washed with 500 µl of AW1, centrifuged at 6000 x *g* for 1 minute, followed by a second wash buffer (AW2) at 20000 x *g* for 3 min. The spin columns were then placed in new collection tubes and dried out residual ethanol. Air-dried spin columns were placed in sterile 1.5 ml Eppendorf tubes for RNA elution. About 50 µl of RNase-free water was added to the spin filter, allowing it to dry for 60 s. This process used RNase-free water instead of AVE elution buffer, which contained 0.04% sodium azide known to affect RNA quality determination in terms of purity (www.qiagen.com).

Finally, another modification was at the elution stage which involved centrifuging 1.5 ml tubes with spin filters for each sample I.D. for 1 min at 20000 x *g* after 1 min incubation. For better viral yield, the resulting flow through RNA collected was again collected and dispensed into the center of their respective duplicate spin filters and similarly centrifuged for 1 min at 20000 x *g*. For quantification, 2 µl of each RNA sample was assessed using Qubit reagents on a Qubit 3.0 fluorometer (Life Technologies, Invitrogen), which selectively measures RNA or DNA. The samples were stored at -20°C before the DNase treatment step.

## 3.4 DNase treatment

Therefore, DNase treatment was performed for all samples using the DNase I M0303S kit (New England Biolabs), according to the manufacturer's guidelines. First, the AccuBlock Digital Dry Bath (Labnet International, Inc., USA) was set to heat to 37°C. Following this, the reaction mix for each sample was set up on the ice. Each sample mixture of 100 µl total volume contained 48 µl of RNA, 10 µl of DNase reaction buffer, 1 µl of RNase-free DNase I, and 41 µl of Nuclease-free water. This was done swiftly on ice to prevent RNA degradation during bench preparations and to ensure uniformity of treatment time for all samples since the DNA degrading enzyme activity could only be activated at 37°C. Quickly, all sample mixtures were incubated for 10 min at 37°C after which 1 µl of 0.5 M EDTA was added to each treated sample for DNase enzyme inactivation. This step is often followed by column purification rather than heat inactivation for RNA samples used in downstream processes.

## 3.5 Purification of RNA and concentration

The removal of degraded DNA and DNase I enzyme from the reaction mixture and further concentration of sample RNA was carried out using the RNeasy Plus Micro Kit (Qiagen Hilden, Germany) following the manufacturer's guidelines. Any residual activity of DNase I will degrade the single-stranded DNA probes necessary for the ribosomal RNA depletion in the latter step. The

resulting RNA samples (as described in section 3.4) purification steps were performed quickly at room temperature with all centrifugation steps done at 22°C. Delving into specific details, 4 volumes of 100% ethanol were initially added to buffer RPE originally supplied as a concentrate to achieve the working solution. Differing slightly from the recommendations, the addition of β-mercaptoethanol or dithiothreitol to buffer RLT was not explored in this protocol. First, a 100 µl starting volume for individual samples was pipetted in pre-labelled sterile 2 ml qsp tubes, and 350 µl each of buffer RLT plus was added to the samples. This was followed by thorough mixing before and after the addition of 250 µl of 100% ethanol.

Briefly, 700 µl of the lysate was transferred into a gDNA Eliminator spin column placed in a 2 ml collection tube with the lids gently closed. A 30 s centrifugation step was carried out using Labnet Prism™ R Refrigerated microcentrifuge (Labnet International, USA) at 8000 x *g* after which the column was discarded and the flow-through collected. Additionally, 350 µl of 80% ethanol was added to the resulting flow-through and thoroughly pipette-mixed. The samples were then transferred to a RNeasy MinElute spin column secured in 2 ml flow-through tubes with closed lids. After a 15 s centrifugation at 8000 x *g*, the flow-through fluid was discarded, and the silica-membrane spin columns were retrieved and further placed into new collection tubes.

At this stage, the expected sample RNAs are bound to the silica-membrane spin filters facilitated using binding buffers. Hence, a series of wash steps and sample concentration is crucial to achieving pure RNA samples with better yield. First, 700 µl Buffer RW1 each was added to the RNeasy MinElute spin columns with closed lids and centrifuged for 15 s at 8000 x *g*. Similarly, the flow-through was discarded and 500 µl of buffer RPE each, was added to the RNeasy MinElute spin columns and spun down at the same centrifugation conditions. Again, the flow-through was discarded, and 500 µl of 80% ethanol used for membrane washing was dispensed into the RNeasy MinElute spin columns and centrifuged at 8000 x *g* for 2 min. Upon complete centrifugation, the RNeasy MinElute spin columns were carefully removed from the collection tube while preventing ethanol carryover resulting from column contact or spillage with the flow-through fluid, and both the flow-through and collection tubes were discarded. As a precautionary measure to ensure the thorough elimination of residual ethanol, an additional step was undertaken by placing the spin columns in new collection tubes with their caps left open following centrifugation at maximum speed of 17000 x *g* for 5 min. As it appears with uncapped tubes with longer centrifugation time at high speed, breakage of lids often occurs. Hence, prior to the latter centrifugation step, it is important to also label the collection tubes accordingly for clarity, position the lids in the opposite direction to the rotor's

rotation, and arrange tubes with one or two empty spaces between them in the centrifuge's cavities to minimize lid breakage. Finally, after the collection tubes had been discarded, the spin columns are now placed in pre-labelled sterile 1.5 ml qsp collection tubes and 12 µl of RNase-free water was added directly to the centre of the membranous spin filter with lids gently closed. The elution of RNA was achieved by a centrifugation step at maximum speed (16000 x *g*) for 1 min and quantification was carried out with 2 µl RNA sample using highly sensitive (Hs) RNA detection reagents on a Qubit fluorometer 3.0 (Life Technologies, Invitrogen). The eluted RNA samples were temporarily stored at - 20°C.

## 3.6 Ribosomal RNA depletion

To enrich viral RNA even further, the avian host ribosomal RNA (rRNA) was removed using the NEBNext ribosomal RNA (rRNA) depletion (Human/Rat/Mouse) kit (New England Biolabs, Massachusetts, USA) in accordance with the manufacturer guidelines. The ribosomal RNA (rRNA) constitutes about 80% to 90% of the total RNA (Palazzo and Lee 2015; Liefting *et al.*, 2021), with only 2.5%–5% each for the messenger RNA (mRNA) and transfer RNA (Buckingham 2019). Therefore, the main essence of rRNA depletion was to prevent the bulk of the libraries and sequencing reads from including rRNA which is not of interest (Shi *et al.*, 2021a). Hence, the purified RNA samples described in section 3.5 were first quantified. The RNA quantification was achieved using a highly sensitive RNA-specific fluorometric method with Qubit 3.0 (Life Technologies, USA) as described in section 3.7 and purity was assessed using the spectrophotometric principle of the BioDrop. Going by the obtained fluorometric results, the RNA samples met the standard input RNA requirement of 10 ng to 1 µg total RNA (DNA-free) in 10 µl total volume stipulated when using the NEBNEXT Globin rRNA depletion kit. The RNA-probe hybridization and rRNA depletion, RNase H degradation, DNase I digestion, and final RNA bead purification are the four key stages of the ribosomal RNA depletion protocol which are majorly characterized by different cycling temperatures, specific buffers, and all bench preparation done on ice.

To begin with, the RNA-DNA Probe hybridization step was prepared on ice using a 200 µl PCR tube each for individual samples. Each PCR tube contained, 10 µl of purified RNA, 3 µl of NEBNext Globin rRNA depletion solution and 2 µl of the probe hybridization buffer. The tube contents were evenly mixed by up and down gentle pipette-mixing 10 times and briefly centrifuged to condense liquid. Quickly, post centrifugation the samples were placed in a pre-heated MiniAmp Plus thermal cycler (Applied Biosystem) automated with the following temperature conditions; heated lid temperature of 105°C to avoid evaporation, 95°C for 2 min, ramping down to 22°C at 0.1°C per

second and a holding time of 5 min at 22°C. After completing cycling time, the tubes were swiftly taken out of the thermal cycler, briefly spun down with a microcentrifuge and placed on ice.

In addition to the completed RNA-DNA probe hybridization stage is the ribonuclease digestion using ribonuclease H, an endoribonuclease which cleaves RNA formed from RNA-DNA hybrids, breaking down the phosphodiester linkages of RNA to DNA probes. The reaction mixture was prepared on ice and each sample contained 2 µl NEBNext thermostable RNase H, 2 µl of RNase H reaction buffer and 1 µl nuclease-free water and 15 µl of RNA-DNA probe sample from the previous step achieving a total volume of 20 µl each. The sample mixture was evenly mixed by gentle pipette-mixing up and down 10 times, briefly centrifuged, and placed in a pre-heated Multigene OptiMax thermal cycler (Labnet International, USA) automated with a heated lid temperature of 40°C, and a 30-min incubation time at 37°C. After completed incubation, the ribosomal RNA bound to the DNA probes in the reaction mixture as anticipated has been fully depleted, however, an additional DNase treatment was crucial to degrade the remaining DNA probe in the sample mixture. First, the samples were removed from the thermal cycler, briefly spun down, and placed on ice. Following this, the DNase I reaction mixture containing 2.5 µl of NEBNext DNase I enzyme, 5 µl DNase I reaction buffer, and 22.5 µl nuclease-free water each, were added to the previously incubated 20 µl RNase H degraded samples. These samples with a total volume of 50 µl each, were gently pipette-mixed, briefly spun down with a microcentrifuge, and placed in a pre-heated Multigene OptiMax thermal cycler (Labnet International, USA) automated with a heated lid temperature of 40°C, and a 30-min incubation time at 37°C. Similarly, after completing incubation, the DNA-digested samples were briefly centrifuged and transferred from PCR tubes to a 96-well PCR plate on ice ready for the following RNA purification stage.

Subsequently, the purification of viral RNA was carried out as an enrichment step toward selectively isolating viral RNA from digested enzymes, salts, buffer reagents, and contaminants contained in the reaction mixture. This was achieved using the Agencourt RNAClean XP beads (Beckman Coulter, USA) which employ the principle of magnetic separation and size selection. To obtain an even solution of the Agencourt RNAClean XP beads, the container holding the sedimented beads was vigorously homogenized using a desktop vortex mixer (Axiology laboratories, South Africa). About 90 µl of the resuspended bead solution, was added to each of the 50 µl RNA samples from the DNase digestion stage. The samples were thoroughly mixed with the bead solution by pipetting up and down not less than 10 times and the plate was further incubated on ice for 15 min. The incubation stage was to enable RNA binding to the beads, thereafter, the PCR plate was gently positioned into a 96-plate

PCR magnetic stand ((PerkinElmer, USA) for a 5-min magnetic separation. A good separation of RNA-bound beads from the solution was characterized by a clear solution and bead pellets stuck to the directional contact of the magnetic cavities. The clear supernatant was completely removed by gentle suctioning without disturbing the bead pellets with the plate still in place on the magnetic stand. Afterwards, 200 µl of newly made 80% ethanol was added to each sample well and incubated for 30 s at room temperature as a wash step. The supernatant was gently pipetted out while avoiding any contact with the beads and this same wash step was repeated. After two complete wash steps, the beads were air-dried for about 5 min to evaporate all residual ethanol, with the plate still positioned on the magnetic rack. The beads were eluted while they still maintained their glossy dark-brown colour, as over-dried beads appear with lighter brown shades with cracks and result in lower RNA recovery.

Finally, the PCR plate was removed from the magnet stand, and immediately the RNA-bound beads were resuspended in solution by adding 7 µl of nuclease-free water in individual wells. The RNA-bead solution was thoroughly mixed by pipetting up and down 13 times, briefly centrifuged using a Hermle MK plate centrifuge (Hermle Labortechnik, Germany), and incubated at room temperature for 2 min. The PCR plate was repositioned on the magnetic base for magnetic separation of the beads and RNA in a clear solution for about four min. After the solution was sufficiently clear, 5 µl each of the clear supernatant containing the recovered RNA from samples was withdrawn and individually transferred to a fresh 96-well PCR plate. The purified RNA samples in the PCR plate were first sealed using a plate micro-seal film, tightly rolled over twice using a speedball Bio-Rad Roller, securing all edges before temporarily storing the RNA-containing plate at -20 °C overnight before proceeding with the cDNA synthesis step.

## 3.7 Whole transcriptome amplification

The whole transcriptome amplification (WTA) method is an efficient, unbiased, and a relatively fast method often used to amplify DNA. This approach involves a first step of reverse transcription RNA into complementary DNA, flanking the primer ligation sites, and final amplification step of the cDNA library independent of poly A-tailing and oligo d(t) priming. The flowchart presented in Figure 3.7.1 depicts the process involved in the amplification of the entire transcriptome utilizing the QIAseq single cell technology (Andreou *et al.,* 2023). Consequently, ensuring a uniform amplification of template RNA without 3′ bias. Hence, using cDNA obtained from the reverse transcription of the polyadenylated mRNA for amplification has remained the gold standard of major transcriptomics studies (Hrdlickova *et al.,* 2017; Andreou *et al.,* 2023).

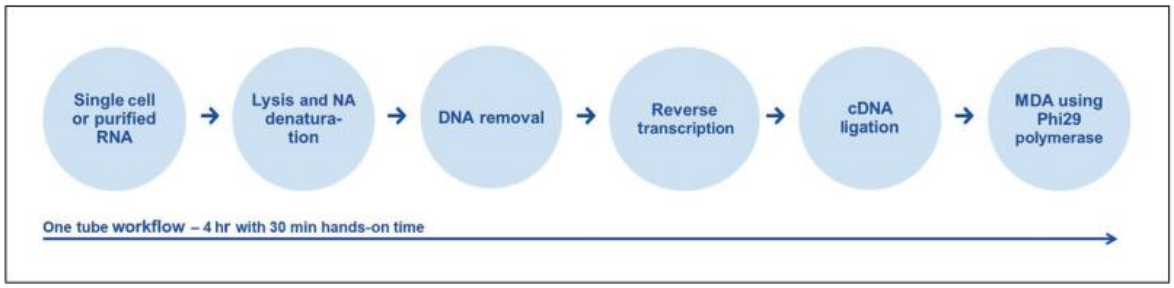



Figure 3.7.1 A procedural overview of the entire transcriptome amplification process utilizing the QIAseq FX single cell technology adapted from Andreou *et al.,* (2023).

The WTA of the ribosomal depleted RNA samples was achieved using QIASeq FX Single Cell RNA Library Preparation Kit (Qiagen, Hilden, Germany) following the manufacturer's guidelines. An input RNA volume of 8 µl was required for the WTA starting material, hence, 5 µl of the purified resultant RNA from each sample from the previous step (section 3.5) were first transferred into 2ml sterile pre-labelled microcentrifuge tubes and an additional 3 µl of nuclease-free water each was added into the samples. The RNA cell samples were first lysed by adding 3 µl of the denaturation buffer following a brief but simultaneous vortexing and centrifugation using a tabletop vortexer (Axiology laboratories, South Africa) and microcentrifuge (Labnet, International, USA) respectively. After the solution had spun down, ensuring that no liquid was above the meniscus, the samples were first incubated for 3 min at 95°C and chilled to 4°C using a Multigene OptiMax thermal cycler (Labnet International, USA). In order to remove all residual genomic DNA that may be present in the sample, 2 µl of gDNA wipe-out buffer was added to the samples, briefly vortexed, completely spun down, and further incubated for 10 min at 42°C. Quickly, a 10-min timer was set, and the Quantiscript reaction mixture for the reverse transcription phase was prepared. This was done by adding 4 µl RT/Polymerase buffer, 1 µl oligo dT primer, 1 µl random primer, and 1 µl Quantiscript RT enzyme mix in a 2ml sterile qsp tube and further scaled up by the number of samples, thoroughly vortexed and briefly centrifuged.

Reverse transcription involves the conversion of RNA into cDNA. At the beep of the timer at 10 min incubation time, the samples were taken out and placed on ice. Following this, 7 µl of the freshly prepared Quantiscript RT mix was added to the RNA samples, briefly vortexed, spun down, and then incubated in an auto-programmed thermal cycler with the following incubation conditions; 60 min at 42°C and 95°C for 3 min respectively and then cooled on ice. Similarly, after the completion of the reverse transcription of the RNA to cDNA, an adapter consisting of short oligonucleotide base pairs was ligated to the priming ends of cDNA to allow for further synthesis of the complementary strand

and amplification. The Adapter ligation step was achieved by adding 10 µl each of the freshly prepared ligation reaction mixture containing 2 µl ligase mix and 8 µl ligase buffer to all transcribed cDNA samples. The samples were removed from the ice, briefly vortexed, and centrifuged, following incubation at 24°C for 30 min. During the 30 min incubation time, the REPLI-g SensiPhi amplification reaction mixture for WTA was prepared by adding 29 µl REPLI-g sc reaction buffer and 1 µl of REPLI-g SensiPhi DNA polymerase, further scaled up by the number of samples and placed on ice. After 30 min of incubation at 24°C, the ligation reaction was stopped by incubating the samples for 5 min at 95°C, and immediately 30 µl of REPLI-g SensiPhi amplification reaction was added to the ligated reaction samples. These samples were thoroughly vortexed, completely spun down by brief centrifugation, and incubated for 2 h at 30°C using a Multigene OptiMax thermal cycler (Labnet International, USA). After 2 h of incubation, the amplification reaction was stopped by incubating the samples for 5 min at 65°C. The WTA products were quantified using Qubit 3.0 fluorometer (Life Technologies, Invitrogen) and stored at -20°C.

## 3.8 Fluorometric Qubit quantification of nucleic acids

The most common methods of quantifying nucleic acids and proteins in molecular studies are often spectrophotometric and fluorometric based. While the spectrophotometric technique relies on ultraviolet-visible absorption of light at a specific wavelength (260 nm) using a BioDrop µLITE micro-volume spectrophotometer (Biodrop, Cambridge, United Kingdom) and the $A_{260}/A_{280}$ ratio concentration determination calculations. The fluorometric measurement principally involves the use of fluorescent intercalating dyes such as Qubit assay reagents and the SYBR Green used in a quantitative Polymerase chain reaction (qPCR) for a more direct and precise determination of sample DNA, RNA, or proteins of interest. In this study, the Qubit assay kits (Life Technologies, USA) were used which comprised of the concentrated assay reagents, dilution buffer and pre-diluted DNA or RNA standards with known concentrations.

The Qubit reagent working solution was prepared following the manufacturers guidelines. First, to ensure nuclease free handling, wearing a new latex glove, the working bench, tubes racks, pipettes and calculator were sterilized using 1% sodium hypochlorite followed by 75 % ethanol and the gloves were discarded. Also, all reactions, particularly the fluorescent dye and samples with cold storage conditions were allowed to thaw, as preparation and quantification were to be carried out at room temperature only. After, wearing a new pair of gloves, 1 µl of Qubit assay reagent containing fluorescent dyes and 199 µl of Qubit buffer makes up the 200 µl solution for one reaction. Hence, to scale up the working reaction, 199 µl of Qubit buffer was multiplied by the number of samples to be

quantified plus 2 for the two standards. Similarly, for the fluorescent dyes, 1 µl was multiplied by the same multiplying number used for the buffer calculation. After, the scaled-up buffer and dye quantities were transferred into a sterile plastic 15 ml centrifuge tube, the solution was thoroughly mixed by 15 s pulse-vortexing method.

Moving on, 500 µl Qubit assay tubes (Thermo Fisher Scientific, USA) were arranged on the racks and labelled on their lids according to samples with two additional tubes, each labelled standard 1 and 2. After thoroughly mixing the working solution, 199 µl each of the working solutions were added to the sample tubes whereas 190 µl each were added to the standard 1 and 2 tubes. Following this, 10 µl of each of the Qubit HS standards was added to labels of standard tubes containing190 µl of the working solution, bringing it to 200 µl volume per tube. On the contrary, 1 µl each of the pipette mixed nucleic acid sample mixture was added to tubes containing 199 µl of working solution according to their labels. All Qubit tubes containing soon to be quantified samples and standards in working solutions were briefly vortexed and incubated for two minutes to allow proper binding of the fluorescent dye to the DNA or RNA. These nuclei molecule-dye binding allows the excitation of fluorescent signals which are translated into concentration during the actual quantification in the desktop Qubit's cavity.

In measuring sample concentrations, the Qubit 3.0 fluorometer was first calibrated using the standards. This was to ensure that the working solution was well formulated, and the concentration readings of standards were accurate. Going into details, the specific molecule type to be quantified selected on the Qubit user interface screen, High sensitivity (HS) RNA assay or HS double stranded (ds) DNA assay for RNA or DNA quantification, respectively. The calibration was done by placing the standard 1 and 2 tubes in the machines cavity, resulting in a confirmatory standard curve. The first sample was then placed in the sample chamber after completing calibration, and the volume of the sample used was imputed, for instance; 1 µl in this case. The sample concentration was read in ng/µl or ng/ml by carefully closing the chamber lid and pressing the "read tube" button.

Finally, it is important to emphasize that in RNA viral metagenomics studies, the specific nuclei material to be quantified as well as the Qubit assay kits used is dependent on the specific step of the workflow. In the case of this avian RNA virome study, quantifications done after RNA extraction, DNase I treatment, initial concentration and rRNA removal were carried out using the Qubit RNA HS assay kit (Life Technologies). On the contrary, post whole transcriptome amplification, through library preparations and quality checks, the DNA was instead quantified using the ds DNA assay kits since the RNA has now been transcribed to double stranded DNA.

## 3.9 Normalization of amplified cDNA products

One of the major objectives of genomics studies is to generate the best quality results with minimal technical variations across samples. Thus, addressing the origins of technical variability that may cause bias in downstream analyses has been shown to be an excellent measure for enhancing protocols (Bacher *et al.*, 2022). This is to say that, in the case of DNA concentration, samples with much higher concentrations may be overrepresented in the resultant library, hence, dominating the total read composition post-sequencing, and thus leading to under-detection of transcripts from samples with lower concentrations. In previous studies, equalization of cDNA concentrations across libraries has been proven to minimize sequencing coverage variation and improve transcriptome diversity by providing more uniform sequencing coverage for all samples (Bogdanova *et al.*, 2008; Bacher *et al.*, 2022). In addition, normalization reduces the bias of sequencing of most abundant transcripts and hence boosts the sequencing throughput of low- and mildly expressed genes (Zhulidov *et al.*, 2005). In preparation for the amplified library normalization, dilution calculations were performed to adjust the obtained Qubit concentrations of the amplified cDNA products to the required input DNA concentration of 200- 1000 ng/µl in 10 µl for all samples as recommended by the QIASeq FX Single Cell RNA Library Preparation Kit (Qiagen, Hilden, Germany). A confirmatory step of spectrophotometric quality assessment and Qubit HS DNA quantification was undertaken to ensure that the diluted samples had been equalized to a uniform concentration.

## 3.10 Library preparation

During library preparation, DNA/cDNA samples are transformed into a collection of cDNA fragments of specified lengths, amplified into clusters and sequenced on a high-throughput next-generation sequencer. Basically, library construction begins with the size selection in which the target DNA of interest is fragmented into specified lengths, often followed by end polishing/repair and ligation of adapters. These adapters are 8-10 bases unique molecular identifiers with T-overhangs that bind to the 3´adenylated ends of the cDNA fragments needed for barcoding, primer annealing and containing sequences that allow their hybridization to the flow cell. Further, a PCR clean-up wash step is often carried out to remove free barcodes, enzymes, nucleotides, and adapter dimers. After the PCR clean-up was done, the libraries were analysed on a microfluidics platform or bioanalyzer for sizing and quantification before normalization and pooling. It is important to understand that the procedures for library preparation may differ based on the specific kit used.

The library construction for this study was achieved using the QIASeq FX Single Cell RNA Library Preparation Kit (Qiagen, Hilden, Germany) following the manufacturer's guidelines. The approach of this kit generates PCR-free libraries from the already amplified cDNA. In preparation for fragmentation, the kit reagents were thawed on ice. Also, the defrosted buffers were quickly vortexed and briefly spun down to homogenize any localized concentrations that may have formed during their storage. It is critical to note that the insert size of final DNA fragments generated by fragmentation is influenced by the amount of DNA input and the fragmentation timeframe. According to the manufacturer's instructions, an input DNA of 200-1000 ng in 10 µl will yield a fragment distribution of about 350 base pairs for a fragmentation time of 15 min. The starting cDNA material used in this study has been normalized as previously mentioned in section 3.8. To begin with, the fragmentation of the Multigene OptiMax automated thermal cycler (Labnet International, USA) was programmed with the following temperature cycling conditions; Lid temperature of 60°C, 4°C for 1 minute, 32°C for 15 min, 65°C for 30 min and a holding temperature of 4°C. The program was allowed to start, ensuring that the heated lead option was on to avoid evaporation, however, it was paused immediately after the temperature reached the 4°C. At this point the fragmentation reaction was prepared on ice by adding 5 µl FX Buffer (10x), 20 µl $H_2O$, 5 µl FX Enhancer, and 10 µl FX Enzyme mix scaled up by the number of reactions. The PCR plate containing the normalized amplified cDNA initially placed on ice was brought forward, 40 µl of the fragmentation reaction mix each was added to the wells containing samples, gently mixed by pipetting up and down for 5 times and sealed with a micro seal film (Biorad, USA). Immediately the plate was spun down using a Hermle MK plate centrifuge (Hermle Labortechnik, Germany) 280 x *g* at 20°C for 30 s and then placed inside the pre-chilled Multigene optimax thermal cycler (Labnet International, USA). The program initially on pause was resumed to commence from the fragmentation at 32°C for 15 min. During the fragmentation step the cDNA was cleaved into fragment lengths of about 350 base pairs. After, the total fragmentation temperature cycling conditions had been completed the PCR plate was taken out of the thermal cycler and placed on ice in preparation for the ligation of adapters.

During this stage of adapter ligation, known 8 to10 oligonucleotides base molecular identifiers with T-overhangs known as barcodes or indices bind to the 3´adenylated ends of the cDNA fragments, serving as tags used to differentiate one sample from another, thus making multiplexing possible. In addition, the adapters contain primer annealing sites in which the sequencing primers for amplification are bound as well as sequences which can hybridize to the flow cell adaptors. In this study, the ligation reaction mix containing 20 µl DNA ligase buffer (5X), 15 µl $H_2O$ sc, and 10 µl

DNA Ligase enzyme were prepared and scaled up to the number of reactions and placed on ice. First, the adapter plate was spun down briefly using a plate centrifuge, the information of the individually matched adapters of each sample was recorded and then computed into the Illumina Manager software version 1.15 (Illumina, Inc., California, USA). This is essential for appropriately classifying and pooling samples according to their barcodes. Thereafter, 5 µl of each DNA adapter was added to their respective samples as previously recorded and thoroughly mixed by pipetting up and down 10 times. Additionally, 45 µl of the ligation reaction mix each was added to the barcode samples, thoroughly pipette-mixed, and incubated at 20°C for 15 min after which the PCR plate containing the indexed libraries was placed on ice.

Following the completed incubation time is the adapter clean-up steps characterized by ethanol washes and bead-based paramagnetic separation. This step was carried out to selectively obtain the fragment lengths of interest from the reaction mixture containing unused enzymes, ligation reaction components, concatemers, and adapter dimers. Immediately after the 15 min incubation, pre-thawed magnetic Agencourt AMPure XP beads (Beckman Coulter, USA) were briefly vortexed to resuspend the beads into a resulting homogenous solution. Subsequently, a considerable working volume of the homogenized Agencourt's beads was transferred into a medium-sized rectangular trough (Thermofisher Scientific, USA), and 80 µl was added to each barcoded library in the PCR-plate wells. The component of each well was thoroughly mixed by pipetting up and down and then, the plate was properly sealed with a micro seal (Biorad, California, USA) and further incubated for 5 min at room temperature in the PCR fume hood. The incubation step was undertaken to allow proper binding of the indexed cDNA library on the silica-coated paramagnetic beads. After incubation, with the cDNA now bound to the beads, the PCR plate was positioned into a 96-plate PCR magnetic stand (PerkinElmer, USA), causing the beads to stick to the outer part of the tube in contact with the magnet separating from the solution. Following a clear separation, the supernatant was gently pipetted out prior to subsequent ethanol wash steps. DNA is insoluble in alcohol, thus the immobilized bead-bound cDNA which formed visible clumps is retained during the ethanol washes. Briefly, with the plate still positioned on the magnetic stand, 200 µl of freshly constituted 80% ethanol each were added to individual well pellet and supernatant was gently pipetted out while avoiding any contact with the beads and this same wash step was repeated. After two complete wash steps, the beads were air-dried for about 5 min to evaporate all residual ethanol until the beads were dry, with the plate still positioned on the magnetic rack. The beads were eluted while maintaining their glossy dark-brown colour, as over-dried beads appear with lighter brown shades with cracks and result in lower RNA

recovery. The PCR plate was removed from the magnet stand, and immediately these bead-bound cDNA was re-suspended in solution by adding 52.5 µl of Buffer EB in individual wells. The cDNA-bead solution was thoroughly mixed by pipetting up and down 13 times, briefly centrifuged using a Hermle MK plate centrifuge (Hermle Labortechnik, Germany), and incubated at room temperature for 2 min. The PCR plate was repositioned on the magnetic base for magnetic separation of the beads and purified indexed cDNA in a clear solution for about four min. After the solution was sufficiently clear, 50 µl each of the clear supernatant containing the recovered barcoded cDNA libraries was withdrawn and individually transferred to a fresh 96-well PCR plate.

Similarly, 50 µl of Agencourt AMPure XP beads was added to the recovered cDNA libraries in the fresh PCR-plate. The same procedure for incubation and ethanol washes steps described in the previous paragraph was followed accordingly. However, for this final elution step after the pelleted bead bound cDNA had dried, 26 µl of buffer EB was added in the same manner placed on the magnetic stand for separation. Finally, after complete separation, marked by a clear supernatant containing the indexed cDNA library, 23.5 µl of the supernatant each were individually transferred in a new PCR plate. These purified libraries in the PCR plate were first sealed using a plate micro-seal film, tightly rolled over twice using a speedball Bio-RAD Roller, securing all of its edges. The quality of the generated libraries was first validated using a Qubit fluorometer and Agilent Bioanalyzer 2100 (Agilent Technologies, USA) for library quantification and fragment size determination respectively, before proceeding downstream with the Illumina sequencing.

## 3.11 Quality assessment of cDNA library products

The two validation steps assayed on the indexed cDNA libraries were based on their concentrations and fragment size distribution. Although the cDNA concentrations were initially quantified prior to the library preparation, nevertheless, as expected the enzymatic shearing of the DNA template, barcoding and series of wash steps may have altered the concentration. In addition, determining the size distribution of the libraries was crucial at this stage to ascertain if cDNA templates were effectively fragmented to the desired size distribution. Most importantly, the parameters of library quality validation are needed for further library normalization, unit conversions to nano Molar and pooling for cluster generation. Following the Qubit 3.0 methodology previously described in section 3.7, the concentrations of libraries were determined fluorometrically, and the fragment sizes assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA).

Library quality evaluations were carried out by analysing 1 µl of each library sample on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The quantification is based on the principle of capillary electrophoresis, in which separation occurs through a glass micro-phoretic chip embedded in a plastic frame with its reverse side comprising of different channels through which nucleic acids are separated with an optical detector and electrode array in the bioanalyzer. First, the gel-dye working mixture stored at 4°C which had been prepared prior to this experiment was taken out and allowed to normalize to room temperature. The mixture was prepared by combining gel and dye reagents from the Bioanalyzer kit, filtering the mixture through a spin filter and a 3-min centrifugation to collect the filtrate. In addition, the system was washed by pipetting 350 µl nuclease-free water into the wash chip and running the chip in the bioanalyzer for 1 min. This wash protocol was carried out before and after every run. After complete wash time, the water was discarded, and the wash chip was saved for future runs. Following this, the electrophoretic chip was carefully placed in the priming station and, 9 µl of the gel-dye mixture was loaded into one of the gel positions marked with a dark circular G on the chip. It was important to touch the bottom of the wells during loading and avoid the expelling bubbles as this may interfere with the electrophoretic separation of nucleic acids through the channels. The syringe was brought up to the 1 ml mark and the lid of the priming station gently shut until a click was heard. The plunger was gently and steadily pushed down until it touched the hole and allowed for 1 min for the gel-dye mix to be evenly distributed. After 1 min, the plunger was then released and gently returned to the 1 ml mark, and 9 µl of the gel was gently added to the other G positions while avoiding dispensing bubbles. Additionally, 5 µl marker solution and 1 µl ladder each were added to the ladder wells which serves to calibrate the lower and upper markers for samples. At this point, 1 µl of randomly selected samples from each group was added to the G wells, vortexed using a special foam cushioning vortexer that came with the Bioanalyzer machine and allowed for 1 min to settle. The chip was then loaded into the Bioanalyzer machine, using the high-sensitivity DNA assay option, and ran for 4 to 5 min. Finally, a comprehensive report was created for each sample containing the electropherogram, fragment size distribution, molarity, concentration, and gel smears.

## 3.12 Sample preparations for Illumina sequencing

Prior to the final stage of sequencing chemistries, the barcoded libraries still need to undergo additional processing steps for successful cluster generation for Illumina sequencing. These processes majorly include normalization, pooling, and denaturation of the cDNA libraries as well as the

incorporation of a control which is critical for optimal generation of high-quality sequencing data. Hence, this pre-sequencing steps will be described in detail in this section.

### 3.12.1 Normalization and pooling of cDNA libraries

During the library preparation step, the enzymatic shearing, tagging and series of wash steps, as presumed may have altered the normalized starting cDNA material, thus requiring another concentration equilibration of the resultant cDNA library samples. Equalization of cDNA concentrations across libraries has been proven to minimize sequencing coverage variation and as well improve transcriptome diversity by providing a more uniform sequencing coverage for all samples (Bogdanova *et al.,* 2008; Bacher *et al.,* 2022). As a result, the Illumina protocol includes a standard normalization of libraries to 4 nM, or 2 nM for cluster generation. However, recent sequencing technologies, such as Illumina, necessitate the use of dsDNA fluorescent dyes specialized for precise quantification of dsDNA. These fluorometric approaches quantify dsDNA concentrations in nanograms per microliter (ng/µl), thus requiring their unit conversion to nanomolar (nM) before normalizing to equimolar concentrations. The conversion of cDNA library sample was achieved using the formula below.

$$Concentrations\ in\ nM\ = \frac{Concentration\ in\ ng/\mu l}{(660\ g/mol\ x\ average\ library\ size\ in\ base\ pair)} X\ 10^6$$

This nM formula considered four parameters which includes the sample Qubit concentrations in ng/µl, the mean mass of sodium salt of DNA for equal proportions of AT and GC pairs, the average library size in base pairs as obtained from the Agilent Bioanalyzer and their multiplication to the right orders of magnitude. After converting to nM, appropriate diluent calculations were performed using the resuspension (RSB) buffer to achieve a 4 nM concentration for samples.

Normalized samples which have been previously barcoded as described in section 3.9 were all pooled into one tube for cluster generation. This multiplexing was achieved by transferring 5 µl equal volume of each library sample into a 2 ml pre-labelled Eppendorf tube, thoroughly vortexed and briefly spun down using a bench top centrifuge. It is important to emphasize that each pooled sample would be individually identified using their unique barcode post-sequencing, a process called de-multiplexing.

### 3.12.2 Denaturation and dilution of the cDNA libraries

In preparation for denaturation and dilution, the hybridization buffer (HT1) was removed from - 20°C and allowed to thaw. After it had thawed, the buffer was stored at 4°C until needed for dilution. First,

5 µl of freshly prepared 0.2 M NaOH (Sigma Aldrich, USA) was added into a sterile 1.5 ml micro centrifuge tube. The normalized 4nM pooled library was briefly vortexed and 5 µl transferred into the 1.5 ml containing NaOH, making it a total volume of 10 µl. The tube content was briefly vortexed to homogenize and centrifuged at 280 *x g* for 60 s. Thereafter, a timer was set, and the tube was incubated at room temperature for 5 min. During this incubation period the double stranded DNA molecules unwind as the hydrogen bonds are broken, consequently resulting in single-stranded DNA molecules, in a process known as Denaturation.

After the completed incubation time, the 10 µl of the 4 nM denatured DNA library was further diluted to 20 picomolar (pM) by adding 990 µl of pre-chilled hybridization buffer HT1 (Illumina, Inc., USA). This was again, diluted further by combining 300 µl of pre-chilled HT1 and 300 µl of the 20 pM libraries to achieve a final library concentration of 10 pM. The final library was thoroughly mixed by inverting the tube up and down repeatedly.

### 3.12.3 Incorporation of the PhiX control

PhiX is an icosahedral, non-tailed bacteriophage with a 5386-nucleotide genome contained in a single stranded DNA. PhiX is often employed as a control for Illumina sequencing due to its short, well-defined genomic sequence. PhiX29 DNA polymerase's high proof-reading ability also helps eliminate erroneous mutations (Esteban *et al.,* 1993). Hence, the denaturation of PhiX control was performed to get optimum cluster density. To do this, a solution of 10 nM Tris-HCl, pH 8.5, was prepared using 2 µl of the 10 nM PhiX library and 3 µl of elution buffer (Qiagen, Hilden, Germany). After adding 5 µl 0.2N NaOH and centrifuging at 280 x *g* for 1 min, the resulting 4 nM PhiX was kept at room temperature for 5 min. The denatured PhiX library (4 nM) was diluted to 20 pM by adding 990 µl of cold HT1 buffer to 10 µl of the sample. A PhiX control spike-in of 15% was used to introduce variety into the library by mixing 90 µl of 20 pM denatured and diluted PhiX with 510 µl of 10 pM denatured and diluted library to make a final volume of 600 µl. Before denaturation and subsequent reactor loading, the final library was first incubated on ice. Diversity in resulting nucleotide sequences is essential for efficient run times and quality data generation. Thus, they are subsequently used in base calling and cycle quality score computations.

## 3.13 Illumina sequencing

Fragments ranging from 350-600 bp were selected for the final library and were sequenced on a MiSeq platform (Illumina, San Diego) for 300 cycles (300 X 2 cycles; paired ends). Chicken faecal samples with a multiplex of 27 libraries and 4 control samples were sequenced.

## 3.14 Quality sequence processing and data assembly

The raw sequencing reads were processed using the genome detective pipeline (Vilsker *et al.,* 2019) with integrates different underlying software for accurate viral data analysis. In this pipeline, the removal of low quality, uninformative reads and adapter trimming was achieved using Trimmomatic (Bolger *et al.,* 2014), whereas in-depth quality control of the generated reads was performed with FQC (Brown *et al.,* 2017) which integrates FASTQ results before and after trimming. The paired-end sequences were assembled into contigs using metaSPADES (Nurk *et al.,* 2017). Scaffolds was classified using BLASTx and BLASTn search for reference sequences against all National Center for Biotechnology Information (NCBI) reference virus databases. Scaffolds with significant hits were joined and aligned by amino acids and nucleotides using Advanced Genome Aligner (AGA) downstream analysis exported as fasta files for further and final contigs. All sequences with full genome or complete RdRp were processed for submission to GenBank.

