GENOTYPING ASSAY BY MULTIPLEX PCR AND HIGH THROUGHPUT MINION SEQUENCING OF INFECTIOUS LARYNGOTRACHEITIS VIRUS (ILTV) UNIQUE SHORT GENOME REGION

by

YI-CHEN LUO

(Under the Direction of Maricarmen Garcia)

ABSTRACT

This study aims to improve the ILTV genotyping assays by amplifying an expanded region of the viral genome through multiplex PCR and adopting the MinION sequencing technology to increase the depth of coverage on informative SNP sites to provide accurate discrimination between ILTV strains. The Unique short (Us) region of the ILTV genome was targeted for multiplex PCR development. Sixteen primer pairs were designed, each capable of amplifying approximately a 1kb fragment within the Us ILTV genome region. The sequence analysis of the U_S sequences of known ILTV isolates were assigned to corresponding genotypes as characterized by full genome analysis and multi-allelic Sanger sequencing assay. Applying the multiplex PCR MinION sequencing assay to clinical samples, demonstrated successful genotyping, and increased the depth of coverage of the sequences allowed to identify four SNPs within the Us sequences not previously identified among currently circulating genotype VI virus.

INDEX WORDS: MinION, nanopore, high throughput sequencing, Infectious Laryngotracheitis, Genotyping assay, Multiplex PCR

GENOTYPING ASSAY BY MULTIPLEX PCR AND HIGH THROUGHPUT MINION SEQUENCING OF INFECTIOUS LARYNGOTRACHEITIS VIRUS (ILTV) UNIQUE SHORT GENOME REGION [1]

by

YI-CHEN LUO

DVM., National Chung-Hsing University, Taiwan, 2017

MS., University of Georgia, 2021

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2023

GENOTYPING ASSAY BY MULTIPLEX PCR AND HIGH THROUGHPUT MINION SEQUENCING OF INFECTIOUS LARYNGOTRACHEITIS VIRUS (ILTV) UNIQUE SHORT (U_S) GENOME REGION

by

YI-CHEN LUO

Major Professor:	Maricarmen Garcia, MS,PhD
Committee:	Holly Sellers, MS,PhD
	James Stanton, DVM,PhD,ACVP

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2023

DEDICATION

For my family and friends, both near and far, who have provided their unwavering support and love:

For my family, who teaches me strength and resilience no matter the situation.

For my friends, who have always along my side when things are difficult and provide wisdom advises to guide me through

ACKNOWLEDGEMENTS

Thank you to my major professor, Dr. Maricarmen Garcia, for your continuous support and guidance and for your profound contribution to my professional and personal growth. The wisdom you have shown to me has beyond anything I have experienced so far, which shall benefit me for my whole life. I further would like to express my gratitude to my committee, Dr. James Stanton, Dr. Holly Sellers, for their advice and support. Finally, I am indebted to the faculty, staff, and students at the Poultry Diagnostic and Research Center who have aided me in every aspect of completing this work; in particular, I would like to acknowledge Ms.April Skipper, Mr. Patrick Whang, Ms.Kelsey Young, Mr. Zhihan Xian and Ms. Sylva Riblet.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

LITERATURE REVIEW

Infectious Laryngotracheitis Virus (ILTV)

Overview and history

The term laryngotracheitis was first used in the early 1920s in the United States to describe the disease in chicken characterized by acute or subacute dyspnea (1). Many names were used to describe the clinical findings of the disease as it first emerged, such as infectious bronchitis, tracheolaryngitis, infectious tracheitis, and avian diphtheria. In 1931, a special committee on poultry disease of American Veterinary Medical Association named the disease Infectious Laryngotracheitis (ILT). Following the discovery of the disease, numerous countries including Canada, Australia, Britain, Sweden, Holland, Poland, Germany, and Finland recognized cases of ILT, which demonstrated that the disease affected the poultry industry worldwide since its birth (2).

Epidemiology

The infectious laryngotracheitis virus (ILTV) is the causative agent of ILT. The virus has a very restricted range of natural hosts, with chickens being its primary target. Although this virus can infect chickens of any age, older flocks are more susceptible and likely to display more severe clinical signs upon initial exposure (2). While there were few records of pheasants, peafowl and

turkey showing susceptibility to the disease, most of other avian species including duck, guinea fowl and pigeon embryos were refractory (3, 4).

The high transmission capacity and ability of ILTV to remain latent are the primary factors that challenge potential programs of disease eradication. The virus can be spread through several means, including contaminated dust, litter, aerosols, biofilms in drinking water line and darkling beetles, and can survive for a long time. For instance, vaccine viruses could be isolated from the biofilms in the drinking water line 3 weeks after the drinking water application and virulent viruses could be isolate from darkling beetles in a complex 6 weeks after the outbreak. These fomites act as vectors, causing re-infection of newly placed flocks in a chicken house (5, 6). Viruses enter via ocular, intranasal, oral and intratracheal routes (7). The virus incubation period can range from 4–- 10 days before identifying any signs of the disease. However, the virus can readily establish a latent infection characterized by the migration of viral genomes migrating to the trigeminal ganglion (8, 9). These latently infected birds serve as potential carriers and can sporadically shed virus when the flock faces stressful conditions such as rehousing or the onset of lay (10).

The acute form of the disease is characterized by excessive hemorrhagic and fibrinous exudate in trachea while the mild subacute form is less fatal and featured by small amount of adherent fibrin present in the trachea (1). During severe forms of the disease, ILTV infection can induce high morbidity (>90%) and variable mortality between 10%-60% (11). In mild forms of the disease, morbidity and mortality are lower (12). Because the disease is frequently observed in densely populated poultry farming regions, it can lead to substantial production losses. These losses include increased mortality rates, reduced egg production, delayed weight gain, and increased susceptibility to other respiratory pathogens (13).

Disease Control

During outbreaks of the disease, preventing the spread of ILTV between successive flocks involves thorough cleaning and an extended downtime. Just like other enveloped virus, ILTV can be readily inactivated outside the host when exposed to high temperature or treatment with disinfectants. All surfaces of house and equipment should be deep cleaned and potential contaminants such as feathers, carcasses and litters should be heated in 38℃ for 100 hours prior to be disposed (14). As for the prevention of ILTV spread between nearby sites, requires industrywide cooperation. Infected flocks should be submitted for processing as soon as possible, and during the trip to the processing plant the route of the haul truck should minimize the number of farms it passes by to decrease the possibility of spreading the virus to susceptible chickens (15).

Two types of live attenuated ILTV vaccines are used to effectively prevent the disease: the chicken embryo origin (CEO) vaccine and the tissue culture origin (TCO) vaccine. The CEO vaccine features a faster infection rate when administered via drinking water application achieving high vaccination coverage of the flock. The CEO vaccines can induce respiratory reactions that affects the flock performance parameters. In contrast, although the TCO vaccine is known for having milder side effects and a slower transmission rate than the CEO vaccine (16), the eye drop single route of application is time consuming and labor force costly (17). Both the CEO and TCO vaccines share the potential to regain virulence after bird-to-bird passage (18). Thereby, the use of live attenuated vaccines for controlling ILTV is only suggested in epidemic regions (14).

The new generation of ILT vaccines aim to provide sufficient protection against the disease while minimizing the risk of virulence reversion and protecting against two economically important disease in poultry at once. One approach involves the use of viral vector recombinant ILTV vaccines carrying ILTV antigens. Recombinant vaccines using Fowlpox virus vectors that

carry either the viral glycoprotein B (gB) alone or in combination with the surface UL32 of ILTV have been tested. Although the rFPV-ILTV gB vaccine was not as effective as attenuated vaccines in preventing virus replication, it did decrease the severity of clinical signs and mortality (19). Similar results were seen using the Herpes virus of turkey (HVT) vector vaccine expressing the ILTV- gD and gI showed suboptimal effect in inhibiting viral replication but promising result to ameliorate clinical signs of the disease. Other vaccination strategies commonly utilized to expand protection is to vaccinate with recombinant vaccines subcutaneously at day of age or in ovo followed by a booster with live attenuated vaccines (17). Other recombinant vaccines have been developed using Newcastle disease virus (NDV) (20) and MEQ-deleted Marek's disease virus (BACΔMEQ) as viral vectors to carry ILT antigens (21).

Taxonomy

ILTV is a member of the *Herpesviridae* family classified under the *Alphaherpesvirinae* subfamily. The international Committee on Taxonomy of Viruses assigned ILTV to genus *Iltovirus* and gave a scientific species name of *Iltovirus gallidalpha1.* Beside ILTV, three other virus species — *Iltovirus cacatuidalpha2, Iltovirus psittacidalpha1 and Iltovirus psittacidalpha1*, were also classified under *Iltovirus* (22). Among, *Psitaccid herpesvirus 1* (PsHV-1), the former of *Iltovirus psittacidalpha1*, was an old member which shown to have strong similarities in structural characteristics and studied for the comparative (23). Now with two new members included in genus *Iltovirus*, studies comparing between these viruses could aid in understanding the evolution of this unique clade of herpesvirus.

Genome Structure

The architecture of ILTV genome falls within the type three arrangement. The genome is arranged as two isoforms, inverted repeat (I_R) and terminal repeat (T_R), flanking the Unique short

(U_S) region in an opposite orientation, and a Unique long region (U_L) being separated from the Unique short (U_S) region by the inverted repeat (24, 25). Whereas the length of the U_L and repeat regions tend to be variable among strains, the length of the U_S region remains to be conserved among strains with approximately13kbp of size (26). The entire dsDNA molecule is approximately 150kbp (148,687bp – 159,744 bp) long with a GC content of 48% and predicted to contain at least 80 open reading frames (27).

Compared to other alphaherpesvirus genomes, three unique features were found in the *Iltovirus* genus genomes. First, the genomes contained a 48kbp internal inversion from the UL22 to UL44 genes within the U_L region. Secondly, the UL47 tegument protein encoding gene translocated from the U_L to the U_S region. Thirdly, viruses in this genus contained five unique ORFs— ORFA, ORFB, ORFC, ORFD and ORFE (23, 28, 29). Furthermore, ILTV possessed distinct features differentiating from Psittacid herpesvirus 1 (PsHV-1), another member of the *Iltovirus* genus. Incliningly, the absence of a widely conserved capsid assembly assisting protein UL16 (23, 30) and the possession of ILTV specific genes – UL0 (31). The UL0 gene is situated adjacent to the UL[-1] gene, and they exhibit several shared characteristics, including expression kinetics, subcellular localization, and, most significantly, homology in their amino acid sequences. These similarities strongly suggest that they are paralogous genes that originated from duplication events (31). Interestingly, UL0 and UL[-1] seemed to participate differently in viral replication. A UL0 deleted ILTV mutant displayed only a minor defect on replication, but the decreased pathogenicity and the ability to elicit immunity make the mutant a potential candidate for a live modified vaccine (32). In contrast, UL[-1] is indispensable. An UL[-1] deletion mutant was deprived of the ability to replicate independently (33). During the search for the function of the five unique open reading frame cluster, only one study suggested that ORFC could be a virulence

related gene that makes ORFC a potential target to study attenuated live vaccine (34). The function of the remaining four unique open reading frames of ILTV remain unknown.

Molecular Diagnostics for Detection of ILTV genomes

Several molecular-based techniques, such as polymerase chain reaction (PCR), nested PCR, real-time PCR, and *in-situ* hybridization, have been developed for diagnosing ILTV. Each of these methods targets different parts of the genome and offers unique benefits (35). A PCR reaction requires target templates, both forward and reverse primers and polymerase enzyme to perform amplification. With a successful amplification and sufficient output of product, a band with a target size should reveal under an electrophoresis examination (36). Similar but in contrast to PCR, a nested PCR contains an additional amplification process which is targeting on the PCR product from the first PCR. Hence, nested PCR provides higher detection sensitivity (37-39). In contrast, a real-time PCR does not require an electrophoresis step after amplification to validate the presence of PCR amplicons. Instead, real-time PCR detects the accumulation of amplicons passing the threshold anytime during the amplification cycles (36, 38, 40). Hence, real-time PCR not only exhibits high sensitivity, but it is the simplest procedure among amplification techniques. Also real-time PCR can provide a relative quantification of nucleic acid load in the sample. These features make real-time PCR a useful tool in facilitating the studies of virus tissue tropism and growth kinetics as well as a tool for diagnostics (38, 40-43). Finally, *in-situ* hybridization techniques utilize probes to interact with target sequence and reveals the location of ILTV nucleic acid on histo-slides, which aid not only in diagnosis but also enable studies related to tropism or pathogenesis as well (44).

Compared to classic virological diagnostic approaches like virus isolation, histopathological examination, and serological methods, molecular-based techniques offer

advantages such as high sensitivity, accuracy, speed, reproducibility. One disadvantage of molecular detection is that they are unable to distinguish between viable and non-viable virions, where careful interpretation of the results in combination with the clinical display and history is required to reach a diagnosis (35, 37, 40). Moreover, i''s important to consider the specificity of the detection assay to other genetically or clinically similar pathogens to ensure accurate diagnosis (38, 45, 46).

ILTV Genotyping

Since there is only one serotype for ILTV (47), the methods to differentiate among ILTV strains relies on genome nucleotide differences (48). Over the years, several genotyping techniques such as PCR-restriction fragment length polymorphism (RFLP), PCR-Sanger sequencing, high-resolution melting (HRM), TaqMan SNPs genotyping assay and whole-genome sequence have been employed to genotype ILTV. Each of these assays have a unique set of advantages and disadvantages. PCR-RFLP technology is a widely used method that involves PCR amplification of single or multiple virus alleles, followed by digestion of the amplification products with restriction enzymes that recognize differences in restriction sites. The distinct RFLP patterns, characterized by variable band sizes on the electrophoresis gel, enables differentiation between genotypes (49-54). While RFLP can effectively identify the presence of co-infections (43), it is time-consuming, requires a substantial amount of intact viral genomic DNA (55). Additionally, the sensitivity of PCR-RFLP is limited to detect nucleotide changes within the restriction enzyme recognition sites. Therefore, any nucleotide changes outside of the recognition pattern sites may be missed (49, 52). Another method utilized to identify nucleotide differences without doing direct sequencing is the high-resolution melting (HRM) analysis that detects nucleotide differences by measuring variations in amplicon melting temperaturesI). It is cost effective, fast and contains the

potential to differentiate recombinants. The HRM assay is sensitive, economical, and fast. However, the sensitivity of the assay comes at the cost of amplifying a short segment to precisely detect differences in Tm. Furthermore, like the RFLP assay, the differences in HRM does not provide information about the actual nucleotide changes (56, 57). To get the exact sequence difference sequencing of the amplicon is necessary (51). Another assay that detects nucleotide differences without using sequencing procedures are TaqMan SNP genotyping assays, which use probes that anneal to known single nucleotide polymorphism (SNP) sites to achieve typing. However, it should be noted that changes on other sites of the designed probe will not be detected and novel nucleotide changes on the SNP site will turn as a false negative result. Therefore, interpretation should be careful, when having negative results (58).

Most genotyping systems were developed to monitor the spreading and evolution of ILTV circulating within a specific country, differentiation between live-modified vaccines (TCO or CEO), differences among virulent strains closely associated to vaccines, among field strains, and among natural emerging recombinant strains (48). The Australian ILTV genotyping system comprises 10 classes (classes 1 to 10). The Australian origin attenuated vaccine strains A20 and SA2 belong to class 1, while the field isolates circulating in commercial flocks V1-99 is grouped to class 2, CSW-1 to class 4 and V1-03 to class 5 respectively (52). The class 3 strains were initially isolated from Victoria state and south Australia commercial flocks (52) which later migrated to backyard flocks and serve as reservoirs that infect commercial poultry (49). The distinctive Australian field isolates – Class 6 were associated with outbreaks in commercial poultry in the State of Victoria in 2007 (49). Class 7 groups the imported European vaccine strain Serva, whereas class 8 and 9 are strains that are closely related to class 7 (49). The origin of class 8 and 9 was uncertain, but as these emerged after the importation and use of the Serva vaccine, it was believed

that these were revertant strains from the Serva vaccine or subpopulations pre-existing in the vaccine that were selected by best fitness to the host (49). In support of the fitness theory, laboratory studies have shown that class 9 strains exhibit greater pathogenicity and transmission compared to the circulating class 2 strains (59) which may explain the rapid establishment of class 9 strains in Australia (60). The class 10 strains belong to a group of natural recombinant viruses that exhibit significant similarity to the class 7 (Serva) and class 1 (SA2) vaccine strains, as well as class 2 field strain viruses. Between 2009 and 2014, class 10 strains were prevalent in the state of New South Wales; however, after 2014 detection of class 10 strains in NSW significantly decreased (60). Targeting the gG, TK, ICP18.5, ICP4, and ORFB-TK genes through RFLP enables the genotyping of Australian ILTV classes 1 to 9 (49, 52). By employing the HRM assay that focuses on the ORFB, UL27, UL36, and US7 genes, it is possible to discern between classes 2, 4, 8, 9, and 10 (56). The Taqman sequencing assay based on SNPs in the UL46, UL36, UL8, UL0, ICP4 and US3 serves to detect recombinants between the class 2 and 4 strains (58). Lastly, benefiting from whole genome sequencing using next generation sequencing technologies, a new emerging subtype – class 7b, a natural recombinant strain with the Serva vaccine and class 8 as the major genome contributing parents and classes 9 and 1 as the minor contributors. It is worth mentioning that identification of the 7b class could be missed if the genotyping assay target genome regions outside of the recombinant breakpoints (61).

The North American strains have been classified in nine genotype groups by targeting on ORFB-TK, gM, ICP4 and UL47 genes by PCR-RFLP (54, 62, 63) or by full genome sequencing (64). The USDA standard challenge strain is assigned to genotype I; the tissue culture origin (TCO) attenuated vaccine is assigned to genotype II; TCO related virulent strains are included within genotype III; chicken embryo origin (CEO) attenuated vaccines are classified as genotype IV;

CEO related virulent strains are identified as genotype V; and virulent strains not related to vaccine strains in commercial poultry are identified as group VI and depository strains from back-yard flocks are identified as genotypes VII to IX (62, 64). Also, in endemic countries including Korea , China , Israel , Taiwan , Myanmar, Argentina and Brazil (TK, gE,gG and ICP4) variable targets, including the TK, ICP4, ORFB, UL47, ICP18.5, gG, gE, gB, gJ and gD genes and the homopolymer stretches at the terminal repeats, have been used to differentiate the local field strains from the commercial vaccines (51, 63, 65-70). PCR-RFLP assays, while effective in differentiating among circulating viral strains within a country or region, full genome sequencing will allow to analyze the evolution of ILTV worldwide (48).

