Towards *in silico* IBV vaccine design: defining the role of polymorphism in viral attenuation

Maciej Grywalski

BSc (Hons) Microbiology

Submitted in fulfilment of the requirements for the Degree of Master of Research.

College of Medical, Veterinary & Life Science

University of Glasgow

July 2016

Abstract

The gammacoronavirus, Infectious Bronchitis Virus (IBV), is a respiratory pathogen of chickens. IBV is a constant threat to poultry production as established vaccines are often ineffective against emerging strains. This requires constant and rapid vaccine production by a process of viral attenuation by egg passage, but the essential forces leading to attenuation in the virus have not yet been characterised. Knowledge of these factors will lead to the development of more effective, rationally attenuated, live vaccines and reduction of the mortality and morbidity caused by this pathogen.

M41 CK strain was egg passaged four times many years ago at Houghton Poultry Research Station and stored as M41-CK EP4 (stock virus at The Pirbright Institute since 1992). It was the first egg passage to have its genome pyrosequenced and was therefore used as the baseline reference. The overall aim of this project was to analyse deep sequence data obtained from four IBV isolates (called A, A1, C and D) each originating from the common M41-CK EP4 (ep4) and independently passaged multiple times in embryonated chicken eggs (figure 1.1). Highly polymorphic encoding regions of the IBV genome were then identified which are likely involved in the attenuation process through the formation of independent SNPs and/or SNP clusters. This was then used to direct targeted investigation of SNPs during the attenuation process of the four IBV passages.

A previously generated deep sequence dataset was used as a preliminary map of attenuation for one virulent strain of IBV. This investigation showed the nucleocapsid and spike as two highly polymorphic encoding regions within the IBV genome with the highest proportion of SNPs compared to encoding region size. This analysis then led to more focussed studies of the nucleocapsid and spike encoding region with the ultimate aim of mapping key attenuating regions and nucleotide positions.

The 454 pyrosequencing data and further investigation of nucleocapsid and spike encoding regions have identified the SNPs present at the same nucleotide positions within analysed A, A1, C and D isolates. These SNPs probably play a crucial role in viral attenuation and universal vaccine production but it is not clear if independent SNPs are also involved in loss of virulence. The majority of SNPs accumulated at different nucleotide positions without further continuation in Sanger sequenced egg passages presenting S2 subunit (spike) and nucleocapsid as polymorphic encoding regions which in nature remain highly conserved.

Table of contents

List of tables

List of figures

Figure 3.9. The bar chart illustrates the accumulation of SNPs found at >50 % frequency in ep110 also visible across Sanger sequenced egg passages. The bars were presented as follow: blue = S1 subunit, red = S2 subunit and green = nucleocapsid. The table below bar chart includes the exact number of SNPs identified in each egg passage of analysed encoding region.

Acknowledgements

I would like to thank Dr Richard Orton and Dr Dan Haydon, my primary supervisors at University of Glasgow. I wish to also thank Dr Pamela Greenwell, Prof Paul Britton and Dr Erica Bickerton for their advice, patience and support during the project.

I would like to thank members of Immunogenetics Group for their support, encouragement and friendship; Dr Clare Grant, Dr John Schwartz, Dr Mark Gibson and Dr Alasdair Allan.

I would also like to thank other staff at The Pirbright Institute; in particular Dr Lynda Moore and Mrs Yvonne Walsh for advice in scientific career planning.

Author`s Declaration

I declare that the work include in this dissertation was continued in accordance with the requirements of the University`s Regulation and according to the Code of Practice for Research Degree Programmes. This dissertation has not been submitted for any other academic award. I also declare that the work is the candidate`s own risk except of where indicated by specific reference in the text. The research project was perform in collaboration with, or with assistance of, others, is indicated as such.

SIGNED: Macej Grywalski **DATE:** 04/12/2016

Definitions/ abbreviations

cDNA	complementary DNA
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP	nucleoside triphosphate
dTTP	deoxythymidine triphosphate
Ep	egg passage
ep #	egg passage number
Freq	Frequency
freq %	frequency percentage
NS	non-synonymous
Nucl. #	nucleotide number
PCR	polymerase chain reaction
RNA	ribonucleic acid
S	synonymous
SNP	single nucleotide polymorphism
Tm	melting temperature

Chapter 1: Introduction

1.1 Infectious Bronchitis Virus (IBV) and other avian pathogens

Research studies have revealed various pathogens to be responsible for initiation of respiratory diseases in poultry industry. A variety of viruses, bacteria and fungi might be involved in infection of the flock respiratory tract making them a significant component of the overall disease incidence in poultry (Glisson, 1998). Interestingly, many cases have shown bacteria as the defined primary component of respiratory disease such as infectious coryza and fowl cholera in chickens. In other cases, bacteria colonizing the respiratory tract, such as *Escherichia coli*, occur as secondary pathogens after viral involvement following Infectious Bronchitis virus infection (Page, 1961). Environmental factors were also proven to have an influence on respiratory disease in poultry. It is well known that disease severity is largely increased in the winter months characterised by low temperatures. The research studies have shown that chickens kept in lower temperatures (7-10 ^{o}C) are more susceptible to respiratory tract diseases rather than chickens maintained at higher temperatures (>24 ^{o}C). Other factors responsible for the large number of viral/bacterial infection in chickens are ventilation, humidity, atmospheric ammonia and the amount of dust in poultry shelters. (Kleven, 1998).

The viral diseases of the flock respiratory system have a huge impact in mortality across the poultry industry due to changes in viral antigenic properties such as Infectious Bronchitis Virus (IBV), an avian pathogen responsible for Infectious Bronchitis (IB) disease. This disease, characterised by respiratory distress, including nasal discharge, coughing and sneezing, results in reduced egg production and severe kidney damage leading to a high mortality rate (Selvarani, 2014). IBV is a Coronavirus, a single-stranded positive-sense enveloped RNA virus, discovered by Schalk and Hawn in 1931 (Schalk, 1931), with a wide distribution in many parts of the globe. The most common strains are Arkansas, Connecticut and Massachusetts (North America), Q1 type (South America), Massachusetts-like type (Europe and Africa) and QX and Q1 IBV in Asia (reviewed by Jackwood, 2012). Although there are effective live attenuated vaccines for most strains, the immune response induced by a vaccine for one strain is often not effective against another strain. The SNPs, insertions and recombination in the IBV genome create serotype specific epitopes which vary between IBV strains, largely due to amino acid differences in the S1 encoding region (reviewed by (Cook et al., 2012)). In addition, recombination in the S1

subunit of two IBV strains can also cause reduced vaccine protection (Jia et al., 1996). This was clearly illustrated in Sweden where Massachusetts-type vaccines did not protect against the Swedish IBV isolates present in the field. The subsequent investigation confirmed the emergence of a new IBV strain evolved from two Australian subgroups (Farsang et al., 2002). This N1/08 strain emerged directly from the recombination of S1 subunit of Australian subgroup 2 (N1/88) and subgroup 1 (N1/03) (Hewson et al., 2014). This can make controlling outbreaks difficult, compounded by the fact that some IBV strains circulate together in small geographic regions and that the risk of infection also depends on the viral dose transmitted, virus strain, age of the chicks and many other environmental factors (Ignjatovic et al., 2002). However, research has shown partial IB disease control using the first accepted IBV live vaccines in the UK, H52 and H120 strains (Dutch field strain of IBV) (MacDonald and McMartin, 1976). Improvements in genetic and bioinformatics techniques combined with specific safety requirements for live IB vaccine production gives an opportunity for the design of a new universal live vaccine - genomic IBV clone (Geerligs et al., 2011, Zanaty, 2013).

Apart from IBV characterised with constant newly strands emerging, Infectious Laryngotracheitis virus (ILTV), Avian influenza virus (AIV) and Newcastle Disease virus (NDV) are also classified as the most common viral agents causing issues in the respiratory tract of birds (Villegas, 1998). Interestingly, ILTV was the first major avian pathogen ever found (1932) for which there has been sufficient vaccine development (reviewed by Bagust and Johnson, 1995). The respiratory disease is associated with haemorrhagic tracheitis, and expectoration of blood-stained mucus in young layers resulting in fast immune response of respiratory tract and very low mortalities (Ayala *et al.*, 2014). In countries with modern intensive poultry industries, ILTV in layer flocks is no longer recognised as a cause of significant economic losses with single outbreaks occurring occasionally (Reddy *et al.*, 2014). However, research studies have confirmed several highly pathogenic strains of AIV resulting as fowl plague in the poultry industry (Keawcharoen *et al.,* 2008). The highly virulent H5N1 type A AIV (Suarez *et al.,* 1998) and its subtype H9N2 which possessed H5N1-like internal genomes caused a significant loss of layer fowl in Asia, Middle East, Europe and Africa (Guan *et al.,* 2000). The subtype H7N7 of type A AIV caused severe respiratory disease in the poultry industry followed by diarrhoea, swollen heads and combs resulting in culling of large chicken population in Netherlands (Fouchier *et al.,* 2004). In general NDV is much less pathogenic that AIV with a main division into Velogenic,

Viscerotropic and Neutropic strains (Bang, *et al.* 1974). Typically, the number of infected birds increases during the winter season compared to a very few cases analysed in summer time (Wajid *et al.*, 2015). Interestingly, Velogenic and Neutropic types were not isolated in USA for many years but reported across Europe, Africa, Asia, Australia and South America (Shittu *et al.*, 2016).

1.2 Structure and replication of airborne avian pathogens

The replication of avian airborne pathogens has a significant impact on disease progression in epithelial cells of the respiratory tract. IBV is a single stranded virus with a 27.6 kb RNA genome which exploits multiple host proteins to achieve successful cell entry. In fact all RNA viruses consist of only a small number of encoding regions so the investigation of them plays a crucial role in understanding the mechanisms responsible for the virus replication cycle inside the host cell, egress and immune evasion (Smith and Denison, 2012, Kong *et al.*, 2010).

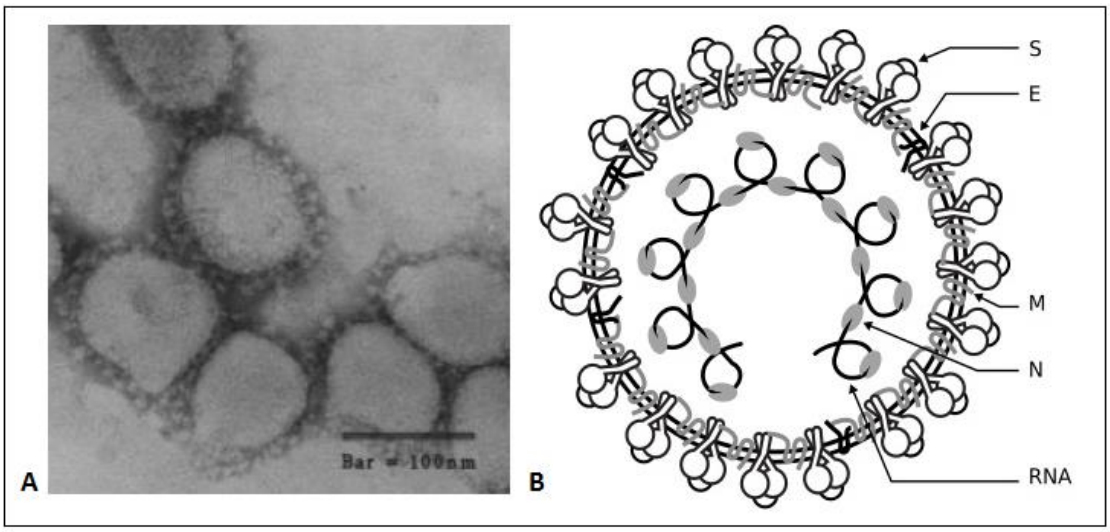


Figure 1.1. The picture presents: A= the image of IBV particles under the microscope, B= the general structure of IBV including spike protein (S), envelope protein (E), membrane protein (M) and nucleocapsid (N) with RNA genome inside (Unknown photographer, 2016).

The figure 1.1 (Unknown photographer, 2016) shows that an IBV particle is comprised of four structural proteins; the spike (S), membrane (M) and small envelope (E) proteins are present in the viral membrane which surrounds the phosphorylated nucleocapsid protein (N) which interacts with the IBV genomic RNA (Fan *et al*., 2005). In addition, IBV also encodes accessory proteins; 3a, 3b, 4b, 5a and 5b with unknown function and two large open reading frames encoding proteins called Replicase 1a and 1b (figure 1.2). IBV assembly is based on association of the nucleocapsid with the M protein and structural proteins at the ER-Golgi intermediate region (ERGIC) (Klumperman *et al*., 1994). The fully formed viral particles bud off from the Golgi apparatus and exit the cell via exocytosis (Tooze *et al*., 1987).

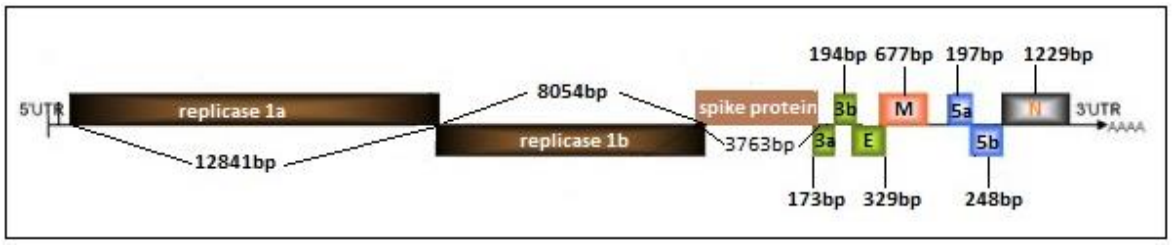


Figure 1.2. Schematic drawing of IBV genome including all gene sizes expressed in base pairs (bp). Picture includes shorter names of genes as follow: 3a= 3a protein, 3b=3b protein, E=envelope protein, M=membrane protein, 5a=5a protein, 5b= 5b protein and N=nucleocapsid.

The diameter of the virus particle is approximately 120nm combined with 20nm glycosylated spike (S) proteins which form large surface projections. The S glycoprotein is constructed with two subunits called amino-terminal S1 and carboxyl terminal S2. The S glycoproteins are assembled with the M proteins into the virion membranes through non-covalent interactions and determine the ability of the virus to attach to the host cells (McBride and Machamer, 2010). The mechanism of viral membrane fusion to the host cell membrane suggests that the specific structure of the S1 subunit spike protein induces substantial changes in the S glycoprotein, allowing the nucleocapsid to be released into the cytoplasm (Tay *et al*., 2012, Wicht *et al*., 2014). As the S1 subunit of glycoprotein is responsible for binding to the host receptor, any SNPs in nucleotide sequence might have an impact on virulence, tissue tropism and persistence in a host (Luo and Weiss, 1998, Koch *et al*., 1990).

Interestingly, IBV is not the only avian airborne pathogen having impact on disease progression in epithelial cells of the respiratory tract. A single stranded, negative sense, RNA Newcastle disease virus causes high mortality rates due to its fast cleavage of F protein responsible for the level of virulence and initiation of infection, followed by replication and spread of disease (Dortmans *et al.,* 2010 and Dortmans *et al.,* 2011). Interestingly NDVs have an influence on the impact of the disease caused by AIV which is also a single stranded, negative sense RNA virus. The AIV virulence and the severity of the disease in chickens is determined by the possible interactions and outcomes of infection (disease and virus shedding) when AIV and NDV co-infect bird populations in the farm (Costa-Hurtado *et al.,* 2015 and Liu *et al.,* 2015). One of the promising ideas is recognition of mechanisms responsible for of host-virus interactions to combat the virus (Reedy *et al.,* 2014). Therefore, novel control strategies are urgently needed to limit outbreaks due to side effects of currently existing therapeutic strategies especially in the rural areas of developing countries.

1.3 Outbreaks in poultry industry - failure of the vaccination programme

Acute respiratory tract infections caused by viral pathogens are the major reason of outbreaks in poultry industry especially in rural areas. The complex form of aetiology in respiratory diseases involves often more than one pathogen at the same time resulting in uncontrolled spread of an outbreak (Roussan *et al.,* 2008). Interestingly, AIV, IBV, NDV and ILTV are classified as major respiratory pathogens because the outbreak can be caused by a viral agent in association with different strains of the same virus. Therefore acute respiratory infections are the leading cause of morbidity, mortality and economic loss in the poultry industry (Mounts *et al.,* 1999). Similar to IBV strains constantly emerging, special attention should be paid to type A AIV known to be responsible for several outbreaks across the world. This avian pathogen causes high a mortality rate in the poultry industry due to its ability to cause infection with subtype H5 and H7 at this same time. The research studies revealed 12 outbreaks since 1959 (eight since 1990) caused by type A AIV in chickens (reviewed by Alexander, 2000) including the fowl plague in Netherlands in 2003 (Koopmans *et al.,* 2004).

Modern live vaccines available against IBV and other avian pathogens present different level of cross protection. Therefore, difficulties in outbreak control were also reported among the poultry industry in Venezuela and China caused by genotype VII NDV. Further investigation has shown large genetic diversity within NDV isolates including 9 class 1 NDV and 10 different lineages of class II NDV making the spread of avian disease even more difficult to stop. In addition, the efficiency of the vaccine protocols against present genotype VII virus in Venezuela and China was compromised by field-associated factors including immunosuppression. A large numbers of subtypes were also reported during the outbreaks of ILTV in Australia during the period of three years (Perozo *et al.,* 2012). The recent losses in the Australian poultry industry have shown that genetic differences in viral pathogens are responsible for significant mortality in domestic chickens (Lee *et al.,* 2012). Four new classes were identified and added to already existing five subtypes of ILTV. The main outbreaks were localised in New South Wales and Victoria where the avian respiratory diseases could possibly be limited by prevention using a range of therapeutic strategies developed over decades (Blacker *et al.,* 2011). The research outcomes have shown humoral immune response against viral pathogens as one of the major solutions across the poultry industry. In addition, antibody production is one of the principal protective strategies in the avian respiratory system against ILTV, AIV and IBV (Swayne, 2009) such as production of antibodies against the functional surface glycoproteins (Capua and Alexander, 2008).

Another effective strategy of public health focuses on the combination of vaccination and non-pharmaceutical intervention solutions in order to maximise surveillance activities (Sambhara and Poland, 2010). The veterinary vaccine production has been available for more than 50 years. Nowadays, development of recombinant genetically engineered techniques provides an opportunity to develop more efficient vaccines against avian viral pathogens. Two major strategies were introduced for the production of viral vaccines including modified live attenuated virus (Moss, 1991) and the other employing chemically inactivated virus. The novel approaches such as protective immunogens available for many viral pathogens in a single vector were created against RNA viruses such as NDV and IBV using reverse genetics system (Huang *et al.,* 2003). A wide range of therapeutics strategies is available due to the constant development of veterinary medicine. However, in underdeveloped countries such as India, the impact of avian respiratory diseases causes a

significant loss to the poultry industry and requires more preventive steps in the case of outbreaks (reviewed by Dhama *et al.*, 2013).

A series of IBV outbreaks amongst commercial poultry occurred in different rural parts of Argentina between 2001 and 2008. During these seven years, 20 serotypes of IBV were isolated and classified into three separate clusters with no relationship to vaccine serotypes, clearly showing the weakness of the Massachusetts serotype vaccination programme used during the outbreak in South America (Rimondi *et al*., 2009). In a separate outbreak in Thailand (January-June 2008), thirteen field isolates were investigated to determine their evolution and origin. The analyses indicated the presence of two IBV groups within Thai IBV isolates which evolved separately but were circulating together, risking the creation of more genetically complicated serotypes (Pohuang *et al*., 2009). A separate case of fast spreading IB disease caused by strain CK/CH/LDL/97I brought major economic losses in Iraq, Jordan and Saudi Arabia in 2011. This situation shows the failure of the vaccination programme and the first reported case of strain CK/CH/LDL/97I originating from Taiwan present in the Middle East. These examples of outbreaks clearly show poor effectiveness of vaccination programmes due to larger IBV diversity (Rimondi *et al*., 2009, Ababneh *et al*., 2012).

1.4 Vaccine development against IBV and other avian pathogens

Current vaccination programmes focus on induction of the low-virulent, live or inactivated vaccine in order to boost protective immunity and minimise the antagonistic response of epithelial cells in the respiratory tract (Kapczynski and King, 2005). The danger of uncontrolled avian airborne viruses spreading can increase due to nature of tracheal epithelial cells which fully support AIV and NDV growth and reinfection in the respiratory system. The available types of live and dead vaccine can possibly reduce shedding of the avian airborne pathogens, morbidity, mortality and transmissibility in infected areas (Hafez *et al.,* 2010) but also decrease the risk of new strains emerging by reducing field strain replication and providing several weeks of protection with a single vaccination of the poultry population (Swayne, 2006). The systematic reduction of poultry loss is possible due to implementation of dry powder aerosol vaccine against viral pathogen. However, the

major development issue is the deposition pattern of nebulized microspheres in the airways of chickens of different ages. The dose of vaccine necessary for sufficient protection of a chicken strongly depends on selection of appropriately sized airborne microspheres and their potential to avoid deposition in air sacs, lungs or penetration into lower respiratory tract (Corbanie *et al.,* 2006).

The available live vaccine against avian respiratory pathogens induced higher protection and has been used for over 50 years (Zaffuto *et al.,* 2008). The development of live vaccination program was used against Turkey rhinotracheitis virus (TRTV) focusing on 30 cycles of alternate passages in 9-day old chicken embryos. Interestingly, the TRTV was identified as still infective presenting only mild, transient respiratory signs with full protection developed after 21 days. The research studies suggested that this attenuated strain has potential for use as a vaccine (Cook and Ellis, 1990). The progress in genetics technology also allowed for more efficient design of live vaccine against newly emerging avian pathogens after identification of fourteen amino acid substitutions which might be responsible for replication and pathogenicity changes (Buynak and Hilleman, 1966). Moreover, further analyses has shown that pathogenic strain was partially attenuated after 50 and 70 egg passages and was fully attenuated after 90 passages, giving mortality and morbidity rates and viral loads in inoculated birds (Sun, *et al.,* 2014 and Enders, *et al.,* 1946).

Until recently, most IBV vaccines were based on live attenuated or killed vaccines derived from classical or variant serotypes. These vaccines were developed from strains originating from the USA such as M41, Ma5, Ark, and Conn and Netherlands, for example, H52, as well as European strains such as 793/B, CR88, and D274. Effective and economically viable vaccines against IBV are available such as recombinant infectious bronchitis virus (IBV) H120 vaccine strain, but multiple combinations of available vaccines are needed because the level of cross-protection against different IBV strains is insufficient (Yang *et al.,* 2016). Poor cross-protection is the result of variation in a major surface protein of the virus (the spike (S) protein). Therefore, highly attenuated nephropathogenic IBV vaccine, K2 strain was introduced to provide significantly higher levels of protection against variant QXIBV compared with chickens vaccinated with H120 in China (Lim *et al.,* 2011). New variant strains of IBV (currently over 50) with differences in the S protein appear regularly

in the field and, through analysis based on the sequence of the S protein, it is impossible to predict which vaccines will induce protection against the newly emerged viruses (Bande *et al.,* 2015). Notably, changes in geographical distribution and tissue tropism have been observed in QX-like strains that initially emerged in China and spread to cause great economic loss to poultry farmers in a region of south-east Asia. Only elaborate and expensive testing in chickens elucidates which vaccine combination is needed to protect against a new strain of IBV (De Wit *et al.,* 2011). The availability of a unique reverse genetics system for IBV has the potential to lead to the development of a new generation of live vaccines. Vaccination-challenge experiments with the same and with different viruses could possibly identify if there are different degrees of protection and also examine the causes of insufficient and unpredictable levels of cross-protection. However, so far no specific vaccine or combination of vaccines will protect against all IBV types, and the only sure way to determine if a vaccination program will work is to test it experimentally in chickens against the variant virus in question (Cook *et al.,* 1999).

Deep Sequencing also called as Massive Parallel Sequencing has important applications to vaccine design against viral pathogens. The excellent example is NGS of the wild-type yellow fever virus strain Asibi presenting a RNA virus quasi-species whereas the live attenuated 17D vaccine virus was identified with limited quasi-species, indicating the excellent safety record of the attenuated vaccine strain (Barrett, 2016). Unfortunately, live vaccine efficiency strongly depends on its ability to multiply in respiratory tract tissues and induce an immune response. Therefore, the crucial point is to identify polymorphisms in major surface protein of the virus (the spike (S) protein) responsible for the virus virulence using NGS technology. The danger with introduction of live vaccines is ultrastructural changes in tracheal epithelium cells such as hypertrophy of the goblet cells, deformation of cilia or even plaque of lysed cells (Mast *et al.,* 2005). Interestingly, the damage to epithelial cells is also noticed in the case of infection with IBV which invades avian upper and lower respiratory tracts facilitating secondary infection with *E.coli* in the trachea (Nakamura *et al.,* 1996). Therefore, the disease outbreak caused by a vaccine strain that has reverted to virulence in unvaccinated or immunocompromised chickens tends to cause a high mortality rate and unpredictable spreading of IBV (Pohjola *et al*., 2014).

1.5 Polymorphism – the key in live vaccine production

Currently live IBV vaccines are normally attenuated by multiple repeat passage of a virulent virus in specific pathogen free (SPF) embryonated chicken eggs (McKinley *et al.*, 2008). Spontaneous mutations also called Single Nucleotide Polymorphisms (SNPs) arise throughout the IBV genome some of which lead to attenuation of the virus. However, as a consequence of this method the attenuated viruses produced by this approach have only a few SNPs that are responsible for loss of virulence and these will differ between vaccine strains (Cook *et al.*, 2010). Two major drawbacks of this method is that once the virus is used to inoculate chickens the mutations within the attenuated vaccine viruses may back-mutate resulting in virulent virus, an undesirable consequence, or that as a consequence of multiple passage the immunogenicity of the attenuated virus will not result in adequate protection. Given the nature of these live attenuated vaccines, further passage must be kept to a minimum to prevent potential loss of immunogenicity. The stocks must be grown in SPF chicken eggs to prevent the introduction of other potential pathogens (Ladmana *et al.*, 2002).

The extracted pyrosequenced viral genome identified with the SNPs after multiple egg passages gave a brief overview of the highly polymorphic regions in IBV genome and helped to identify the polymorphic similarities and differences in viral genome of four egg isolates. Further investigation of previous egg passages by Sanger sequencing was required to determine and understand the process of SNP accumulation over the time. In fact, compared to the virulent strains, an attenuated virus has only nine to eighteen SNPs at the consensus sequence level but their role in loss of virulence must be confirmed by further investigation to determine which SNPs in the IBV genome are responsible for virulence and/or immunogenicity (Britton *et al*., 2012). The reverse genetics system presents many advantages including stability of the recombinant vaccinia virus genome supported by efficient replication in tissue culture. The process mainly focuses on generating a recombinant IBV using full-length cDNA which is cloned into the vaccinia virus where *in vitro* assembly is not a necessary step. The main advantage of using the vaccinia virus as a vector is the ability to generate modifications in the IBV cDNA using homologous recombination. This particular method enables production of insertion and deletion sequences from the vaccinia virus (Casais *et al*., 2001, Britton *et al*., 2005). Because of

substantial non-synonymous SNPs in the IBV spike glycoprotein, a recombinant IBV was created using the genetically modified Beaudette strain containing the Beaudette spike encoding region (Beau-R) and the virulent M41 strain (BeauR-M41(S)) which was observed to cause minimal damage to the ciliated epithelium cells of the trachea (Hodgson *et al*., 2004). Increasing our knowledge of the factors that influence the infectious ability of IBV will lead to the development of more effective live vaccines and reduction of the mortality caused by pathogens on farms across the world (Lim et al., 2012). During the last decade an increasing effectiveness of live vaccines was observed due to improvements in understanding the genetic modifications and genes responsible for IBV pathogenicity (Bourogâ, 2014).

1.6 Questions addressed by the MSc of Virology by Research:

- **What are the highly polymorphic encoding regions of IBV?**

The investigation of genotypic diversity between IBV strains is an essential step to understand the differences in genome structure. In this thesis, experiments focused on analysing four IBV isolates each subsequently passaged a further 106 times in embryonated chicken eggs. The SNPs produced after egg passage four (ep4) and 110 (ep110) were identified using 454 pyrosequencing in order to determine the highly polymorphic encoding regions of the IBV genome. Traditional Sanger sequence data of IBV genome from the four independent egg isolates were also analysed to determine the pattern of SNPs among IBV genes (nucleocapsid and spike encoding region) during egg passage. The investigation of highly polymorphic encoding regions was based on 454 pyrosequencing data and was combined with analyses of regular Sanger sequence data obtained from every 10 egg passages in order to detect SNP accumulation (figure 1.2).

- **Which of these regions are common or unique in the IBV genome within the four egg isolates?**

A key to understanding the complicated process of attenuation involves comparing the four viral populations focusing on nucleotide sequences obtained from pyrosequenced ep4, ep110 and Sanger sequenced egg passages (ep10-100) in order to determine similarities

and differences between these IBV isolates. The analyses of viral pyrosequencing data obtained after multiple passages in embryonated chicken eggs were performed by using bioinformatics tools to create a map of polymorphic encoding regions in the IBV genome from the four egg isolates. These maps were then compared to investigate the possible appearance of independent SNPs and/or SNP clusters characterized as unique or common in IBV isolates.

- **When does the process of SNP accumulation occur in the viral genome during passage in embryonated chicken egg?**

The analysis of 454 pyrosequencing data from the pathogenic ep4 of IBV subsequently passaged 106 times in embryonated eggs generated the first sequence data about possible IBV genome SNP accumulation due to the relaxation of immune pressure resulting in attenuation. The Sanger sequence data from previous egg passages was considered as principal information about accumulation and distribution of SNPs over time. This information allowed to determine the pattern and frequency of accumulated SNPs in particular those encoding regions of viral genome during the egg passages.

1.7 The aim of MSc of Virology by Research

The main aim of the project was to determine the polymorphism during the 110 egg passages and create a "map" of SNPs in investigated IBV isolates. Firstly, the generated deep sequence dataset (ep4 and ep110) was used to identify the interesting regions within the viral genome that accrue high frequency polymorphisms during the process of attenuation. Secondly, a Sanger sequenced data set was obtained from previous egg passages (ep10-100) to examine the contribution of Nucleocapsid and Spike protein to attenuation process. The analysed two data sets then led to more focused studies with the ultimate aim of investigation SNPs accumulation over the time and possibly, quantifying their effects in loss of viral pathogenicity.