## 3.15 Addressing the issues of index-hopping and contaminants

In accordance with good laboratory practices and aseptic working measures, various studies have demonstrated that contamination cannot be totally avoided in metagenomics-based microbiome research (Asplund *et al.,* 2019; Jurasz *et al.,* 2021; Lou *et al.,* 2023). The presence of internal and external contaminants has posed a significant challenge in disguising artifacts from main result outcomes in microbiome and virome datasets (Hornung *et al.,* 2019; Zinter *et al.,* 2019; Jurasz *et al.,* 2021). Notably, among other issues of contaminants in sequencing data, the problem of external reagents contamination (kitome) and internal index switching or hopping has been reported by several studies to be poorly managed. Therefore, it is noteworthy that a mere 30% of microbiome studies have included controls (Hornung *et al.,* 2019); consequently, researchers have emphasized the need to address contamination in microbiome sequencing studies (Salter *et al.,* 2014; Eisenhofer *et al.,* 2019; Zinter *et al.,* 2019; Lou *et al.,* 2023).

In this study, viruses were assumed to be contaminants due to index-hopping from another library if the total read count was less than 0.1% of the most abundant read count of the same virus(es) (Wille *et al.,* 2021). In addition, viruses detected in the negative control libraries of sterile water or reagent mix incorporated at different stages, and/or belonging to the same clades as those detected in blank libraries (Edmonds and Williams 2017; Porter *et al.,* 2021; Lou *et al.,* 2023) was presumed to have originated from contamination that is most likely linked to laboratory reagents (Porter *et al.,* 2021; French *et al.,* 2022). Hence, those viruses were eliminated from the chicken libraries and excluded

from all further downstream investigations. Therefore, in addition to the inclusion of negative controls at key steps, the decontamination process of reagent master mixes holds a possibility to enhance the precision and sensitivity of inferences made in the microbiome (Stinson *et al.,* 2019) and metagenomic studies (Asplund *et al.,* 2019; Zinter *et al.,* 2019; Porter *et al.,* 2021).

## 3.16 Analysis of RNA viral diversities and abundance as a function of age and season

### 3.16.1 Viral abundance determination

To adjust for variations in read depth across libraries, the measure of abundance was expressed as the read count of reads per million. This was achieved by dividing the read count by the total number of reads in the library and then multiplying the quotient by one million. The differential viral abundance analysis was carried out from the generated fasta-QC files using the edgeR package (R core team, 2021). The relative abundance values for viruses were calculated for all viral families found present in each age group characterized by each season. These values were further used for pheat map (Gu *et al.,* 2016), dplyr (Wickham *et al.,* 2019), and reshape2 tools representations in ggplot2 (Wickham *et al.,* 2016).

### 3.16.2 Diversity measures and statistical significance

The exclusive use of singular numerical indices to ascertain the community structure and ecological status of a site, sample or environment reduces the true significance of its biological diversity. In accordance with existing literature, it has been recommended to employ multiple indices for the assessment of diversity (Beaugrand *et al.,* 2002; Bandeira *et al.,* 2013). The fundamental principles underlying the community diversity theory are predicated upon two salient attributes, namely the numerical abundance of species and the evenness of their distribution within an ecological system (Washington 1984; Fedor and Zvaríková 2019). Therefore, the utilization of an index or multiple indices that establish a connection between these two elements of diversity is of utmost importance.

#### 3.16.2.1 Alpha diversity measures

Alpha diversity of the RNA viruses was measured to determine the average species diversity in the GIT of the studied chickens. This analysis of alpha diversity, which pertains to the diversity within each sample, involved the examination of richness, or the number of viral species, sobs (number of observed genera) as well as the Shannon- Weiner and Simpson indices. The alpha diversity measures, which is viral diversity within each sample explored were Sobs (number of observed genera),

Shannon-Weiner and Simpson index was computed (average of 1000 replicates) on sample sequences from each sample as well as each group. The Shannon-Weiner and Simpson indices are utilized to evaluate the richness and evenness of a community. It is noteworthy, however, that the Shannon index places greater emphasis on richness, while the Simpson index prioritizes evenness (Fedor and Zvaríková 2019). The Kruskal-Wallis's rank sum test, a non-parametric method for conducting one-way analysis of variance, was employed to ascertain differences in alpha diversity richness for more than two groups. This approach is distinct from Mann-Whitney U, which only considers two groups. A post-hoc analysis, specifically the Bonferroni method and Turkey's test for multiple comparisons was conducted as a followup to determine which groups means were significantly different from each other.

#### 3.16.2.2 Beta diversity measures

The Bray-Curtis dissimilarity index matrix was used as input for ordination analysis to compare the compositions of RNA viral communities of all faecal samples and groups. The beta diversity metrics for age and season were determined using the abundances of the principal coordinate analysis (PCoA) based on the Bray-Curtis dissimilarity index (Bray and Curtis 1957) and Jaccard presence of absence theorem (Real and Vargas 1996; Verma and Aggarwal 2020). This was carried out on all sub-samples from each group. These dissimilarity matrices were used as input for statistical testing of significance that were computed as PCoA plots. Furthermore, the statistical evaluation of the Bray-Curtis beta diversity metrics results was conducted using the Adonis method. This is a permutational multivariant analysis of variance Adonis under a reduced model for a) age b) season and c) age and season. variation in richness (alpha diversity). Overall, the chosen $\alpha$ was 0.05, hence a p-value $< 0.05$ was considered significantly different for drawing inferences from the results of all the statistical analysis conducted.

Generally, the effects of age and seasons on gut virome diversities investigated using ecological diversity metrics; alpha and beta diversity was to determine if at least one age group or season are significantly different from each other. Thus, study speculated that the age group with younger chickens may have a greater number of viruses or more diverse viruses than those at mature stages. Similarly, for seasons, that their diversity and abundance may differ.

## 3.17 Data visualization

The data visualizations and statistical analyses employed in the present study were mainly performed using the R Software (version 4.3.0) (R core Team). The paired end reads from all 27 libraries was collected after sequencing quality cleaning (QC) of the raw reads using FASTQC. These abundance data generated as reads and taxonomy tables were used as input files for further visualization on R, while utilizing a suite of packages including Vegan (Oksanen *et al.,* 2013), ggplot2 (Wickham *et al.,* 2016), cowplot (Wilke *et al.,* 2019), pheatmap (Kolde and Kolde 2018), and tidy2verse (Wickham and Wickham 2019). Determination of differences between the sample groups observed in relation to chicken age and seasons was explored. Linear models and analysis of variance were employed to test for these differences in richness and diversity indices across groups (age and season).

## 3.18 Evolutionary relationship of the identified RNA viruses

The analysis of evolutionary phylogenies was conducted with viruses having full-length genomes with ≥ 90% of genome coverage or partial genomes that have complete RNA-dependent RNA polymerase (RdRp) genes. Multiple sequence alignment was carried out using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) (Katoh and Standley 2013) comparing them with global sequences retrieved from GenBank and best hit results of BLAST searches e-value $10^{-3}$. In summary, the nucleotide sequences of the complete phylogeny were performed using the neighbour-joining method and maximum likelihood approach depending on each specific virus case and a bootstrapping re-sampling investigation was done using 1000 replicates to measure the tree topologies (Tamura *et al.,* 2021). In the construction of each phylogenetic tree, through the maximum likelihood method, the Molecular Evolutionary Genetics Analysis (MEGA v11.0) software option finds the best-fit substitution models of evolution was employed. The resulting trees of phylogenies were visualized and annotated using FigTree v1. 3.1(Rambaut 2009).

## 3.19 Viral classification

Virus name assignment was done using the latest 2022 MSL38 International Committee on Taxonomy of Viruses (ICTV) taxonomy release (https://ictv.global/taxonomy) and the host association was done using virus host database (https://www.genome.jp/virushostdb/). The detection of novel viral species is based upon the presence of either less than 80% RdRp protein identity or less than 80% genome identity when compared to viruses that have been previously described (Wille *et*

*al.,* 2021) in agreement with the ICTV species demarcation criteria of specific viral species and families.

# CHAPTER 4. Results

## 4.1 Viral enrichment of samples

### 4.1.1 The enrichment of viral RNA

The results of the quantification and quality assessment of the extracted RNA after RNA extraction and DNase treatment (concentrated) are presented in Table 4.1.1. The RNA extraction process was carried out on 28 samples (27 chicken samples and control N1) excluding the negative control N2 consisting of the final elution buffer. However, it is worth noting that 12 of the chicken faecal samples had concentration values that were considered too low to be quantified on the Qubit 3.0 fluorometer (Table 4.1.1).

Table 4.1.1 Evaluation of chicken RNA concentrations and quality ($A_{260}/A_{280}$) and quantification post-extraction and DNase treatment

| Sample tag | Sample ID | Quantification after RNA extraction (ng/µl) | Quality 260/280 ratio after RNA extraction | Quantification after concentration of DNase treated samples (ng/µl) | Quality 260/280 ratio after DNase treatment |
|---|---|---|---|---|---|
| 2S1 | CH/2022/2S1 | 1.12 | 1.93 | 3.90 | 2.09 |
| 2S2 | CH/2022/2S2 | Too low | 1.76 | Too low | 1.98 |
| 2S3 | CH/2022/2S3 | 0.90 | 2.30 | 3.22 | 2.00 |
| 2S4 | CH/2022/2S4 | Too low | 2.40 | 0.30 | 2.00 |
| 2S5 | CH/2022/2S5 | Too low | 1.78 | Too low | 1.99 |
| 4S1 | CH/2022/4S1 | Too low | 1.77 | 1.05 | 2.01 |
| 4S2 | CH/2022/4S2 | 0.70 | 1.79 | 1.90 | 2.20 |
| 4S3 | CH/2022/4S3 | Too low | 2.36 | Too low | 2.06 |
| 4S4 | CH/2022/4S4 | Too low | 1.99 | Too low | 2.00 |
| 4S5 | CH/2022/4S5 | 2.00 | 1.89 | 3.09 | 2.10 |
| 7S1 | CH/2022/7S1 | Too low | 1.98 | Too low | 2.20 |
| 7S2 | CH/2022/7S2 | 4.86 | 2.10 | 6.25 | 2.01 |
| 4W1 | CH/2022/4W1 | Too low | 1.76 | 0.29 | 2.30 |
| 4W2 | CH/2022/4W2 | 5.95 | 1.99 | 7.49 | 2.19 |
| 4W3 | CH/2022/4W3 | 6.40 | 1.87 | 10.70 | 2.03 |
| 4W4 | CH/2022/4W4 | 6.30 | 1.80 | 9.04 | 2.00 |
| 4W5 | CH/2022/4W5 | 9.06 | 1.96 | 11.27 | 2.01 |

| | | | | | |
|---|---|---|---|---|---|
| 7W1 | CH/2022/7W1 | 0.80 | 1.75 | 3.2 | 1.99 |
| 7W2 | CH/2022/7W2 | 3.30 | 1.88 | 4.02 | 2.06 |
| 7W3 | CH/2022/7W3 | Too low | 1.81 | 0.28 | 2.03 |
| 7W4 | CH/2022/7W4 | 5.4 | 2.17 | 6.32 | 2.00 |
| 2W1 | CH/2022/2W1 | Too low | 1.83 | 1.02 | 2.02 |
| 2W2 | CH/2022/2W2 | 17.40 | 1.99 | 21.29 | 2.00 |
| 2W3 | CH/2022/2W3 | Too low | 1.87 | 0.98 | 2.13 |
| 2W4 | CH/2022/2W4 | 16.19 | 2.01 | 19.40 | 1.99 |
| 2W5 | CH/2022/2W5 | Too low | 1.79 | 0.85 | 2.10 |
| 7W5 | CH/2022/7W5 | 4.20 | 1.75 | 5.33 | 1.97 |
| N1 | Control/PBS/A/E | Too low | - | Too low | - |
| N2 | Control/extraction | Too low | - | Too low | - |

The highest RNA concentration recorded was from sample 2W2 with 17.40 ng/µl after extraction and 21.29 ng/µl after concentration. Sample 2W4 followed closely, demonstrating a comparable RNA concentration with 2W2. In addition, samples such as 4S2 (0.70 ng/µl), 7W1 (0.80 ng/µl), and 2S3 (0.90 ng/µl) had very low RNA concentrations after extraction. As anticipated, after purifying and concentrating the DNase-treated samples, there was a significant increase in viral RNA concentrations across all samples. It is noteworthy that 58% of the samples with too-low RNA concentrations, mainly from the winter samples (2W1, 2W3, 2W5, 4W1, 7W3, 2S4, and 4S1) were at this point quantifiable with observed numerical values, although they showed quite low concentrations (> 0.25 ng/µl). Nonetheless, the outcomes of the remaining five summer samples and negative control (N1) were still too low to be read by the Qubit fluorometer (Table 4.1.1).

A further assessment of the quality of the extracted RNA samples revealed that the purity ($A_{260}/A_{280}$) was found to be within the range of 1.76 and 2.40. It was observed that RNA quality improved after purification, as expected, with values ranging from 1.97 to 2.20 (Table 4.1.1). In general, the mean RNA integrity values of the samples after the extraction and subsequent purification procedures were 1.91 and 2.05, respectively.

### 4.1.2 The ribosomal RNA removal

Table 4.1.2 presents an assessment of the quality ($A_{260}/A_{280}$) estimate and quantity of chicken viral RNA before and after rRNA removal.

Table 4.1.2 Evaluation of chicken viral RNA quality ($A_{260}/A_{280}$) and quantity after rRNA removal

| Sample tag | Sample ID | RNA concentration | RNA concentration | Quality 260/280 | Percentage RNA concentration |
|---|---|---|---|---|---|

| | | (ng/µl) before rRNA removal | (ng/µl) after rRNA removal | ratio after rRNA removal | decrease per sample (%) |
|---|---|---|---|---|---|
| 2S1 | CH/2022/2S1 | 3.90 | 1.30 | 1.93 | 66.67 |
| 2S2 | CH/2022/2S2 | Too low | Too low | 1.76 | - |
| 2S3 | CH/2022/2S3 | 3.22 | 1.10 | 2.30 | 65.84 |
| 2S4 | CH/2022/2S4 | 0.30 | Too low | 2.40 | - |
| 2S5 | CH/2022/2S5 | Too low | Too low | 1.78 | - |
| 4S1 | CH/2022/4S1 | 1.05 | 0.35 | 1.77 | 66.67 |
| 4S2 | CH/2022/4S2 | 1.90 | 0.50 | 1.79 | 73.68 |
| 4S3 | CH/2022/4S3 | Too low | Too low | 2.36 | - |
| 4S4 | CH/2022/4S4 | Too low | Too low | 1.99 | - |
| 4S5 | CH/2022/4S5 | 3.09 | 1.01 | 1.89 | 67.31 |
| 7S1 | CH/2022/7S1 | 6.25 | 2.1 | 2.10 | 66.40 |
| 7S2 | CH/2022/7S2 | Too low | Too low | 1.98 | - |
| 4W1 | CH/2022/4W1 | 0.29 | Too low | 1.76 | - |
| 4W2 | CH/2022/4W2 | 7.49 | 3.03 | 1.99 | 59.55 |
| 4W3 | CH/2022/4W3 | 10.70 | 3.59 | 1.87 | 66.45 |
| 4W4 | CH/2022/4W4 | 9.04 | 4.01 | 1.80 | 55.64 |
| 4W5 | CH/2022/4W5 | 11.27 | 3.85 | 1.96 | 65.84 |
| 7W1 | CH/2022/7W1 | 3.2 | 1.10 | 1.75 | 65.63 |
| 7W2 | CH/2022/7W2 | 4.02 | 1.34 | 1.88 | 66.67 |
| 7W3 | CH/2022/7W3 | 0.28 | Too low | 1.81 | - |
| 7W4 | CH/2022/7W4 | 6.32 | 2.21 | 2.17 | 65.03 |
| 2W1 | CH/2022/2W1 | 1.02 | 0.45 | 1.83 | 55.88 |
| 2W2 | CH/2022/2W2 | 21.29 | 11.40 | 1.99 | 45.45 |
| 2W3 | CH/2022/2W3 | 0.98 | Too low | 1.87 | - |
| 2W4 | CH/2022/2W4 | 19.40 | 7.56 | 2.01 | 61.03 |
| 2W5 | CH/2022/2W5 | 0.85 | Too low | 1.79 | - |
| 7W5 | CH/2022/7W5 | 5.33 | 2.17 | 1.75 | 59.28 |
| N1 | Control/PBS/A/E | Too low | Too low | - | - |
| N2 | Control/extraction | Too low | Too low | - | - |
| N3 | Control/rRNA | Too low | Too low | - | - |

A significant reduction in concentrations of RNA was observed in all samples after the depletion of rRNA (Table 4.1.2). In addition, it was observed that in 60% (17/27) of the samples, there was a significant reduction of over 50% in the concentration of RNA, although, sample 4W2 exhibited quite

a lower percentage decrease, about 45.5%. The concentration of the purified samples with "too-low-outcomes" (2S4, 2W3, 2W5, 4W1, and 7W3), which experienced an increase after DNase treatment, also exhibited a significant decrease post rRNA depletion. It was observed that the concentration of these samples was below the quantifiable threshold of 0.25 ng/µl as determined by the Qubit 3.0 fluorometer (Table 4.1.2).

## 4.2 Complementary DNA synthesis and whole transcriptome amplification.

The results obtained after WTA were observed to be very high to be quantified. Table 4.2.1 depicts the results of the cDNA concentrations of all samples, after transcriptome amplification, dilution and normalization.

Table 4.2.1 Fluorometric quantification of cDNA post amplification and normalization

| Sample tag | Sample ID | Concentration of cDNA after amplification (before dilution) | Concentration of cDNA after 10x dilution (ng/ml) | Concentration of normalized cDNA (ng/µl) |
|---|---|---|---|---|
| 2S1 | CH/2022/2S1 | Too high | 52500 | 83.4 |
| 2S2 | CH/2022/2S2 | Too high | 49100 | 87.0 |
| 2S3 | CH/2022/2S3 | Too high | 50700 | 90.0 |
| 2S4 | CH/2022/2S4 | Too high | 42800 | 88.9 |
| 2S5 | CH/2022/2S5 | Too high | 42000 | 94.8 |
| 4S1 | CH/2022/4S1 | Too high | 50100 | 89.8 |
| 4S2 | CH/2022/4S2 | Too high | 45600 | 98.9 |
| 4S3 | CH/2022/4S3 | Too high | 48500 | 96.0 |
| 4S4 | CH/2022/4S4 | Too high | 51300 | 80.6 |
| 4S5 | CH/2022/4S5 | Too high | 52400 | 86.4 |
| 7S1 | CH/2022/7S1 | Too high | 35300 | 95.8 |
| 7S2 | CH/2022/7S2 | Too high | 23800 | 96.7 |
| 4W1 | CH/2022/4W1 | Too high | 24600 | 98.9 |
| 4W2 | CH/2022/4W2 | Too high | 51600 | 89.8 |
| 4W3 | CH/2022/4W3 | Too high | 51200 | 94.4 |
| 4W4 | CH/2022/4W4 | Too high | 51400 | 96.8 |
| 4W5 | CH/2022/4W5 | Too high | 49600 | 91.2 |
| 7W1 | CH/2022/7W1 | Too high | 44400 | 93.6 |
| 7W2 | CH/2022/7W2 | Too high | 52400 | 89.0 |
| 7W3 | CH/2022/7W3 | Too high | 41500 | 98.6 |

| | | | | |
|---|---|---|---|---|
| 7W4 | CH/2022/7W4 | Too high | 36800 | 91.8 |
| 2W1 | CH/2022/2W1 | Too high | 51300 | 94.2 |
| 2W2 | CH/2022/2W2 | Too high | 51700 | 95.4 |
| 2W3 | CH/2022/2W3 | Too high | 49000 | 95.8 |
| 2W4 | CH/2022/2W4 | Too high | 51200 | 89.0 |
| 2W5 | CH/2022/2W5 | Too high | 51800 | 87.2 |
| 7W5 | CH/2022/7W5 | Too high | 43800 | 98.8 |
| N1 | Control/PBS/A/E | Too high | 13100 | 87.0 |
| N2 | Control/extraction | Too high | 3900 | 90.4 |
| N3 | Control/rRNA | Too high | 1560 | 81.9 |
| N4 | Control/cDNA/WTA | Too high | 11300 | 82.0 |

It was observed from the results in Table 4.1.2 that the concentrations of all samples, including controls N1 to N4, were "too high" after amplification. After the ten-fold dilution, most of the samples exhibited cDNA concentrations ranging from 41000 to 52500 ng/ml, except for samples 7S1,7S2, 4W1, 7W4 and the control samples (N1-N4) with lower concentrations. Furthermore, the result validation of the cDNA concentrations and the quality of the amplified cDNA were assessed, utilizing spectrophotometry as elaborated in Appendix 8.2. Based on the spectrophotometric data, it was determined that the average purity of amplified cDNA for all samples was 1.8. Subsequently, the normalization of varying concentrations, after dilution, resulted in values increasing within the acceptable range of 80.0 and 98.9 ng/ul for all samples (Table 4.1.2). These concentrations are deemed sufficient as starting material for the downstream QIAseq FX single-cell library preparation.

## 4.3 Generation of cDNA libraries

The quantification of standardized amplified cDNA with respect to the baseline are shown in Table 4.3.1.

Table 4.3.1 Quantification of normalized amplified cDNA concentration to library standard

| Sample tag | Concentration of cDNA after 10x dilution (ng/ µl) | Expected starting cDNA concentration in 10 µl as per library kit (ng/µl) | Expected chosen cDNA concentration range in 10 µl (ng/µl) | Confirmed concentration of normalized cDNA per 1 µl of sample (ng/µl) | Confirmed concentration of normalized cDNA range in 10 µl (ng/µl) |
|---|---|---|---|---|---|
| 2S1 | 52.5 | 200-1000 | 800-1000 | 83.4 | 833.9 |

| | | | | | |
|---|---|---|---|---|---|
| 2S2 | 49.1 | 200-1000 | 800-1000 | 87.0 | 869.8 |
| 2S3 | 50.7 | 200-1000 | 800-1000 | 90.0 | 900.0 |
| 2S4 | 42.8 | 200-1000 | 800-1000 | 88.9 | 889.1 |
| 2S5 | 42.0 | 200-1000 | 800-1000 | 94.8 | 947.8 |
| 4S1 | 50.1 | 200-1000 | 800-1000 | 89.8 | 898.0 |
| 4S2 | 45.6 | 200-1000 | 800-1000 | 98.9 | 989.1 |
| 4S3 | 48.5 | 200-1000 | 800-1000 | 96.0 | 959.9 |
| 4S4 | 51.3 | 200-1000 | 800-1000 | 80.6 | 805.8 |
| 4S5 | 52.4 | 200-1000 | 800-1000 | 86.4 | 864.3 |
| 7S1 | 35.3 | 200-1000 | 800-1000 | 95.8 | 957.5 |
| 7S2 | 23.8 | 200-1000 | 800-1000 | 96.7 | 966.4 |
| 4W1 | 24.6 | 200-1000 | 800-1000 | 98.9 | 988.7 |
| 4W2 | 51.6 | 200-1000 | 800-1000 | 89.8 | 898.2 |
| 4W3 | 51.2 | 200-1000 | 800-1000 | 94.4 | 943.9 |
| 4W4 | 51.4 | 200-1000 | 800-1000 | 96.8 | 968.2 |
| 4W5 | 49.6 | 200-1000 | 800-1000 | 91.2 | 911.0 |
| 7W1 | 44.4 | 200-1000 | 800-1000 | 93.6 | 936.3 |
| 7W2 | 52.4 | 200-1000 | 800-1000 | 89.0 | 890.0 |
| 7W3 | 41.5 | 200-1000 | 800-1000 | 98.6 | 986.4 |
| 7W4 | 36.8 | 200-1000 | 800-1000 | 91.8 | 917.5 |
| 2W1 | 51.3 | 200-1000 | 800-1000 | 94.2 | 940.9 |
| 2W2 | 51.7 | 200-1000 | 800-1000 | 95.4 | 954.2 |
| 2W3 | 49.0 | 200-1000 | 800-1000 | 95.8 | 957.8 |
| 2W4 | 51.2 | 200-1000 | 800-1000 | 89.0 | 980.1 |
| 2W5 | 51.8 | 200-1000 | 800-1000 | 87.2 | 872.0 |
| 7W5 | 43.8 | 200-1000 | 800-1000 | 98.8 | 987.6 |
| N1 | 13.1 | 200-1000 | 800-1000 | 87.0 | 869.8 |
| N2 | 3.9 | 200-1000 | 800-1000 | 90.4 | 903.9 |
| N3 | 1.6 | 200-1000 | 800-1000 | 81.9 | 818.6 |
| N4 | 11.3 | 200-1000 | 800-1000 | 82.0 | 801.4 |

The concentrations of the normalized samples fall within the expected threshold, with the least sample concentration being 801.4 ng/µl in 10 µl for a control sample (N4) (Table 4.3.1). Overall, an average sample concentration of 916.1 ng/µl in 10 µl after normalization was achieved.

### 4.3.1 Fragmentation and indexing of library

The concentrations result of the indexed cDNA libraries are illustrated in Table 4.3.1.1

Table 4.3.1.1 Determination of the concentrations of labelled cDNA libraries

| Sample tag | Sample ID | Concentration of indexed cDNA libraries (ng/µl) |
|---|---|---|
| 2S1 | CH/2022/2S1 | 5.38 |
| 2S2 | CH/2022/2S2 | 6.94 |
| 2S3 | CH/2022/2S3 | 6.58 |
| 2S4 | CH/2022/2S4 | 12.60 |
| 2S5 | CH/2022/2S5 | 15.80 |
| 4S1 | CH/2022/4S1 | 17.80 |
| 4S2 | CH/2022/4S2 | 19.20 |
| 4S3 | CH/2022/4S3 | 8.44 |
| 4S4 | CH/2022/4S4 | 9.06 |
| 4S5 | CH/2022/4S5 | 10.80 |
| 7S1 | CH/2022/7S1 | 10.10 |
| 7S2 | CH/2022/7S2 | 14.90 |
| 4W1 | CH/2022/4W1 | 12.70 |
| 4W2 | CH/2022/4W2 | 9.84 |
| 4W3 | CH/2022/4W3 | 11.00 |
| 4W4 | CH/2022/4W4 | 22.00 |
| 4W5 | CH/2022/4W5 | 12.10 |
| 7W1 | CH/2022/7W1 | 13.30 |
| 7W2 | CH/2022/7W2 | 12.60 |
| 7W3 | CH/2022/7W3 | 11.20 |
| 7W4 | CH/2022/7W4 | 8.92 |
| 2W1 | CH/2022/2W1 | 5.48 |
| 2W2 | CH/2022/2W2 | 9.92 |
| 2W3 | CH/2022/2W3 | 8.94 |
| 2W4 | CH/2022/2W4 | 14.10 |
| 2W5 | CH/2022/2W5 | 9.82 |
| 7W5 | CH/2022/7W5 | 18.60 |
| N1 | Control/PBS/A/E | 13.0 |
| N2 | Control/extraction | 8.96 |
| N3 | Control/rRNA | 9.64 |
| N4 | Control/cDNA/WTA | 3.14 |

The quantification results of the indexed complementary DNA libraries showed good concentrations were obtained, with the highest concentration of 22.0 ng/µl coming from a 4-week winter sample (4W4). Overall, of all the 31 samples, analysed including controls (N1 to N4), it was found that about 54.8% of these samples, or 17 in total, exhibited concentrations of 10 ng/µl or higher (Table 4.3.1.1).

The remaining samples, all displayed concentrations of 5 ng/µl or higher, except for a WTA control sample (N4). Notably, this control sample had the lowest cDNA concentration observed at 3.14 ng/µl.

### 4.3.2 Quality evaluation of cDNA library products

Based on the cDNA concentrations of the indexed samples (Table 4.3.1.1), the results of the fragment size quality analysis of selected samples from each age group and season including a control (N1; PBS/antibiotic/enzyme) were performed and are depicted in Appendix 8.3.1. The fragment size results of these samples (2S1, 2W4, 2W5, 4S2, 4W2, 7S31, 7W2, 7W4, N1) were observed to be between 150 and 600 bp (Appendix 8.3.1). The result depicted in Appendix 8.3.1 and 8.3.2 includes two bioanalyzer inherent controls denoted as C1 and C2 found in the last 2 lanes since only eleven samples can be validated at a given time (Appendix 3.3.2). The remaining samples that were subsequently analyzed in batches had similar results with those illustrated in Appendix 3.3.2. Overall library fragment sizes in each sample were seen to fall between 150 and 650 bp, however the average library size for all sample was 355 bp average.

### 4.3.3 Normalization of indexed cDNA libraries

The results obtained from the conversions of indexed cDNA libraries (ng/µl) to nM as well as their normalization are presented in Table 4.3.3.1.

Table 4.3.3.1 The mean library size, concentration in ng/µl and nanomolar of selected barcoded samples

| Sample tag | Sample ID | Barcoded DNA library concentration (ng/µl) | Bioanalyzer's mean library size (bp) | Barcoded DNA library concentration (nM) | Normalization of cDNA for cluster generation (nM) |
|---|---|---|---|---|---|
| 2S1 | CH/2022/2S1 | 5.38 | 326 | 25.0 | 4 |
| 2S2 | CH/2022/2S2 | 6.94 | 423 | 24.9 | 4 |
| 2S3 | CH/2022/2S3 | 6.58 | 348 | 28.6 | 4 |
| 2S4 | CH/2022/2S4 | 12.60 | 322 | 59.3 | 4 |
| 2S5 | CH/2022/2S5 | 15.80 | 304 | 78.7 | 4 |
| 4S1 | CH/2022/4S1 | 17.80 | 411 | 65.6 | 4 |
| 4S2 | CH/2022/4S2 | 19.20 | 328 | 88.7 | 4 |
| 4S3 | CH/2022/4S3 | 8.44 | 338 | 37.8 | 4 |
| 4S4 | CH/2022/4S4 | 9.06 | 306 | 44.9 | 4 |

| | | | | | |
|---|---|---|---|---|---|
| 4S5 | CH/2022/4S5 | 10.80 | 400 | 40.9 | 4 |
| 7S1 | CH/2022/7S1 | 10.10 | 322 | 47.5 | 4 |
| 7S2 | CH/2022/7S2 | 14.90 | 342 | 66.0 | 4 |
| 4W1 | CH/2022/4W1 | 12.70 | 430 | 44.7 | 4 |
| 4W2 | CH/2022/4W2 | 9.84 | 414 | 36.0 | 4 |
| 4W3 | CH/2022/4W3 | 11.00 | 308 | 54.1 | 4 |
| 4W4 | CH/2022/4W4 | 22.00 | 359 | 92.9 | 4 |
| 4W5 | CH/2022/4W5 | 12.10 | 328 | 55.9 | 4 |
| 7W1 | CH/2022/7W1 | 13.30 | 342 | 58.9 | 4 |
| 7W2 | CH/2022/7W2 | 12.60 | 325 | 58.7 | 4 |
| 7W3 | CH/2022/7W3 | 11.20 | 316 | 53.7 | 4 |
| 7W4 | CH/2022/7W4 | 8.92 | 424 | 31.9 | 4 |
| 2W1 | CH/2022/2W1 | 5.48 | 358 | 23.2 | 4 |
| 2W2 | CH/2022/2W2 | 9.92 | 353 | 42.6 | 4 |
| 2W3 | CH/2022/2W3 | 8.94 | 350 | 38.7 | 4 |
| 2W4 | CH/2022/2W4 | 14.10 | 305 | 70.0 | 4 |
| 2W5 | CH/2022/2W5 | 9.82 | 340 | 43.8 | 4 |
| 7W5 | CH/2022/7W5 | 18.60 | 353 | 79.8 | 4 |
| N1 | Control/PBS/A/E | 13.0 | 317 | 62.1 | 4 |
| N2 | Control/extraction | 8.96 | 406 | 33.4 | 4 |
| N3 | Control/rRNA | 9.64 | 415 | 35.1 | 4 |
| N4 | Control/cDNA/WTA | 3.14 | 386 | 12.32 | 4 |

Based on the results obtained, it can be observed that a higher initial concentration in ng/µl is positively correlated with a higher concentration in nM. The most notable effect, however, was observed with regards to the mean library size, where the concentration in nM exhibited a negative correlation with the average size of fragments. All samples were normalized to 4 nM for generated clusters (Table 4.3.3.1).

## 4.4 Illumina sequencing

The initial visualization of data quality prior to analysis revealed that the sequence data obtained from the Illumina Miseq platform attained a high Phred quality score of 38 (nearly 40). In addition, a total of 9,426276 sequences, were generated using the Illumina Miseq with a mere 703860 sequences being filtered out. A notable viral recovery rate was achieved, wherein from a total of 8,722,416 quality reads, 3,869638 reads greater than 1000 nt had hits with viruses. Overall, from the outcomes of the distribution of 8,722,416 obtained sequence reads, it was extrapolated that slightly above half of the total reads were non-viral (Figure 4.4.1). Interestingly, among the remaining 3,869638 sequence

reads, viral reads with hits to eukaryotic viruses far outweighed those aligning to phages (0.05%), insect viruses (0.04%) and plant viruses (0.03%) with a 99.87% prevalence as depicted in Figure 4.4.1.

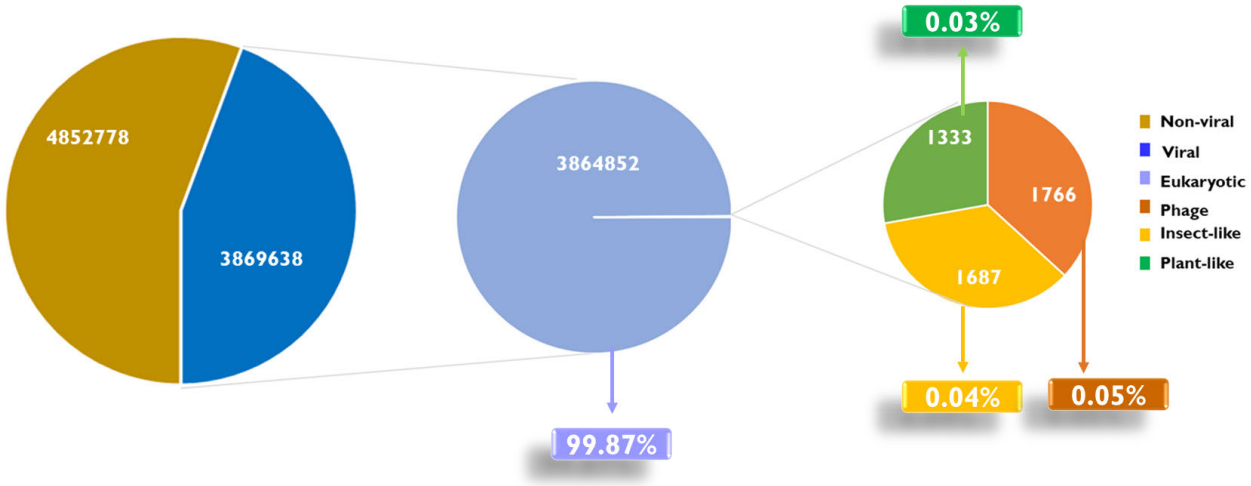



Figure 4.4.1 The distribution of viral and non-viral sequence reads obtained from studied chickens.

The obtained viral reads were collated from 27 individual chicken libraries. The viral reads composition in percentages of the 27 individual libraries of asymptomatic chickens' faecal matter is illustrated in Figure 4.4.2.

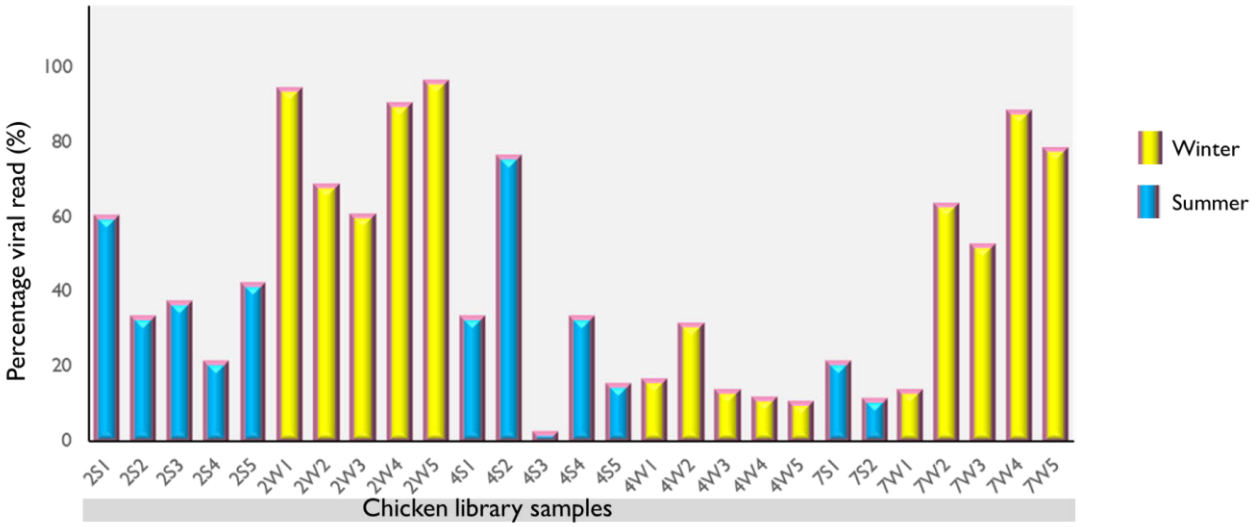



Figure 4.4.2 The percentage distribution of viral reads obtained from chicken faeces indicated by their respective 27 library.

Based on the results of these reads assembled into contigs, it was observed that the winter samples obtained at 2 weeks developmental age, namely 2W5, 2W1and 2W4 exhibited the highest percentage

of viral reads of 96%, 94% and 90%, respectively. Conversely, it is noteworthy that the samples 4S3 (2%), 4W5 (10%), and 7S2 (11%) showed very low viral percentage reads (Figure 4.4.2). The viral reads from the 27 chicken samples were subsequently *de novo* assembled into 4334 contigs.

## 4.5 Chicken enteric virome

In accordance with the taxonomic assignment of viruses, the identified viruses were classified based on their genome type, as well as their viral family and their respective genera. The taxonomic analysis of the 4334 viral contigs obtained from chicken faecal samples revealed 48 viral species from 4 kingdoms, 6 phyla, 8 classes, 11 order, 15 viral families, 21 genera and some unclassified viruses (Table 4.5.1). In addition, the results of the study based on genome type identified a total of 43 viral species pertaining to RNA viruses, 2 DNA viruses (tail phages belonging to the *Siphoviridae* family) and 3 unclassified viruses (Figure 4.5.1). Within the viruses with RNA genomes, 25 of them were dsRNA (52%), whereas 18 were ssRNA viruses (38%). The remaining 10% was shared between dsDNA (4%) and unclassified viruses (6%) as shown in Figure 4.5.1.