ILTV whole genome database

The very first assembled ILTV genome was obtained as a composite of 14 independently published Sanger sequences, resulting in a genome length of 148,665 base pairs (23). Later, with the advent of next generation sequencing (NGS) technology, a whole virus genome consensus was built by mapping de novo assembled contigs to the previously established composite genome. As a result, the whole genome sequence of the ILTV Serva strain (European CEO vaccine), spanning length 137,693 bp, excluding the terminal repeat region, was released. Upon comparison with the previous concatenated sequence, the Serva whole genome sequence exhibited the presence of four substantial insertions in the UL29, UL36, and the inverted and terminal repeat regions respectively. These insertions were confirmed to be authentic genomic sequences through the utilization of Sanger sequencing (27). As the Serva strain sequence (Accession no: HQ_630064) was the first whole genome sequence assembled de novo, subsequent viral whole genomes have been obtained by utilizing the Serva strain genome as reference sequence to assemble contigs and to facilitate consensus building (57, 71, 72).

Currently, a total of 85 whole genome sequences of viruses from the United States (18 isolates) (26, 73-76), Australia (23 isolates) (27, 57, 60, 61, 71, 76-78), China (5 isolates) (46, 79), Italy (5 isolates) (72), Russia (3 isolates) (80), Germany (1 isolate), Korea (4 isolates) (81), Peru (1 isolate)(82) and Canada (25 isolates) (64, 73, 83, 84). Most of these sequences were generated using second generation sequencing technology, where the genome fragmentation process during library preparation introduces several challenges for genome assembly. Genome fragmentation can favorably brake sequences at certain regions, such as the GC rich regions (85) increasing the difficulty to reassemble them. Consequently, despite massive number of reads and sufficient depth of coverage, unbridgeable gaps may persist in the more fragmented region (27). Also, short reads may not be able to resolve structural variants. An example is ILTV strains with variant I_R/T_R sequences where precise mapping of short reads into highly similar assembled contigs within

inverted and terminal repeats is inaccurate. Hence, sequence variants in the I_R/T_R regions of the ILTV genome need to be confirmed by Sanger sequencing (72). During the process of adapting the virus to grow in tissue culture or embryos, may introduce or select for variants that improve the fitness of the virus to grow in vitro but are not characteristic of the original sequence from the clinical sample (86). In an ideal scenario, it is to be able to do direct sequencing from clinical samples to introduce minimal bias. However, obtaining comprehensive coverage of the entire ILTV genome directly from clinical samples can be challenging due to the substantial presence of host DNA in the sample composition (87).

The feasibility of obtaining whole virus genome sequences facilitated to resolve the ILTV genome structure. When whole genomes of the US strains, the USDA, 81658, 63140 and 1874C5, were compared to the Serva strain genome, a 1,015 bp sequence of the Serva vaccine genome was aligned as part of U_L region rather than the terminal repeat genome region. After verification by

Sanger sequencing, Spatz et al., (2012) proposed that the Serva strain sequence should contain 112,915 bp in the U_L region and 13,818 bp in the repeated regions to resolve the assembly error of the truncated repeat region (26). Other sequence differences identified among ILTV genomes included a three 855bp long repeat copies deletions detected upstream of the ICP4 gene promoters in the LT-Ivax® p20 (passages 20) isolates as compared to LT-Ivax® p1 (passage 1) (74). A three CCT repeat deletions detected in the low complexity regions of the UL52 genes in the Serva vaccine strains as compared to Italian wild-type strains (72). Lastly, the major driving force of DNA virus evolution – recombination (86). Through comparison of whole genome genomes and using multiple recombination detection tools allowed to infer potential recombination events and the break points in the viral genome. For instance, the Australian class 10 strain was determined to have emerged from three recombination events in the U_L region and one event spanning the I_R and U_S region (60). Sabir et al detected two recombination genomic regions in the U_L (38kbp-undetermined ends, 61kbp-107kbp) of class 7b strain (61).

Nucleic Acid Sequencing

First generation sequencing

In the 1970s, two methods established the foundation for the development of what is recognize today as Sanger sequencing. The first “plus and minus” method developed by Sanger and Coulson contained two phases of polymerase synthesis reactions. In the first phase, assuming that the template amplification process was not synchronized, random lengths of oligonucleotide products from the initiation site of primers was generated. The resulting mixture was subsequently subjected to two distinct treatments, referred t“ as ”plus“ or "m”nus," during the second

phase.“The ”plus" method incorporated only one of the four deoxyribonucleoside triphosphate (A, T, C, or G) at a time, yielding products with single terminating residue. The “minus” method, on the contrary, excluded only one of the four deoxyribo-triphosphates at a time, extending all the terminates except for the one that was not included. By consolidating the information on the polyacrylamide gel from the eight types of product mixture generated by the “plus and minus” methods, the DNA sequence could be inferred (88). The second method, pioneered by Maxam and Gilbert, employed a different approach by focusing on removal rather than addition (89). The core concepts utilize chemicals with the ability to selectively remove nucleotide bases: hydrazine eliminates pyrimidines but under high salt concentration it exclusively targets cytosine; acids remove all purines while dimethyl sulfate specifically cleaves guanine. Similarly, the DNA sequence could be inferred by comparing the length of cleavage fragments on polyacrylamide gel. This method was more widely adopted than the former, which was then further considered as the birth of first- generation sequencing (90).

After the ground-breaking Sanger sequencing technology, today Sanger sequencing utilizes the “chain-termination” chemistry incorporating fluorescent label deoxyribonucleoside triphosphates and the 2’,3’- dideoxy nucleotides. The inclusion’of’2',3'-dideoxy nucleotide analogues in the DNA leads to the termination of the polymerase synthesis process, producing DNA fragments of partial lengths. By integrating the principles of“the "plus and ”inus" method mentioned earlier, one of ’he’2',3'-dideoxy nucleotides (A, T, C, or G) was incorporated alongside the remaining three types of deoxyribonucleoside triphosphates during the polymerase reaction. This resulted in four distinct combination patterns, allowing for the generation of amplicons terminated with four different residues at various positions. Together, these fractions of amplicons facilitated the inference of the DNA sequence (91). Because the Sanger sequencing accuracy,

robustness, and ease of use, it remained for two decades as the predominant sequencing technology. The integration of further emerging techniques had brought the technology a step forward. For instance, the “shot-gun sequencing” which involved cloning and sequencing overlapping sequences extended the previous limit in sequencing length (less than 1 kbp) and enabled the assembly of longer continuous DNA fragments (90, 92). The introduction of the polymerase chain reaction technique (PCR) greatly facilitated sequencing by providing high concentrations of DNA templates necessary for sequencing (93).

Second generation sequencing

Simultaneously to the advancement of large-scale dideoxy sequencing endeavors, another technique emerged, paving the way for the initial wave of the next generation DNA sequencers with increased throughput. Just as Sanger sequencing, the new technology “s a "sequence-by-synt”esis" (SBS) technique, relying on the incorporation of nucleotides into the template during polymerization reactions to generate observable output. Since the successful incorporation of nucleotides is detected by the concurrent release of pyrophosphate, the technique is also known as pyrosequencing (94, 95). Pyrosequencing utilizes chemical conversion with two enzymes to detect strand extensions: ATP sulfurylase converts the pyrophosphates into ATP, while the production of ATP is used as a substrate of luciferase to generate visible signals (96, 97). By repeatedly adding one of the four deoxynucleotides and enzymatically removing the unincorporated dNTP in each sequencing cycle, the DNA sequence can be inferred by detecting signals from the correct complementary bases incorporated (94, 95). The beneficial features of this technique included sequencing in real-time, without the need for size selection via electrophoresis and the possibility to sequence multiple reads in parallel (94, 95, 98). However, the weakness of this technology was the accuracy of homopolymer sequences. Though the proportionally strong signal allowed to infer

the presence for homopolymers (98), the nonlinear increase in signal poses challenges in determining the exact numbers of nucleotides in a homopolymer (94). Secondly, the restriction on read length (200-300bp) as compared to the shotgun-Sanger sequencing technology posed algorithmic challenges for sequence assembly (99).

The 454-sequencing system (Life Science Corp) was the first commercialized next generation sequencing technology (99). It first fragmented the genomic DNA and then ligated these fragments to adapter sequences for adhesion to micro-scale beads followed by PCR amplification in water-in-oil emulsion droplets (100, 101). These DNA coated beads were flushed and fit into an array of picoliter-scale wells where the sequencing cycles took place and charge-coupled device (CCD) attached to detect the optical signal (101). The parallel in vitro amplification and sequencing on a micrometer scale is what came to define the second-generation sequencing technology (99). After the achievements of 454 sequencing, the most widely applied second generation sequencing technology is Illumina. Instead of bead-based emulsion PCR, Illumina featured bridge PCR on a solid substrate to which forward and reverse primers for amplifications were tethered nearby. As a result, when the fragmented DNA library, constructed with the forward and reverse adapter sequences at both ends, was flushed through the lawn of sequencing surface, the DNA sequence would arch over on the solid substrate and prime the two ends for a two-round PCR amplification. While the first round performed a forward amplifying orientation, the second round read from the alternative direction and together generated the pair-end (PE) read data which largely improved the accuracy of sequencing (90, 99). However, the accuracy of the pair-end approach came at the cost of sequencing length restriction of approximately 35 bp long. As a result, such short reads relied on the existence of reference sequences to perform accurate assembly (90, 99, 102). Instead of adding one type of nucleotide at a time during the repeated sequencing cycle,

Illumina add four types of modified nucleotides at once. The modified nucleotides, also known as "reversible terminators", contained a cleavable moiety at the 3'hydroxyl position which assured a single-base incorporation in each cycle, and a fluorescent label which allowed to identify the type of nucleotide added (90, 102). Other second-generation sequencing technology platforms including the sequencing by oligonucleotide ligation and detection (SOLiD) and Ion Torrent which utilizes rapid detection of proton release with microprocessor chips during polymerization (90). Irrespective of the variations among platforms, the fundamental steps shared by short-read parallel sequencing technologies involve DNA fragmentation, adapter ligations, DNA library surface attachment, and in-situ amplification (103).

Third generation sequencing

Despite the enhancements in read lengths brought about by the release of Ill'mina's HiSeq and MiSeq instruments, the maximum reads length remained at 250bp (104, 105). The short read lengths render it unsuitable for de novo assembly, gene isoform detection, and identification of epigenomic modifications (106). Therefore, in response to the requirements, third-generation sequencing technology emerged, offering long-read and single molecule sequencing features. Now there are two major third generation sequencing platforms, the single molecule real time (SMRT) platform from Pacific Biosciences (PacBio) and nanopore sequencing technology from oxford nanopore technology (ONT).

The PacBio was the first commercially available long-read sequencer capable of yielding up to 300kbp long reads (103). Utilizing a sequence-by-synthesis methodology, akin to other second-generation sequencing techniques, this process stands out with its elimination of pauses between read steps. As a result, leading to a substantial enhancement in sequencing rate (107). The DNA template for sequence is a circular molecule referred to as SMRTbell, which is composed of

a double-stranded segment containing the sequence of interest and two single-stranded hairpin adapters ligated at both ends for primer bind. During the sequencing process, SMRTbells are introduced into a sequencing unit known as a zero-mode waveguide (ZMW). As the strand displacing polymerase, harbored in each ZMW, initiates the replication from hairpin loop of the SMRTbell (108), ZMW records the continuous signals emitted from the four fluorescent-labeled triphosphates as each nucleotide incorporation occurs. By interpreting the uninterrupted light pulses, DNA sequence could be inferred (109). Meanwhile, kinetic variations during the process could be analyzed and allow for detection of epigenetic modification, such as DNA methylation (106). Because of the circular structure of SMRTbell, the replication process continues to extend until the exhaustion of polymerase, which means the targeted template could be sequenced more than once. The sequencing process captured within a ZMW is known as continuous long read (CLR), and the segment consisting exclusively of the repeatedly sequenced specific template is referred to as subreads. Subreads serve to build and improve the accuracy of sequence consensus, as the random errors existing in each subreads should be corrected through inter-comparison. Due to the limited lifespan of each polymerase, a trade-off must be considered between the length of the template sequence and the depth of coverage in SMRT sequencing. The longer the targeted template sequence is, the less subreads could be obtained to improve accuracy (103).

The concept of nanopore sequencing technology emerged early in the 1990s when researchers detected transient changes in ionic current signals as single-stranded RNA or DNA molecules passed through a lipid membrane via an ionic channel protein derived from Staphylococcus. Initially, the technology solely allowed inferences on the length of polymer molecule based on the duration of the signal changes (110). The key breakthrough was improving the base recognition through the genetic modification of the nanopore protein, which amplified the

signal disparities among nucleotides and substantially enhanced the power of base discrimination (111). Another significant advancement was achieved through the incorporation of polymerase motor proteins, or a biotin-streptavidin complex attached to the DNA template to reduce the translocation speed of ssDNA. The immobilization and ratcheting of ssDNA through the nanopore played a crucial role in extending the observation time for the detection of subtle changes in electric signals (112-114). The integration of these two technologies has paved the way for nanopore DNA strand sequencing. Nonetheless, the ongoing challenge lies in achieving comparable levels of accuracy to those attained by short-read sequencing technologies. When the technology first emerged commercially in 2014, the substantial error rate was up to 40% (115), but efforts in refinements of device and kit chemistry throughout the years has substantially increased the sequencing accuracy (over 99% in the best chemistry kit) and translocation rate (up till 450 bp/s) (116). Moreover, ONT offers a range of scalable device options including MinION, GridION and PromethION, catering to diverse sequencing throughput requirements ranging from low to high (117). Among, the MinION sequencer are becoming more and more popular in both laboratories and field investigations due to its affordability and portability (118).

Data Analysis of nanopore sequencing

In addition to optimizing in hardware and library chemistry, bioinformatic tools continued to make progress on enhancing accuracy as well as increasing the speed for analysis (119). Ever since the advent of the technology, ONT has continuously introduced a series of base callers with various models, consistently enhancing their performance. The initial base calling method employed a hidden Markov model (HMM) to estimate the probability of subsequent k-mers sequences using previous segmented event data, integrating all available information from the current measurements. Second-generation models also inferred k-mers sequences, but only

directly from raw current data. However, due to the nature of nanopore base calling with k-mers, predicting regions with homopolymers longer than the detection unit of k-mers became more challenging and often leads to base call errors (120). The subsequent base callers adopted a flip-flop model utilized neural networks that allow to infer sequences to a resolution of a single base instead of k-mers. The next generation base callers, Guppy, incorporated the flip-flop model and customized base-calling models to allow more accurate base prediction with specific applications (119).

Error corrections can be important analysis approach in assuring the data quality as well. Two type of error corrections are available – the "hybrid correction" and "the self-correction" (119). The self-correction method aligns the long reads against one another and perform corrections based on the consensus reads. By exclusively utilizing long-read data, this approach effectively circumvents biases that may arise from fragmentation during second-generation sequencing (121). In contrast, "hybrid correction" approach harnesses the strength from both error prone long-reads of third generation sequencing and accurate short-reads from second generation sequencing (122). As the long reads are suitable to deal with large range complexity or structural variants while the short reads sharpen the localized details, this approach were proven with higher accuracy and longer contiguity than relying solely on one type of sequencing approach (123).

Application of MinION sequencing

With the feature of high-throughput, real-time sequencing, and portability, MinION has been a very useful tool in infectious pathogen detection and controlling disease outbreak. For instance, a more sensitive diagnosis of bacterial meningitis can be delivered with less than 10 hours using nanopore sequencing, significantly outperforming the conventional diagnostic methods that usually take up to 7 days (124). Furthermore, it has been well-documented the application of

nanopore sequencing in on-site surveillance, enabling the monitoring of rapid spreading viral diseases such as Ebola virus and Zika virus outbreaks (125, 126).

The long read features allow nanopore sequencing to accurately identify repetitive genomes and complex genomic structures. For example, the ultra-long MinION read sequencing had assisted in closing the unsolvable gaps and assembled a single contig of major histocompatibility complex (MHC) containing long tandem repeats. Ultimately, yielded the first MHC genome being haplotyped in full length (127). Additionally, transcriptomic data analysis may also rely on long reads to reconstruct the complex full-length gene isoforms diversity. In a B cell transcriptomics study, hundreds of splicing events were identified, and thousands of transcriptions start, and end sites were newly annotated. Most of these genes were identified as unique surface receptors distinguishing B cell forms from other immune cells, which may turn a brand-new page in understanding the B cell diversity and the importance of these receptors to B cell biology (128).

Reference

1. Graham R, Thorp F, James W. 1930. Subacute or chronic infectious avian laryngotracheitis. The Journal of Infectious Diseases:87-91.

2. Cover MS. 1996. The early history of infectious laryngotracheitis. Avian Dis 40:494-500.

3. Crawshaw GJ, Boycott BR. 1982. Infectious laryngotracheitis in peafowl and pheasants. Avian Diseases:397-401.

4. Winterfield R, So IG. 1968. Susceptibility of turkeys to infectious laryngotracheitis. Avian Diseases 12:191-202.

5. Ou S, Giambrone J, Macklin K. 2011. Infectious laryngotracheitis vaccine virus detection in water lines and effectiveness of sanitizers for inactivating the virus. Journal of Applied Poultry Research 20:223-230.

6. Ou S-C, Giambrone J, Macklin K. 2012. Detection of infectious laryngotracheitis virus from darkling beetles and their immature stage (lesser mealworms) by quantitative polymerase chain reaction and virus isolation. Journal of Applied Poultry Research 21:33-38.

7. Beltrán LG, Williams S, Zavala G, Guy JS, García M. 2017. The route of inoculation dictates replication patterns of infectious laryngotracheitis virus (ILTV) pathogenic strain and chicken embryo origin (CEO) vaccine. Avian Pathology:1-26.

8. Williams RA, Bennett M, Bradbury JM, Gaskell RM, Jones RC, Jordan FT. 1992. Demonstration of sites of latency of infectious laryngotracheitis virus using the polymerase chain reaction. J Gen Virol 73 (Pt 9):2415-20.

9. Bagust TJ, Calnek BW, Fahey KJ. 1986. Gallid-1 herpesvirus infection in the chicken. 3. Reinvestigation of the pathogenesis of infectious laryngotracheitis in acute and early post-acute respiratory disease. Avian Dis 30:179-90.

10. Hughes C, Gaskell R, Jones R, Bradbury J, Jordan F. 1989. Effects of certain stress factors on the re-excretion of infectious laryngotracheitis virus from latently infected carrier birds. Research in Veterinary Science 46:274-276.

11. Jordan F. 1966. A review of the literature on infectious laryngotracheitis (ILT). Avian Diseases 10:1-26.

12. Timurkaan N, Yilmaz F, Bulut H, Ozer H, Bolat Y. 2003. Pathological and immunohistochemical findings in broilers inoculated with a low virulent strain of infectious laryngotracheitis virus. Journal of veterinary science 4:175-180.

13. García M. 2017. Current and future vaccines and vaccination strategies against infectious laryngotracheitis (ILT) respiratory disease of poultry. Veterinary microbiology 206:157-162.

14. Swayne DE. 2013. Diseases of poultry. John Wiley & Sons.

15. Dufour-Zavala L. 2008. Epizootiology of infectious laryngotracheitis and presentation of an industry control program. Avian Dis 52:1-7.