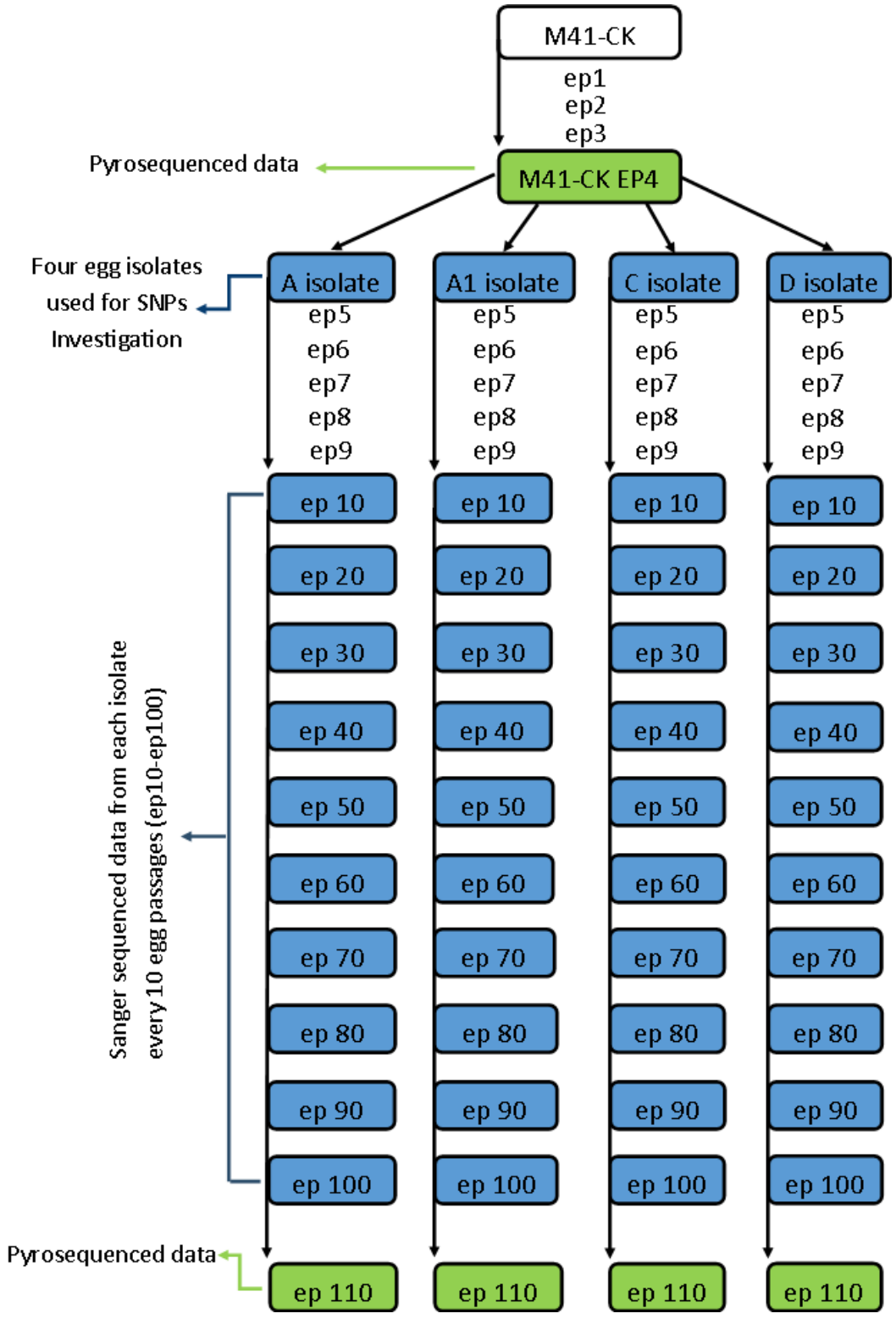


Figure 1.3. The summary of A, A1, C and D isolates originating form common M41 CK EP4 strain (ep4), independently passaged multiple times in embryonated chicken egg. White = original viral genome, green = Pyrosequenced egg isolates, blue = Sanger sequenced egg isolates.

Chapter 2: Materials and Methods

2.1 Introduction

This chapter explains in detail the materials and methods used during the investigation of IBV SNP accumulation in egg passages. M41 CK strain was passaged four times at The Pirbright Institute many years ago (data unavailable). The existing egg passage four (ep4) of M41 CK strain was the starting inoculum (main stock virus) for the four isolate groups originating from ep4 and subsequently passaged additional 106 times in embryonated chicken eggs by Dr Sarah Keep (The Pirbright Institute) between February 2010 and September 2011 (data not published yet). 454 pyrosequencing was then used to determine highly active SNP regions of IBV genome in starting egg passage 4 (ep4) and egg passage 110 (ep110) of each isolate. The importance in 454 pyrosequencing of ep4 and ep110 was the ability to sequence different DNA template separately in such a way that each sequence readout was linked to an initial single molecule DNA template which was amplified separately from other templates. This method was crucial for accurate identification the polymorphisms between ep4 and ep110. In the course of this investigation two encoding regions (nucleocapsid and spike) were detected with a high number of SNPs compared to the rest of the IBV genome. Further experiments focused on investigating SNPs in the nucleocapsid and spike encoding regions from Sanger sequenced egg passages using sequence analysis software. The main template for using Sanger sequencing was a pure cDNA sample (PCR amplified) with a primer binding site obtained after viral RNA extraction. The designed primer amplified a specific product which enables the sequence to be analysed by bioinformatics software. The outcomes of analysed IBV isolates sequence data from the nucleocapsid and spike encoding regions delivered principal information regarding the accumulation of SNPs over time. The data set collected after Sanger sequencing of each egg isolate every 10 egg passages gave a brief understanding of SNP appearance in the viral genome in the absence of immune pressure which in the future will be a foundation for deeper investigation of polymorphism in previous egg passages using 454 pyrosequencing. Published methods are referred to where appropriate. The details of primer sequences used for amplification of IBV encoding regions are included in Appendix 1.

2.2 Verification of viral attenuation in four egg isolates

The attenuation of four egg isolates was confirmed after infecting twelve 8-day old Rhode Island Red chickens administering eye drops (ten doses) of isolates (A-ep110, A1-ep110, C-ep110 and D-ep110) and demonstrating a lack of post-infection clinical signs on days 3-7. The four birds were culled on days 4 and 6 post-infection and ciliary activity and virus presence in trachea were assessed. The remaining birds were culled on day 7 post-infection and tissues harvested to check for virus presence. If during the period of observation, more than two chickens died from causes not attributable to the viral agent, the test would be repeated. These observations were made based on previous work confirming viral attenuation through chicken egg passages at Compton Laboratory (The Pirbright Institute) in April 2014. The viral attenuation was confirmed as no chicken showed clinical signs, in particular respiratory signs, and no chicken died from causes attributable to the egg passaged virus.

2.3 Standard methods for DNA and RNA screening

2.3.1 Viral RNA storage and primer design

The four ep4 isolates that came from embryonated chicken eggs were named A, A1, C and D. The RNA samples used to investigate the SNPs in IBV egg isolates were extracted from every 10 egg passages of each egg isolates using RNeasy Plus Mini Kit according to manual and stored in a freezer at -80 ^{o}C. The two sets of primers were designed based on Massachusetts strain M41 to investigate the best set of primers for encoding region amplification. The assay included using literature review and Primer3 website (Untergasser *et al.,* 2012) to design the best primers for amplification of the spike (S1 subunit and S2 subunit) and nucleocapsid encoding regions.

2.3.2 RNA concentration and complementary DNA synthesis (cDNA)

The assay was performed using a NanoDrop™ 2000 UV-Vis Spectrophotometer to measure concentration of total RNA and also detect the level of purity using 1 µl of each sample. The cDNA synthesis was performed according to the protocol (Superscript™ III Reverse Transcriptase) using RNA samples. The synthesis of the first strand of cDNA was followed by adding 1 µl of Random Primers (50-250 ng), 1 µl of 10 mM dNTP Mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH) and RNA sample (maximum of 11 µl with concentration <5000 ng), heated up to 65 ^{o}C for 5 minutes and incubated on ice for at least one minute. The contents of the tube were collected and mixed with 4 µl of 5x First–Strand Buffer, 1 µl 0.1 M DDT, 1 µl water and 1 µl Superscript™ III RT (200 units/µl). The contents were incubated at 25 ^{o}C for five minutes and 50 ^{o}C for 60 minutes. The final step was to inactivate the reaction by heating at 70 ^{o}C for 15 minutes.

2.3.3 PCR optimisation

Taq (Thermus aquaticus) DNA polymerase (Invitrogen) was used to investigate the optimum PCR condition by determining the best annealing temperature for the previously designed primers. The reaction contained all Invitrogen reagents except primers: 5x Green GoTaq flexi buffer, 25 mM magnesium chloride, 10 mM each dNTP, 10 mM each primer, 1 µl template (216 ng/µl) and 1.25 units *Taq* in a total volume of 25 µl.

Table 2.1. The range of annealing temperatures used for the investigated amplicons.

Amplicon	Primer name	Tm (^{o}C)	The annealing temperatures (^{o}C)
S1 subunit	S1_forward	56.1	54, 57, 60, 63
S1 subunit	S1_reverse	58.7	54, 57, 60, 63
S2 subunit part1	S2(p1)_forward	57.1	46, 48, 50, 52
S2 subunit part1	S2(p1)_reverse	58.8	46, 48, 50, 52
S2 subunit part2	S2(p2)_forward	72.3	62, 64, 66, 68, 70, 72
S2 subunit part2	S2(p2)_reverse	71.2	62, 64, 66, 68, 70, 72
Nucleocapsid	N_forward	58.4	51, 54, 57, 60
Nucleocapsid	N_reverse	58.7	51, 54, 57, 60

According to the information collected about the structure of the oligonucleotides and their Tm, the best annealing temperature and PCR conditions were chosen for improved and successful amplification of the encoding regions of interest. The PCR reaction for encoding regions of interest was prepared by pre-denaturation for 4min at 94 ^{o}C, followed by 30 cycles of denaturation at 94 ^{o}C for 60 s, annealing at various temperatures (table 2.1) for 40 s and extension at 70 ^{o}C for 120 s. The final extension step was performed at 72 ^{o}C for 10 min followed by analysing the PCR products on 1.0 % agarose gel.

2.3.4 Separation of PCR amplicons by agarose gel electrophoresis

Agarose gel electrophoresis was used to examine the ability of designed primers to correctly amplify the chosen encoding regions and to visualise amplicons. The separation of DNA fragments based on length was performed using a 1.0 % agarose gel made in1x TAE buffer. Syber Safe DNA Gel Stain (3.5 µl/1 g of LB agar) was added before the gel solidified. Gels were run for approximately 60 minutes at 85V then photographed under UV light. The Syber Safe Stain intercalated with the DNA to fluoresce and if necessary enabled bands to be cut out from the gel for purification.

2.3.5 The analyses of the PCR products

PCR products were purified using GFX PCR, DNA and Gel Band Purification Kit or PCR product purification (Qiagen QIAquick Spin Handbook) according to the manual. During the assay, the elution step was performed by using water (30 µl per sample) to extract DNA seen as bands on the agarose gel image with the exception of the A isolate samples (ep10-50) where 50 µl of water was used. The concentration of extracted DNA (ng/ µl) and the level of contamination of samples were then measured using a Nanodrop^{TM}. The amplified genome sequences (minimum 1 ng/µl/100base pair) from four isolates were sent for Sanger sequencing with the previously designed set of primers. The details of DNA concentration and level of sample purity are available in Appendices 10, 11 and 12.

2.3.6 Testing primers under specific conditions

The two samples with the highest concentration of DNA (minimum 1 ng/µl/100bp) were sent for Sanger sequencing with designed primers for each encoding region of interest. This step was required to test the ability of primers to amplify the IBV encoding regions in PCR condition provided by supplier (Source Bioscience).

2.4 Methods for analysing Sanger sequencing data

This part of the investigation used Sanger Sequencing technology for better understanding of polymorphisms in previous egg passages (ep10-100) of each egg isolate (A, A1, C and D). In DNA sequencing the nucleotide sequence of DNA is determined and appropriate treatments are used to generate DNA fragments that end at the four bases: adenine (A), guanine (G), cytosine (C) and thymine (T). Then the fragments were subjected to polyacrylamide gel electrophoresis so that molecules with one nucleotide difference in length were separated on the gel (Sanger *et al.*, 1977).

2.4.1 Chromas Lite

In the first stage of investigation into the viral SNPs, raw Sanger sequence data from every 10 egg passages of four isolates (A, A1, C and D) received from Source Bioscience was analysed using the bioinformatics program Chromas Lite (Li-Cor, 2014). The detection of SNPs in the sequence of interest was possible due to chromatograms presented in a form of visible but not overlapping picks indicating the presence of SNPs in the genome. The primers designed for amplification of specific encoding regions across the IBV genome were able to amplify nearly the whole encoding region presenting clear and visible chromatograms enabling independent SNPs to be easily spotted. However, sometimes the amplification process was disturbed by secondary or tertiary structures in the sequence of interest. Therefore, Sanger sequencing results were less effective leading to points in the nucleotide sequence where readings with Chromas Lite were stated as “N” (unknown nucleotide). Single errors in Sanger sequence reads were fixed by manual correction based

on personal judgment regarding the type of nucleotide present in specific genome positions. The last stage of analysing the Sanger sequence data with Chromas Lite focused on cutting, from the sequence, parts which were amplified but encoded other genes in IBV genome. The fully analysed Sanger sequence data was exported as a separate Fasta file for further investigation.

2.4.2 BioEdit

BioEdit (Hall, 2013) was used to visualise the sequences to detect any SNPs in the viral genome. The investigation of these SNPs required a full genome sequence of IBV (ep4) which was used as a reference sequence in the egg passages of the four investigated isolates. The full genome sequence was aligned together with the encoding region of interest and egg passage sequences amplified with the set of designed primers using BioEdit. This step played a crucial role in investigating the SNPs as the SNPs could be easily detected in the Sanger sequence data and identified at specific nucleotide positions by comparison with the full IBV genome.

2.4.3 Excel – 454 Pyrosequencing and Sanger sequencing data sets

The list of SNPs detected in the Sanger sequence data was saved as excel file for each egg isolate (Microsoft Corporation, 2010). The table was organised for easy reading of the information relating to SNPs detected in S1 subunit, S2 subunit (part 1) and nucleocapsid encoding regions. Each file consisted of the reference sequence (ep4) and SNPs after every 10 egg passages including nucleotide genome position, details of SNP, frequency in the case of ep4 and ep110 and also mutation type specification (Non-synonymous/ Synonymous). The identification of SNPs was done by comparison to the outcomes of 454 pyrosequencing in each egg isolates (ep110).

However, a second set of data had to be generated due to large amount of SNPs in investigated sequences which could not be previously detected in 454 pyrosequencing data. Therefore, each isolate and its egg passages consist of two sets of SNP tables relating to

each genome of interest. Similar to the 454 pyrosequencing data, information generated by Sanger sequencing includes the reference sequence (ep4) and SNPs after every 10 egg passages including nucleotide genome position, details of SNP but without the frequency in the case of ep4 and also no mutation type specification.

2.5 The 454 pyrosequencing analyses

The IBV isolate at ep4 as well as the four IBV isolates at ep110 (A ep110, A1 ep110, C ep110 and D ep110) were sequenced at the Liverpool Centre for Genomic Research (University of Liverpool) using 454 Pyrosequencing (Titanium chemistry, Roche). The 454 data was generated by universal library preparation for each sample and through emulsion-based clonal amplification of the DNA library fragments were put onto micron-sized beads. As a result of the amplification of the DNA fragments reads were filtered out during sequencing signal processing and mapped to the closest reference genome (M41-give accession number DQ834384) using Burrows–Wheeler Aligner (BWA) (Li and Durbin, 2009). The percentage gene coverage was calculated using the bedtools package (Quinlan and Hall, 2010). Single nucleotide polymorphisms (SNPs) and indels were then called using the varscan2 software package (Koboldt, *et al.*, 2012).

The analysed and mapped 454 pyrosequencing data was used to investigate SNP appearance in the IBV genome. It was possible to establish the number and SNP frequency in ep110 (A, A1, C and D isolate) after 106 egg passages by performing a series of simulations generated from reference egg passage (ep4) and comparison to its genome sequence. The information obtained from the 454 pyrosequencing data showed the map of SNPs across the genome of four isolates (A, A1, C and D), identified the independent SNPs and/or SNP clusters, calculated the frequency and helped make a decision about further analysis of the spike (S1 subunit, S2 subunit) and nucleocapsid encoding regions.

Chapter 3: Results

3.1 Introduction

This chapter explains the results obtained during the investigation of primer design, amplification, 454 pyrosequencing in ep4 and ep110 but also the Sanger sequencing of spike and nucleocapsid. These two encoding regions were used during the investigation of independent SNPs and SNP clusters (defined as the group of a minimum three SNPs within the IBV genome which are close to each other (<50 bp) and separated by the gap of 100 bp from the nearest SNP) in egg passages (ep10-100) of four egg isolates. The SNP frequency was calculated after performing a series of simulations generated from reference sequence of ep4. In addition, the reporting threshold of 5% was established due a high possibility of SNP error rate and >50% as the mark of SNP at consensus sequence level in 454 pyrosequencing data set. The appendices 2-5 include details of gel images relating to the investigation into the best annealing temperature for the designed set of primers. The gel images showing the amplification process of investigated spike and nucleocapsid encoding regions are included in appendices 6-8. Appendices 9-12 presented in a form of the tables relating to Nanodrop^{TM} measurements of RNA/DNA concentration and the purity of extracted and amplified IBV encoding regions. The additional SNPs identified in nucleocapsid encoding region of C isolate are presented in appendix 13. The table of IUPAC Ambiguity Codes used for identification minor SNPs are included in appendix 14.

3.2 The 454 pyrosequencing of four egg isolates after 110 egg passages

3.2.1 Polymorphism observed across each egg isolate

Figure 3.1 presents the total number of SNPs at > 5 % and >50 % frequency in each isolate across the whole IBV genome, with the highest number of SNPs detected in A isolate (45) and the lowest in A1 isolate (36). In addition, the bar chart also clearly indicates the low number of >50 % SNP frequency (consensus level) found within egg isolates compared to much higher number of >5 % SNP frequency, as expected. The numbers of >50 % SNP frequency varies from the highest number of 18 SNPs in A isolate to the lowest of nine SNPs in C isolate. Interestingly, C isolate has shown the highest number of SNPs (36) at

>5 % frequency presenting the lowest number of mutations (nine) at >50 % frequency. This observation might indicate that the process of attenuation is possible with only nine consensus SNPs, so shows that not many consensus level SNPs are needed to eliminate the virulence in egg isolates.

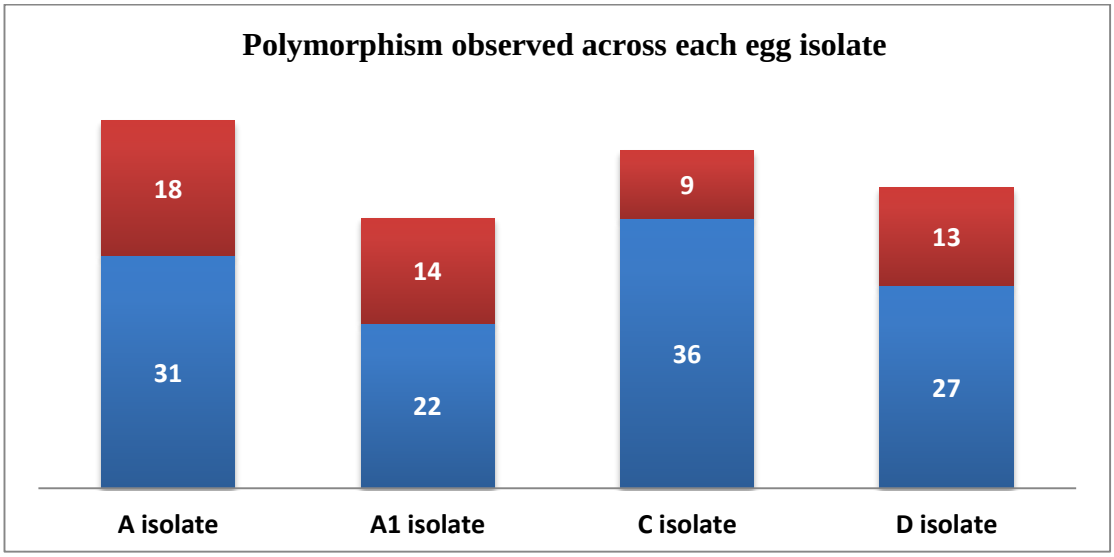


Figure 3.1. The total number of SNPs after 110 egg passages for each egg isolate with frequency >5 % marked as blue and >50 % marked as red colour bar.

The results showed a high accumulation of SNPs in the spike and nucleocapsid encoding regions at the low SNP frequency (>5 %) but also an increased SNPs accumulation with high frequency (>50 %) compared to the rest of the IBV encoding regions. Furthermore, the investigation of SNPs in the spike encoding region has shown S2 subunit to be more polymorphic than the S1 subunit. Moreover, figure 3.1 suggests that the attenuation process might be caused by as low as nine SNPs at consensus sequence level in the quasispecies due to the low number of SNPs with frequency >50 % observed in each egg isolate compared to the number of SNPs identified with >5 % frequency.

Table 3.1. The table includes the number of mutations with >5 % frequency observed in each gene of each isolate, the total number of mutations observed in the gene and amount of SNPs overlapping within egg isolates in particular gene. SNPs observed in two or more samples were counted only once in each case.

The SNPs with >5 % freq. identified in genes of each egg isolate including shared mutations.									
Isolates / Gene	**A**	**A1**	**C**	**D**	**Total**	**Observed in one isolate**	**Observed in two isolates**	**Observed in three isolates**	**Observed in four isolates**
Replicase 1a	**6**	**11**	**15**	**10**	**42**	**34**	**4**	0	0
Replicase 1b	**5**	**3**	**7**	**5**	**20**	**16**	**2**	0	0
Spike	**5**	**2**	**4**	**4**	**15**	**6**	**3**	**1**	0
3a	0	0	0	0	0	0	0	0	0
3b	0	0	0	0	0	0	0	0	0
E	0	0	**2**	0	0	0	0	0	0
M	**1**	0	**2**	0	**3**	**3**	0	0	0
5a	**2**	0	0	0	**2**	**2**	0	0	0
5b	**1**	**2**	0	0	**3**	**3**	0	0	0
Nucleocapsid	**11**	**4**	**6**	**8**	**29**	**15**	0	**2**	**2**

The full details of shared SNPs identified within single gene of egg isolates were included in table 3.1 (>5 % freq.) and table 3.2 (>50 % freq.). The analysis of 454 pyrosequencing data clearly indicated the presence of unique and common SNPs at >5 % frequency shared between the egg isolates in ep110 (table 3.1). The highest polymorphism unique for each gene was identified within replicase 1a (34 SNPs), replicase 1b (16 SNPs) and nucleocapsid encoding region (15 SNPs). Moreover, analysis of the deep sequenced dataset also indicated unique polymorphisms for egg isolates in spike (six SNPs), M and 5b (three SNPs) and 5a encoding region (two SNPs). Interestingly, the phenomenon of SNPs appearance at the same nucleotide position in more than one egg isolate were observed in encoding regions of nucleocapsid with appearance of two SNPs within three and four egg isolates at >5 % frequency. Further investigation of the deep sequenced data has also shown spike encoding region with three SNPs observed across two egg isolates and one SNP identified within three egg isolates. Furthermore, the polymorphisms present identified mutations in replicase 1a (four SNPs) and replicase 1b (two SNPs) shared within two egg isolates. Further investigation clearly shows lack of any common SNPs within the egg isolates in encoding regions of E, M, 5a and 5b, encoding regions previously identified with a small number of unique mutation in ep110 of A, A1 and C isolate.

Table 3.2. The table includes number of mutations with >50 % frequency observed in each gene of each isolate, the total number of mutations observed in the gene and number of SNPs overlapping within egg isolates in particular gene. SNPs observed in two or more samples were counted only once in each case.

The SNPs with >50 % freq. identified in genes of each egg isolate including shared mutations.									
Isolates / Gene	**A**	**A1**	**C**	**D**	**Total**	**Observed in one isolate**	**Observed in two isolates**	**Observed in three isolates**	**Observed in four isolates**
Replicase 1a	**7**	**10**	**4**	**5**	**26**	**22**	**2**	0	0
Replicase 1b	**1**	**2**	**1**	**1**	**5**	**4**	**1**	0	0
Spike	**4**	**2**	**3**	**4**	**13**	**4**	**3**	**1**	0
3a	0	0	0	0	0	0	0	0	0
3b	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0
5a	0	0	0	0	0	0	0	0	0
5b	**2**	0	0	0	**2**	**2**	0	0	0
Nucleocapsid	**4**	0	**1**	**3**	**8**	**8**	0	0	0

Furthermore, analysed pyrosequencing data with >50 % frequency in ep110 shows reduction of mutations in all investigated encoding regions (table 3.2). The deep sequenced dataset clearly shows the highest number of unique SNPs identified in replicase 1a (22 SNPs) and nucleocapsid (eight SNPs), within replicase 1b and spike encoding regions (four SNPs) and also 5b encoding region (two SNPs). Interestingly, further investigation of shared mutations confirmed the presence of SNPs at consensus sequence level within three (one SNP) and two egg isolates (three SNPs) in spike at >50 % frequency rate in ep110. In addition, 454 pyrosequencing data presents the shared SNPs at >50 % frequency rate within two egg isolates in replicase 1a (two SNPs), replicase 1b (one SNP) but lack of any polymorphism in 3a, 3b, E, M, 5a, 5b and nucleocapsid encoding region within more than one egg isolate. Interestingly, spike was identified with the highest number of shared SNPs within the deep sequenced dataset at >50 % frequency, where nucleocapsid encoding region shared mutations within four egg isolates only at >5 % frequency. Therefore, both genes could be the potential keys encoding regions in a process of attenuation.

3.2.2 The most variable regions of IBV genome

Analysis of the deep sequence data identified four IBV encoding regions that had a larger number of SNPs than the other encoding regions of IBV genome including replicase 1a, replicase 1b, spike and nucleocapsid. Each egg isolate (A, A1, C and D) was treated individually focusing on pyrosequencing datasets to calculate the total number of synonymous and non-synonymous mutations in different genes after equal coverage was applied to analysed genomic regions and isolates. The shared SNPs identified in encoding region of more than one egg isolate were counted for each egg isolate separately and presented together with total number of SNPs above a frequency of 5 % in figure 3.2. The generated data shows a substantial number of SNPs established on the level of 30 non-synonymous (NS) SNPs (2.3/1000 bp) and 12 synonymous (S) SNPs (0.9/1000 bp) in replicase 1a encoding region and high number of 13 NS (11/1000 bp) and 16 S (13/1000 bp) in the nucleocapsid encoding region. The replicase 1b and spike encoding regions also showed increased number of SNPs with >5 % frequency compared to the rest of IBV encoding regions (replicase 1b: 12 NS at 1.5/1000 bp and 8 S at 1/1000 bp and spike 8 NS at 2.3 /1000 bp and 7 S at 2/1000 bp). The M, 5a and 5b encoding regions present a very low level of detected SNPs with maximum of two NS SNPs (8.0/1000 bp) in 5b and only one S SNP (4/1000 bp) in M encoding region. In addition, further analyses of deep sequenced data also showed that 3a, 3b and E encoding regions were identified with no SNPs relating to frequency greater than 5 %.

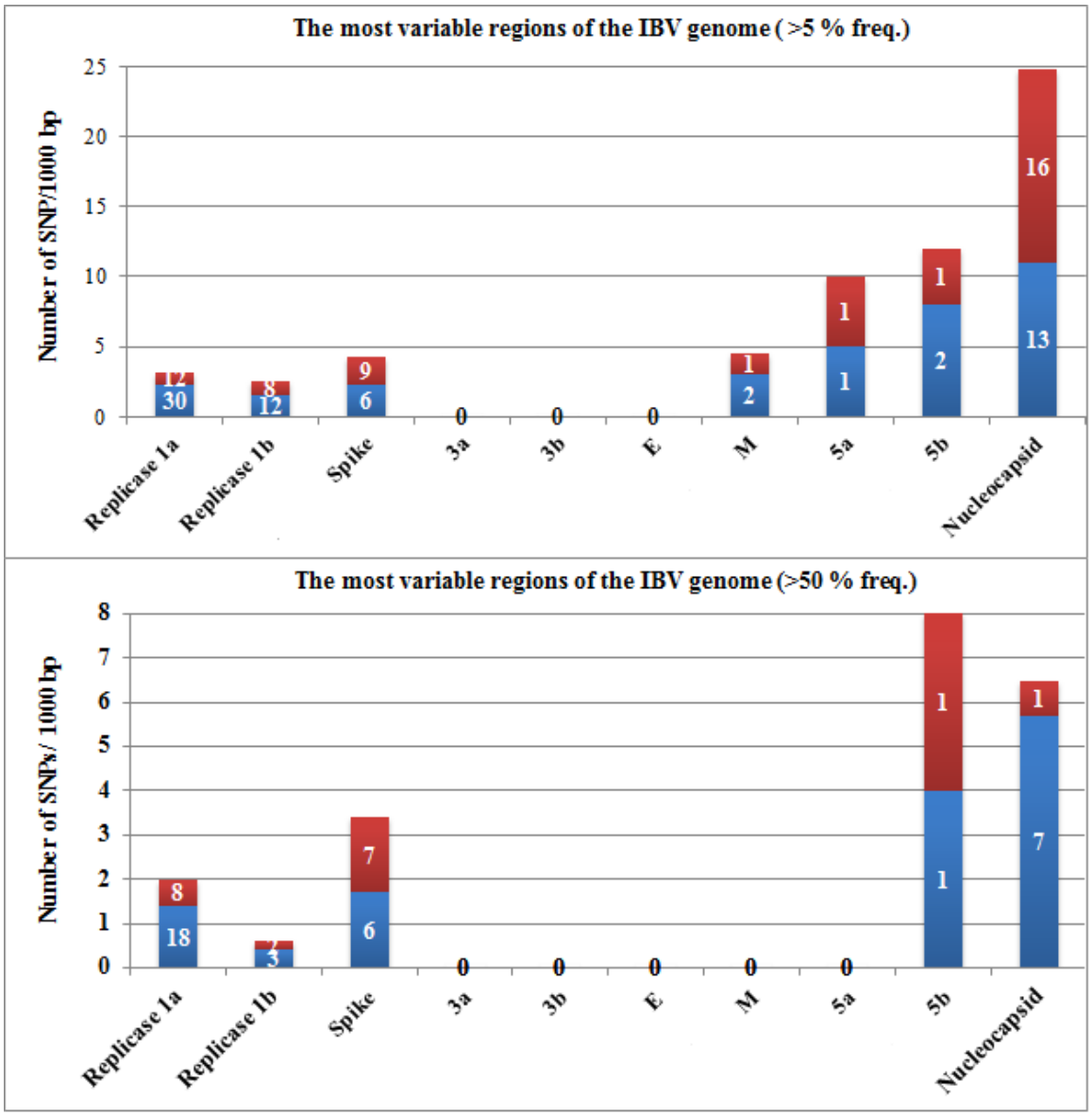


Figure 3.2. The total number of SNPs (non-synonymous and synonymous) with frequency >5 % (top) and >50 % (bottom) presented in different encoding regions of four egg isolates. The SNPs were expressed in the form of two coloured bars: blue as non-synonymous and red as synonymous SNPs per 1000 bp including absolute numbers of mutations in each encoding region.

The results of deep sequencing relating to SNP frequency >50 % (i.e. consensus level) showed the spike and nucleocapsid as two encoding regions with the highest number of SNPs with six NS SNPs, established as 1.7/1000 bp (spike) and seven NS SNPs at 5.7/1000 bp (nucleocapsid). In addition, both encoding regions also presented synonymous SNPs identified as seven SNPs (2/1000 bp) in the spike and one SNP (0.8/1000 bp) in the nucleocapsid. The number of SNPs detected in replicase 1a was a little bit lower with the

increased SNP frequency of >50 % presenting as 18 SNPs at 1.4/1000 bp (NS) and eight SNPs at 0.6/1000 bp (S). The number of SNPs detected in replicase 1b was reduced and established as three SNPs at 0.4/1000 bp (NS) and two SNPs at 0.2/1000 bp (S). These data clearly shows the highest reduction of SNPs when increasing frequency from >5 % to >50 % was in replicase 1b (equally of 75 % reduction of NS and S) and nucleocapsid (56 % reduction of NS and 94 % of S).