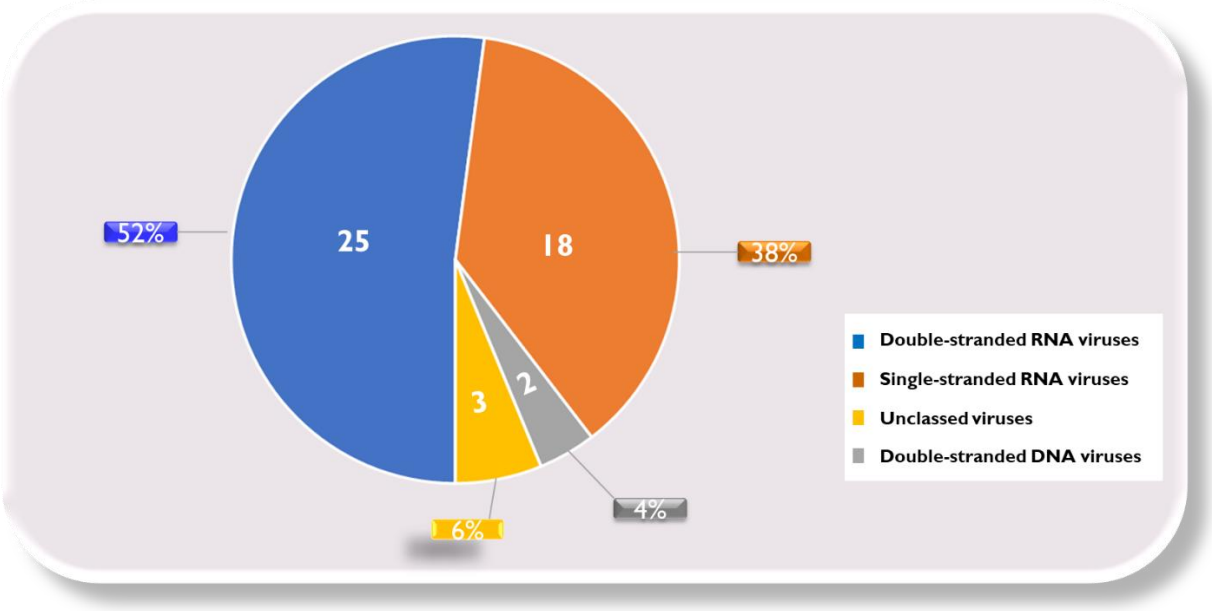



Figure 4.5.1 A representation of genome type distribution of viruses identified from the investigated chicken faecal samples.

Furthermore, at the family level, members of seven viral families *Coronaviridae, Picornaviridae, Reoviridae, Astroviridae, Caliciviridae, Picorbirnaviridae and Retroviridae* were found to be abundant across samples. It is noteworthy that the most abundant viral families are *Picornaviridae, Reoviridae,* and *Coronaviridae* with 1485, 762 and 646 viral contigs, respectively (Table 4.5.1).

Table 4.5.1 Taxonomic classification of chicken viral contigs from faeces based on their respective collection time points

| **Sample groups ID** | | | **2S** | **4S** | **7S** | **2W** | **4W** | **7W** | **Sum of contigs** | **Sum of identified viral genera** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Collection time (weeks)** | | | **2** | **4** | **7** | **2** | **4** | **7** | | **21 classified genera and unclassified viruses** |
| **Sum of Contigs of samples** | | | 886 | 835 | 208 | 658 | 801 | 946 | 4334 | |
| | | | | | | | | | | |
| **Assigned RNA virus species** | **Genome** | **Viral families** | | | | | | | | ***Genera*** |
| ***Avian coronavirus*** | **Single-stranded RNA** | ***Coronaviridae* (646)** | 111 | 141 | 41 | 40 | 182 | 131 | 646 | *Gammacoronavirus* |
| ***Avisivirus B*** | | ***Picornaviridae* (1485)** | 10 | 12 | 0 | 0 | 0 | 43 | 65 | *Avisivirus* |
| ***Chicken picornavirus 1*** | | | 59 | 55 | 0 | 34 | 59 | 45 | 252 | *Unclassified Picornaviridae* |
| ***Gallivirus A*** | | | 63 | 52 | 9 | 53 | 57 | 68 | 302 | *Gallivirus* |
| ***Megrivirus A*** | | | 0 | 0 | 11 | 0 | 11 | 4 | 26 | *Megrivirus* |
| ***Megrivirus C*** | | | 0 | 54 | 1 | 5 | 55 | 130 | 245 | |
| ***Orivirus A*** | | | 27 | 24 | 0 | 31 | 0 | 15 | 97 | *Orivirus* |
| ***Quail picornavirus QPV1/HUN/2010*** | | | 45 | 53 | 0 | 15 | 7 | 37 | 157 | *Unclassified Picornaviridae* |
| ***Sicinivirus A*** | | | 70 | 81 | 16 | 33 | 93 | 48 | 341 | *Sicinivirus* |
| ***Chicken astrovirus*** | | ***Astroviridae* (173)** | 48 | 38 | 0 | 31 | 48 | 8 | 173 | *Avastrovirus* |
| ***Bavaria virus*** | | ***Caliciviridae* (131)** | 47 | 36 | 0 | 48 | 0 | 0 | 131 | *Bavovirus* |
| ***Avian leukosis virus*** | | ***Retroviridae* (304)** | 30 | 28 | 8 | 0 | 22 | 33 | 121 | *Alpharetrovirus* |
| ***Rous sarcoma virus*** | | | 57 | 34 | 2 | 18 | 32 | 40 | 183 | |
| ***Avihepevirus magniiecur*** | | ***Hepeviridae* (107)** | 0 | 24 | 0 | 14 | 36 | 33 | 107 | *Avihepevirus* |
| ***Avian orthoreovirus*** | **Double-stranded RNA** | ***Reoviridae* (762)** | 105 | 3 | 0 | 85 | 62 | 31 | 286 | *Orthoreovirus* |
| ***Rotavirus F*** | | | 0 | 0 | 0 | 78 | 29 | 17 | 124 | *Rotavirus* |
| ***Rotavirus G*** | | | 94 | 90 | 30 | 67 | 46 | 25 | 352 | |
| ***Infectious bursal disease virus*** | | ***Birnaviridae* (4)** | 0 | 0 | 0 | 0 | 0 | 4 | 4 | *Avibirnavirus* |
| ***Chicken picobirnavirus*** | | ***Picobirnaviridae* (391)** | 6 | 10 | 8 | 8 | 21 | 6 | 59 | Unclassified *Picobirnavirus* |
| ***Orthopicobirnavirus hominis*** | | | 5 | 4 | 2 | 16 | 17 | 2 | 46 | *Orthopicobirnavirus* |
| ***Otarine picobirnavirus*** | | | 10 | 9 | 5 | 8 | 8 | 8 | 48 | *Unclassified Picobirnavirus* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Picobirnavirus dog/KNA/2015* | | | 17 | 11 | 4 | 9 | 6 | 12 | 59 | |
| *Picobirnavirus green monkey/KNA/2015* | | | 10 | 5 | 0 | 3 | 12 | 6 | 36 | |
| *Picobirnavirus sp.* | | | 2 | 31 | 0 | 16 | 10 | 8 | 67 | |
| *Porcine picobirnavirus* | | | 17 | 17 | 0 | 15 | 17 | 10 | 76 | |
| *Escherichia virus DE3* | Double-stranded DNA | *Siphoviridae* (53) | 0 | 0 | 0 | 0 | 0 | 22 | 22 | *Lambdavirus* |
| *Lambdavirus lvO276* | | | 0 | 0 | 0 | 31 | 0 | 0 | 31 | |
| *Aspergillus fumigatus partitivirus 2* | Double-stranded RNA | *Partitiviridae* (78) | 9 | 3 | 11 | 0 | 6 | 2 | 31 | *Gammapartitivirus* |
| *Botryotinia fuckeliana partitivirus 1* | | | 0 | 0 | 7 | 0 | 0 | 0 | 7 | *Unclassified Partitiviridae* |
| *Cryptosporidium parvum virus 1* | | | 10 | 0 | 0 | 0 | 0 | 0 | 10 | *Cryspovirus* |
| *Fusarium poae virus 1* | | | 0 | 0 | 0 | 0 | 3 | 0 | 3 | *Betapartitivirus* |
| *Penicillium aurantiogriseum partitivirus 1* | | | 0 | 0 | 3 | 0 | 0 | 0 | 3 | *Unclassified Partitiviridae* |
| *Penicillium aurantiogriseum partiti-like virus* | | | 0 | 2 | 0 | 0 | 0 | 1 | 3 | *Unclassified Partitiviridae* |
| *Penicillium stoloniferum virus F* | | | 0 | 0 | 2 | 0 | 4 | 0 | 6 | *Gammapartitivirus* |
| *Pythium nunn virus 1* | | | 0 | 0 | 10 | 0 | 0 | 0 | 10 | *Unclassified Partitiviridae* |
| *Sclerotinia sclerotiorum partitivirus S* | | | 0 | 0 | 0 | 0 | 1 | 0 | 1 | *Unclassified Partitiviridae* |
| *Ustilaginoidea virens partitivirus 2* | | | 0 | 0 | 3 | 0 | 0 | 0 | 3 | *Unclassified Partitiviridae* |
| *Verticillium dahliae partitivirus 1* | | | 0 | 0 | 1 | 0 | 0 | 0 | 1 | *Unclassified Partitiviridae* |
| *Festuca pratensis amalgavirus 1* | Double-stranded RNA | *Amalgaviridae* | 0 | 0 | 0 | 0 | 0 | 5 | 5 | *Unclassified Amalgaviridae* |
| *Tomato mosaic virus* | Single-stranded RNA | *Potyviridae* | 0 | 0 | 3 | 0 | 0 | 0 | 3 | *Potyvirus* |
| *Pepper mild mottle virus* | | *Virgaviridae* (68) | 0 | 0 | 25 | 0 | 0 | 0 | 25 | *Tobamovirus* |
| *Tobacco mild green mosaic virus* | | | 0 | 0 | 6 | 0 | 0 | 37 | 43 | |
| *Eimeria tenella RNA virus 1* | Double-stranded RNA | *Totiviridae* (52) | 22 | 0 | 0 | 0 | 0 | 24 | 46 | Unclassified *Victorivirus* |
| *Scheffersomyces segobiensis virus L* | | | 6 | 0 | 0 | 0 | 0 | 0 | 6 | *Totivirus* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Picornavirales Tottori-HG1* | **Single-stranded RNA** | **Unclassified viruses** | 0 | 18 | 0 | 0 | 8 | 7 | 33 | Unclassified *Picornavirales* |
| *Hubei orthoptera virus 1* | | | 0 | 0 | 0 | 0 | 0 | 26 | 26 | **Unclassified RNA virus ShiM-2016** |
| *Hubei picorna-like virus 24* | | | 0 | 0 | 0 | 0 | 0 | 5 | 5 | |
| *Wuhan insect virus 22* | | | 0 | 0 | 0 | 0 | 0 | 2 | 2 | |
| **Total enteric viruses** | | **Viral families** | **Total group viral contigs** | | | | | | **Total contigs** | |
| **48** | | **15** | 886 | 835 | 208 | 658 | 801 | 946 | 4334 | |

The brown highlight indicates the age group with the highest viral contigs abundance for each virus presented in rows, while the ash colour tone means that a specific virus(es) is entirely absent in that group(s). In addition, values enclosed in brackets for viral families equates the total number of contigs for each viral family.

Based on the data shown in Table 4.5.1, a contrasting effect of total contigs was observed for the 7-week chickens. In the six sample groups in Table 4.5.1, the viral contigs for the 7 weeks of winter samples were the highest, while the viral contigs for the 7 weeks of summer samples were the lowest. Hence, the effect of overall viral contigs categorized by age and season of the investigated chickens were further described.

There was no distinct trend in the dynamics of total viral contigs across various collated sample time points in the six categories analyzed, as shown in Table 4.5.1. A look at the week two summer (2S) and winter (2W) groups showed that they had total viral contigs of 886 and 658, respectively. Hence, there seems to be an observed decrease in viral contigs at the same collection age (2 weeks) between these seasons at week 2. The observed pattern was way different for the 4 weeks chickens which had comparable ranges of total viral contigs, which were 835 and 801, respectively. Another contrasting phenomenon exhibited was in the 7-week samples, showing an opposite effect. It is noteworthy that the 7W group exhibited a significantly higher number of total viral contigs (946) in contrast to the 7S group (208), with a nearly 4-fold increase. Based on the two independent sample collection seasons, the findings indicate that during the summer season, there was a negligible reduction of viral contigs in the 2S and 4S categories. However, a drastic decrease of viral contigs was observed in the 4S and 7S chicken groups that were examined. The winter samples exhibited a consistent rise in the number of viral contigs from the 2W group through 4W and 7W. Specifically, the total viral contigs for the winter groups were 658, 801, and 946, respectively (Table 4.5.1).

In addition, it is notable that certain viral families and/or species within the examined chicken samples exhibited unique trends with respect to age and seasonal variations. For instance, there was a decrease in viral contigs with increasing age noted in *Rotavirus G, Avian orthoreovirus*, and *Bavaria virus*. Interestingly, some viral species have been observed to exhibit a seasonal distribution, whereby they are detected in summer samples but are entirely absent in winter samples, or conversely. It was observed that *Rotavirus F, genus Lamdavirus, Infectious bursal disease virus, Festuca pratensis almagavirus 1, Fusarium poae virus 1,* and unclassified RNA virus ShiM-2016 were not detected in the summer samples but were found present in the winter samples (Table 4.5.1). Similarly, the *Tomato mosaic virus, Pepper mild mottle* virus, and some partitiviruses (*Botryotinia fuckeliana partitivirus 1, Cryptosporidium parvum* virus 1*, Pythium nunn* virus 1*, Ustilaginoidea virens partitivirus 2,* and *Verticillium dahliae partitivirus 1*) exhibit a seasonal pattern, being present exclusively in the summer samples (Table 4.5.1).

Overall, it could be concluded that about 20 RNA virus species were found in more than ten different chicken samples. as shown in Appendix 8.5. However, 24 distinct viruses were found in four or fewer samples, with 14 of the 24 detected in just one sample (Appendix 8.5).

### 4.5.1 Host range of identified viruses from chicken faeces

The present study classified the identified viruses based on their established host specificity as illustrated in Figure 4.5.2. It was observed that the viruses profiled exhibited a remarkably broad host range with the avian viruses having the highest prevalence of 48% (23), surpassing viruses linked to other hosts (Figure 4.5.2). Also, the results showed that fungal viruses constituted 15% (7) of the total virus host obtained and were primarily made up of unclassified partitiviruses. Additionally, occurrence of viruses derived from mammalian and plant hosts was found to be equivalent, with each accounting for 12.5% of identified virus host distribution (Figure 4.5.2). It is noteworthy that the 6 mammalian viruses belonged to family *Picobirnaviridae* (Table 4.5.1). On the contrary, the plant viruses emerged from various viral families, namely *Virgaviridae, Potyviridae, Amalgaviridae,* and *Partitiviridae*. Furthermore, it is noteworthy that phages, constituting 4% of the host-virus population, were found to be the least abundant. The phages from the *Lamdavirus* genus were represented by only two viruses, whereas the insect viruses accounted for 8% of the population (Figure 4.5.2), with 3 *Unclassified RNA virus ShiM-2016* viruses being identified (Table 4.5.1).

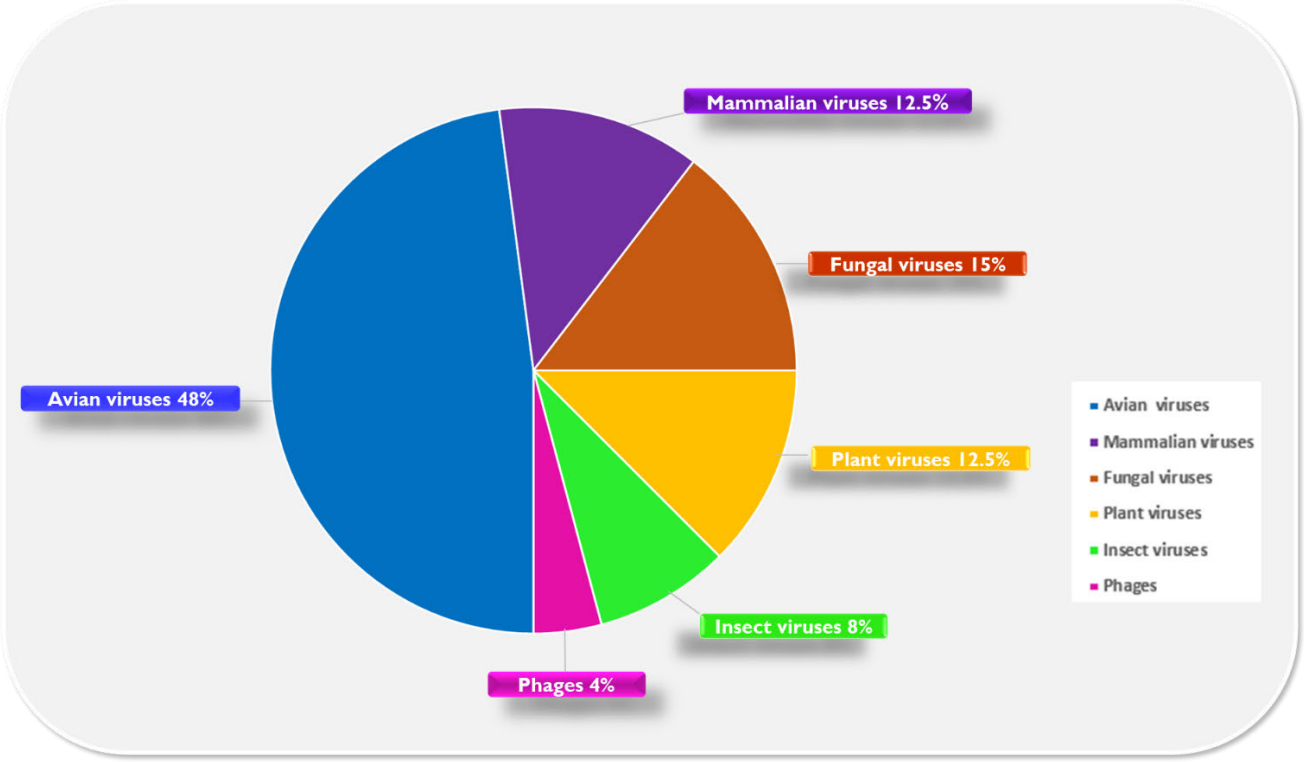



Figure 4.5.2 Host distribution of the identified viruses from chicken faeces

## 4.6 Diversity of RNA viruses identified from chicken faeces

The results of the viral occurrence at the family level revealed about 15 viral families including those with unclassified Picornavirales and patitiviruses (Figure 4.6.1). The family *Picornaviridae* was 100% prevalent across the 27 samples. This was closely followed by the families *Reoviridae*, *Astroviridae*, and *Picobirnaviridae*, occurring in 24 (88.9%) and 22 (81.5%) chicken samples respectively. In addition, the family, *Retroviridae* was observed to have averagely occurred in across sample with 66.7% occurrence while occurring in 18 out of 27 chicken faecal samples. However, families *Birnaviridae* and *Siphoviridae* had the lowest incidence, occurring in just one sample each, in comparison with all other families identified.

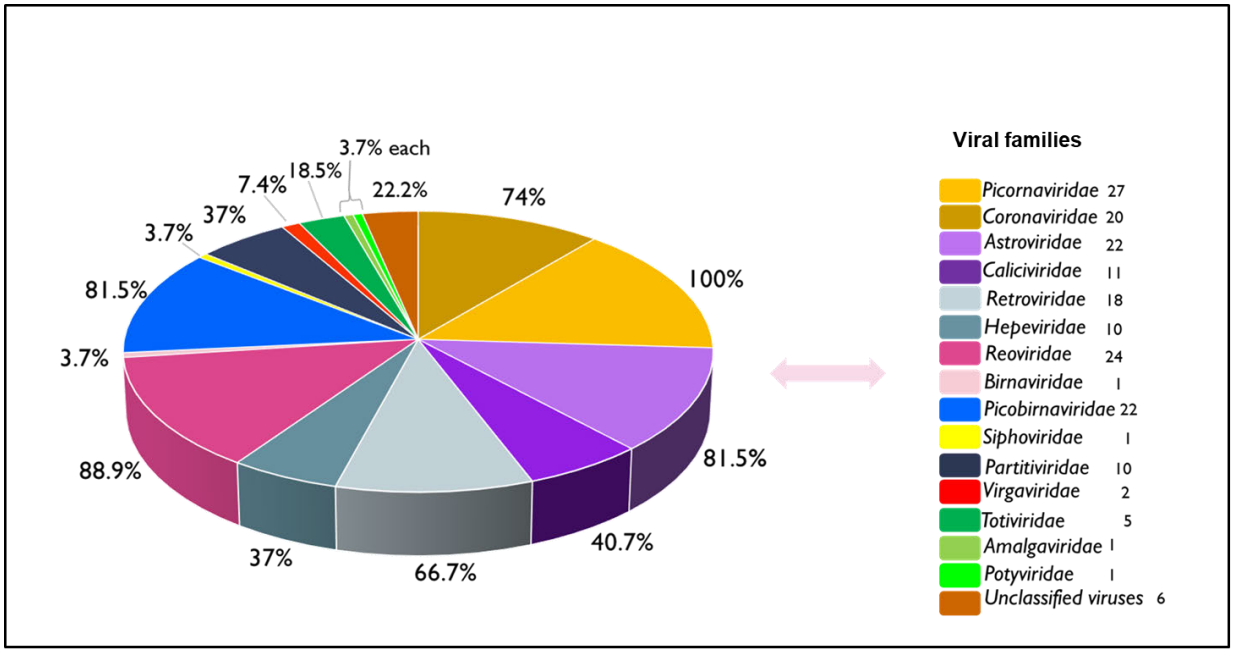



Figure 4.6.1 Occurrence rate of viral families across 27 chicken samples

In addition, the recovery rate of the individual viral genera obtained from chicken virome analysis showed that the viruses recovered were distributed among 21 different genera, excluding the unclassified viruses (Figure 4.6.2). The two most abundant genera are *Sicinivirus* and *Gallivirus*, all from the *Picornaviridae*. About eight other genera occurred in more than 50% of the entire samples. Nevertheless, in addition to some unclassified viruses mainly from *Partitiviridae* and RNA virus shiM-2016, five genera had very low occurrence, namely *Betapartitivirus, Avibirnavirus, Potyvirus, Cryspovirus* and *Totivirus* occurring in less than two samples each (Table 4.5.1). A close examination at the viral species level showed that viruses belonging to the genera with low occurrence (*Botryotinia fuckeliana partitivirus 1,Cryptosporidium parvum virus 1, Fusarium poae virus 1, Penicillium aurantiogriseum partitivirus 1, Pythium nunn virus 1,Sclerotinia sclerotiorum partitivirus S, Ustilaginoidea virens partitivirus 2,Verticillium dahliae partitivirus 1, Scheffersomyces segobiensis virus L*) are mainly, fungal and plant viruses, with exception to the infectious bursal disease virus with avian host tropism, occurring in only one sample (Figure 4.6.3). Viruses with higher incidence at the genus level have their specific viruses such *as Sicinivirus A, Gallivirus A, Rotavirus G* and *Chicken astrovirus* also represented in Figure 4.6.3.

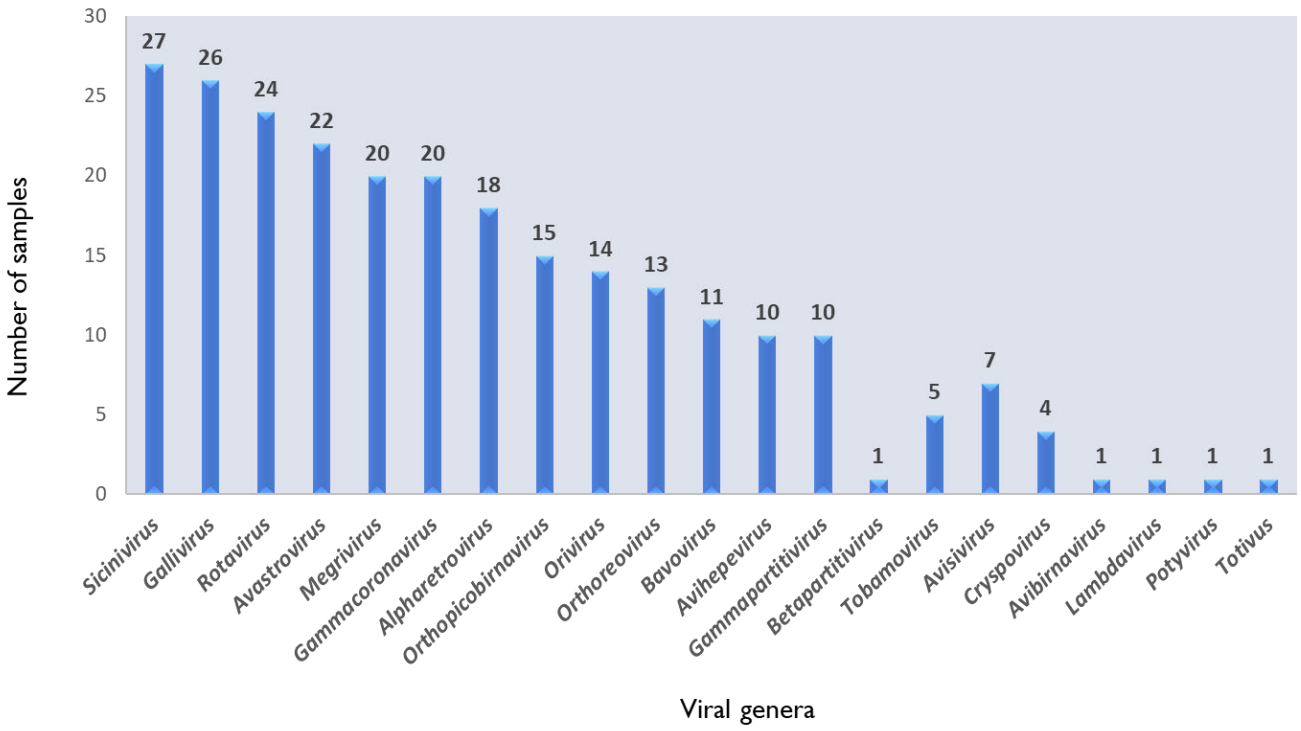



Figure 4.6.2 Recovery rate of identified viruses across chicken faecal samples at the genus level.

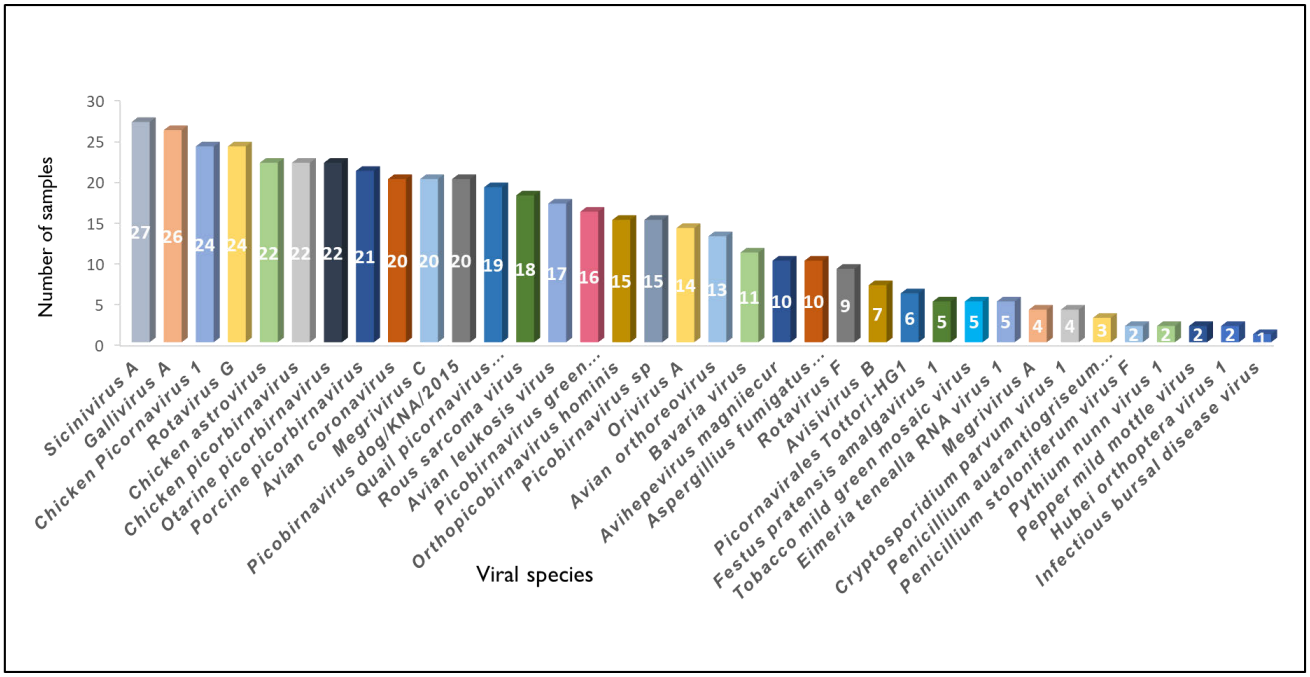



Figure 4.6.3 Occurrence rate of the individual viruses across the 27 chicken faecal samples at the species level

### 4.6.1 Avian RNA viruses

A total of 3129015 viral reads were assigned to avian RNA viruses from the 27 investigated chicken faecal samples as shown in Table 4.6.1. The 11 avian RNA viral families identified include *Picornaviridae, Coronaviridae, Reoviridae, Caliciviridae, Astroviridae, Retroviridae, Hepeviridae, Birnaviridae*, *Picobirnaviridae, Totiviridae* and *Partitiviridae.* Among these avian viral families, *Picornaviridae, Reoviridae* and *Coronaviridae* were found to be most abundant across samples accounting for 1553143, 632918 and 167789 reads, respectively (Table 4.6.1). These avian RNA viruses included both double-stranded and single-stranded RNA viruses as described in Table 4.6.1.

Table 4.6.1 Taxonomic classification of avian RNA viruses identified from chicken faecal samples

| **No of avian viral species** | **Genome type** | | **Total viral contigs** | **Total viral reads** | **Sum of identified avian viral genera** |
|---|---|---|---|---|---|
| | | | | | **15 genera + unclassified viruses** |
| **Avian RNA viral species** | | **Viral families** | | | ***Genera*** |
| ***Avian coronavirus*** | **Single-stranded RNA** | ***Coronaviridae*** | 646 | 167789 | *Gammacoronavirus* |
| ***Avisivirus B*** | | ***Picornaviridae* (1553143 reads)** | 65 | 2584 | *Avisivirus* |
| ***Chicken picornavirus 1*** | | | 252 | 66122 | *Unclassified Picornaviridae* |
| ***Gallivirus A*** | | | 302 | 100247 | *Gallivirus* |
| ***Megrivirus A*** | | | 26 | 2303 | *Megrivirus* |
| ***Megrivirus C*** | | | 245 | 1121717 | |
| ***Orivirus A*** | | | 97 | 22023 | *Orivirus* |
| ***Quail picornavirus QPV1/HUN/2010*** | | | 157 | 43107 | *Unclassified Picornaviridae* |
| ***Sicinivirus A*** | | | 341 | 195040 | *Sicinivirus* |
| ***Chicken astrovirus*** | | ***Astroviridae*** | 173 | 666244 | *Avastrovirus* |
| ***Bavaria virus*** | | ***Caliciviridae*** | 131 | 12759 | *Bavovirus* |
| ***Avian leukosis virus*** | | ***Retroviridae*** | 121 | 4478 | *Alpharetrovirus* |
| ***Rous sarcoma virus*** | | | 183 | 2201 | |
| ***Avihepevirus magniiecur*** | | ***Hepeviridae*** | 107 | 12126 | *Avihepevirus* |
| ***Avian orthoreovirus*** | **Double-stranded RNA** | ***Reoviridae* (632918)** | 286 | 14446 | *Orthoreovirus* |
| ***Rotavirus F*** | | | 124 | 205450 | *Rotavirus* |
| ***Rotavirus G*** | | | 352 | 413022 | |

| | | | | | |
|---|---|---|---|---|---|
| ***Infectious bursal disease virus*** | | ***Birnaviridae*** | 4 | 8 | *Avibirnavirus* |
| ***Chicken picobirnavirus*** | | ***Picobirnaviridae*** | 59 | 20631 | Unclassified *Picobirnavirus* |
| ***Scheffersomyces segobiensis virus L*** | | ***Totiviridae*** | 6 | 9197 | *Totivirus* |
| ***Aspergillus fumigatus partitivirus 2*** | | ***Partitiviridae*** | 31 | 47521 | *Gammapartitivirus* |
| **Total avian viruses** | | **Avian viral families** | | | |
| **21** | | **11** | | 3129015 | **15 genera + unclassified viruses** |

### 4.6.1.1 Novel avian RNA viruses

The present study identified 173 viral contigs associated with the genus *Avastrovirus*, particularly the species, chicken astrovirus (CAstV) as shown in Table 4.61. While this virus occurred in 23 of the samples with genome coverage >90% in 10 samples, it was entirely absent in the 7 weeks summer samples (Table 4.5.1). Two novel CAstV each with nearly complete genome coverage of 99.9%, 99.9%, 99.6%, 98.8% and 99.0% were recovered from the 7W2, 7W3, 2W4, 7W1, and 7W4 samples, respectively (Table 4.6.1.1). The first novel CAstV from 2W4 and 7W2 had a mere 58.3%, 57.8% nucleotide identity (nt) and 47.7 and 47.4% amino acids (aa) identity, respectively with all identified CAstVs. The second novel CAstV obtained from this study was from 7W3, 7W1 and 7W4 with about 78.8%, 63.2%, 77.6% nt identity and 82.5%, 59.1%, 81.6%, aa identity, respectively (Table 4.6.1.1). Figure 4.6.1.1 depicts the coverage map result of near whole genome of CAstV obtained.

Table 4.6.1.1 Novel chicken astroviruses identified from chicken faecal samples

| **Viruses** | **Sample** | **Genome coverage (%)** | **Nucleotide identity (%)** | **Amino-acid identity (%)** |
|---|---|---|---|---|
| *Chicken astrovirus* | 7W2 | 99.9 | 58.3 | 47.8 |
| | 2W4 | 99.6 | 57.8 | 47.4 |
| *Avian nephritis virus* | 7W1 | 98.8 | 63.2 | 59.1 |
| | 7W3 | 99.9 | 78.8 | 82.5 |

| | | | | |
|---|---|---|---|---|
| | 7W4 | 99.0 | 77.6 | 81.6 |

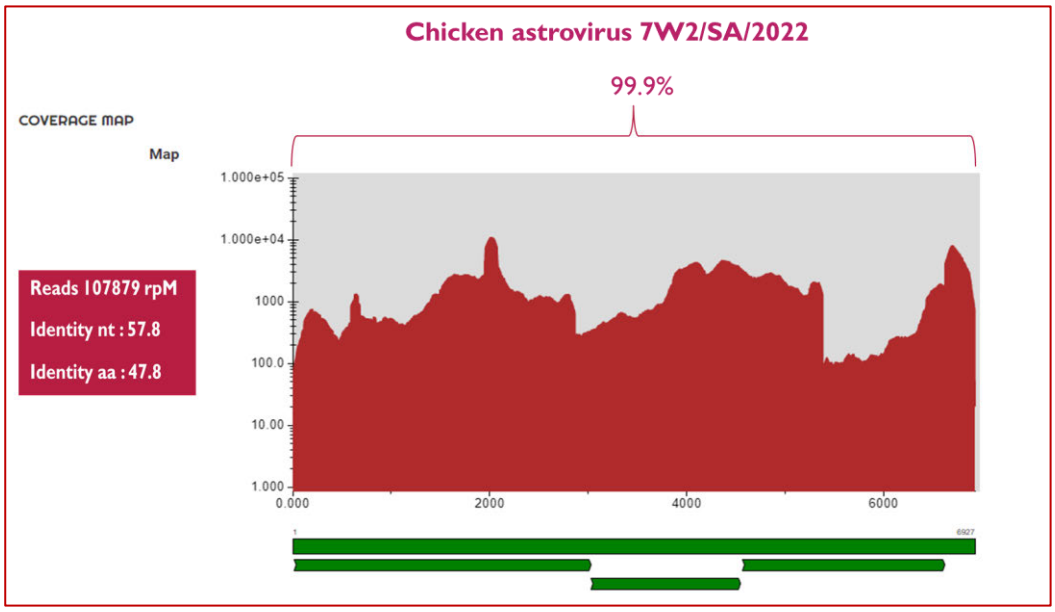



Figure 4.6.1.1 Whole genome coverage map of the novel chicken astrovirus genome identified from the 7W2 sample

### 4.6.1.2 Double-stranded avian RNA viruses

In this study, three dsRNA avian viral families (*Reoviridae*, *Birnaviridae* and *Picobirnaviridae*) comprising five viral species (*Avian orthoreovirus, Rotavirus F, Rotavirus G, Infectious bursal disease virus, Chicken picobirnavirus, Scheffersomyces segobiensis virus L,* and *Aspergillus fumigatus partitivirus 2*) were obtained from the chicken enteric virome (Table 4.6.1). These avian viral species generated a total of 862 dsRNA viral contigs (Table 4.6.1) and the results obtained from individual dsRNA viruses are described below.

#### 4.6.1.2.1 *Reoviridae*

The three reoviral species obtained in this study are from two genera, *Orthoreovirus* and *Rotavirus*. The *Orthoreovirus* in this study was identified from both summer and winter samples totalling to 286 contigs (Table 4.6.1). The partial and nearly complete coding gene sequences of Segment L1 (95%), L2 (98.7%), L3 (99.0%) and M2 (87.5%) of *Avian orthoreovirus* were obtained from sample 2W1. The viral contigs obtained from this virus was observed to decrease with increased age (Table 4.5.1). *Rotavirus F* and *Rotavirus G* are the two rotaviral species obtained with 124 and 352 contigs,

respectively. The nucleotide and protein-coding sequences of all eleven segments of Rotavirus F (98-100%) and Rotavirus G (96.8-100%) were obtained from 2W1, 2W2, and 2W2 for the former, and 2W3 and 7W5 for the latter. Overall, members of this viral family demonstrated an inverse correlation with age.

#### 4.6.1.2.2 *Birnaviridae*

In this study, *Infectious bursal disease virus or Gumboro disease virus* with four viral contigs from only one sample was identified. This sample was a 7-week winter sample (Table 4.5.1).

#### 4.6.1.2.3 *Picobirnaviridae*

Chicken picobirnaviruses in addition to other mammalian types, still under unclassified isolates by the ICTV nomenclature were obtained from this study. *Chicken picorbirnavirus* had a total of 59 contigs from nearly all the 27 sample age groups and season except 5 samples (7S2, 2W4, 2W5, 7W4, and 7W5). However, only the segment RNA 2 which is the polymerase gene were obtained across samples with 99.9%, 99.6%, and 97.8% gene segment coverage from samples 2S2, 4S5 and 4S1, respectively (Table 4.5.1).