16. Rodriguez-Avila A, Oldoni I, Riblet S, Garcia M. 2007. Replication and transmission of live attenuated infectious laryngotracheitis virus (ILTV) vaccines. Avian Dis 51:905-11.

17. Palomino-Tapia VA, Zavala G, Cheng S, García M. 2019. Long-term protection against a virulent field isolate of infectious laryngotracheitis virus induced by inactivated, recombinant, and modified live virus vaccines in commercial layers. Avian Pathology 48:209-220.

18. Guy JS, Barnes HJ, Smith L. 1991. Increased virulence of modified-live infectious laryngotracheitis vaccine virus following bird-to-bird passage. Avian Dis 35:348-55.

19. Tong GZ, Zhang SJ, Wang L, Qiu HJ, Wang YF, Wang M. 2001. Protection of chickens from infectious laryngotracheitis with a recombinant fowlpox virus expressing glycoprotein B of infectious laryngotracheitis virus. Avian Pathol 30:143-8.

20. Zhao W, Spatz S, Zhang Z, Wen G, Garcia M, Zsak L, Yu Q. 2014. Newcastle disease virus (NDV) recombinants expressing infectious laryngotracheitis virus (ILTV) glycoproteins gB and gD protect chickens against ILTV and NDV challenges. J Virol 88:8397-406.

21. Gimeno IM, Cortes AL, Faiz NM, Hernandez-Ortiz BA, Guy JS, Hunt HD, Silva RF. 2015. Evaluation of the Protection Efficacy of a Serotype 1 ’arek's Disease Virus-

Vectored Bivalent Vaccine Against Infectious Laryngotracheitis and ’arek's Disease. Avian Dis 59:255-62.

22. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Dempsey DM, Dutilh BE, García ML, Curtis Hendrickson R. 2022. Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). Archives of virology 167:2429-2440.

23. Thureen DR, Keeler CL, Jr. 2006. Psittacid herpesvirus 1 and infectious laryngotracheitis virus: Comparative genome sequence analysis of two avian alphaherpesviruses. J Virol 80:7863-72.

24. Gatherer D, Depledge DP, Hartley CA, Szpara ML, Vaz PK, Benkő M, Brandt CR, Bryant NA, Dastjerdi A, Doszpoly A. 2021. ICTV virus taxonomy profile: herpesviridae 2021. Journal of General Virology 102:001673.

25. Leib DA, Bradbury JM, Hart CA, McCarthy K. 1987. Genome isomerism in two alphaherpesviruses: Herpesvirus saimiri-1 (Herpesvirus tamarinus) and avian infectious laryngotracheitis virus. Brief report. Arch Virol 93:287-94.

26. Spatz SJ, Volkening JD, Keeler CL, Kutish GF, Riblet SM, Boettger CM, Clark KF, Zsak L, Afonso CL, Mundt ES, Rock DL, Garcia M. 2012. Comparative full genome analysis of four infectious laryngotracheitis virus (Gallid herpesvirus-1) virulent isolates from the United States. Virus Genes 44:273-85.

27. Lee S-W, Markham PF, Markham JF, Petermann I, Noormohammadi AH, Browning GF, Ficorilli NP, Hartley CA, Devlin JM. 2011. First complete genome sequence of infectious laryngotracheitis virus. BMC genomics 12:1-6.

28. Ziemann K, Mettenleiter TC, Fuchs W. 1998. Gene arrangement within the unique long genome region of infectious laryngotracheitis virus is distinct from that of other alphaherpesviruses. J Virol 72:847-52.

29. Wild MA, Cook S, Cochran M. 1996. A genomic map of infectious laryngotracheitis virus and the sequence and organization of genes present in the unique short and flanking regions. Virus Genes 12:107-16.

30. Fuchs W, Mettenleiter TC. 1999. DNA sequence of the UL6 to UL20 genes of infectious laryngotracheitis virus and characterization of the UL10 gene product as a nonglycosylated and nonessential virion protein. J Gen Virol 80 (Pt 8):2173-82.

31. Ziemann K, Mettenleiter TC, Fuchs W. 1998. Infectious laryngotracheitis herpesvirus expresses a related pair of unique nuclear proteins which are encoded by split genes located at the right end of the UL genome region. J Virol 72:6867-74.

32. Veits J, Luschow D, Kindermann K, Werner O, Teifke JP, Mettenleiter TC, Fuchs W. 2003. Deletion of the non-essential UL0 gene of infectious laryngotracheitis (ILT) virus leads to attenuation in chickens, and UL0 mutants expressing influenza virus haemagglutinin (H7) protect against ILT and fowl plague. J Gen Virol 84:3343-52.

33. Nadimpalli M, Lee S, Devlin J, Gilkerson J, Hartley C. 2017. Impairment of infectious laryngotracheitis virus replication by deletion of the UL [-1] gene. Archives of virology 162:1541-1548.

34. Garcia M, Spatz SJ, Cheng Y, Riblet SM, Volkening JD, Schneiders GH. 2016. Attenuation and protection efficacy of ORF C gene-deleted recombinant of infectious laryngotracheitis virus. J Gen Virol 97:2352-62.

35. Gowthaman V, Kumar S, Koul M, Dave U, Murthy TGK, Munuswamy P, Tiwari R, Karthik K, Dhama K, Michalak I. 2020. Infectious laryngotracheitis: Etiology, epidemiology, pathobiology, and advances in diagnosis and control–a comprehensive review. Veterinary Quarterly 40:140-161.

36. Vogtlin A, Bruckner L, Ottiger HP. 1999. Use of polymerase chain reaction (PCR) for the detection of vaccine contamination by infectious laryngotracheitis virus. Vaccine 17:2501-6.

37. Humberd J, Garcia M, Riblet SM, Resurreccion RS, Brown TP. 2002. Detection of infectious laryngotracheitis virus in formalin-fixed, paraffin-embedded tissues by nested polymerase chain reaction. Avian Dis 46:64-74.

38. Davidson I, Raibstein I, Altory A. 2015. Differential diagnosis of fowlpox and infectious laryngotracheitis viruses in chicken diphtheritic manifestations by mono and duplex real-time polymerase chain reaction. Avian Pathol 44:1-4.

39. Thilakarathne DS, Hartley CA, Diaz-Méndez A, Coppo MJ, Devlin JM. 2020. Development and application of a combined molecular and tissue culture-based approach to detect latent infectious laryngotracheitis virus (ILTV) in chickens. Journal of Virological Methods 277:113797.

40. Mahmoudian A, Kirkpatrick NC, Coppo M, Lee SW, Devlin JM, Markham PF, Browning GF, Noormohammadi AH. 2011. Development of a SYBR Green quantitative polymerase chain reaction assay for rapid detection and quantification of infectious laryngotracheitis virus. Avian Pathol 40:237-42.

41. Santander Parra S, Nunez L, Buim MR, Astolfi-Ferreira CS, Piantino Ferreira AJ. 2018. Development of a qPCR for the detection of infectious laryngotracheitis virus (ILTV) based on the gE gene. British poultry science 59:402-407.

42. Callison SA, Riblet SM, Oldoni I, Sun S, Zavala G, Williams S, Resurreccion RS, Spackman E, Garcia M. 2007. Development and validation of a real-time Taqman PCR assay for the detection and quantitation of infectious laryngotracheitis virus in poultry. J Virol Methods 139:31-8.

43. Creelan JL, Calvert VM, Graham DA, McCullough SJ. 2006. Rapid detection and characterization from field cases of infectious laryngotracheitis virus by real-time polymerase chain reaction and restriction fragment length polymorphism. Avian Pathol 35:173-9.

44. Nielsen OL, Handberg KJ, Jorgensen PH. 1998. In situ hybridization for the detection of infectious laryngotracheitis virus in sections of trachea from experimentally infected chickens. Acta Vet Scand 39:415-21.

45. Chacon JL, Ferreira AJ. 2008. Development and validation of nested-PCR for the diagnosis of clinical and subclinical infectious laryngotracheitis. J Virol Methods 151:188-93.

46. Zhao Y, Kong C, Cui X, Cui H, Shi X, Zhang X, Hu S, Hao L, Wang Y. 2013. Detection of infectious laryngotracheitis virus by real-time PCR in naturally and experimentally infected chickens. PLoS One 8:e67598.

47. Shibley G, Luginbuhl R, Helmboldt C. 1962. A study of infectious laryngotracheitis virus. I. Comparison of serologic and immunogenic properties. Avian Diseases 6:59-71.

48. Spatz SJ, Garcia M, Riblet S, Ross TA, Volkening JD, Taylor TL, Kim T, Afonso CL. 2019. MinION sequencing to genotype US strains of infectious laryngotracheitis virus. Avian Pathology 48:255-269.

49. Blacker HP, Kirkpatrick NC, Rubite'A, O'Rourke D, Noormohammadi AH. 2011. Epidemiology of recent outbreaks of infectious laryngotracheitis in poultry in Australia. Aust Vet J 89:89-94.

50. Han MG, Kim SJ. 2003. Efficacy of live virus vaccines against infectious laryngotracheitis assessed by polymerase chain reaction-restriction fragment length polymorphism. Avian Dis 47:261-71.

51. Han MG, Kim SJ. 2001. Analysis of Korean strains of infectious laryngotracheitis virus by nucleotide sequences and restriction fragment length polymorphism. Vet Microbiol 83:321-31.

52. Kirkpatrick NC, Mahmoudian'A, O'Rourke D, Noormohammadi AH. 2006. Differentiation of infectious laryngotracheitis virus isolates by restriction fragment length polymorphic analysis of polymerase chain reaction products amplified from multiple genes. Avian Dis 50:28-34.

53. Ojkic D, Swinton J, Vallieres M, Martin E, Shapiro J, Sanei B, Binnington B. 2006. Characterization of field isolates of infectious laryngotracheitis virus from Ontario. Avian Pathol 35:286-92.

54. Oldoni I, Garcia M. 2007. Characterization of infectious laryngotracheitis virus isolates from the US by polymerase chain reaction and restriction fragment length polymorphism of multiple genome regions. Avian Pathol 36:167-76.

55. Choi EJ, La TM, Choi IS, Song CS, Park SY, Lee JB, Lee SW. 2016. Genotyping of infectious laryngotracheitis virus using allelic variations from multiple genomic regions. Avian Pathol 45:443-9.

56. Fakhri O, Hartley CA, Devlin JM, Browning GF, Noormohammadi AH, Lee SW. 2019. Development and application of high-resolution melting analysis for the classification of infectious laryngotracheitis virus strains and detection of recombinant progeny. Arch Virol 164:427-438.

57. Fakhri O, Devlin JM, Browning GF, Vaz PK, Thilakarathne D, Lee S-W, Hartley CA. 2020. Genomic recombination between infectious laryngotracheitis vaccine strains occurs under a broad range of infection conditions in vitro and in ovo. PLoS One 15:e0229082.

58. Loncoman CA, Hartley CA, Coppo MJ, Vaz PK, Diaz-Mendez A, Browning GF, Lee SW, Devlin JM. 2017. Development and application of a TaqMan single nucleotide

polymorphism genotyping assay to study infectious laryngotracheitis virus recombination in the natural host. PLoS One 12:e0174590.

59. Lee SW, Hartley CA, Coppo MJ, Vaz PK, Legione AR, Quinteros JA, Noormohammadi AH, Markham PF, Browning GF, Devlin JM. 2015. Growth kinetics and transmission potential of existing and emerging field strains of infectious laryngotracheitis virus. PLoS One 10:e0120282.

60. Agnew-Crumpton R, Vaz PK, Devlin 'M, O'Rourke D, Blacker-Smith HP, Konsak-Ilievski B, Hartley CA, Noormohammadi AH. 2016. Spread of the newly emerging infectious laryngotracheitis viruses in Australia. Infect Genet Evol 43:67-73.

61. Sabir AJ, Olaogun 'M, O'Rourke D, Fakhri O, Coppo MJC, Devlin JM, Konsak-Ilievski B, Noormohammadi AH. 2020. Full genomic characterisation of an emerging infectious laryngotracheitis virus class 7b from Australia linked to a vaccine strain revealed its identity. Infect Genet Evol 78:104067.

62. Oldoni I, Rodriguez-Avila A, Riblet S, Garcia M. 2008. Characterization of infectious laryngotracheitis virus (ILTV) isolates from commercial poultry by polymerase chain reaction and restriction fragment length polymorphism (PCR-RFLP). Avian Dis 52:59-63.

63. Chacon JL, Ferreira AJ. 2009. Differentiation of field isolates and vaccine strains of infectious laryngotracheitis virus by DNA sequencing. Vaccine 27:6731-8.

64. Perez Contreras A, van der Meer F, Checkley S, Joseph T, King R, Ravi M, Peters D, Fonseca K, Gagnon CA, Provost C. 2020. Analysis of whole-genome sequences of infectious laryngotracheitis virus isolates from poultry flocks in Canada: evidence of recombination. Viruses 12:1302.

65. Kim HR, Kang MS, Kim MJ, Lee HS, Kwon YK. 2013. Restriction fragment length polymorphism analysis of multiple genome regions of Korean isolates of infectious laryngotracheitis virus collected from chickens. Poult Sci 92:2053-8.

66. Yan Z, Li S, Xie Q, Chen F, Bi Y. 2016. Characterization of field strains of infectious laryngotracheitis virus in China by restriction fragment length polymorphism and sequence analysis. J Vet Diagn Invest 28:46-9.

67. Davidson I, Nagar S, Ribshtein I, Shkoda I, Perk S, Garcia M. 2009. Detection of wild- and vaccine-type avian infectious laryngotracheitis virus in clinical samples and feather shafts of commercial chickens. Avian Dis 53:618-23.

68. Chang P-C, Shieh HK, Shien J-H, Kang S-W. 2000. A homopolymer stretch composed of variable numbers of cytidine residues in the terminal repeats of infectious laryngotracheitis virus. Avian diseases:125-131.

69. Yang Z, Murata S, Fujisawa S, Takehara M, Katakura K, Hmoon MM, Win SY, Bawm S, Konnai S, Ohashi K. 2020. Molecular detection and genetic characterization of infectious laryngotracheitis virus in poultry in Myanmar. BMC Vet Res 16:453.

70. Craig MI, Rojas MF, van der Ploeg CA, Olivera V, Vagnozzi AE, Perez AM, Konig GA. 2017. Molecular Characterization and Cluster Analysis of Field Isolates of Avian Infectious Laryngotracheitis Virus from Argentina. Front Vet Sci 4:212.

71. Lee SW, Devlin JM, Markham JF, Noormohammadi AH, Browning GF, Ficorilli NP, Hartley CA, Markham PF. 2011. Comparative analysis of the complete genome sequences of two Australian origin live attenuated vaccines of infectious laryngotracheitis virus. Vaccine 29:9583-7.

72. Piccirillo A, Lavezzo E, Niero G, Moreno A, Massi P, Franchin E, Toppo S, Salata C, Palu G. 2016. Full Genome Sequence-Based Comparative Study of Wild-Type and Vaccine Strains of Infectious Laryngotracheitis Virus from Italy. PLoS One 11:e0149529.

73. Spatz S, García M, Fuchs W, Loncoman C, Volkening J, Ross T, Riblet S, Kim T, Likens N, Mettenleiter T. 2023. Reconstitution and Mutagenesis of Avian Infectious Laryngotracheitis Virus from Cosmid and Yeast Centromeric Plasmid Clones. Journal of Virology 97:e01406-22.

74. Garcia M, Volkening J, Riblet S, Spatz S. 2013. Genomic sequence analysis of the United States infectious laryngotracheitis vaccine strains chicken embryo origin (CEO) and tissue culture origin (TCO). Virology 440:64-74.

75. Chandra YG, Lee J, Kong B-W. 2012. Genome sequence comparison of two United States live attenuated vaccines of infectious laryngotracheitis virus (ILTV). Virus Genes 44:470-474.

76. Loncoman CA, Hartley CA, Coppo MJ, Vaz PK, Diaz-Méndez A, Browning GF, García M, Spatz S, Devlin JM. 2017. Genetic Diversity of Infectious Laryngotracheitis Virus during In Vivo Coinfection Parallels Viral Replication and Arises from Recombination Hot Spots within the Genome. Applied and environmental microbiology 83:e01532-17.

77. Lee SW, Markham PF, Coppo MJ, Legione AR, Markham JF, Noormohammadi AH, Browning GF, Ficorilli N, Hartley CA, Devlin JM. 2012. Attenuated vaccines can recombine to form virulent field viruses. Science 337:188.

78. Lee S-W, Devlin JM, Markham JF, Noormohammadi AH, Browning GF, Ficorilli NP, Hartley CA, Markham PF. 2013. Phylogenetic and molecular epidemiological studies reveal evidence of multiple past recombination events between infectious laryngotracheitis viruses. PloS one 8:e55121.

79. Zhao Y, Kong C, Wang Y. 2015. Multiple comparison analysis of two new genomic sequences of ILTV strains from China with other strains from different geographic regions. PLoS One 10:e0132747.

80. Kozlov A, Zinyakov N, Kolosov S, Mudrak N, Chvala I. 2016. Analysis of whole genome sequence of s”ra”n" 0" of avian infectious laryngotracheitis virus. Veterinary Science Today:55-58.

81. La TM, Choi EJ, Lee JB, Park SY, Song CS, Choi IS, Lee SW. 2019. Comparative genome analysis of Korean field strains of infectious laryngotracheitis virus. PLoS One 14:e0211158.

82. Morales Ruiz S, Bendezu Eguis J, Montesinos R, Tataje-Lavanda L, Fernández-Díaz M. 2018. Full-genome sequence of infectious laryngotracheitis virus (Gallid Alphaherpesvirus 1) strain VFAR-043, isolated in Peru. Genome announcements 6:10.1128/genomea. 00078-18.

83. Wei X, Shao Y, Han Z, Sun J, Liu S. 2020. Glycoprotein-C-gene-deleted recombinant infectious laryngotracheitis virus expressing a genotype VII Newcastle disease virus fusion protein protects against virulent infectious laryngotracheitis virus and Newcastle disease virus. Vet Microbiol 250:108835.

84. Elshafiee EA, Hassan MSH, Provost C, Gagnon CA, Ojkic D, Abdul-Careem MF. 2022. Comparative full genome sequence analysis of wild-type and chicken embryo origin vaccine-like infectious laryngotracheitis virus field isolates from Canada. Infect Genet Evol 104:105350.

85. Poptsova M', Il'icheva IA, Nechipurenko DY, Panchenko LA, Khodikov MV, Oparina NY, Polozov RV, Nechipurenko YD, Grokhovsky SL. 2014. Non-random DNA fragmentation in next-generation sequencing. Sci Rep 4:4532.

86. Renner DW, Szpara ML. 2018. Impacts of genome-wide analyses on our understanding of human herpesvirus diversity and evolution. Journal of virology 92:e00908-17.

87. Asif K, O'Rourke D, Shil P, Steer-Cope PA, Legione AR, Marenda MS, Noormohammadi AH. 2022. Rapid typing of infectious laryngotracheitis virus directly from tracheal tissues based on next-generation sequencing. Archives of Virology 167:1151-1155.

88. Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology 94:441-448.

89. Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. Proceedings of the National Academy of Sciences 74:560-564.

90. Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107:1-8.

91. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences 74:5463-5467.

92. Anderson S. 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic acids research 9:3015-3027.

93. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

94. Ronaghi M, Uhlén M, Nyrén P. 1998. A sequencing method based on real-time pyrophosphate. Science 281:363-365.

95. Hyman ED. 1988. A new method of sequencing DNA. Analytical biochemistry 174:423-436.

96. Nyrén P, Lundin A. 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Analytical biochemistry 151:504-509.

97. Nyrén P. 1987. Enzymatic method for continuous monitoring of DNA polymerase activity. Analytical biochemistry 167:235-238.

98. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry 242:84-89.

99. Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nature biotechnology 26:1135-1145.

100. Tawfik DS, Griffiths AD. 1998. Man-made cell-like compartments for molecular evolution. Nature biotechnology 16:652-656.

101. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

102. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. nature 456:53-59.

103. Hu T, Chitnis N, Monos D, Dinh A. 2021. Next-generation sequencing technologies: An overview. Hum Immunol 82:801-811.

104. Illumina. HiSeq 2500 Specification

105. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics 13:1-13.

106. Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 13:278-89.

107. Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. Hum Mol Genet 19:R227-40.

108. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38:e159.

109. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133-138.

110. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. 1996. Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences 93:13770-13773.

111. Stoddart D, Heron AJ, Klingelhoefer J, Mikhailova E, Maglia G, Bayley H. 2010. Nucleobase recognition in ssDNA at the central constriction of the α-hemolysin pore. Nano letters 10:3633-3637.

112. Cockroft SL, Chu J, Amorin M, Ghadiri MR. 2008. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. Journal of the American Chemical Society 130:818-820.

113. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. 2012. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. Nature biotechnology 30:344-348.

114. Deamer D, Akeson M, Branton D. 2016. Three decades of nanopore sequencing. Nat Biotechnol 34:518-24.

115. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res 25:1750-6.

116. Technology ON. Introduction to library preparation chemistry.

117. technology ON. product brochure.

118. Walter MC, Zwirglmaier K, Vette P, Holowachuk SA, Stoecker K, Genzel GH, Antwerpen MH. 2017. MinION as part of a biomedical rapidly deployable laboratory. J Biotechnol 250:16-22.

119. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol 39:1348-1365.

120. Rang FJ, Kloosterman WP, de Ridder J. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome biology 19:90.

121. Salmela L, Walve R, Rivals E, Ukkonen E. 2017. Accurate self-correction of errors in long reads using de Bruijn graphs. Bioinformatics 33:799-806.

122. Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. Bioinformatics 30:3506-14.

123. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, Dvornicic M, Soldo JP, Koh JY, Tong C, Ng OT, Barkham T, Young B, Marimuthu K, Chng KR, Sikic M, Nagarajan N. 2019. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol 37:937-944.

124. Moon J, Kim N, Kim TJ, Jun JS, Lee HS, Shin HR, Lee ST, Jung KH, Park KI, Jung KY, Kim M, Lee SK, Chu K. 2019. Rapid diagnosis of bacterial meningitis by nanopore 16S amplicon sequencing: A pilot study. Int J Med Microbiol 309:151338.

125. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouedraogo N, Afrough B, Bah A, Baum JH, Becker-Ziaja B, Boettcher JP, Cabeza-Cabrerizo M, Camino-Sanchez A, Carter LL, Doerrbecker J, Enkirch T, Dorival IGG, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallash E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, VitorIano I, Yemanaberhan RL, Zekeng EG, Trina R, Bello A, Sall AA, Faye O, et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature 530:228-232.

126. Faria NR, Quick J, Claro IM, Theze J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, du Plessis L, Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli A, Vinhal LC, Henriques CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Munoz-Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO, Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, de Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D, Moura LC, Dhalia R, Franca RF, et al. 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 546:406-410.

127. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature biotechnology 36:338-345.

128. Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, DuBois RM, Forsberg EC, Akeson M, Vollmers C. 2017. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. Nat Commun 8:16027.

CHAPTER 2

GENOTYPING ASSAY BY MULTIPLEX PCR AND HIGH THROUGHPUT MINION SEQUENCING OF INFECTIOUS LARYNGOTRACHEITIS VIRUS (ILTV) UNIQUE SHORT GENOME REGION [1]

[1]Yi-Chen Luo, [1]Maricarmen García, [1]April Skipper, [1]Sylva M. Riblet , [1]Patrick Wang

[2]Kelsey T. Young, [2]James B. Stanton [3]Zhihan Xian . To be submitted to *Journal of Clinical Microbiology*.

[1] Department of Population Health, College of Veterinary Medicine, Poultry Diagnostic & Research Center, University of Georgia, Athens, GA.

[2] Department of Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA.

[3] Center for Food Safety, University of Georgia, Griffin, GA.

Key words: MinION, nanopore, high throughput sequencing, Infectious Laryngotracheitis, Genotyping assay, Multiplex PCR

Abbreviations: C_T = cycle threshold; PCR = polymerase chain reaction ; ILTV=Infectious Laryngotracheitis Virus; SNP= Single Nucleotide Polymorphism; IR=inverted repeat; TR=terminal repeat; UL=unique long; US=unique short; bp=base pair; ORF=open reading frame; Hasegawa–Kishono–Yano =HKY; Tm=melting temperature; DNA= Deoxyribonucleic acid; gC=glycoprotein C; SPRI= Solid Phase Reversible Immobilization; tissue culture origin TCO= tissue culture origin; CEO=chicken embryo origin.

ABSTRACT

This study aims to develop an upgraded ILTV genotyping assay by sequencing an expanded viral genome region and increase the number of informative single nucleotide polymorphism (SNPs) sites to provide more accurate genotyping of ILTV strains. Also, by adopting the MinION sequencing technology, we seek to obtain higher sequencing depth of coverage on SNP sites. The Unique Short (U_s) region was chosen as the target for developing a multiplex PCR assay, utilizing overlapping primer pairs which amplifies approximately 1 kbp fragment. The assay yielded 15 functional primer pairs, encompassing a genome sequence of 12,706 bp long and including a total of 133 SNPs which differentiate among the previously identified nine ILTV genotypes. MinION sequencing of the U_S region provided more than 40k reads per sample and resulted in > 200 depth of coverage on each SNP site which correctly differentiated known ILTV strains into genotypes previously assigned by whole genome sequencing and multiallelic PCR Sanger sequencing assays. The novel multiplex PCR MinION sequencing assay was applied to clinical samples collected from current outbreaks of the disease. Clinical samples were successfully genotyped, and six additional unique SNPs were identified in current circulating isolates. Therefore, the established multiplex PCR MinION ILTV genotyping assay demonstrated to provide increased resolution and accuracy as compared to the multiallelic PCR Sanger sequencing assay.

INTRODUCTION

Infectious Laryngotracheitis (ILT) is a respiratory disease of chickens with worldwide distribution (1). The pathogen Infectious laryngotracheitis virus (ILTV) is a double stranded DNA virus classified under genus Iltovirus of family Herpesviridae (2), with a genome of approximately 150 kbp, which consists of two singular regions – the Unique Long (UL) region (110 kbp) and a Unique Short (US) region (13 kbp). The US region is flanked by two repeat regions – an inverted repeat (IR) (13 kbp) and a terminal repeat (TR) (13kbp). The ILTV genome has a predicted coding capacity of 76 to 80 no spliced open reading frames (ORFs) (3, 4). Since ILTV exists as a single serotype (5), the methods to differentiate among ILTV strains relies on genome nucleotide differences (6). ILTV strains from the United States have been classified into nine genotype groups. The USDA standard challenge strain is assigned to genotype I; the tissue culture origin (TCO) vaccine is assigned to genotype II; virulent strains closely related to the TCO vaccine were included within genotype III; chicken embryo origin (CEO) vaccines and circulating CEO strains are classified as to genotype IV; virulent strains closely related to the CEO vaccines are identified as genotype V; and wild-type virus circulating in poultry are assigned within genotypes VI whereas the wild-type in backyard flocks are categorized in VII to IX (7, 8). The early typing method by PCR amplification followed by restriction fragment length polymorphism (PCR-RFLP) utilizes restriction enzymes that cleave DNA amplicons at specific nucleotide sequences to create a variety of fragment lengths (9). Though effective, PCR-RFLP requires several restriction enzymes, large quantities of viral DNA and laborious procedures to accurately reveal differences among strains. Despite the amount of effort, the final output of RFLP is limited in displaying the extent of genetic diversities between strains as differences in RFLP pattern could not reveal a full extent of genetic diversities. Sites may escape the detection due to their inability to be recognized

by restriction enzymes. Genotyping assays that involve sequencing of the amplification product provide better resolution than RFLP since polymorphic sites detection are no longer restrained by restriction enzymes and allow to identify all SNPs within the amplification product. Therefore, to achieve a more accurate virus genotyping by sequencing, increasing the number of SNPs is required. For instance, a recently developed single allele PCR-sequencing assay that amplifies the ORFA-B genome region is capable to separate strains from the United States into five major groups (genotype I-III, IV, V,VI,VII-IX) (6, 10). This assay has a discriminatory power comparable to a more extended previously described multi-allelic PCR-sequencing analysis (11).

An epidemiologic genotyping study in 2008 using the multi-allelic PCR-RFLP assay found that out of 46 samples collected from outbreaks of the disease in the US, 22 samples belong to genotype V, 15 samples to genotype VI, six to genotype IV and one to genotype III (7). In 2020, after the advent of high throughput sequencing, full genome sequence of ILTV isolates from Canada showed that 24 viral strains were grouped with the CEO related genotype V, three strains were similar to the CEO vaccine strain genotype IV, three were similar to wild type genotype VI and two strains belong to genotypes VII to IX (10). Overall ILTV genotype data from US and Canada showed that although CEO vaccine related strains seemed to be more prevalent, it is not uncommon to see wild-type genotype VI virus also related to outbreaks of commercial poultry. A constant concern is the potential emergence of natural recombinant viruses arising from genome exchanges between CEO (Genotype IV) and TCO (Genotype III) vaccines or recombination between Genotype V and co-circulating viruses belonging to Genotype VI. Natural recombinants of ILTV were first reported in Australia. These emerged from the recombination events between Australian vaccines A20 and SA2 and the Serva strain a CEO vaccine mainly used in Europe and Australia. Full genome sequences of Australian isolates, revealed that genome exchange events

between A20, SA2 and the Serva strains occurred more frequently in the repeat regions (IR and TR) and some in the UL region of the ILTV genome, but not in the US region (12). These natural recombinants in Australia continue to recombine while circulating in the field and favorable phenotypes with higher fitness, transmissibility and pathogenicity had replace previous predominant strains (13). In a recent report full genome sequence from Canadian isolates indicated that strain BC-10-1122 may be a natural recombinant virus with the TCO vaccine as the major parental strain and CEO vaccine Poulvac ILT (Zoetis, Animal Health) and Nobilis Laryngo-vac (Merck Animal Health) as minor parental strains. Although authors did not identify the recombination site, the major genome differences for the BC-10-1122 strain resides in nucleotide positions 1 to 15,393 bp within the UL region which share 99.41% nucleotide identity with TCO vaccine strain and 77.09% identity with Poulvac ILT(Zoetis). The remaining regions of the genome share equal identity with CEO and TCO vaccine strains (8). The emergence of natural recombinants between TCO and CEO vaccines introduces a challenge to the existing genotyping method which is based on partial sequences of multiple genome regions. For instance, using the single allele PCR-sequencing assay amplifying the ORFA-B genome region (21,295bp-23,676bp) developed by Spatz et. al (6), the BC-10-1122 strain from Canada would have been cluster within the CEO vaccine (Genotype IV) clade and none of the genetic information from the main parental donor in the recombination event the TCO vaccine would be detected (10). Ideally, full-genome sequencing provides the most rigorous genotyping analysis (8, 14, 15) and allows to identify novel recombinant viruses (11). However, the quantity and quality of viral DNA from clinical samples hinder to acquire full and accurate ILTV genome sequences by next generation sequencing protocols. Asif et al (2022) utilized Illumina sequencing to obtain ILTV genome reads from trachea samples of experimentally infected chickens. Although large number of quality reads were

obtained, only a limited amount of them were successfully mapped to the ILTV genome and whole genome assembly was not successful (16).

Because accurate full genome sequences are very difficult to obtain from clinical samples, most laboratories still rely on separate target amplifications and PCR-sequencing analysis of those regions. Currently in our laboratory clinical samples received from outbreaks have been genotyped by separate amplifications of multiple alleles followed by sequencing. The alleles amplified are the glycoprotein M (gM) open reading frame (ORF), partial sequence of the gB ORF, and a region that encompass the ILTV specific ORF A and B genes. One limitation of this genotyping scheme is the restricted number of SNPs within the three regions does not allow to accurately infer how currently circulating genotypes (II/III, IV, V, and VI) are evolving as compared to those that were circulating between 2004 and 2008. The overall objective of this study is to upgrade the current ILTV genotyping scheme by expanding the region targeted for sequencing and utilize MinION to obtain deeper sequencing of informative single nucleotides. Sequences of the expanded genome regions should maintain the capacity to differentiate among current known viral genotypes and in addition to provide discriminatory power to identify variations within the prevalent circulating genotypes. Also, for effective surveillance, the new method should be sensitive enough to perform viral genotyping directly from clinical samples.

MATERIALS AND METHODS

Single Nucleotide Polymorphism (SNP) Analysis

To identify suitable alleles, 39 representative ILTV complete genome sequences from North America isolates including outbreak related viruses and vaccine strains were obtained from GenBank. The genome library was composed of two genotype I, four genotype II, one genotype III, eight genotype IV, thirteen genotype V, eight genotype VI and one genome from genotype VII, VIII, and X (Table 1). The genotype I USDA strain genome (Accession No:MN518177) was selected as the reference genome sequence and was compared to the remaining 38 query genomes to perform SNP calling. The analysis pipeline for SNP calling consists of four python scripts. (i) The first script identifies all the query genome FASTA files in the analysis folder and performs SNP calling on each of the query genome using software Mummer3 (17). The program aligns only the forward strands to avoid bias between forward and reverse strands of the inverted repeats. The SNPs positions in the reference strain as well as nucleotide/gap information of the reference and query sequences at the corresponding position is the output. Lastly, this script extracts the SNPs information from each of the 38 query genomes excluding all the sites that contain gaps in either the reference or query sequence and consolidates all the SNP calling information across the 38 query genomes, into two output files. The first output file contains a list of all SNP positions, while the second output file is composed of matrices that show corresponding nucleotide/gap positions for each query genome. (ii) The SNP matrices are categorized based on the genotype of the query genome. The outputs are multiple files that contains SNPs matrices from each of the genotype category (group I, II, III, IV, V, VI and VII-IX). (iii) SNP matrices with informative SNPs sites of closely related genotypes (II and III, IV and V, V and VI, VI and VII-IX) and changes within strains that belong to the same genotype were produced. The outputs are lists of SNP positions that

have the potential to differentiate between closely related viral genotypes or between strains within a genotype. (iv) The last script serves to exclude potential SNP calling bias. The tool Repeat Masker (18) was utilized to screen for low complexity regions, such as poly-purine or poly-pyrimidine stretches and regions with high A/T and G/C content, and simple repeat regions composed of di- to hexameric repeats on the USDA reference genome (Accession No:MN518177) (Table 2). With this information, SNPs that are in low complexity regions which prone for introduction of sequencing errors (19) were excluded from the SNP analysis. Lastly, because the genome repeat regions (I_R & T_R) are variable in length, ranging between 12,811 to 17,254 bp among strains. To address for the bias, we only included SNP sites located in the I_R and excluded the T_R region from the SNP analysis.

Phylogenetic analysis

To identify target sequences within the ILTV genome appropriate for genotyping, phylogenetic analysis was conducted on partitioned genome alignments. Considering the longer the fragments the more primer pairs are required to amplify via a multiplex PCR the target region, fragments of approximately 17kbp to 20kbp in length were analyzed for their ability to cluster strains of the same genotype together and to separate strains of closely related genotypes. The USDA reference genome was first annotated using Geneious Prime (v11.0.4) and then searched for U_L overlapping segments ranging from 17kbp to 20kbp, with each of the identified segment containing intact open reading frames for the analysis. As the length of U_S (13kbp) and I_R (17 kbp) regions were close to our goal, these two regions were analyzed by itself. The partitioned U_L segments, U_S, and I_R were used to construct phylogenetic trees by the Neighbor Joining method and Hasegawa–Kishono–Yano (HKY) substitution model using Geneious Prime (v11.0.4) to determine the clustering ability between the nine genotypes.

Multiplex PCR primer design for the ILTV genome U_S region amplification

A primer design software—Primal scheme (20) was utilized to design primers pairs for overlapping amplicons that tiled up the target US region of 13,094 bp. This tool was applied to the US sequences of 39 North America viral sequences in search for primer pairs that could anneal with specificity to all the ILTV genotype and will not interact with each other to use the primers in a multiplex amplification reaction. The multiplex PCR scheme was designed with a desired amplicon length of 1,000 bp, an overlap of at least 100 bp and a narrow melting temperIe (Tm) ranging between 59°C to 62°C, as predicted by software Primer 3 (21). The primer pairs output from Primal Scheme were validated to be free of hairpin formation through Geneious Prime (v.11.0.4). To prevent primer pairs from yielding amplification products shorter than the approximately desired 1,000 bp, alternate primer pairs were assigned to two primer pools for amplification. The multiplex PCRs for either pool of primers were run at a high annealing temperature and at a low primer concentration to reduce unspecific primer binding, and extended annealing time to promote PCR efficiency (20).

Optimal concentration of primers for effective multiplex PCR

To maintain a maximum primer concentration of 1μM per multiplex per PCR (22), individual primer concentration should not exceed 0.06μM. Individual primer pair concentrations between 0.015 to 0.06 μM were tested in single PCR. Additionally, it was important to validate the efficacy and sensitivity of these primers to amplify varying genotype sequences at moderate amounts of viral DNA under low primer concentration. The DNA from five ILTV isolates propagated in chicken kidney cells – TX J2 P7 (genotype VI), 1874C5 Ct P4 (genotype VI), Laryngo vac® (Zoetis) (genotype IV), LT-IVAX® (MSD Animal Health) (genotype II) and 63140 P4 (genotype V) were selected as templates and diluted to obtain working stocks with Ct values

ranging from 28.3 to 30.6. Working primer pair concentrations of 0.015µM, 0.03µM and 0.06µM were tested in single PCR to determine the lower concentration at which the primers were still able to reliable amplify. Single PCRs were assembled as follow, 12.5µL of Q5-Hotstart High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) (25), primer pairs at three different concentrations 0.015µM, 0.03µM and 0.06µM and nuclease free water with a final 25 µL reaction volume. The thermocycling conditions were programed for denaturing at 98°C for 30 seconds and with 40 amplification cycles at 98°C for 15 seconds and 68°C for 5 minutes. The PCR products were electrophoresed in 1% agarose gel with GelRed Nucleic Acid Stain (Millipore, MA, USA) to visually examine a band at approximately 1,000 bp of size. Primer pairs that did not produced amplicons under the three concentrations tested indicated that the primer pair was not suitable for multiplex PCR and needed to be redesigned.