3.2.3 Polymorphism identified in spike encoding region

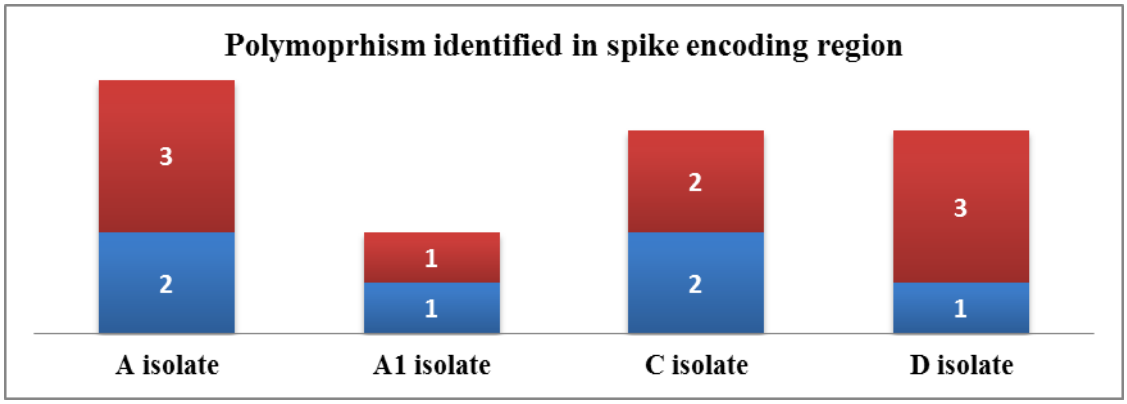


Figure 3.3. The number of SNPs in the spike encoding region after 110 egg passages with frequency >5 %. The graph illustrates identified SNPs in the S1 subunit (blue) and the S2 subunit (red) of spike protein encoding region in the four investigated egg isolates.

When considering the size of each encoding region in IBV, the spike and nucleocapsid accumulate the most SNPs which might imply a role in viral attenuation of the investigated four egg isolates. Figure 3.3 shows the number of SNPs at >5 % frequency detected in the two subunits of the spike encoding region. The A isolate was identified with two SNPs in the S1 subunit comparing to three SNPs found in S2 subunit. Furthermore, also D isolate clearly presented a greater number of SNPs in the S2 subunit compared to S1 subunit. Interestingly A1 and C isolates have shown an equal number of SNPs shared between subunits of the spike encoding region. The results of deep sequenced data might be due to the smaller size of the S1 subunit (1607 bp) compared to the S2 subunit (1855 bp) enhancing lower number of mutations. However, this observation shows definitely a unique situation as the S2 subunit is more conserved than the S1 subunit in wild IBV strains.

3.2.4 Number of shared SNPs in spike encoding region

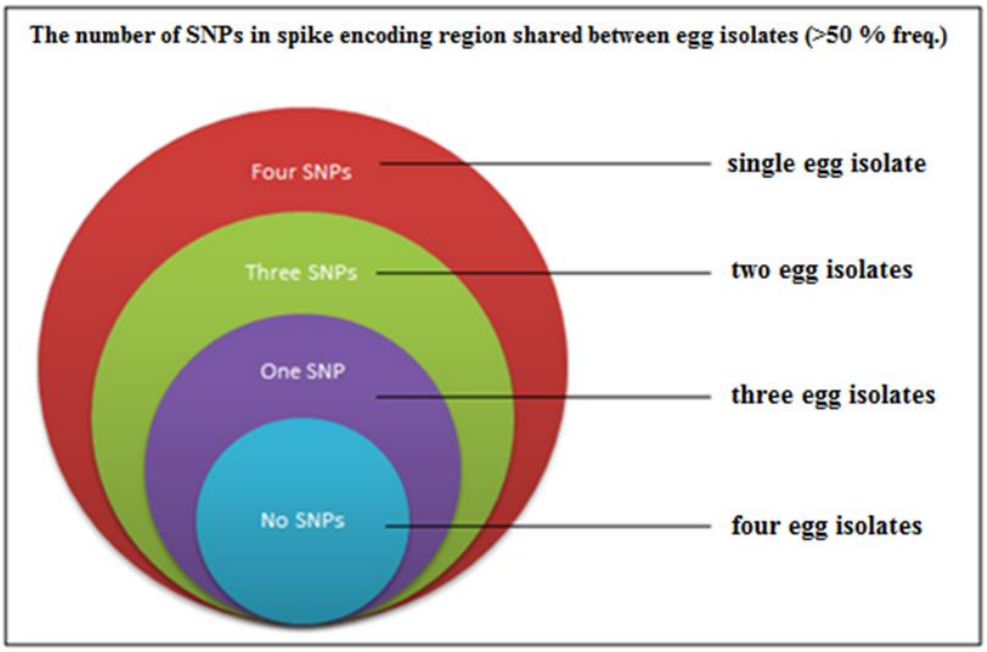


Figure 3.4 The Venn diagram illustrates polymorphism in spike encoding region at >50 % frequency in ep110. The number of SNPs shared within the single, two, three and four egg isolates were included inside the coloured circles.

The further analyses of pyrosequencing data suggested that the spike encoding region plays a key role in the process of attenuation. The figure 3.4 presents the total number of shared mutations within four egg isolates. The S1 subunit has been identified with non-synonymous mutation at nucleotide position 20720 in A, C and D isolate. S2 subunit has shown three synonymous SNPs at consensus sequence level (nucleotide position 22224 and 22240) shared between C and D isolate and at nucleotide position 22421 in A and A1 isolate. In addition, spike encoding region was also identified with four independent SNPs found only in a single egg isolate. The only one common > 50 % frequency SNP (nucleotide position 20720) in S1 subunit shared between three egg isolates might indicate the crucial role of unique SNPs at consensus sequence level but also minor SNPs in process of attenuation. Interestingly, S1 subunit known as highly polymorphic in wild IBV strains presents only one independent mutation (>50 % freq.) at nucleotide position 20429 in C isolate. In addition, the S2 subunit was found with three independent SNPs at nucleotide positions 22285 and 22504 (A isolate) and nucleotide position 22459 (D

isolate). These data show that a few changes in the viral genome might play a significant role in the attenuation process of egg isolates.

3.3 The preparation of samples for Sanger sequencing

3.3.1 The concentration of RNA samples

The RNA concentration after extraction of viral samples varied from 1.2 ng/µl (A1-ep100) to 9.9 ng/µl (D-ep50). The measurements of 260/280 ratio were an acceptable level (1.8-2.0) but 260/230 value presented in Appendix 9 is appreciably lower (<0.64) than expected (2.0-2.2) for non-contaminated nucleic acids suggesting the presence of contaminants such as carbohydrates, ethanol or guanidine isothiocyanate used for RNA isolation.

3.3.2 The outcomes of primer design

The two sets of primers were designed for amplification of whole S1 encoding sequence with the forward primer (5`-TGA AAA ACT GAA CAA AAG ACA GA -3`) and the reverse primer (5`- GGG CAA CTT GTT ACA TTT TCA C -3`) based on information found in Zhu, *et al.,* 2007. The second set of primers: forward primer (5`- GCT CTA GAA TGT TGG TAA CAC CTC TT -3`) and reverse primer (5`- CGG GAT CCT TAA TAA CTA ACA TAA GG -3`) were used according to Zhou, *et al.,* 2003. The primer design for the S2 subunit was more complicated due to the larger size of the encoding region (approx. 1800 bp). In this case the S2 subunit was divided into two separate but overlapping parts with two sets of primers for each encoding nucleotide sequence. The first part of the S2 subunit (892 bp) was amplified with the first set of oligos: forward primer (5`- GGT GGT ATT ACT GCA GCT GGT GC-3`) and reverse primer (5`- CTG ACC CTT TGT CAC AGT CTA GAC -3`) and second set of primers: forward primer (5`- ATG GTT TGC TTG TGT TGC CTC CC-3`) and reverse primer (5`- TTC CAG GAG CTA AGG GTA CAG CC -3`) designed on the Primer3 website. According to the information obtained from pyrosequenced ep4, the design of primers for S2 subunit (part 2) encoding sequence was possible with the first set of forward primer (5`- ATG GTT TTG CTT GTG TTG CCT -

3`) and reverse primer (5`- AGG AAG GAC GTG GGA CTT TG -3`) and the second set of primers: forward primer (5`- TGG TTT GCT TGT GTT GCC TC-3`) and reverse primer (5`-GGA AGG ACG TGG GAC TTT G -3`). The amplification of the nucleocapsid encoding region was successful due to two set of primers designed manually. The several basic rules of primer design were applied: primers were 18-24 nucleotides in length with a GC content of 40-60 % and an annealing temperature (Tm) >55 ^{o}C. The amplification assay was done with the first set of primers: forward primer (5`- CCA GGG GAA AAC TTG TGA GG -3`) and reverse primer (5`- TAC CGT TCG TTT CCA GGC TA -3`) and second set of primers: forward primer (5`- CCG CGT GTA CCT CTC TAG TA -3`) and reverse primer (5`- GGT CTA CCG TTC GGT TCC AG -3`).

3.3.3 The identification the optimum annealing temperature of primers

3.3.3.1 Spike encoding region (S1 subunit)

The primers for the S1 subunit were designed based on the information found in Zhu *et al.*, 2007 and Zhou *et al.*, 2003. The two sets of primers designed for amplification of the S1 subunit were tested at a different range of temperatures (51-67 ^{o}C) and investigated using gel electrophoresis with a 1 % agarose gel. The first set of primers failed to amplify the S1 subunit. The investigation of the second set of primers showed a very thin band appearing on the gel image with the appropriate size at a range of annealing temperatures between 54-63 ^{o}C. This observation led to the PCR and the gel electrophoresis being repeated for the second set of primers with a narrow range of annealing temperatures between 54-63 ^{o}C. According to the results presented in Appendix 2 the second set of primers was able to amplify the S1 subunit completely at an optimum annealing temperature of 57 ^{o}C, producing strong bands with the predicted PCR product size on the gel.

3.3.3.2 Spike encoding region (S2 subunit)

Both sets of primers were tested with the same range of temperatures (46-52 ^{o}C) due to similar melting temperature (57.1-58.8 ^{o}C) established using the Primer3 website. The appearance of bands with the correct bp size of product was observed in the cases of both primer sets. The amplification of the S2 subunit (part 1) was the most effective at an annealing temperature of 50 ^{o}C (the second set of primers) shown in Appendix 3. The primer design for amplification of S2 subunit (part2) was tested with a wide range of temperatures for more accurate investigation of the optimum annealing temperature demonstrating the most efficient amplification at 68 ^{o}C (the first set of primers) and failure of amplification process with the second set of primers (Appendix 4).

3.3.3.3 Nucleocapsid encoding region

The successful amplification of the nucleocapsid encoding region was observed in the case of both primer sets (Appendix 5). The strongest bands were presented with the first set of primers in the range of temperatures between 51-60 ^{o}C. The second set of primers also amplified the correct IBV encoding region but the PCR product was less visible on the gel image at annealing temperatures of 51-60 ^{o}C. In addition gel bands at annealing temperature between 54-57 ^{o}C are visible but not as bright compared with the first set of primers. Therefore, the further amplification of nucleocapsid encoding region was continued with the first set of primers in annealing temperature of 57 ^{o}C.

3.4 Results of cDNA synthesis and amplification

3.4.1 Introduction

This section shows in details the results of primers testing for spike and nucleocapsid encoding region. The investigation of the ability of the primers to amplify those encoding regions under specific conditions was necessary to investigate independent SNPs and SNP clusters in egg isolates. In addition, the spike encoding region was divided due to large bp size in three parts: S1, S2 (part 1) and S2 (part 2) subunits. The process of amplification spike and nucleocapsid focused on every 10 egg passages from ep10 to ep100 in each isolate. The gel images showing the amplified products are included in appendices 6-8. Appendices 9-12 included information related to Nanodrop™ measurements of RNA/DNA concentration and the purity of extracted and amplified IBV encoding regions.

3.4.2 Testing primers under specific conditions

The samples represent four encoding regions with relatively high DNA concentration ranging from 80.9 ng/µl (D isolate-ep50) to 29.5 ng/µl (A isolate-ep20). The contamination of samples after gel band purification was low shown by a 260/280 value at the level between 1.81(D isolate-ep80) and 2.39 (D isolate ep50) but 260/230 values were identified with a result much lower (<0.63) than expected 2.0-2.2 unit.

Table 3.3 Summary of samples sent for test sequencing.

Sample #	Amplicon	Isolate/Ep	Conc. (ng/µl)	260/280	260/230
1	S1 subunit	A/20	29.5	1.97	0.16
2	S1 subunit	D/80	29.8	1.89	0.12
3	S2 subunit (p1)	A/20	39.9	1.99	0.41
4	S2 subunit (p1)	D/50	80.9	2.39	0.63
5	S2 subunit (p2)	A/20	19.4	2.32	0.03
6	S2 subunit (p2)	C/50	31.1	1.97	0.15
7	Nucleocapsid	D/40	34.7	1.82	0.28
8	Nucleocapsid	D/80	33.7	1.81	0.23

3.4.3 Spike encoding region - S1 subunit

The gel image presents very strong bands in every 10 egg passages of A isolate in ep10-50 and less visible bands in ep60-100 (Appendix 6). The A1 isolate showed only partial success of S1 subunit amplification with visible bands produced after ep60 to ep90. However, the C isolate was found with very effective amplification due to the size of the bands and their dark colour suggesting that the designed primers were able to amplify the gene from most egg passages excluding ep50 and ep100. The most efficient results of the amplification process were found with the D isolate which presented visible and strong bands in all egg passages except ep10. Furthermore, lack of visible bands representing non-specific products allowed for the second step focusing on purification of investigated isolates. Appendix 10 includes Nanodrop^{TM} readings relating to the concentration and purity of amplified products, highlighting the highest concentration of DNA found with the D isolate ep70 (86.5ng/µl) and the lowest in A1 isolate ep80 (16.9 ng/µl). The regular Sanger sequencing could be performed in all cases regarding to A, A1, C and D isolate where majority of PCR products ware successfully purified. The measurements of 260/280 ratio was found to be at a satisfactory level (1.8) but 260/230 ratio for most of DNA samples presented in appendix 10 are lower (<1.72) than expected 2.0-2.2 unit indicating sample contamination with carbohydrates, ethanol or Guanidine HCL used for DNA isolation.

3.4.4 Spike encoding region - S2 subunit (part 1)

The gel image of the cDNA S2 subunit (part 1) amplification (Appendix 7) showed partial success in the case of A and A1 isolate. The early ep10-50 of the A isolate and later ep60-100 of A1 isolate were successfully amplified resulting in strong, visible bands. Low quality bands were found with A isolate ep60-100 and A1 isolate ep20-50 (ep10 of A1 isolate was not involved in experiment due to low RNA concentration). The amplification of the C isolate egg passages was successful shown by the appearance of strong bands on the agarose gel. The complete amplification of the S2 subunit (part 1) was also observed in the D isolate where bands were presented in all investigated egg isolates except for ep10. The highest concentration of DNA was found with A isolate ep20 (50.4 ng/µl) and the lowest was in A1 isolate ep20 (8.9 ng/µl). Appendix 11 also indicates the values of

Nanodrop™ where 260/280 values were detected at a satisfactory level but 260/230 ratio was lower (<1.42) than expected level of 1.8-2.0 unit.

3.4.5 Spike encoding region - S2 subunit (part 2)

The S2 subunit (part 2) was not amplified except for two samples which were sent for test Sanger sequencing (A isolate ep20 and C isolate ep50). However, the failure of the Sanger sequencing reads showed that primers cannot be used for further investigation. In addition, the information obtained for 454 pyrosequencing data suggested that the S2 subunit (part 2) did not show high polymorphism presenting five SNPs at >5 % frequency and one SNP at >50 % frequency within four egg isolates.

3.4.6 Nucleocapsid encoding region

The amplification of the nucleocapsid encoding region was performed with the first set of designed primers which presented the strongest ability to amplify this particular encoding region of the viral genome. The gel images in Appendix 8 confirmed that most of the egg passages from the four egg isolates could be further analysed for the detection of SNPs. The investigation showed amplification failure only in ep10 of the D isolate with no visible bands. According to the Appendix 12 the highest concentration of DNA was found with ep40 (45.7 ng/µl) and the lowest in ep20 (15.7 ng/µl) of the A1 isolate. The measurements of 260/280 ratio was found in most of cases to be at a satisfactory level of 1.8-2.0 unit. However, a few samples of A, A1 and C isolate presented 260/280 values at >2.0 level which might indicate a problem with the regular Sanger sequencing reads. In addition, 260/230 values investigated with the Nanodrop™ measurements were lower (<0.84) than expected 2.0 unit (Appendix 12).

3.5 Outcomes of sequencing analysis

3.5.1 Introduction

This section explains in detail the results obtained during the investigation of SNPs using Sanger sequencing (ep10-ep100) in spike and nucleocapsid encoding regions of four isolates (A, A1, C and D), with comparisons to the SNPs identified from the 454 pyrosequencing data (ep4 and ep110). The analyses of 454 pyrosequencing data clearly showed lack of >50 % frequency SNPs shared between four isolates. However, S1 and S2 subunit showed SNPs correlation within more than one isolate at consensus sequence level. Further investigation, presents numerous SNPs identified with Sanger sequencing as independent SNPs and/or SNP clusters within the egg isolates but also numerous SNPs (C or T residue) within egg passages of each isolate. The additional SNPs of C isolate were included in appendix 13. Appendix 14 presents the table used for identification of minor SNPs using regular Sanger sequencing method within egg passages.

3.5.2 The SNP accumulation in S1 subunit

The investigation of SNPs in the S1 subunit detected with the 454 pyrosequencing data showed five SNPs found in the A isolate, four SNPs in the C and D isolate and one SNP in the A1 isolate. An interesting observation was made at nucleotide position 20720 where a non-synonymous SNP with a frequency of 99.5 % in ep110 was observed in the pyrosequencing data in the A, C and D isolates resulting in a change of amino acid sequence (lysine to arginine) as seen in table 3.4. However, the SNP accumulation in Sanger sequenced egg passages varies in each of the investigated egg isolates showing the A and C isolate with ambiguous SNPs (the ambiguity code R codes for A or G) in ep20-60 which at some point were classified as SNP at consensus sequence level. Interestingly, high SNP frequency rate in D isolate presents consensus sequence SNPs in all Sanger sequenced samples (ep50-70 and ep90) clearly illustrating similarities in SNP accumulation within egg isolates presented with ~100 % SNP frequency in ep110.

Table 3.4. The SNPs found in more than one egg isolate with specification of mutation NS/S (Non-synonymous/Synonymous) indicated to the sequence of ep4. Table also includes the % SNP frequency for ep4 and ep110. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

The nucleotide position 20720															
NS/S	**Ep # / Egg isolate**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
NS	**A**	A	100			R		G						G	99
NS	**C**	A	100		R		R		R		G	G		G	99
NS	**D**	A	100					G	G	G		G		G	100
The nucleotide position 21630															
NS/S	**Ep # / Egg isolate**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
S	**A**	T	100		C									C	1.7
S	**C**	T	100											C	1.1
S	**D**	T	100				C							C	1.1

In addition, further investigation showed the appearance of a SNP in the viral genome when the SNP frequency in the ep110 pyrosequencing data reached only <1.7 % and could not be detected in ep4. Interestingly, this low frequency SNP was found within the A, C and D isolates at nucleotide position 21630 as one of the total five SNPs observed in spike encoding region of more than one egg isolate. The single consensus sequence SNP might explain the very low SNP frequency in ep110 by showing through Sanger sequenced egg passages the accumulation in ep20 without further appearance of SNP. The S1 subunit of the spike encoding region was identified only with two SNPs that were found in more egg isolates suggesting some SNPs correlation between the four investigated egg isolates. Furthermore, the ≥99 % SNP frequencies in ep110 might have an influence on the appearance of the greater number of SNPs in the Sanger sequence data from previous egg passages comparing to low SNP frequency (≤1.7%) mirroring little SNP activity in previous egg passages.

Further analyses of sequence data also detected a number of SNPs unique to the investigated isolates as seen in table 3.5. The A isolate was identified with three minor SNPs with the highest SNP frequency (14.8 %) estimated at nucleotide position 21543 in ep20 and ep110 causing a change from aspartate to glutamate in the amino acid sequence of the viral population. The A1 egg isolate showed only one unique SNP at nucleotide

position 20429 as ambiguous SNP (C or T nucleotide) detected in ep50-60 followed by change at consensus sequence level in ep80 and ep100 resulting in 100 % SNP frequency in ep110. Further analyses of the C and D isolate confirmed the presence of minor SNPs with a low frequency (<1.2 %) in ep110 and no SNP accumulation in Sanger sequenced egg passages. Interestingly, the C isolate also presented a SNP at nucleotide positon 21489 with a frequency of 12.1 % in ep110 and a minor SNP in ep90. The number of unique SNPs (S1 subunit) presented in each egg isolate was found only in minor part of viral populations presented by a <15 % SNP frequency (except of A1 isolate). These data show that the frequency rate in ep110 depends on low SNP accumulation in previous egg passages investigated by regular Sanger sequencing.

Table 3.5. The minor SNPs and consensus sequence SNPs identified in all screened egg passages unique for each egg isolate. Specification of mutation NS/S (Non-synonymous/Synonymous) was indicated to the sequence of ep4. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

A isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**21543**	T	100		A									A	14.8
S	**21579**	T	100											C	2.8
S	**21618**	G	100											A	1
A1 isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**20429**	C	100					Y	Y		T		T	T	100
C isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**21489**	T	100									W		A	12.1
NS	**21928**	C	100											T	1.2
D isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**20760**	T	100											A	1.1
S	**20871**	C	100											T	1.2

3.5.3 The additional SNPs activity in S1 subunit

Further investigation into Sanger sequenced egg passages of the analysed egg isolates detected a number of SNPs previously not identified in the 454 pyrosequencing data. The A isolate had a small number of SNPs presented with a random appearance across the investigated egg passages (table 3.6.). In fact there are two mutations that continue from ep30 to ep50 but the analyses of SNPs after ep50 was impossible due to failure of Sanger sequencing reads. Interestingly, the ep10 in the A isolate showed a unique SNPs cluster built up within an 84 bp encoding region. The nine SNPs at consensus sequence level were identified between nucleotide positions 20536 and 20620 without its further identification in later egg passages. Furthermore, deeper analyses have detected the T residue replacement by a C residue or vice versa accounted as 78 % of the total SNPs observed in SNP cluster of ep10 A isolate. Further investigation also identified two minor SNPs within the SNP cluster at the nucleotide position 20536 and 20595 in the A isolate (represented with ambiguity codes in Appendix 14). The further sequence analyses presented an appearance of SNP without further continuation in later egg passages but also independent SNPs visible in more than one egg passages. In addition, three closely related consensus sequence SNPs were identified in ep30 at nucleotide positions 21271, 21273 and 21309 showing a formation of SNP cluster without continuation in further egg passages. However, there is an exception to previous example showing SNPs at nucleotide positions 21838 and 21865 with visible continuation of SNPs in the nucleotide residues of ep30 and ep50. This SNP accumulation was closely investigated presenting a SNP cluster formed by mix of minor and consensus sequence SNPs within nucleotide position 21803 -21865 found in ep50 of A isolate. However, the further observation was not possible due to failure of Sanger sequencing in ep60-100. These data have shown a major difference in accumulation of SNPs identified in early egg passages and/or in ep110. Considering the available Sanger sequence data, the additional SNPs previously not detected in ep110 (deep sequenced data) were present in the early egg passages of A isolate but also in later egg passages of A1 isolate.

Table 3.6. The Sanger sequencing outcomes of A and A1 isolate presenting independent SNPs and SNP clusters identified in S1 subunit after indicating to the sequence of ep4. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

A isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
20536	G	R									
20554	T	C									
20555	C	T									
20558	C	T									
20559	A	C									
20595	T	W									
20601	T	C									
20610	T	C									
20620	T	C									
21271	G			A							
21273	G			A							
21309	G			A							
21323	A					G					
21803	A					M					
21832	A					M					
21838	A			T		T					
21865	G			A		A					
A1 isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
20399	T									G	
20400	G									A	
20429	C					Y		T			
20430	T									A	
20440	C									T	
20443	T									C	
20485	T									G	
20495	C									A	
20496	A									G	
20497	T									C	
20503	C							M			
20520	T									C	
20521	A									G	
20526	A									T	
20540	C									A	
20590	T									C	
20604	A									T	
20910	G					A					
20914	A					G					
20918	A					W					
20925	A					C					
20932	G					T					
20940	T					A					
20947	G					T					
20952	T					A					
21746	C							T			
21765	C							G			
21767	A							T			
21793	C							M			
21908	C							G			

The failure of the Sanger sequencing reads between ep20-40 in A1 isolate prevented determination of the polymorphism in the early stages of egg passages (table 3.6). However, the investigation of ep50, ep70 and ep90 showed only one case of SNP continuation within the egg passages at nucleotide position 20429 with minor SNP (ep50) and change at consensus sequence level (ep70). Furthermore, the A1 isolate was also analysed in the light of SNP clusters showing the highly polymorphic encoding region (eight SNPs) between nucleotide positions 20910 and 20952 in ep50 separated by 480 bp from the nearest independent SNP. In addition, the confirmed SNP cluster presented more variable SNPs compared to the highly polymorphic encoding region identified in ep10 of the A isolate. The analyses of Sanger sequenced data showed that SNPs involving pairs of A or G and T or A residues accounted for 50 % in a SNP cluster (ep50 of the A1 isolate). In addition, this particular SNP accumulation was identified as being mostly SNPs at consensus sequence level with the exception of the nucleotide position 20918 shown as minor SNP (presented with ambiguity code). Furthermore, ep70 presented a SNP cluster composed of five SNPs between nucleotide positions 21746 and 21793 with no additional polymorphic genomic positions within 115 bp. The SNP cluster (ep70) was identified in 75 % as the consensus sequence SNPs with the exception to nucleotide position 21793 defined as a minor SNP (found as IUPAC Ambiguity Code). Furthermore, the phenomenon of SNP cluster in the viral genome was confirmed in ep90 with 15 consensus sequence SNPs detected between nucleotide positions 20399 and 20604. Interestingly, the ratio of purines and pyrimidines was calculated as eight to seven showing an equal share relating to two groups of nitrogenous bases within SNP cluster. Furthermore, this large number of SNPs was identified only in ep90. The further analyses of Sanger sequenced data could not confirmed differences in SNP accumulation between the egg isolates due to missing data for late passages of A isolate and early passages of A1 isolate. However, these Sanger sequenced datasets suggest that a degree of SNP accumulation can random and its presence in egg passages strongly depends on analysed egg isolate. The investigation of independent SNPs and SNP clusters in viral genome detected a similarity of SNP appearance regarding to a small number of independent SNPs in the A and A1 isolate and their accumulation in further egg passages.

Interestingly, a similar observation was made in the C isolate where the appearance of SNP clusters and independent SNPs were analysed in table 3.7. The ep20 of the C isolate was found as a mixture of one independent SNP and SNP cluster separated from each other by 577 bp. The highly polymorphic encoding region (11 SNPs) formed by consensus sequence SNPs was identified between nucleotide positions 21084-21137 (53 bp). In addition, a very unique situation was spotted within the SNP cluster between nucleotide positions 21210-21265 with an insertion of three nucleotides (G, T and A) in the C isolate. The nature of the highly polymorphic encoding regions in ep20 was also confirmed in ep40 as SNP cluster (four SNPs) shown between nucleotide positions 21272-21350 with 533 bp gap to the next independent SNP. The investigation of ep60 has shown no SNP cluster formation across the S1 subunit encoding region but identified independent SNPs. Interestingly, ep90 was confirmed as the egg passage that included a small polymorphic encoding region formed of three neighbourhood residues identified between genomic positions 21165-21167 with a residue switch (A and T nucleotide) at nucleotide positions 21165 and 21166. The highly polymorphic encoding regions were also confirmed as two clusters of eleven and five consensus sequence SNPs between nucleotide positions 20451-20491 and 20595-20627 in ep80. In addition, similar to the SNP cluster of ep90 (A1 isolate), the purines and pyrimidines in two SNP clusters of ep90 was balanced at the ratio of five to six and two to three. Furthermore, the observation confirmed ep80 as the only egg passage without independent SNPs in C isolate and the biggest SNP accumulation in a form of cluster identified in the S1 subunit encoding region. Further investigation showed that SNPs and their features strongly depend on the analysed IBV isolates and their egg passages. However, the unique SNPs found in egg passages of C isolate mirror the situation of SNP clusters and independent SNPs analysed in the early egg passages of A isolate and in the later egg passages of A1 isolate. Further investigation confirmed lack of correlation between single SNPs in four analysed isolates but showed a pattern of SNP cluster appearance in the investigated egg passages of A, A1 and C isolate.

Table 3.7. Results of Sanger sequencing reads relating to the additional SNPs and SNP clusters found in S1 subunit of C isolate after indicating to the sequence of ep4. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP, green = SNP at consensus sequence level and NIL = gap.

C isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
20448	T						A				
20451	C								G		
20452	G								C		
20460	C								T		
20461	T								G		
20464	C								T		
20465	A								G		
20468	G								A		
20481	A								T		
20483	C								A		
20485	A								C		
20489	G						R				
20491	T								C		
20595	C								T		
20600	T								C		
20602	T								G		
20625	A								G		
20627	G								C		
21084	T		C								
21103	C		G								
21116	G		T								
21124	G		A								
21129	G		C								
21130	C		A								
21131	A		T								
21137	T		A								
21165	T									A	
21166	A									T	
21167	C									A	
21210	NIL		G								
21224	NIL		T								
21265	NIL		A								
21272	G				A						
21280	T				A						
21288	T				A						
21350	G				A						
21490	T									W	
21838	A				T					T	
21839	A									T	
21842	A		T								
21853	G						T				
21865	T						A				

Table 3.8. The S1 subunit of D isolate with additional SNPs not identified previously with 454 pyrosequencing reads indicated to the sequence of ep4. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP, green = SNP at consensus sequence level and NIL = gap.

D isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
20357	T						A				
20358	A						T				
20488	G					R					
20964	C									A	
20965	T									C	
21254	A						C				
21270	G					A					
21329	A						T				
21342	A						G				
21343	T						A				
21350	A					C					
21373	G					A					
21606	C				T						
21607	NIL				T						
21617	C				A						
21619	G				C						
21627	A				C						
21633	A				C						
21634	G				A						
21635	T				A						
21637	A				T						
21822	T						G				
21825	A						C				
21831	T						G				
21832	A					M					
21853	A				T						
21859	A						T				
21867	C						G				
21895	A					M					
21980	G				A						
21986	G						A				

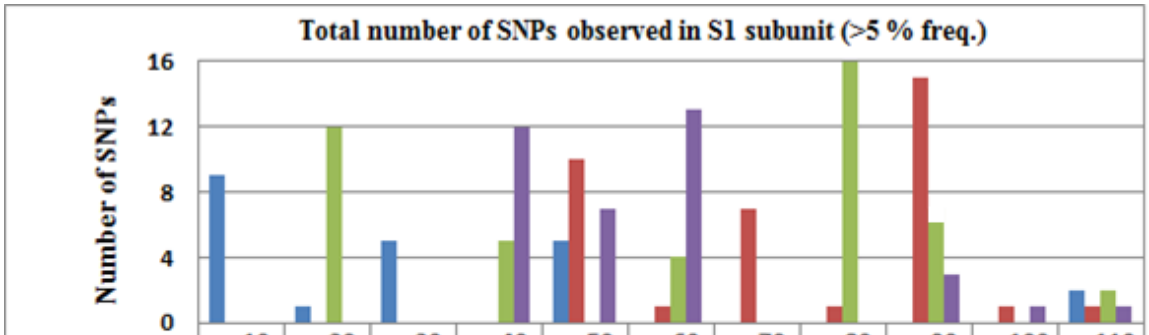
The constantly changing composition of the viral genome was also investigated as the SNP dynamics in the D isolate. The analysed Sanger sequenced datasets showed a mixture of SNP clusters with various lengths and independent SNPs across the S1 subunit encoding region (table 3.8). Interestingly, the largest SNP cluster was shown in ep40 with ten SNPs at consensus sequence level in the viral genome between nucleotide positions 21606-21637. Furthermore, the analyses of SNP composition in this highly polymorphic encoding region showed 40 % of C residues and 30 % of each A and T residues across the investigated ep40. In addition, this particular SNP cluster is the only cluster where G

residue is not observed making it unique among SNP active encoding regions in four egg isolates. The SNP cluster formation usually involved a minor SNPs or/and SNPs at consensus sequence level but insertions of additional SNPs in investigated Sanger sequenced egg passages were rare. However, nucleotide position 21607 was identified with insertion of a T residue in the SNP cluster visible in ep40 (D isolate) causing a frameshift mutation. The insertion in viral genome could possibly be responsible for the attenuation process by changing the reading frame resulting in a completely different translation from the original. The Sanger sequence reads identified independent SNPs in ep50-60 without further continuation of genomic change in later egg passages. Interestingly, further investigation also analysed the appearance of SNPs in ep60 resulting in identification of two SNP clusters combined with three and six consensus sequence SNPs between nucleotide positions 21329-21343 and 21822-21986. Furthermore, ep70 and ep90 were an exception among Sanger sequenced egg passages of the D isolate showing the low amount of polymorphism previously undetected with 454 pyrosequencing in ep110. The further analyses has shown unique situation relating to SNP appearance in S1 subunit encoding region without further accumulation (SNP clusters and independent SNPs) between ep40-60 not identified in investigated A, A1 and C isolate. However, similar to previously analysed egg isolates, the D isolate has shown a number of independent SNPs with random appearance across the S1 subunit but also SNP clusters without further accumulation in further egg passages.

3.5.4 The summary of SNP accumulation in S1 subunit

The total number of SNPs identified within four isolates of S1 subunit across all investigated egg passages at >5 % frequency was included in figure 3.5. The Sanger sequenced data clearly shows differences in SNP accumulation but also in the number of genomic changes in egg passages of A, A1, C and D isolate which might be involved in the process of attenuation. According to the bar chart below C isolate was shown as one of the most polymorphic across all investigated isolates. Interestingly, Sanger sequenced data identified rapid SNP accumulation in a form of visible peak in ep20 (12 SNPs) followed by the reduction of SNPs in ep40 (five SNPs) - ep60 (four SNPs) and rapid SNPs accumulation in a case of ep80 (16 SNPs) in C isolate. Interestingly, C isolate was found as only one across four egg isolates with visible polymorphism in ep20 and ep80 and

decreased number of SNPs in ep30-70. The D isolate was identified with polymorphism between ep40-60 where analyses of Sanger sequence data observed SNP accumulation in ep40 (12 SNPs) and ep60 (13 SNPs).



	ep10	ep20	ep30	ep40	ep50	ep60	ep70	ep80	ep90	ep100	ep110
A isolate	9	1	5	0	5	0	0	0	0	0	2
A1 isolate	0	0	0	0	10	1	7	1	15	1	1
C isolate	0	12	0	5	0	4	0	16	6	0	2
D isolate	0	0	0	12	7	13	0	0	3	1	1

Figure 3.5. The total number of SNPs across all investigated egg isolates with the final reads of ep110 (>5 % freq.). The diagram presents number of genomic mutations in each of analysed egg passage. The bar chart illustrates polymorphism across four egg isolates presented by different colours. The table below bar chart presents exact number of SNPs identified in egg passage and egg isolate of interest. The zero numbers marked as red colour relates to failure of Sanger sequencing reads.

Interestingly, D isolate was characterised by rapid decreasing of SNPs in ep70 with a continuous trend of small numbers of polymorphisms in later egg passages such as ep90 (three SNPs). Similar to D isolate, A1 isolate was found with SNP accumulation only in ep50 (10SNPs), ep70 (seven SNPs) and ep90 (15SNPs). Interestingly, further observation identified the phenomenon of rapid SNPs accumulation in ep90 (A1 isolate) and ep80 (C isolated) as the highest number of genomic changes ever observed in S1 subunit of single egg passage. The lowest SNP accumulation across all investigated egg passages was seen in the A isolate. However, Sanger sequenced data clearly showed the A isolate with rapid SNP accumulation in ep10 (nine SNPs). Furthermore, intensive polymorphism in ep10 was not continued in further egg passages showing decreasing tendency of genomic changes in ep30 and ep50 (five SNPs) and lack of ability to investigate polymorphisms in later egg passages due to failure of Sanger sequencing reads. Interestingly, Sanger sequenced data of S1 subunit clearly shows a tendency of at least two visible bars due to significant SNP

accumulation in four egg isolates. In addition, each egg isolate has been identified with rapid accumulation but also quick reduction of SNPs across all investigated egg passages such as A1 and C isolates presented intensive polymorphism in ep90 comparing to SNP reduction in A and D isolates. Pyrosequencing data of S1 subunit clearly identified all four egg isolates with similar numbers of mutations in ep110 suggesting that additional SNP appearance in previous egg passages do not have an influence on number of SNPs in ep110 of S1 subunit.

3.5.5 The SNPs investigation in S2 subunit (part 1)

The 454 pyrosequencing data showed that the first 982 bp of the S2 subunit encoding region presents high polymorphism in all four isolates. Therefore, further investigations included Sanger sequenced egg passages to analyse the occurrence of SNPs within the viral populations. The studies of known wild IBV strains show the S2 subunit, as more conserved and larger part of spike encoding region (built up with 1855 bp) compared to the S1 subunit found as highly polymorphic within its size of 1607 bp. However, the analyses of pyrosequencing data identified the D isolate with seven SNPs with various frequencies in the viral population of the S2 subunit compared to three SNPs across the S1 subunit encoding region. A similar observation was made in the A1 isolate detected with three SNPs relating to the S2 subunit and only one SNP identified in the S1 subunit. This observation suggests unusual occurrences in viral genome where, in lack of immune pressure, the conserved region of S2 subunit was identified with more mutations than S1 subunit which is highly polymorphic under immune pressure.

The results of the 454 pyrosequencing data also confirmed the presence of common SNPs shared between investigated egg isolates (table 3.9). This situation is clearly visible at nucleotide position 22224 where a synonymous SNP was identified in the C and D isolate with >50 % SNP frequency in ep110. Interestingly, only D isolate (no data of ep10-100 for C isolate) showed minor SNPs which in ep70-80 returned to the original nucleotide found in ep4 but was confirmed as SNPs at consensus sequence level in ep110 with a SNP frequency of 56 %. A common residue change was also observed at nucleotide position

22421 in the A and A1 isolate where the pyrosequencing data identified both egg isolates with a SNP at consensus sequence level with vary SNP frequency in ep110.

Table 3.9. The SNPs found in more than one egg isolate (S2 subunit) with specification of mutation NS/S (Non-synonymous/Synonymous) after indicating to the sequence of ep4. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

The nucleotide position 22224															
NS/S	**Ep # / Egg isolate**	**4**	**Freq %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
S	**C**	G	80											T	52
S	**D**	G	88		K		K		K	G	G			T	56
The nucleotide position 22240															
NS/S	**Ep # / Egg isolate**	**4**	**Freq %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
S	**C**	G	100											T	58
S	**D**	G	100								T	T	T	T	52
The nucleotide position 22421															
NS/S	**Ep # / Egg isolate**	**4**	**Freq %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
S	**A**	C	100			M		M	M	M	A	A		A	52
S	**A1**	C	100					A	A	A				A	98

Interestingly, the number of SNP appearing in more than one egg isolate of S2 subunit (three SNPs) was similar to the S1 subunit encoding region (two SNPs). However, the SNP frequency was less diverse in S2subunit (52-98 %) comparing to S1 subunit (1.1-100 %) which might indicate a substantial impact on SNP accumulation in Sanger sequenced egg passage. Furthermore, the high SNP frequency observed in deep sequenced data of ep110 seems to have an influence on the degree of SNP accumulation presenting minor SNP which gradually become a SNP at consensus sequence level or accumulate directly at consensus sequence level in earlier egg passages.

Further analyses also confirmed a number of SNPs that occurred only in one of the investigated isolates **(**shown in a table 3.10). The relatively high SNP frequency (54 %) was obtained from nucleotide position 22459 in ep110 of the D isolate showing a non-synonymous SNP (leucine with isoleucine) in ep70 and a minor SNP (represented with ambiguity code) in ep80-100 across viral population. The A isolate was identified with two independent SNPs at nucleotide positions 22285 and 22504 across the analysed S2 subunit encoding region with a ≤50 % SNP frequency in ep110. Further investigation into

SNPs using 454 pyrosequencing also showed four synonymous SNPs with a very low frequency (≤2.2 %) detected only in ep110 of the D isolate. Interestingly, the SNP with frequency ~50 % in ep110 are visible as a mixture of minor SNPs and SNPs at consensus sequence level within regular Sanger sequenced egg passages. However, the SNPs accumulation not always seems to be logical showing sometimes the SNP at consensus sequence level with a change to minor SNP in later egg passages (D isolate – nucleotide position 22459).

Table 3.10. The outcomes of unique SNP investigation in A, C and D isolate from Sanger sequenced egg passages and final results of 454 pyrosequencing in S2 subunit (part 1). The specification of mutation NS/S (Non-synonymous/Synonymous) was indicated to the sequence of ep4. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

A isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
NS	**22285**	G	100							K				T	50
NS	**22504**	G	100											A	5
C isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
NS	**22201**	T	100											A	4.6
D isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq %**
NS	**22459**	C	100							A	M	M	M	A	54
S	**22476**	T	100											A	2.2
S	**22633**	C	100											T	1.5
S	**22696**	C	100											T	1.7
S	**22717**	T	100											C	1

3.5.6 Other polymorphism within S2 subunit (part 1)

Further analyses of the S2 subunit (part1) encoding region showed SNPs previously not detected using 454 pyrosequencing data in ep110. In addition, a substantial number of SNPs investigated in the S2 subunit were common for more than one egg isolate (table 3.11).

Table 3.11. The shared SNPs identified in S2 subunit (part1). The information included reference nucleotide position in ep4, a SNP change in each of egg isolates and the details of detected SNPs. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

The nucleotide position 21994											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	A			R				R			
A1	A			R							
D	A		R							R	
The nucleotide position 22002											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	T							K			
D	T								K		
The nucleotide position 22006											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	A							R			
A1	A			R							
D	A						R				
The nucleotide position 22048											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	G							R			
A1	G					R					
C	G		A								
D	G								R		
The nucleotide position 22051											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	G							R			
A1	G					R					
The nucleotide position 22117											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	G						R	R	R	R	
A1	G					R					
D	G		R				R				
The nucleotide position 22162											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	G			R	R	R	R	R	R		
D	G				R	R	R	R	R		
The nucleotide position 22561											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	C			Y	Y	Y	Y	Y	Y	Y	
A1	C		Y		Y	Y					
D	C						Y	Y	Y	Y	Y
The nucleotide position 22768											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A1	T		C	C						C	
C	T		C								
D	T		C		C		C		C		

Interestingly, the A isolate was found to have the highest number of SNPs commonly identified between the four egg isolates. Further investigation showed that out of 27 SNPs identified in the A isolate, eight SNPs were also confirmed within the A1, C or D isolate. An interesting example is the SNP at nucleotide position 22048 within the SNP clusters of A isolate (ep70) and C isolate (ep20) also identified as independent SNP in A1, and D isolate. The generated Sanger sequence data confirmed a similarity between the A, A1 and D isolate relating to common SNPs in the viral population such as SNP detected at nucleotide position 22006, clearly visible in a SNP cluster of the A isolate (ep70), but also as an independent minor SNP in the D isolate (ep60) and the A1 isolate (ep30). The common SNPs were also observed at nucleotide position 22117 as a minor SNP found in the A isolate (ep60-90), the A1 isolate (ep50) and D isolate (ep20, ep60). The analyses of Sanger sequence data identified the phenomenon of common SNPs when comparing the A and D isolate at nucleotide position 22002 where a minor SNP (represented with ambiguity code) was observed in ep70 of the A isolate and ep80 of the D isolate. Interestingly, the majority of SNPs identified across A, A1, C and D isolate have been shown as independent SNPs with appearance in different Sanger sequenced egg passages for each egg isolate. Furthermore, there is also a small correlation between SNPs observed at this same nucleotide position showing SNP continuation within a few egg passages in more than one egg isolate.

The further investigation of egg isolates identified a number of SNPs that were unique to each viral population, identified as a SNP cluster or as independent SNPs in the S2 subunit (part1) encoding region. The analyses of sequence data in the A isolate showed no additional SNPs in ep20 but confirmed independent SNPs in ep30 and ep50 without further continuation in later egg passages (table 3.12). Interestingly ep40 and ep60 were identified with no additional SNP or/and SNP clusters. However, the later egg passages such as in ep80 was shown with SNP cluster (six SNPs at consensus sequence level) identified between nucleotide positions 22350 and 22370 without continuation in further egg passages. Interestingly the A isolate has shown SNP accumulation in early egg passages of S1 subunit comparing to polymorphism found in later egg passages of S2 subunit supported by the appearance of SNP cluster and a number of independent SNPs.

Table 3.12. The unique SNPs analysed in S2 subunit (part 1) of A, A1and D isolate. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

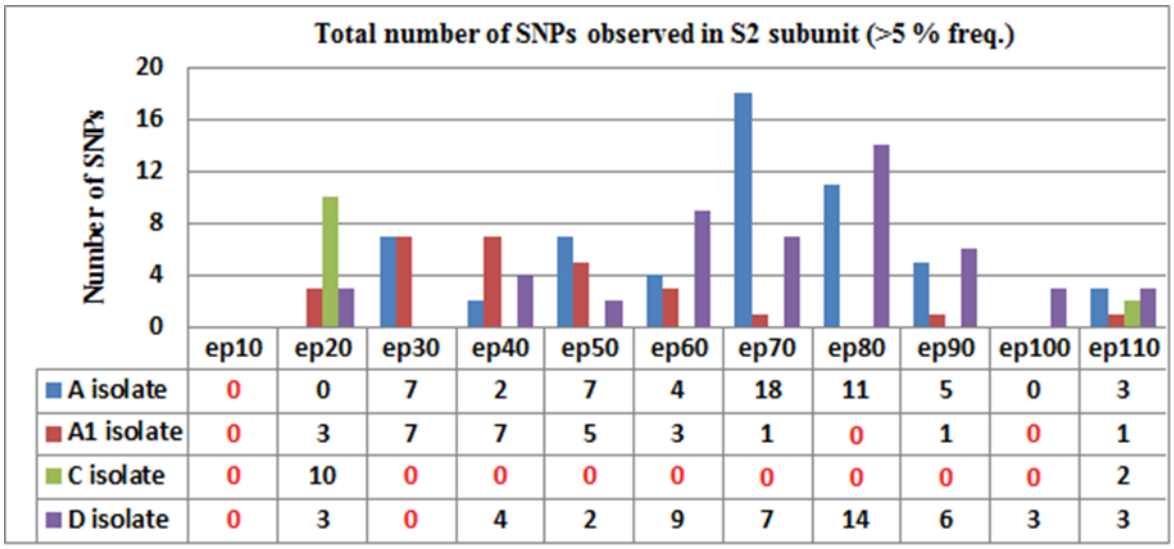
A isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
21994	G							R			
22002	T							K			
22006	A							R			
22048	G							R			
22051	G							R			
22117	G							R			
22162	G							R			
22285	G								T		
22312	G			R							
22350	G								T		
22354	G								A		
22358	T								A		
22365	A								T		
22370	C								T		
22426	G					A			A		
22430	G					A					
22455	C					Y					
22486	G			K							
22540	C			Y							
22593	C									Y	
22643	C									Y	
22738	T					C					
22766	C							T			
A1 isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
22242	T				A						
22243	T				A						
22272	C				A						
22295	G				T						
22343	C						G				
22441	C			T							
22442	G			A							
22443	T			A							
22448	T				A						
22462	C			T							
22616	C				T		T				
22769	C									G	
D isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
22002	T								K		
22048	G								R		
22162	G				R	R	R	R	R		
22388	C									A	
22436	G							T			
22459	C						S		S		
22545	C								Y		
22561	C						Y	Y	Y	Y	Y

According to the analysed Sanger sequenced data the unique SNPs were identified in early egg passages of the A1 isolate showing ep30 with four consensus sequence SNPs identified between nucleotide positions 22441-22462 (table 3.12). Interestingly, the SNP accumulation in this encoding region was equally dominated by T and A residue appearance in the SNP cluster. In addition, a similar conclusion was made in ep40 with four consensus sequence SNPs between nucleotide positions 22242-22295 where A and T residues accounted for 100 % of SNPs. Furthermore, SNP cluster in ep40 was separated by 150 bp gap from two consensus sequence SNPs found at nucleotide positions 22448 and 22616. Interestingly, further investigation showed lack of unique SNP in ep50 of the A1 isolate and decreasing polymorphism in ep60 (two SNPs) and in ep90 (one SNP) across S2 subunit encoding region. The analysed sequence data clearly showed the differences in additional SNP accumulation and continuation between the A isolate (ep70-80) and A1 isolate (ep30-40) in S2 subunit. The difference was also spotted between S1 and S2 subunit of A1 isolate presenting different level of SNP appearance within Sanger sequenced egg passages relating to the SNP clusters and independent SNPs.

The high polymorphism was detected in the C isolate (ep20) shown as SNP cluster (seven consensus sequence SNPs) between nucleotide positions 21999 and 22105 separated by a 540 bp gap from the next independent SNP in S2 subunit (part1) – data not shown. Further investigation clearly showed an equal domination of A and G residues confirmed to be 84 % within the SNP cluster. However, further investigation of polymorphism in the C isolate could not be continued due to the failure of the Sanger sequencing reads in all egg passages. The analysed sequence data identified in ep20 and ep40 of D isolate a lack of polymorphisms with an exception of nucleotide position 22162 (ep40-80) detected as a minor SNP (represented with ambiguity code-R). The further investigation of Sanger sequencing reads clearly indicated genome changes in later egg passages of the D isolate by increasing numbers of SNPs in ep60, with three minor SNPs randomly spread across the investigated S2 subunit (part1) encoding region. Interestingly, the SNP detected at nucleotide position 22561 was also confirmed with accumulation in ep70-100. The further investigation identified ep80 as highly active egg passage presenting unique minor SNPs without further continuation. Furthermore, the investigation of SNPs previously not detected using 454 pyrosequencing data showed decreasing polymorphism in ep90. The SNP accumulation of D isolate in the later egg passages was similar to the observation made in the A isolate presented with polymorphism in ep70-80 across S2 subunit encoding

region. In addition, the pattern of SNP accumulation was also spotted between S1 and S2 subunit of D isolate shown as highly active only in later egg passages. Furthermore, the investigation of the D isolate identified unique SNP continuation (minor SNPs) at two nucleotide position (22162 and 22561) in Sanger sequenced egg passages never seen before in previous egg isolates including ep110.

3.5.7 The overview of polymorphism across S2 subunit (part 1)



	ep10	ep20	ep30	ep40	ep50	ep60	ep70	ep80	ep90	ep100	ep110
A isolate	0	0	7	2	7	4	18	11	5	0	3
A1 isolate	0	3	7	7	5	3	1	0	1	0	1
C isolate	0	10	0	0	0	0	0	0	0	0	2
D isolate	0	3	0	4	2	9	7	14	6	3	3

Figure 3.6. The total number of SNPs across all investigated egg isolates in S2 subunit (part 1) with the final reads of ep110 (>5 % freq.). The diagram presents number of genomic mutations in each of analysed egg passage. The bar chart illustrates polymorphism in a form of bars across four egg isolates presented by different colours. The table below bar chart presents exact number of SNP identified in egg passage (ep10-110) and egg isolate of interest. The zero numbers marked as red colour relate to failure of Sanger sequencing reads.

The Sanger sequence data has shown crucial differences in accumulation of SNPs across egg passages in S1 and S2 subunits. Interestingly, A1 egg isolates has shown visible and rapid accumulation of mutation in ep20-30 comparing to A and D isolate identified with accumulation but also reduction of SNPs within ep20-60 (figure 3.6).

Interestingly, Sanger sequenced data has shown a phenomenon of increasing polymorphism (A isolate) and decreasing number of mutations (D isolate) or vice versa within ep20-60. A1 isolate has been found with growing SNP numbers in ep20-30. Interestingly, the number of SNPs in ep30-40 of A1 isolate did not change (seven SNPs) but started to decrease continuously presenting one SNP in ep90. The further investigation of genomic changes in A, A1 and D isolate have shown ep60 as a crucial point where polymorphism change rapidly and strongly depends on egg isolate. The most visible SNP accumulation was identified in ep70 of A isolate (18 SNPs) but also ep80 of D isolate (14 SNPs). However, these two visible bars were continued as rapid reduction of polymorphism followed by detection of only three SNPs in ep100 (D isolate) and lack of any SNPs in ep100 (A isolate). Interestingly, further analyses of pyrosequencing data have shown similar level of SNP accumulation in ep110 varies from one SNP (A1 isolate) to three SNPs (D isolate).

The comparison of two spike subunits could not be fully possible due to failure of Sanger sequencing assay in S1 subunit (few egg passages of A, A1, C and D isolates) and S2 subunit (C isolate). However, the figure 3.5 and figure 3.6 clearly showed the differences of SNP accumulation in egg passages between two subunits of spike encoding region. S1 subunit was found with visible bars presented during investigation of egg isolates without an impact relating to further accumulation/reduction of SNPs shown as crucial point in ep60 of S2 subunit. In addition, constant changes in number of SNPs across spike encoding regions were more visible in S1 subunit. However, the total number of genomic mutations was greater in S2 subunit presenting increased polymorphisms in ep70 (A isolate) and ep80 (D isolate). Interestingly, both subunits of spike encoding region were identified with decreasing polymorphisms in ep100 across four egg isolates. In addition, the situation of ep110 presenting higher number of SNPs in S2 subunit comparing to S1 subunit pyrosequencing data does not explain the sudden reduction of SNP numbers in later egg passages of spike encoding region. According to the results the highest polymorphism was observed in A and D isolate (three SNPs) and the lowest in A1 isolate (one SNP) of ep110.

3.5.8 The common SNPs found only in the nucleocapsid

Further investigation indicated SNPs in nucleocapsid encoding region of pyrosequencing data (ep4) at <43 % SNP frequency and Sanger sequenced egg passages (table 3.13). Interestingly, the SNPs with a low frequency were identified in ep110 of four egg isolates at the nucleotide position 26792 and 26793 also visible in most of investigated Sanger sequenced egg passages (A, C, D isolate) and ep70-90 of A1 isolate (nucleotide position 26792). Furthermore, the SNP frequency of ep110 varied across investigated egg isolates from ~20 % in A and A1 isolate to ~40 % in C and D isolate. In addition, a similar situation was detected at nucleotide position 26793 identified in the C and D isolate as having 31-36 % SNP frequency in ep110 whereas A and A1 isolate showed a decreased SNP frequency (17-26 %). Interestingly, the low frequency (>43 %) identified in ep110 (nucleotide position 26792-26793) might be a sign of error in the 454 pyrosequencing data due to presence of poly A tract in the viral genome. The further analysis across the nucleocapsid indicated the presence of SNPs which are visible in ep4 at low SNP frequency (<30 %) such as the nucleotide position 26795 with 15 % SNP frequency in ep4 of A1, C and D isolate. There was a lack of SNPs in Sanger sequenced egg passages except for deep sequenced ep110 identified with the increasing SNP frequency in D isolate (36 %) and C isolate (23 %) but stable in A1 isolate (15 %) compared to 15 % SNP frequency rate in ep4. Interestingly, the nucleocapsid encoding region has shown a SNP accumulation with SNP frequency (<43 %) never seen before in the spike encoding region. The polymorphism could be identified in all previous egg passages with a SNP frequency as low as ~30 % in all investigated egg isolates. These data presented unusual accumulation of SNPs with low SNP frequency at genomic positions in more than one egg isolate.

Table 3.13. The common SNPs appearance within the egg isolates in the nucleocapsid encoding region with SNP frequency ≤ 33 % in ep4. The specification of mutation NS/S (Non-synonymous/Synonymous) was included. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

The nucleotide position 26791														
NS/S	**Ep # / Egg isolate**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**A**	C	2.9											C	8.7
S	**A1**	C	2.6											C	13.1
S	**D**	C	2.6											C	8.6
The nucleotide position 26792														
NS/S	**Ep # / Egg isolate**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**A**	G	31	G	G	G		G	G	G	G	G	G	G	29
S	**A1**	G	31							G	G	G		G	20
S	**C**	G	31	G	G	G	G			G	G	G	G	G	40
S	**D**	G	31		G	G	G	G	G	G	R	G	G	G	43
The nucleotide position 26793														
NS/S	**Ep # / Egg isolate**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**A**	A	33	A	A	A		A	A	A	A	A	A	A	26
S	**A1**	A	33							A	A	A		A	17
S	**C**	A	33	A	A	A	A			A	A	A	A	A	36
S	**D**	A	33		A	A	A	A	A	A	A	A	A	A	31
The nucleotide position 26795														
NS/S	**Ep # / Egg isolate**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**A1**	C	15											C	15
S	**C**	C	15											C	23
S	**D**	C	15											C	36

The next interesting observation was made at nucleotide position 27124 where SNPs in viral genome were observed during the egg passages in D isolate (table 3.14). This phenomenon was observed only once across the nucleocapsid encoding region with a substantial difference in frequency between ep4 (1.2 %) and ep110 (100 %) also detectable as consensus sequence SNPs in ep70-100. However, during the investigation of the SNPs within the viral populations, the high SNP frequency in ep110 mirror the situation of SNPs appearance between egg passages of the A and D isolate. A good example is the non-synonymous SNP analysed in the A isolate at nucleotide position 26845 where, with 99 % frequency in ep110, the minor SNP was identified in ep50 with further SNP continuation (ep60-90) at consensus sequence level across the viral population. Further investigation into the D isolate led to a very similar observation, confirmed as a non-synonymous minor

SNP at nucleotide position 26006 detected in ep40-70 with change at consensus sequence level in ep80 and 99 % SNP frequency in ep110.

Table 3.14. The SNPs in the nucleocapsid encoding region of A, C and D isolate relating to SNP frequency (25-100 %) in ep110. NS/S = Non-synonymous/Synonymous. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

A isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**26312**	G	100											A	60.3
S	**26394**	C	100							Y	Y	T		T	98
NS	**26557**	T	100						C	C	C	C		C	99
NS	**26729**	C	100	Y		Y		Y	T	T	Y			T	40.8
NS	**26845**	C	100					Y	T	T	T	T		T	99
C isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**26213**	G	100		A								R	A	25
NS	**26224**	G	100										K	T	62
NS	**26563**	A	100										G	G	44
D isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**26006**	G	100				R	R	R	R	A	A	A	A	99
NS	**27022**	A	100								M	C	C	C	86.8
NS	**27124**	A	1.2							A	A	A	A	A	100

The investigation of pyrosequenced data has shown that SNPs appearing within the quasispecies of Sanger sequenced egg isolates can be identified with decreased SNP frequency in ep110. The deep sequence data presented genomic positions in nucleocapsid encoding region where the SNP frequency varies (25-100 %) as seen in table 3.14. The A isolate perfectly illustrates the dynamics of the viral genome regarding to non-synonymous SNPs at nucleotide position 26729 detected in early egg passages as the minor SNP with a change at consensus sequence level in ep60-70 bur relatively low SNP frequency in ep110 (40.8 %). The interesting fact is that the non-synonymous SNP at nucleotide position 26312 was undetected in Sanger sequenced egg passages but was identified with a relatively high frequency (60.3 %) in ep110. The C isolate was found with mutation at nucleotide position 26224 where a non-synonymous SNP was detected in ep100 (IUPAC Ambiguity Code-K) resulting in change at consensus sequence level presented by 62 %

frequency rate in ep110. However, the further SNPs identification in previous egg passages is an open question due to the failure of the Sanger sequencing reads. The analysed Sanger sequenced data also identified interesting SNPs at nucleotide position 27022 in the D isolate observed in ep80 as minor SNP, ep90-100 with a change at consensus sequence level and the final confirmation in ep110 (freq.86 %).

The 454 pyrosequencing data of A and A1 isolates also showed a number of SNPs with low frequency (<20 %) where in most cases the SNPs were undetectable during the Sanger sequenced egg passages within all four egg isolates. The A isolate presents four different nucleotide positions as a mixture of synonymous and non-synonymous SNPs with a low SNP frequency (≤ 5.8 %) detected only in ep110 (table 3.15).The further investigation also confirmed two low (≤1.5 %) SNP frequency identified within the viral population of ep110 in A1 isolate. Further investigation of Sanger sequenced data identified a number of genome positions in the A isolate such as nucleotide position 26002 where a minor non-synonymous SNP (represented with ambiguity code) was confirmed in all investigated passages with a 6.9 % SNP frequency in ep110. The D isolate presents the only one SNP at nucleotide position 26064 identified as minor SNP in ep80 with low SNP frequency (<5.2 %) in ep110. The 454 pyrosequencing data has shown different level of SNP accumulation in ep110 of four egg isolates. The further investigation using regular Sanger Sequencing method has shown that high SNP frequency (>85 %) presents a regular and logical SNP accumulation in previous egg passages mostly in a form of minor SNPs which at some point become the SNP at consensus sequence level across nucleocapsid encoding region. Furthermore, the decreasing SNP frequency in ep110 has shown a dynamic but very often random SNP accumulation in Sanger sequenced egg passages. Interestingly, the SNP frequency (<60 %) in ep110 does not ensure appearance of SNPs in the viral population of previous egg passages. In addition, nucleotide positions with SNP frequency (>15 %) were mostly identified as too low for SNP detection using the Sanger sequencing method. However, the exceptions are present suggesting that a viral population might develop a minor SNP in the genome which stays at constant level during the Sanger sequenced egg passages without further increasing of accumulation.