### 4.6.1.3 Single-stranded avian RNA viruses

Within the scope of the present study, a total of 14 viral species, belonging to six distinct viral families and 10 viral genera excluding unclassified ssRNA viruses were obtained from the chicken enteric virome (Table 4.6.1). These families including *Coronaviridae, Picornaviridae, Caliciviridae, Astroviridae, Retroviridae*, and *Hepeviridae,* have yielded about 50% of the entire contigs across all families including non-avian viral contigs. The total contigs accrued from this ssRNA is about 2846 out of an overall of 4334 viral contigs (Table 4.6.1). In addition to the ten viral genera (*Gammacoronavirus, Avisivirus, Gallivirus, Megrivirus, Orivirus, Sicinivirus, Avastrovirus, Bavovirus, Alpharetrovirus, Avihepevirus*), there were two unclassified viruses namely, *Chicken picornavirus 1* and *Quail picornavirus QPV1/HUN/2010* as shown in Table 4.6.1. The specific results derived from the ssRNA viral families in this study are outlined as follows.

#### 4.6.1.3.1 *Coronaviridae*

In this study, an avian coronavirus or Infectious bronchitis virus, a member of *Gammacoronavirus* genera in subgenus *Igacovirus* and subfamily *Orthocoronavirinae* was identified from this study. This viral species had about 646 contigs recovered from this study, across the 3 ages groups and two seasons. This virus demonstrated a 74% prevalence occurring in the about 20 samples out of 27

samples (Appendix 8.5). The *Avian coronavirus* obtained from samples 2S4 and 2S5 had a nearly complete genome with 99.3% and 99.0% nucleotide sequence coverage, respectively. Notably, complete coding sequence (CDS) and 100% sequence coverage of the OR1ab polyprotein, NSP1 to NSP16 proteins, 3a, 3b, 5a, and 5b proteins including the spike, envelope, membrane and hypothetical proteins were recovered from sample 2S4.

#### 4.6.1.3.2 *Picornaviridae*

The finding from this study demonstrated the family Picornaviridae to be most abundant family with a total of 1485 viral contigs from eight species in five *genera* (Table 4.6.1). These viral genera include *Megrivirus, Gallivirus, Sicinivirus, Avisivirus* and *Orivirus*, in addition unclassified viruses as shown in Table 4.6.1. The species identified in this study includes, *Megrivirus A and C, Gallivirus A, Sicinivirus A, Avisivirus B, Orivirus A, Chicken picornavirus 1 and Quail picornavirus QPV1/HUN/2010.* The prevalence of individual species of the picornaviruses identified in this study is shown in Table 4.6.1.3

Table 4.6.1.3 Prevalence of picornavirus *genera* in chicken faeces

| Family | Picornavirus species identified | Number of incidences (N= 27) | Percentage occurrence (%) |
|---|---|---|---|
| ***Picornaviridae*** | *Sicinivirus A* | 27 | 100.0 |
| | *Gallivirus A* | 26 | 96.3 |
| | *Megrivirus A* | 4 | 14.8 |
| | *Megrivirus C* | 20 | 74.0 |
| | *Avisivirus B* | 7 | 25.9 |
| | *Orivirus A* | 14 | 51.9 |
| | *Chicken picornavirus 1* | 24 | 88.9 |
| | *Quail picornavirus QPV1/HUN/2010* | 19 | 70.4 |

The species *Sicinivirus A* had a 100% prevalence (Table 4.6.1.3) across the 27 samples with a total contigs of 341 (Table 4.6.1). Partial nucleotide sequence (96.0) of *Sicinivirus A* was obtained from sample 7W2 with 144420.35 reads per million (rpm). In addition to the nucleotide sequence the nearly complete DF53_gp1 CDS (99.5%) and complete protein sequence of the RdRp gene crucial for its replication known as $3D^{pol}$, was recovered alongside the complete sequence of the VP0 to VP3, 2A, 2B, 2C, 3A, 3B and 3C proteins. *Megrivirus A* and *Megrivirus C*, all from the genus *Megrivus* was obtained, with the latter being more prevalent across samples than the former (Table 4.6.1.3), accounting for 26 and 245 contigs, respectively (Table 4.6.1). *Megrivirus A* had a partial nt sequence from 4W1 and the results of BLAST alignment at e-value $10^{-3}$ showed its best hits with a *Melegrivirus A* virus conserved in Turkey from Hungary (KF961188.1) and a reference sequence of Turkey hepatitis virus 2993D at 77.1 and 78.3% identity respectively. However, the *Megrivirus C* identified in this study had a 93.1 nt sequence coverage, with 1121717 reads as shown in Table 4.6.1. In addition, the *Gallivirus A* species obtained had a 96.3% prevalence across samples (Table 4.6.1.3), with 5303 reads (Table 4.6.1) and a 94.4% nt sequence coverage from sample 2S3. The complete (100%) protein sequences of 3A, 3B,3c and 3D (RdRp) genes of this *Gallivirus A* was recovered. Furthermore, *Avisivirus B* viral species occurred in 7 samples only at lower coverages (<80). Nevertheless, the complete 3D RdRp was recovered from 2W2 sample. The alignment results showed its best hit of 85.8% identity with *Chicken picornavirus 2 isolate 44C* from Hong Kong (KF979333.1). Similarly, Orivirus A showed a 51.9% incidence across samples (Table 4.6.1.3) with the highest nt sequence coverage of 96.1% from 2W2 sample. The complete PI35_gp1 CDS and polyprotein genes of *Orivirus A* was recovered from this sample. Additionally, two unclassified avian picornaviruses isolated from this study are *Chicken picornavirus 1* and *Quail picornavirus QPV1/HUN/2010* with 88.9 and 70.4% prevalence across samples, respectively (Table 4.6.1.3). *Chicken picornavirus 1* had very low coverage (59.8%), while *Quail picornavirus QPV1/HUN/2010* was observed to have its highest nt sequence coverage of 94.2% from 4W5 sample. The complete QPV_gp1 CDS and polyprotein of *Quail picornavirus QPV1/HUN/2010* were recovered, with results demonstrating maximum hit with a *Phacovirus sp* from Brazilian chickens.

#### 4.6.1.3.3 *Caliciviridae*

The calicivirus identified from this avian study belongs to the only species in the *Bavovirus genera.* This Bavarian virus identified had about 131 contigs and a total of 12759 reads (Table 4.6.1), with its highest nt sequence coverages of 99.3 and 98.5% emerging from 2S4 and 2S5 samples. The complete CDS, polyprotein and VP2 gene segments of this *Bavaria virus* were retrieved. Further analysis using

the polyprotein revealed this virus was most similar with 85% nt and 96.1% aa identity to a calicivirus identified from American chickens (KY120883.1). It was thus seen from this study that the *Bavaria virus* often occurred within samples from younger chickens after which they gradually disappeared through 4 weeks, becoming non-existent at 7 weeks.

#### 4.6.1.3.4 *Astroviridae*

In this study the chicken astroviruses (CAstVs) obtained had a total of 173 viral contigs (Table 4.6.1). This CAstVs occurred in 23 of the samples with genome coverage >90% in 10 samples, it was entirely absent in the 7 weeks summer samples (Appendix 8.5). The identified viral CAstV species belong to the Genus *Avastrovirus* as shown in Table 4.6.1. Two different CAstV each with nearly complete genome coverage of 99.9%, 99.9%, 99.6%, 98.8% and 99.0% were recovered from the 7W2, 7W3, 2W4, 7W1, and 7W4 samples, respectively (Table 4.6.1.1). These viruses were believed to be novel viuses as their results showed only 58.3%, 57.8% nucleotide identity (nt) and 47.7 and 47.4% amino acids (aa) identity with all known CAstV for first, while the second only shared 78.8%, 63.2%, 77.6% nt identity and 82.5%, 59.1%, 81.6%, aa identity from samples 2W4 and 7W2 had, respectively (Table 4.6.1.1).

#### 4.6.1.3.5 *Retroviridae*

Among the 8 species designated in the genus *Alpharetrovirus* by the ICTV, only the *Avian leukosis virus* and *Rous Sarcoma virus* were obtained in this study. These retroviruses accrued a total of 121 and 183 contigs for *Avian leukosis virus* and *Rous sarcoma virus*, respectively (Table 4.6.1). Although these viruses were identified in very low genome coverage between 36.3-63.1%, the two viruses appeared to have high hit up to 99.6% with the same clones *Avian leukosis virus subgroup E* from the UK (MT263508.1) and USA (MF817820.1).

#### 4.6.1.3.6 *Hepeviridae*

In this study, *Avian hepatis E virus* or its new name *Avihepevirus magniiecur* was identified in 10 samples, mainly from the winter season in this study yielding about 107 contigs and 12126 reads (Table 4.6.1). A nearly complete genome of 98.8% was obtained from 7W4. The complete CDS (BA77_gp1, gp2 and gp3) and non-structural polyprotein were retrieved. Additional BLAST analyses with its recovered complete genome showed that it was mostly related to a novel *Avihepevirus magniiecur* virus from chickens and pheasants in France (ON922634.1) with about 83.7% identity.

### 4.6.2 Non-avian viruses

The non-avian RNA viruses categorized in the current study are those whose viral tropism has been established in hosts other than avians. The results of these non-avian viral hosts identified in this study encompass a range of organisms, including mammals, plants, fungi, arthropods and bacteria, among others. In this study, 27 non-avian viral species were identified with majority of them occurring in very low abundances (Table 4.6.2.1). These non-avian viruses spanned across seven viral families namely *Picobirnaviridae, Amalgaviridae, Potyviridae, Virgaviridae, Siphoviridae, Partitiviridae, Totiviridae*, in addition to an unclassified Picornavirales as categorized below in Table 4.6.2.1.

Table 4.6.2.1 Non-avian viruses identified from chicken faecal samples

| No of avian viral species | Genome type | Host | Viral families | Sum of contigs | Sum of identified viral genera |
|---|---|---|---|---|---|
| | | | | 4334 | |
| **Non-avian RNA virus species** | | | | | ***Genera*** |
| ***Orthopicobirnavirus hominis*** | ***Double-stranded RNA*** | **Mammalian viruses** | ***Picobirnaviridae* (332)** | 46 | *Orthopicobirnavirus* |
| ***Otarine picobirnavirus*** | | | | 48 | *Unclassified Picobirnavirus* |
| ***Picobirnavirus dog/KNA/2015*** | | | | 59 | |
| ***Picobirnavirus green monkey/KNA/2015*** | | | | 36 | |
| ***Picobirnavirus sp.*** | | | | 67 | |
| ***Porcine picobirnavirus*** | | | | 76 | |
| ***Picornavirales Tottori-HG1*** | **Single-stranded RNA** | | **Unclassified virus** | 33 | Unclassified *Picornavirales* |
| ***Escherichia virus DE3*** | ***Double-stranded DNA*** | **Phages** | ***Siphoviridae*** | 22 | *Lambdavirus* |
| ***Lambdavirus lvO276*** | | | | 31 | |
| ***Botryotinia fuckeliana partitivirus 1*** | | **Fungal viruses** | | 7 | *Unclassified Partitiviridae* |
| ***Cryptosporidium parvum virus 1*** | | | | 10 | *Cryspovirus* |
| ***Fusarium poae virus 1*** | | | | 3 | *Betapartitivirus* |

| | | | | | |
|---|---|---|---|---|---|
| ***Penicillium aurantiogriseum partitivirus 1*** | | | | 3 | *Unclassified Partitiviridae* |
| ***Penicillium aurantiogriseum partiti-like virus*** | | **Fungal viruses** | | 3 | *Unclassified Partitiviridae* |
| ***Penicillium stoloniferum virus F*** | | | | 6 | *Gammapartitivirus* |
| ***Pythium nunn virus 1*** | | | | 10 | *Unclassified Partitiviridae* |
| ***Sclerotinia sclerotiorum partitivirus S*** | | | | 1 | *Unclassified Partitiviridae* |
| ***Ustilaginoidea virens partitivirus 2*** | | | | 3 | *Unclassified Partitiviridae* |
| ***Verticillium dahliae partitivirus 1*** | | | | 1 | *Unclassified Partitiviridae* |
| ***Festuca pratensis amalgavirus 1*** | **Double-stranded RNA** | **Diet associated viruses** | ***Amalgaviridae*** | 5 | *Unclassified Amalgaviridae* |
| ***Tomato mosaic virus*** | **Single-stranded RNA** | | ***Potyviridae*** | 3 | *Potyvirus* |
| ***Pepper mild mottle virus*** | | | ***Virgaviridae*** | 25 | *Tobamovirus* |
| ***Tobacco mild green mosaic virus*** | | | | 43 | |
| ***Eimeria tenella RNA virus 1*** | **Double-stranded RNA** | | ***Totiviridae*** | 46 | Unclassified *Victorivirus* |
| ***Hubei orthoptera virus 1*** | | | **Unclassified viruses** | 26 | **Unclassified RNA virus ShiM-2016** |
| ***Hubei picorna-like virus 24*** | | | | 5 | |
| ***Wuhan insect virus 22*** | | | | 2 | |
| **Total enteric viruses** | | | **Viral families** | **Total contigs** | |
| **27** | | | **7** | **4334** | **8** |

### 4.6.2.1 Mammalian RNA viruses

The mammalian viruses obtained from the investigated chicken samples includes an unclassified Picornavirales Tottori-HG1 and non-avian unclassified *Picobirnaviridae* members (Table 4.6.2.1). The *Picornavirales Tottori-HG1 virus*, a swine virus had 33 contigs (Table 4.6.2.1) with genome coverage ranging within 38-45%. In contrast, non-avian members of *Picobirnaviridae*;

*Orthopicobirnavirus hominis* from humans, *Otarine picobirnavirus* segment 2 from California sea lion, *Picobirnavirus dog/KNA/2015* from dog, *Picobirnavirus green monkey/KNA/2015* from African green monkey, *Picobirnavirus species,* and *Porcine picobirnavirus* from pigs occurred in moderate abundance across samples with high coverage ranging from 60-85%. These non-avian picorbirnaviruses accrued an overall 332 contigs (Table 4.6.2.1).

#### 4.6.2.2 Diet-associated RNA virus

The food-related viruses obtained were seen to come from three viral families, excluding those grouped under unclassified RNA virus ShiM-2016 by the ICTV. The three families occurred in very read abundances and includes *Amalgaviridae, Potyviridae, Virgaviridae* (Table 4.6.2.1). The viral species includes *Festuca pratensis amalgavirus 1* from *Amalgaviridae, Tomato mosaic virus* from *Potyviridae, Pepper mild mottle virus* and *Tobacco mild green mosaic virus* from *Virgaviridae.* In addition, the species in the *unclassified RNA virus ShiM-2016* are *Hubei orthoptera virus 1*, *Hubei picorna-like virus 24* and *Wuhan insect virus 22* (Table 4.6.2.1).

#### 4.6.2.3 Tailed DNA phages

The dsDNA phages identified from this study are from one viral family, the Siphoviridae. The two tailed phage viruses are *Escherichia virus DE3* and *Lambdavirus lvO276*, both from *Lambdavirus genera* (Table 4.6.2.1). These phages had moderately high read abundances (1766) accumulating a total of 53 contigs as shown in Table 4.6.2.1. However, these phages, *Escherichia virus DE3* and *Lambdavirus lvO276* occurred only in 7W3 and 2W4 winter samples, respectively (Appendix 8.5).

#### 4.6.2.4 Fungal viruses

In this investigation, 10 distinct fungal viruses were identified from two viral families, *Partitiviridae* and *Totiviridae* (Table 4.6.2.1). *Totiviridae* was the smaller group with just one virus, *Eimeria tenella RNA virus 1*. On the other hand, the larger group *Partitiviridae* had 2 classified viruses under *Gammapartitivirus* and *Betapartitivirus* genera. The remaining 8 viruses were unclassified partitiviruses according to the ICTV classification as of June 2023. The Gammapartitiviruses include *Penicillium stoloniferum virus F* and Fusarium poae virus 1 (Table 4.6.2.1). In addition, the unclassified partitiviruses includes *Botryotinia fuckeliana partitivirus 1, Cryptosporidium parvum virus 1, Penicillium aurantiogriseum partitivirus 1, Penicillium aurantiogriseum partiti-like virus, Pythium nunn virus 1, Sclerotinia sclerotiorum partitivirus S, Ustilaginoidea virens partitivirus 2, and Verticillium dahliae partitivirus 1*. Overall, these fungal viruses occurred in low contig

abundance and in less than three samples, except *Aspergillus fumigatus partitivirus 2* that was found in 9 samples with about 31 contigs that includes both seasons (Appendix 8.4).

## 4.7 Variations in avian RNA viral abundance and diversity

Analysis of gut RNA virome structure of the faecal samples of chickens in this study revealed differences in diversity and abundances of viral species across the 27 chicken libraries investigated. A high proportion of the relative standardised abundance estimates of sequence reads across all the libraries were classed as RNA viruses and the results of their relative abundance from individual sample and group are presented in Figure 4.7.1. The results showed that the most relatively abundant viral species reads across age groups are *Avian coronavirus, Rotavirus G. Chicken astrovirus, Megrivirus C* and *Sicinivirus A* (Figure 4.7.1). In addition, *Bavaria virus, Chicken picobirnavirus Chicken Picornavirus 1* and *Aspergillius fumigatus partitivirus 2* were moderately represented. However, the remaining viral species had very low abundance in individual samples, including those assigned in the other group as shown in Figure 4.7.1. It was interesting to see the changing dynamics of members of the *Reoviridae,* particularly, *Rotavirus G* and *Avian orthoreovirus,* which first increased then drastically decreased with increasing age (Figure 4.7.1). It was also revealed that the winter sample were mainly characterized by *Chicken astrovirus* and *Rotavirus G*. Overall, avian viruses appear to be well represented in all chicken faecal samples.

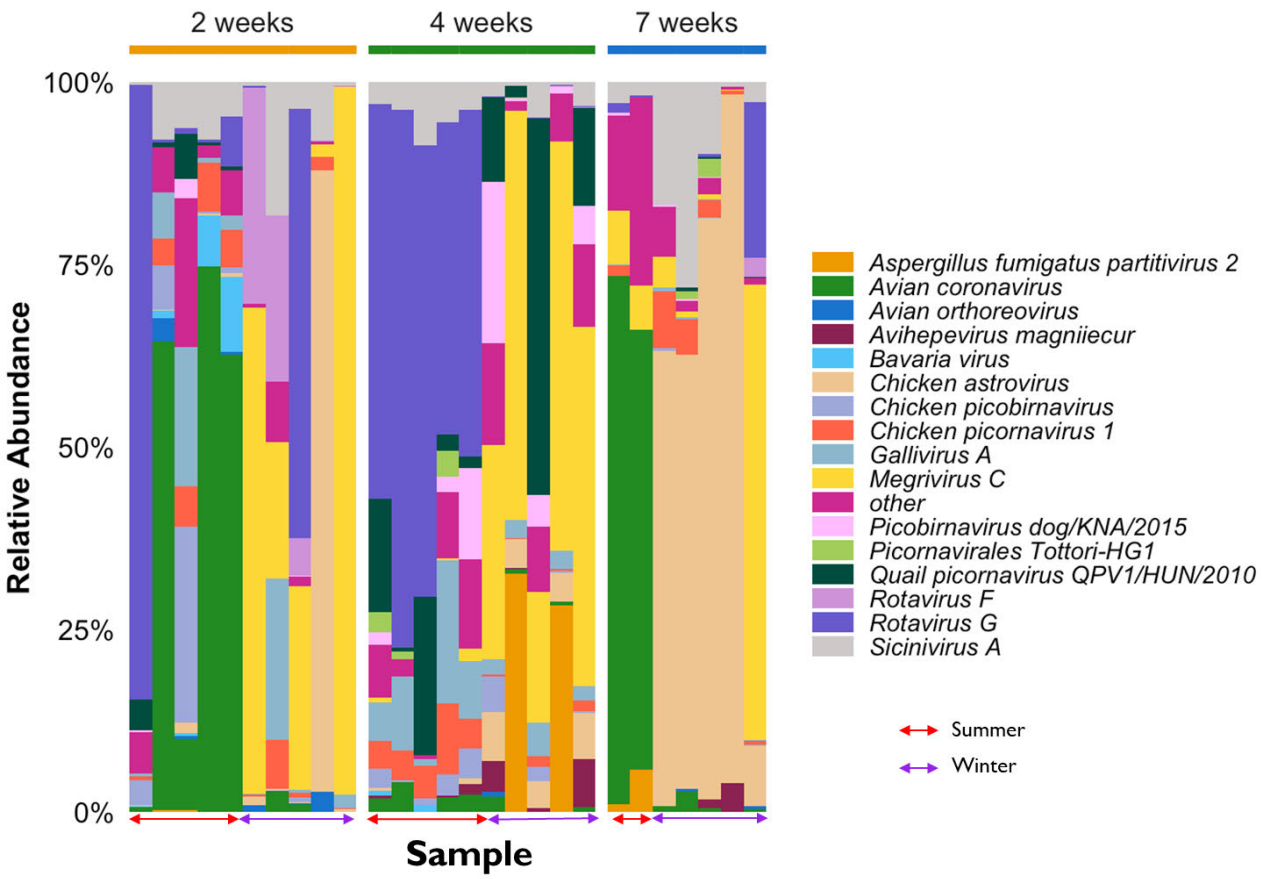



Figure 4.7.1 The relative abundance profile of major RNA viral species across the chicken faecal samples at different ages (2, 4 and 7 weeks) and seasons (Summer and winter)

### 4.7.1 Determining the effect of chickens age and collection season on virome diversity

#### 4.7.1.1 Alpha diversity

The results of the observed alpha diversity principally based on expected count for age and season are presented in Figure 4.7.1.1.1. For age groups 2, 4 and 7 weeks, the total counts (observed) within ages are very similar with interquartile range positively skewed, with most values within the lower quartile (Figure 4.7.1.1.1A). For the two seasons, summer and winter, a slight difference was observed at 2 weeks in the within sample counts abundance with the summer sample a bit higher than the winter sample (Figure 4.7.1.1.1B). This is not the case with summer and winter samples at week 4, as they demonstrated to have similar viral counts (Figure 4.7.1.1.1B). However, a notable shift was observed for the winter and summer samples at week 7, the viral count for the winter sample was higher than that of the summer samples (Figure 4.7.1.1.1B).

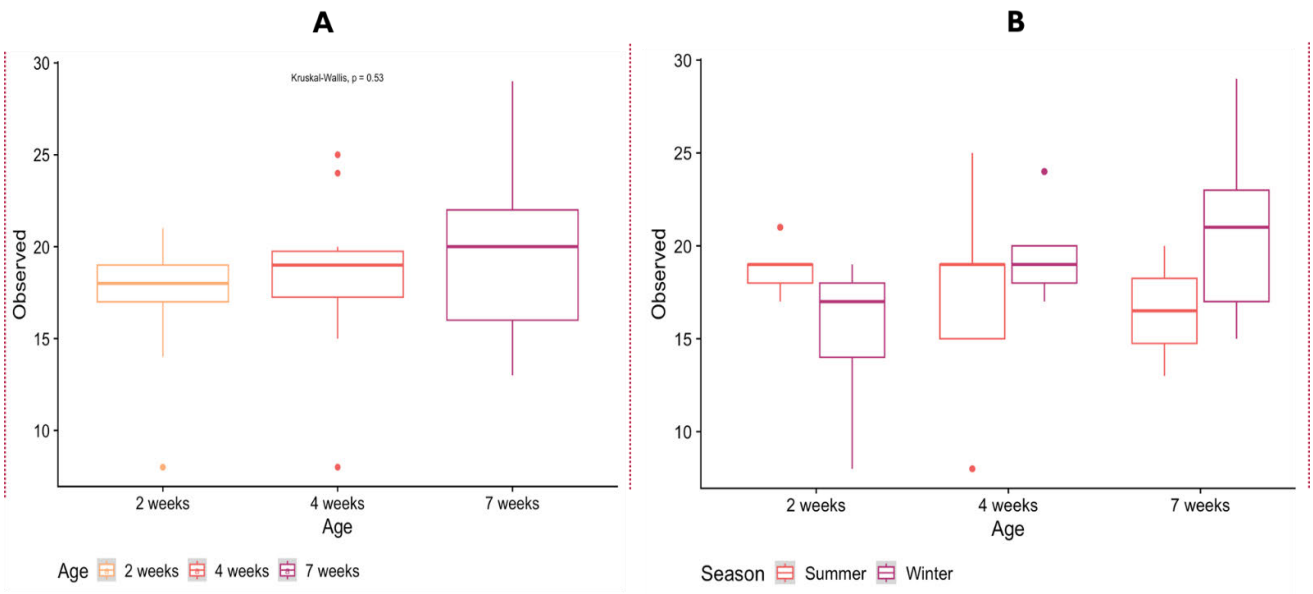



Figure 4.7.1.1.1 The observed alpha diversity metric determination of chicken RNA virome as a function of (A) age, and (B) season.

Similarly, the result of alpha diversity (within samples) using metrics that examines both species richness as well as evenness are illustrated in the box plot denoted as Figure 4.7.1.1.2. The Shannon's index considers richness more, while the Simpson's index places emphasis on the evenness of the viral species.

For Shannon diversity, the 4 weeks' samples had the highest species richness among the three age groups studied (2, 4, 7 weeks). The Shannon's richness based on season showed that the species dynamics of summer were evidently much higher than winter samples at 2 weeks (Figure 4.7.1.1.2C). In contrast to the trend observed at week 2 summer and winter samples, here, the summer and winter samples observed at week 4 appeared to be similar. However, at week 7, a slightly higher species abundance was observed for its summer samples than its winter samples.

The Simpson's diversity findings were remarkably comparable to the Shannon's diversity results, with the most species even samples detected at week 4 for age groups. (Figure 4.7.1.1.2D). In addition, the results of diversity measures based on seasons revealed that same species evenness was seen at both summer and winter sample at weeks 4 and 7. However, at week 2, the samples had more species unevenness for their summer and winter collection seasons (Figure 4.7.1.1.2D).

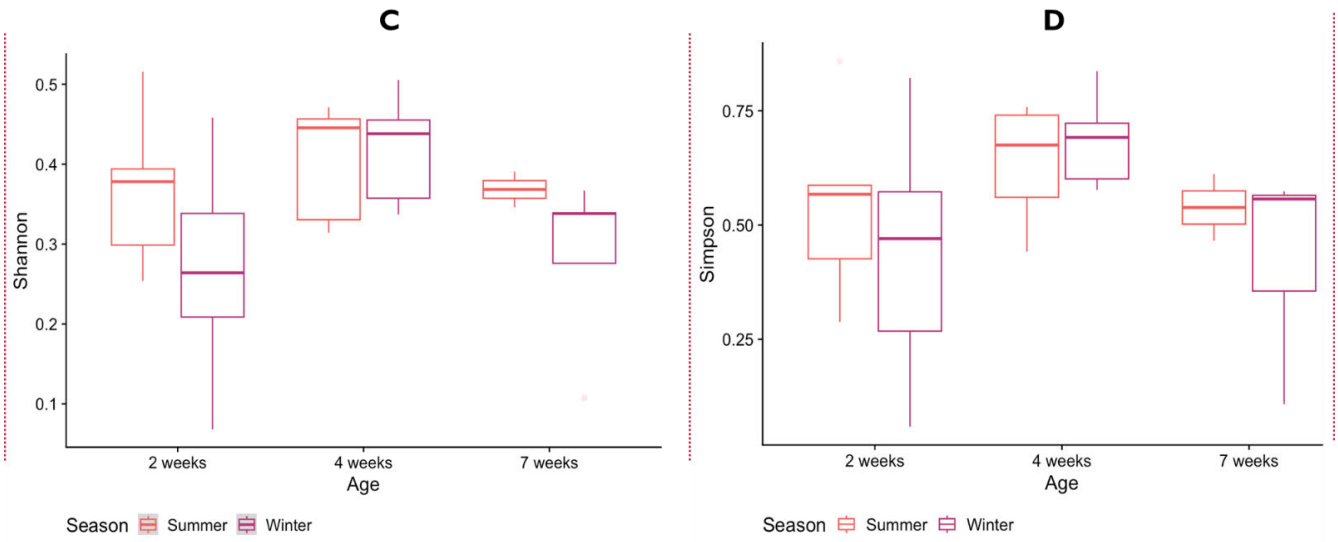



Figure 4.7.1.1.2 The alpha diversity metric determination of chicken RNA virome as a function of age and season using (C) Shannon-Wiener index, and (D) Simpson's index

Overall, the alpha diversity for age (p= 0.116; df = 2) and seasons (p = 0.172; df =1) were not significant. Further analysis of variance using the non-parametric Kruskal-Wallis's H rank sum test on alpha diversity within groups showed that the values for age (p = 0.146; $R^2$ = 3.849; df = 2) and observed season (p = 0.241; $R^2$ = 1.371; df = 1) were not significant. A post-hoc analysis, specifically the Bonferroni method and Turkey's multiple comparisons conducted as a follow up to determine which groups means were not significantly different from each other, showed that the alpha diversity of all individually paired group was not significant (p > 0.05) (Appendix 8.7).

### 4.7.1.2 Beta diversity

The Bray-Curtis dissimilarity index matrix employed as ordination analysis compared the clustering of RNA viral communities of all faecal sample groups using distance matrices. The results of the beta diversity metrics for age and season determined using the abundances of the principal coordinate analysis (PCoA) based on the Bray-Curtis and Jaccard showed that with exception of few outliers, the individual samples from weeks 4 and 7 samples clustered distinctively by their ages (Figures 4.7.1.2A and B). However, for both abundance (Bray-Curtis) and presence/absence (Jaccard) theorems, it was revealed that the 2 weeks samples were more radically spread, clustering with both the weeks 4 and 7 samples (Figures 4.7.4A and B).

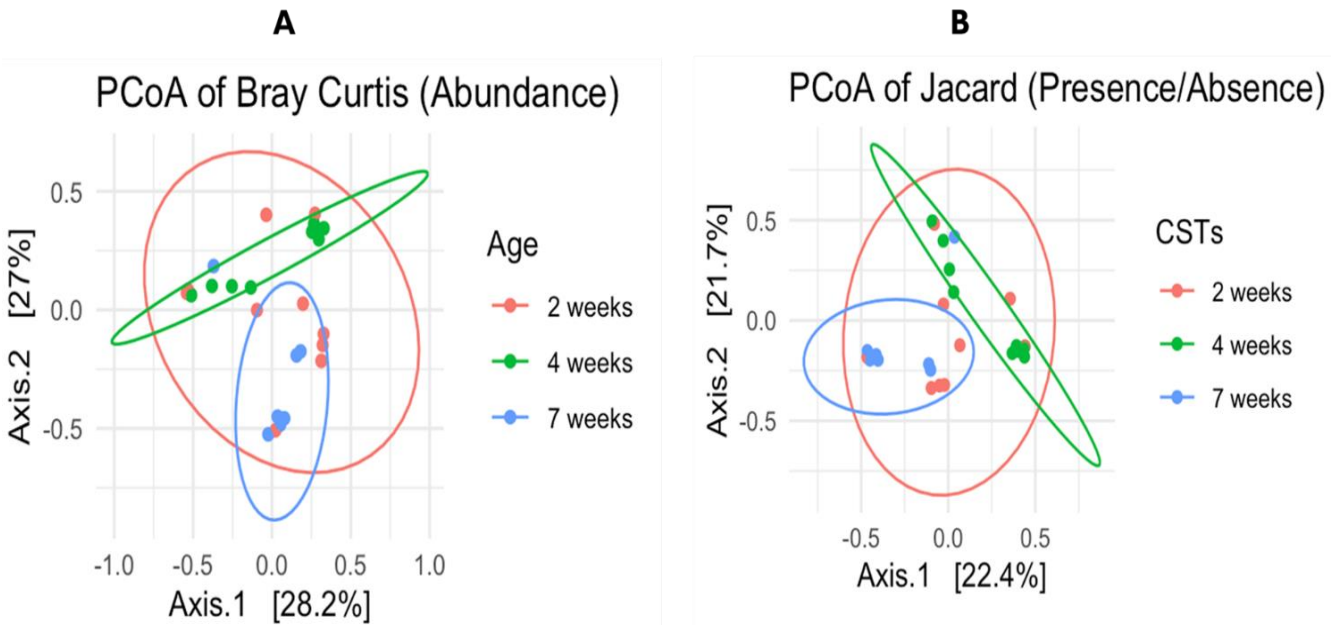



Figure 4.7.1.2 Principal co-ordinate analysis on chicken viral abundance and diversity as functions of age and season, presented as A) Bray-Curtis's similarity index and B) Jaccard index. The Age mean he individual age of birds whose samples were obtained Which CSTs means the chicken seasonal collection timepoints.

Furthermore, the statistical evaluation of the Bray-Curtis beta diversity metrics showed that the beta diversity between age groups were significant (P = 0.01099; $R^2$ = 0.17085; df = 2), and consistent with that of the two seasons revealing high significant difference between the seasons (P = 0.000999; $R^2$ = 0.24107; df = 1) (Appendix 8.8). Overall, based on the data sets in this study, the beta diversity measures revealed that there is a significant difference in the viral abundance and diversities between age groups (2, 4 and 7 weeks) and seasons (Summer and winter) with p-value < 0.05. Hence, the null hypothesis was rejected.

## 4.8 Evolutionary relationships of some selected RNA viruses identified from faecal samples of chickens

In this section, only the results of viruses with nearly complete genome coverage or complete conserved RdRp gene responsible for their replication are presented. Of the 48 viral species identified in this study, only 21 viral species had genome coverage >80%. Among these 21 viruses, only those with genome coverage > 90 were considered. Hence the results of the genetic relationship of 5 avian viral species identified from chicken faecal sample are presented below.

### 4.8.1 *Avihepevirus magniiecur*

The phylogenetic results of the near complete genome (98.8%) of *Avihepevirus magniiecur* obtained from chicken samples in this study showed that it shares 83.7% identity with a recently identified complete genome of a novel virus (ON922634.1) from Ringed-necked Pheasant (*Phasianus colchicus*) and chickens (*Gallus gallus domesticus*) in France. The phylogenetic analysis results showed that this virus alongside the novel virus from France (ON922634.1) clustered distinctively in a new clade away from all other clades including those previously isolated from chickens (MZ736614.1, MW924815 and MK050107.1) (Figure 4.8.1.1).

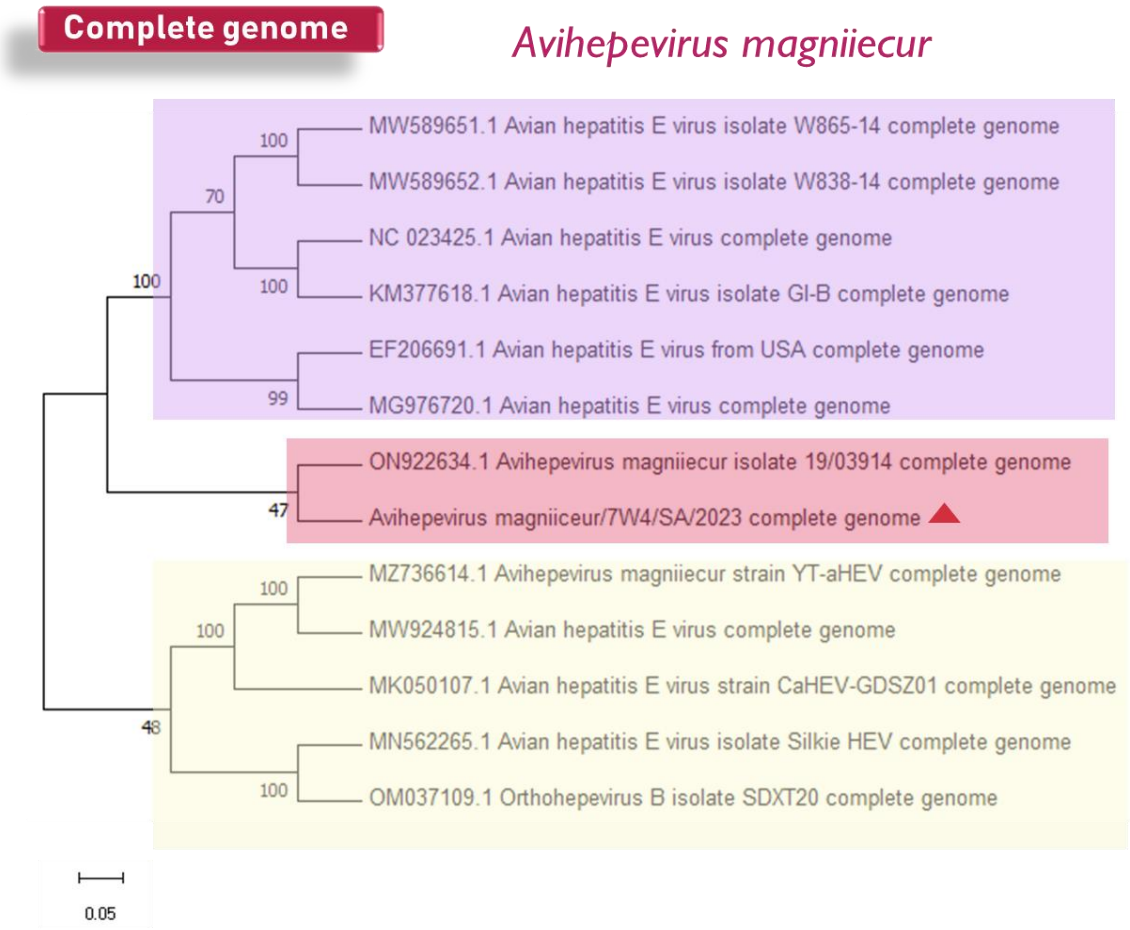



Figure 4.8.1.1 Phylogenetic analysis of *Avian hepatitis E virus*, using the neighbour-joining method of the complete genome was inferred along with representative members of Human avian Hepatitis E. The tree was midpoint rooted for clarity and the branch length support was estimated using 1000 bootstrap replicates. The label with a red triangle denoted the viral specie identified from chicken sample in this study.

### 4.8.2 *Bavaria virus*

The phylogenetic result of *Bavaria virus* with 99.3% genome coverage from chicken sample 2S4 is depicted in Figure 4.8.2.1. This result showed that this virus is most similar with the provisional reference sequence (RefSeq) of novel *Calicivirus chicken/V0021/Bayern/2004* (NC_075411.1) virus from Germany. In addition, it was observed that this virus clustered with the only member of this genus *Bavovirus*. Further alignment of the virus against the non-redundant (nr) NCBI database also showed it shared only 85% nt and 96.1% aa with Chicken calicivirus strain RS/BR/2015 in a different clade from *Nacovirus* (Figure 4.8.2.1).