Redesign Primers

Primer pairs that failed to amplify the viral template at a 0.06µM concentration or yield of MinION sequence reads was 3% or less were redesigned. The redesigned primers should withhold the conditions of yielding amplification product at approximately 1000 bp in size, with at least 100bp overlap with the adjacent amplicons and annealing temperatures ranging from 59℃ to 62℃. Primer were redesigned using Geneious Prime (v11.0.4) and PrimerPooler (v1.85) (23) was used to assess the hybridization potential within or between a primer pair sets using the thermodynamic value $\triangle G$ (Gibbs free energy; kcal/mol) as an estimator of primer interaction.

Clinical Samples

A total of 13 samples associated to outbreaks of the disease collected between Aug-2021 to March 2023 that were previously genotyped by PCR multi-locus and Sanger sequencing were tested by the multiplex PCR MinION sequencing. Total DNA was extracted from tracheal

scrapings and swabs, using QIAamp DNA Blood Kits (QIAGEN, Hilden, Germany). Before the multiplex PCR MinION sequencing genotyping assays, viral genome load per DNA sample was estimated by qPCR of the gC gene (24).

DNA Quantification

Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA) was used for DNA quantification. Two standards dilutions, Standard 1 (0 ng/µL in TE buffer) and Standard 2 (10ng/µL in TE buffer) were used to generate a linear standard curve within the core quantification range of 0.2 -100 ng DNA prior to reading any samples. The DNA concentration for each sample was routinely estimated by measurements of 2µl and 5µl of sample to adjust for high and low DNA contents.

Multiplex PCR amplicon library preparation for MinION

Overview

The efficacy of each primer pairs in the multiplex PCR was evaluated by simultaneously sequencing pools of amplicons from the multiplex reactions. Library preparation for MinION sequencing was performed following the steps below: (i) Multiplex PCR, (ii) End prepping, (iii) Barcode ligation, (iv) Adapter ligation and (v) flowcell priming, library loading and sequencing. Between each step, SPRI magnetic beads washes were performed to eliminate enzymes, primers, barcodes, and adapters. DNA quantifications are required after washes to assure appropriately normalized amount of DNA is carried on to the next step. Washes and normalization steps are essential to prepare a successful library for sequencing.

Unique short (Us) multiplex PCR

Two pools of eight primer pairs mix (pool1 and pool2) with each prepared at a final concentration of 10µM. The multiplex PCR was assembled with 25µL of Q5-Hotstart High-

Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA), 5 µL of 10µM pool 1 or pool 2 primer pairs mixes, 5µL of DNA sample template, and 15 µL nuclease free water for a final volume of 50 µL per reaction. A two-step thermocycling program of denaturing at 98°C for 30 seconds, followed by 40 cycles of 98°C for 15 seconds and annealing and extension at 68°C for 5 minutes. The multiplex PCR product for pool 1 and pool 2 were visually confirmed in 1% agarose gel with Gel-Red Nucleic Acid Stain (Millipore Sigma, Burlington, MA). To remove primers from pool 1 and pool 2 PCR products washes with a 0.6:1 bead (KAPA Pure Beads; Roche Sequencing Store) to sample ratio were performed. PCR products from pool 1 and pool 2 were quantified with Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific) and normalized to adjust for DNA yield differences.

Solid Phase Reversible Immobilization (SPRI) beads-based washes

During library preparation steps, magnetic beads washes were performed to eliminate primers, barcodes, and adapters. Solid phase reversible immobilization (SPRI) beads are negatively charged, resuspended in a salt and polyethylene glycol (PEG) buffer. The salt cations serve as a bridge between the negatively charged DNA and SPRI beads, whereas the PEG serves as a crowding reagent that squeezes out water molecules to keep the DNA attached to the bead to achieve size selection and purification of nucleic acid (25). To select for PCR products at approximate 1,000bp size, a 0.6:1 bead (KAPA Pure Beads; Roche Sequencing Store) to DNA volume ratio was required after Multiplex PCR and End-prepping, whereas a 0.5:1 bead-to-DNA volume ratio was optimal for barcode and adapter ligated PCR products and eliminating un-ligated barcodes and adapters. The required amount of bead suspension was added to the DNA sample in lo-bind tubes and incubated at room temperature for 15 minutes. Beads and buffer separation was conducted on a magnetic rack (MagJET Separation Rack, Thermo Scientific, Waltham, MA) and

washed with freshly prepared 80% ethanol twice. Once washed the tube was opened to allow the residual ethanol to evaporate. Then the lo-bind tubes were removed from the magnetic rack and the bead pellets were resuspended using nuclease free water and incubated at 37℃ for 15 minutes. During the resuspension process, periodically flicking the tube could assist in eluting. The tubes were transferred back to the magnetic rack and the supernatant containing DNA was utilized in the next step.

End prep

End prepping serves to repair the 5' and 3' ends and adds a single adenosine at the 3' end of a DNA molecule. The 3' end dA tails are meant to prevent DNA amplicons from being ligated to one another and facilitates the ligation process of the native barcodes. End prep reaction was performed in a 54.5 µl reaction containing 48µl of 1,000 – 2,000 fmol (equivalent of 13.75 – 27.5ng/µl) of the multiplex PCR product mix, 3.5µl of Next Ultra II End-prep reaction buffer (New England Biolabs, Ipswich, MA), and 3µl of Next Ultra II End-prep Enzyme Mix (New England Biolabs, Ipswich, MA) for 5 minutes at 20°C followed with 5 minutes at 65°C. After the end prepping reaction, products were washed with 0.6:1 beads-to-sample ratio followed by DNA quantification.

Barcode ligation

Native barcodes are unique sequences composed of 24 nucleotides that serve to identify each sample in the library. A total of 100-200 fmol of End-prepped DNA per sample in 22.5 µl was barcoded in a reaction containing 25µl of Blunt TA ligase (New England Biolabs, Ipswich, MA) and 2.5µl of assigned native barcodes (EXP NBD-104; Oxford Nanopore Technology, Oxford, United Kingdom). The barcoding reaction was incubated at room temperature for 10 to 15 minutes, followed by washes in a 0.5x bead-to-DNA ratio. After washes, DNA was quantified

and normalized with a total amount between 100-200 fmol per sample. Then, equal volumes of barcoded DNA from all samples were pooled to create the library.

Adapter ligation

The end-prepped and barcoded library (65µl) was combined with 5µl of Sequencing adaptors (AMXII; Oxford Nanopore Technology, Oxford, United Kingdom), 20µl of NEBNext Quick Ligation Reaction buffer and 10µl of Quick T4 DNA ligase (New England Biolabs, Ipswich, MA) as suggested by ONT protocol. Then 100µl of the adapter ligated library was wash with 50µl of beads (0.5x ratio) resuspended in short fragment buffer (SFB; Oxford Nanopore Technology, Oxford, United Kingdom) to eliminate non-ligated adaptors and eluted with 15µl elution buffer (EB; Oxford Nanopore Technology, Oxford, United Kingdom).

Flowcell priming and library loading

Flow cell (FLOW-MIN106D R9.4.1; Oxford Nanopore Technology, Oxford, United Kingdom), priming is a process of replacing the storage buffer with priming mix which consists of flush tether (FLT) and flush buffer (FB; Oxford Nanopore Technology). Priming mix can be loaded approximately 5 minutes apart via the priming port. The purpose for the first priming is to replace most of the storage buffer and prepare the flowcell for sequencing whereas the second priming process is to ensure no air bubbles present on the SpotON port prior loading DNA library. Hence, the first priming process loads 800µl of priming mix with the SpotON port remained closed while the second priming process loads 200µl of priming mix with the SpotON port lid opened to push out any air bubbles. A total of 12µl of adapter ligated library is combined with 37.5µl fuel containing sequencing buffer (SQB, SQK-LSK109; Oxford Nanopore Technology) and 25.5µl loading beads (LB, SQK-LSK109; Oxford Nanopore Technology) to form a 75µl library loading mix. The library loading mix should be prepared shortly prior to loading, as the fuel immediately

started to be catalyzed at the presence of adapters and the loading beads settle in a short time. Immediately after the second priming and ensuring equilibrated to room temperature, the library was loaded dropwise via SpotON port.

MinION data analysis pipeline

Overview

The MinION sequencing device captured electrical disruptions signals when DNA passed through the nanopore embedded on a flowcell (26). In this study, MinKNOW (v.22.08.06) from Oxford Nanopore Technology (ONT) was selected as the connection manager that allowed the computer to access the electrical signal from sequencing device and saved as raw data, FAST5 files. FAST5 files must undergo a serial analysis process to convert the electrical level signals into nucleotide sequences. The term for the nucleotide converting process is named base calling. After base calling the data is demultiplexed by barcode tags on each individual reads and sorted to different files correspond to each sample in the library. Lastly, these sequences were mapped to the ILTV reference genome sequence to obtain a consensus sequence for phylogenetic analysis and genotyping. Five shell scripts were written to perform the sequential analysis mentioned in real-time sequencing and archive the intermediate analysis products for the purpose of backtracking.

Base-calling and demultiplexing

For real-time analysis of the data, script A_fast5_guppy.sh archived the raw data under the "fast5" directory in chronological order and identified these directories in numerical order. Then, the software Guppy (v6.4.6) (27) performed base-calling for each file in each directory. The output data was saved in a FASTQ format under the corresponding archived directory number under a

newly generated directory “fastq” as the output path. During the base-calling process, Guppy selected for reads obtaining q-score > 7 by default and saved them under the “pass” directory in the output path. Otherwise, the reads with low q-score were archived under the “fail” directory under the output path. Porechop (v0.2.4) (28) was used at default setting to search for adapter/barcode sequences within each sequence read and either trimmed off the adapter sequences at the end of a read or break down chimeric reads if the adapter sequences were detected in the middle. Trimmed reads were sorted by barcodes and all the reads with the same barcode were concatenated to obtain a final fastq file.

Sequence taxonomy classification

To prevent mapping or assembly bias generated from non-ILTV sequence amplification, reads needed to be classified and designated with a taxonomy ID prior to consensus assembly. Centrifuge (v1.0.4) (33) is a pre-build local database index system that adopted a high-speed microbial classification tool to designate a taxonomy ID to individual read in the demultiplexed FASTQ. Reads assigned with *Herpesvirus* or *Gallid alphaherpesvirus 1* taxonomy IDs were selectively extracted to proceed for further analysis.

ILTV Reads mapping and Building an ILTV consensus sequence

The mapping reference sequence for the mapper tool BWA (v.0.7.17)(29) was created using the 16 primer targeted PCR amplicon sequences obtained from USDA (Accession No: MN518177). Each individual read of a sample was aligned to this reference sequence using the default settings, resulting in the generation of a bam file that contains sequence alignment information. Bam file then went through an indexing and sorting process using Samtools (v.1.6) (30) to obtain summary statistics of number of reads that were successfully mapped to each of the 16 PCR amplicon sequence and more importantly to apply genome viewer tools on the mapped

alignment to navigate through each genomic region. The indexed and sorted bam file was imported to Geneious Prime (v.11.0) to generate 16 PCR contigs files for each amplicon before primer trimming. The primer sequences for each contig were trimmed off to prevent the masking of any true nucleotide differences in these regions. Since the multiplex PCR assay was designed with overlapping, trimming off the primers did not affect sequence depth of coverage. As there is a 30 bp long interspersed duplication in the US9 amplicon of genotype VI strains (J2 Accession No:MF417808), the US9 contigs were submitted for a second mapping process using Geneious Prime (v.11.0) with inbuilt Geneious mapper at medium sensitivity and selected the "find short insertion and large deletions up to" function to search for the insertion that is less than 40 bp in the reads.

After trimming primers, all the reads from the 16 contigs files were submitted to a second round of mapping for consensus building using all the sequences from reads assigned to different contigs. The reference used for the second round of mapping depended on whether insertion sequences were identified in the US9 contigs. The J2 (Accession No:MF417808) US sequence was used as reference when the presence of the interspersed duplication was detected in the US9 amplicon sequences, otherwise the US sequence of USDA reference (Accession No: MN518177) was used. Considering the consensus accuracy, the following parameters were selected: (i) A minimum mapping score of ≥ 30 was determined to exclude reads with poor mapping quality. (ii) Consensus calling threshold Highest Quality (75%). In addition to default parameters, to acknowledge the stochastic base call errors and the homopolymer errors by MinION sequencing, the consensuses were further manually curated when disagreements between the built consensus and the reference sequence occurred to resolve the in/dels detected in reads. However, the curation process still withholds to abide by the 75% composition consensus calling criteria.

Computer Specs

Dell Precision 5820 Tower XCTO Base with Itel Core i9-10980XE(18 core, 36 threads, 3.00 GHz to 4.80 Ghz Turbo, 165W), NVIDIA® RTX™A4000, and 1TB M.2 PCIe NVM Class 40 Solid State Drive. Operating System: Ubuntu 20.04 LTS, 64-bit.

RESULTS

Informative SNP sites distribution

The distribution of informative SNPs across the USDA genome regions – U_L, I_R, and U_S among closely related ILTV genotype groups and within genotype VI strains is shown in Figure 1A. Figure 1B shows separately the distribution of informative SNPs between closely related genotype groups and SNPs within the genotype VI respectively. Figures 1A and 1B show that the informative SNPs across the ILTV genome are preferentially distributed toward the 5' and 3' ends of the genome, coinciding with the U_S, I_R and the first 10,000bp of the U_L region.

Genotype clustering analysis of partitioned Genome U_L segments, U_S, and I_R

The U_L genome region from 39 North American ILTV sequences were fragmented into 91 overlapping segments with a length between 17k to 20 kbp encompassing complete ORFs. The U_S and I_R regions were not fragmented as the size of each is 13k and 17k, respectively. Phylogenetic analysis was performed on U_L partitioned segments, the U_S and I_R regions to know whether the sharing informative SNPs among the closely related genotype group (II/III, IV/V, V/VI, and VI/ VII-IX) in each partition can generate clusters departing the closely related genotype groups of interest.

Based on the phylogenetic analysis of U_L partitions, I_R and U_S sequences, two main categories of phylogenetic trees were identified and are listed in Table 3. Category I sequences

were separated into five clusters (II/III), IV, V, VI, (VII-IX). On the other hand, sequences in category II were separated into four clusters (II/III), (IV/V), VI, (VII-IX). The primary distinction between these two categories lies in the fact that category II fails to cluster genotype V separately from genotype IV on a phylogenetic tree. Examples of phylogenetic trees representative of categories I and II are shown in Figure 2A and 2B respectively. Category I comprises three regions on the ILTV genome, from position 1 to 71,249 and from 73,233 to 109,587 of the U_L, and 126,833 to 139,926 of the U_S. In total, there are 48 U_L partitions that encompass 17 to 20 kbp within the genome coordinates of 1 to 71,249, and 21 partitions betwe–n 73,233 - 109,578 within category I. On the other hand, category II includes sequences within three regions of the ILTV genome, from position 4,680 to 26,186 and from 63,589 to 91,050 of the U_L and from 139,927 to 157,108 of the I_R. Within the genome coordinate range–of 4,680 - 26,186 there are four partitions, while between 63,589 -91,050 there are 16 partitions. Supplemental Figure 1 maps across the genomes the partitions (17 to 20 kbp) found within phylogetic trees in categories I and II. The green lines represent sequence partitions within category I phylogenetic trees that clusters genotype IV and V sequences apart. Whereas the red lines represent sequence partitions within category II phylogenetic trees that are uncapable of clustering genotype V from genotype IV apart. The two orange rectangles on the ILTV U_L from 10,000 to 21,300 and from 57,771 to 92,184 positions of the genome indicate genome regions that contain sparsely distributed informative SNPs that differentiate genotype IV and V. However, sequence partitions (17 to 20 kbp) within these regions are indicated as red lines because do not include sufficient SNPs to cluster genotypes IV and V sequences separately.

Primer Design

A total of 16 primer pairs were designed to overlap the US region. Of the originally 16 primer pairs designed by Primer Primal scheme (20), primer sets, US1, US6, and US16 were redesigned. US1 failed to amplify at concentrations lower than 0.06μM in single-plex PCR whereas all other originally designed primers were able to amplify the products under lower concentration (Fig. 3). The US6 and US16 primer were redesigned because these did not produce enough reads after MinION sequencing of multiplex PCR products. The redesigned primers were named as, US1_beta, US6_beta and US16_beta respectively and their ability to amplify viral DNA from five ILTV isolates representative of four genotypes at a primer concentration 0.06μM was assessed. Several US16_beta primer pairs cand idates that appeared to be suitable after in silico analysis, were unsuccessful in amplifying under the required conditions. Although the US16_beta primer pairs did not provide high number of reads compared to the other 15 primer pairs, the redesigned US16_beta primer pair that presented the lesser potential for primer dimer formation in silico was included in the multiplex primer library to normalize the primer concentration in pool1 and pool2.

The sequence, annealing temperature, and amplicon sizes of primers adopted in the multiplex PCR are listed in Table 4. The amplicon sizes produced by the designed multiplex primers range between 980 to 1,172 bp. Primer lengths fall between 17 to 23 bp and the annealing temperatures range within 60 ±1 ℃ (Table 4). Table 5 shows the sequenced reads obtained from the 16 primer pairs out of a MinION sequencing run from a library containing five different samples. The result among the five barcoded samples agreed that primer pairs US7 and US12 being the most efficient primers and the redesigned US6_beta and US16_beta primers are the least efficient primers in the multiplex reaction.

Validation of the U_S multiplex PCR and MinION sequencing assay

Since the US16 primer pair was not efficient in providing sufficient reads as compared to the other 15 primer pairs, the SNPs within the US16 amplicon were excluded from the analysis. As a result, by comparing the 39 NCBI deposited whole genome sequences in Table 1, a total of 133 informative SNP sites were identified within the target region spanning 12,706 base pairs of the US ILTV genome. The 133 SNPs were separated in groups based on their ability to differentiate between closely related genotypes that were previously identified by multi-allelic PCR and Sanger sequencing and confirmed by whole genome analysis (6, 8, 31).