Table 3.15. The reduced polymorphism in the nucleocapsid encoding region of four egg isolates after Sanger sequencing and Pyrosequencing. NS/S = Non-synonymous/Synonymous was included. White = not sequenced, yellow = failure of sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

A isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**26002**	G	100	R	R	R	R	R	R	R	R	R	R	A	6.9
NS	**26140**	C	100							Y				T	14.9
S	**26331**	G	100											A	1.3
NS	**26398**	C	100											A	2.2
NS	**26845**	C	100											T	5.8
NS	**26919**	A	100											C	1.9
A1 isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
NS	**26291**	C	100											T	1.4
NS	**26347**	C	100											A	1.5
C isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**26328**	C	100											T	1.5
S	**26361**	C	100											T	1
D isolate															
NS/S	**Ep # / Nucl. #**	**4**	**Freq. %**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**Freq. %**
S	**26064**	T	100								Y			C	5.2

3.5.9 The additional SNPs of nucleocapsid encoding region

The further investigation of previous egg passages using Sanger sequencing highlighted a substantial number of SNPs across the nucleocapsid encoding region previously not detected in the 454 pyrosequencing data. The uncovered SNPs were identified in the C isolate (133) but also A isolate (38), D isolate (36) and within A1 egg passages (15). In most cases the SNPs detected across the investigated isolates were an independent SNPs identified only once within one isolate. However, a small fraction of SNPs showed as continued SNPs in the viral genome clearly visible in a Sanger sequenced egg passages of more than one egg isolates (table 3.16). The interesting sequence data was analysed at nucleotide position 25924 identified as SNP visible in the A isolate (ep20-100), D isolate (ep20-30, ep50-100), A1 isolate (ep50), and C isolate (ep100). In addition, a very similar

situation of SNPs identified in all four egg isolates was spotted at genome position 26101 and 26772 identified as a highly polymorphic with a different degree of SNP accumulation within four egg isolates. Further investigation of regular Sanger sequencing data also identified the polymorphism only in egg passages of the A, C and D isolate across the nucleocapsid such as the nucleotide position 26948 with the minor SNP (presented as ambiguity code) identified in the D isolate (ep20-70, ep90-100), A isolate (ep70-90), and C isolate (ep40). The polymorphism was also identified at nucleotide position 27128 with various degrees of SNP accumulation within the A, A1 and D isolate due to replacement of the C residue with a T residue forming a consensus sequence SNPs (A isolate) or mixture of minor SNPs (A1 and D isolate) within the viral population. In some cases the mutation in viral genome was detectable only in two egg isolates such as the A and C isolate with SNPs at nucleotide positions 26067 and 26068. Interestingly, SNP identification at nucleotide position 26068 confirmed unusual appearances of two nucleotides relating to a change of a C residue in ep30 of the A isolate (G residue) and the replacement in ep60 of the A1 isolate (T residue). The overlapping SNPs were also found as common for A and D isolate showing minor SNPs accumulation in Sanger sequenced egg passages identified at the nucleotide position 26887. These sequence data shows that some SNPs could not be detected using 454 pyrosequencing reads because of their disappearance before the final investigation of ep110. In addition, analysed Sanger sequenced data clearly presents the dynamics at common nucleotide positions with consensus sequence SNPs visible in early egg passages and a logical continuation of these SNPs in more than one investigated egg isolate. However, there are a few nucleotide positions in a nucleocapsid encoding region with a polymorphism showing a continuation of SNPs without the change at consensus sequence level. Furthermore, the analyses of Sanger sequence data have shown a number of independent SNPs identified at this same genomic position but in different egg passages of A, A1, C and D isolate.

Table 3.16. The common SNPs identified in the nucleocapsid encoding region compared to ep4 reference sequence. The table presents the minor SNPs or consensus sequence SNPs in four egg isolates on specific nucleotide position. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

The nucleotide position 25924											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	A		C	C	C	C	C	C	C	C	C
A1	A					C					
C	A										C
D	A		C	C		C	C	C	C	C	C
The nucleotide position 26067											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	G										K
C	G						T				
The nucleotide position 26068											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	C			G							
C	C						T				
The nucleotide position 26101											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	G					R					
A1	G					R	A		R		
C	G						A				
D	G			R	R	R	R	R			
The nucleotide position 26772											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	C			Y		Y		Y			
A1	C	Y	Y	Y	Y	Y	Y		Y	Y	Y
C	C							A			
D	C		Y	Y			Y	Y			
The nucleotide position 26887											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	C			Y	Y	Y	Y	Y	Y	Y	
D	C			Y	Y	Y					
The nucleotide position 26948											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	C							Y	Y	Y	
C	C				Y						
D	C		Y	Y	Y	Y	Y	Y		Y	Y
The nucleotide position 27128											
Ep # / Egg isolate	**4**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	C		Y			Y	T			Y	
A1	C		Y			Y					
D	C		Y	Y	Y	Y	Y	Y	Y	Y	Y

The analysis of the nucleocapsid encoding region in four egg isolates clearly shows that the 399 bp between nucleotide positions 26729-27130 is highly polymorphic in the case of C or T residues (Table 3.17). One of the most polymorphic encoding regions in A isolate was identified at nucleotide position 26729 with a minor SNP in ep30 and ep50 and a visible SNP at consensus sequence level (T residue) identified in ep70. Further analysis of the A isolate Sanger sequenced data confirmed the presence of nine polymorphic nucleotide positions identified in more than one egg passage regarding to the C or T residues mixture. This data has shown the most variable encoding region at nucleotide position 27060 with visible minor SNPs identified within egg passages of A, A1 and C isolate. A high frequency of C or T residues was also found at nucleotide position 26887 presenting SNP appearance in the A isolate (ep30, ep50, ep70) and D isolate (ep30-50). Further investigation of the highly active nucleocapsid encoding region also detected an extra three nucleotide positions (26855, 26948 and 26954) with minor SNPs occurring in A and D isolate. The analysis of Sanger sequence data identified three nucleotide positions where SNPs were visible only once within a 399 bp highly polymorphic region with lack of any accumulation in later egg passages of A isolate. Furthermore, the minor SNPs were also identified in A1 isolate continued in a few Sanger sequenced egg passages within the 355 bp highly polymorphic region. The highest activity of C or T residue unique only for A1 isolate was identified at nucleotide positions: 26888 (ep20-ep80), 26949 (ep40-50) and 27128 (ep20, ep50). Interestingly, further analyses of the nucleocapsid encoding region detected two nucleotide positions 26856 and 26937 with appearance of a C or T residue mixture only in ep20 without continuation of SNP accumulation in later egg passages. The phenomenon of a SNP appearance relating to an A or C residue mixture at nucleotide position 26919 in the A1 isolate (ep60-70) was only detected once at nucleotide position 26981 of C isolate.

Table 3.17. The highly active encoding region of nucleocapsid with majority of C/T residue mixture detected in egg isolates. The SNPs found in more than one isolate were highlighted as pink. White = not sequenced, yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

The nucleocapsid encoding region with majority of C or T residue mixture											
Nucleotide	**Isolate**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
26729	A			Y		Y		T			
26730	A							T	Y		
26772	A			Y		Y		Y			
26772	D		Y	Y			Y	Y			
26773	A1		Y		Y	Y	Y		Y		
26797	D			Y							
26855	A			Y							
26855	D						Y				
26856	A1		Y								
26887	A			Y		Y		Y			
26887	D			Y	Y	Y					
26888	A1		Y	Y	Y	Y	Y	Y	Y		
26891	A	Y								Y	
26919	A1						M	M			
26936	D									Y	Y
26937	A1		Y								
26948	A							Y			
26948	D		Y	Y	Y	Y	Y	Y		Y	Y
26949	A1				Y	Y					
26950	C	Y	Y	Y							
26954	A							Y		Y	
26954	D									Y	
26955	C	Y	Y	Y						Y	
26955	D					Y		Y			
26959	A							Y		Y	
26978	D							Y		Y	
26979	D							Y			
26980	C										T
26981	A					Y					
26981	C				M						
26981	D									Y	
26983	C		Y								
26987	A			Y						Y	
26994	C									Y	
27003	C								T	Y	
27050	A			Y							
27055	C									W	
27055	D				Y						
27059	D		Y	Y	Y	Y	Y	Y	Y	Y	Y
27060	A	Y		Y		Y		Y		Y	
27060	A1		Y		Y		Y		Y		
27060	C	Y	Y					Y	Y	Y	
27128	A		Y			Y	T			Y	
27128	A1		Y			Y					
27128	D		Y	Y	Y	Y	Y	Y	Y	Y	Y
27130	C		Y	Y				Y	Y	Y	

The analyses of Sanger sequence data identified seven unique C or T residue nucleotide positions in the C isolate across a 257 bp polymorphic region, with a variable frequency of appearance in different egg passages. High polymorphism was observed at nucleotide position 27130 with single accumulation of minor SNP identified in ep20-30 and ep70-90. In addition, continuation of SNPs was also investigated at genome positions 26950 (ep10-30) of the C isolate and 26955 (ep10-30, ep90) in the C isolate and D isolate (ep50, ep70). The further analyses of the C isolate genome confirmed the presence of four independent minor SNPs identified between nucleotide positions 26889-26994 without further SNP accumulation in later egg passages. Interestingly, the analyses of sequence data also identified nucleotide positions 26980 (ep100) and 27003 (ep80) with a consensus sequence SNP (T residue). In addition, the SNP cluster (C or T residue mixture) of the C isolate was also identified with a number of different SNPs in ep90 such as the A or T residues mixture such as a nucleotide position 27055.

The further investigation also confirmed a polymorphism (C or T residue mixture) unique for D isolate with various SNP accumulations across analysed egg passages. A clear continuation of SNP was visible at nucleotide position 27059 and 27128 in ep20-100 of the D isolate. In addition, further investigation of Sanger sequence data also confirmed two nucleotide positions where minor SNP (C or T residue mixture) was detected in two egg passages at nucleotide position 26936 (ep90-100) and 26978 (ep70 and ep90). The highly polymorphic encoding region was also screened in light of the independent SNPs with appearance in one egg passage such as nucleotide positions 26797 (ep30) and 26979 (ep70). Interestingly, the highly polymorphic region (C or T residue mixture) includes a few nucleotide positions previously investigated within A, A1, C and D isolate. In addition, the major part of the highly polymorphic region is built up with mutations occurring as the independent SNPs across a 399 bp encoding region with further continuation as minor SNPs or SNPs at consensus sequence level in two or more egg passages.

The investigation of the A isolate has shown dynamic SNPs in the viral genome where the C or T residue mixture accounted for more than half of total additional SNPs investigated in Sanger sequenced data. However, very often egg passages were identified with various forms of SNPs including independent SNPs and/or SNP clusters such as ep10 (table 3.18).

Table 3.18. The table illustrates the additional polymorphism in a form of independent SNPs and SNP clusters in the nucleocapsid encoding region of A isolate indicated to the sequence of ep4. Yellow = failure of Sanger sequencing reads, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

A isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
25926	G							A			
25983	A									M	
25984	G									S	
26012	C					Y	Y	Y			
26068	C			G							
26086	G			R							
26095	G									R	
26101	G					R					
26111	A			C							
26113	C									T	
26120	A				T						
26126	A			T							
26138	G			A							
26140	C							Y			
26143	NIL			T							
26150	A			T							
26187	A										T
26194	T										C
26123	G										T
26128	G										T
26235	G	R				R					
26269	G							A			
26276	G	T									
26277	T	A									
26291	C								T		
26298	T							A		Y	
26334	T							A			
26341	T							T			
26364	A										
26403	G	R									
26411	G						A				
26450	C							T			
26512	NIL								T		
26527	A					G					
26528	NIL							A			
26551	NIL							A			
26565	T	G									
26566	A									G	
26724	C									S	

In addition, different egg passages were detected with various levels of polymorphism excluding a C or T residue polymorphic region. Interestingly, ep30 was identified as an egg passage without independent SNPs, presenting SNP clusters in a form of seven consensus sequence SNPs between nucleotide positions 26068-26150 (82 bp) with no further continuation in later egg passages. Moreover, a reduced polymorphism was analysed in ep40-60 identified with few independent SNPs at consensus sequence level comparing to ep70 found with formation of a SNP cluster between nucleotide positions 26269-26341 (72 bp) and few additional SNPs. Interestingly, ep90 was identified with seven SNPs where 85 % of these genomic changes were not accumulated in further egg passages. The polymorphism was substantially reduced in ep100 showing no SNPs appearing in the C or T residue abundant encoding region or any independent SNPs unique for this viral population. However, further investigation showed a SNP cluster between nucleotide position 26187-26218 presenting as a formation of four consensus sequence SNPs within a 31 bp. The A isolate was identified as having a substantial number of independent SNPs without accumulation in later egg passages. In addition, SNP formation was shown only twice, presenting the A isolate with reduced accumulation of SNPs within a particular egg isolate.

The identified C or T residue-polymorphic region of the A1 isolate was established to be identical as in the A isolate (54 % of all additional SNPs). Interestingly, the SNP cluster was identified at consensus sequence level in ep40 between genomic positions 26168-26198 without any additional independent SNPs in the nucleocapsid encoding region (table 3.19). The analysed Sanger sequenced egg passages showed only SNPs at nucleotide positions 26223 and 26254 in ep50 without formation of a SNP clusters across the viral population and lack of any polymorphism in ep60. In addition, a very low polymorphism was also confirmed in ep70 presenting a small SNP cluster (three consensus sequence SNPs) between nucleotide positions 26674-26684 (10 bp encoding region) separated from the nearest minor SNP by 235 bp gap. Further SNP investigation in ep90 could not be continued due to the failure of the regular Sanger sequencing reads. However, little polymorphism in previous egg passages showed that A1 isolate presented a reduced number of SNPs compared to the rest of the egg isolates.

Table 3.19. The table of polymorphism identified as additional SNPs in the nucleocapsid encoding region of A1 and D isolate. White = not sequenced, orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

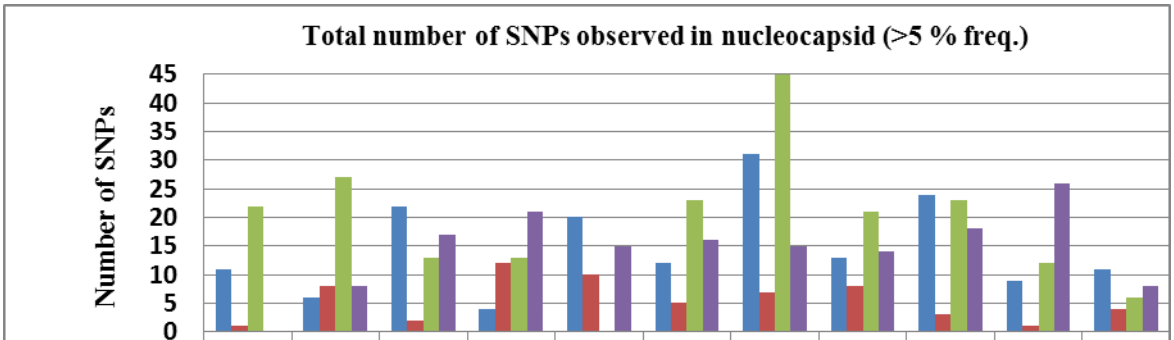
A1 isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
26168	G				C						
26174	G				A						
26184	C				T						
26186	C				T						
26190	T				A						
26196	G				C						
26198	A				G						
26233	C					T					
26254	G					C					
26398	C							Y			
26547	G								A		
26674	G							A			
26680	T							A			
26684	G							T			
D isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
26121	G										R
26123	A										W
26174	G								T	T	
26175	A								T		
26179	C				A						
26180	C								G		
26181	A								T		
26184	C										A
26197	T										A
26210	C										A
26214	A										T
26217	A										T
26218	G										T
26224	G										T
26225	C									T	T
26226	C										T
26228	G				R						
26233	C						Y				T
26235	G					R					
26241	G										T
26249	G										T
26292	C				T						
26342	A			T							
26347	C			Y							
26369	C			T							
26373	NIL					A					
26380	G				A						
26406	NIL				A						
26413	G				A						
26417	G				T						
26577	NIL										
26573	NIL						G				
26580	C					T					
26589	NIL						A				
26603	NIL				T						

The SNPs previously not detected in ep110 using 454 pyrosequencing are a substantial part of the D isolate showed as a SNP cluster between nucleotide positions 26342-26369 in ep30. In addition, SNP accumulation was also confirmed in ep40 (SNP cluster) identified at consensus sequence level between nucleotide positions 26292-26417 (125 bp). Further analyses of Sanger sequence data confirmed reduced numbers of SNPs in ep50-60, presented by consensus sequence and minor SNPs. The ep80 was identified with a SNP cluster of four mutations at consensus sequence level between nucleotide positions 26174-26181. Interestingly, ep80 showed the first identified genomic change at nucleotide position 26174 with further continuation as consensus sequence SNP (T residue) in ep90. The most interesting fact is the large number of SNPs presented in ep100 (16 SNPs) including SNP cluster (12 SNPs) between nucleotide positions 26184-26249 (65 bp) classified as unique for the D isolate, never been seen before in investigated egg isolates. Interestingly, all Sanger sequenced SNP clusters were visible only once within its egg passage without further continuation. Furthermore, the SNP cluster built up at consensus sequence level in ep100 was not confirmed in attenuated ep110 but could have an influence on its SNPs accumulation and frequency rate.

The C isolate clearly showed the unique polymorphism in ep10 identified a SNP cluster formed by minor SNPs and SNPs at consensus sequence level between nucleotide positions 26048-26184 (136 bp) with the distance of 553 bp to the nearest independent SNP (appendix 13). Furthermore, this same observation was confirmed in ep20 regarding to the appearance of two SNP clusters within nucleotide position 26177-26239 (six consensus sequence SNPs and one minor SNP) separated by a 278 bp gap from the second SNP cluster presenting 10 SNPs between nucleotide positions 26517-26610. Furthermore, ep10 and ep20 were analysed as some of the most polymorphic egg passages shown by unique and independent SNPs (A or T residue mixture) occurring as the most common mutations. Interestingly, a common characteristic relating to the appearance of close mutations were also shared between ep30 and ep40 showing a SNP cluster (six residues) found between nucleotide position 26179-26198 (ep30) and nucleotide position 26052-26108 (ep40). The accumulation of SNPs was also confirmed in ep60 identified with a highly polymorphic encoding region formed by 21 SNPs between nucleotide positions 26039-26104 (56 bp) with no additional independent SNPs. The same situation was presented in ep70 with two highly polymorphic encoding regions, 22 SNPs identified between nucleotide positions 26051-26167 (116 bp) and 20 SNPs found between

nucleotide positions 26725-26810 (85 bp) across the investigated viral population. Further analyses showed the distance between these two clusters as 558 bp proving no correlation between two highly polymorphic regions of ep70. The further investigation of ep80 has shown common characteristic to have a mixture of SNPs at consensus sequence level within a short bp length presented by two SNP clusters between nucleotide positions 26170- 26198 (28 bp) and 26683-26720 (37 bp). Further analyses also confirmed consensus sequence SNPs in the SNP cluster of ep90 between nucleotide positions 26083-26195 which in fact fully overlaps with the first SNP cluster of ep80 in the C isolate. Interestingly, ep100 could not be confirmed with a presence of SNP clusters but was identified with a number of unique SNPs across the viral population.

3.5.10 The summary of polymorphism in nucleocapsid encoding region

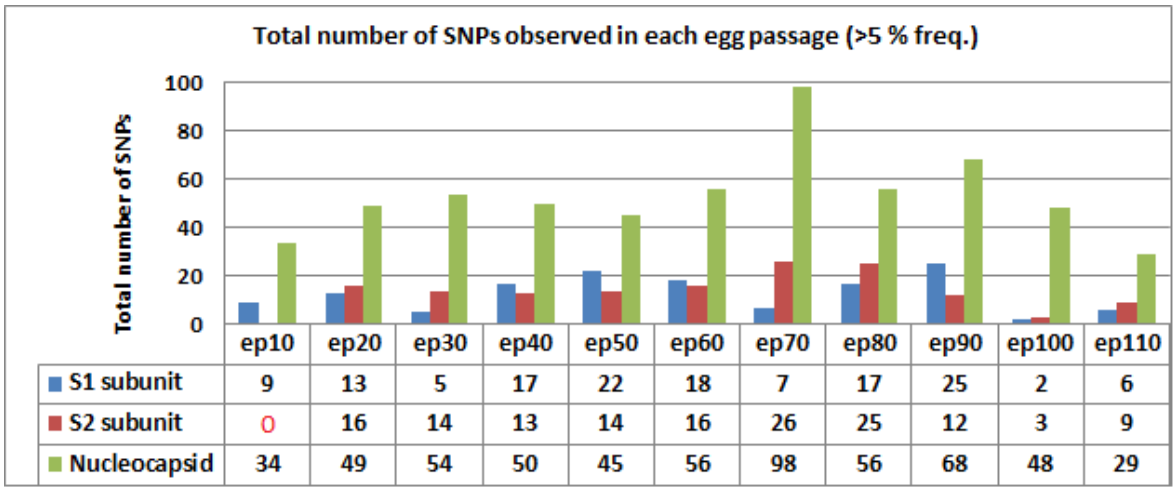


	ep10	ep20	ep30	ep40	ep50	ep60	ep70	ep80	ep90	ep100	ep110
A isolate	11	6	22	4	20	12	31	13	24	9	11
A1 isolate	1	8	2	12	10	5	7	8	3	1	4
C isolate	22	27	13	13	0	23	45	21	23	12	6
D isolate	0	8	17	21	15	16	15	14	18	26	8

Figure 3.7. The total number of SNPs across all investigated egg isolates in nucleocapsid encoding region with the final reads of ep110 (>5 % freq.). The diagram presents number of genomic mutations in each of analysed egg passage (ep10-110). The bar chart illustrates polymorphism across four egg isolates (A, A1, C and D isolate) presented by different colours. The table below bar chart presents exact number of SNPs identified in egg passage and egg isolate of interest.

The nucleocapsid encoding region was identified with the highest number of SNPs in all investigated egg isolates. Interestingly the pattern of SNP accumulation was found similar to the S2 subunit (part 1) where SNPs were visible in ep10-50 without rapid increasing of genomic mutations. However, C isolate presents high polymorphism in ep10-20 with maximum of 27 SNPs identified using Sanger sequencing. Interestingly, the number of SNPs rapidly decrease in ep30-ep50 to reach the highest number of identified SNPs in ep70 (45 SNPs) and reduced again with final result of 12 SNPs in ep100 (figure 3.7). The similar observation was made in Sanger sequenced data of A isolate with the highest number of SNPs (31) in ep70. However, A isolate has shown visible accumulation but also reduction of SNPs every 20 egg passages. The further analyses of Sanger sequenced data highlighted phenomenon similar to observations made within S2 subunit of A and D isolate. The SNP accumulation in nucleocapsid encoding region of A and C isolates increase or decrease in opposite manners between ep20-50 with two increased accumulation of SNPs in ep70 and ep90. Interestingly, SNP accumulation in A1 and D isolate was observed with a pattern with minimal polymorphism in ep10 and gradual reduction of genomic mutations from ep40 to ep80. Interestingly ep80 is a crucial point where SNP accumulation increases in D isolate from 14 SNPs (ep80) to 26 SNPs (ep100) but decreases in A1 isolate with final result of one SNP in ep100. Interestingly, reduced polymorphism in ep90-100 of A, A1 and C isolates did not have an influence on SNPs measurement in ep110 comparing to increasing polymorphism in ep90-100 of D isolate. Furthermore, analyses of pyrosequencing data (ep110) has shown increased number of SNPs in A and A1 isolates but decreasing tendency in C and D isolates comparing to polymorphism found in ep100.

3.5.11 The analyses of polymorphism observed across all genes.

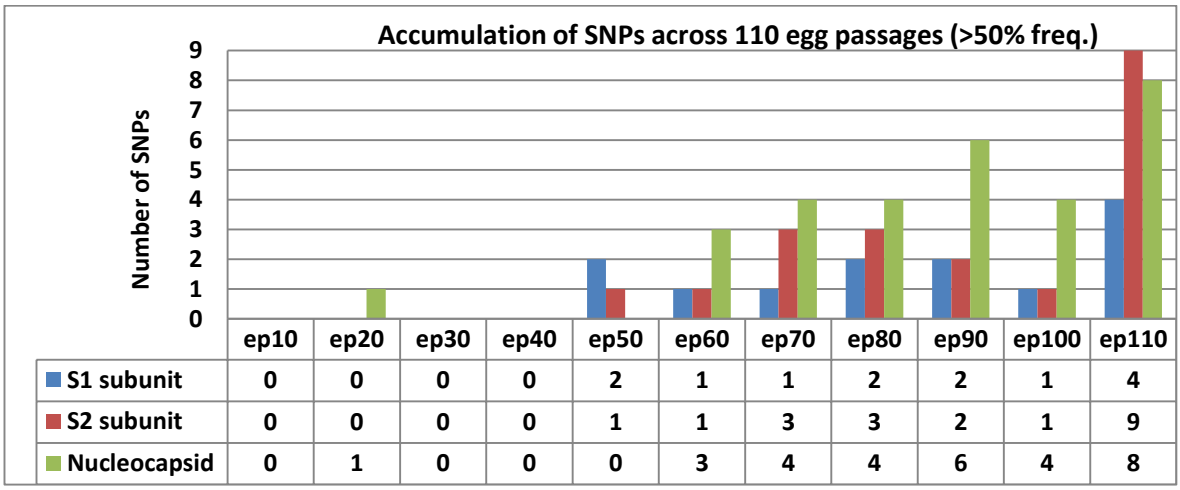


	ep10	ep20	ep30	ep40	ep50	ep60	ep70	ep80	ep90	ep100	ep110
S1 subunit	9	13	5	17	22	18	7	17	25	2	6
S2 subunit	0	16	14	13	14	16	26	25	12	3	9
Nucleocapsid	34	49	54	50	45	56	98	56	68	48	29

Figure 3.8. The bar chart presents the total sum of SNPs identified within four egg isolates in Sanger sequenced encoding regions including >5 % frequency of pyrosequenced datasets. The bars were presented as follow: blue = S1 subunit, red = S2 subunit and green = nucleocapsid. The table below bar chart includes the exact number of SNPs identified in each egg passage of analysed encoding region. The zero numbers marked as red colour relates to failure of Sanger sequencing reads.

The Sanger sequenced data showed the scale of total SNP accumulation in previous egg passages within three encoding regions at >5 % frequency. The figure 3.8 presents visible pattern of polymorphism between S1 subunit and nucleocapsid encoding region. Both genes were found with number of investigated SNPs in ep10-20 showing increasing SNP accumulation in nucleocapsid (49 SNPs) and S1 subunit (13 SNPs) at >5 % frequency. However, ep30 was found as the crucial point showing decreasing polymorphism in S1 subunit (five SNPs) and increasing number of SNPs in nucleocapsid (54 SNPs). Further analyses of previous egg passages also identified ep50 and ep70 as the breaking points presenting nucleocapsid encoding region and S1 subunit with opposite tendency of SNP accumulation in viral genome. According to the results included in bar chart above, ep50 in nucleocapsid was classified with a relatively high number of mutations (45 SNPs) but with increasing SNP tendency in further egg passages and maximum of 98 SNPs identified in ep70. Interestingly, ep70 in S1 subunit was identified as one of the lowest points in SNPs accumulation presenting only seven mutations. However, the extreme difference in SNP accumulation visible in ep70 between S1 subunit and nucleocapsid led to another interesting situation. The analysed Sanger sequenced data has shown increasing number of

SNPs in S1 subunit (25 SNPs) and nucleocapsid (68 SNPs) in ep90 and decreased SNP accumulation in ep100 classified as six SNPs in S1 subunit and 29 SNPs in nucleocapsid encoding region. Interestingly, S2 subunit was found with visible SNP accumulation in early egg passages presenting a maximum of 16 SNPs in ep20. Further investigation of Sanger sequenced data identified increased polymorphism in ep70 presented by 26 mutations with decreasing SNP accumulation in ep80-100. Interestingly, deep sequenced data presented increased number of SNPs in ep110 at >5 % frequency presented final value of six SNPs (S1subunit) and nine SNPs (S2 subunit). However, significant accumulation of mutations in nucleocapsid encoding region identified in Sanger sequenced egg passages was reduced to 29SNPs found in pyrosequencing data of ep110 at >5 % frequency. The reduction of polymorphism in ep110 can be possibly explained by selection of SNPs which play a key role in process of attenuation under lack of immune pressure. This observation was present in all encoding regions of interest showing that reduction of viral pathogenicity could possibly depends on few genomic changes across viral quasispecies.



	ep10	ep20	ep30	ep40	ep50	ep60	ep70	ep80	ep90	ep100	ep110
S1 subunit	0	0	0	0	2	1	1	2	2	1	4
S2 subunit	0	0	0	0	1	1	3	3	2	1	9
Nucleocapsid	0	1	0	0	0	3	4	4	6	4	8

Figure 3.9. The bar chart illustrates the accumulation of SNPs found at >50 % frequency in ep110 also visible across Sanger sequenced egg passages. The bars were presented as follow: blue = S1 subunit, red = S2 subunit and green = nucleocapsid. The table below bar chart includes the exact number of SNPs identified in each egg passage of analysed encoding region.

Figure 3.9 shows the significant level of >50 % frequency SNPs accumulation in later egg passages of S1 subunit, S2 subunit and nucleocapsid encoding region. The Sanger sequence datasets presented very low polymorphism in early egg passages with maximum

of one SNP as seen in ep10-40. The number of mutations at >50 % frequency were observed with gradual accumulation in ep80 within three encoding regions with the highest activity observed in nucleocapsid (four SNPs). In addition, the polymorphism in nucleocapsid encoding region seems to be continued with the maximum measurement of seven SNPs in ep90. Figure 3.9 also presents a very similar pattern of SNP accumulation and reduction in S1 and S2 subunits which starts in ep50 (one SNP) followed by increasing number of mutation at maximum point of three SNPs in S2 subunit (ep80) and progressive reduction visible in ep100 (one SNP) of both spike subunits. In addition, S1 and S2 subunits were identified with SNPs accumulation at this same level in ep60 (one SNP) and ep90 (two SNPs).Interestingly, all three encoding regions were also found with decreasing number of SNPs in ep100 before the final pyrosequencing reads in ep110 at >50 % frequency. Interestingly, the lower number of polymorphisms in ep100 did not mirror the situation of SNP accumulation in ep110 presented with nine SNPs observed in S2 subunit, eight SNPs in nucleocapsid and four SNPs found in S1 subunit.

Chapter 4: Discussion

4.1 Introduction

The final confirmation of pathogenicity loss in egg isolates was confirmed after infecting Rhode Island Red chickens, identified with lack of any post-infection clinical signs. The analyses of sequenced data from 110 egg passages showed build-up of SNPs in the nucleocapsid and spike encoding regions which might be involved in the attenuation of the four egg isolates. Further investigation showed a number of SNPs commonly visible within more than one egg isolate. In addition, appearance of these SNPs was identified at different stages of egg passage depending on the analysed isolate and SNP frequency in ep110. Furthermore, the appearance of additional SNPs previously not detected in ep110 was also confirmed as a substantial part of the investigation. These findings have clearly shown the viral genome as a dynamic environment which, in the absence of immune pressure presents a mixed formation of SNP clusters and independent SNPs within particular egg passages. In addition, SNP accumulation within egg passages is not a phenomenon visible in all investigated egg isolates suggesting the influence of numerous factors affecting the polymorphism.