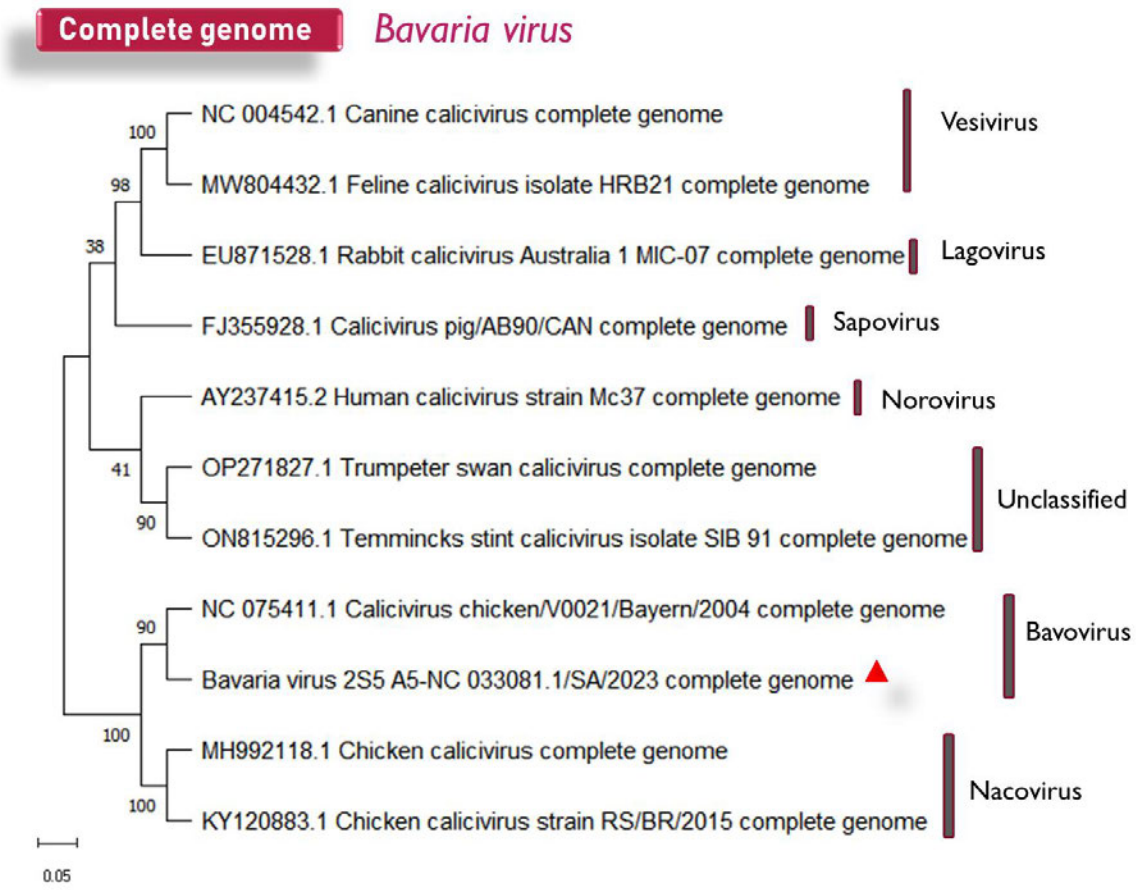



Figure 4.8.2.1 Phylogenetic analysis of *Bavaria virus*, using the neighbour-joining method of the complete genome was inferred along with representative genera of *Caliciviridae*. The tree was midpoint rooted for clarity and the branch length support was estimated using 1000 bootstrap replicates. The label with a red triangle denoted the viral specie identified from chicken sample in this study.

### 4.8.3 Picornaviruses

The phylogenetic result of the polyprotein gene of *Orivirus A* (96.1%,), *Quail picornavirus QPV1/HUN/2010* (94.2%,) and *Chicken megrivirus* (99.8.1%) from chicken sample 2W2,4W5 and 2W3 respectively, are depicted in Figure 4.8.3.1. *Chicken megrivirus* was seen to cluster with polyprotein of *Chicken megrivirus* strain chicken/B21-CHV/2012/HUN (YP_009513232.1) obtained from Hungarian chickens, *Quail picornavirus QPV1/HUN/2010* was in a sister clade with a novel *Quail picornavirus QPV1/HUN/2010* isolated from domesticated *Coturnix coturnix* (YP_004935357) while clustering with *Picornavirales sp* (ULF99733.1) isolated from faecal samples of Chinese chickens. Similarly, *Orivirus A* clustered with an *Orivirus A* strain Pf-CHK1/OrV-A2 isolated from cloacal samples of Hungarian chickens.

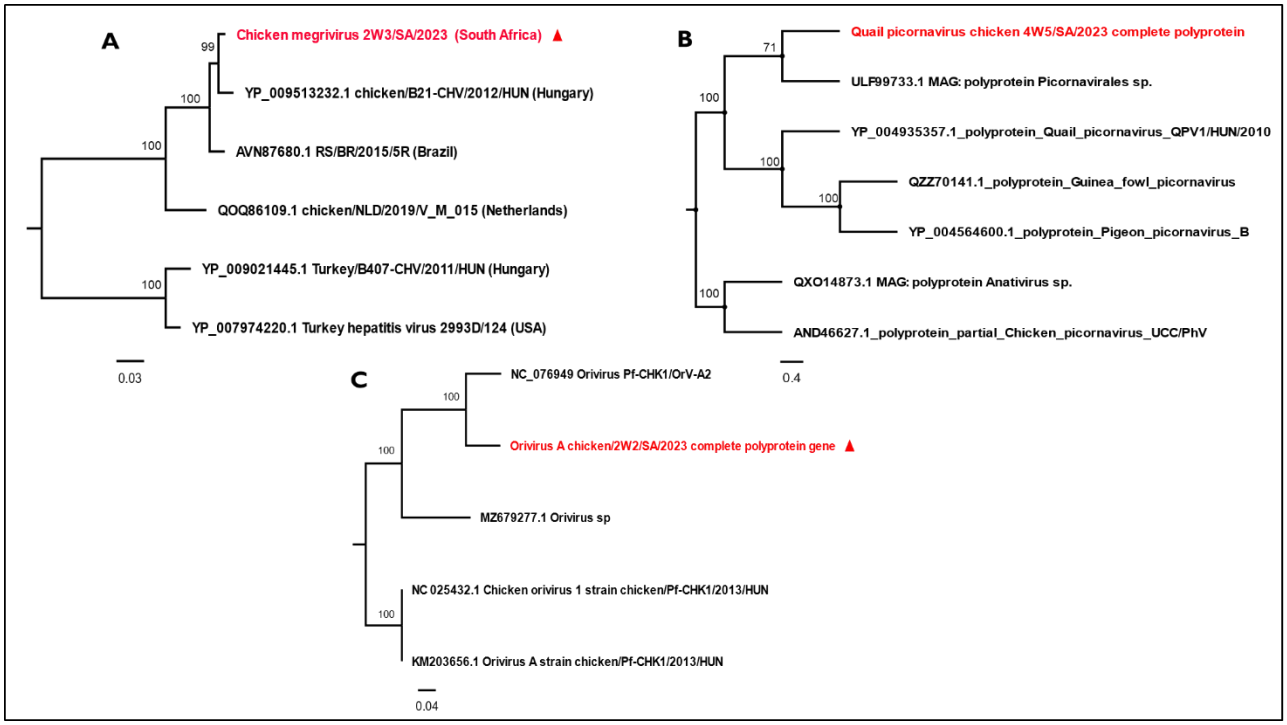



Figure 4.8.3.1 Phylogenetic analysis of members of the *Picornaviridae* family, using the maximum likelihood using the polyprotein gene (Jone-Taylor-Thornton model). The trees were midpoint rooted for clarity and the branch length support was estimated using 1000 bootstrap replicates. Viruses in red fonts with red triangle were identified from chicken faecal sample in this study. A. Phylogeny of *Chicken megrivirus*. B. Phylogeny of *Quail picornavirus QPV1/HUN/2010* using the polyprotein gene C. Phylogeny of *Orivirus A* using the polyprotein gene.

In addition, the phylogenetic results of 3D polyprotein of other picornaviruses identified in this study: *Sicinivirus A, Gallivirus A, Avisivirus B, Megrivirus C2* are shown in Figure 4.8.3.2. Phylogenetic

analysis of *Sicinivirus A* showed its closest relative to be 3D pol of strain UCC001, obtained from chicken caecal samples from Ireland (YP_009021777.1), *Gallivirus A* was most related to 3D pol of *Gallivirus A1* isloate 518C (YP_009055057.1) from Hong Kong chickens. Also the Avisivirus B was seen to be most related to the 3D polyprotein of *Chicken picornavirus 2* isolate 44C (YP_0055013.1) from Hong Kong chickens while the *Megrivirus C2* was observed to be most related to strain Chicken/MG9567/Brazil/2012 *Chicken megrivirus* from Brazil.

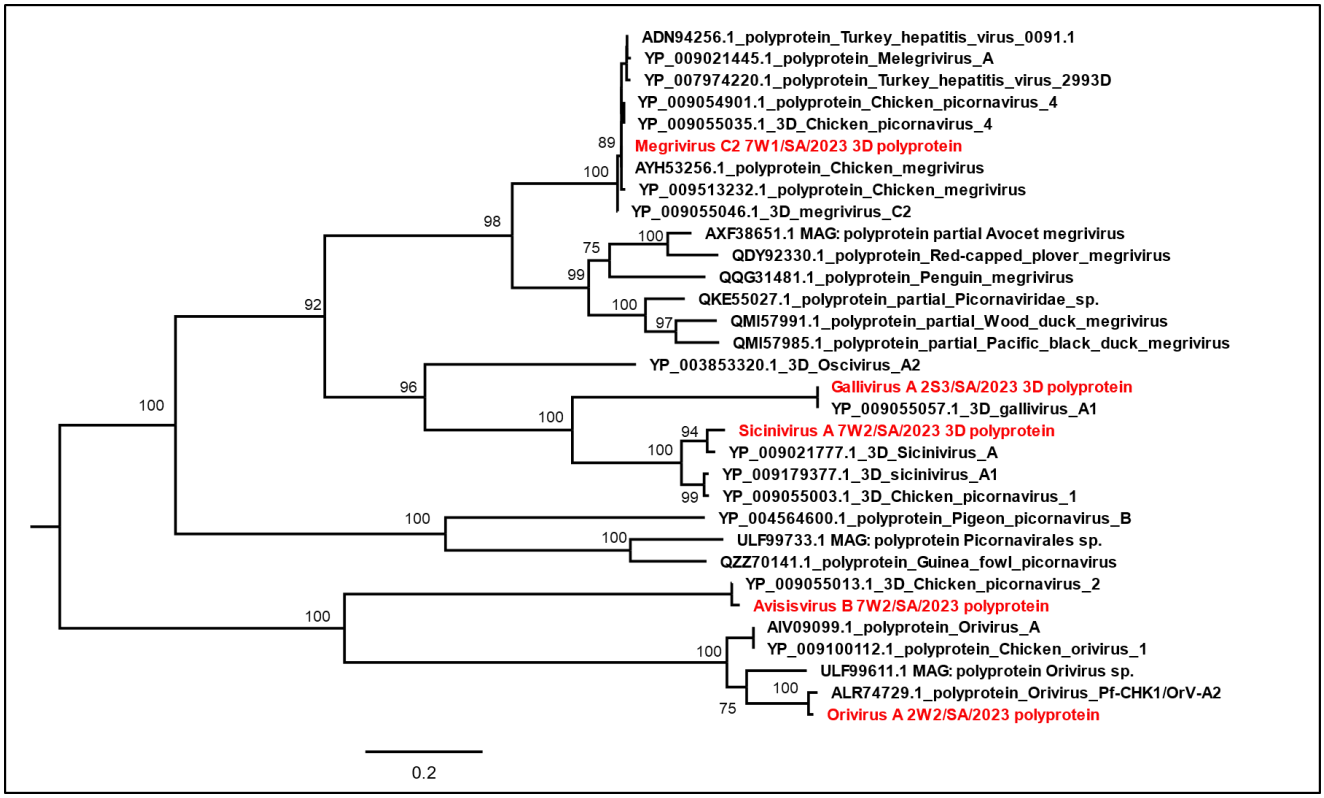



Figure 4.8.3.2 Phylogenetic analysis of the 3D polyprotein of picornaviruses, *Gallivirus A*, *Sicinivirus A*, *Avisivirus B*, *Megrivirus C2* and *Orivirus A*, using the maximum likelihood method (Jone-Taylor-Thornton model). The trees were midpoint rooted for clarity and the branch length support was estimated using 1000 bootstrap replicates. Viruses in red fonts were identified from chicken faecal samples in this study.

### 4.8.4 Chicken astroviruses

The phylogenetic results of the nearly complete genome of two novel CAstV each with nearly complete genome coverage of 99.9%, 99.9%, 99.6%, 98.8% and 99.0% were recovered from the 7W2, 7W3, 2W4, 7W1, and 7W4 samples, respectively (Table 4.6.1.1). The first novel CAstV obtained from this study was from 7W3, 7W1 and 7W4 with about 78.8%, 63.2%, 77.6% nt identity

and 82.5%, 59.1%, 81.6%, aa identity, respectively. The second novel Avian nephritis virus from 2W4 and 7W2 had a mere 58.3%, 57.8% nucleotide identity (nt) and 47.7 and 47.4% amino acids (aa) identity, respectively with all identified *Avian nephritis virus*.

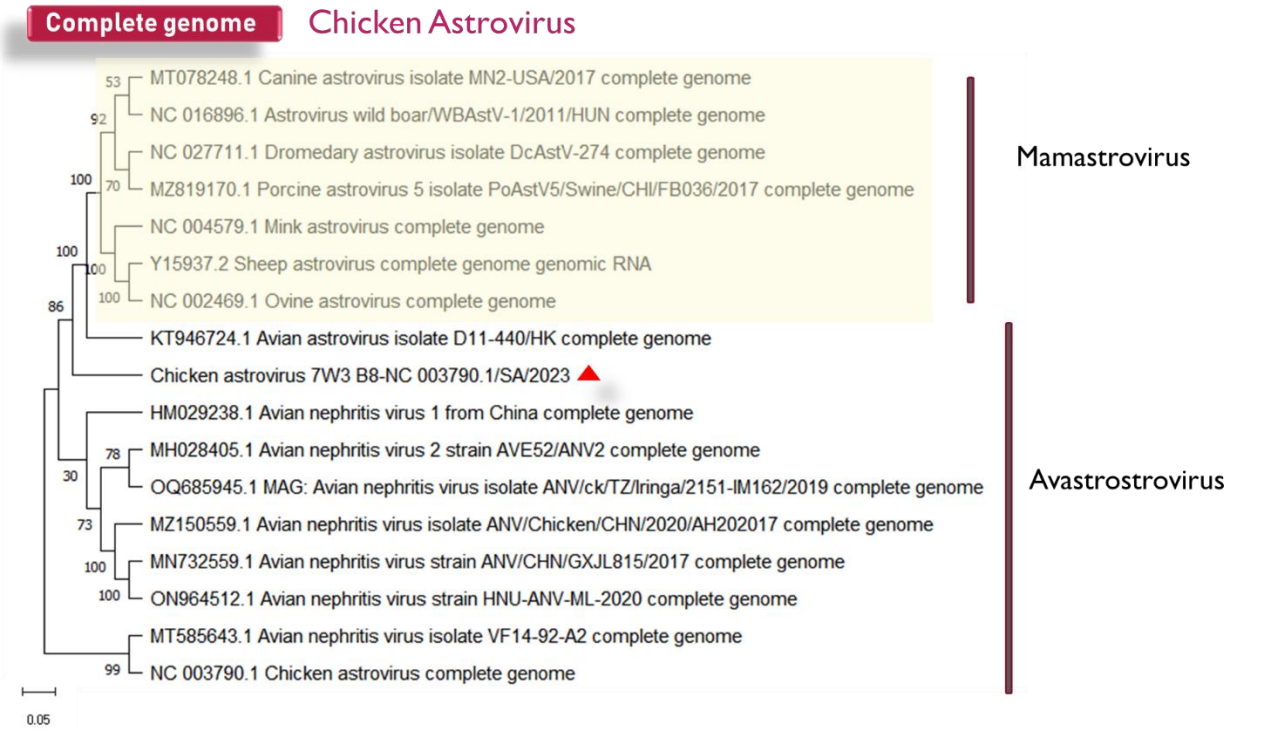



Figure 4.8.4.1 Phylogenetic analysis of novel *Chicken astrovirus* using the neighbour-joining method of the complete genome was inferred along with representative genera of *Astroviridae*. Trees were midpoint rooted for clarity, only and branch length support was estimated using 1000 bootstrap replicates.

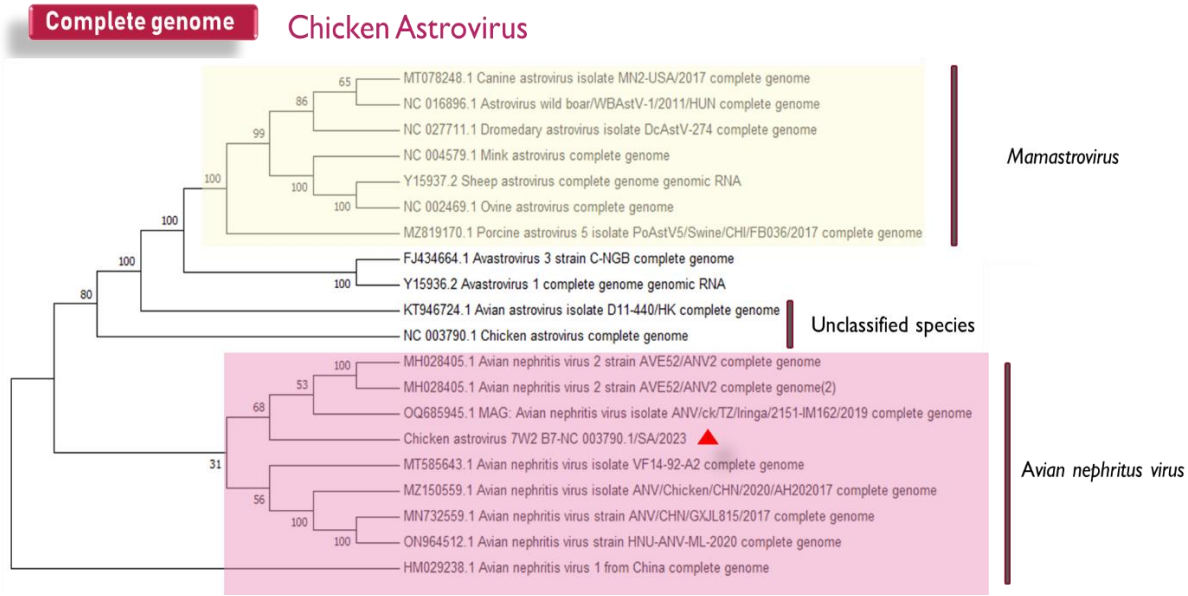


Figure 4.8.4.2 Phylogenetic analysis of novel *Avian nephritis virus* using the neighbour-joining method of the complete genome was inferred along with representative genera of *Astroviridae*. Trees were midpoint rooted for clarity, only and branch length support was estimated using 1000 bootstrap replicates.

# CHAPTER 5. Discussion

Metagenomic NGS offer combined advantages of speed, sensitivity, automation, and high-throughput deep sequencing that has allowed unprecedented advances in the characterization of complex microbial communities including viruses (Chang *et al.,* 2020; Kang *et al.,* 2021). The present metagenomic investigation used faeces to analyze the intestinal RNA viruses found in South African chickens. Research efforts examining the virome and bacteriome profiles have adopted the use of faecal samples in both domestic (Lima *et al.,* 2017; Castro *et al.,* 2018; Kwok *et al.,* 2022) and wild bird species (Vibin *et al.,* 2018; Sarker 2021; Wang *et al.,* 2022). Furthermore, faecal sampling is an ideal method, for animal-based studies, since it allows for continuity, temporal analysis, and monitoring of the digestive tract microbiome at various phases without the need to kill the animal (Stanley *et al.,* 2015; Videvall *et al.,* 2018). However, faecal samples include bacterial and fungal loads that are often more abundant than the target viruses of interest. This necessitates the need to reduce the bulk burden of these unwanted nuclei by targeted enrichments to retrieve viral genetic material of interest. In this study, the major viral enrichment strategies employed included antibiotics treatment, homogenization, filtration and nuclease treatment.

## 5.1 Faecal sample enrichment

Avian studies have used antibiotics treatment as a way of reducing non-viral nuclei including studies in pigs (Theuns *et al.,* 2016), domestic (Walker *et al.,* 2019) and wild (Canuti *et al.,* 2019; Wille *et al.,* 2021) birds. In this study, the choice of antibiotic was considered from three antibiotic classes to ensure a broader depletion range namely, aminoglycosides (streptomycin and gentamicin), beta-lactamase (penicillin) and a polyene antifungal class (amphotericin B). In addition, a homogenization step was also employed, without the use of beads. According to the findings by Conceição-Neto *et al.,* (2015), the process of homogenizing faecal matter using beads has been observed to cause significant depletion of viruses, attributed to their breakdown by the beads, which consequently results in the release of their nuclei material. For the enrichment step through filtration, studies have shown that filters with pore sizes of 0.45 and 0.8 μm are ideal for removing bacteria and eukaryotic cells while recovering viruses, whereas filters with pore sizes of 0.22 μm filter off large viruses, particularly mimiviruses and some members of the *Nidovirales* (Conceição-Neto *et al.,* 2015; Vibin *et al.,* 2018). Hence, both 0.45 and 0.8 μm were employed in this study for the filtration step. However, it is notable that despite the adoption of filtration, certain host bacteria and other organisms that possess sizes smaller than the pore size and/or have their nuclei exposed will persist in the filtrate

(Conceição-Neto *et al.,* 2015). Finally, taking advantage of the rigidity of the viral capsid, an additional treatment step using nuclease enzymes was employed in this study to further deplete free-floating host background nuclei. Overall, the contributing effects of these enrichment steps were crucial for optimal recovery of viruses from the faecal samples.

## 5.2 Viral RNA enrichment

The observed too low outcomes of viral RNA from 12 of the 27 samples after extraction excluding control N1 (PBS + antibiotics + nuclease enzymes) showed that viral RNA occurred in very low concentrations in chicken faecal samples. This observation is consistent with the consensus by researchers that there are comparatively small amounts of RNA viruses in samples in comparison to their bacteria counterparts or competing genomic DNA (Zhang *et al.,* 2019; Fitzpatrick *et al.,* 2021). The observed low outcomes, which appear to be below the quantification threshold, may be attributed to the targeted design of the Qubit® RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA) used in this study. This kit was designed to ensure both precision and accuracy in the measurement of RNA sample concentrations that fall within the range of 0.25 ng/µl to 100 ng/µl (Ogunbayo *et al.,* 2023). Furthermore, the sample-to-sample RNA concentration variations could be explained by the inherent nuclei content differences in original faecal matter, since they were collected from different chickens at diverse time points and not necessarily a result of bias from extraction. Moreover, samples were processed in parallel. Overall, the RNA integrity results obtained in the study could be regarded as significant as they agree with the standard 260/280 ratio of 2.0 for high quality RNA samples (Roy *et al.,* 2020), hence, were acceptable for downstream processes.

Importantly, this study was rationalized to target only the RNA virome of chickens, thus requiring a conscious exclusion of DNA viruses while enriching for the target (viral RNA). Although, the RNA extraction stage was carried out using QIAamp viral RNA mini kit, nevertheless, this kit was not equipped to completely distinguish between viral RNA and cellular DNA. In addition, QIAamp viral RNA mini kit has been demonstrated to extract DNA as effectively as it extracts RNA (Devaney *et al.,* 2016). Furthermore, the presence of DNA in the sample potentially hampers the efficient removal of host globin mRNA and rRNA as stipulated in the protocol for rRNA depletion by New Englands Biolabs (NEB), thus the DNA removal step remains critical for this study. Expectedly, the significant increase in RNA concentrations across samples after RNA extraction and post removal of DNA, purification and concentration could be attributed to the impact of additional concentration of sample volumes, resulting in a final volume of 14 µl. In addition, this increase may also be ascribed to

suboptimal removal of impurities which could have been more effectively managed with the additional purification procedure. Interestingly, the observed RNA integrity result was seen to be consistent with the improved 260/280 RNA integrity results obtained after purification in other similar studies (Vishnuraj *et al.,* 2020; Nandi Jui *et al.,* 2022).

## 5.3 Viral mRNA enrichment

Sequencing the entire RNA sample, including rRNA results in a significant loss of sequencing reads, which pertain to host nucleic acids or other impurities in the nucleic acid sample (Andreou *et al.,* 2023). This, in turn, escalates the sequencing expenses. Studies have shown that rRNA makes up a significant proportion of the overall RNA, ranging from 80 to 90% of a viral RNA genome (Palazzo and Lee 2015; Buckingham 2019; Liefting *et al.,* 2021). As a result of the considerable competing amounts of these host genetic materials in viral samples, the identification of viruses may be impeded. Hence, rRNA removal was carried out in this study. The significant decrease in RNA concentration post-rRNA depletion in 17 of the 27 chicken faecal samples could be mean that that the chicken RNA samples were dominated by their host rRNA. Reports from Liefting *et al.,* (2021) and Andreou *et al.,* (2023), showed that rRNA constitute two-third of the total RNA in microbiome samples. The resultant non-rRNA concentration, highlights the significance of this phase in RNA viral metagenomic research, as well as the efficacy of using specific DNA probes for rRNA depletion. In general, the results of the 260/280 RNA quality assessment are congruent with the established benchmark of 2.0 for pure RNA (Roy *et al.,* 2020), making them suitable to be used in downstream procedures.

## 5.4 Complementary DNA synthesis and whole transcriptome amplification

Expectedly, the observed too high cDNA concentrations of faecal samples, after the amplification of the entire transcriptome showed that the enrichment steps employed were effective from the cDNA results obtained. These excessively high cDNA concentrations observed could be attributed to the phenomenon that the direct transcription amplification results of cDNA far exceed the quantifiable threshold of the Qubit fluorimeter, which is 0.25 ng/µl to 100 ng/µl (Ogunbayo *et al.,* 2023). This could also be true for the control samples despite the absence of any nuclei-template and is consistent with previous studies (Andreou *et al.,* 2023; Smit *et al.,* 2022) that WTA holds the ability to amplify even relatively small amounts of RNA (cDNA) while maintaining high fidelity. Whole transcriptome amplification produces high-fidelity results not just with relatively small amounts of RNA but also

with degraded RNA samples (Tomlins *et al.,* 2006). Therefore, the use of amplified cDNA from polyadenylated mRNA has remained the gold standard of major transcriptomics (Hrdlickova *et al.,* 2017; Andreou *et al.,* 2023). However, it is also important to note that the stepwise enrichment reagents, PCR reagents, enzymes and random primers present in the control samples also undergo amplification during the WTA steps and might have contributed to the excessively high cDNA concentrations observed with the chicken faecal samples in this study. This observation regarding the presence of these artifacts other than the nucleic material contributing to high cDNA concentration agrees with the reports of Mogotsi *et al.* (2019) on the faecal samples of infants.

## 5.5 Libraries and sequencing quality

As a measure towards ensuring that samples are not over-diluted post transcriptome amplification and to further enrich more viruses in this study, the high concentration cut-offs (800- 1000 ng/ml), adapted from the stipulated standard range (200 to 1000 ng/ml) employed showed cDNA libraries with significantly high concentration as anticipated. This was made possible due to the use of high concentration of cDNA starting material, and a wide acceptable concentration range allowed for libraries prepared from QIAseq kit. The consistent and uniform concentration results obtained across all generated cDNA samples in this study shows that the QIASeq FX RNA kit was an ideal choice of library preparation kit. This is indicative of the moderately high indexed cDNA concentrations (up to 22 ng/µl) obtained after library preparation and the average sample concentration of 916.1 ng/µl in 10 µl after normalization obtained from this study. The observed high Phred score of 38 (nearly 40) obtained from all samples can be attributed to a good sequence base calling accuracy of 99.99%, with an error rate of 0.01 per 1 in 10,000 reads and is consistent with recent reports (Mogotsi *et al.,* 2020; Ogunbayo *et al.,* 2023) with Phred score greater than thirty (>30). Overall, obtaining high-quality RNA viral sequences from avian specimens through the use of NGS approach involves making careful, informed choices that best align with the underlying rationale of the study at each specific stage of the NGS methodology.

## 5.6 The chicken gut RNA virome

In this study, the total paired end reads of 9,426276 obtained from the mNGS analyses are consistent with those of total sequence reads obtained from previous metagenomic studies on chicken gut virome (Devaney *et al.,* 2016; Shah *et al.,* 2016; Lima *et al.,* 2019). The moderately lower proportion of viral reads (55.6%) to non-viral reads (44.4%) was anticipated and aligns with previous metagenomics

associated avian species (François and Pybus 2020; Truchado *et al.,* 2020; Kwok *et al.,* 2022). However, within this proportion, the 99% prevalence hits to eukaryotic viruses implies that the targeted enrichment for non-ribosomal targets was successful. Despite the lower percentage of viral to non-viral reads, the high viral diversity can be deduced from the identification of about 48 viruses spanning across 4 kingdoms, 6 phylla, 8 classes, 11 orders, 15 viral families, 21 genera and some unclassified viruses including 27 dsRNA and 17 ssRNA, and 2 dsDNA tail phages from *Siphoviridae.* The absence of avian DNA viruses associated with the gut of chicken could mean that the enrichment and DNase treatment procedures undertaken were effective. However, the presence of two dsDNA tail phages was not expected, although they occurred in very low read abundances. Nevertheless, studies on gut virome reported the presence of tail phages from families *Myoviridae Podoviridae* and *Siphoviridae* in American (Day *et al.,* 2015) and UK (Devaney *et al.,* 2016) chickens, despite the former employing DNA removal procedures.

Importantly, the recovery of seven prevalent viral families' (*Coronaviridae, Picornaviridae, Reoviridae, Astroviridae, Caliciviridae, Picorbirnaviridae,* and *Retroviridae*) among samples could mean that members of these viral families are more often detected from chickens. The occurrence of the identified viral families in this study is in agreement with previous reports on metagenomic investigation of chicken gut virome (Day *et al.,* 2015; Devaney *et al.,* 2016; Shah *et al.,* 2016; Lima *et al.,* 2019). These studies have commonly identified all the viral families with exception to only few studies reporting member of the *Retroviridae* mainly in diseased chicken presented with Rou sarcoma or avian leukosis lesions (Bande *et al.,* 2016; Mason *et al.,* 2020; Chen *et al.,* 2021; Chen and Li 2022). Notably, the observed most abundant viral families (*Picornaviridae, Reoviridae,* and *Coronaviridae*) have been reported from similar studies on chicken virome from Netherlands (Kwok *et al.,* 2022), Brazil (Lima *et al.,* 2017; Lima *et al.,* 2019), UK (Devaney *et al.,* 2016), and Switzerland (Kubacki *et al.,* 2022). In addition, an Iranian tracheal virome study on broiler chicken reported only two of these families, *Picornaviridae and Coronaviridae* in high abundance (Rajeoni *et al.,* 2021). Similarly, the considerably high prevalence of avian viruses (48%), surpassing viruses linked to other hosts, can be attributed to the targeted enrichment strategy which significantly reduced the presence of non-target microbes. The high genome coverage (> 90%) of most of these avian viruses obtained except for infectious bursal disease virus, indicates good biological sample content and effective storage condition that prevented viral degradation. Studies have also shown that the accuracy and quality of the sequencing data obtained in microbiome studies, can be influenced by the storage of the samples prior to nucleic acid extraction, and through other downstream analyses to sequencing

(Fouhy *et al.,* 2015; Marotz *et al.,* 2021). Some recent studies have identified a variety of picornavirus genera in seemingly healthy bird species from different orders particularly, from the *Galliformes* (Lima *et al.,* 2017), *Anseriformes* (Wille *et al.,* 2021), *Charadriiformes* (Vibin *et al.,* 2020a), and Psittaciformes (Sarker 2021). Hence, it could be deduced that picornaviruses are normal flora of the chicken's gut since studies have reported them to be prevalent in healthy and/or diseased states (Devaney *et al.,* 2016; Lima *et al.,* 2019; Kubacki *et al.,* 2022) and across different developmental stages (Shah *et al.,* 2016; Kwok *et al.,* 2022).

Many viruses exhibit host tropism limited to specific species or families of organisms, plants, insects' animals or human. Nevertheless, certain viruses have demonstrated a remarkable ability to replicate in multiple hosts through the acquisition of adaptive features. Furthermore, the range of hosts that a pathogen can infect plays a crucial role in determining its classification and transmission patterns (Sallinen, *et al.,* 2020). In addition, it is widely believed that diversification and evolution processes are influenced by the pathogen's host range. The identification of *Aspergillus fumigatus partitivirus 2*, a fungal virus known to cause opportunistic immune related infections in chicken, is of concern. In humans, this pathogen causes mild to severe lung related infections leading to bleeding and often gotten through the inhalation of its spores in air. Recent studies have shown that this virus lacks host specificity (Cheng *et al.,* 2020; Nururrozi *et al.,* 2020; Lofgren *et al.,* 2022) which can now be attributed to its occurrence and replication in the studied chickens. Overall, the data obtained from this metagenomic study of asymptomatic South African chickens reveals high faecal RNA viral diversity that may be circulating in poultry farms. In addition, it could be highlighted that even in the absence of an observable overt disease condition, the gut RNA virome of chicken are highly diversified. Hence proper measure should be put in place for the routine testing of key pathogenic viral agents at live poultry bird markets. Furthermore, the coinfection of some pathogenic viruses observed, particularly from *Rotavirus F* and *G* for the winter samples and *Avian leukosis virus* and *Rous sarcoma virus* indicates, may indicate that some of these viruses may be more widely spread than imagined in commercial flocks. Additionally, the identification of important pathogens of public health concern and novel viruses from the present study including, mammalian, plant and fungal viruses are further discussed.

## 5.7 Pathogenic avian viruses identified from South African chickens

The RNA viruses categorized as avian RNA viruses in the present investigation are RNA viruses widely known and reported by previous studies to replicate in bird's species. Over the years, different

RNA viruses have been associated with specific/defined disease conditions in birds, while for many of them, their disease pathogenesis is yet to be elucidated or explored. For instance, the majority of *Picornaviridae* members and Picobirnaviruses have not been proven to cause diseases in birds. Pathogenic chicken RNA viruses have been implicated in enteritis (Lima *et al*.,2019; Kubaicki *et al*., 2022), malabsorption syndrome, immune suppression (Khataby et al., 2020; Van Borm et al., 2021), big liver, and spleen disease (BLSD) (Luo *et al*., 2021; Zhang *et al*., 2022) among others which result in the underperformance, morbidity and even death of chickens. In this section, a discussion of established pathogenic avian RNA viruses identified from this study is explored, while others of non-avian hosts or those whose disease pathogenesis has not been confirmed are discussed in other sections. However, this is not a confirmation that all others not discussed in this section do not cause significant diseases in avian hosts.

*Avihepevirus magniiecur*, commonly known as *Avian Hepatitis E virus* (aHEV), is an emerging zoonotic pathogen (Park *et al.,* 2016; Nan *et al.,* 2017; Sun *et al.,* 2019) with significant public health concern. This virus has over the years shown infections in humans, pigs, rabbits, and dogs among others, with about 6 genotypes currently existing (Zhang *et al.,* 2022). Recent studies have implicated aHEV to be a major cause of hepatitis-splenomegaly syndrome (HSS) or big liver and spleen disease (BLSD) in chickens (Iqbal *et al.,* 2019; Ouoba *et al.,* 2019; Osamudiamen *et al.,* 2021; Matos *et al.,* 2022; Zhang *et al.,* 2022; Wang *et al.,* 2023). The identification of this virus in ten chicken samples in both summer (4S1 and 4S5) and winter (2W3, 4W1, 4W2, 4W3, 4W5, 7W2. 7W3, and 7W4) periods at high genome coverage (98.8%) may imply that this virus may be circulating in South African poultry farms, with a potential threat of infecting humans. Interestingly, studies have attributed major aHEV infections to the consumption of contaminated water or undercooked meat (Park *et al.,* 2016). Nevertheless, this aHEV has been initially identified from apparently healthy chickens in the western Africa (Osamudiamen *et al.,* 2021).

Retrotranscribing viruses are viruses capable of reverse transcribing their genome into DNA in host cells and encompasses about 11 genera and 68 species assigned by the ICTV as of June 2023. Retroviruses causing infections in birds are conserved within the *Alpharetrovirus*. These retroviruses are known to cause cancerous infections in chickens and turkeys. Major cancerous infections occurring in avian species including chickens, have been majorly attributed to two major retroviruses, *Rous Sarcoma virus* (Khare *et al.,* 2022) and *Avian leukosis virus* (Liang *et al.,* 2019; Wang *et al.,* 2019; Freick *et al.,* 2022). The identification of these viruses from both summer and winter samples

in this study, though at lower coverage showed that they may be circulating in commercial poultry farms in South Africa.

Reoviruses are economically important wildly spread pathogens causing substantial morbidity and mortality in poultry. Studies have attributed *Avian Orthoreovirus* to arthritis and tenosynovitis (Yan *et al.,* 2021; Jiang *et al.,* 2022; Mirzazadeh *et al.,* 2022) while *Rotavirus G* and *F* have been associated with acute gastroenteritis in poultry (Lima *et al.,* 2019; de Oliveira *et al.,* 2021; Kubacki *et al.,* 2022; Pinheiro *et al.,* 2023). These reoviruses cause significant lesions, leading to immune suppression, malabsorption syndrome and water droppings in poultry which impacts their state of health (Mirzazadeh *et al.,* 2022). The higher viral abundance of *Avian Orthoreovirus and Rotavirus G* at week 2 summer and winter samples *and Rotavirus F* at week 2 winter samples may indicate that these chickens may have suffered from these viruses. Recent studies have demonstrated that younger chicks are more highly susceptible to reoviral infections, particularly *Rotavirus G and F* (Vibin *et al.,* 2020a; de Oliveira *et al.,* 2021; Pinheiro *et al.,* 2023). Hence, the observed decrease in viral abundance of reoviruses (*Avian Orthoreovirus, Rotavirus G* and *Rotavirus F*) from week 2 to week 7, is in agreement with the findings from these studies. This could be explained by the high susceptibility of juvenile chickens to reoviral infection particularly rotaviruses, which are listed among the major causes of death in young chickens. Seasonally, while *Rotavirus G* prevailed in both seasons, *Rotavirus F* only occurred in winter samples which may imply a seasonal emergence of *Rotavirus F* in South African chickens. In addition to reoviruses, *Infectious bronchitis virus* or *Avian coronavirus* are transmissible causative agents of diarrheal related infections in young birds and immune suppression resulting from frequent nasal discharge, conjunctivitis in poultry (Hassan *et al.,* 2022; Najimudeen *et al.,* 2022). The family *Coronaviridae* is currently made up of 3 genera, *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus*. Notably, in addition to the ICTV (International Committee on Taxonomy of Viruses) newly proposed genus, only gammacoronaviruses cause infection in birds. Recent studies have established its pathogenesis as a nephropathogene which cause death in chickens and turkeys (Wang *et al.,* 2020; Van Borm *et al.,* 2021; Yin *et al.,* 2022a). The occurrence of the pathogen, *Avian coronavirus* in 20 out of 27 studied chicken samples implies that despite presenting no observable disease, the studied chicken may have suffered enteritis at different development ages. This may be attributed to high-rate viral transmission of *Infectious bronchitis virus*/ *Avian coronavirus* among birds, particularly in multi-age commercial farm structures.