To evaluate the effectiveness of the multiplex PCR and MinION sequencing, five viral isolates of known genotype available in our laboratory depository were selected. Total DNA was extracted from viral stocks of strain J2 (Genotype VI) (Accession No:MF417808), 1874C5 (Genotype VI) (Accession No: JN542533), Laryngo Vac (Genotype IV, Vaccine strain Zoetis Animal Health) (Accession No: JQ083494), LT-IVAX (Genotype II, Tissue culture vaccine strain, Merck Animal Health) (Accession No: JN580312) and 63140 (Genotype V) (Accession No: JN542536). To mimic the viral genome load frequently detected in clinical samples from outbreaks of the disease, viral DNA load was normalized to a viral DNA concentration equivalent to Ct values ranging between 28 to 30 (Table 6). Individual samples were barcoded for library preparation by native ligation chemistry and finally sequenced. The number of reads, number of mapped reads and depth of coverage are presented in Table 6. The data showed that by sequencing more than 40k of reads per sample, the depth of coverage at each site surpassed 50X, meeting the minimum criteria for consensus establishing for this study. The coverage depth across the US region agreed among isolates and the deeper coverage at amplicon junctions after trimming off primer pairs displayed successful tiling of amplicons (Figure 4). The informative SNP sites for

genotype groups were built in silico from the NCBI database, and the MinION depth of coverage for the informative SNPs are listed in Table 7 to 10. Table 7 lists the seven informative SNPs in the US region that distinguish between genotype II (TCO vaccine) from genotype IV (CEO vaccine) and the depth of coverage for the seven SNPs of the LT-IVAX (TCO vaccine, MSD Animal Health) and Laryngo Vac (CEO Vaccine, Zoetis) commercial vaccines when tested by the multiplex PCR MinION sequencing assay. Table 8 lists the three informative SNPs depth of coverage that differentiate genotype IV Laryngo Vac®(Zoetis) from genotype V virulent isolate 63140 (lab isolate). Table 9 lists the 17 informative SNPs depth of coverage and the 30bp long insertion in the US region that allows the discrimination between virulent genotype V (63140 isolate) and virulent genotype VI (J2 and 1874C5 isolates). Table 10 lists the 27 informative SNPs depth of coverage that allow the discrimination between the virulent genotype VI isolates (J2 and 1874C5) and the NCBI sequences for genotypes VII-IX which are field strains that originated from backyard flocks in the late 1980's and early 1990's. A total of eighty-three SNPs and a 30bp long insertion were identified in silico between genotype VII, VIII, and IX backyard flock isolates which are additional differences that allow to easily discriminate among these archived viral strains (data not shown). Overall, the SNP analysis of the five ILTV isolates sequenced with US multiplex PCR MinION assay aligned with the data present in the NCBI database. However, with one exception in the US9 amplicon of isolate 1874C5, where a 30bp insert was identified in the current consensus compared to the reference sequence in the 1874C5 NCBI database (Accession No: JN542533). The presence of inserted sequence obtained by the MinION sequencing was validated by Sanger sequencing (Eurofins Genomics, Louisville, KY, USA) of the US9 amplicon for the 1874cC5 strain. There was agreement between the MinION sequencing results and previous genotyping results for known ILTV strains (Figure 5, Table 7-10). Furthermore, the novel assay

provided with significant depth of coverage, ensuring their accuracy of the assay. The LT-IVAX (TCO vaccine) sequencing yield more than 3000x depth of coverage at each informative SNPs site, followed by the CEO vaccine Laryngo Vac with at least 2000x, then isolates J2 and 1874C5 with more than 700x depth of coverage. Whereas genotype V isolate 63140 yield a depth across informative SNPs above 200x.

Sequencing and genotyping of clinical samples

Thirteen tracheal samples obtained from cases of the disease were genotyped by both the multi-allelic Sanger sequencing and the multiplex PCR MinION sequencing. Overall, both assays agreed in that 11 of the samples were genotyped as group VI, one as genotype II/III and one as genotype IV. As presented in Figure 5, except for sample S04N166, all other clinical samples (red text) displayed branch length differences within the respective genotype cluster of known North American NCBI reference sequences (black text). Table 11 presented unique nucleotide changes identified in clinical samples with a depth of coverage per nucleotide ≥ 1000x, which indicates these are established SNPs in currently circulating viruses. Among the unique SNPs, three SNPs yield synonymous amino acid changes in sample S22N02 (type IV) a C to T (position 129,966), in 8 of 11 samples categorized as genotype VI a T to A (position 138,699), and in sample S04N128 a C to T (position 139,517). There were three samples with SNPs that resulted in non-synonymous changes. In sample S04N48 (genotype VI) a T to C change (position130,277) which predicts a change from aromatic (F) to polar (S) amino acid in the tegument protein UL47. In sample S03N30 (genotype II/III) a C to T change (position 134,028) predicts an exchange of nonpolar amino acids (A to V) in glycoprotein J (US5). Lastly, in sample S01N17 (genotype VI) a G to T change predicts an exchange of negatively charged amino acids (E to D) in glycoprotein J (US5). The unique

nucleotide changes identified in each of the clinical sample were further verified via sanger sequencing (Eurofins genomics, KY, USA).

CHAPTER 3

DISCUSSION AND CONCLUSION

The highly contagious Infectious laryngotracheitis virus (ILTV) poses a significant threat to the poultry industry, resulting in considerable economic losses due to increased mortality rates, reduced production, and elevated susceptibility toward other respiratory pathogens. The control of the disease relies on prompt implementation of biosecurity programs supported by rapid and accurate molecular diagnostics and vaccine administration. However, the use of attenuated ILTV live vaccines complicated the process for its latency and the potential to regain virulence. Therefore, genome sequencing become an indispensable tool to track the source of an outbreak as well as serving to monitor genome changes of ILTV. The current genotyping approach in the lab adopted sanger sequencing three partial genes – ORFAB, gB and gM to classify an ILTV strain into one of the nine known genotypes. However, depending on the phase of infection and sample quality, the sequencing result may be hindered due to the potential impact on the ILTV DNA load and integrity. Moreover, the restriction on throughput of sanger sequencing poses an additional obstacle to detect the already few numbered SNPs. Altogether, the overall objective of this study is to increase the throughput of genotyping assay applicable for clinical samples by expanding the breadth of genome coverage to include more SNPs and intensifying the depth of sequencing coverage to enhance the confidence on SNP detection.

The first aim of this study was to identify a continuous region on ILTV genome, spanning from 13k to 20k, that is suitable for North American ILTV genotyping assay development. We selected 39 North American representative ILTV strains from nine known genotypes to perform

SNP calling and generated sliding window partitions at the step of ORFs for phylogenetic tree construction and closely related genotype cluster analysis. As a result, 91 overlapping partitions in the UL region, one in the US region and one in the IR region was generated. A total of 69 partitions in the UL and one in the US region were identified and can be used to develop genotyping assays because are able to accurately separate the known nine ILTV genotypes into five main clusters namely the (II/III), IV, V, VI and (VII-IX). Although 69 sequence segments from the category I of U_L region holds the potential to develop a genotyping assay (Table 3), the U_S region was eventually selected. Selection of the Us region as target of the multiplex PCR was mainly based on the phylogenetic analysis, which showed that overall, the U_S region holds a longer tree branch length as compared to the U_L phylogenetic trees, which suggests that there are more nucleotide changes between the closely related genotype groups of interests. In addition, the length of the U_S region of 13,094 bp is shorter than the partition sequences analyzed from the U_L region (ranging from 17 kbp to 20 kbp). This simplify the primer designing process because fewer primer pairs are required to amplify the selected genome region. However, this analysis acknowledges that there are other regions on the ILTV genome which are suitable for developing multiplex PCR genotyping assays. Though possessing abundant SNPs, the I_R region was excluded as a suitable region for genotyping assays for various reasons. First, the repeat regions contain multiple homopolymer stretches (32), which is known to be the major contributor of MinION sequencing errors (33, 34). Secondly, the repeat regions failed to separate genotypes IV and V which their discrimination is of epidemiological importance (Figure 2B). Finally, though the T_R and I_R regions are highly similar, the two are not identical (35). Such similarity could lead to ambiguous short read alignment between the T_R and I_R regions. This may introduce bias during the process of consensus building,

variant calling and even genotyping. Therefore, the repeat region was not considered appropriate for the current short read tiling approach.

The multiplex PCR method offered the advantage of employing multiple primers targeting different sites. This approach is beneficial toward cases with compromised DNA sample quality, as even if a few pairs of primer failed to anneal, the other primers may retain the ability to amplify and thereby improved the detection sensitivity. Given that the intended application was on clinical samples with variable quality of viral genome load, the adoption of the multiplex PCR method was justified. However, despite rigorous in silico and in bench analysis of the multiplex PCR primers, noticeable variations in primer efficiency were observed. Although factors such as melting temperature, primer secondary structures, primer dimer formation, and template length, were considered during the design process (as indicated in Table 4) (20, 23, 36), factors inherent to the viral genome sequences were not considered during the design phase. A study analyzing the contributing factors to the success or failure of sequencing 1438 human exons revealed that the regionalized GC content (a window of 21bp) could actually be a better predictor for PCR success than primer design. The higher regionalized GC content was associated to a lower likelihood of successful product amplification. Therefore, it is crucial to pay greater attention to the analysis on template sequences and refine the sequencing protocol accordingly, particularly for templates with higher regionalized GC content (37). For instance, DMSO or betaine addition may assist in stabilizing the DNA strand separation to allow successful amplification in the high GC/AT ratio templates (38). Though the GC content of ILTV genome is 48 % (39), which may not be considered high as compared to other Herpes viruses possessing more than 70% of GC composition across the genome (40). However, Herpes viruses were found to have unevenly distributed GC content across the genome, with higher GC content in the coding regions and

reduced GC content in the intergenic regions (41). In ILTV the GC content of the selected U_S region is 53%, which is higher than the average GC content of the whole ILTV genome. Significantly, the US16 amplicon exhibited the highest GC composition among all the examined amplicons in silico, accounting for 58%. This high GC content could potentially be a contributing factor to the unsuccessful amplification observed. Additionally, template secondary structure, such as formation of hairpin structures due to inverted repeats within the sequenced genome, can lead to polymerase disassociation, resulting in arrested amplification or polymerase slippage (42). Efforts dedicated to reducing variations in primer efficiency will be required as part of future assay optimization.

The total coverage of the non-continuous multi-allelic Sanger sequencing ILTV genotyping assay, which amplified partial segments within the ORF_AB, gB and gM genes, is limited to 4,686 bp of the viral genome which represents 3% of the ILTV genome (U_L + U_S +I_R) and includes a total of 27 informative SNPs to discriminate between the nine ILTV genotypes (I-IX). Specifically, 13 SNPs are identified in the ORF_AB, eight SNPs in the gB and six SNPs in the gM. Two SNPs allow the differentiation between the genotype I (USDA challenge virulent) and II/III (TCO vaccine related strains), three SNPs discerning between genotype II /III (TCO vaccine related strains) and genotype IV (CEO vaccines), three SNPs discriminate between genotype IV (CEO vaccines) and genotype V (CEO related virulent strains), nine SNPs discriminate between genotype V (CEO related virulent strains) and genotype VI (virulent field strains in commercial poultry) and 19 SNPs discriminate between the genotype VI and backyard poultry archive strains genotypes VII-IX (6, 8, 31). The current U_S multiplex PCR MinION sequencing assay included a total of 133 SNPs to discriminate between the nine ILTV genotypes (I-IX). Precisely, two SNPs allow the differentiation between the genotype I (USDA challenge

virulent) and II/III (TCO vaccine related strains), seven SNPs discerning between genotype II /III (TCO vaccine related strains) and genotype IV (CEO vaccines), three SNPs discriminate between genotype IV (CEO vaccines) and genotype V (CEO related virulent strains), eighteen SNPs discriminate between genotype V (CEO related virulent strains) and genotype VI (virulent field strains in commercial poultry) and 27 SNPs discriminate between the genotype VI and backyard poultry archive strains genotypes VII-IX. Moreover, compared to the previous partial gM, gB and ORF_AB gene multi-allelic genotyping assay, the current assay provides not only accuracy in genotyping but also increased the depth (≥ 200X) of coverage on each informative SNPs sites resulting in stronger support for SNP calling. Lastly, the improved sensitivity of the current assay enables sequencing and genotyping of samples with low viral loads (up to Ct value 30), surpassing the previous assay's requirement of samples with Ct value ≤ 28 (data not shown).

The success to sequence and genotyping directly from clinical samples by the multiplex PCR MinION sequencing assay demonstrated the sensitivity and flexibility of the assay on samples with a wide range of viral genome load (Ct value 17 to 33) (data not shown). Notably, sequencing of the U_S genome region of currently circulating strains, revealed that these viruses are carrying six unique nucleotide changes not found in previously sequenced North American strains. Among, the changes identified in the UL47 and US5 coding region resulted in non-synonymous changes. Despite being a non-essential gene for cell culture, deletion in the UL47 gene showed in vitro growth defects. Chickens infected with the ΔUL47 deletion mutants exhibited attenuation in vivo (43). The US5 gene encoded for the late protein glycoprotein J (gJ), is required for the egress of ILTV virions of cell (44). However, the exact impact of these mutations on the current circulating outbreak viruses remains to be discovered.

To guide future research devoted in developing and implementing genotyping assays using the multiplex PCR MinION sequencing, it is essential to address the challenges faced in this project. During the process, we noticed a minor proportion of the amplicon reads containing overhanging sequences at the end, which was consisted of partial adapter sequences, lengthy homopolymer stretches, and unidentified sequences. It was suspected that these overhangs may be artifacts generated during the library preparation process or due to sequences stalling in the MinION sequencing pore. Though these overhangs were minority, it could become problematic at sites of junction of two amplicons where the primer efficiencies were drastically different (Table 5, Figure 4). The minor proportion of overhangs resulting from extensively represented amplicons may overwrite the authentic sequence information from amplicons having lower depth of coverage. This could pose a challenge to our consensus determining approach by utilizing a "pileup" strategy that relies on calling bases consisting of 75% or more on each site. Therefore, these overhangs should be either masked or trimmed off and excluded from the consensus calling process to prevent biases. Another challenge encountered is the difficulty in mapping short reads with structural variants. Typically, structural variants refer to genomic alterations that are longer than 50bp, including deletions, duplications, insertions, inversions, translocations and interspersed duplications, which can complicate read mapping to reference sequence (45, 46). In the case of this study, an interspersed duplication was identified in the genotype VI strain-J2 (Accession No: MF417808) as compared to the USDA reference genome (Accession No:MN518177), only with a shorter length of 30bp. Without using any structural variant searching parameter, the ordinary mapping algorithm could not properly align the duplication to the correct location for consensus calling. The method proposed in this assay showed a way to build a consensus sequence with the potential presence of structural variant in the US9 amplicon.

Conclusions:

With more whole ILTV genome sequences being released and the advancement in high throughput sequencing technology, systemically analyzing these data could assist in disease surveillance and monitoring the evolution of the virus. This study reports a systematic approach to analyze the ILTV genome sequences to search for potential genome regions for the development of future genotyping assays to increase the resolution of the genotyping to samples. Also, in this study a rapid library preparation method and analysis pipeline for sequencing were established. The multiplex PCR and MinION sequencing method achieved high accuracy by its high depth of coverage and permitted simultaneous sequencing of multiple samples. Moreover, the result of the current paper provides flexibility to expand the current genotyping assay and may even apply the approach of analysis to worldwide ILTV sequences if monitoring global circulation is required.

REFERENCE

1. Bagust T, Jones R, Guy J. 2000. Avian infectious laryngotracheitis. Revue scientifique et technique (International Office of Epizootics) 19:483-492.

2. Davison AJ. 2010. Herpesvirus systematics. Veterinary microbiology 143:52-69.

3. Fuchs W, Veits J, Helferich D, Granzow H, Teifke JP, Mettenleiter TC. 2007. Molecular biology of avian infectious laryngotracheitis virus. Vet Res 38:261-79.

4. Thureen DR, Keeler CL, Jr. 2006. Psittacid herpesvirus 1 and infectious laryngotracheitis virus: Comparative genome sequence analysis of two avian alphaherpesviruses. J Virol 80:7863-72.

5. Shibley G, Luginbuhl R, Helmboldt C. 1962. A study of infectious laryngotracheitis virus. I. Comparison of serologic and immunogenic properties. Avian Diseases 6:59-71.

6. Spatz SJ, Garcia M, Riblet S, Ross TA, Volkening JD, Taylor TL, Kim T, Afonso CL. 2019. MinION sequencing to genotype US strains of infectious laryngotracheitis virus. Avian Pathology 48:255-269.

7. Oldoni I, Rodriguez-Avila A, Riblet S, Garcia M. 2008. Characterization of infectious laryngotracheitis virus (ILTV) isolates from commercial poultry by polymerase chain reaction and restriction fragment length polymorphism (PCR-RFLP). Avian Dis 52:59-63.

8. Perez Contreras A, van der Meer F, Checkley S, Joseph T, King R, Ravi M, Peters D, Fonseca K, Gagnon CA, Provost C. 2020. Analysis of whole-genome sequences of infectious laryngotracheitis virus isolates from poultry flocks in Canada: evidence of recombination. Viruses 12:1302.

9. Oldoni I, Garcia M. 2007. Characterization of infectious laryngotracheitis virus isolates from the US by polymerase chain reaction and restriction fragment length polymorphism of multiple genome regions. Avian Pathol 36:167-76.

10. Barboza-Solis C, Contreras AP, Palomino-Tapia VA, Joseph T, King R, Ravi M, Peters D, Fonseca K, Gagnon CA, van der Meer F. 2020. Genotyping of Infectious Laryngotracheitis Virus (ILTV) isolates from Western Canadian provinces of Alberta and British Columbia based on partial Open Reading Frame (ORF) a and b. Animals 10:1634.

11. Choi EJ, La TM, Choi IS, Song CS, Park SY, Lee JB, Lee SW. 2016. Genotyping of infectious laryngotracheitis virus using allelic variations from multiple genomic regions. Avian Pathol 45:443-9.

12. Lee S-W, Devlin JM, Markham JF, Noormohammadi AH, Browning GF, Ficorilli NP, Hartley CA, Markham PF. 2013. Phylogenetic and molecular epidemiological studies reveal evidence of multiple past recombination events between infectious laryngotracheitis viruses. PloS one 8:e55121.

13. Agnew-Crumpton R, Vaz PK, Devlin JM, O'Rourke D, Blacker-Smith HP, Konsak-Ilievski B, Hartley CA, Noormohammadi AH. 2016. Spread of the newly emerging infectious laryngotracheitis viruses in Australia. Infect Genet Evol 43:67-73.

14. Spatz SJ, Volkening JD, Keeler CL, Kutish GF, Riblet SM, Boettger CM, Clark KF, Zsak L, Afonso CL, Mundt ES, Rock DL, Garcia M. 2012. Comparative full genome analysis of four infectious laryngotracheitis virus (Gallid herpesvirus-1) virulent isolates from the United States. Virus Genes 44:273-85.

15. Chandra YG, Lee J, Kong B-W. 2012. Genome sequence comparison of two United States live attenuated vaccines of infectious laryngotracheitis virus (ILTV). Virus Genes 44:470-474.

16. Asif K, O'Rourke D, Shil P, Steer-Cope PA, Legione AR, Marenda MS, Noormohammadi AH. 2022. Rapid typing of infectious laryngotracheitis virus directly from tracheal tissues based on next-generation sequencing. Archives of Virology 167:1151-1155.

17. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome biology 5:1-9.

18. Smit A, Hubley R, Green P. 2021. RepeatMasker Open-4.0. 2013–2015. 2015. Google Scholar:289-300.

19. Torresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, Gruca A, Grynberg M, Kajava AV, Promponas VJ, Anisimova M, Jakobsen KS, Linke D. 2019. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res 47:10994-11006.

20. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nature protocols 12:1261-1276.

21. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. Nucleic acids research 40:e115-e115.

22. Grunenwald H. 2003. Optimization of Polymerase Chain Reactions. PCR Protocols:89.

23. Brown SS, Chen Y-W, Wang M, Clipson A, Ochoa E, Du M-Q. 2017. PrimerPooler: automated primer pooling to prepare library for targeted sequencing. Biology Methods and Protocols 2:bpx006.

24. Callison SA, Riblet SM, Oldoni I, Sun S, Zavala G, Williams S, Resurreccion RS, Spackman E, Garcia M. 2007. Development and validation of a real-time Taqman PCR assay for the detection and quantitation of infectious laryngotracheitis virus in poultry. J Virol Methods 139:31-8.

25. Stortchevoi A, Kamelamela N, Levine SS. 2020. SPRI Beads-based Size Selection in the Range of 2-10kb. Journal of biomolecular techniques: JBT 31:7.

26. Patel A, Belykh E, Miller EJ, George LL, Martirosyan NL, Byvaltsev VA, Preul MC. 2018. MinION rapid sequencing: Review of potential applications in neurosurgery. Surg Neurol Int 9:157.

27. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:129.

28. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132.

29. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.

31. Elshafiee EA, Hassan MSH, Provost C, Gagnon CA, Ojkic D, Abdul-Careem MF. 2022. Comparative full genome sequence analysis of wild-type and chicken embryo origin vaccine-like infectious laryngotracheitis virus field isolates from Canada. Infect Genet Evol 104:105350.

32. Chang P-C, Shieh HK, Shien J-H, Kang S-W. 2000. A homopolymer stretch composed of variable numbers of cytidine residues in the terminal repeats of infectious laryngotracheitis virus. Avian diseases:125-131.

33. Delahaye C, Nicolas J. 2021. Sequencing DNA with nanopores: Troubles and biases. PLOS ONE 16:e0257521.

34. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, Corbett CR. 2018. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. Sci Rep 8:10931.

35. Piccirillo A, Lavezzo E, Niero G, Moreno A, Massi P, Franchin E, Toppo S, Salata C, Palu G. 2016. Full Genome Sequence-Based Comparative Study of Wild-Type and Vaccine Strains of Infectious Laryngotracheitis Virus from Italy. PLoS One 11:e0149529.

36. Chavali S, Mahajan A, Tabassum R, Maiti S, Bharadwaj D. 2005. Oligonucleotide properties determination and primer designing: a critical examination of predictions. Bioinformatics 21:3918-3925.

37. Benita Y, Oosting RS, Lok MC, Wise MJ, Humphery-Smith I. 2003. Regionalized GC content of template DNA as a predictor of PCR success. Nucleic acids research 31:e99-e99.

38. Baskaran N, Kandpal RP, Bhargava AK, Glynn MW, Bale A, Weissman SM. 1996. Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome research 6:633-638.

39. Lee S-W, Markham PF, Markham JF, Petermann I, Noormohammadi AH, Browning GF, Ficorilli NP, Hartley CA, Devlin JM. 2011. First complete genome sequence of infectious laryngotracheitis virus. BMC genomics 12:1-6.

40. Sewatanon J, Srichatrapimuk S, Auewarakul P. 2007. Compositional bias and size of genomes of human DNA viruses. Intervirology 50:123-32.

41. Brown JC. 2007. High G+ C content of herpes simplex virus DNA: proposed role in protection against retrotransposon insertion. The open biochemistry journal 1:33.

42. Canceill D, Viguera E, Ehrlich S. 1999. Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency.

43. Helferich D, Veits J, Teifke JP, Mettenleiter TC, Fuchs W. 2007. The UL47 gene of avian infectious laryngotracheitis virus is not essential for in vitro replication but is relevant for virulence in chickens. Journal of general virology 88:732-742.

44. Mundt A, Mundt E, Hogan RJ, García M. 2011. Glycoprotein J of infectious laryngotracheitis virus is required for efficient egress of infectious virions from cells. Microbiology Society.

45. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. 2019. Structural variant calling: the long and the short of it. Genome biology 20:1-14.

46. Soylev A, Le TM, Amini H, Alkan C, Hormozdiari F. 2019. Discovery of tandem and interspersed segmental duplications using high-throughput sequencing. Bioinformatics 35:3923-3930.

Table 1. North American ILTV strains Full genome sequences downloaded from NCBI.

Strain Name	Genotype	Source of Virus	Accession Number	U_L	I_R	U_S	T_R
USDA 09/2019	I	Challenge strain	MN518177	1-109,578	109,579-126,832	126,833-139,926	139,927-157,180
USDA	I	Challenge strain	JN542534	1-109,593	109,594-124,127	124,128-137,222	137,223-151,756
CAN/BC-10-1122	II	TCO vaccine	MT797249	1-109,626	109,627-123,285	123,286-136,483	136,384-150,021
TCO high passage	II	TCO vaccine	JN580314	1-109,575	109,576-123,408	123,409-136,502	136,503-150,335
TCO IVAX	II	TCO vaccine	JN580312	1-109,575	109,576-125,973	125,974-139,067	139,068-155,465
TCO low passage	II	TCO vaccine	JN580315	1-109,575	109,576-125,973	125,974-139,067	139,068-155,465
81658	III	Virulent Isolate	JN542535	1-109,575	109,576-123,408	123,409-136,502	136,503-150,335
3.26.90	IV	CEO vaccine	MF417809	1-112,915	112,916-126,738	126,739-139,832	139,833-153,655
CAN/ON-2462241	IV	CEO vaccine	OK573459	1-112,963	112,964-126,722	126,723-139,820	139,821-153,589
CAN/ON-2462242	IV	CEO vaccine	OK624781	1-112,984	112,985-126,724	126,725-139,822	139,823-153,565
CEO high passage	IV	CEO vaccine	JN580316	1-112,915	112,916-126,734	126,735-139,828	139,829-153,647
CEO Low passage	IV	CEO vaccine	JN580317	1-112,915	112,916-126,731	126,732-139,825	139,826-153,641
CEO TRVX	IV	CEO vaccine	JN580313	1-112,915	112,916-126,734	126,735-139,828	139,829-153,647
Laryngo Vac	IV	CEO vaccine	JQ083494	1-112,930	112,931-126,728	126,729-139,826	139,827-153,624
LT Blen	IV	CEO vaccine	JQ083493	1-112,930	112,931-126,727	126,728-139,825	139,826-153,623
CAN/QC-230414	V	CEO vaccine	OL661344	1-112,921	112,922-126,727	126,728-139,821	139,822-153,629
14.939	V	Virulent Isolate	MF417811	1-112,917	112,918-126,726	126,727-139,820	139,821-153,629

63140/C/08/BR	V	Virulent Isolate	JN542536	1-112,915	112,916-126,727	126,728-139,821	139,822-153,633
CAN/AB-15A	V	Virulent Isolate	MT797239	1-112,916	112,917-126,735	126,736-139,829	139,830-153,647
CAN/AB-S42	V	Virulent Isolate	MT797241	1-113,050	113,051-126,714	126,715-139,808	139,809-153,469
CAN/AB-S45	V	Virulent Isolate	MT797242	1-112,917	112,918-126,727	126,728-129,821	139,822-153,630
CAN/AB-S50	V	Virulent Isolate	MT797243	1-112,918	112,919-126,731	126,732-139,825	139,826-153,641
CAN/AB-S61	V	Virulent Isolate	MT797244	1-112,918	112,919-126,731	126,732-139,825	139,826-153,642
CAN/AB-S77	V	Virulent Isolate	MT797246	1-112,917	112,918-126,729	126,730-139,823	139,824-153,633
CAN/AB-S84	V	Virulent Isolate	MT797247	1-113,063	113,064-126,730	126,731-139,824	139,825-153,497
CAN/AB-T85	V	Virulent Isolate	MT797248	1-112,917	112,918-126,727	126,728-139,821	139,822-153,631
CAN/QC-1990662	V	Virulent Isolate	MT797250	1-112,918	112,919-126,732	126,733-139,826	139,827-153,621
CAN/QC-2154822	V	Virulent Isolate	MT797251	1-111,060	111,061-124,672	124,723-137,666	137,667-151,326
CAN/AB-S63	VI	Virulent Isolate	MT797245	1-113,921	113,922-126,750	126,751-139,875	139,876-152,703
CAN/QC-2175807	VI	Virulent Isolate	MT797252	1-112,949	112,950-126,740	126,741-139,865	139,866-153,662
CAN/QC-2236832	VI	Virulent Isolate	OL354140	1-113,941	113,942-126,751	126,752-139,876	139,877-152,687
CAN/QC-2301711	VI	Virulent Isolate	OK646550	1-113,931	113,932-126,739	126,740-139,864	139,865-152,674
CAN/QC-2470159	VI	Virulent Isolate	ON598585	1-113,019	113,020-126,725	126,726-139,850	139,851-153,566
CAN/QC-2551439	VI	Virulent Isolate	ON598586	1-112,930	112,931-126,748	126,749-139,873	139,874-153,691
J2	VI	Virulent Isolate	MF417808	1-112,908	112,909-126,747	126,748-139,872	139,873-153,711
VFAR-043	VI	Virulent Isolate	MG775218	1-113,917	113,918-126,746	126,747-139,871	139,872-152,701
6.48.88	VII-IX	Virulent Isolate	MF417810	1-113,212	113,213-127,054	127,055-140,180	140,181-154,022

S2.816	VII-IX	Virulent Isolate	MF417807	1-113,213	113,214-127,044	127,045-140,170	140,171-154,001
CAN/AB-S20	VII-IX	Virulent Isolate	MT797240	1-113,953	113,954-126,769	126,827-139,866	139,867-152,691

Table 2. Annotation of the repeats and low complexity sequences on USDA (Accession No: MN518177) 09/2019 ILTV genome.

Repeat class	Position begins	Position end
Simple repeat	6,928	6,992
Simple repeat	7,990	8,022
Simple repeat	12,119	12,162
Simple repeat	21,283	21,313
Simple repeat	35,714	35,753
Low complexity	112,631	112,681
Low complexity	115,670	115,707
Simple repeat	121,574	121,608
Simple repeat	126,784	126,827
Simple repeat	141,151	145,185
Simple repeat	151,052	151,089
Simple repeat	154,078	154,128

Table 3. Closely related genotype clustering analysis for the partitioned segments in UL, US, and IR region of ILTV genome.

Region	Category	Genotype clustered	Genome Coordinate†	Number of Partitioned segments
U_L	I	(II/III), IV, V, VI, (VII-IX)*	1 – 71,249	48
U_L	I	(II/III), IV, V, VI, (VII-IX)	73,233 – 109,578	21
U_L	II	(II/III), (IV/ V), VI, (VII-IX)	4,680 – 26,186	4
U_L	II	(II/III), (IV/ V), VI, (VII-IX)	63,589 – 91,050	16
U_S	I	(II/III), IV, V, VI, (VII-IX)	126,833-139,926	1
I_R	II	(II/III), (IV/V), VI, (VII-IX)	139,927-157,108	1

†Using the genome coordinate of ILTV USDA strain (Accession No: MN518177)

*Genotypes within a paratheses are indicating that they are clustered together in the phylogenetic tree, despite differences in SNPs within the cluster.

Table 4. Primers designed and amplicon size for U_S multiplex PCR.

Primer Name	Primer Sequence	Tm(°C)	Product size
US1F_beta	CCGTTCGCCATCATCCGA	59.9	1,033 bp
US1R_beta	TGTATGCGAGCCGTCAACTT	60	
US2F	TTAGAGAAGATGCGGTTTCGGC	61	1,003 bp
US2R	AGTACGAATCCCTGGCGAGTAT	60.8	
US3F	TCACAGGGACATCAAACTCGAA	59.6	1,000 bp
US3R	GTCCACTTACCGCCAAATTCCT	60.9	
US4F	CGACGCAAATCACAGAGGGATA	60.5	1,004 bp
US4R	AGTCATCGAAATTGCCGTCTGC	61.8	
US5F	TTCCATTTTACTGGCGCGGC	61.9	989 bp
US5R	CGGTCAGCGGGTACACTTTAT	60.1	
US6F_beta	TTAAGCCGGTGATAGACGAGC	59.9	1,172 bp
US6R_beta	GGCGGTGAAAAGACAACCAT	59	
US7F	CCAATTGTCTAGACATGCCCCC	61	979 bp
US7R	CCCAATGTACTGCGCTTTTGTC	60.7	
US8F	TCGGGTCTTAGTTAACGCCTCT	60.9	1,018 bp
US8R	ATACCACGGTTCCTGACTCTGG	61.5	
US9F	CCCGGTGAGTCAGAAAATACACT	60.3	1,036 bp
US9R	CCGTTACTTCGCCATCGTTTGA	61.5	
US10F	GGACTTATCCAGTGCAAACGAC	59.6	1,020 bp
US10R	TCGCAGCCTCTTGATACCCTAG	61.3	
US11F	TCGTACTGCTTTCTTCCTTCGC	60.9	1,011 bp

US11R	GCATGGAGACGGCATTAGAACT	60.7	
US12F	AGAAGACTCGGAGGACGACATG	61.5	1,000 bp
US12R	CAACGTACGCTTGCCAGTTG	60.4	
US13F	TCGGTGTATCAATTCGCGAACA	60.7	980 bp
US13R	CCTTCATCTTGCACGCGTTC	60.2	
US14F	TAGTTCTCGCCTCTTGTCTTGC	60.4	996 bp
US14R	CCGTGTAGCTCCATCCCAAAAT	60.7	
US15F	CTGGATTGTTCCCGCGATACTC	61.1	1,065 bp
US15R	TGCTGCTTGAGAATCCGGAAG	60.7	
US16F_beta	ACCAATGGACGAAACGGGTAA	59.9	1,109 bp
US16R_beta	GAAGACACGCCCTCCGC	60.5	

Table 5. Number of reads yield for the 16 primer pairs included in the multiplex PCR on ILTV strain tested.

Primer pairs	ILTV isolates				
	63140	**TCO**	**J2**	**Laryngo vac**	**1874C5**
US1_beta	4634	3359	2696	4435	4963
US2	7317	5591	4421	6913	7646
US3	3008	1729	1646	2368	2766
US4	7471	5790	6227	7453	6194
US5	3133	3748	2951	4445	5045
US6_beta	298	926	784	334	796
US7	9633	7304	6468	9291	8894
US8	3170	3219	2813	2901	3530
US9	3984	3388	2509	4463	4147
US10	4584	3631	2990	4752	4349
US11	2887	1400	1157	2436	2535
US12	8749	7701	6748	6754	7042
US13	1956	2427	1678	3129	3799
US14	3623	3555	2554	3749	3201
US15	3726	4793	4387	4265	5753
US16_beta	368	822	310	185	446

Table 6. Sequencing statistics of U_S multiplex PCR MinION sequencing assay using the five ILTV isolates.

	Ct	Total reads	(%) Reads mapped	Mean depth	Minimum depth†
63140	29.3	62,311	100%	3896	193
IVAX	28.2	53,773	100%	3675	559
J2	29.0	46,109	100%	3323	214
Laryngo vac	30.1	62,503	100%	3873	128
1874C5	29	64,539	100%	4090	340

† the minimum depth does not include the bases of the first (US1F_beta) and last (US16_beta) primer binding sites, which have no tiling overlaps.

Table 7. Informative SNPs differentiating genotypes II (TCO) and IV (CEO) and the depth of coverage using MinION sequencing technology.

No	USDA coordinate	NCBI reference base		Sequencing Depth of coverage	
		TCO (II)[a]	CEO (IV)[b]	IVAX TCO[c]	Laryngo Vac CEO[d]
1	127799	C	T	5495	6592
2	127911	A	G	5501	6096
3	129968	A	G	5479	6582
4	132739	T	C	9862	10790
5	133133	C	T	3068	2524
6	138343	G	A	3342	3307
7	138870	G	A	4516	3932

[a] Consensus based on four NCBI genome sequences listed in Table 1.

[b] Consensus based on eight NCBI genome sequences listed in Table 1.

[c] Sequencing run conducted using ILTV IVAX TCO vaccine (accession no: JN580312).

[d] Sequencing run conducted using ILTV Larygo Vac CEO vaccine (accession no: JQ083494).

Table 8. Informative SNPs differentiating genotypes IV (CEO) and V (CEO related virulent) and the depth of coverage using MinION sequencing technology.

No	USDA coordinate	NCBI reference base		Sequencing Depth of coverage	
		CEO (IV)[a]	CEO related virulent (V)[b]	Laryngo Vac CEO[c]	63140[d]
1	127562	A	G	4284	4114
2	128816	A	G	2219	2692
3	138343	A	G	3306	3391

[a]Based on eight NCBI genome sequences listed in Table 1

[b] Based on thirteen NCBI genome sequences listed in Table 1

[c]The sequencing run were conducted using ILTV isolate Larygo Vac CEO (accession no: JQ083494)

[b]The sequencing run were conducted using ILTV isolate 63140 (accession no: JN542536)

Table 9. Informative SNPs differentiating genotypes V (CEO related virulent) and VI (commercial poultry) and the depth of coverage using MinION sequencing technology.

No	USDA coordinate	NCBI reference base		Sequencing Depth of coverage		
		CEO related virulent (V)[a]	commercial poultry (VI)[b]	63140[c]	J2[d]	1874C5[e]
1	126983	G	A	4438	2549	4662
2	128488	A	C	5500	3238	5279
3	130801	T	C	2353	2010	3809
4	131053	C	A	2900	2769	4656
5	131115	A	G	280	755	772
6	131233	A	G	285	764	781
7	131459	C	A	286	767	784
8	131511	T	G	273	722	742
9	131653	C	T	290	717	723
10	131825	G	C	262	764	772
11	131826	G	T	280	762	772

12	132060	-	T	-[f]	6127	8415
13	133020	C	T	2970	2663	3347
14	134219	-	†	-	>2000	>2000
15	134305	G	A	3807	2395	3961
16	136965	C	A	9982	7900	10075
17	139492	C	T	3511	4140	5442
18	139564	G	A	3609	3903	5596

[a] Based on thirteen NCBI genome sequences listed in Table 1

[b] Based on eight NCBI genome sequences listed in Table 1

[c] The sequencing run were conducted using ILTV isolate 63140 (accession no: JN542536)

[d]The sequencing run were conducted using ILTV isolate J2 (accession no:MF417808)

[e] The sequencing run were conducted using ILTV isolate 1874C5(accession no:JN542533)

[f] Site with a trinucleotide insertion uniquely in genotype VI. Therefore, no information on depth of coverages in genotype

†30 bp Insertion AAA TTA CTC AGA CTC CGA GTA CGG TAC CGG

Table 10. Informative SNPs differentiating wildtypes genotypes VI (commercial poultry) and VII-IX (wildtype strains).