4.2 Verification of vaccine efficiency based on animal testing

One of the most effective ways to confirm the loss of viral pathogenicity is infecting birds with multiple egg passaged viral agent. In this case, the attenuation was confirmed after infecting four birds only (Rhode Island Red chickens) for each isolate. This methodology is commonly used for production of many poultry vaccines based on live-attenuated vaccine strains. According to results of viral attenuation verification, 110 passages were required to achieve a virus strain to become attenuated sufficiently to comply with safety requirements for live IB vaccines. So far, the types of mechanism causing a virus to attenuate are not fully recognised. However, the investigation of SNPs accumulated during multiple egg passages identified potential SNPs which are probably involved in loss of viral pathogenicity. Hung and Wang, 2007 has confirmed presence of genomic mutations (SNPs) in QX-like infectious bronchitis virus strain after only 80 egg passages in the spike 1 and 2 subunits after sequencing the 3’ 7.3 kb of the genome of an attenuated and parent virus strain from which an attenuated strain was derived. However, there was lack of any SNPs identified in the nucleocapsid but the envelope protein and the membrane protein.

This information confirms that the attenuation process and the encoding region involved strongly depend on IBV strain and the number of egg passages required for loss of pathogenicity.

Geerligs *et al.,* 2011 confirmed a clear effect of increased egg passage level on the results of ciliostasis scores which decreased after 25 passages. Interestingly, the results of ciliostasis increased again until ep 50 with constant reduction straight after. These results clearly demonstrate that the IBV isolate was attenuated during passaging and the data for passage level 80 can be considered with respect to safety for the respiratory tract and kidneys of tested birds. Interestingly, investigation of viral attenuation (Crinion & Hofstad, 1971) also confirmed 12.5% female chickens with lesions in the oviducts on the day of hatching. IBV caused lesions only in chickens vaccinated at 1 day of age, but no lesions were found in chickens vaccinated at 8 days of age. These data confirm finding in the egg passages of M41 strain investigated in the Pirbright Institute. The evidence of attenuation after infecting Rhode Island Red chickens clearly shows the presence of virus in trachea followed by lack of post-infection clinical signs on day 3-7.

The universal live vaccines against constantly emerging IBV strains do not yet exist. In order to find ways to protect poultry against IBV strains, combinations of existing vaccine strains have been tested using birds. The results of that practice showed that combined vaccines could possibly confer protection against newly IBV strains challenge. The disadvantage of this practice is the case of accidental production of more pathogenic strains mutated from existing IBV agents (Hopkins & Yoder, 1984). Luckily, it is not unusual for chickens to be vaccinated with different combinations of vaccines to generate broad protection. Other aspects, such as degree of attenuation and interactions and compatibilities with other vaccines, also should be taken into account (Terregino *et al.*, 2008).

4.3 The failure of Sanger sequencing assay in previous egg passages

The failure of Sanger sequencing reads was a significant issue in investigation of polymorphism in a number of previous egg passages including S1 subunit, S2 subunit (part1), and S2 subunit (part2) encoding regions regarding to four egg isolates. The cDNA synthesis was a problem in successfully obtaining complete Sanger sequencing data due to inappropriate heating of the sample to 65 ^{o}C for five minutes and incubating them on ice for at least one minute. However, this issue was solved successfully with visible bands on agarose gel after the amplification process of cDNA samples, synthesized once again. Furthermore, purification of samples was done by using GFX PCR, DNA and Gel Band Purification Kit or PCR product purification (Qiagen QIAquick Spin Handbook) following the instructions from the manual. Interestingly, the purification process was only partially successful showing ethanol contamination of S1 subunit samples based on 260/230 ratio identified as the main issue in Sanger sequencing assay presenting low bars throughput of A isolate (ep40, ep60-100), A1 isolate (ep10-40), C isolate (ep10, ep30, ep50, ep70, ep100) and D isolate (ep10-30, ep80). Sometimes the Sanger sequencing reads (S2 subunit–part1) were visible but the signal stopped abruptly showing a secondary structure (GC and AT rich templates caused the DNA to loop and form hairpins) causing failure of Sanger sequence reads in A isolate (ep10), A1 isolate (ep10, ep80 and ep100), C isolate (ep10, ep30-100) and D isolate (ep10 and ep30). The designed primers were only a problem in S2 subunit (part2) encoding region due to no priming site present in the viral genome but also primer dimer formation. Moreover, failure of S2 subunit (part 2) Sanger sequencing was identified as a main issue in completion of polymorphism in previous egg passages followed by satisfactory amplification of S1 subunit, S2 subunit (part1). Interestingly, nucleocapsid encoding region was successfully Sanger sequenced in all egg passages of four egg isolates comparing to S1 and S2 subunit (part1).

4.4 The cases of overlapping SNPs in investigated IBV encoding regions

SNPs identified at the same nucleotide position in all investigated egg isolates might indicate a substantial influence on viral attenuation. Interestingly, the spike encoding region presented SNPs at common nucleotide positions in a maximum of three egg isolates identified as non-synonymous (S1subunit) and synonymous (S2 subunit) mutations. The

S1 subunit is responsible for viral neutralisation, attachment to the host cell but also viral pathogenicity. Therefore, substantial SNPs in the spike encoding region may be correlated with the viral attenuation. In addition, the identified SNPs in both subunits were found with a high frequency (>50 %) in ep110 suggesting a strong influence of polymorphism within the viral population. Furthermore, the high SNP frequency (80 %) in ep4 (unknown level of pathogenicity) showed no correlation with increasing SNP frequency in ep110 but had an impact on SNP accumulation in Sanger sequenced egg passages. In addition, S1 subunit was also found with only one synonymous SNP (1.1-1.7 % freq.) within three egg isolates compared to S2 subunit identified with three synonymous SNPs (52-98 % freq.) suggesting visible differences in genome changes within two subunits of spike encoding region. However, it is reasonable to conclude that only a small fraction of SNPs are crucial for attenuation of pathogenicity. Therefore, the differences in SNP accumulation within the spike encoding region of the four egg isolates suggested that the attenuation process is driven by many forces and might be possible due to SNP accumulation in other encoding regions.

A high number of overlapping SNPs was also identified within the nucleocapsid encoding region with two SNPs commonly visible within three and four egg isolates suggesting a substantial influence of nucleocapsid in the process of building up the attenuated egg isolates. Interestingly, shared SNPs were identified in all cases with a relatively low SNP frequency (<50 %) in ep110 suggesting involvement of the nucleocapsid in early stages of viral attenuation, developing in a small fraction of the viral population and having an impact on the further appearance of SNPs in the spike encoding region. In addition, these sequence datasets have shown a correlation between SNPs in previous egg passages and SNP frequency in ep110. Furthermore, sometimes a lower frequency (>17 %) was identified as having a substantial SNP accumulation in previous egg passages compared to SNPs at >50 % frequency in ep110 without further signs of SNPs activity in Sanger sequenced egg passages. The analyses of deep sequence data confirmed that SNP appearance in single genome positions could be a random phenomenon clearly visible with a low frequency in ep110. Therefore, SNPs should be carefully identified as an early sign of the attenuation process leading to a substantial impact on viral pathogenicity. Similar to the spike, the nucleocapsid encoding region was identified as having only two common synonymous SNPs with a frequency of 31-33 % (ep4) and 20-43 % (ep110) within the investigated egg isolates. This phenomenon might be a sign presenting influence of

synonymous SNPs in the attenuation process occurring in four egg isolates after multiple egg passages. In addition, research studies suggested that synonymous SNPs exempt from functional constraints at the protein level, and its rate reflects the mutation rate to a great extent especially in the light of non-synonymous changes across viral genome. The variation of SNPs identified in four egg isolates could be caused by different replication frequencies, which are affected strongly by the infection and transmission mode in embryonated chicken eggs (Hanada *et al.,* 2004). Interestingly, common SNPs identified in the nucleocapsid encoding region were less visible in previous egg passages compared to SNPs detected in the spike encoding region potentially due to the low SNP frequency in ep110 (nucleotide position 26791) showing reduced polymorphism. Furthermore, the synonymous SNP analysed at nucleotide position 26793 in all four egg isolates presented an activity which might have an impact on SNP appearance in a neighbouring nucleotide position 26792 in the viral populations of A, A1, C and D isolate. However, the situation of genomic changes identified in extreme close locations was found only once in the nucleocapsid without further explanation or any confirmation relating to the spike encoding region.

Further investigation also classified a number of overlapping SNPs previously not detected by the 454 pyrosequencing reads in ep110. Interestingly, there is a lack of overlapping SNPs in the case of the S1 subunit showing this highly active encoding region without common SNPs in four egg isolates. However, the S2 subunit was defined as a more conserved part of the spike encoding region responsible for virus to cell and cell to cell fusion. Interestingly, analysed Sanger sequence data showed a highly polymorphic encoding region within the first 892 bp of S2 subunit with substantial SNP correlation in the A, C and D isolate at eight nucleotide positions identified as independent SNPs. However, the analyses also classified a number of SNPs such as nucleotide position 22224 visible as polymorphic across investigated egg passages of C and D isolate. Furthermore, the majority of SNPs might indicate that the attenuation process could be dependent on a selection of minor SNPs within the viral population resulting in a reduction of pathogenicity in quasispecies. Similar to the identified number of SNPs in ep110, the highest numbers of additional SNPs shown by Sanger sequencing were detected in the A isolate. These findings showed the influence of additional SNPs which might be crucial in building up the attenuation process observed in four egg isolates.

4.5 The influence of differences between SNP frequency in ep4 and ep110

The low SNP frequency of ep4 seems to have an influence on SNPs and their frequency at the consensus sequence level in ep110 but also the SNPs appearance in the Sanger sequenced egg passages. The confirmation of above statement was identified in nucleocapsid encoding region where SNP accumulation was strongly dependent on the changes in frequency between ep4 and ep110. In addition, in most of the cases the relatively reduced (<30 %) SNP frequency in ep4 was continued in all egg isolates resulting in a lower (<29 %) frequency of ep110 having also an influence on SNP accumulation in Sanger sequenced egg passages. This phenomenon was observed at nucleotide positions 26792 and 26793 where the dynamics of SNP accumulation were not predictable showing relatively low SNP frequency in ep4 followed by SNP accumulation in all egg passages which could be a sign of the attenuation process. However, the low frequency (>43 %) identified in ep110 after the appearance of SNPs at consensus sequence level in previous egg passages (nucleotide position 26792-26793) might be due to potential error in 454 pyrosequenced data caused by poly A tract present in viral genome. This issue could be solved by Sanger sequencing of nucleocapsid encoding region to analyse nucleotide position 26792-26793 in order to confirm polymorphism in viral genome. Interestingly, the A1 isolate presents polymorphism only in ep70-90 (nucleotide position 26792) which could be triggered by other factors (undefined yet) influencing the SNP appearance in later egg passages. The lower frequency in ep4 followed by SNP accumulation in all Sanger sequenced egg passages could be a sign of selection the random SNPs across the viral population. In addition, the further observation confirmed the link between the SNP frequency in ep4 and its substantial impact on SNP appearance in early egg passages and the final measurement of SNP frequency in ep110. The reason might be the low SNP accumulation in the viral population which could not be continued further as a selection of minor SNPs but also the fact of less sensitive detection of mutation by Sanger sequencing comparing to deep sequencing method. This observation was confirmed with very low but increasing frequency between ep4 (2.6 %) and ep110 (8.6 %) identified at nucleotide position 26791 in D isolate with lack of any polymorphism in Sanger sequenced egg passages. Interestingly, further investigation also identified the C and D isolate (nucleotide position 26795) with a low (15 %) frequency in ep4 leading to a higher (23-36 %) frequency in ep110 without sign of SNPs activity in previous egg passages. The observation of constantly changing frequency between ep4 and ep110 has shown no

correlation in appearance of SNPs in viral genome of Sanger sequenced egg passages. Further investigation, also confirmed no influence of polymorphism in previous egg passages with increasing SNP frequency between ep4 and ep110. The results of investigation have shown the dynamics of viral genome after numerous egg passages which might have a place due to lack of immune pressure in embryonated chicken eggs.

4.6 The SNPs potentially involved in process of attenuation

Based on the information analysed from deep sequence data, the process of attenuation seems to have different mechanisms of formation in egg isolates. The SNPs found at consensus sequence level might be involved in loss of virulence due to appearance in more than 50 % of viral population. Interestingly, none of investigated encoding regions have showed genomic changes shared within four egg isolates after multiple egg passages at >50 % frequency in ep110. Interestingly, S1 subunit as the highly polymorphic part of spike encoding region was found with non-synonymous SNP shared within three egg isolates. In addition, S2 subunit identified as highly mutative under lack of immune pressure presented three synonymous genomic changes visible in two egg isolates at consensus sequence level. This observation might indicate the significant influence of spike as encoding region implying in process of attenuation. It is understandable that non-synonymous genomic changes will have a greater impact on viral virulence due to possible changes in amino acid sequence. However, it is believed that synonymous SNPs might have an influence on further genomic changes leading to loss of virulence in egg isolates (Studer *et al.,* 2013). Furthermore, the SNP rate at nonsynonymous sites is roughly 100-fold less than that at synonymous sites in RNA viruses although the greater influence of natural selection at synonymous sites is likely to increase the variance in viral genome by appearance of non-synonymous SNPs (Holmes, 2003). Therefore, all SNPs identified at >50 % frequency in ep110 must be consider as potential key mutations responsible for driving the process of attenuation under lack of immune pressure. Further analyses of 454 pyrosequencing datasets clearly shows fundamental meaning of unique SNPs at consensus sequence level classified in each egg isolate as mutations potentially involved in key genomic changes after multiple egg passages. The spike and nucleocapsid encoding regions were also identified within four encoding regions with the highest percentage of unique SNPs classified at >50 % frequency. However, there is no evidence as to how many

SNPs (non-synonymous and/or synonymous) are required for successful attenuation and loss of virulence in investigated egg isolates. In addition, the nucleocapsid has shown twice the number of >50 % frequency SNPs comparing to spike encoding region which might play a crucial role in loss of virulence within four egg isolates. Interestingly, none of these SNPs were shared in more than one isolate which partially explains the complicated mechanism responsible of attenuation process. In addition, spike encoding region also has shown few SNPs identified as unique for particular egg isolate suggesting induction of severe changes in analysed viral population structure. Therefore, constantly changing situation in viral genome after multiple egg passages can go in many different directions leading to this same conclusion and final attenuation in major part of viral population.

Interestingly, loss of virulence can be trigger by appearance of SNPs with >5 % frequency which by "domino effect" might cause another genomic changes in viral population. The above hypothesis is supported by the results of deep sequence data analysed in nucleocapsid encoding region found with two SNPs shared within four egg isolates at >5 % frequency. In addition, significantly lower frequency within viral population has shown SNPs which are also shared within three egg isolates in nucleocapsid encoding region. It is believed that appearance of mutations in a small fraction of larger viral quasispecies might have an influence on development of genomic changes which at some point will be classified at consensus sequence level. This situation had a place in spike encoding region where shared >5% frequency SNPs were present in deep sequenced analyses of >50 % frequency SNPs in ep110. Interestingly, further investigation has shown that above situation was confirmed only in a case of spike encoding region, whereas other encoding regions were classified with decreasing number of SNPs regarding to gradual increasing in frequency. The attenuation process followed by loss of virulence might begin in shared polymorphic genome positions at extremely low SNP frequency (<5 %) within three egg isolates as showed in S1 subunit. However, there is no evidence or confirmed link between attenuation process and genomic changes investigated in 1-5 % of viral quasispecies. Therefore, SNPs at <5 % frequency identified using 454 pyrosequencing data were included in deep sequenced datasets but not considered as crucial in process of attenuation under lack of immune pressure.

4.7 The different stages of SNP accumulation in investigated egg isolates

The four egg isolates were identified with a substantial SNPs accumulation in early and/or later egg passages using Sanger sequencing. Further analyses had indicated that SNP appearance in egg passages strongly depended on the egg isolate but also the encoding region of interest. The identification of SNPs in the S1 subunit clearly showed the SNP accumulation at consensus sequence level in early egg passages of the A (data available only for ep10-50). Further observation identified SNPs which were present in early egg passages without information about further continuation (A isolate - data not available) suggesting that SNPs occur at consensus sequence level but their continuation varies in egg isolates. Interestingly, the visible SNPs within a single egg passage could be spotted as a part of SNP cluster which, in the absence of immune pressure, were not continued as independent mutations in later egg passages. However, Sanger sequenced data has shown A1 (no data for ep20-40) and D isolate with very low SNP activity in early egg passages but the presence of minor SNPs and/or SNPs at consensus sequence level in later egg passages. Interestingly, C isolate has shown a large number of minor SNPs and SNPs at consensus sequence level appearing independently or in SNP clusters across all Sanger sequenced egg passages. Its seems that the increased activity of mutations occur in the viral genome across investigated egg passages make a significant impact on the attenuation process presenting C isolate as the most attenuated egg isolate. These data suggest that the SNP accumulation under lack of immune pressure have significant effect on level of attenuation and the way in which particular egg isolate lost its virulence. In addition, the high SNP frequency at genomic positions in ep110 seems to be necessary for the process of attenuation but selection of those SNPs has been shown as random or driven by other factors responsible for lack of viral pathogenicity. Furthermore, the viral populations were also identified with a visible continuation of fewer minor SNPs which graduated to SNPs at consensus sequence level suggesting that the attenuation in quasispecies might be due to progressive SNP accumulation visible in Sanger sequenced egg passages. Interestingly, this idea can be accepted by the SNP examples found in the S1 subunit of the A and A1 isolate showing the early egg passages as an accumulation of minor SNPs with a change to SNP at consensus sequence level in later egg passages and a strong SNP frequency (≥99 %) in ep110.

Interestingly, the observation of SNP appearance in the S1 subunit was also confirmed in the S2 subunit (part 1) relating to various levels of SNP accumulation as potentially key for investigating the attenuation process in egg isolates. In addition, the mechanism of slow selection of the mutations was confirmed in the A isolate as a synonymous SNPs clearly visible only in later egg passages. However, the identification of progressive SNPs could be confirmed in the S2 subunit with a high SNP frequency (>52 %) in ep110 having an influence on SNPs accumulation visible as minor SNPs in previous egg passages (ep30-70) and SNPs at consensus sequence level in ep80-90 of A isolate. Interestingly, polymorphism found in the D isolate confirmed the fact of viral genome described as highly dynamic environment with the SNP accumulation identified as minor SNP in early egg passages which turns back to the original nucleotide from ep4 in ep70-80 of the S2 subunit (part 1) with the final 56 % SNP frequency in ep110. These differences of SNPs appearance with the high frequency within two egg isolates suggest that the polymorphism in viral genome strongly depends on many other factors which in the lack of immune pressure present selection of minor SNPs. In addition, process of attenuation might be successful due to the combination of unique and commonly found SNPs in the viral genome after 110 egg passages. Interestingly, not all investigated egg isolates were identified with visible additional SNPs and their continuation in egg passages potentially having an influence on the attenuation process. However, this phenomenon was observed few times in the case of the S1 and S2 subunit (part 1) suggesting a substantial change in SNP occurrence during selection of minor SNPs in the spike encoding region. The dynamics of SNP accumulation has shown as a difference within investigated egg isolates supported by number of SNPs and the degree of continuation within the viral populations. The A and D isolate were identified with various lengths of SNP accumulation across Sanger sequenced egg passages supporting the idea of SNP selection continued in later egg passages. Further investigation identified two nucleotide positions in both egg isolates with SNPs (mixture of C or T residues and A or G residues) detected in ep30-80. The SNP accumulation was also confirmed as a substantial part of the A1 isolate with more specific selection of minor SNPs in the viral genome compared to the A and D isolate where Sanger sequence data indicated SNP accumulation in fewer egg passages but at consensus sequence level.

The highly polymorphic encoding region of the nucleocapsid was investigated in light of SNP accumulation and the differences relating to selection of SNPs in the viral population of ep110. The analyses of Sanger sequence data clearly showed that the nucleocapsid encoding region presented different types of SNP accumulation previously identified in the spike encoding region such as graduate SNP accumulation and the appearance of independent SNPs without a further continuation in later egg passages. The nucleocapsid is known as highly conserved in nature compared to the nucleocapsid of egg isolates found with the great polymorphism (Schelle *et al.*, 2005). Interestingly, the majority of SNPs were identified with further continuation due to SNP appearance at consensus sequence level in later egg passages. This phenomenon could be observed in each of the investigated egg isolates with a difference in the number of egg passages required for SNP visibility in a major part of the viral population. Further Sanger sequencing showed that the A isolate was generally identified with SNP accumulation in early egg passages and its continuation as the consensus sequence SNP in later egg passages. However, in some cases the SNPs were observed among egg passages without further identification in later egg passages before a final appearance in ep110 of the A and D isolate. However, the majority of SNPs with further accumulation have been shown at consensus sequence level with relatively high frequency in ep110 (>50 %) as seen in the nucleocapsid encoding region of the A1 isolate. The SNP dynamics in viral population presents different levels of SNP accumulation in all four egg isolates. Furthermore, a high frequency (>98 %) in ep110 seems to have an impact on SNP accumulation in Sanger sequenced egg passages with further identification as SNPs at consensus sequence level in the ep110. However, a lower frequency (<20 %) did not show a substantial impact on the appearance of consensus sequence SNPs and their continuation. In addition, the nucleocapsid encoding region was identified as a logical SNP accumulation in numerous egg passages of A1 and D isolate, also detected in A and C isolate as independent SNPs. These data suggest that a lack of immune pressure in embryonated chicken eggs is not the only one force involved in building up viral attenuation. Further investigation of the viral genome after 110 egg passages might indicate the influence of independent SNPs or SNP clusters which could cause further polymorphism in nucleocapsid encoding region.

4.8 The unique SNP cluster identified in the nucleocapsid encoding region

An interesting SNP accumulation was also detected in a part of the nucleocapsid encoding region identified as minor SNPs (C or T residues) visible at numerous nucleotide positions in many egg passages of four egg isolates. The appearance of multiple SNPs within relatively short encoding region in egg isolates (241-399 bp) is believed to have an influence on nucleocapsid polymorphism in the viral populations. These findings might suggest that the substantial SNPs in the nucleocapsid encoding region might alter the independent SNP appearance under lack of immune pressure in rest of nucleocapsid and the spike encoding regions due to changes of amino acid composition and /or alteration of the reading frames. The high polymorphism of C or T residues in all investigated egg isolates could be responsible for selection of the SNPs which were successfully continued in later egg passages. Furthermore, the most visible SNP accumulation of C or T residues was identified in the A (17 SNPs) and D (14 SNPs) isolate with relatively high detection rates compared to other SNPs within C or T residue abundant encoding region which might be an effect of substantial SNPs activity relating to lack of immune pressure. Furthermore, this theory seems to make sense after identification of the same highly polymorphic encoding region confirmed in the A1 isolate (eight SNPs) and the C isolate (11 SNPs) as C or T residue mixture. These differences in the level of diversity among SNP clusters in the egg isolates could be explained as a random accumulation of minor SNPs without a logical pattern or explanation. In addition, further investigation confirmed a correlation between appearances of C or T residue in a single or multiple egg passages within the four egg isolates. The percentage of these SNPs escalated to around 40 % in the case of continued SNPs and 60 % as independent SNPs (A and D isolate). The situation changed completely in the A1 isolate showing 75% of SNPs as continuing and 25 % as independent SNPs. According to these analysed sequence data the C isolate was also found with 27 % of continued SNPs and 63 % of independent SNPs. These differences might be due to independently passaged viral populations in embryonated chicken eggs resulting in various numbers of independent or continued SNPs. Further analyses of Sanger sequence data also presented various degrees of C or T residue mixture accumulation in egg passages among quasispecies having a potential impact on final SNP appearance in ep110 and its SNP frequency. Furthermore, the nucleotide positions 27059 and 27128 were found as an exception where SNP accumulation of C or T residues was observed in all egg passages of

the D isolate suggesting that the viral genome can be defined as a highly active SNP environment.

4.9 The SNP clusters and/ or independent SNPs within one isolate

The phenomenon of a highly polymorphic encoding region (C or T residue mixture) clearly visible in the form of SNP accumulation was identified in the nucleocapsid encoding region within the Sanger sequenced egg passages of the A, A1, C and D isolate. Interestingly, dynamics of SNPs (C or T residue mixture) were identified as random event due to a substantial number of all additional SNPs (50 %) in viral populations occurring without logical pattern across nucleocapsid encoding region. Furthermore, the absence of immune pressure in embryonated chicken eggs could have an influence in the form of a rapid response presenting a large number of independent SNPs in the viral genome visible in the early stages of Sanger sequenced egg passages across the spike and nucleocapsid encoding regions. In addition, some viral populations were identified with a low polymorphism in early egg passages such as C isolate showing that a single stranded RNA viral genome might alter by SNP accumulation during later egg passages screened by regular Sanger sequencing (Rafiei *et al.*, 2009). Therefore, SNP appearance in the form of SNP clusters or independent SNPs in later egg passages could be a sign of slower adaptation within the viral population relating to absence of immune pressure. Interestingly, the presence of SNP clusters within egg passages were detected in all four egg isolates and could be a different way of selection the minor SNPs, possibly continued by accumulation of single SNPs visible in a form of the C or T residue mixture. Further observations showed the same ability in the nucleocapsid encoding region of the C isolate where overlapping SNP clusters were observed between ep60-70 and also ep80-90. The analysis of Sanger sequenced data showed that formation of SNP clusters might be due to selection of the one or two minor SNPs which are shared within a number of egg passages.

Furthermore, S2 subunit of A isolate showed the appearance of SNP formations which might be triggered by the accumulation of independent SNPs continued in Sanger sequenced egg passages which at some point caused more polymorphism within neighbouring nucleotide positions in a particular egg passage. Interestingly, the formation of SNP clusters is not always caused by the further continuation of SNPs presented in

polymorphic encoding regions. The further analysis of Sanger sequence data also identified a large number of independent SNPs which are not associated with the presence of SNP clusters in single egg passages but presents a various degree of SNP accumulation with or without further continuation in later egg passages. The large number of SNPs has shown the viral population as a constantly active genomic environment in nucleocapsid and spike encoding regions possibly responsible for the process of attenuation. The observed situation has shown that SNPs and their accumulation in viral genome might be due to random selection with no correlation to SNP clusters or absence of immune pressure in chicken eggs (Dolz *et al.,* 2008 and Feng *et al.,* 2015).

4.10 The SNPs activity due to varying SNP frequency in ep110

The analysis of sequence data has shown a very little correlation between changes in the SNP frequency of ep110 and SNP accumulation in Sanger sequenced egg passages. However, the visible process of SNP accumulation can be detected as minor SNPs in early egg passages and consensus sequence SNPs in later egg passages of the viral populations. In addition, further investigation identified nucleotide positions where SNP appearance was detected in most of the egg passages screened by regular Sanger sequencing showing that number of SNPs in quasispecies increased over time. However, there is evidence for SNP appearance directly at consensus sequence level in early egg passages which seems to be continued in the viral population of later egg passages. Furthermore, the high SNP frequency in ep110 showed a similar trend of SNP accumulation but only in later egg passages presented by rapid accumulation in quasispecies. These data proved that high SNP frequency has only a partial influence on SNP appearance within the viral population. There is no visible pattern which could be identified as a logical explanation of SNP accumulation in Sanger sequenced egg passages relating to >80 % SNP frequency (ep110). This same observation was identified with the <80 % SNP frequency in ep110 showing minor SNP or SNPs at consensus sequence level and its different stages of accumulation in previous egg passages. However, the majority of SNPs identified in different egg passages depends mostly on the selection of minor SNPs driven by other factors responsible for viral attenuation. Furthermore, a reduced SNP frequency (<20 %) has also shown a decreasing number of detected SNPs across the viral population. In addition, further investigation could not find a correlation in SNP accumulation within egg passages relating to the

random SNPs with a low frequency (<20 %) in ep110. However, these Sanger sequence data presented SNPs which appeared more than once during the egg passages suggesting that low percentage of SNP frequency does not play a key role in selection of SNPs in viral population. In addition, the SNP activities in viral genome cannot be defined as predicable events and similar to random SNPs; it is impossible to make a prediction of SNP appearance. Furthermore, the analysis of sequenced data clearly showed that a relatively high (>50 %) SNP frequency has an increased influence on SNP accumulation in previous egg passages comparing to low (>20 %) frequency in ep110. However, this tendency cannot be used for investigation of exact SNP appearance relating to regular Sanger sequenced data of previous egg passages.

4.11 Conclusion

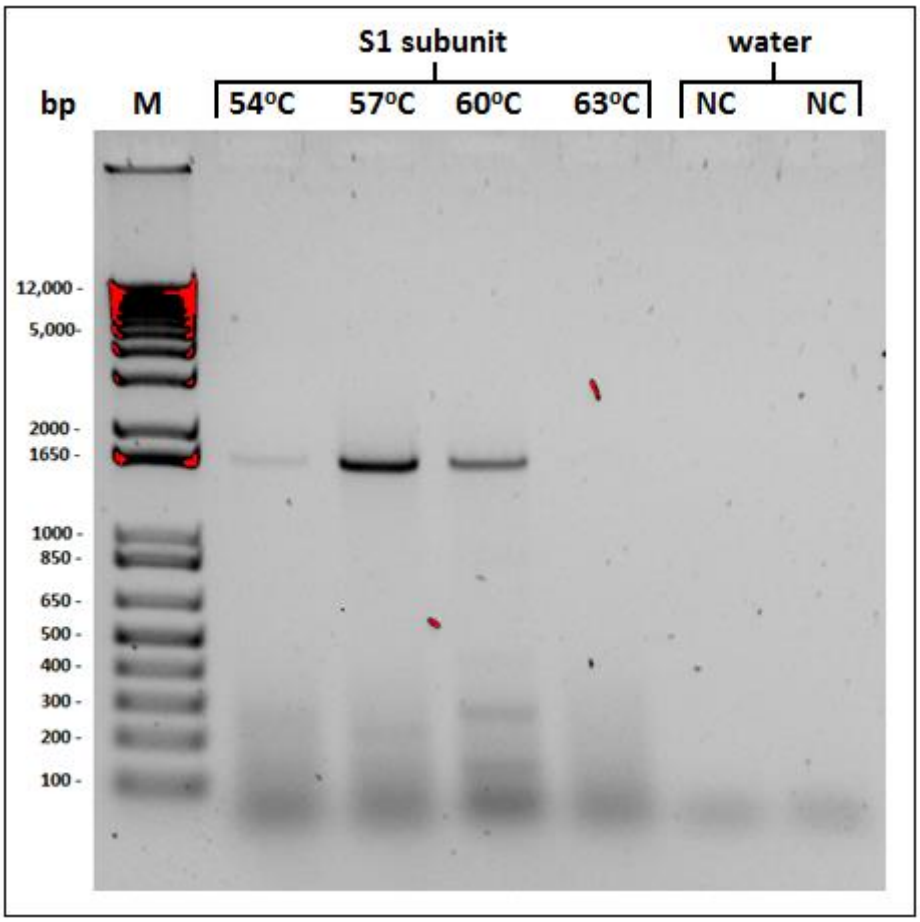
The analyses of 454 pyrosequencing data obtained from ep4 and ep110 was used as a preliminary map of attenuation for one virulent strain of IBV. The generated a deep sequence data but also Sanger sequencing outcomes from previous egg passages showed viral populations with dynamic SNPs across the viral genome. Interestingly, further analyses of Sanger sequenced data present a substantial number of SNP clusters and independent SNPs previously not identified in ep110 of A, A1, C and D isolates. In addition, the high activity in viral genomes identified in previous egg passages of spike and nucleocapsid encoding regions might have an impact on SNP appearance in pyrosequenced ep110s. Therefore, the multiple ways of SNP formation and their accumulation among egg passages were analysed as the potential reason of attenuation. Interestingly, only a fraction of identified SNPs were continued as polymorphisms in previous egg passages. In addition, those genomic mutations identified as independent SNP and/or SNP clusters in regular Sanger sequenced data were not found in ep110 deep sequenced data. Furthermore, the 454 pyrosequencing dataset has shown the limited number of SNPs relating to frequency greater than 50 % in each egg isolate. In addition, the large number of SNP clusters and also the independent SNPs in Sanger sequenced data might have an influence on further changes in the viral genome defined with deep sequence data. Interestingly, dynamics in previous egg passages were observed within all egg isolates but only as few as nine SNPs with >50 % frequency might be a trigger for viral attenuation. However, the process of SNP accumulation in previous egg passages and its influence on SNP appearance in pyrosequenced ep110 is unknown. The genomic changes in early egg passages and its

continuation in later egg passages might have no connection with the SNPs identified with various frequencies in ep110. Definitely, the common SNPs with >50% frequency (ep110) identified in nucleocapsid and spike encoding region of more than one egg isolate can be classified as potential candidates playing a key role in attenuation process and should be considered as crucial in universal vaccine development after further testing. Therefore, the conclusion clearly states that only a few SNPs in viral genome might be responsible for attenuation process but the forces responsible for the degree of SNP accumulation and its number is still the point of the discussion.