Furthermore, within the poultry industry, there exist two avian astroviruses that have been recognized and thoroughly characterized for their economic significance, which are *Avian nephritis virus* (ANV) and *chicken astrovirus* (CAstV) (Bulbule *et al.,* 2013). These viruses exhibit serological dissimilarities while maintaining antigenic similarities amongst themselves (Raji and Omar 2022). *Chicken astrovirus* (CAstV) is an avian virus that affects chickens and belongs to the *Avastrovirus* genus of the Astroviridae family. CAstV has been linked to a variety of diseases and syndromes in young broiler chickens, including malabsorption syndrome (Devaney *et al.,* 2016; Kang *et al.,* 2018), visceral gout disease (Li *et al.,* 2022; Luan *et al.,* 2022), and white chick syndrome (Sajewicz-Krukowska and Domanska-Blicharz 2016; Kang *et al.,* 2018; Palomino-Tapia *et al.,* 2020). In this study, the occurrence of the pathogen, CAstV in 23 out of 27 studied chicken faecal samples implies that there might be high prevalence of CAstV in the studied chickens. Phylogenetically, the novel CAstV from 7W2 sample with 99.9% genome coverage was related to an Avian astrovirus isolate D11-440/HK, of Hong Kong feral pigeon. With the high recombination rates associated with CAstV (Palomino-Tapia *et al.,* 2020), this may imply that the novel CAstV identified may have evolved from avian astrovirus from pigeon and the known CAstV from chicken, as a result of mutations and recombination. In addition, different studies have implicated *Avian nephritis virus* with enteritis in poultry including chickens (Gogoi *et al.,* 2017; Nuñez *et al.,* 2018; Yin *et al.,* 2021; Xia *et al.,* 2023). The Avian *nephritis virus,* identified from chicken faecal sample 7W3 with 99.9% was most related to two avian nephritis isolates; an ANV/CK/TZ/Iringa from the cloacal sample of free-range Tanzanian chickens and a Brazilian AVE52/ANV2 isolate. This could imply that the *Avian nephritis virus* from this study may be a recently evolved African strain, sharing only 78.8 nt and 82.5% aa identify with all existing Avian nephritis viruses. It could also suggest that the novel *Avian nephritis virus* to be an imported strain, since Brazil is a major poultry importing country for the South African nation. Studies have reported novel chicken astroviruses from China (Yin *et al.,* 2021; Luan *et al.,* 2022), Iran (Mafigholami *et al.,* 2021), Malaysia (Raji *et al.,* 2022), and Poland (Sajewicz-Krukowska and Domanska-Blicharz 2016). Additionally, Similar studies has identified *Avian nephritis virus* from chickens including novel ones (Gogoi *et al.,* 2017; Yin *et al.,* 2022b; Xia *et al.,* 2023). Taken together, considering the increasing number of astroviruses of avian origin that are awaiting classification by the ICTV and continuous identification of novel avian astroviruses correlated with increased sampling of unexplored region from different, it can be concluded that CAstVs and Avian *nephritis virus* possess rich genetic diversity which is not fully known and can further be explored.

In addition to the single-stranded avian viral pathogens identified in the present study is a member of the double-stranded family *Birnaviridae*, with the ability to replicate in fishes, mammals and a variety of birds. The commonest birnaviruses causing acute infection in poultry is an *Avibirnavirus* generally known as Gumboro disease virus. While infections in younger birds may be mild, lethal cases are recorded for birds 3 to 7 weeks of age (Fan *et al.*, 2019; Orakpoghenor *et al.*, 2020; Lian *et al.*, 2022). In this study, the identification of *Infectious bursal disease virus or* Gumboro disease virus, an *Avibirnavirus* is of great concern. It is also surprising since these chickens' received vaccine against *Gumboro disease* virus before 2 weeks of age. Nevertheless, since this virus was detected from only one sample (7W5), it could be speculated that it may be vaccine strain or as a result of missed vaccination. However, studies have reported *Infectious bursal disease virus* to be higher in birds of 3-6 weeks of age, often causing bursal lesions and immune suppression resulting to susceptibility to other secondary infections, which may lead to the death (Fan *et al.*, 2019; Orakpoghenor *et al.*, 2020; Lian *et al.*, 2022). Finally, another pathogen identified in this study which has gained significant attention is a fungal virus, *Aspergillus fumigatus partititivirus*. This virus is an opportunistic pathogen causing immune suppression and respiratory related diseases in chickens (Cheng *et al.*, 2020; Nururrozi *et al.*, 2020). Recent studies have shown that this virus lacks host specificity (Lofgren *et al.*, 2022) which could be explained by its infection in the studied chickens.

## 5.8 Mammalian RNA viruses identified from the study

In this study, the identification of mammalian viruses from all age groups and seasons of commercially bred chicken faecal samples may be attributed to the close interactions between animals, interspecies spillover of RNA viruses and chance transmission events attributed to viruses. The characterization of unclassified *Picornavirales*, *Picornavirales Tottori-HG1* from swine may imply that this virus may be circulating in poultry farms in Durban, KwaZulu-Natal province, quite unnoticed. Interestingly, recent metagenomic investigation on the Oral RNA virome of backyard swine farms from the KwaZulu-Natal Province, South Africa, identified *Picornavirales Tottori-HG1*, with results showing that it originated from Japan (Chauhan *et al.*, 2022). Therefore, it is of utmost importance to elucidate the mechanisms underlying the transmission dynamics and spillover events pertaining to various viruses affecting farm animals, particularly with a focus on swine viruses, in order to comprehensively understand their spillover and impact on poultry populations on a global level. In addition, the identification of diverse mammalian picorbirnaviruses (PBVs) from this study is not utterly surprising. This can be attributed to the lack of host specificity of the viral family with only one genus *Picobirnavirus* and two genogroups, GI and GII. Currently, there is contradictory

information regarding the real host of PBVs, owing to the short sequences used to identify PBVs which are not typical of the whole RdRp gene phylogeny, hence making their defined identification and classification difficult (Knox *et al.,* 2018). Surprisingly, this virus has members with unsegmented genomes and those whose segments have evolved into segments (Delmas *et al.,* 2019; Ullah *et al.,* 2022). Not only have PBVs been found in humans, invertebrates, and birds, but they have also been hypothesized to be phages (Boros *et al.,* 2018; Krishnamurthy and Wang, 2018) and eukaryotic fungal viruses (Yinda *et al.,* 2018; Kleymann *et al.,* 2020). However, as of June 2023, nearly all PBVs have remained unclassified by the ICTV including *Chicken picobirnavirus.* In this study, the identification of *Orthopicobirnavirus* hominis, *Otarine picobirnavirus* from California sea lion, *Picobirnavirus dog/KNA/2015* from dog, *Picobirnavirus green monkey/KNA/2015* from African green monkey, and *Porcine picobirnavirus* from swine, further explains the highly diversed nature of PBVs. Although, the direct and indirect interactions between some of these mammal and poultry cannot be totally ignored. Particularly, humans are responsible for feeding these commercial chickens, dogs are common pets, monkeys are highly diverse in South Africa, and some commercial poultry farms rear other animals such as pig. Nevertheless, it remains unclear and difficult to explain how the *Otarine picobirnavirus* may have occurred in the studied chicken. Notwithstanding, previous studies have identified *Otarine picobirnavirus, Picobirnavirus dog/KNA/2015* and *Picobirnavirus dog/KNA/2015* in human respiratory samples (Ogunbayo *et al.,* 2022) and swine faecal samples (Chauhan *et al.,* 2022) in South Africa. In addition, recent studies have reported *Picobirnavirus dog/KNA/2015* (Atasoy *et al.,* 2022; Chauhan *et al.,* 2022; Ogunbayo *et al.,* 2022).

## 5.9 Diet-associated RNA viruses identified from the study

The identification of *Tomato mosaic virus* from tomato, *Pepper mild mottle virus* from pepper, *Tobacco mild green mosaic virus* from tobacco and *Festuca pratensis amalgavirus 1* from meadow fescue (*Festuca pratensis)* can be associated with different plant sources of the diet regimens eaten by these chickens. Recent avian metagenomics studies have reported the occurrence of food associated viruses (Wille *et al.,* 2021; French *et al.,* 2022; Shan *et al.,* 2022). French et al. (2022) have reported the discovery of an invertebrate Nelson astrovirus-like 1, exhibiting a significant amino acid similarity of 99.6%, within the avian species of robins and tuii birds. Additionally, the study revealed the presence of two unidentified non-host bastroviruses in tomits and song thrushes, which are believed to be associated with the birds' diets. Similarly, the identification of insect-like viruses *Hubei orthoptera virus 1, Hubeic picornavirus 24* and *Wuhan insect virus* from this study may be attributed to use of insects as feed ingredients for poultry. The use of insects has gained attention due

to their potential as a sustainable and alternative protein source in animal diets (Elahi *et al.,* 2022). In South Africa, the absence of a national regulatory structure governing the use of insects as animal feeds, including for chickens, the insect-rearing industries have predominantly capitalized raising insects as animal feeds for the poultry industries. Hence this may be associated with the identification of insect viruses in recent avian metagenomics studies (Wille *et al.,* 2020; Shan *et al.,* 2022). Shan *et al*. (2022) identified food associated members of *Dicistroviridae* and *Iflaviridae* while Wille et al. (2020) discovered the Taggert virus, a nairovirus from chinstrap penguins, as well as a new Bruthen virus from the family *Phenuviridae*, a sister family of Phlebovirus that consists of tick association viruses. Since some RNA viruses previously thought to be conserved in plants, arthropods, and insects have been isolated from birds, further research is needed to determine the true host of these viruses and the effect of diet sources on the diversity and abundance of RNA viruses in birds.

## 5.10 Effect of age and season on viral diversities and abundance

Due to the perception of viruses as obligate parasites and/or pathogens, many virome studies have remained descriptive majorly identifying and describing the viruses, while ignoring some factors that may trigger and/or modulate the occurrence of these viruses in avian hosts, including chickens. Hence, it is imperative to acknowledge that there exist numerous significant factors with the capability to modulate the occurrence or emergence of viruses within avian species. Previous studies on wild birds have shown that virus-virus interaction (Wille *et al.,* 2018), seasonality (Lambrecht *et al.,* 2016; Lisovski *et al.,* 2017; Vibin *et al.,* 2020b) and age structure (Wille *et al.,* 2021; Hill *et al.,* 2023) including host specie may impact on the diversity of viruses in birds.

In the present study, it was anticipated that there would be the emergence of distinctive patterns in terms of age and seasonal variations pertaining to specific viral families and/or species found in the chicken samples under investigation. The notable prevalence of *Reoviridae* members, namely *Rotavirus G*, and *F* occurring in both winter and summer samples, and *Avian orthoreovirus*, with high abundance in only week 2 can be ascribed to the heightened vulnerability of juvenile avian specimens to gastroenteritis infections. This susceptibility arises from the ongoing maturation of their immune system, which is characterized by low immunity to infections.

Recent studies have reported a high susceptibility rate of rotaviral infections in younger birds, (Vibin *et al.,* 2020a; de Oliveira *et al.,* 2021; Kubacki *et al.,* 2022; Pinheiro *et al.,* 2023), which is a major cause of runting-stunting/malabsorption syndrome resulting in delayed development and lower productivity in chickens Also, the observed significant decrease in reoviruses at 7 weeks may be

further explained by the stable adaptation features exhibited by these mature chickens with full grown immune systems. The observed presence of the members of *Picornaviridae*, particularly, *Sicinivirus A* throughout all age groups and seasons, may imply that major members of this viral family are commensal and hence may form part of the normal flora of the chicken gut virome. Moreover, different studies have reported picornaviruses to be prevalent in both healthy (Lima *et al.,* 2017; Vibin *et al.,* 2020b; Wille *et al.,* 2021; Kwok *et al.,* 2022) and diseased (Devaney *et al.,* 2016; Lima *et al.,* 2019; de Oliveira *et al.,* 2021; Kubacki *et al.,* 2022) bird species including chickens.

Interestingly, the observed seasonal distribution exhibited by some viral species, such as *Rotavirus F, genus Lamdavirus, Infectious bursal disease virus, Festuca pratensis almagavirus 1, Fusarium poae virus 1,* and unclassified RNA virus ShiM-2016 may be attributed to their emergence during the winter period. Similarly, the exclusive occurrence of *Tomato mosaic virus, Pepper mild mottle virus*, and some partitiviruses (*Botryotinia fuckeliana partitivirus 1, Cryptosporidium parvum virus 1, Pythium nunn virus 1, Ustilaginoidea virens partitivirus 2,* and *Verticillium dahliae partitivirus 1*) during the summer periods, may either mean that the viruses are more predominant during the summer or may be associated with the diet (milled grains) taken by these chickens which may carry these fungal viruses. Based on available studies as at the time of this study, there was limited information on the effect of season on viral abundance and diversity in domesticated birds. However, a recent Australian study on wild birds (Pacific black ducks, Chestnut teals, Grey teals and Wood ducks) explored the effect of season on the abundance of the viral families *Picornaviridae* and *Parvoviridae* (Vibin *et al.,* 2020b*)*. Findings from their study showed that picornaviruses were mainly found during late autumn to late winter months while parvoviruses were found throughout the year. In addition, Vibin *et al.,* (2020b) demonstrated that members of these two viral families varied not only in virus composition across species and time but also in their abundances, despite the sharing the same habitat at the times of sampling. Overall, in the current study, at viral family levels, all viral families occurred throughout the two seasons explored with exception to some non-avian viral families, *Siphoviridae, Amalgaviridae, Partitiviridae, Potyviridae* and *Totiviridae* comprising of fungal and plant viruses. However, the viral diversity varied at species level with some viruses occurring in winter period while being entirely absent in summer or vice versa. In addition, viral abundance varied with age rather season, where the effect of age chicken may have been higher than the seasonal effect in the abundance of some of these viruses. Nevertheless, their occurrence by mere chance events cannot be entirely ruled out, particularly with viruses that occurred in lower abundance

and in <3 samples. To the best of our knowledge, the present study is the first to explore the effect of season on viral abundance in chickens, particularly from the African continent.

**The impact of age and season on the Alpha diversity of viruses**

With the observed results within sample groups exhibiting similar diversity levels for Shannon and Simpson's indices, the lack of distinct trend across age groups and seasons can be explained further. First, the viral abundance pattern within each age group can be homogeneous since these chickens are grouped, housed, fed and vaccinated together, however most commercially bred chickens are multi-age in the same farm. In addition, the inconsistence and higher abundance shown by members of the older chicken at week 7 may be attributed to their lowered immunity arising from different systemic and recurrent viral infections they may have suffered, leading to different survival mechanism by individual chicken in this age group. This can be explained further by the issue of chance events as has been observed for some transient viral infections, thus leading to chicken individual biased survival features or resistance to some viruses even within the same group. Furthermore, other environmental factors, such as sanitation, lack of biosecurity measures and the introduction of birds from other commercial farms may influence viral abundance. Overall, the observed ($P > 0.05$) no significant effect observed for alpha diversity indicates that regardless of age or seasons, the overall viral abundance of samples within the same group are similar/homogenous. In addition, viral abundance was higher in juvenile chickens (2 weeks) between the three age groups attributed to their immature immune systems resulting in increased susceptibility to infections. Similar no significant alpha diversity result was observed from a recent study that investigated the impact of age in an *Anseriformes*, Ruddy turnstones with P-value > 0.05 (Wille *et al.,* 2021).

**The impact of age and season on the beta diversity of viruses**

The observed distinct clustering for both indices (Jaccard and Bray-Curtis) between viruses at week 4 and week 7 may indicate a marked difference in viral diversity between the acclimatization stage (week 4) and the mature stage (week 7) of the chicken. Hence it appears that there is no relationship/ limited connectivity/viral sharing between weeks 4 and 7. However, for week 2, the observed clustering results indicate significant viral connectivity/sharing between week 2 and week 4 as well as week 2 and week 7. Hence it can be deduced that, some of the viruses identified from this chicken at week 2 followed through to week 4 and week 7. Overall, the observed beta diversity (between groups) was statistically significant for both age and season ($p<0.05$) and indicates that viral abundance and diversities are influenced by age and season between sample groups for the studied chicken faecal samples. Hence it could be deduced that the diversity and abundance between studied

chicken samples groups may differ and may be dependent on distinct features characterising each group. Therefore, between group of one or more distinct features, the diversity and abundance may be heterogenous. Different avian studies have demonstrated marked differences in viral abundance and diversities between sample sets characterized by different features such as age structure, seasonality, host species and latitude (Lambrecht *et al.*, 2016; Lisovski *et al.*, 2017; Vibin *et al.*, 2020b; Hill *et al.*, 2023). These differences in diversities and abundance between groups whether age or season, may be attributed to factors such as varying temperature changes for season, different feed formulations in form of diet eaten by these chickens at different ages as well as different vaccines and/or treatment administered to chickens in a specific group which may influence viral abundance and diversities.

# CHAPTER 6. Conclusion and future perspectives

## 6.1 Conclusion

In this study, the faecal RNA virome of apparently healthy South African chickens was analyzed at three developmental ages and over two seasons. The well-planned viral enrichment and purification strategies adopted in this study were key to the high recovery rate of avian viruses (48%) using mNGS. The complete genome of novel chicken astroviruses (CAstV) and many previously known viruses, including pathogenic avian viruses, mammalian, fungal and plant viruses were identified in the present study. The two novel astroviruses of avian origin obtained from chicken faecal sample were *Chicken astrovirus* mostly related to a Hong Kong DII-440/HK strain and *Avian nephritis virus* most related to two avian nephritis isolates; an ANV/CK/TZ/Iringa from the cloacal sample of free-range Tanzanian chickens and a Brazilian AVE52/ANV2 isolate. In addition, the detection of mammalian viruses such as *Picornavirales Tottori-HG1, Porcine picobirnavirus, Picobirnavirus dog/KNA/2015, Picobirnavirus green monkey/KNA/2015* in this study could mean a cross species spillage of mammalian RNA viruses to South African chickens. Furthermore, the identification of *Avihepevirus magniiecur* from chickens in this study is of concern, since this virus has the capability to infect humans, hence should be monitored against potential zoonotic outbreak. The detection of picornaviruses, avian coronaviruses, and reoviruses was consistent across samples and some of the viruses such as picornaviruses detected in this study may have a commensal role in the avian virome or become pathogenic under certain conditions. The results from this chicken RNA viral metagenomics study showed that despite the lack of discernible clinical manifestations, RNA viruses remain prevalent within the GIT of chickens. This suggests that chickens may experience recurrent viral infections that could significantly impact the development of their normal gut virome and overall well-being. The relative abundance profiles of investigated chicken viruses were higher in the 2 weeks summer, 4 weeks summer and 7 weeks winter samples. Given that the virome of individual age groups and seasons both changed over the time of chicken development and that multiple co-infections may have occurred, since commercial poultry farms are multi-age, it is difficult to define the emergence of some of these pathogenic viruses, in terms of chicken age or collection time points. Although, the age and seasons considered had no effect on within sample abundance (alpha diversity), a significant effect was however observed between age groups and seasons (beta diversity) based on the studied samples. These diversity results suggest that viral abundance and diversities in chickens in the same age group and season may be similar, however between groups characterized by different age or seasons, varies, hence their viral abundance and diversity are different. The results of phylogenetic

analysis of avian RNA viruses suggests relatedness to viruses previously identified in other countries; For instance, *Bavaria virus* mostly related to German *chicken/V0021/Bayern/2004*, *Avihepevirus magniiecur* mostly related to a novel *Avihepevirus magniiecur* 19/03914 from France, *Chicken astrovirus* related to DII-440/HK from Hong Kong, and *Avian nephritis virus* most related to Avian nephritis strains from Tanzania (ANV/CK/TZ/Iringa) and Brazil (ANE/ANV2). Overall, considering that, the utilization of viral metagenomics in studying the RNA virome of chickens was efficient in the present investigation, yielding diverse viruses including novel viruses. Hence, mNGS is a useful tool and approach with the capability to determine total viruses present in faecal samples of bird and can be explored for its implementation in areas that aim to study viral diseases in birds targeted at enhancing avian gut health.

## 6.2 Future perspectives

Since this study only considered asymptomatic South African chickens, it is highly recommended that future studies should look to compare virome of healthy and diseased South African chickens. This consideration should ensure that comparative investigations include additional measures of ascertaining health status of birds grouped as healthy. In addition, while the present study employed the use of a total 27 chicken faecal samples, future studies should look to increase sample size and as well sample from different provinces of South Africa, since geographical location may influence virome diversity and abundance. Importantly, this study only looked at two of the four known seasons in South Africa (summer and winter), hence, it may be interesting if future research considers sampling across all four seasons to see their robust impact on the dynamics of viral abundance and diversity. Furthermore, there is scarcity of information on the virome of many South African domesticated animals, pets and wild birds, hence future studies should look to decipher the viruses present in these animals. This may be important to track viral host spillage of animal viruses with potential zoonotic threat. Moreso, with recent studies frequently identifying porcine viruses in chicken or vice versa, turkey and/or duck viruses in chicken or vice versa and even dog viruses in chickens, the impacts of interspecies jumping of viruses in farms rearing diverse farm animals or bird species requires investigation. In addition to common city birds, there are so many wild bird's endemic to South Africa whose virome could be explored through mNGS. Information obtained from these unique birds' categories may assist towards better understanding of avian virome. In conclusion, given the recurrent identification of plant-derived viral sequences in different avian virome investigations, future research may endeavour to determine the impact of dietary factors on viral prevalence and diversity within South African poultry populations. Such investigations could

shed light on the potential regulatory role of diet in shaping the composition of the gut virome, thereby presenting a promising avenue for further scientific exploration.

# REFERENCES


Abaidullah, M., Peng, S., Kamran, M., Song, X. and Yin, Z. 2019. Current findings on gut microbiota mediated immune modulation against viral diseases in chicken. *Viruses*, 11 (8): 681.

Abd El-Hack, M. E., El-Saadony, M. T., Alqhtani, A. H., Swelum, A. A., Salem, H. M., Elbestawy, A. R., Noreldin, A. E., Babalghith, A. O., Khafage, A. F. and Hassan, M. I. 2022. The relationship among avian influenza, gut microbiota and chicken immunity: an updated overview. *Poultry Science*, 101 (9): 102021.

Abolnik, C., Ostmann, E., Woods, M., Wandrag, D. B., Grewar, J., Roberts, L. and Olivier, A. J. 2021. Experimental infection of ostriches with H7N1 low pathogenic and H5N8 clade 2.3. 4.4 B highly pathogenic influenza A viruses. *Veterinary Microbiology*, 263: 109251.

Abolnik, C., Strydom, C., Rauff, D. L., Wandrag, D. B. R. and Petty, D. 2019. Continuing evolution of H6N2 influenza a virus in South African chickens and the implications for diagnosis and control. *BioMed Central Veterinary Research*, 15 (1): 1-20.

Ackermann, H.-W. 2003. Bacteriophage observations and evolution. *Research in Microbiology*, 154 (4): 245-251.

Akinyemi, F. T., Ding, J., Zhou, H., Xu, K., He, C., Han, C., Zheng, Y., Luo, H., Yang, K. and Gu, C. 2020. Dynamic distribution of gut microbiota during embryonic development in chicken. *Poultry Science*, 99 (10): 5079-5090.

Alexander, D. J. 2000. Newcastle disease and other avian paramyxoviruses. *Revue Scientifique et Technique*, 19 (2): 443-462.

Alqazlan, N., Alizadeh, M., Boodhoo, N., Taha-Abdelaziz, K., Nagy, E., Bridle, B. and Sharif, S. 2020. Probiotic lactobacilli limit avian influenza virus subtype H9N2 replication in chicken cecal tonsil mononuclear cells. *Vaccines*, 8 (4): 605.

Andreou, I., Storbeck, M., Hahn, P., Rulli, S. and Lader, E. 2023. Optimized Workflow for Whole Genome and Transcriptome Next-Generation Sequencing of Single Cells or Limited Nucleic Acid Samples. *Current Protocols*, 3 (5): e753.

Arndt, C., Davies, R., Gabriel, S., Harris, L., Makrelov, K., Modise, B., Robinson, S., Simbanegavi, W., Van Seventer, D. and Anderson, L. 2020. Impact of Covid-19 on the South African economy. *Southern Africa-Towards Inclusive Economic Development Working Paper*, 111.

Asplund, M., Kjartansdóttir, K. R., Mollerup, S., Vinner, L., Fridholm, H., Herrera, J. A., Friis-Nielsen, J., Hansen, T. A., Jensen, R. H. and Nielsen, I. B. 2019. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clinical Microbiology and Infection*, 25 (10): 1277-1285.

Atasoy, M. O., Isidan, H. and Turan, T. 2022. Genetic diversity, frequency and concurrent infections of picobirnaviruses in diarrhoeic calves in Turkey. *Tropical Animal Health and Production*, 54 (2): 127.

Bacher, R., Chu, Cara, Bolin, J. M., Knight, P., Thomson, J. A., Stewart, R. and Kendziorski, C. 2022. Enhancing biological signals and detection rates in single-cell RNA-seq experiments with cDNA library equalization. *Nucleic Acids Research*, 50 (2): e12-e12.

Bande, F., Arshad, S. S. and Omar, A. R. 2016. Isolation and metagenomic identification of avian leukosis virus associated with mortality in broiler chicken. *Advances in Virology*, 2016.

Bandeira, B., Jamet, J.-L., Jamet, D. and Ginoux, J.-M. 2013. Mathematical convergences of biodiversity indices. *Ecological Indicators*, 29: 522-528.

Beaugrand, G., Reid, P. C., Ibanez, F., Lindley, J. A. and Edwards, M. 2002. Reorganization of North Atlantic marine copepod biodiversity and climate. *Science*, 296 (5573): 1692-1694.

Bogdanova, E. A., Shagin, D. A. and Lukyanov, S. A. 2008. Normalization of full-length enriched cDNA. *Molecular BioSystems*, 4 (3): 205-212.

Bolger, A. M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15): 2114-2120.

Boros, Á., Pankovics, P., Adonyi, Á., Fenyvesi, H., Day, J. M., Phan, T. G., Delwart, E. and Reuter, G. 2016. A diarrheic chicken simultaneously co-infected with multiple picornaviruses: Complete genome analysis of avian picornaviruses representing up to six genera. *Virology*, 489: 63-74.

Bray, J. R. and Curtis, J. T. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27 (4): 326-349.

Brown, J., Pirrung, M. and McCue, L. A. 2017. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, 33 (19): 3137-3139.

Brucker, M. C. 2020. Novel viruses, zoonotic infections, and travel health. *Nursing for Women's Health*, 24 (2): 65-66.

Buckingham, L. 2019. *Molecular diagnostics: fundamentals, methods and clinical applications*. FA Davis.

Bulbule, N., Mandakhalikar, K., Kapgate, S., Deshmukh, V., Schat, K. and Chawak, M. 2013. Role of chicken astrovirus as a causative agent of gout in commercial broilers in India. *Avian Pathology*, 42 (5): 464-473.

Callanan, J., Stockdale, S. R., Shkoporov, A., Draper, L. A., Ross, R. P. and Hill, C. 2018. RNA phage biology in a metagenomic era. *Viruses*, 10 (7): 386.

Callanan, J., Stockdale, S. R., Shkoporov, A., Draper, L. A., Ross, R. P. and Hill, C. 2020. Expansion of known ssRNA phage genomes: from tens to over a thousand. *Science Advances*, 6 (6): 5981.

Canuti, M., Kroyer, A. N., Ojkic, D., Whitney, H. G., Robertson, G. J. and Lang, A. S. 2019. Discovery and characterization of novel RNA viruses in aquatic North American wild birds. *Viruses*, 11 (9): 768.

Cao, Z., Sugimura, N., Burgermeister, E., Ebert, M. P., Zuo, T. and Lan, P. 2022. The gut virome: A new microbiome component in health and disease. *EBioMedicine*, 81: 104113.

Carnaccini, S., Palmieri, C., Stoute, S., Crispo, M. and Shivaprasad, H. 2022. Infectious laryngotracheitis of chickens: Pathologic and immunohistochemistry findings. *Veterinary Pathology*, 59 (1): 112-119.

Carrasco-Hernandez, R., Jácome, R., López Vidal, Y. and Ponce de León, S. 2017. Are RNA viruses candidate agents for the next global pandemic? A review. *Information on Laboratory Animals for Research Journal*, 58 (3): 343-358.

Castro, C. M. O., Chagas, E. H. N., Bezerra, D. A. M., Ribeiro, A. F., da Silva, S. P., Cruz, A. C. R., Júnior, E. C. S., Silva, R. R. and Mascarenhas, J. D. A. P. 2018. A proposed new strain of avian picornavirus in broiler chicken from Brazil. *Genome Announcements*, 6 (12): e00012-00018.

Chang, Wei-Shan, Eden, J.-S., Hall, J., Shi, M., Rose, K. and Holmes, E. C. 2020. Metatranscriptomic analysis of virus diversity in urban wild birds with paretic disease. *Journal of Virology*, 94 (18): e00606-00620.

Chang, C.-M., Coville, J.-L., Coquerelle, G., Gourichon, D., Oulmouden, A. and Tixier-Boichard, M. 2006. Complete association between a retroviral insertion in the tyrosinase gene and the recessive white mutation in chickens. *BioMed Central Genomics*, 7 (1): 1-15.

Chauhan, R. P., San, J. E. and Gordon, M. L. 2022. Metagenomic Analysis of RNA Fraction Reveals the Diversity of Swine Oral Virome on South African Backyard Swine Farms in the uMgungundlovu District of KwaZulu-Natal Province. *Pathogens*, 11 (8).

Chen, Li, H.-W., Cong, F. and Lian, Y.-X. 2021. Avian leukosis virus subgroup J infection alters viral composition in the chicken gut. *Federation of European Microbiological Societies Letters*, 368 (10): fnab058.

Chen and Li, H. 2022. Avian leukosis virus subgroup J infection influences the gut microbiota composition in Huiyang bearded chickens. *Letters in Applied Microbiology*, 74 (3): 344-353.

Cheng, Z., Li, M., Wang, Y., Chai, T., Cai, Y. and Li, N. 2020. Pathogenicity and immune responses of Aspergillus fumigatus infection in chickens. *Frontiers in Veterinary Science*, 7: 143.

Chowdhury, S., Aleem, M. A., Khan, M. S. I., Hossain, M. E., Ghosh, S. and Rahman, M. Z. 2021. Major zoonotic diseases of public health importance in Bangladesh. *Veterinary Medicine and Science*, 7(4), 1199-1210.

Cisek, A. A., Dąbrowska, I., Gregorczyk, K. P. and Wyżewski, Z. 2017. Phage therapy in bacterial infections treatment: one hundred years after the discovery of bacteriophages. *Current Microbiology*, 74 (2): 277-283.

Clavijo, V. and Flórez, M. J. V. 2018. The gastrointestinal microbiome and its association with the control of pathogens in broiler chicken production: A review. *Poultry Science*, 97 (3): 1006-1021.

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C. K., Lavigne, R., Maes, P. and Van Ranst, M. 2015. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Scientific Reports*, 5 (1): 1-14.

Cui, H., Pan, S., Xu, X., Ji, J., Ma, K., Yao, L., Kan, Y., Bi, Y. and Xie, Q. 2022. Molecular characteristics of novel chaphamaparvovirus identified in chickens. *Poultry Science*, 102 (3): 102449.

Dai, M., Xie, T., Feng, M. and Zhang, X. 2022. Endogenous retroviruses transcriptomes in response to four avian pathogenic microorganisms infection in chicken. *Genomics*, 114 (3): 110371.

Day, J. M., Oakley, B. B., Seal, B. S. and Zsak, L. 2015. Comparative analysis of the intestinal bacterial and RNA viral communities from sentinel birds placed on selected broiler chicken farms. *PLoS One*, 10 (1): e0117210.

de Oliveira, L. B., Stanton, J. B., Zhang, J., Brown, C., Butt, S. L., Dimitrov, K., Afonso, C. L., Volkening, J. D., Lara, L. J. and de Oliveira, C. S. F. 2021. Runting and stunting syndrome in broiler chickens: histopathology and association with a novel picornavirus. *Veterinary Pathology*, 58 (1): 123-135.

Delahoy, M. J., Wodnik, B., McAliley, L., Penakalapati, G., Swarthout, J., Freeman, M. C. and Levy, K. 2018. Pathogens transmitted in animal feces in low-and middle-income countries. *International Journal of Hygiene and Environmental Health*, 221 (4): 661-676.

Delmas, B., Attoui, H., Ghosh, S., Malik, Y. S., Mundt, E., Vakharia, V. N. and Ictv Report, C. 2019. ICTV virus taxonomy profile: *Picobirnaviridae*. *Journal of General Virology*, 100 (2): 133-134.

Devaney, R., Trudgett, J., Trudgett, A., Meharg, C. and Smyth, V. 2016. A metagenomic comparison of endemic viruses from broiler chickens with runting-stunting syndrome and from normal birds. *Avian Pathology*, 45 (6): 616-629.

Deviatkin, A. A., Kholodilov, I. S., Vakulenko, Y. A., Karganova, G. G. and Lukashev, A. N. 2020. Tick-Borne Encephalitis Virus: An Emerging Ancient Zoonosis? *Viruses*, 12 (2).

Ding, J., Jiang, T., Zhou, H., Yang, L., He, C., Xu, K., Akinyemi, F. T., Han, C., Luo, H. and Qin, C. 2020. The Gut Microbiota of Pheasant Lineages Reflects Their Host Genetic Variation. *Frontiers in Genetics*, 11: 859.

Edmonds, K. and Williams, L. 2017. The role of the negative control in microbiome analyses. *The Federation of American Societies for Experimental Biology Journal*, 31: 940.943-940.943.

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R. and Weyrich, L. S. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in Microbiology*, 27 (2): 105-117.

Elahi, U., Xu, C.-c., Wang, J., Lin, J., Wu, S.-g., Zhang, H.-j. and Qi, G.-h. 2022. Insect meal as a feed ingredient for poultry. *Animal Bioscience*, 35 (2): 332.

Elferink, M. G., Vallée, A. A., Jungerius, A. P., Crooijmans, R. P. and Groenen, M. A. 2008. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BioMed Central genomics*, 9 (1): 1-9.

Esteban, J. A., Salas, M. and Blanco, L. 1993. Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *Journal of Biological Chemistry*, 268 (4): 2719-2726.

Fan, L., Wu, T., Hussain, A., Gao, Y., Zeng, X., Wang, Y., Gao, L., Li, K., Wang, Y. and Liu, C. 2019. Novel variant strains of infectious bursal disease virus isolated in China. *Veterinary Microbiology*, 230: 212-220.

Fasanmi, O. G., Odetokun, I. A., Balogun, F. A. and Fasina, F. O. 2017. Public health concerns of highly pathogenic avian influenza H5N1 endemicity in Africa. *Veterinary World*, 10 (10): 1194.

Fedor, P. and Zvaríková, M. 2019. Biodiversity indices. *Encyclopedia of Ecology*, 2: 337-346.

Feschotte, C. and Gilbert, C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics*, 13 (4): 283-296.

Fitzpatrick, A. H., Rupnik, A., O'Shea, H., Crispie, F., Keaveney, S. and Cotter, P. 2021. High throughput sequencing for the detection and characterization of RNA viruses. *Frontiers in Microbiology*, 12: 190.

Fouhy, F., Deane, J., Rea, M. C., O'Sullivan, Ó., Ross, R. P., O'Callaghan, G., Plant, B. J. and Stanton, C. 2015. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. *PLoS One*, 10 (3): e0119355.

François, S. and Pybus, O. G. 2020. Towards an understanding of the avian virome. *The Journal of General Virology*, 101 (8): 785.

Freick, M., Schreiter, R., Weber, J., Vahlenkamp, T. W. and Heenemann, K. 2022. Avian leukosis virus (ALV) is highly prevalent in fancy-chicken flocks in Saxony. *Archives of Virology*, 167 (4): 1169-1174.

French, R.K., Filion, A., Niebuhr, C.N. and Holmes, E.C., 2022. Metatranscriptomic comparison of viromes in endemic and introduced passerines in New Zealand. *Viruses*, 14 (7): 1364.

Garcia-Gutierrez, E., Mayer, M. J., Cotter, P. D. and Narbad, A. 2019. Gut microbiota as a source of novel antimicrobials. *Gut Microbes*, 10 (1): 1-21.

Gardy, J. L. and Loman, N. J. 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, 19 (1): 9-20.

Gogoi, S. M., Gulhane, A. B., Deshpande, A. A. and Balaguru, P. 2017. Isolation and Identification of Avian Nephritis Virus from Commercial Broiler Chickens. *Journal of Animal Research*, 7 (2): 299-305.

Grond, K., Sandercock, B. K., Jumpponen, A. and Zeglin, L. H. 2018. The avian gut microbiota: community, physiology and function in wild birds. *Journal of Avian Biology*, 49 (11): e01788.

Gruwell, J., Fogarty, C., Bennett, S., Challet, G., Vanderpool, K., Jozan, M. and Webb Jr, J. 2000. Role of peridomestic birds in the transmission of St. Louis encephalitis virus in southern California. *Journal of Wildlife Diseases*, 36 (1): 13-34.

Gu, Z., Eils, R. and Schlesner, M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32 (18): 2847-2849.

Hameed, M., Wahaab, A., Nawaz, M., Khan, S., Nazir, J., Liu, K., Wei, J. and Ma, Z. 2021. Potential Role of Birds in Japanese Encephalitis Virus Zoonotic Transmission and Genotype Shift. *Viruses*, 13 (3).

Hassan, K. E., Shany, S. A., Ali, A., Dahshan, A.-H. M., Azza, A. and El-Kady, M. F. 2016. Prevalence of avian respiratory viruses in broiler flocks in Egypt. *Poultry Science*, 95 (6): 1271-1280.

Hassan, M. S., Najimudeen, S. M., Ali, A., Altakrouni, D., Goldsmith, D., Coffin, C. S., Cork, S. C., van der Meer, F. and Abdul-Careem, M. F. 2022. Immunopathogenesis of the Canadian Delmarva (DMV/1639) infectious bronchitis virus (IBV): Impact on the reproductive tract in layers. *Microbial Pathogenesis*, 166: 105513.

Hauser, N., Gushiken, A. C., Narayanan, S., Kottilil, S. and Chua, J. V. 2021. Evolution of Nipah virus infection: past, present, and future considerations. *Tropical Medicine and Infectious Disease*, 6 (1): 24.

Hill, S. C., François, S., Thézé, J., Smith, A. L., Simmonds, P., Perrins, C. M., Van Der Hoek, L. and Pybus, O. G. 2023. Impact of host age on viral and bacterial communities in a waterbird population. *The International Society for Microbial Ecology Journal*, 17 (2): 215-226.

Hornung, B. V., Zwittink, R. D. and Kuijper, E. J. 2019. Issues and current standards of controls in microbiome research. *Federation of European Microbiological Societies Microbiology Ecology*, 95 (5): fiz045.

Hrdlickova, R., Toloue, M. and Tian, B. 2017. RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA*, 8 (1): e1364.

Iqbal, T., Rashid, U., Shahzad, N., Afroz, A., Malik, M. F. and Idrees, M. 2019. Molecular Detection of Hepatitis E Virus in Layer Chickens: A Possible Public Health Risk in Pakistan. *Pakistan Journal of Zoology*, 51 (6).

Ishag, H. Z. A., Terab, A. M. A., El Tigani-Asil, E. T. A., Bensalah, O. K., Khalil, N. A. H., Khalafalla, A. I., Al Hammadi, Z. M. A. H., Shah, A. A. M. and Al Muhairi, S. S. M. 2022. Pathology and Molecular Epidemiology of Fowl Adenovirus Serotype 4 Outbreaks in Broiler Chicken in Abu Dhabi Emirate, UAE. *Veterinary Sciences*, 9 (4): 154.

Jeffries, C., Mansfield, K., Phipps, L., Wakeley, P., Mearns, R., Schock, A., Bell, S., Breed, A., Fooks, A. and Johnson, N. 2014. Louping ill virus: an endemic tick-borne disease of Great Britain. *The Journal of General Virology*, 95 (5): 1005.

Jern, P., Sperber, G. O. and Blomberg, J. 2005. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology*, 2 (1): 1-12.

Jiang, X., Yao, Z., He, D., Wu, B., Wei, F., Li, G., Wu, Q., Tang, Y. and Diao, Y. 2022. Genetic and pathogenic characteristics of two novel/recombinant avian orthoreovirus. *Veterinary Microbiology*, 275: 109601.

Johnson, C. K., Hitchens, P. L., Pandit, P. S., Rushmore, J., Evans, T. S., Young, C. C. and Doyle, M. M. 2020. Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proceedings of the Royal Society B*, 287 (1924): 20192736.

Jurasz, H., Pawłowski, T. and Perlejewski, K. 2021. Contamination issue in viral metagenomics: problems, solutions, and clinical perspectives. *Frontiers in Microbiology*, 12: 745076.

Kalkowska, D.A., Boender, G.J., Smit, L.A., Baliatsas, C., Yzermans, J., Heederik, D.J. and Hagenaars, T.J., 2018. Associations between pneumonia and residential distance to livestock farms over a five-year period in a large population-based study. *PLoS One*, 13 (7): e0200813.

Kanda, R., Tristem, M. and Coulson, T. 2013. Exploring the effects of immunity and life history on the dynamics of an endogenous retrovirus. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 368 (1626): 20120505.

Kang, K.-i., Linnemann, E., Icard, A. H., Durairaj, V., Mundt, E. and Sellers, H. S. 2018. Chicken astrovirus as an aetiological agent of runting-stunting syndrome in broiler chickens. *Journal of General Virology*, 99 (4): 512-524.

Kang, K., Hu, Y., Wu, S. and Shi, S. 2021. Comparative Metagenomic Analysis of Chicken Gut Microbial Community, Function, and Resistome to Evaluate Noninvasive and Cecal Sampling Resources. *Animals*, 11 (6): 1718.

Kapusta, A. and Suh, A. 2017. Evolution of bird genomes—a transposon's-eye view. *Annals of the New York Academy of Sciences*, 1389 (1): 164-185.

Kariithi, H. M., Welch, C. N., Ferreira, H. L., Pusch, E. A., Ateya, L. O., Binepal, Y. S., Apopo, A. A., Dulu, T. D., Afonso, C. L. and Suarez, D. L. 2020. Genetic characterization and pathogenesis of the first H9N2 low pathogenic avian influenza viruses isolated from chickens in Kenyan live bird markets. *Infection, Genetics and Evolution*, 78: 104074.

Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30 (4): 772-780.

Khare, V. M., Saxena, V. K., Pasternak, M. A., Nyinawabera, A., Singh, K. B., Ashby Jr, C. R., Tiwari, A. K. and Tang, Y. 2022. The expression profiles of chemokines, innate immune and apoptotic genes in tumors caused by Rous Sarcoma Virus (RSV-A) in chickens. *Genes and Immunity*, 23 (1): 12-22.

King, A. M., Lefkowitz, E. J., Mushegian, A. R., Adams, M. J., Dutilh, B. E., Gorbalenya, A. E., Harrach, B., Harrison, R. L., Junglen, S. and Knowles, N. J. 2018. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Archives of Virology*, 163 (9): 2601-2631.

Knox, M. A., Gedye, K. R. and Hayman, D. T. 2018. The challenges of analysing highly diverse picobirnavirus sequence data. *Viruses*, 10 (12): 685.

Kock, R. and Heuer, C., 2019. Prevention and control of diseases at the interface of livestock, wildlife and humans. *Veterinary Sciences*, 6 (1): 11.

Kolde, R. and Kolde, M. R. 2018. Package 'pheatmap'. *R package*, 1 (10).

Krishnamurthy, S. R., Janowski, A. B., Zhao, G., Barouch, D. and Wang, D. 2016. Hyperexpansion of RNA bacteriophage diversity. *PLoS Biology*, 14 (3): e1002409.

Kubacki, J., Qi, W. and Fraefel, C. 2022. Differential Viral Genome Diversity of Healthy and RSS-Affected Broiler Flocks. *Microorganisms*, 10 (6): 1092.

Kuchinsky, S. C., Frere, F., Heitzman-Breen, N., Golden, J., Vázquez, A., Honaker, C. F., Siegel, P. B., Ciupe, S. M., LeRoith, T. and Duggal, N. K. 2021. Pathogenesis and shedding of Usutu virus in juvenile chickens. *Emerging Microbes and Infections*, 10 (1): 725-738.

Kwok, K. T., de Rooij, M. M., Messink, A. B., Wouters, I. M., Smit, L. A., Cotten, M., Heederik, D. J., Koopmans, M. P. and Phan, M. V. 2022. Establishing farm dust as a useful viral metagenomic surveillance matrix. *Scientific Reports*, 12 (1): 16308.

Lambrecht, B., Marché, S., Houdart, P., Van Den Berg, T. and Vangeluwe, D. 2016. Impact of age, season, and flowing vs. stagnant water habitat on avian influenza prevalence in mute swan (Cygnus olor) in Belgium. *Avian Diseases*, 60 (1): 322-328.

Lebdah, M., Alshaya, D. S., Jalal, A. S., Mousa, M. R., Radwan, M. M., Samir, M., Adel, A., Albaqami, N. M., El-Saadony, M. T. and El-Tarabily, K. A. 2022. Molecular characterization of aviadenovirus serotypes and pathogenicity of the identified adenovirus in broiler chickens. *Poultry science*, 101 (12): 101918.

Lee, D, M., Pedroso, A. A. and Maurer, J. J. 2023. Bacterial composition of a competitive exclusion product and its correlation with product efficacy at reducing Salmonella in poultry. *Frontiers in Physiology*, 13: 2737.

Lee, Sunjin, Kwon, T., Chae, S.-J., Kim, J.-H., Kang, Y. H., Chung, G. T., Kim, D.-W. and Lee, D.-Y. 2016. Complete genome sequence of bacteriophage MA12, which infects both Campylobacter jejuni and Salmonella enterica serovar Enteritidis. *Genome Announcements*, 4 (6): e00810-00816.

Leung, S. S. Y., Carrigy, N. B., Vehring, R., Finlay, W. H., Morales, S., Carter, E. A., Britton, W. J., Kutter, E. and Chan, H.-K. 2019. Jet nebulization of bacteriophages with different tail morphologies–Structural effects. *International Journal of Pharmaceutics*, 554: 322-326.

Li, J., Davis, B. W., Jern, P., Dorshorst, B., Siegel, P. B. and Andersson, L. 2019. Characterization of the endogenous retrovirus insertion in CYP19A1 associated with henny feathering in chicken. *Mobile DNA*, 10 (1): 1-8.

Li, L., Sun, M., Zhang, Y. and Liao, M. 2022. A review of the emerging poultry visceral gout disease linked to avian Astrovirus infection. *International Journal of Molecular Sciences*, 23 (18): 10429.

Lian, J., Wang, Z., Xu, Z., Pang, Y., Leng, M., Tang, S., Zhang, X., Qin, J., Chen, F. and Lin, W. 2022. Pathogenicity and molecular characterization of infectious bursal disease virus in China. *Poultry Science*, 101 (1): 101502.

Liang, X., Gu, Y., Chen, X., Li, T., Gao, Y., Wang, X., Fang, C., Fang, S. and Yang, Y. 2019. Identification and characterization of a novel natural recombinant avian leucosis virus from Chinese indigenous chicken flock. *Virus Genes*, 55: 726-733.

Liefting, L. W., Waite, D. W. and Thompson, J. R. 2021. Application of Oxford Nanopore Technology to Plant Virus Detection. *Viruses*, 13 (8): 1424.

Lima, D. A., Cibulski, S. P., Finkler, F., Teixeira, T. F., Varela, A. P. M., Cerva, C., Loiko, M. R., Scheffer, C. M., Dos Santos, H. F. and Mayer, F. Q. 2017. Faecal virome of healthy chickens reveals a large diversity of the eukaryote viral community, including novel circular ssDNA viruses. *Journal of General Virology*, 98 (4): 690-703.

Lima, D. A., Cibulski, S. P., Tochetto, C., Varela, A. P. M., Finkler, F., Teixeira, T. F., Loiko, M. R., Cerva, C., Junqueira, D. M. and Mayer, F. Q. 2019. The intestinal virome of malabsorption syndrome-affected and unaffected broilers through shotgun metagenomics. *Virus Research*, 261: 9-20.

Lisovski, S., Hoye, B. J. and Klaassen, M. 2017. Geographic variation in seasonality and its influence on the dynamics of an infectious disease. *Oikos*, 126 (7): 931-936.

Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R., Dobson, A. P., Hudson, P. J. and Grenfell, B. T. 2009. Epidemic dynamics at the human-animal interface. *Science*, 326 (5958): 1362-1367.

Loc-Carrillo, C. and Abedon, S. T. 2011. Pros and cons of phage therapy. *Bacteriophage*, 1 (2): 111-114.

Lofgren, L. A., Lorch, J. M., Cramer, R. A., Blehert, D. S., Berlowski-Zier, B. M., Winzeler, M. E., Gutierrez-Perez, C., Kordana, N. E. and Stajich, J. E. 2022. Avian-associated Aspergillus fumigatus displays broad phylogenetic distribution, no evidence for host specificity, and multiple genotypes within epizootic events. *G3 (Bethesda)*, 12 (5): p.jkac075

Long, K. E., Hastie, G. M., Ojkić, D. and Brash, M. L. 2017. Economic impacts of white chick syndrome in Ontario, Canada. *Avian Diseases*, 61 (3): 402-408.

Lou, Y. C., Hoff, J., Olm, M. R., West-Roberts, J., Diamond, S., Firek, B. A., Morowitz, M. J. and Banfield, J. F. 2023. Using strain-resolved analysis to identify contamination in metagenomics data. *Microbiome*, 11 (1): 36.

Luan, Q., Han, Y., Yin, Y. and Wang, J. 2022. Genetic diversity and pathogenicity of novel chicken astrovirus in China. *Avian Pathology*, 51 (5): 488-498.

Lubinga, M. H., Mazibuko, N., Ngqangweni, S., Potelwa, Y. X. and Nyhodo, B. 2018. Impact of export promotion and market development on social welfare in South Africa: Evidence from the agricultural sector. *African Evaluation Journal*, 6 (2): 1-7.

Mafigholami, N., Ghalyanchilangeroudi, A., Hosseini, H., Haghbin Nazarpak, H. and Morshed, R. 2021. Detection and Molecular Characterization of Chicken Astrovirus from Broiler Flocks in Iran: The First Report. *Iranian Journal of Virology*, 15 (1): 1-6.

Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J. and Belshaw, R. 2012. Env-less endogenous retroviruses are genomic superspreaders. *Proceedings of the National Academy of Sciences*, 109 (19): 7385-7390.

Majewska, J., Kaźmierczak, Z., Lahutta, K., Lecion, D., Szymczak, A., Miernikiewicz, P., Drapała, J., Harhala, M., Marek-Bukowiec, K. and Jędruchniewicz, N. 2019. Induction of phage-specific antibodies by two therapeutic staphylococcal bacteriophages administered per os. *Frontiers in Immunology*, 10: 2607.

Malik, Y. S., Dhama, K. and Singh, R. K. 2018. Suppl-2, M1: Emerging and Zoonotic Virus Challenges of Developing Nations. *The Open Virology Journal*, 12: 42.

Marimwe, M. C., Fosgate, G. T., Roberts, L. C., Tavornpanich, S., Olivier, A. J. and Abolnik, C. 2021. The spatiotemporal epidemiology of high pathogenicity avian influenza outbreaks in key ostrich producing areas of South Africa. *Preventive Veterinary Medicine*, 196: 105474.

Marotz, C., Cavagnero, K. J., Song, S. J., McDonald, D., Wandro, S., Humphrey, G., Bryant, M., Ackermann, G., Diaz, E. and Knight, R. 2021. Evaluation of the effect of storage methods on fecal, saliva, and skin microbiome composition. *Msystems*, 6 (2): e01329-01320.

Mason, A. S., Fulton, J. E., Hocking, P. M. and Burt, D. W. 2016. A new look at the LTR retrotransposon content of the chicken genome. *BioMed Central genomics*, 17 (1): 1-13.

Mason, A. S., Lund, A. R., Hocking, P. M., Fulton, J. E. and Burt, D. W. 2020. Identification and characterisation of endogenous Avian Leukosis Virus subgroup E (ALVE) insertions in chicken whole genome sequencing data. *Mobile DNA*, 11 (1): 1-13.

Matos, M., Bilic, I., Tvarogová, J., Palmieri, N., Furmanek, D., Gotowiecka, M., Liebhart, D. and Hess, M. 2022. A novel genotype of avian hepatitis E virus identified in chickens and common pheasants (*Phasianus colchicus)*, extending its host range. *Scientific Reports*, 12 (1): 21743.

Maurin, M., Fenollar, F., Mediannikov, O., Davoust, B., Devaux, C. and Raoult, D. 2021. Current status of putative animal sources of SARS-CoV-2 infection in humans: wildlife, domestic animals and pets. *Microorganisms*, 9 (4): 868.

Methner, U. and Rösler, U. 2020. Efficacy of a competitive exclusion culture against extended-spectrum β-lactamase-producing Escherichia coli strains in broilers using a seeder bird model. *BioMed Central Veterinary Research*, 16 (1): 143.

Mirzazadeh, A., Abbasnia, M., Zahabi, H. and Hess, M. 2022. Genotypic characterization of two novel avian orthoreoviruses isolated in Iran from broilers with viral arthritis and malabsorption syndrome. *Iranian Journal of Veterinary Research*, 23 (1): 74.

Mogotsi, M. T., Mwangi, P. N., Bester, P. A., Mphahlele, M. J., Seheri, M. L., O'Neill, H. G. and Nyaga, M. M. 2020. Metagenomic Analysis of the Enteric RNA Virome of Infants from the Oukasie Clinic, North West Province, South Africa, Reveals Diverse Eukaryotic Viruses. *Viruses*, 12 (11): 1260.

Nabil, N. M., Tawakol, M. M. and Hassan, H. M. 2018. Assessing the impact of bacteriophages in the treatment of Salmonella in broiler chickens. *Infection Ecology and Epidemiology*, 8 (1): 1539056.

Naguib, M. M., Ulrich, R., Kasbohm, E., Eng, C. L., Hoffmann, D., Grund, C., Beer, M. and Harder, T. C. 2017. Natural reassortants of potentially zoonotic avian influenza viruses H5N1 and H9N2 from Egypt display distinct pathogenic phenotypes in experimentally infected chickens and ferrets. *Journal of Virology*, 91 (23): e01300-01317.

Najimudeen, S. M., Barboza-Solis, C., Ali, A., Buharideen, S. M., Isham, I. M., Hassan, M. S., Ojkic, D., Van Marle, G., Cork, S. C. and van der Meer, F. 2022. Pathogenesis and host responses in lungs and kidneys following Canadian 4/91 infectious bronchitis virus (IBV) infection in chickens. *Virology*, 566: 75-88.

Nan, Y., Wu, C., Zhao, Q. and Zhou, E.-M. 2017. Zoonotic hepatitis E virus: an ignored risk for public health. *Frontiers in Microbiology*, 8: 2396.

Nandi Jui, B., Sarsenbayeva, A., Jernow, H., Hetty, S. and Pereira, M. J. 2022. Evaluation of RNA isolation methods in human adipose tissue. *Laboratory Medicine*, 53 (5): e129-e133.

Newmana, H. and Abrahamsb, S. 2018. Zoonotic viral infections in South Africa: an overview. *Research and Review Insights*, 2: 1-7.

Nkukwana, T. 2018. Global poultry production: Current impact and future outlook on the South African poultry industry. *South African Journal of Animal Science*, 48 (5): 869-884.

Nuñez, L. F. N., Parra, S. H. S., De la Torre, D., Catroxo, M. H., Buim, M. R., Chacon, R. V., Ferreira, C. S. A. and Ferreira, A. J. P. 2018. Isolation of avian nephritis virus from chickens showing enteric disorders. *Poultry Science*, 97 (10): 3478-3488.

Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27 (5): 824-834.

Nururrozi, A., Yanuartono, Y., Widyarini, S., Ramandani, D. and Indarjulianto, S. 2020. Clinical and pathological features of aspergillosis due to Aspergillus fumigatus in broilers. *Veterinary World*, 13 (12): 2787.

Oakley, B. B., Lillehoj, H. S., Kogut, M. H., Kim, W. K., Maurer, J. J., Pedroso, A., Lee, M. D., Collett, S. R., Johnson, T. J. and Cox, N. A. 2014. The chicken gastrointestinal microbiome. *Federation of European Microbiological Societies Microbiology Letters*, 360 (2): 100-112.

Ocejo, M., Oporto, B. and Hurtado, A. 2019. 16S rRNA amplicon sequencing characterization of caecal microbiome composition of broilers and free-range slow-growing chickens throughout their productive lifespan. *Scientific Reports*, 9 (1): 2506.

Ogali, I. N., Wamuyu, L. W., Lichoti, J. K., Mungube, E. O., Agwanda, B. and Ommeh, S. C. 2018. Molecular characterization of Newcastle disease virus from backyard poultry farms and live bird markets in Kenya. *International Journal of Microbiology*, 2018.

Ogunbayo, Ayodeji, Mogotsi, M. T., Sondlane, H., Nkwadipo, K. R., Sabiu, S. and Nyaga, M. M. 2022. Metagenomic Analysis of Respiratory RNA Virome of Children with and without Severe Acute Respiratory Infection from the Free State, South Africa during COVID-19 Pandemic

Reveals Higher Diversity and Abundance in Summer Compared with Winter Period. *Viruses*, 14 (11): 2516.

Ogunbayo, A. E., Sabiu, S. and Nyaga, M. M. 2023. Evaluation of extraction and enrichment methods for recovery of respiratory RNA viruses in a metagenomics approach. *Journal of Virological Methods*, 314: 114677.

OIE. 2021. *World organization of animal health: Highly pathogenic avian influenza situation report 2021. World Animal Health Information System of the World Organisation for Animal Health*. Available:https://www.oie.int/app/uploads/2021/12/hpai-situation-report-20211213.pdf (Accessed on 28 January, 2022).

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H. and Wagner, H. 2013. Package 'vegan'. *Community Ecology Package, version*, 2 (9): 1-295.

Orakpoghenor, O., Oladele, S. B. and Abdu, P. A. 2020. Infectious bursal disease: Transmission, pathogenesis, pathology and control-an overview. *World's Poultry Science Journal*, 76 (2): 292-303.

Osamudiamen, F. T., Akanbi, O. A., Zander, S., Oluwayelu, D. O., Bock, C.-T. and Klink, P. 2021. Identification of a Putative Novel Genotype of Avian Hepatitis E Virus from Apparently Healthy Chickens in Southwestern Nigeria. *Viruses*, 13 (6): 954.

Ouoba, J. B., Traore, K. A., Mâ, A. K., Ngazoa, S., Rouamba, H., Doumbia, M., Traore, A. S., Roques, P. and Barro, N. 2019. Distribution and molecular characterization of avian hepatitis E virus (aHEV) in domestic and wild birds in Burkina Faso. *Journal of Veterinary Medicine and Animal Health*, 11 (2): 45-50.

Palazzo, A. F. and Lee, E. S. 2015. Non-coding RNA: what is functional and what is junk? *Frontiers in Genetics*, 6: 2.

Palomino-Tapia, V., Mitevski, D., Inglis, T., Van der Meer, F., Martin, E., Brash, M., Provost, C., Gagnon, C. A. and Abdul-Careem, M. F. 2020. Chicken Astrovirus (CAstV) molecular studies reveal evidence of multiple past recombination events in sequences originated from clinical samples of White Chick Syndrome (WCS) in Western Canada. *Viruses*, 12 (10): 1096.

Park, W. J., Park, B. J., Ahn, H. S., Lee, J. B., Park, S. Y., Song, C. S., Lee, S. W., Yoo, H. S. and Choi, I. S. 2016. Hepatitis E virus as an emerging zoonotic pathogen. *Journal of Veterinary Science*, 17 (1): 1-11.

Patzina-Mehling, C., Falkenhagen, A., Gadicherla, A. K., Grützke, J., Tausch, S. H. and Johne, R. 2020. Whole genome sequence analysis of cell culture-adapted rotavirus A strains from chicken. *Infection, Genetics and Evolution*, 81: 104275.

Payne, L. N. 1998. Retrovirus-induced disease in poultry. *Poultry Science*, 77 (8): 1204-1212.

Peacock, T. P., Penrice-Randal, R., Hiscox, J. A. and Barclay, W. S. 2021. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. *Journal of General Virology*, 102 (4): 001584.

Phan, M. V., Ngo Tri, T., Hong Anh, P., Baker, S., Kellam, P. and Cotten, M. 2018. Identification and characterization of Coronaviridae genomes from Vietnamese bats and rats based on conserved protein domains. *Virus Evolution*, 4 (2): vey035.

Pinheiro, M. S., Dias, J. B., Petrucci, M. P., Travassos, C. E., Mendes, G. S. and Santos, N. 2023. Molecular Characterization of Avian Rotaviruses F and G Detected in Brazilian Poultry Flocks. *Viruses*, 15 (5): 1089.

Popescu, M., Van Belleghem, J. D., Khosravi, A. and Bollyky, P. L. 2021. Bacteriophages and the Immune System. *Annual Review of Virology*, 8 (1): 415-435.

Porter, A. F., Cobbin, J., Li, C.-X., Eden, J.-S. and Holmes, E. C. 2021. Metagenomic identification of viral sequences in laboratory reagents. *Viruses*, 13 (11): 2122.

Pradeep, M., Reddy, M. and Kannaki, T. 2020. Molecular identification and characterization of chicken parvovirus from Indian Chicken and Association with Runting and Stunting Syndrome. *Indian Journal of Animal Research*, 54 (12): 1517-1524.

Principi, N., Silvestri, E. and Esposito, S. 2019. Advantages and limitations of bacteriophages for the treatment of bacterial infections. *Frontiers in Pharmacology*, 10: 513.

Rahman, M. T., Sobur, M. A., Islam, M. S., Ievy, S., Hossain, M. J., El Zowalaty, M. E., Rahman, A. T. and Ashour, H. M. 2020. Zoonotic Diseases: Etiology, Impact, and Control. *Microorganisms*, 8 (9).

Rajeoni, A. H., Ghalyanchilangeroudi, A., Khalesi, B., Madadi, M. S. and Hosseini, H. 2021. The tracheal virome of broiler chickens with respiratory disease complex in Iran: the metagenomics study. *Iranian Journal of Microbiology*, 13 (3): 337.

Raji, A. A., Ideris, A., Bejo, M. H. and Omar, A. R. 2022. Molecular characterization and pathogenicity of novel Malaysian chicken astrovirus isolates. *Avian Pathology*, 51 (1): 51-65.

Raji, A. A. and Omar, A. R. 2022. An insight into the molecular characteristics and associated pathology of chicken astroviruses. *Viruses*, 14 (4): 722.

Rambaut, A. 2009. FigTree version 1. 3.1. *http://tree. bio. ed. ac. uk/software/figtree/*.

Real, R. and Vargas, J. M. 1996. The probabilistic basis of Jaccard's index of similarity. *Systematic Biology*, 45 (3): 380-385.

Roy, D., Tomo, S., Modi, A., Purohit, P. and Sharma, P. 2020. Optimising total RNA quality and quantity by phenol-chloroform extraction method from human visceral adipose tissue: A standardisation study. *MethodsX*, 7: 101113.

Sah, R., Mohanty, A., Reda, A., Siddiq, A., Mohapatra, R. K. and Dhama, K. 2022. Marburg virus re-emerged in 2022: recently detected in Ghana, another zoonotic pathogen coming up amid rising cases of Monkeypox and ongoing COVID-19 pandemic-global health concerns and counteracting measures. *Veterinary Quarterly*, 42 (1): 167-171.

Sajewicz-Krukowska, J. and Domanska-Blicharz, K. 2016. Nearly full-length genome sequence of a novel astrovirus isolated from chickens with 'white chicks' condition. *Archives of Virology*, 161 (9): 2581-2587.

Sakuma, S., Uchida, Y., Kajita, M., Tanikawa, T., Mine, J., Tsunekuni, R. and Saito, T. 2021. First Outbreak of an H5N8 Highly Pathogenic Avian Influenza Virus on a Chicken Farm in Japan in 2020. *Viruses*, 13 (3): 489.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J. and Walker, A. W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BioMed Central Biology*, 12: 1-12.

SAPA. 2019. *South African Poultry Association, Annual report. SAPA Board Chairperson's report*. Available: http://www.sapoultry.co.za/pdf-docs/sapa-annual-report.pdf (Accessed on 12 October 2022).

SAPA. 2021. *Avian flu update: Poultry Bulletin Official Magazine of South African Poultry Association August/ September. Poultry Bulletin, .* Available: http://www.sapoultry.co.za/pdf-bulletin/poultry-bulletin-2021-aug-sep2021.pdf (Accessed on 18 January 2023).

Sarker, S. 2021. Metagenomic detection and characterisation of multiple viruses in apparently healthy Australian Neophema birds. *Scientific Reports*, 11 (1): 20915.

Saxena, S. K., Shukla, S., Kumar, S., Maurya, V. K. and Ansari, S. 2023. Introductory Chapter: Recent Trends in Emerging and Reemerging Viral Contagions–The Day after Today. *Viral Outbreaks-Global Impact and Newer Horizons*.

Shah, J. D., Desai, P. T., Zhang, Y., Scharber, S. K., Baller, J., Xing, Z. S. and Cardona, C. J. 2016. Development of the intestinal RNA virus Community of Healthy Broiler Chickens. *PLoS One*, 11 (2): e0150094.

Shan, T., Yang, S., Wang, H., Wang, H., Zhang, J., Gong, G., Xiao, Y., Yang, J., Wang, X. and Lu, J. 2022. Virome in the cloaca of wild and breeding birds revealed a diversity of significant viruses. *Microbiome*, 10 (1): 1-21.

Shi, Huajuan, Zhou, Y., Jia, E., Pan, M., Bai, Y. and Ge, Q. 2021a. Bias in RNA-seq library preparation: current challenges and solutions. *BioMed Research International*, 2021.

Shi, Ying Tao, Jie Li, B., Shen, X., Cheng, J. and Liu, H. 2021b. The Gut Viral Metagenome Analysis of Domestic Dogs Captures Snapshot of Viral Diversity and Potential Risk of Coronavirus. *Frontiers in Veterinary Science*, 8: 752.

Shi, Yun, H., Zhu, X., Li, W. L., Mak, J. W., Wong, S. H., Zhu, S. T., Guo, S. L., Chan, F. K., Zhang, S. T. and Ng, S. C. 2021c. Modulation of gut microbiota protects against viral respiratory tract infections: a systematic review of animal and clinical studies. *European Journal of Nutrition*, 1-24.

Sood, U., Gupta, V., Kumar, R., Lal, S., Fawcett, D., Rattan, S., Poinern, G. E. J. and Lal, R. 2020. Chicken gut microbiome and human health: past scenarios, current perspectives, and futuristic applications. *Indian Journal of Microbiology*, 60 (1): 2-11.

Stanley, D., Geier, M. S., Chen, H., Hughes, R. J. and Moore, R. J. 2015. Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. *BioMed Central Microbiology*, 15 (1): 1-11.

Stinson, L. F., Keelan, J. A. and Payne, M. S. 2019. Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses. *Letters in Applied Microbiology*, 68 (1): 2-8.

Sun, P., Lin, S., He, S., Zhou, E.-M. and Zhao, Q. 2019. Avian hepatitis E virus: with the trend of genotypes and host expansion. *Frontiers in microbiology*, 10: 1696.

Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 38 (7): 3022-3027.

Theuns, S., Conceição-Neto, N., Zeller, M., Heylen, E., Roukaerts, I. D., Desmarets, L. M., Van Ranst, M., Nauwynck, H. J. and Matthijnssens, J. 2016. Characterization of a genetically heterogeneous porcine rotavirus C, and other viruses present in the fecal virome of a non-diarrheic Belgian piglet. *Infection, Genetics and Evolution*, 43: 135-145.

Theuns, S., Desmarets, L. M., Heylen, E., Zeller, M., Dedeurwaerder, A., Roukaerts, I. D., Van Ranst, M., Matthijnssens, J. and Nauwynck, H. J. 2014. Porcine group A rotaviruses with heterogeneous VP7 and VP4 genotype combinations can be found together with enteric bacteria on Belgian swine farms. *Veterinary Microbiology*, 172 (1-2): 23-34.

Tomlins, S. A., Mehra, R., Rhodes, D. R., Shah, R. B., Rubin, M. A., Bruening, E., Makarov, V. and Chinnaiyan, A. M. 2006. Whole transcriptome amplification for gene expression profiling and development of molecular archives. *Neoplasia*, 8 (2): 153-162.

Truchado, D. A., Llanos-Garrido, A., Oropesa-Olmedo, D. A., Cerrada, B., Cea, P., Moens, M. A., Gomez-Lucia, E., Doménech, A., Milá, B. and Pérez-Tris, J. 2020. Comparative metagenomics of Palearctic and Neotropical avian cloacal viromes reveal geographic bias in virus discovery. *Microorganisms*, 8 (12): 1869.

Turner, D., Kropinski, A. M. and Adriaenssens, E. M. 2021. A roadmap for genome-based phage taxonomy. *Viruses*, 13 (3): 506.

Ullah, K., Mehmood, A., Chen, X., Dar, M. A., Yang, S. and Zhang, W. 2022. Detection and molecular characterization of picobirnaviruses in the wild birds: Identification of a novel picobirnavirus possessing yeast mitochondrial genetic code. *Virus Research*, 308: 198624.

Uyeki, T. M., Hui, D. S., Zambon, M., Wentworth, D. E. and Monto, A. S. 2022. Influenza. *Lancet*, 400 (10353): 693-706.

Van Borm, S., Steensels, M., Mathijs, E., Vandenbussche, F., van den Berg, T. and Lambrecht, B. 2021. Metagenomic sequencing determines complete infectious bronchitis virus (avian Gammacoronavirus) vaccine strain genomes and associated viromes in chicken clinical samples. *Virus Genes*, 57 (6): 529-540.

Veit, W. 2021. Experimental philosophy of medicine and the concepts of health and disease. *Theoretical Medicine and Bioethics*, 42 (3): 169-186.

Verma, V. and Aggarwal, R. K. 2020. A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective. *Social Network Analysis and Mining*, 10: 1-16.

Vibin, J., Chamings, A., Collier, F., Klaassen, M., Nelson, T. M. and Alexandersen, S. 2018. Metagenomics detection and characterisation of viruses in faecal samples from Australian wild birds. *Scientific Reports*, 8 (1): 1-23.

Vibin, J., Chamings, A., Klaassen, M. and Alexandersen, S. 2020a. Metagenomic characterisation of additional and novel avian viruses from Australian wild ducks. *Scientific Reports*, 10 (1): 1-14.

Vibin, J., Chamings, A., Klaassen, M., Bhatta, T. R. and Alexandersen, S. 2020b. Metagenomic characterisation of avian parvoviruses and picornaviruses from Australian wild ducks. *Scientific Reports*, 10 (1): 12800.

Vidaña, B., Busquets, N., Napp, S., Pérez-Ramírez, E., Jiménez-Clavero, M. Á. and Johnson, N. 2020. The role of birds of prey in West Nile virus epidemiology. *Vaccines*, 8 (3): 550.

Videvall, E., Strandh, M., Engelbrecht, A., Cloete, S. and Cornwallis, C. K. 2018. Measuring the gut microbiome in birds: comparison of faecal and cloacal sampling. *Molecular Ecology Resources*, 18 (3): 424-434.

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L. C., Vanden Eynden, E. and Vandamme, A.-M. 2019. Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics*, 35 (5): 871-873.

Virgin, H. W. 2014. The virome in mammalian physiology and disease. *Cell*, 157 (1): 142-150.

Virgin, H. W., Wherry, E. J. and Ahmed, R. 2009. Redefining chronic viral infection. *Cell*, 138 (1): 30-50.

Vishnuraj, M., Devatkal, S., Vaithiyanathan, S., Mendiratta, S., Kumar, R. U., Srinivas, C. and Sowmya, M. 2020. Chicken RNA Integrity Assessment through Capillary Electrophoresis for miRNA Expression Study using RT-qPCR.

Walker, P. J., Tesh, R. B., Guzman, H., Popov, V. L., Travassos da Rosa, A. P., Reyna, M., Nunes, M. R., De Souza, W. M., Contreras-Gutierrez, M. A. and Patroca, S. 2019. Characterization of three novel viruses from the families Nyamiviridae, Orthomyxoviridae, and Peribunyaviridae, isolated from dead birds collected during West Nile Virus Surveillance in Harris County, Texas. *Viruses*, 11 (10): 927.

Wang, He, Y., Yan, X., Sun, Y., Yi, L., Tu, C. and He, B. 2023. Virome Profiling of Chickens with Hepatomegaly Rupture Syndrome Reveals Coinfection of Multiple Viruses. *Viruses*, 15 (6): 1249.

Wang, Shi, M., He, C., Lin, L., Li, H., Gu, Z., Li, M., Gao, Y., Huang, T. and Mo, M. 2019. A novel recombinant avian leukosis virus isolated from gamecocks induced pathogenicity in Three-Yellow chickens: a potential infection source of avian leukosis virus to the commercial chickens. *Poultry Science*, 98 (12): 6497-6504.

Wang, Yan, Cui, X., Chen, X., Yang, S., Ling, Y., Song, Q., Zhu, S., Sun, L., Li, C. and Li, Y. 2020. A recombinant infectious bronchitis virus from a chicken with a spike gene closely related to that of a turkey coronavirus. *Archives of Virology*, 165 (3): 703-707.

Wang, J., Xiao, J., Zhu, Z., Wang, S., Zhang, L., Fan, Z., Yali, D., Hu, Z., Peng, F. and Shen, S. 2022. Diverse viromes in polar regions: a retrospective study of metagenomic data from Antarctic animal feces and Arctic frozen soil in 2012–2014. *Virologica Sinica*.

Wang, Z., Qu, L., Yao, J., Yang, X., Li, G., Zhang, Y., Li, J., Wang, X., Bai, J. and Xu, G. 2013. An EAV-HP insertion in 5′ flanking region of SLCO1B3 causes blue eggshell in the chicken. *PLoS Genetics*, 9 (1): e1003183.

Washington, H. 1984. Diversity, biotic and similarity indices: a review with special relevance to aquatic ecosystems. *Water Research*, 18 (6): 653-694.

Wells, K., Morand, S., Wardeh, M. and Baylis, M. 2020. Distinct spread of DNA and RNA viruses among mammals amid prominent role of domestic species. *Global Ecology and Biogeography*, 29 (3): 470-481.

Whitton, C., Bogueva, D., Marinova, D. and Phillips, C. J. 2021. Are we approaching peak meat consumption? Analysis of meat consumption from 2000 to 2019 in 35 countries and its relationship to gross domestic product. *Animals*, 11 (12): 3466.

WHO. 2023.Cummulative number of avian influenza A (H5N1) reported to WHO, 2003-2023. Available: https://cdn.who.int/media/docs/default-source/influenza/human-animal-interface-risk-assessments/cumulative-number-of--confirmed-human-cases-for-avian-influenza-a(h5n1)-

reported-to-who--2003-2023.pdf?sfvrsn=c6600b55_1&download=true (Accessed on June 12, 2023).

Wickham, H., Chang, W. and Wickham, M. H. 2016. Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version*, 2 (1): 1-189.

Wickham, H., François, R., Henry, L. and Müller, K. 2019. dplyr: A Grammar of Data Manipulation. R package version 0.8. 0.1.13.

Wickham, H. and Wickham, M. H. 2019. Package 'tidyverse'. Available*:http://tidyverse. tidyverse. org*, 1-5.

Wilke, C. O., Wickham, H. and Wilke, M. C. O. 2019. Package 'cowplot'. *Streamlined plot theme and plot annotations for 'ggplot2*,

Wille, M., Eden, J. S., Shi, M., Klaassen, M., Hurt, A. C. and Holmes, E. C. 2018. Virus–virus interactions and host ecology are associated with RNA virome structure in wild birds. *Molecular Ecology*, 27 (24): 5263-5278.

Wille, M., Harvey, E., Shi, M., Gonzalez-Acuña, D., Holmes, E. C. and Hurt, A. C. 2020. Sustained RNA virome diversity in Antarctic penguins and their ticks. *The International Society for Microbial Ecology Journal*, 14 (7): 1768-1782.

Wille, M., Shi, M., Hurt, A. C., Klaassen, M. and Holmes, E. C. 2021. RNA virome abundance and diversity is associated with host age in a bird species. *Virology*, 561: 98-106.

Wragg, D., Mwacharo, J. M., Alcalde, J. A., Wang, C., Han, J.-L., Gongora, J., Gourichon, D., Tixier-Boichard, M. and Hanotte, O. 2013. Endogenous retrovirus EAV-HP linked to blue egg phenotype in Mapuche fowl. *PLoS One*, 8 (8): e71393.

Xia, J.-Y., Liao, J.-Y., Liu, C.-X., Xiong, W.-J. and Xiao, C.-T. 2023. Identification and characterization of a novel avian nephritis virus variant in chickens with enteritis in Hunan province, China. *Archives of Virology*, 168 (2): 1-6.

Yadav, M. P., Singh, R. K. and Malik, Y. S. 2019. Epidemiological Perspective in Managing Viral Diseases in Animals. In: *Recent Advances in Animal Virology*. Springer, 381-407.

Yamanaka, A., Kirino, Y., Fujimoto, S., Ueda, N., Himeji, D., Miura, M., Sudaryatma, P. E., Sato, Y., Tanaka, H. and Mekata, H. 2020. Direct transmission of severe fever with thrombocytopenia

syndrome virus from domestic cat to veterinary personnel. *Emerging Infectious Diseases*, 26 (12): 2994.

Yan, Wei, Sun, C., Zheng, J., Wen, C., Ji, C., Zhang, D., Chen, Y., Hou, Z. and Yang, N. 2019. Efficacy of fecal sampling as a gut proxy in the study of chicken gut microbiota. *Frontiers in Microbiology*, 10: 2126.

Yan, T., Guo, L., Jiang, X., Wang, H., Yao, Z., Zhu, S., Diao, Y. and Tang, Y. 2021. Discovery of a novel recombinant avian orthoreovirus in China. *Veterinary Microbiology*, 260: 109094.

Yin, H.-c., Liu, Z.-d., Zhang, W.-w., Yang, Q.-z., Yu, T.-f. and Jiang, X.-j. 2022a. Chicken intestinal microbiota modulation of resistance to nephropathogenic infectious bronchitis virus infection through IFN-I. *Microbiome*, 10 (1): 1-17.