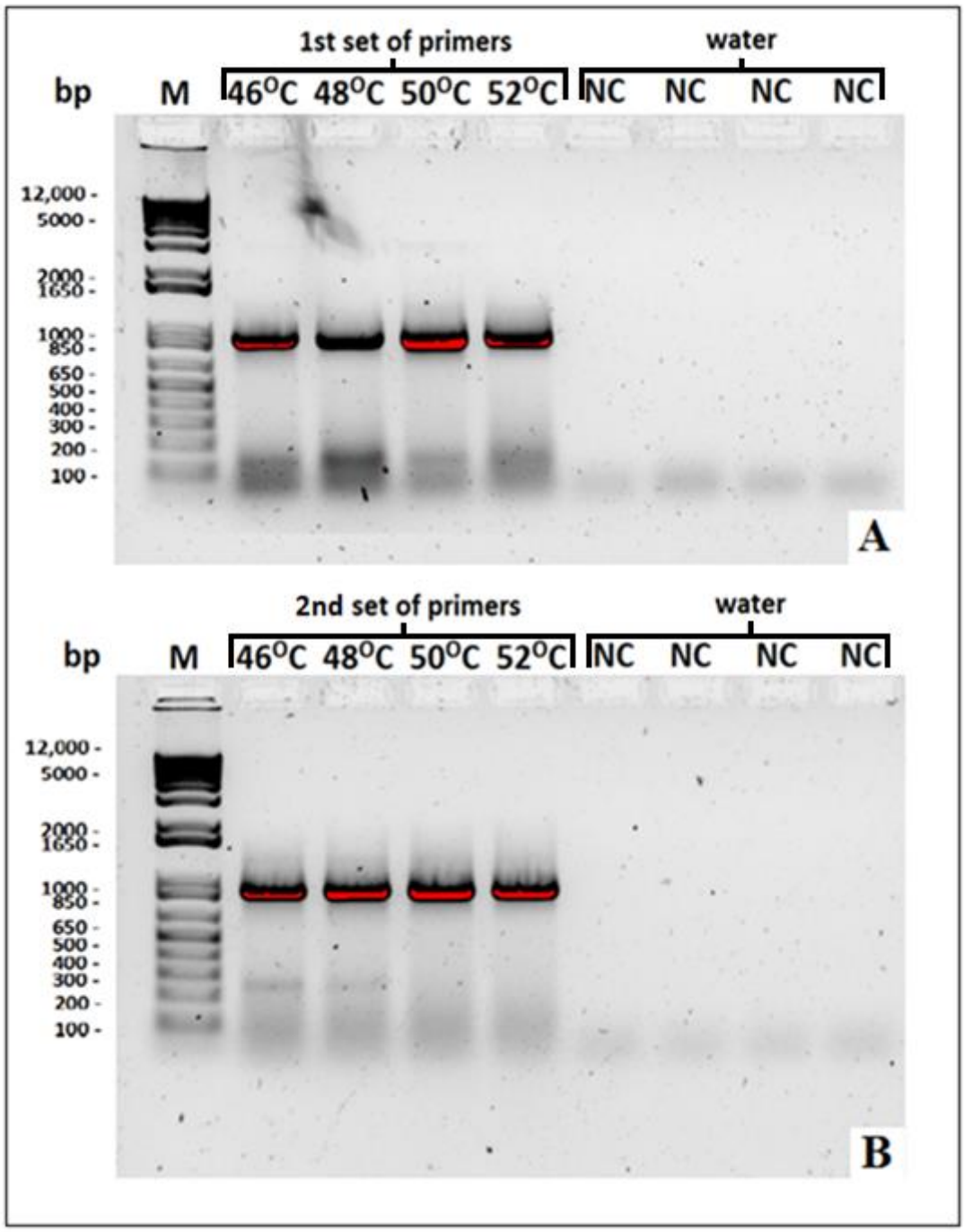
Appendices

Appendix 1. The details of primers used in amplification of IBV genome regions with the highest number of identified SNPs. The GC % content varies from 40-60 % within 18-24nucletides. The annealing temperature varies from 50 ^{o}C in the case of primers for S2 subunit (part 1) to 68 ^{o}C for S2 subunit (part 2) primers.

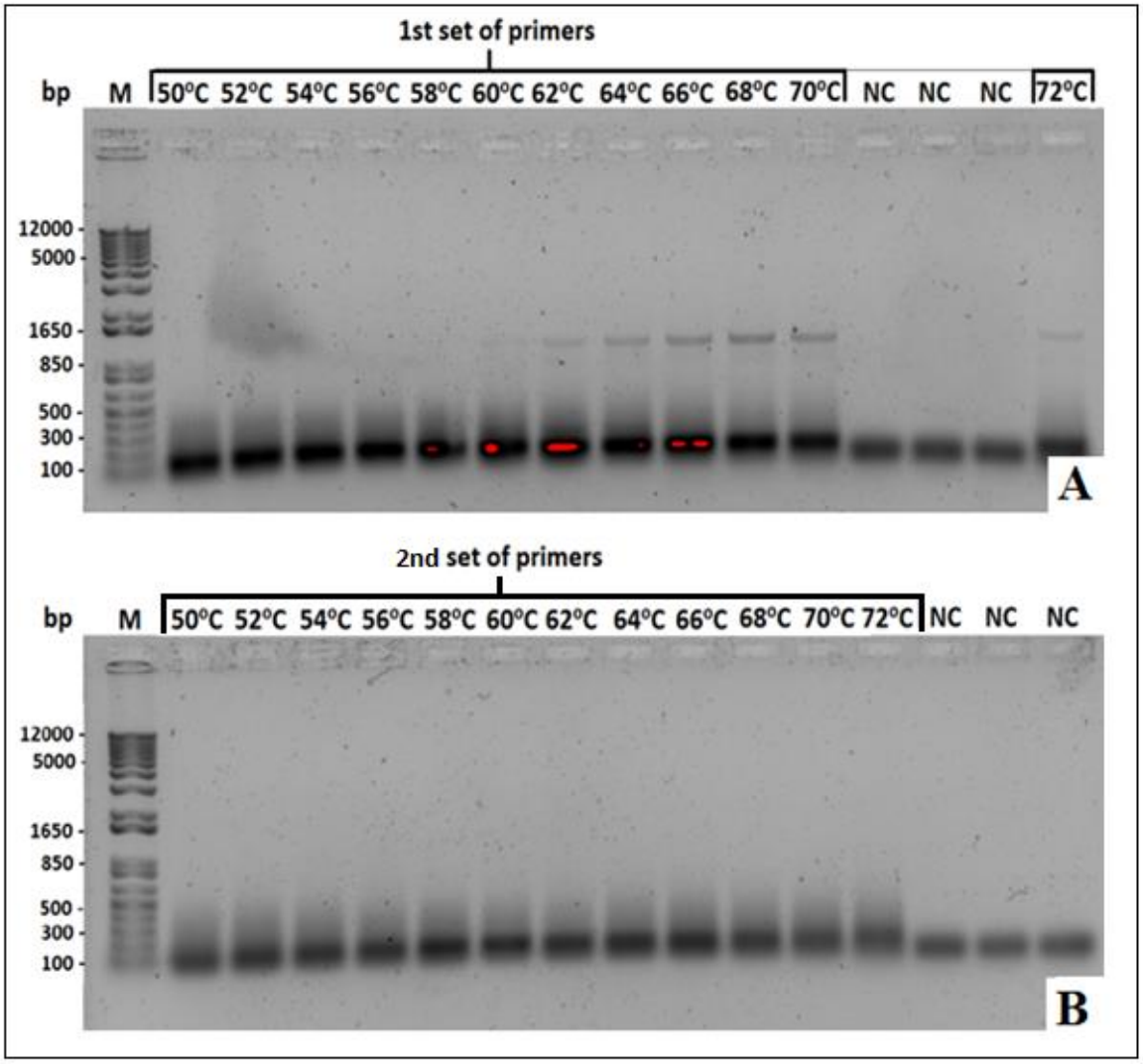
Amplicon	Primer name	Annealing temp (^{o}C)	Sequence 5` to 3`
S1 subunit	S1_forward	57	TGAAAACTGAACAAAAGACAGA
S1 subunit	S1_reverse	57	GGGCAATTTGCAACATTTTCAG
S2 subunit part1	S2(p1)_forward	50	ATGGTTTGCTTGTGTTGCCTCCC
S2 subunit part1	S2(p1)_reverse	50	TTCCAGGAGCTAAGGGTACAGCC
S2 subunit part2	S2(p2)_forward	68	ATGGTTTGCTTGTGTTGCCT
S2 subunit part2	S2(p2)_reverse	68	AGGAAGGACGTGGGACTTTG
Nucleocapsid	N_forward	57	CCAGGGGAAAACTTGTGAGG
Nucleocapsid	N_reverse	57	TACCGTTCGTTTCCAGGCTA



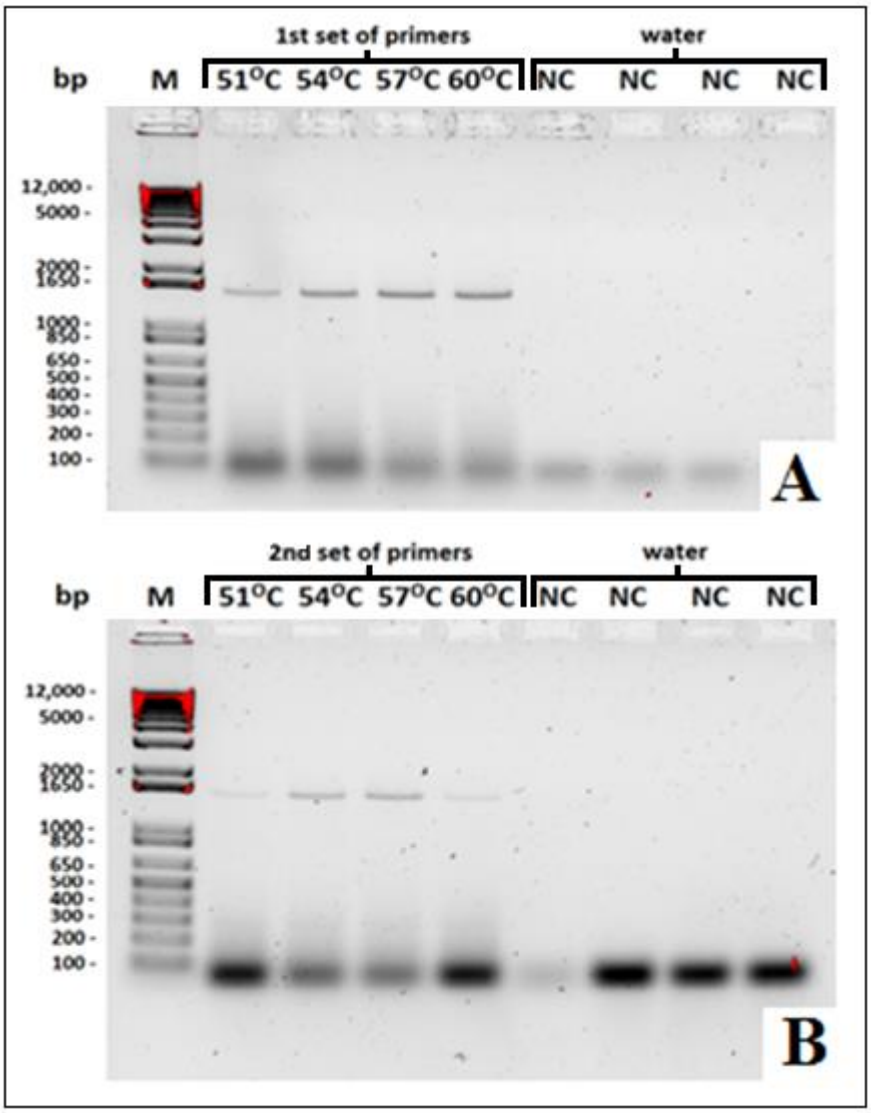
Appendix 2. The best optimum annealing temperature for 2nd set of primers for successful amplification of S1 subunit. 1 % agarose gel was used. Titles from left to right for gel image: 1kb plus ladder (M), 54 ^{o}C, 57 ^{o}C, 60 ^{o}C, 63 ^{o}C and negative control (NC, water). Approximate base pair size of product is shown on the left.



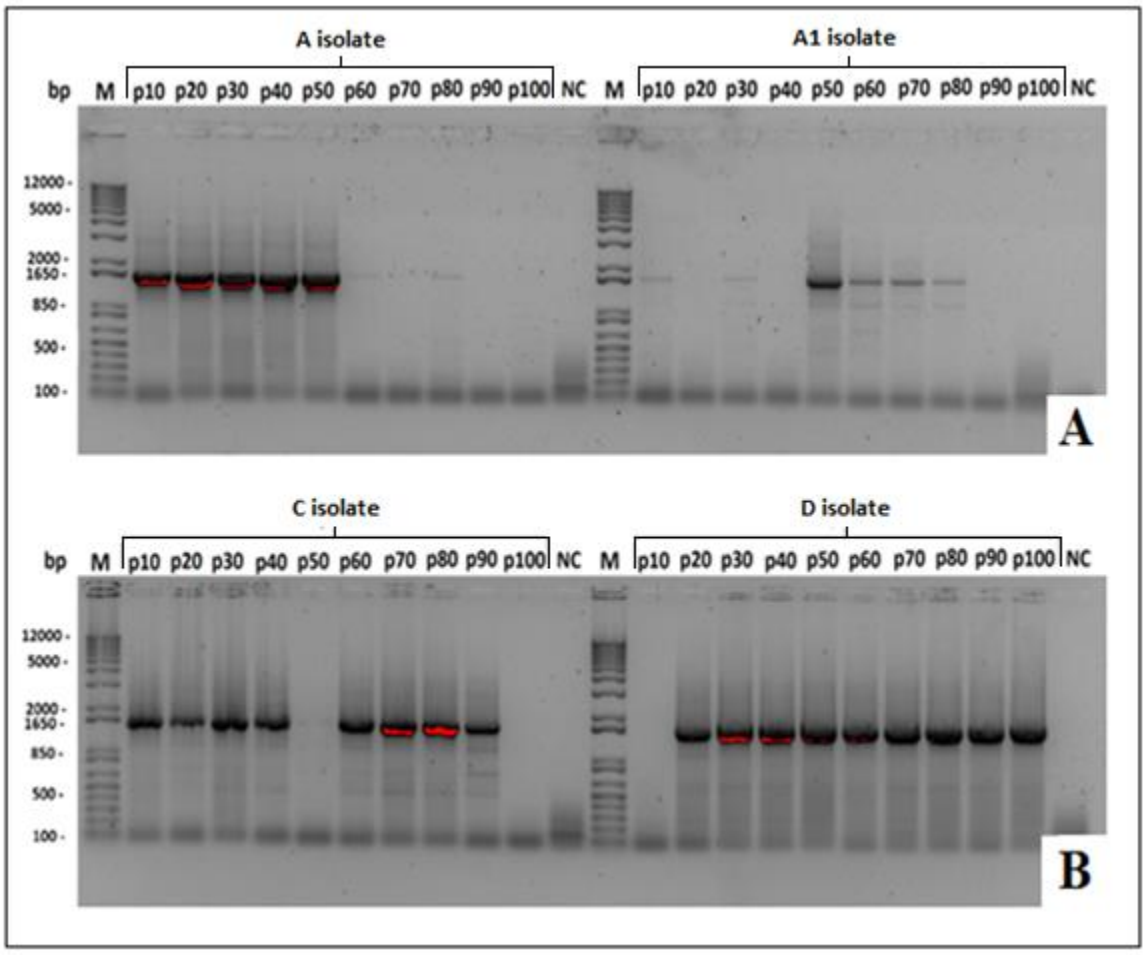
Appendix 3. The gel image S2 subunit part1 amplification. 1 % agarose gel was used. Blot A = 1st set of primers; Blot B 2nd set of primers. Titles from left to right in gel image: 1 kB plus ladder, 46 °C, 48 °C, 50 °C, 52 °C and four negative controls (NC, water). Approximate base pair size of product is shown on the left.



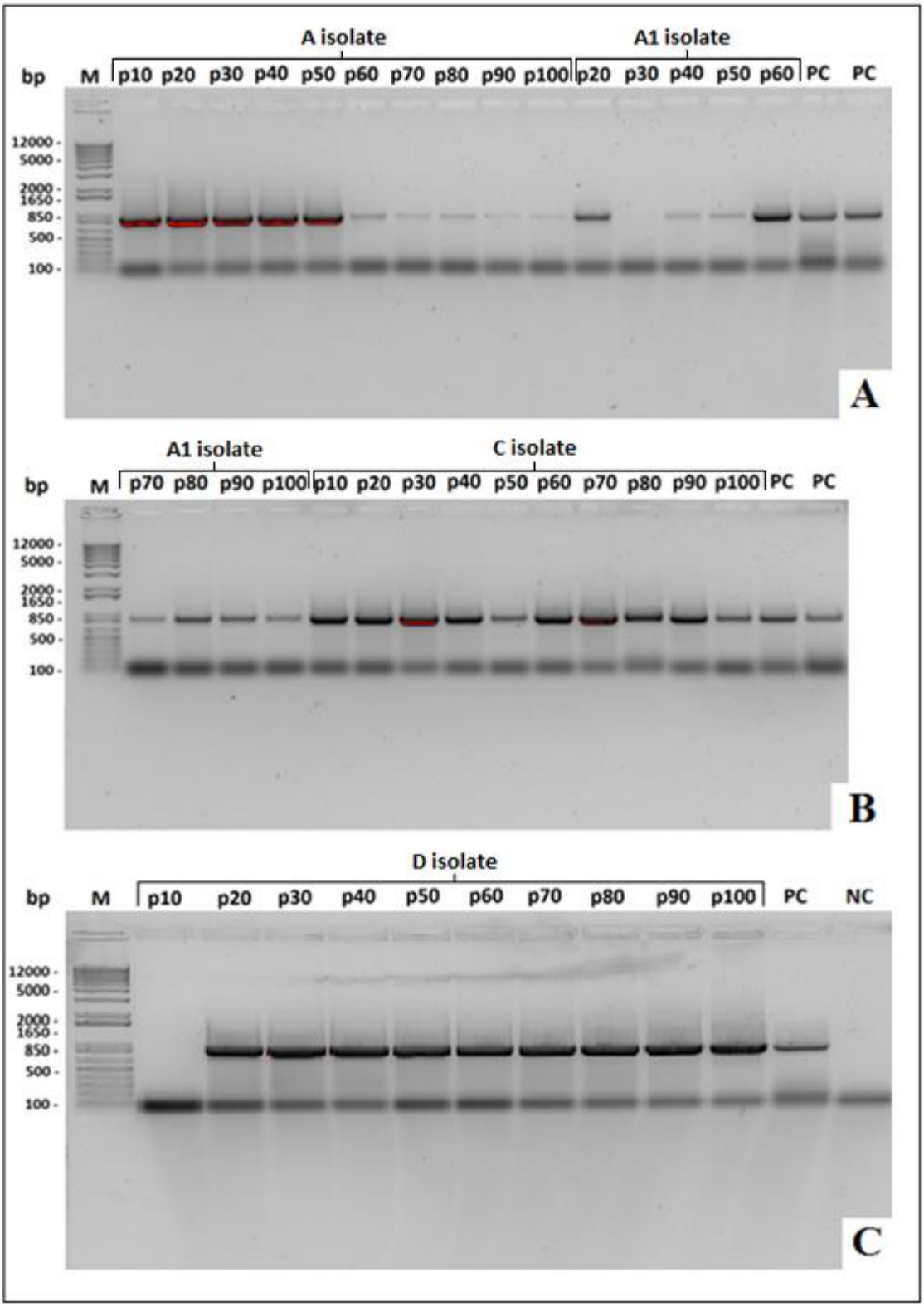
Appendix 4. The results of investigation the best optimum annealing temperature for S2 subunits (part 2). 1 % agarose gel was used. Titles from left to right in gel image: 1kb plus ladder (M), 50 ^{o}C, 52 ^{o}C, 54 ^{o}C, 58 ^{o}C, 60 ^{o}C, 62 ^{o}C, 64 ^{o}C, 66 ^{o}C, 68 ^{o}C, 70 ^{o}C and three negative controls (NC, water). Approximate base pair size of product is shown on the left. Blot A = 1st set of primers; Blot B = 2nd set of primers.



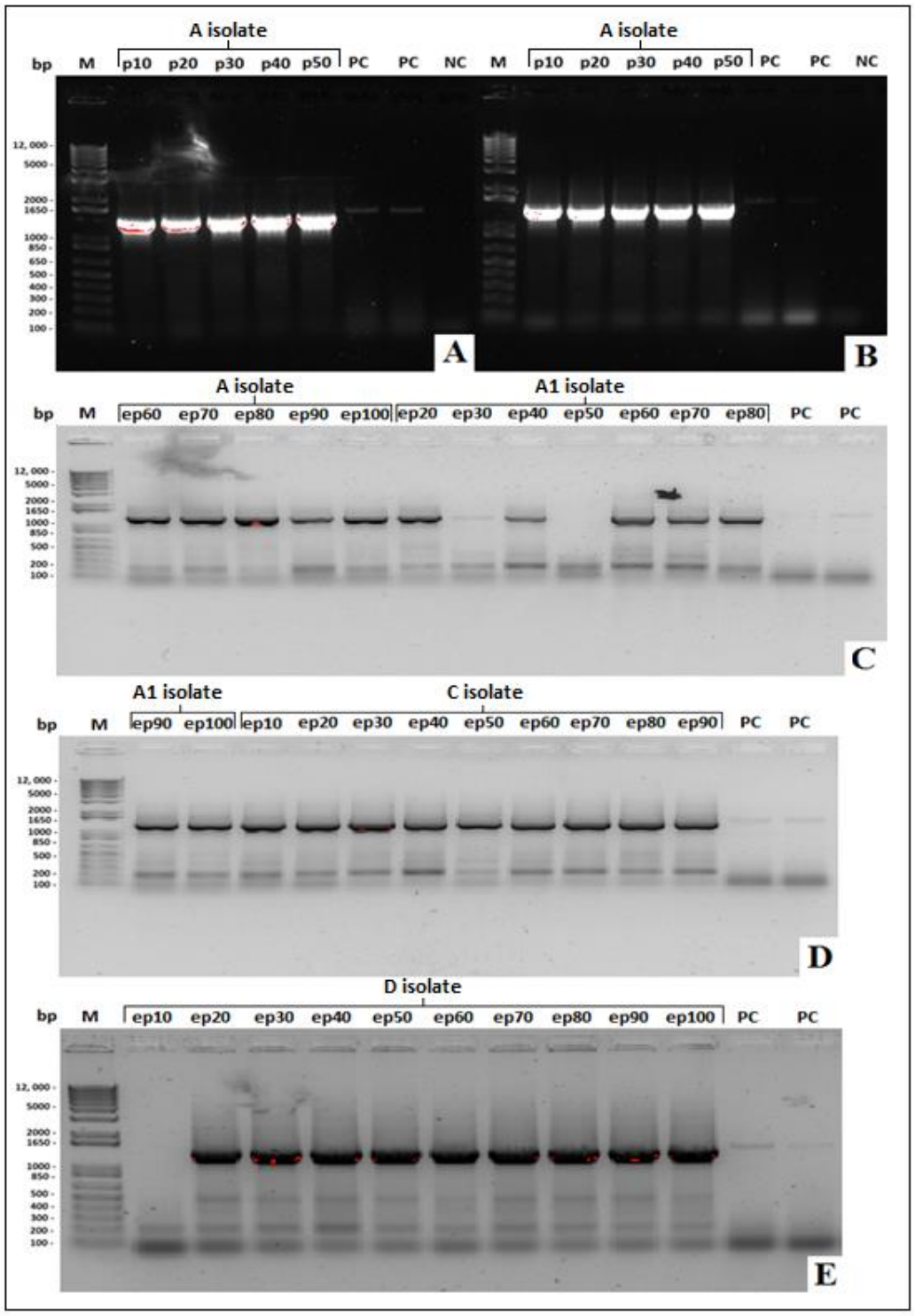
Appendix 5. The agarose gel image showing the nucleocapsid amplification with two set of primers. 1 % agarose gel was used to investigate the best optimum annealing temperature. Blot A = 1st set of primers; Blot B = 2nd set of primers. Titles from left to right in gel image: 1kb plus ladder (M), 51 ^{o}C, 54 ^{o}C, 57 ^{o}C, 60 ^{o}C and four negative controls (NC, water). Approximate base pair size of product is shown on the left.



Appendix 6. The agarose gel image of S1subnit amplification with 2nd set of primers. 1 % agarose gel with the best optimum annealing temperature at 57 ^{o}C. M = 1kb plus ladder (approximate product size in bp are indicated). Blot A = A and A1 isolate; Blot B= C and D isolate. A and B blots including egg passages (ep10-100) for each egg isolate. NC= negative control (water).



Appendix 7. The agarose gel image of S2 subunit (part 1) amplification with 2nd set of primers. 1 % agarose gel with the best optimum annealing temperature at 50 ºC. M = base pair size marker (approximate product size in bp are indicated). Blot A = A isolate and part of A1; Blot B = A1 and C isolate; Blot C = D isolate. Blots A, B and C including egg passages (ep10-100) for each egg isolate. PC = positive control (Vaccinia virus); NC = negative control.



Appendix8. The agarose gel image of Nucleocapsid amplification. 1 % agarose gel with the best optimum annealing temperature at 60 ^{o}C. M = base pair size marker (approximate product size in bp are indicated). Blot A = 1st set of primers (A isolate); Blot B = 2nd set of primers (A isolate), Blot C = 1st set of primers (A and A1 isolate), Blot D = 1st set of primers (A1 and C isolate) and Blot E = 1st set of primers (D isolate). Blots A-E including egg passages for all investigated isolates. PC = positive control (Vaccinia virus); NC = negative control (water).

Appendix 9. The Nanodrop™ measurements of RNA samples extracted from egg passages. The "n/a" written in red colour presents samples excluded from measurements.

A isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
1	10	2.8	1.58	0.01
2	20	3.4	1.78	0.03
3	30	7.7	1.65	0.27
4	40	4.5	1.41	0.14
5	50	5.3	1.91	0.03
6	60	3.0	2.01	0.03
7	70	5.3	1.89	0.11
8	80	7.2	1.80	0.04
9	90	3.5	2.18	0.02
10	100	5.8	2.01	0.04
A1 isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
11	10	**n/a**	**n/a**	**n/a**
12	20	6.9	1.82	0.08
13	30	4.2	1.81	0.10
14	40	3.7	2.06	0.06
15	50	4.7	1.83	0.03
16	60	5.7	1.87	0.05
17	70	3.1	2.04	0.04
18	80	7.6	1.87	0.09
19	90	2.1	2.69	0.03
20	100	1.2	3.42	0.01
C isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
21	10	4.8	2.17	0.03
22	20	3.7	1.73	0.04
23	30	3.5	1.91	0.11
24	40	4.1	1.83	0.64
25	50	4.2	1.62	0.03
26	60	4.7	1.45	0.01
27	70	3.1	1.82	0.06
28	80	6.5	1.66	0.07
29	90	2.3	1.75	0.05
30	100	1.5	2.36	0.02
D isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
31	10	4.3	1.87	0.03
32	20	6.0	1.81	0.07
33	30	5.1	1.68	0.08
34	40	5.7	1.82	0.07
35	50	9.9	1.98	0.11
36	60	4.3	1.88	0.01
37	70	4.7	1.76	0.08
38	80	7.4	1.75	0.12
39	90	7.8	1.87	0.05
40	100	2.5	2.72	0.04

Appendix 10. Nanodrop™ readings of the amplification process (S1 subunit -2nd set of primers). The "n/a" written in red colour presents samples excluded from measurements.

A isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
1	10	30.9	1.82	0.26
2	20	58.7	1.75	1.57
3	30	53.9	1.73	1.36
4	40	78.2	1.72	1.34
5	50	66.1	1.73	1.37
6	60	33.5	1.62	0.76
7	70	22.4	1.61	0.67
8	80	41	1.80	0.33
9	90	32.9	1.62	0.89
10	100	37.7	2.11	0.04
A1 isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
11	10	**n/a**	**n/a**	**n/a**
12	20	20.5	1.58	0.96
13	30	21.2	1.59	0.74
14	40	23	1.73	0.99
15	50	26.2	1.6	0.82
16	60	36.4	1.66	0.99
17	70	30.3	1.65	0.82
18	80	16.9	1.75	1.21
19	90	19.4	1.71	1.13
20	100	23.6	1.58	0.92
C isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
21	10	26.7	1.66	0.81
22	20	27.8	1.63	0.98
23	30	61.6	1.69	1.06
24	40	32.3	1.68	1.13
25	50	18.8	1.78	1.16
26	60	21.3	1.74	1.25
27	70	34.7	1.71	1.37
28	80	44.1	1.81	1.72
29	90	36.7	1.61	0.71
30	100	21.5	1.62	0.88
D isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
31	10	17.6	1.75	1.49
32	20	49.7	1.74	1.25
33	30	68.4	1.79	1.50
34	40	61.4	1.84	1.91
35	50	63.4	1.78	1.52
36	60	45.3	1.75	1.35
37	70	86.5	1.72	1.26
38	80	67.1	1.82	1.86
39	90	76.1	1.73	1.38
40	100	80.1	1.74	1.36

Appendix 11. The summary of S2 subunit (part1) DNA concentration – 2nd set of primers by Nanodrop™ measurement. The "n/a" written in red colour presents samples excluded from measurements.

A isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
1	10	36.4	1.92	1.42
2	20	50.4	1.85	0.43
3	30	36.1	1.83	1.11
4	40	28.5	1.82	0.56
5	50	29.9	1.79	0.97
6	60	36.4	1.84	1.43
7	70	31.1	1.84	1.12
8	80	42.5	1.84	1.51
9	90	36.3	1.91	1.33
10	100	39.8	1.86	0.84
A1 isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
11	10	**n/a**	**n/a**	**n/a**
12	20	8.9	2.27	0.05
13	30	30.8	1.84	0.33
14	40	39.5	1.87	0.18
15	50	43.5	1.90	0.40
16	60	23.9	1.78	0.16
17	70	12.2	2.14	0.11
18	80	9.8	1.64	0.06
19	90	11.2	1.96	0.01
20	100	28	2.17	0.03
C isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
21	10	25.3	1.96	0.03
22	20	20.9	1.85	0.16
23	30	37.2	1.79	0.15
24	40	17.1	1.91	0.06
25	50	37.1	1.83	0.08
26	60	19.9	1.89	0.08
27	70	25.9	1.98	0.09
28	80	16.8	1.79	0.12
29	90	15.8	1.96	0.04
30	100	9.6	2.01	0.03
D isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
31	10	**n/a**	**n/a**	**n/a**
32	20	28	1.99	0.07
33	30	31.4	1.93	0.28
34	40	31.8	1.96	0.06
35	50	39.9	3.31	0.26
36	60	25.5	2.31	0.03
37	70	22.8	1.95	0.42
38	80	17.6	2.02	0.25
39	90	16.5	2.10	0.14
40	100	29.9	1.97	0.09

Appendix 12. The summary from Nanodrop™ readings of nucleocapsid DNA concentration after amplification process. The "n/a" written in red colour presents samples not included in measurement.

A isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
1	10	20.8	2	0.07
2	20	24.8	2.15	0.05
3	30	39.2	1.94	0.15
4	40	31.9	1.82	0.17
5	50	30	1.80	0.10
6	60	25.3	1.89	0.12
7	70	31.5	1.84	0.23
8	80	31.6	1.91	0.15
9	90	29.4	2.10	0.05
10	100	21.4	1.87	0.11
A1 isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
11	10	**n/a**	**n/a**	**n/a**
12	20	15.7	2.65	0.02
13	30	28.7	1.91	0.18
14	40	45.7	1.91	0.15
15	50	40.3	1.93	0.13
16	60	36.8	1.96	0.14
17	70	39.8	1.96	0.17
18	80	29.5	2.02	0.09
19	90	40.2	1.93	0.08
20	100	18.8	2.13	0.05
C isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
21	10	17.9	2.06	0.04
22	20	13.7	2.29	0.02
23	30	28.4	2.01	0.05
24	40	18.3	2.15	0.02
25	50	16.1	2.13	0.02
26	60	20.5	2.18	0.03
27	70	28.6	1.96	0.03
28	80	24.6	2.22	0.03
29	90	21.1	2.09	0.03
30	100	22.3	1.80	0.05
D isolate				
Sample #	**Egg passage #**	**Concentration (ng/µl)**	**260/280 ratio**	**260/230 ratio**
31	10	**n/a**	**n/a**	**n/a**
32	20	18.2	1.6	0.53
33	30	27	1.6	0.83
34	40	29.6	1.71	0.69
35	50	20.5	1.69	0.13
36	60	22.2	1.6	0.73
37	70	29.9	1.65	0.84
38	80	26.7	1.66	0.79
39	90	21.7	1.77	0.08
40	100	19.9	1.72	0.58

C isolate											
Nucleotide #	**ep4**	**ep10**	**ep20**	**ep30**	**ep40**	**ep50**	**ep60**	**ep70**	**ep80**	**ep90**	**ep100**
25969	G									R	
25970	T									K	
26039	T						C				
26045	C						T				
26048	A	W									
26049	G						C				
26051	A						C	M			
26052	G				T						
26053	T							Y			
26056	G						T				
26057	C						T				
26058	T							W			
26060	C	W									
26061	A							Y			
26062	C	W					T				
26066	C							S			
26067	G						T				
26068	C						T				
26069	C						G				
26077	G	R									
26081	G				T		W				
26082	T						C	R			
26083	A						C			M	
26088	T				A						
26089	NIL	T									
26091	T	W									
26092	C						G				
26096	A						W				
26100	A				T		T				
26101	G						A				R
26102	A						G				
26103	A						G				
26104	A						G				
26105	A				T						
26106	A				C						
26107	T							K			
26108	C				G						
26118	G										
26122	G							T			
26123	C										
26124	A							C			
26125	G							C			
26132	T							G			
26133	A	C									
26137	G	T									
26145	A							T			
26146	A							T			
26147	G							A			
26149	T							C			
26150	A							T			
26151	G							K			
26156	A										
26156	A							C			
26158	G	C						T			
26159	C							G			
26164	T							G			
26167	A							T			
26170	T								G		

26172	G	T								T	
26173	A									T	
26177	A	C	W								
26179	C	A	A	A					A	A	
26180	C									T	
26181	C	T								A	
26182	A	T									
26183	A								C		
26184	T	C		A						C	
26185	T								A	A	
26188	A			T						T	
26189	G									C	
26191	A								C		
26193	A			T							
26194	T								G		
26195	T			A						G	
26198	T			C					G		
26200	T		A								
26211	A		T								
26227	C		T								
26238	A		C								
26239	A		C								
26285	C										Y
26365	G										A
26376	NIL										G
26513	NIL										T
26517	G		T								
26526	G		A								
26537	T		C								
26549	G		A								
26568	A		C								
26573	G		A								
26586	G		A								
26603	G		A								
26609	C		G								
26610	A		G								
26677	G			A							
26683	A								T		
26687	T								G		
26691	G								A		
26693	T								G		
26699	G								A		
26700	A								G		
26701	T								A		
26718	NIL								T		
26720	A								G		
26725	G							C			
26727	C							G			
26730	G							T			
26732	T							G			
26737	G	A									
26738	A							G			
26740	T							A			
26742	C							G			
26747	C							A			
26748	C							A			
26750	A							G			
26754	A		W					T			
26755	C							T			
26762	G									A	

26765	T							G			
26766	G							C			
26767	C							G			
26768	T							G			
26773	T							A			
26774	C			M							
26784	T	W									
26804	A							G			
26807	A							G			
26810	T							A			
26822	A				C						
26823	NIL				A						
26857	C									M	
26879	A		W								

Appendix 13. The table below presents the polymorphism in a form of independent SNPs and SNP clusters found as additional SNPs in the nucleocapsid encoding region of C isolate. Orange = minor SNP, red = lack of SNP and green = SNP at consensus sequence level.

Appendix 14. The table used for identification minor SNPs found within egg passages using regular Sanger sequencing method.

IUPAC Code	Meaning	IUPAC Code	Meaning
A	A	**Y**	C or T
C	C	**K**	G or T
G	G	**V**	A or C or G
T	T	**H**	A or C or T
M	A or C	**D**	A or G or T
R	A or G	**B**	C or G or T
W	A or T	**N**	G or A or T or C
S	C or G	-	-

Glossary

Consensus sequence SNP or SNP at consensus sequence level – defined nucleotide change observed at investigated genome position in sequence alignment with >50 % SNP frequency.

Early egg passages – term used to describe the SNP investigation in Sanger sequenced data from ep10 to ep50.

Later egg passages – the egg passages from ep60 to ep100 screened with Sanger sequencing.

Minor SNP – the occurrence of more than one different nucleotide molecules at this same genomic position relating to genome of viral population with <50 % SNP frequency.

SNP cluster – the group of a minimum three SNPs within the IBV genome which are close to each other (<50 bp) and separated by the gap of 100 bp from the nearest SNP

List of references

1. Ababneh, M., Dalab, A. E., Alsaad, S. & Al-Zghoul, M. 2012. Presence of Infectious Bronchitis Virus Strain CK/CH/LDL/97I in the Middle East. *Vet. Sci*. 22-28.
2. Alexander, D.J., 2000. A review of avian influenza in different bird species. *Avian Virology*, 74, 3-13.
3. Ayala, S., Revelo, M., Barragán, V., Chiriboga, J., Torres, A., Iván Santiana, I. and Trueba, G. 2014. Detection and characterization of infectious laryngotracheitis virus from an outbreak of respiratory disease in Ecuador. *Intern. J. Appl. Res. Vet. Med.* 3, 345-248.
4. Bagust, T. J., and Johnson, M. A. 1995. Avian infectious laryngotracheitis: virus-host interactions in relation to prospects for eradication. *Avian Pathology*, 24, 373-391.
5. Bang, I. A., Foard, M. and Bang, D. G. 1974. Acute Newcastle viral infection of the upper respiratory tract of the chicken. *Am. J. Pathol*. 76, 333-348.
6. Barrett, A. D. T. 2016. Next Generation sequencing of yellow fever 17D vaccine virus. *Clin. Vaccine Immunol.* 5, 196-211.
7. Blacker, H. P., Kirkpatrick, N. C., Rubite, A., O`Rourke, D., Noormohammadi, A. H. 2011. Epidemiology of recent outbreaks of infectious laryngotracheitis in poultry in Australia. *Australian Veterinary Journal*, 3, 89-96.
8. Bourogaa, H., Larbi, I., Miled, K., Hellal, Y.K., Hassen, J., Behi, I., Nsiri, J. & Ghram, A. 2014. Evaluation of protection conferred by a vaccination program based on the H120 and CR88 commercial vaccines against a field variant of avian infectious bronchitis virus. *J. Appl. Poult. Res*, 23.
9. Britton, P., Armesto, M., Cavanagh, D. & Keep, S. 2012. Modification of the avian coronavirus infectious bronchitis virus for vaccine development. *Bioeng. Bugs*, 3, 114-9.
10. Britton, P., Evans, S., Dove, B., Davies, M., Casais, R. & Cavanagh, D. 2005. Generation of a recombinant avian coronavirus infectious bronchitis virus using transient dominant selection. *J. Virol. Methods*, 123, 203-11.
11. Buynak, E. B. and Hilleman, M. R. 1966. Live attenuated Mumps Virus vaccine 1. Vaccine development. *Exp Biol Med.* 123 (3), 768-775.

12. Capua, I. and Alexander, D. J., 2008. Avian influenza vaccine and vaccination in birds. *Vaccine,* 26, 70-73.
13. Casais, R., Thiel, V., Siddell, S. G., Cavanagh, D. & Britton, P. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *Journal of Virology*, 75, 12359-12369.
14. Cook J. K. A., Orbell S. J., Woods M. A., Huggins M.B. 1999. Breadth of protection of the respiratory tract provided by different live-attenuated infectious bronchitis vaccines against challenge with infectious bronchitis viruses of heterologous serotypes. *Avian Pathology*, 28, 477-485.
15. Cook, J. K. A., Jackwood, M. & Jones, R. C. 2012. The long view: 40 years of infectious bronchitis research. *Avian Pathology*, 41, 239-250.
16. Cook, J. K.A. and Ellis, M. M. 1990. Attenuation of turkey rhinotracheitis virus by alternative passage in embryonated chicken eggs and tracheal organ cultures. *Avian pathology,* 19, 181-185.
17. Corbanie, E. A., Matthijs, M. G. R., Van Eck, J. H. H., Remon, J. P., Landman, W. J. M. and Varvaet, C. 2007. Deposition of differently sized airborn microspheres in the respiratory tract of chickens. *Avian Pathology,* 35, 475-485.
18. Costa-Hurtado, M., Afonso, C. L., Miller, P. J., Shepherd, E., Cha, R. M., Smith, D., Spackman, E., Kapczynski, D. R., Suarez, D. L., Swayne, D. E. and Pantin-Jackwood, M. J. 2015. Previous infection with virulent strains of Newcastle disease virus reduces highly pathogenic avian influenza virus replication, disease, and mortality in chickens. *Veterinary Research*, 46.
19. Crinion, R.A.P. and Hofstad, M.S. 1971. Pathogenicity of two embryo-passage levels of avian infectious bronchitis virus for the oviduct of young chickens of various ages. Avian Diseases, 16, 967–973.
20. De Wit, J. J., Cook, J. K. A. and Van Der Heijden, H. M. J.F. 2011. Infectious bronchitis virus variants: a review of the history, current situation and control measures. *Avian pathology,* 40 (3), 223-235.
21. Dhama, K., Chakraborty, S., Mohd, M., Wani, Y., Verma, A. K., Deb, R., Tiwari, R. and Kapoor, S. 2013. Novel and emerging therapies safeguarding health of humans and their companion animals: a review. *Pakistan Journal of Biological Sciences*, 16, 101-111.
22. Dolz, R., Pujols, J., Ordonez, G., Porta, R. & Majo, N. 2008. Molecular epidemiology and evolution of avian Infectious Bronchitis Virus in Spain over a fourteen-year period. *Virology,* 374, 50-59.

23. Dortmans, J. C. F. M., Rottier, P. J. M., Koch, G. Peeters, B. P. H. 2011. Passing of a Newcastle disease virus pigeon variant in chickens results in selection of viruses with mutations in the polymerase complex enhancing virus replication and virulence. *Journal of General Virology*, 92, 336-345.
24. Dortmans, J. C. F. M., Rottier, P. J. M., Koch, G. Peeters, B. P. H. 2010. The viral replication complex is associated with the virulence of Newcastle disease virus. *Journal of Virology*, 84, 10113-10120.
25. Enders, J.F., Levens, J. H., Stokes Jr, J., Maris, E. P. and Berenberg, W. 1946. Attenuation of virulence with retention of antigenicity of Mumps virus after passage in the embryonated egg. *The Journal of Immunology*, 54 (3), 283-291.
26. Fan, H., Ooi, A., Tan, Y. W., Wang, S., Fang, S., Liu, D. X. & Lescar, J. 2005. The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure*, 13, 1859-68.
27. Farsang, A., Ros, C., Renstrom, L. H. M., Baule, C., Soos, T. & Belak, S. 2002. Molecular epizootiology of infectious bronchitis virus in Sweden indicating the involvement of a vaccine strain. *Avian Pathology*, 31, 229-236.
28. Feng, K., Xue, Y., Wang, J.,Chen, W., Chen, F., Bi, Y. & Xie, Q. 2015. Development and efficacy of a novel live-attenuated QX-like nephropathogenic Infectious Bronchitis Virus vaccine in China. *Vaccine*, 233, 1113-20.
29. Fouchier, R. A. M., Schneeberger, P. M., Rozendaal, F. W., Broekman, J. M., Kemink, S. A. G., Munster, V., Kuiken, T., Rimmelzwaan, G. F., Schutten, M., Van Doornum, G. J. J., Koch, G., Bosman, A., Koopmans, M. and Osterhaus, A. D. M. E. 2004. Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *PNAS*, 5, 1356-1361.
30. Geerlig, H. J., Boelm, G. J., Meinder, C. A. M., Stuurman, B. G. E., Symons, J., Tarres-Call, J., Bru, T., Vila, R., Mombarg, M., Karaca, K., Wijmenga, W. & Kumar, M. 2011. Efficacy and safety of an attenuated live QX-like infectious bronchitis virus strain as a vaccine for chickens. *Avian Pathology*, 40, 93-102.
31. Glisson, J. R. 1998. Bacterial respiratory diseases of poultry. *Poultry Science*, 77, 1139-1142.

32. Guan, Y., Shortridge, K. F., Krauss, S., Chin, P. S., Dyrting, K. C., Ellis, T. M., Webster, R. G. and Peiris, M. 2000. H9N2 influenza viruses possessing H5N1-like internal genomes continue to circulate in poultry in southeastern China. *Journal of Virology*, 20, 9371-9380.
33. Hafez, M. H., Arafa, A., Abdelwhab, E. M., Selim, A., Khoulosy, S. G., Hasson, M. K. and Aly, M. M. 2010. Avian influenza H5N1 virus infections in vaccinated commercial and backyard poultry in Egypt. *Poultry Industry*, 89, 1609-1613.
34. Hall, T. 2013. BioEdit (Version 7.2.5) [Computer program]. Available at http://www.mbio.ncsu.edu/BioEdit/page2.html (Accessed 9 January 2015).
35. Hanada, K., Suzuki, Y., and Gojobori, T. 2004. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol. Biol. Evol.* 21(6), 1074–1080.
36. Hewson, K. A., Noormohammadi, A. H., Devlin, J. M., Browning, G. F., Schultz, B. K. & Ignjatovic, J. 2014. Evaluation of a novel strain of infectious bronchitis virus emerged as a result of spike gene recombination between two highly diverged parent strains. *Avian Pathol*. 43, 249-57.
37. Hodgson, T., Casais, R., Dove, B., Britton, P. & Cavanagh, D. 2004. Recombinant infectious bronchitis coronavirus Beaudette with the spike protein gene of the pathogenic M41 strain remains attenuated but induces protective immunity. *Journal of Virology*, 78, 13804-13811.
38. Holmes, E. C, 2003. Molecular clocks and the puzzle of RNA virus origins. *Journal of Virology*, 77 (7), 3893–3897.
39. Hopkins, S.R. and Yoder, H.W. 1984. Increased incidence of airsacculitis in broilers infected with Mycoplasma synoviae and chicken-passaged infectious bronchitis vaccine virus. Avian Diseases, 28, 386–396.
40. Huang, Y. P. and Wang, C. H. 2007. Sequence changes of infectious bronchitis virus isolates in the 3’ 7.3 kb of the genome after attenuating passage in embryonated eggs. Avian Pathology, 36, 59–67.
41. Huang, Z., Elankumaran, S., Panda, A. and Samal, S. K. 2003. Recombinant Newcastle disease virus as a vaccine vector. *Poultry Science*, 82, 899–906.
42. Ignjatovic, J., Ashton, D. F., Reece, R., Scott, P. & Hooper, P. 2002. Pathogenicity of Australian strains of avian infectious bronchitis virus. *Journal of Comparative Pathology*, 126, 115-123.

43. Jackwood, M. W. 2012. Review of infectious bronchitis virus around the world. *Avian Dis*. 56, 634-41.
44. Jia, W., Wang, X., Parrish, C. R. & Naqi, S. A. 1996. Analysis of the serotype-specific epitopes of avian infectious bronchitis virus strains Ark99 and Mass41. *J. Virol*. 70, 7255-9.
45. Kapczynski, D. R. and King, D. J. 2005. Protection of chickens against overt clinical disease and determination of viral shedding following vaccination with commercially available Newcastle disease virus vaccines upon challenge with highly virulent virus from the California 2002 exotic Newcastle disease outbreak. *Vaccine*, 23, 3424-3433.
46. Keawcharoen, J., Van Riel, D., Van Amerongen, G., Bestebroer, T., Beyer, W. E., Van Lavieren, R., Osterhaus, A. D. M. E., Fouchier, R. A. M. and Kuiken, T. 2008. Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5N1). *Emerging Infectious Disease*, 4, 600-605.
47. Kleven, S. H. 1998. Mycoplasmas in the etiology of multifactorial respiratory disease. *Poultry Science*, 77, 1146-1149.
48. Klumperman, J., Locker, J. K., Meijer, A., Horzinek, M. C., Geuze, H. J. & Rottier, P. J. 1994. Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J. Virol*. 68, 6523-34.
49. Koboldt, D. C., Zhang, Q., Larson, D., Shen, E.D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. & Wilson, R. K. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 22, 568–576.
50. Koch, G., Hartog, L., Kant, A. & Van Roozelaar, D. J. 1990. Antigenic domains on the peplomer protein of avian infectious bronchitis virus: correlation with biological functions. *J. Gen. Virol*. 71 (9), 1929-35.
51. Kong, Q., Xue, C., Ren, X., Zhang, C., Li, L., Shu, D., Bi, Y. & Cao, Y. 2010. Proteomic analysis of purified coronavirus infectious bronchitis virus particles. *Proteome Sci*. 8, 29.
52. Koopmans, M., Wilbrink, B., Conyn, M., Natrop, G., Van der Nat, H. Vennema, H., Mejier, A., Van Steenbergen, J., Fouchier, R., Osterhaus, A. and Bossman, A. 2004. Transmission of H7N7 avian influenza A type virus to human beings during a large outbreak of commercial poultry farms in the Netherlands. *The Lancet*, 363, 587-593.

53. Ladmana, B. S., Popea, C. R., Zieglerb, A., Swieczkowskic, T., Callahana, J. M., Davisond, S. and Gelb J. 2002. Protection of Chickens After Live and Inactivated Virus Vaccination Against Challenge with Nephropathogenic Infectious Bronchitis Virus PA/Wolgemuth/98. Avian Diseases 46, 938-944.
54. Lee, S., Markham, P. F., Coppo1, M. J. C. Legione1, A. R., Markham, J. F., Noormohammadi, A. H., Browning, G. F., Ficorilli, N., Hartley, C. A., Devlin, J. M. 2012. Attenuated vaccine can recombine to form virulent filed viruses. *Science*, 337, 188.
55. Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.
56. Li-Cor, A. 2014. Chromas Lite (Version 2.1.1) [Computer program]. Available at http://technelysium.com.au/wp/chromas/ (Accessed 19 December 2014).
57. Lim, T. H., Kim, M. S., Jang, J. H., Lee, D. H., Park, J. K., Young, H. N., Lee, J. B., Park, S. Y., Choi, I. S. & Song, C. S. 2012. Live attenuated nephropathogenic infectious bronchitis virus vaccine provides broad cross protection against new variant strains. *Poult. Sci*. 91, 89-94.
58. Lim, T. H., Kim, M.S., Jang, J.H., Lee, D. H., Park, J. K., Youn, H. N., Lee, J. B., Park, S. Y., Choi, I. S. and Song, C. S. 2011. Live attenuated nephropathogenic infectious bronchitis virus vaccine provides broad cross protection against new variant strains. *The Journal of Applied Poultry Research*, 91 (1), 89-94.
59. Liu, Q., Mena, I., Ma, J., Bawa, B., Krammer, F., Lyoo, Y. S., Lang, Y., Morozov, I., Mahardika, G. N., Ma, W., García-Sastre, A. and Richta, J. A. 2015. Newcastle disease virus-vectored H7 and H5 live vaccines protect chickens from challenge with H7N9 or H5N1 avian influenza viruses. *Journal of Virology*, 89, 7401-7408.
60. Luo, Z. & Weiss, S. R. 1998. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. *Virology*, 244, 483-94.
61. Macdonald, J. W. & Mcmartin, D. A. 1976. Observations on the effects of the H52 and H120 vaccine strains of infectious bronchitis virus in the domestic fowl. *Avian Pathol*, 5, 157-73.

62. Mast, J., Nanbru, C., Van Den Berg, T. and Meulemans, G. 2005. Ultrastructural changes of the tracheal epithelium after vaccination of day-old chickens with the La Sota strain of Newcastle disease virus. *Vet. Pathol*. 42, 559-565.
63. Mcbride, C. E. & Machamer, C. E. 2010. A single tyrosine in the severe acute respiratory syndrome coronavirus membrane protein cytoplasmic tail is important for efficient interaction with spike protein. *J. Virol*. 84, 1891-901.
64. McKinley, E. T., Hilt, D. A. and Jackwood, M. W. 2008. Avian coronavirus infectious bronchitis attenuated live vaccines undergo selection of subpopulations and mutations following vaccination. Vaccine 26, 1274–1284.
65. Microsoft Corporation. 2010. Excel (Version 2010) [Computer program]. Available at https://products.office.com/en-us/office-2010 (Accessed 23 January 2015).
66. Moss, B. 1991.Vaccina virus: a tool for research and vaccine development. *Science*, 252, 1662- 1667.
67. Mounts, A. W., Kwong, H., Izurieta, H. S., Ho, Y., Au, T., Lee, M., Bridges, C. B., Williams, S.W., Mak, K. H., Katz, J. M., Thompson, W. W., Cox, N. J. and Fukuda, K. 1997. Case-control study of risk factors for avian influenza A (H5N1) disease, Hong Kong, 1999. *The Journal of Infection Disease.* 180, 505-8.
68. Nakamura, K., Imai, K. and Tanimura, N. 1996. Comparison of the effects of Infectious Bronchitis and Infectious Laryngotracheitis on the chicken respiratory tract. *J. Comp. Path.* 114, 11-21.
69. Page, L. A. 1961. Respiratory diseases in chickens. *California Agriculture*, 3, 7-8
70. Perozo, F., Marcano, R. and Afonso, C. L, 2012. Biological and phylogenetic characterization of a genotype VII Newcastle disease virus from Venezuela: efficacy of field vaccination. *Journal of Clinical Microbiology,* 1204-1208.
71. Pohjola, L. K., Ek-Kommonen, S. C., Tammiranta, N. E., Kaukonen, E. S., Rossow, L. M. & Huovilainen, T. A. 2014. Emergence of avian infectious bronchitis in a non-vaccinating country. *Avian Pathol*. 43, 244-8.
72. Pohuang, T., Chansiriporncha, N., Tawatsin, A. & Sasipreeyajan, J. 2009. Detection and molecular characterization of infectious bronchitis virus isolated from recent outbreaks in broiler flocks in Thailand. *Journal of Veterinary Science*, 10, 219-223.

73. Quinla, A. R., & Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.
74. Rafiei, M. M., Vasfi-Marandi, M., Bozorgmehri-Fard, M. H & Ghadi, S. 2009. Identifiaction of different serotypes of Infectious Bronchitis Viruses in allantoic fluid samples with single and multiplex RT-PCR. *Iranian Journal of Virology*, 3, 24-29.
75. Reddy, V. R. A. P., Steukers, L., Li, Y., Fuchs, W., Vanderplasschen, A. and Nauwynck, H. J. 2014. Replication characteristics of infectious laryngotracheitis virus in the respiratory and conjunctival mucosa. *Avian pathology*, 5, 450-457.
76. Rimondi, A., Craig, M. I., Vagnozzi, A., Konig, G., Delamer, M. & Pereda, A. 2009. Molecular characterization of avian infectious bronchitis virus strains from outbreaks in Argentina (2001-2008). *Avian Pathol*. 38, 149-53.
77. Roussan, D. A., Haddad, R. and Khawaldeh, G. 2008. Molecular survey of avian respiratory pathogens in commercial broiler chicken flocks with respiratory diseases in Jordan. *Poultry Science*, 87, 444-448.
78. Sambhara, S. and Poland, G. A., 2010. H5N1 avian influenza: preventive and therapeutic strategies against a pandemic. *Annu. Rev. Med.* 61, 187-98.
79. Sanger, F., Nicklen, S. and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 12, 5463–5467.
80. Schalk, A. F. H. & Hawn, M. C. 1931. An apparently new respiratory disease of baby chicks. *Journal of the American Veterinary Association*, 78, 413-422.
81. Schelle, B., Karl, N., Ludewig, B., Siddell, S. G. & Thiel, V 2005. Selective replication of Coronavirus genomes that express Nucleocapsid protein. *J. Viro.* 79, 6620-30.
82. Selvarani, R., Ramakrishnan, M. & Panda, S.K. 2014. Molecular characterization of Infectious Bronchitis Virus from eastern India. *Int. J. Agric. Sc. & Vet. Med.* 2, 102-106.
83. Shittu, I., Sharma, P., Volkening, J. D., Solomon, P., Sulaiman, L. K., Joannis, T. M., Smith, E. C. & Denison, M. R. 2012. Implications of altered replication fidelity on the evolution and pathogenesis of coronaviruses. *Current Opinion in Virology*, 2, 519-524.
84. Studer, R. A., Dessailly, B. H. and Orengo, C. A. 2013. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochemical Journal*, 449 (3), 581-594.

85. Suarez, D. L., Perdue, M. L., Cox, N., Rowe, T., Bender, C., Huang, J. and Swayne, D. E. 1998. Comparison of highly virulent H5N1 influenza A virus isolated from humans and chickens from Hong Kong. *Journal of Virology*, 8, 6678-6688.
86. Sun, L., Li, Y., Zhang, Y., Han, Z., Xu, Y., Kong, X. and Liu, S. 2014. Adaptation and attenuation of Duck Tembusu Virus strain Du/CH/LSD/110128 following serial passage in chicken embryos. *Clin. Vaccine Immunol.* 21 (8), 1046-53.
87. Swayne, D. E., 2006. Principles of vaccine protection in chickens and domestics waterfowl against avian influenza. *Ann. N. Y. Acad. Sci.* 1081, 174-181.
88. Swayne, D.E., 2009. Avian influenza vaccine and therapies for poultry. *Comparative Immunology, Microbiology and Infectious Disease.* 32, 351-363.
89. Tay, F. P., Huang, M., Wang, L., Yamada, Y. & Liu, D. X. 2012. Characterization of cellular furin content as a potential factor determining the susceptibility of cultured human and animal cells to coronavirus infectious bronchitis virus infection. *Virology*, 433, 421-30.
90. Terregino, C., Toffan, A., Beato, M.S., de Nardi, R., Vascellari, M.Meini, A. 2008. Pathogenicity of a QX strain of infectious bronchitis virus in specific pathogen free and commercial boiler chickens, and evaluation of protection induced by a vaccination programme based on the Ma5 and 4/91 serotypes. Avian Pathology, 37, 487–493.
91. Tooze, J., Tooze, S. A. & Fulller, S. D. 1987. Sorting of progeny coronavirus from condensed secretory proteins at the exit from the trans-Golgi network of AtT20 cells. *J. Cell Biol*. 105, 1215-26.
92. Unknown photographer, 2016. Avian Infectious Bronchitis Virus [ONLINE]. Available at: http://ytpo.net/viruses/virus.php?id=36&name=avian%20infectious%20bronchitis%20virus&search=virus [Accessed 15 May 2016].
93. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., Rozen, S. G. 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Research* 40 (15), 115
94. Villegas, P. 1998. Viral diseases of the respiratory system. *Poultry Science*, 77, 1143-1145.

95. Wajid, A., Wasim, M., Rehmani, S. F., Bibi, T., Ahmed, N., Afonsod, C. L. 2015. Complete genome sequence of a recent panzootic virulent Newcastle disease virus from Pakistan. *Genome Announcements*, 3, 1-2.

96. Wicht, O., Burkard, C., De Haan, C. A., Van Kuppevelt, F. J., Rottier, P. J. & Bosch, B. J. 2014. Identification and characterization of a proteolytically primed form of the murine coronavirus spike proteins after fusion with the target cell. *J. Virol*. 88, 4943-52.

97. Yang, X., Zhou, Y., Fu L., Ji, G., Zeng, F., Zhou, L., Gao, W. and Wang, H. 2016. Recombinant infectious bronchitis virus (IBV) H120 vaccine strain expressing the hemagglutinin-neuraminidase (HN) protein of Newcastle disease virus (NDV) protects chickens against IBV and NDV challenge. *Archives of Virology*, 161 (5), 1209–1216.

98. Zaffuto, K. M., Estevez, C. N. and Afonso, C. L. 2008. Primary chicken tracheal cell culture system for the study of infection with avian respiratory viruses. *Avian Pathology*, 37, 25-31.

99. Zanaty, A. E. A. 2013. Evaluation of the Protection conferred by heterologous attenuated live infectious bronchitis viruses against an Egyptian variant IBV [EG/1212B]. *Journal of American Science*, 9, 599-506.

100. Zhou, J., Wu, J., Cheng, L., Zheng, X., Gong, H., Shang, S. & Zhou, E. 2003. Expression of immunogenic S1 Glycoprotein of Infectious Bronchitis Virus in transgenic potatoes. *J. Virol.* 77, 9090-9093.

101. Zhu, J. G., Qian, H. D., Zhang, Y. L., Hua, X. G., Wu, Z. L. 2007. Analysis of similarity of the S1 gene in infectious bronchitis virus (IBV) isolates in Shanghai, China. *Arch. Med. Vet.* 39, 223-228.