Yin, L., Zhou, Q., Huang, J., Mai, K., Yan, Z., Wei, X., Shen, H., Li, Q., Chen, L. and Zhou, Q. 2022b. Characterization and pathogenicity of a novel avian nephritis virus isolated in China. *Avian Pathology*, 51 (1): 87-96.

Yin, L., Zhou, Q., Mai, K., Huang, J., Yan, Z., Wei, X., Shen, H., Li, Q., Chen, L. and Zhou, Q. 2021. Isolation and characterization of a novel chicken astrovirus in China. *Poultry Science*, 100 (9): 101363.

Żbikowska, K., Michalczuk, M. and Dolka, B. 2020. The use of bacteriophages in the poultry industry. *Animals*, 10 (5): 872.

Zhang, Yong-Zhen, Chen, Y.-M., Wang, W., Qin, X.-C. and Holmes, E. C. 2019. Expanding the RNA virosphere by unbiased metagenomics. *Annual Review of Virology*, 6: 119-139.

Zhang, Zhao, H., Chi, Z., Cui, Z., Chang, S., Wang, Y. and Zhao, P. 2022. Isolation, identification and genome analysis of an avian hepatitis E virus from white-feathered broilers in China. *Poultry Science*, 101 (3): 101633.

Zhang, Y.-Z., Wu, W.-C., Shi, M. and Holmes, E. C. 2018. The diversity, evolution and origins of vertebrate RNA viruses. *Current Opinion in Virology*, 31: 9-16.

Zhulidov, P. A., Bogdanova, E. A., Shcheglov, A. S., Shagina, I. A., Wagner, L. L., Khazpekov, G. L., Kozhemyako, V. V., Lukyanov, S. A. and Shagin, D. A. 2005. A method for the preparation of normalized cDNA libraries enriched with full-length sequences. *Russian Journal of Bioorganic Chemistry*, 31 (2): 170-177.

Zinter, M., Mayday, M., Ryckman, K., Jelliffe-Pawlowski, L. and DeRisi, J. 2019. Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome*, 7 (1): 1-5.

# APPENDICES

## Appendix 8.1 Ethical clearance

### Appendix 8.1.1 Approval letter from the Animal Research Ethics Committee, University of KwaZulu-Natal, South Africa.

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

21 August 2020

**Dr Saheed Sabiu**
**Dept. of Biotech & Food Tech**
**Durban University of Technology**
**South Africa**

Dear Dr Sabiu,

**Protocol reference number: AREC/012/020**
**Project title:** Deciphering the fecal virome of healthy South African chickens using NGS technique.

**Full Approval – Research Application**

With regard to your revised application received on 30 April 2020, the Animal Research Ethics Committee has accepted the documents submitted and **FULL APPROVAL** for the protocol has been granted.

**Please note: There must be adherence to national and institutional COVID-19 regulations and guidelines at all times.** Researchers will be personally responsible and liable for non-adherence to national regulations. If in doubt, please contact the Research Ethics Chair and/or the University Dean of Research for advice.

**Please note: Any Veterinary and Para-Veterinary procedures must be conducted by a SAVC registered VET or SAVC authorized person.**

**Any alteration/s to the approved research protocol, i.e Title of Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment/modification prior to its implementation. In case you have further queries, please quote the above reference number.**

Please note: Research data should be securely stored in the discipline/department for a period of 5 years.

**The ethical clearance certificate is only valid for a period of one year from the date of issue. Renewal for the study must be applied for before 20 August 2021.**

**Attached to the Approval letter is a template of the Progress Report that is required at the end of the study, or when applying for Renewal (whichever comes first). An Adverse Event Reporting form has also been attached in the event of any unanticipated event involving the animals' health / wellbeing.**

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully

.....

**Dr Sanil D Singh, BVSc, MS, PhD**
**Chair: Animal Research Ethics Committee**

**/kr**
**cc Supervisor:** Dr Phikelelani Ngubane
**cc BRU Manager:** Dr Jaca

**Animal Research Ethics Committee (AREC)**
**Ms Karen Reinertsen (Administrator)**
**Westville Campus, Govan Mbeki Building**
**Postal Address:** Private Bag X54001, Durban 4000
**Telephone:** +27 (0) 31 260 8850 **Facsimile:** +27 (0) 31 260 4609 **Email:** animalethics@ukzn.ac.za
**Website:** http://research.ukzn.ac.za/Research-Ethics/Animal-Ethics.aspx

1910 - 2010
100 YEARS OF ACADEMIC EXCELLENCE

Founding Campuses: Edgewood Howard College Medical School Pietermaritzburg Westville

**Appendix 8.1.2 Research approval of the Department of Agriculture, Land and Rural Development, Republic of South Africa.**

agriculture, land reform & rural development

Department:
Agriculture, Land Reform and Rural Development
REPUBLIC OF SOUTH AFRICA

Directorate Animal Health, Department of Agriculture, Land Reform and Rural Development Private Bag X138, Pretoria 0001
**Enquiries:** Mr Herry Gololo • Tel: +27 12 319 7532 • Fax: +27 12 319 7470 • E-mail: HerryG@dalrrd.gov.za
**Reference:** 12/11/1/5/4 (1531AC)

Dr Saheed Sabiu
Department of Biotechnology and Food Technology
Steve Biko Campus
Durban University of Technology
Tel: 031 333 5330
Email: SabiuS@dut.ac.za

Dear Dr Sabiu,

**RE: PERMISSION TO DO RESEARCH IN TERMS OF SECTION 20 OF THE ANIMAL DISEASES ACT, 1984 (ACT NO 35 OF 1984)**

Your application received on 4 June 2020 requesting permission under Section 20 of the Animal Disease Act, 1984 (Act No. 35 of 1984) to perform a research project or study, refers. I am pleased to inform you that permission is hereby granted to perform the following study, with the following conditions:

**Conditions:**

1. This permission does not relieve the researcher of any responsibility which may be placed on him by any other act of the Republic of South Africa;
2. The research project is approved as per the application form received 4 June 2020 and the correspondence thereafter. Written permission from the Director: Animal Health must be obtained prior to any deviation from the conditions approved for this research project under this Section 20 permit. Please apply in writing to HerryG@dalrrd.gov.za;
3. Naturally voided faecal samples may be collected from healthy chickens in the Durban area, for which a state veterinary letter has been received. It is the researcher's responsibility to obtain confirmation from the responsible state veterinarian of the relevant area that the targeted farm is still not under any restriction for disease control purposes prior to collection of samples. Records must be kept for five years for auditing purposes;

4. The faecal samples may be sent to the Biomedical Resource Unit (BRU), School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, for nucleic acid extraction. Extracted nucleic acid may be outsourced for further sequencing;
5. The following viral family genomes may be targeted as indicated: Viruses of the families *Adenoviridae, Caliciviridae, Circoviridae, Parvoviridae, Picobirnaviridae, Picornaviridae* and *Reoviridae*. As specified, the study will only target these viral families and not specific diseases;
6. Study results are considered as research results and may not be distributed or communicated as diagnostic results;
7. Any detection of suspicion of a controlled or notifiable animals disease must be reported to the responsible state veterinarian and the sample results verified by a SANAS accredited and DAH approved laboratory;
8. All potentially infectious material utilised, collected or generated during the study is to be destroyed at the completion of the study using the specified waste contractor;
9. Records must be kept for five years for auditing purposes;
10. If required, an application for an extension must be made by the responsible researcher at least one month prior to the expiry of this Section 20 permit. Please apply in writing to HerryG@dalrrd.gov.za.

**Title of research/study:** *"Deciphering the fecal virome of healthy South African chickens using next-generation sequencing based molecular characterization technique"*
**Researcher:** Dr Saheed Sabiu
**Institution:** Department of Biotechnology and Food Technology, Durban University of Technology in collaboration with the Biomedical Resource Unit (BRU), School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal
**Our ref Number**: 12/11/1/5/4 (1531AC)
**Your ref**: -
**Expiry date:** December 2022

Kind regards,

**DR. MPHO MAJA**
**DIRECTOR: ANIMAL HEALTH**
**Date:** 2020 -08- 21

**Appendix 8.1.3 Approval of the department of Agriculture and rural development KwaZulu-Natal province, South Africa.**

agriculture
& rural development
Department:
agriculture
& rural development
PROVINCE OF KWAZULU-NATAL

To whom it may concern:

*Letter of no restriction from KwaZulu Natal Department of Agriculture and Rural Development on the study titled deciphering the fecal virome of healthy South African chickens using next Generation Sequencing (NGS) technique*

The department of Agriculture and rural development in KwaZulu Natal hereby grants approval to Dr. S. Saheed to collect faecal samples from chicken farms in Durban, KwaZulu Natal, South Africa.

This study will assist Dr. Saheed in his study to characterize the faecal virome of South Africans chickens; therefore, we have no objection in granting permission for the study.

Kindly note that should the study results have any controlled or notifiable diseases, the department must be notified timeously.

Sincerely

Dr. Cameron Kutwana (*BVSc*)

Director Veterinary Services Area 1

KZN Department of Agriculture and Rural Development

Tel: 035 7806704

Cell: 0829220249

Richards's bay

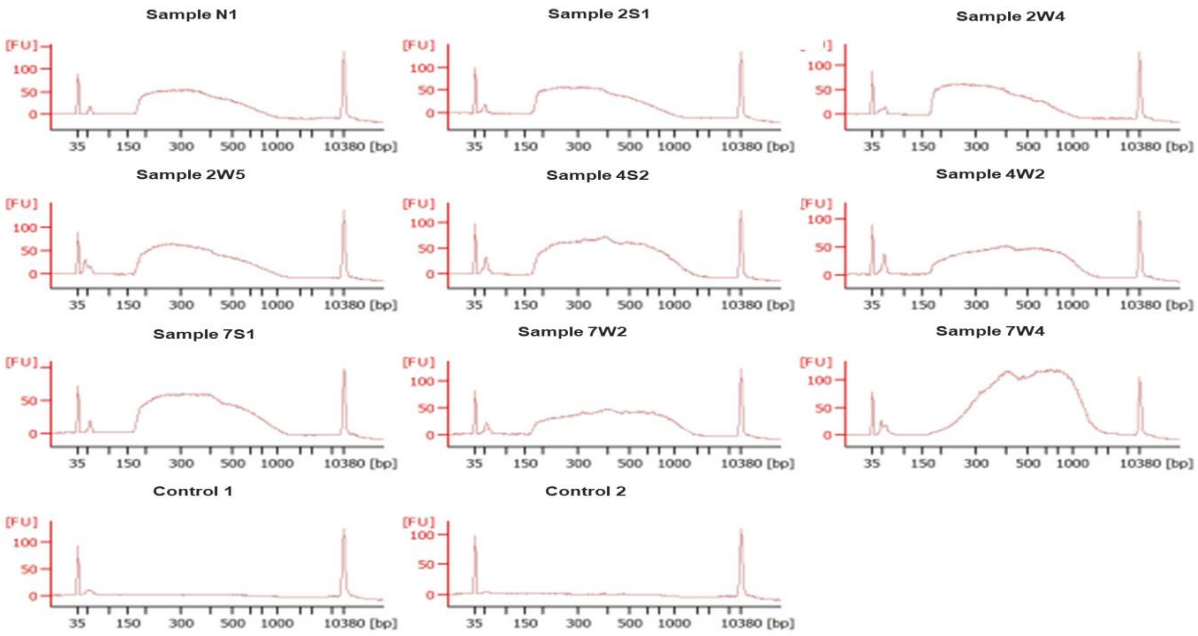
DEPARTMENT : AGRICULTURE & RURAL DEVELOPMENT
PROVINCE OF KWAZULU-NATAL
30 JUN 2020
ETHEKWINI DISTRICT

## Appendix 8.2 Spectrophotometric evaluation of RNA concentrations and 260/280 quality after whole transcriptome amplification.

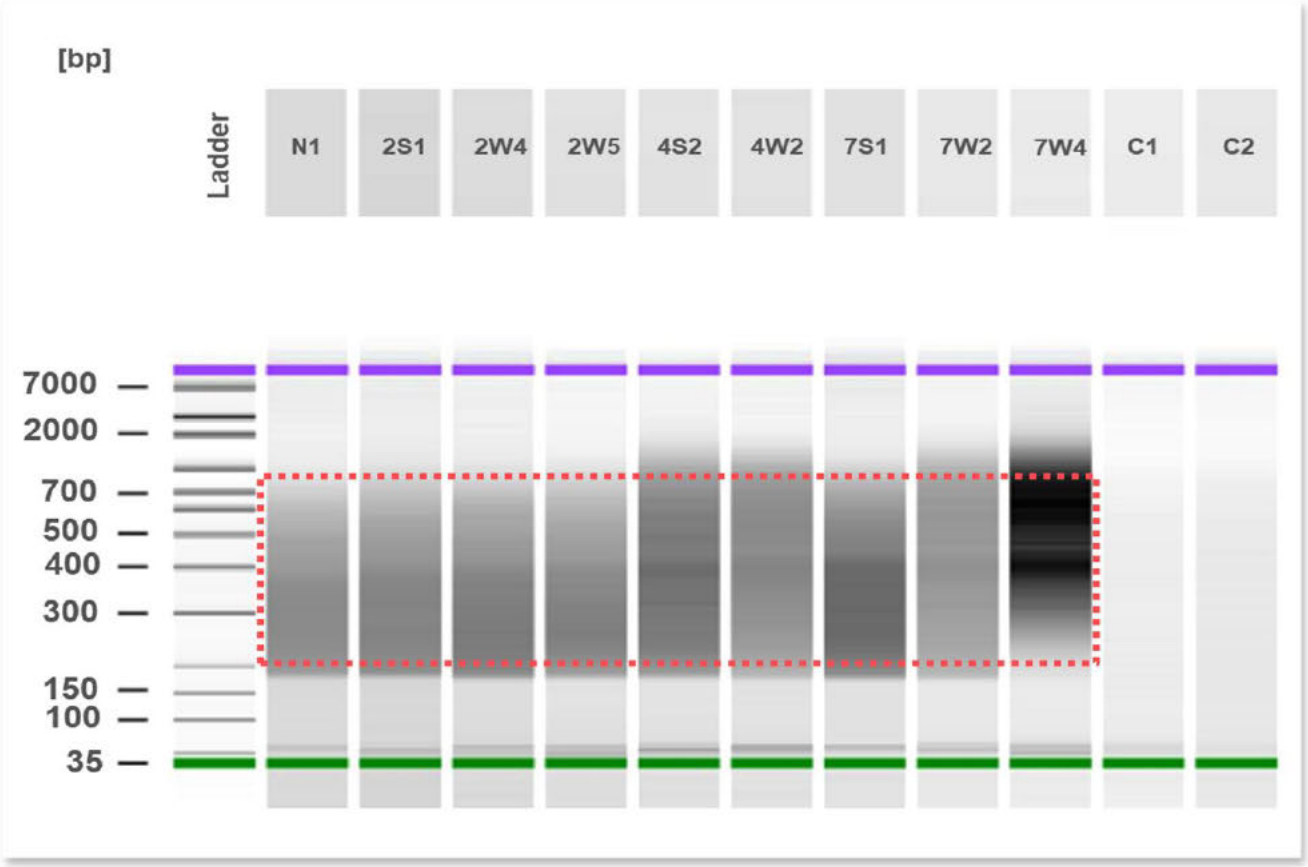
| Sample tag | Sample ID | Concentration of cDNA after amplification (before dilution) (ng/µl) | Quality 260/280 ratio after WTA |
|---|---|---|---|
| A1 | CH/2022/2S1 | 1577.0 | 1.74 |
| A2 | CH/2022/2S2 | 1472.5 | 1.92 |
| A3 | CH/2022/2S3 | 1520.3 | 1.83 |
| A4 | CH/2022/2S4 | 1284.1 | 1.85 |
| A5 | CH/2022/2S5 | 1320.2 | 1.78 |
| A6 | CH/2022/4S1 | 1502.1 | 1.75 |
| A7 | CH/2022/4S2 | 1378.0 | 1.78 |
| A8 | CH/2022/4S3 | 1463.3 | 1.83 |
| A9 | CH/2022/4S4 | 1535.2 | 1.78 |
| A10 | CH/2022/4S5 | 1571.7 | 1.75 |
| A11 | CH/2022/7S1 | 1069.0 | 1.77 |
| A12 | CH/2022/7S2 | 834.0 | 1.71 |
| B1 | CH/2022/4W1 | 748.2 | 1.74 |
| B2 | CH/2022/4W2 | 1548.1 | 1.79 |
| B3 | CH/2022/4W3 | 1533.7 | 1.78 |
| B4 | CH/2022/4W4 | 1539.5 | 1.79 |
| B5 | CH/2022/4W5 | 1487.3 | 1.80 |
| B6 | CH/2022/7W1 | 1331.1 | 1.77 |
| B7 | CH/2022/7W2 | 1571.2 | 1.94 |
| B8 | CH/2022/7W3 | 1245.0 | 1.80 |
| B9 | CH/2022/7W4 | 1103.2 | 1.82 |
| B10 | CH/2022/2W1 | 1539.1 | 1.79 |
| B11 | CH/2022/2W2 | 1551.0 | 1.90 |
| B12 | CH/2022/2W3 | 1470.7 | 1.76 |
| C1 | CH/2022/2W4 | 1535.2 | 1.78 |
| C2 | CH/2022/2W5 | 1553.7 | 1.83 |
| C3 | CH/2022/7W5 | 1313.1 | 1.80 |
| N1 | Control/PBS/A/E | 421.2 | 1.88 |
| N2 | Control/extracton | 146.0 | 1.94 |
| N3 | Control/rRNA | 74.3 | 1.90 |
| N4 | Control/cDNA/WTA | 365.1 | 1.89 |

# Appendix 8.3 Bioanalyzer results of indexed cDNA libraries

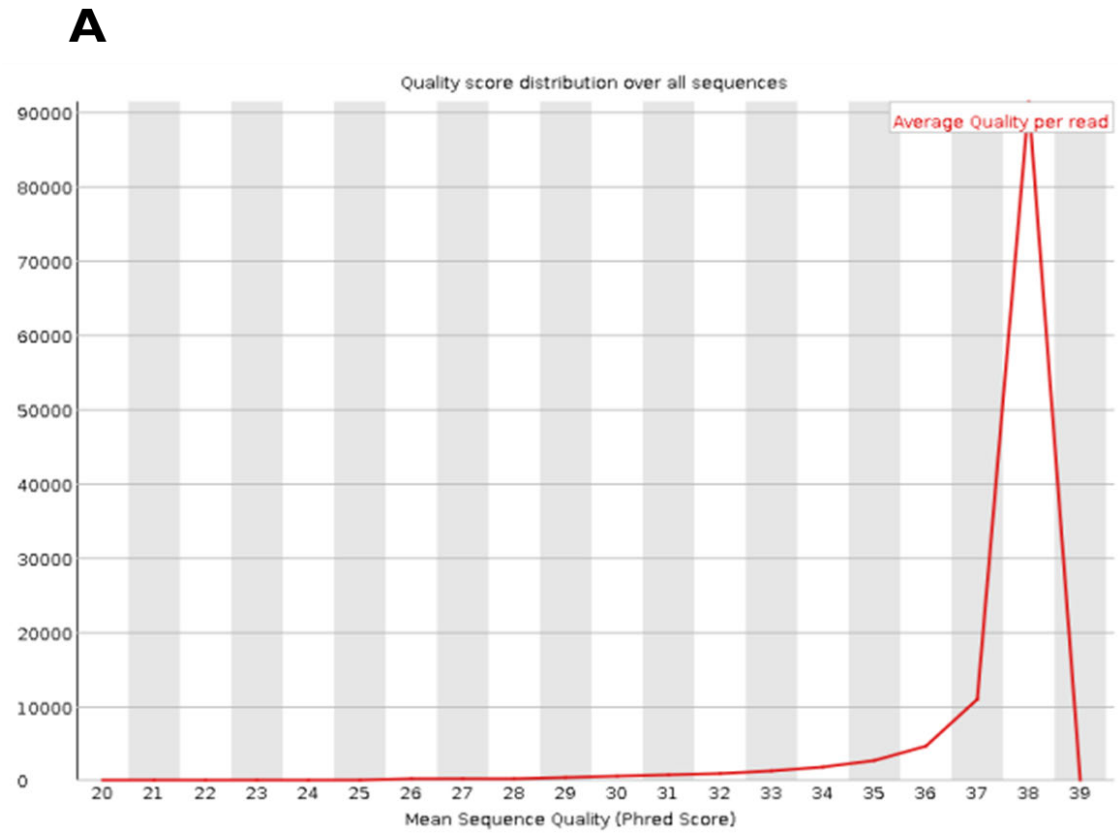
## Appendix 8.3.1 The fragment size distribution electropherogram profile of the barcoded cDNA libraries.

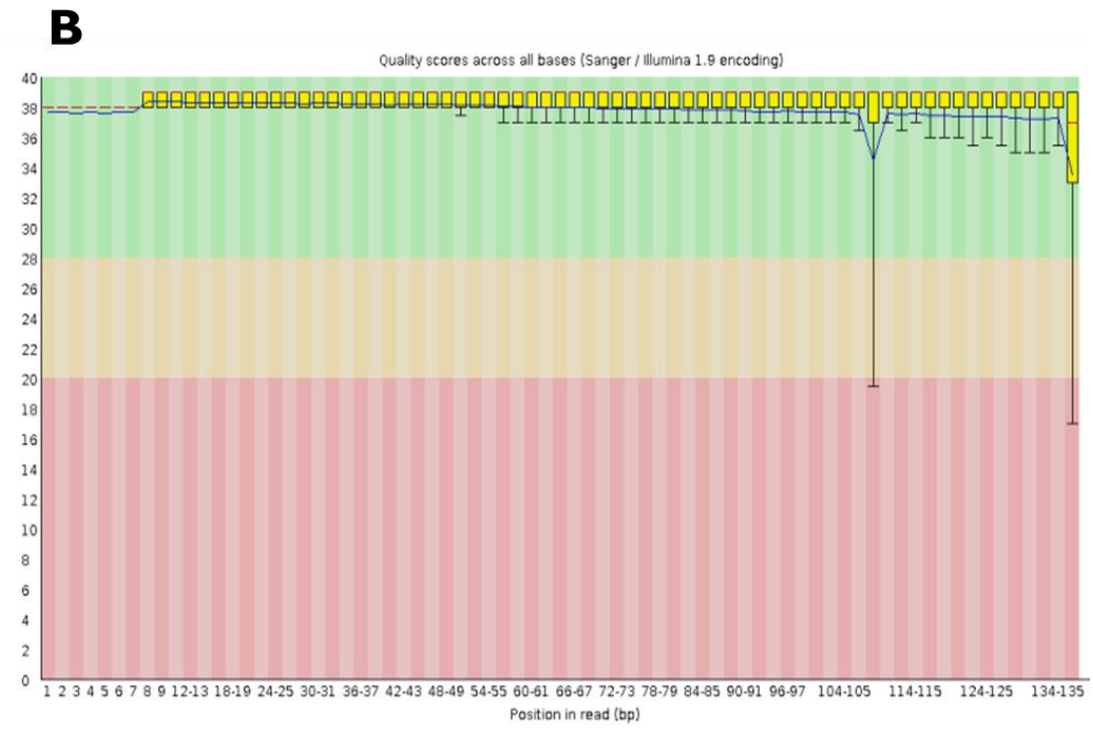


**Appendix 8.3.2 The thick-smear gel image of indexed cDNA libraries size distributions**



**Appendix 8.4 Quality report of sequence distribution of chicken a) Per base sequence quality b) Mean sequency quality (Phred score)**



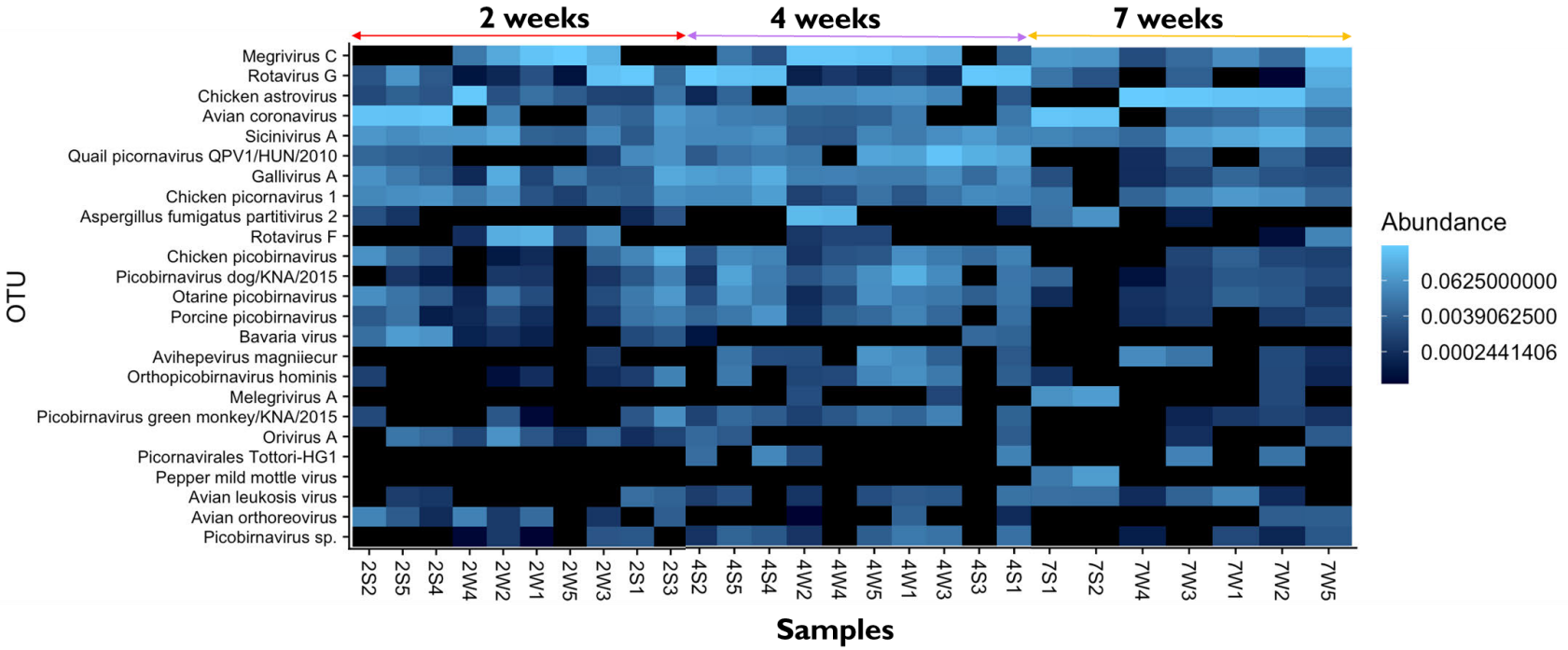
# Appendix 8.5 Taxonomic classification of chicken viral contigs from faeces based on respective 27 samples utilized.

| Sample ID | | 2S1 | 2S2 | 2S3 | 2S4 | 2S5 | 4S1 | 4S2 | 4S3 | 4S4 | 4S5 | 7S1 | 7S2 | 2W1 | 2W2 | 2W3 | 2W4 | 2W5 | 4W1 | 4W2 | 4W3 | 4W4 | 4W5 | 7W1 | 7W2 | 7W3 | 7W4 | 7W5 | Sum of contigs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sum of contigs of samples | | 220 | 164 | 196 | 146 | 154 | 232 | 155 | 94 | 156 | 198 | 95 | 113 | 124 | 150 | 174 | 135 | 75 | 218 | 181 | 122 | 141 | 139 | 182 | 207 | 240 | 143 | 174 | 4328 |
| *Avian coronavirus* | *Coronaviridae* | 55 | 6 | 42 | 3 | 5 | 47 | 13 | 0 | 37 | 44 | 0 | 41 | 0 | 9 | 31 | 0 | 0 | 55 | 53 | 0 | 27 | 47 | 31 | 18 | 38 | 0 | 44 | 646 |
| *Avisivirus B* | *Picornaviridae* | 10 | 0 | 0 | 0 | 0 | 5 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 9 | 9 | 14 | 0 | 65 |
| *Chicken picornavirus 1* | *Picornaviridae* | 13 | 15 | 9 | 9 | 13 | 10 | 14 | 10 | 6 | 15 | 0 | 0 | 8 | 5 | 7 | 5 | 9 | 16 | 8 | 9 | 13 | 13 | 14 | 8 | 9 | 14 | 0 | 252 |
| *Gallivirus A* | *Picornaviridae* | 20 | 11 | 6 | 14 | 12 | 15 | 10 | 6 | 12 | 9 | 9 | 0 | 14 | 8 | 14 | 9 | 8 | 15 | 8 | 10 | 11 | 13 | 8 | 13 | 19 | 8 | 20 | 302 |
| *Megrivirus A* | *Picornaviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 7 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 26 |
| *Megrivirus C* | *Picornaviridae* | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 19 | 15 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 14 | 6 | 13 | 10 | 12 | 27 | 23 | 24 | 54 | 2 | 245 |
| *Orivirus A* | *Picornaviridae* | 11 | 0 | 7 | 5 | 4 | 9 | 6 | 0 | 0 | 9 | 0 | 0 | 2 | 1 | 2 | 13 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 7 | 97 |
| *Quail picornavirus QPV1/HUN/2010* | *Picornaviridae* | 3 | 13 | 5 | 11 | 13 | 3 | 10 | 12 | 13 | 15 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 4 | 1 | 1 | 0 | 1 | 0 | 4 | 12 | 9 | 12 | 157 |
| *Sicinivirus A* | *Picornaviridae* | 13 | 15 | 20 | 13 | 9 | 21 | 17 | 15 | 9 | 19 | 4 | 12 | 9 | 4 | 8 | 3 | 9 | 20 | 16 | 16 | 22 | 19 | 12 | 2 | 9 | 14 | 11 | 341 |
| *Chicken astrovirus* | *Astroviridae* | 12 | 4 | 13 | 9 | 10 | 16 | 9 | 0 | 0 | 13 | 0 | 0 | 2 | 10 | 13 | 1 | 5 | 11 | 6 | 10 | 17 | 4 | 2 | 1 | 1 | 2 | 2 | 173 |
| *Bavaria virus* | *Caliciviridae* | 18 | 15 | 12 | 1 | 1 | 18 | 9 | 9 | 0 | 0 | 0 | 0 | 16 | 19 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 131 |
| *Avian leukosis virus* | *Retroviridae* | 8 | 0 | 8 | 9 | 5 | 7 | 10 | 0 | 0 | 11 | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 7 | 6 | 8 | 0 | 1 | 11 | 7 | 10 | 5 | 0 | 121 |
| *Rous sarcoma virus* | *Retroviridae* | 17 | 8 | 12 | 10 | 10 | 8 | 13 | 0 | 0 | 13 | 2 | 0 | 10 | 0 | 8 | 0 | 0 | 10 | 11 | 11 | 0 | 0 | 11 | 12 | 11 | 6 | 0 | 183 |
| *Avihepevirus magniiecur* | *Hepeviridae* | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 4 | 10 | 15 | 0 | 7 | 0 | 12 | 5 | 2 | 14 | 107 |
| *Avian orthoreovirus* | *Reoviridae* | 0 | 44 | 21 | 8 | 32 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 29 | 19 | 21 | 0 | 30 | 1 | 0 | 0 | 0 | 0 | 28 | 0 | 0 | 34 | 286 |
| *Rotavirus F* | *Reoviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 11 | 13 | 22 | 21 | 0 | 22 | 0 | 7 | 0 | 0 | 3 | 0 | 0 | 14 | 124 |
| *Rotavirus G* | *Reoviridae* | 12 | 13 | 23 | 28 | 18 | 13 | 13 | 35 | 17 | 12 | 23 | 7 | 18 | 26 | 11 | 3 | 9 | 13 | 2 | 7 | 3 | 0 | 0 | 2 | 34 | 0 | 10 | 352 |
| *Infectious bursal disease virus* | *Birnaviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 |
| *Chicken picobirnavirus* | *Picobirnaviridae* | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 5 | 2 | 1 | 8 | 0 | 4 | 3 | 1 | 0 | 0 | 6 | 4 | 1 | 6 | 4 | 2 | 1 | 3 | 0 | 0 | 59 |
| *Orthopicobirnavirus hominis* | *Picobirnaviridae* | 2 | 2 | 1 | 0 | 0 | 3 | 0 | 0 | 0 | 1 | 2 | 0 | 1 | 12 | 3 | 0 | 0 | 1 | 4 | 4 | 4 | 4 | 0 | 2 | 0 | 0 | 0 | 46 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Otarine picobirnavirus* | *Picobirnaviridae* | 1 | 1 | 4 | 3 | 1 | 3 | 2 | 2 | 0 | 2 | 5 | 0 | 2 | 1 | 4 | 1 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 3 | 0 | 48 |
| *Picobirnavirus dog/KNA/2015* | *Picobirnaviridae* | 3 | 0 | 4 | 9 | 1 | 2 | 8 | 0 | 0 | 1 | 4 | 0 | 2 | 3 | 4 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 6 | 1 | 1 | 4 | 0 | 59 |
| *Picobirnavirus green monkey/KNA/20* | *Picobirnaviridae* | 4 | 4 | 2 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 3 | 2 | 5 | 2 | 1 | 3 | 0 | 0 | 36 |
| *Picobirnavirus sp.* | *Picobirnaviridae* | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 26 | 2 | 0 | 0 | 2 | 3 | 2 | 9 | 0 | 3 | 3 | 1 | 0 | 3 | 3 | 3 | 0 | 2 | 0 | 67 |
| *Porcine picobirnavirus* | *Picobirnaviridae* | 3 | 7 | 3 | 2 | 2 | 3 | 4 | 0 | 6 | 4 | 0 | 0 | 4 | 4 | 4 | 3 | 0 | 3 | 2 | 3 | 7 | 2 | 0 | 6 | 2 | 2 | 0 | 76 |
| *Escherichia virus DE3* | *Siphoviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 0 | 0 | 22 |
| *Lambdavirus lvO276* | *Siphoviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 |
| *Aspergillus fumigatus partitivirus 2* | *Partitiviridae* | 2 | 3 | 2 | 0 | 2 | 3 | 0 | 0 | 0 | 0 | 5 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 4 | 0 | 0 | 0 | 2 | 0 | 0 | 31 |
| *Botryotinia fuckeliana partitivirus 1* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| *Cryptosporidium parvum virus 1* | *Partitiviridae* | 4 | 2 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| *Fusarium poae virus 1* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| *Penicillium aurantiogriseum partitivirus 1* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| *Penicillium stoloniferum virus F* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| *Pythium nunn virus 1* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| *Sclerotinia sclerotiorum partitivirus S* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Ustilaginoidea virens partitivirus 2* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| *Verticillium dahliae partitivirus 1* | *Partitiviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Festuca pratensis amalgavirus 1* | *Amalgaviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 5 |
| *Tomato mosaic virus* | *Potyviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Pepper mild mottle virus* | *Virgaviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 |
| *Tobacco mild green mosaic virus* | *Virgaviridae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 13 | 6 | 4 | 0 | 43 |
| *Eimeria tenella RNA virus 1* | *Totiviridae* | 0 | 0 | 0 | 9 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 5 | 8 | 0 | 0 | 46 |
| *Scheffersomyces segobiensis virus L* | *Totiviridae* | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| *Picornavirales Tottori-HG1* | *Unclassified virus* | 0 | 0 | 0 | 0 | 0 | 4 | 5 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 5 | 2 | 0 | 0 | 33 |
| *Hubei orthoptera virus 1* | *Unclassified virus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 11 | 0 | 0 | 0 | 26 |
| *Hubei picorna-like virus 24* | *Unclassified virus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 5 |
| *Penicillium aurantiogriseum partiti-like virus* | *Unclassified virus* | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| *Wuhan insect virus 22* | *Unclassified virus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |

**Appendix 8.6 The overall abundance of RNA viral species from 27 chicken study samples at 2, 4 and 7 weeks**



## Appendix 8.7 Alpha diversity Statistical significance

**A) Alpha diversity : Kruskal-Wallis H rank sum test** – Differences in independent sample (within group)

**1. Age**

Data: Shannon by age

Kruskal-Wallis chi-squared =3.849

Degrees of freedom (Df) = 2

**P-value = 0.146**

**2. Season**

Data: Shannon by age

Kruskal-Wallis chi-squared = 1.3314

Degrees of freedom (Df) = 1

**P-value = 0.2416**

**Null hypothesis:** Independent sample in a group all have the same abundance and diversity hence come from the same population **(no difference/effect within sample in a group)**

**Alternate hypothesis:** At least more than one of the Independent samples in a group does not have the same abundance and as other samples and hence viral abundance and diversity is not equal ≠ within group **(There is a difference within samples in a group)**

**P value greater than 0.05, Thus there is no statistically significant difference in the observed groups, and we failed to reject the null hypothesis.**

**B) Dunn's Q test of multiple comparisons using rank sum: Equivalent post hoc test of Bonferroni**

**I: Age**

| Group pairs | Mean rank difference | P-value |
|---|---|---|
| 4 weeks-2 weeks | 6.0000000 | 0.2729 |
| 7 weeks-2 weeks | - 0.474286 | 1.0000 |
| 7 weeks- 4 weeks | -6.4714286 | 0.2941 |

**11: Season**

| Group pairs | Mean rank difference | P-value |
|---|---|---|
| Winter-Summer | -3.6 | 0.2416 |

**Significant codes**: 0 ' *** ' ; 0.001 ' ** ' ; 0.01 ' * ' ; 0.05 ' . ' ; 0.1 ' ' ; 1

## Appendix 8.8 Beta diversity statistical significance

### Beta Diversity statistical significance

**Permutation test for Adonis under reduced model (1000)**

**a) Age**

| | Df | Sum of squares | R squared | F | P(>F) |
|---|---|---|---|---|---|
| Age | 2 | 1.4469 | 0.17085 | 2.4727 | 0.01099* |
| Residual | 24 | 7.0216 | 0.82915 | | |
| Total | 26 | 8.4684 | 1.00000 | | |

**b) Season**

| | Df | Sum of squares | R squared | F | P(>F) |
|---|---|---|---|---|---|
| Season | 1 | 2.0415 | 0.24107 | 7.941 | 0.000999*** |
| Residual | 25 | 6.4270 | 0.75893 | | |
| Total | 26 | 8.4684 | 1.00000 | | |

**Significant codes**: 0 ‘***’ ; 0.001 ‘**’ ; 0.01 ‘*’ ; 0.05 ‘.’ ; 0.1 ‘ ’ ; 1

**c) Age and Season**

| | Df | Sum of squares | R squared | F | P(>F) |
|---|---|---|---|---|---|
| Age | 2 | 1.4469 | 0.17085 | 4.2263 | 0.000999*** |
| Season | 1 | 1.9763 | 0.23337 | 11.5455 | 0.000999*** |
| Age :Season | 2 | 1.4507 | 0.17131 | 4.2377 | 0.000999*** |
| Residual | 21 | 3.5946 | 0.42447 | | |
| Total | 26 | 8.4684 | 1.00000 | | |