No	USDA coordinate	NCBI reference base		Sequencing Depth of coverage	
		commercial poultry (VI)[a]	Wildtype strains (VII-IX)[b]	J2[c]	1874C5[d]
1	127891	A	C	4320	7475
2	129388	-	CCG	-[e]	-
3	130107	C	T	6008	5912
4	130723	A	C	2140	4099
5	131053	A	C	2767	4656
6	131055	C	T	2760	4633
7	131340	T	G	760	779
8	131653	T	C	715	723
9	131825	C	G	764	772
10	131826	T	G	762	772
11	132017	T	C	6300	8519
12	132278	G	A	6229	8584

13	133158	G	C	2693	3400
14	133736	C	T	2404	3979
15	134244	A	G	2373	3917
16	134305	A	G	2395	3961
17	135081	A	G	2893	4637
18	135600	G	T	1105	2423
19	135612	C	T	1100	2437
20	135664	C	A	1121	2441
21	135891	G	A	1122	2438
22	136987	A	T	7836	9958
23	137066	A	C	1603	3577
24	137317	A	G	1627	3648
25	137452	T	C	1597	3630
26	138514	T	C	2467	3080
27	139492	T	C	4140	5442

[a] Based on eight NCBI genome sequences listed in Table 1

[b] Based on three NCBI genome sequences listed in Table 1

[c] The sequencing run were conducted using ILTV isolate J2(accession no:MF417808)

[d] The sequencing run were conducted using ILTV isolate 1874C5(accession no:JN542533)

[e] Site with a trinucleotide insertion uniquely in genotype VII-IX strains. Therefore, no information on depth of coverages in genotype VI.

Table 11. Unique nucleotide changes and amino acids substitution identified in clinical samples as compared to 39 North American strains using U_S multiplex PCR MinION sequencing assay.

Sample ID	Genotype	USDA coordinates	DOC[a]	Nucleotide change	Amino acid change	Gene
S03N30	II or III	134018	4352	C > T	A > V	US5
S22N02	IV	129966	3640	C > T	Synonymous	UL47
S04N48	VI	130277	1413	T > C	F > S	UL47
S01N17	VI	134214	4870	G > T	E > D	US5
S01N44	VI	138688	10555	T > A	Synonymous	US8
S04N48	VI	138688	6840	T > A	Synonymous	US8
S04N128	VI	138688	8820	T > A	Synonymous	US8
S04N161	VI	138688	5805	T > A	Synonymous	US8
S04N162	VI	138688	5120	T > A	Synonymous	US8
S04N163	VI	138688	6559	T > A	Synonymous	US8
S04N164	VI	138688	5574	T > A	Synonymous	US8
S04N165	VI	138688	5182	T > A	Synonymous	US8
S04N167	VI	138688	10188	T > A	Synonymous	US8

S04N128	VI	139517	5223	C > T	Synonymous	US9

[a] Depth of coverage (DOC)

* The letter in the parentheses represents the unique SNP identified on the coordinates

†Using USDA (Accession No: MN518177) genome coordinates

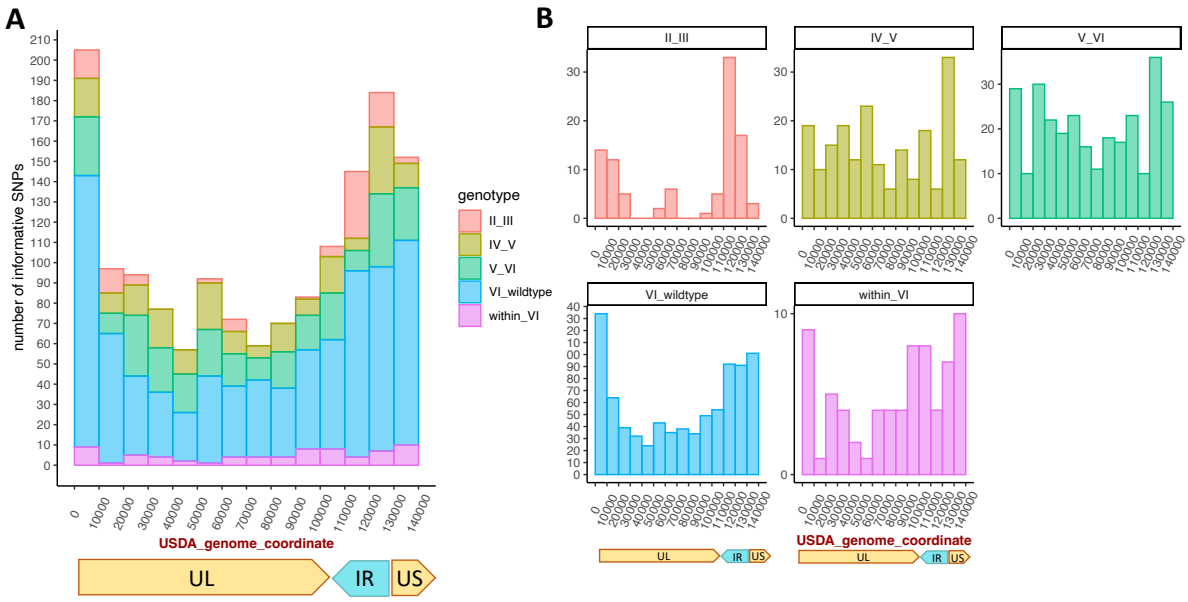
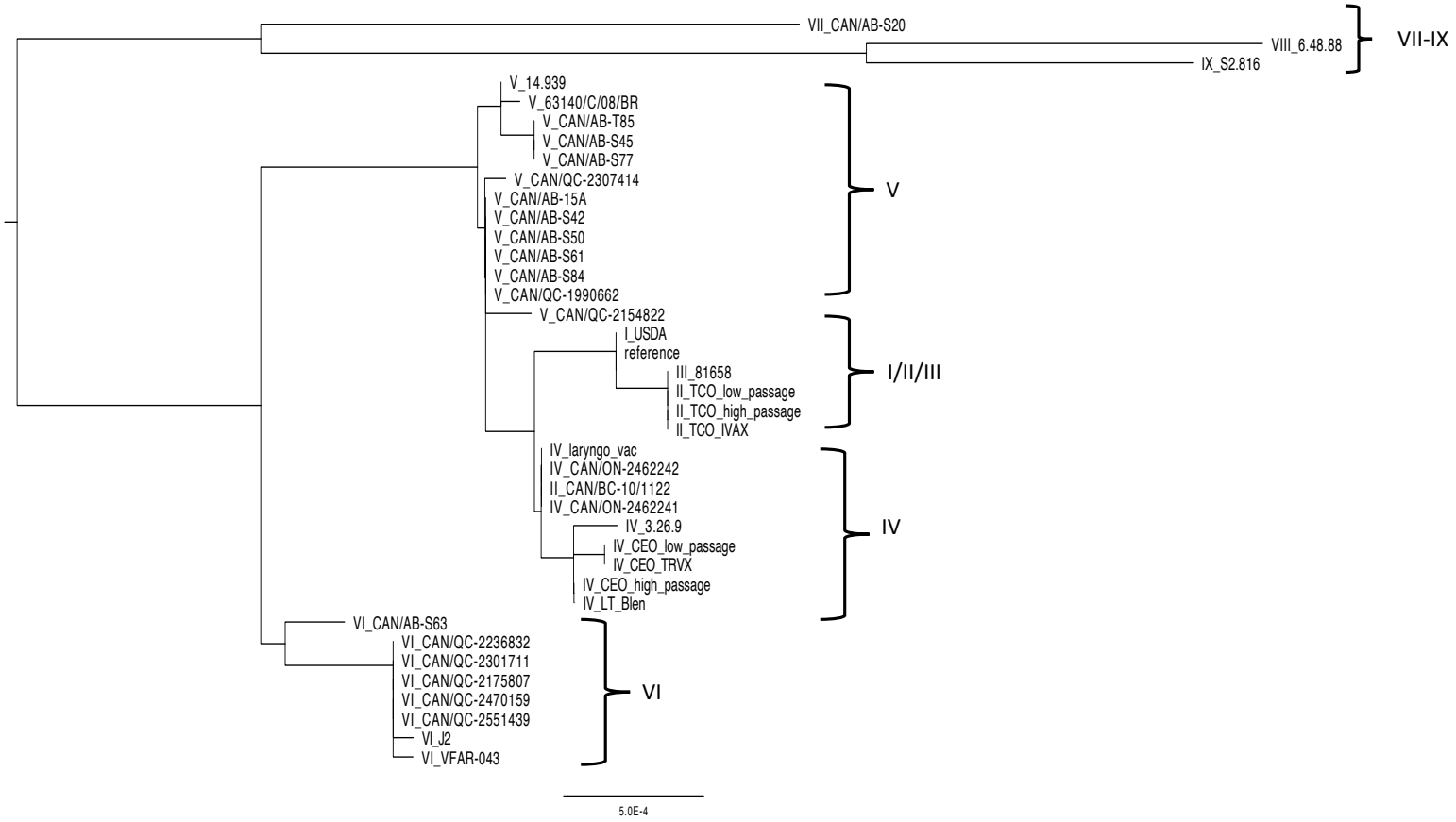


Figure 1. Histograms showing the informative SNP distribution from 1bp to 137,222 bp of USDA coordinate (Accession No. MN518177). Sites polymorphic between type II-III (red), type IV - V(yellow), V-VI (green), VI-IX (blue) and within genotype VI (pink) are charactorized with different color blocks. The x-axis refers to the position of the genome while the Y-axis refers to the amount of SNPs for every 10,000 bp. (A) Histogram showing SNP distribution of all of the closely related genotyping categories. (B) Histograms showing SNP destribution of each closely related viral genotype pairs.

A


VII_CAN/AB-S20
VIII_6.48.88
IX_S2.816
VII-IX
V_14.939
V_63140/C/08/BR
V_CAN/AB-T85
V_CAN/AB-S45
V_CAN/AB-S77
V_CAN/QC-2307414
V_CAN/AB-15A
V_CAN/AB-S42
V_CAN/AB-S50
V_CAN/AB-S61
V_CAN/AB-S84
V_CAN/QC-1990662
V
V_CAN/QC-2154822
I_USDA
reference
III_81658
II_TCO_low_passage
II_TCO_high_passage
II_TCO_IVAX
I/II/III
IV_laryngo_vac
IV_CAN/ON-2462242
II_CAN/BC-10/1122
IV_CAN/ON-2462241
IV_3.26.9
IV_CEO_low_passage
IV_CEO_TRVX
IV_CEO_high_passage
IV_LT_Blen
IV
VI_CAN/AB-S63
VI_CAN/QC-2236832
VI_CAN/QC-2301711
VI_CAN/QC-2175807
VI_CAN/QC-2470159
VI_CAN/QC-2551439
VI_J2
VI_VFAR-043
VI
5.0E-4

B

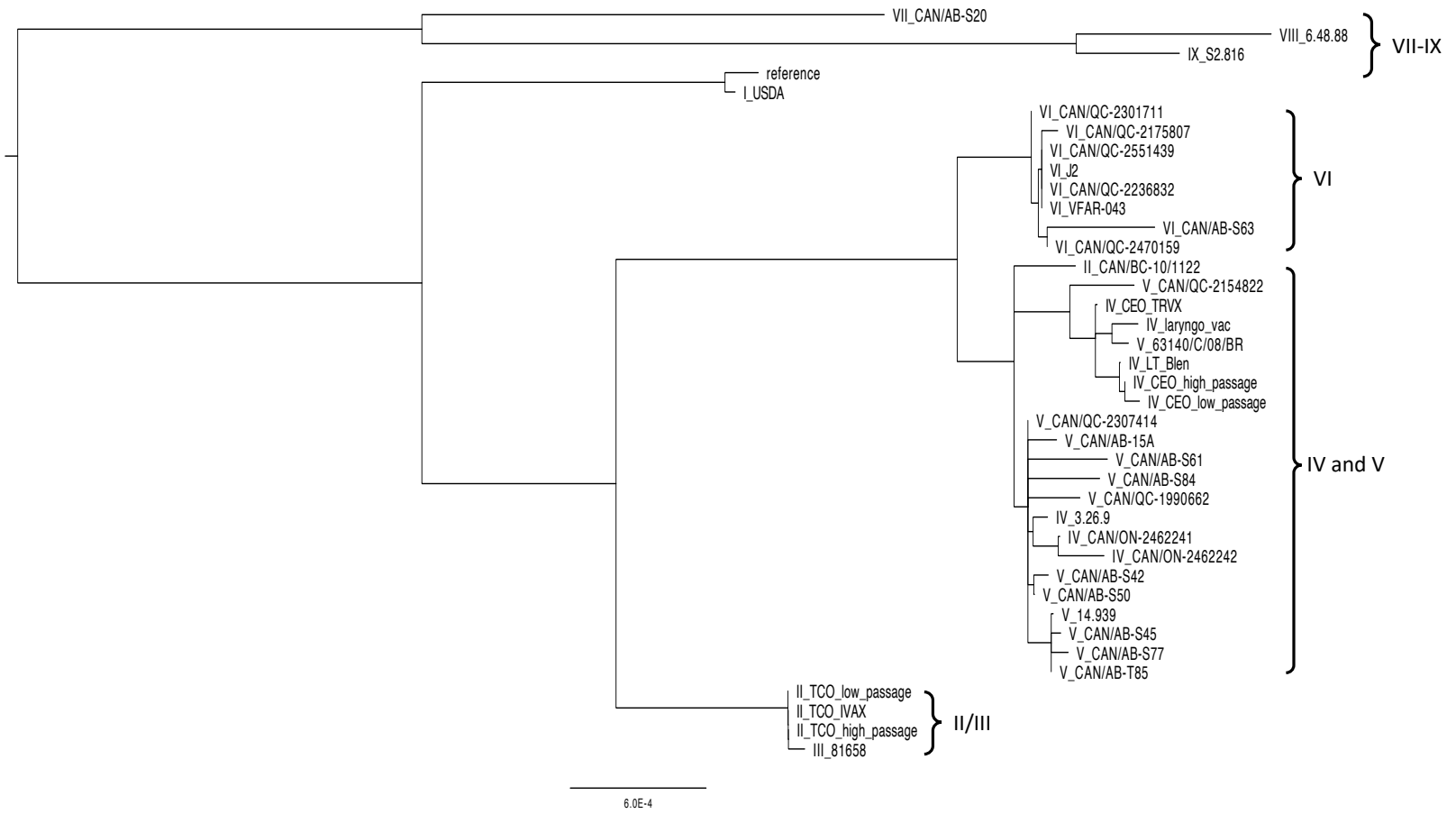


Figure 2. Phylogenetic trees of the two categories for genotype clustering from partitioned segments. Tip labels on the trees included the genotypes (Roman number) and strain name. The trees were constructed with Neighbor Joining clustering algorithm using HKY substitution model using Geneious Prime (v.11.0.4). The scale bar indicates the number of substitutions per site. (A) A

phylogenetic tree example of category I, capable of clustering genotype of interest II/III, IV, V, VI and VII-IX into five groups. The segment of U_S region including ORFs US2 to US9(13,094 bp long, from 126,833-139,926 of USDA coordinate) was used to construct this tree. (B) A phylogenetic tree example of category II. While departing the genotypes, II/III, IV/V to the remaining VI and VII-IX clusters, it could not discern clusters between genotype IV and V. As a result, this category is capable of clustering genotypes of interest into four groups. The segment of I_R region including ORFs sORF4/3 to ICP4 (17,254 bp long, from 109,579-126,832 of USDA coordinate) was used to construct this tree.

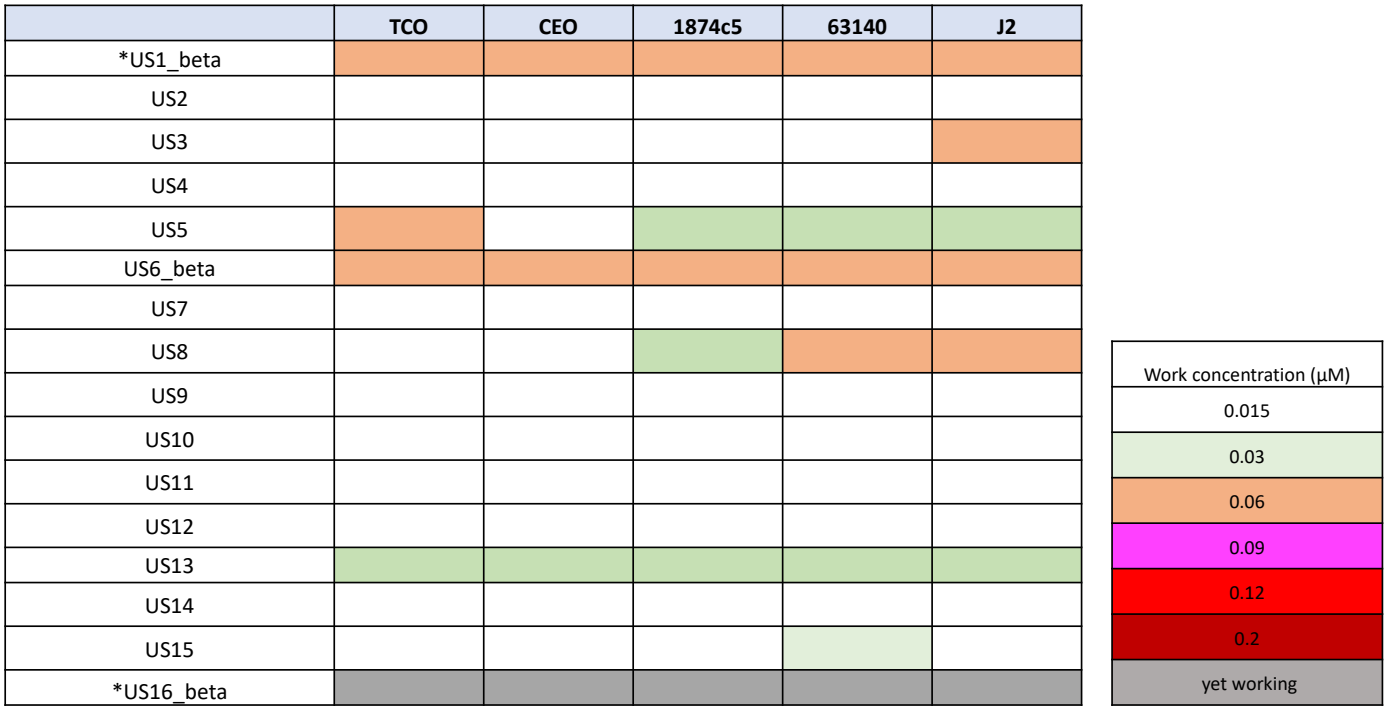


Figure 3. Working concentration of U_S multiplex PCR primers. The working primer concentration was validated against five ILTV strains – TX J2 P7 (genotype VI), 1874C5 Ct P4 (genotype VI), Laryngo vac® (Zoetis) (genotype IV), LT IVAX (genotype II) and 63140 P4 (genotype V), using 68°C as annealing temperature. The color gradient scale indicates the lowest primer working concentration (µM/µl) against each testing strains.

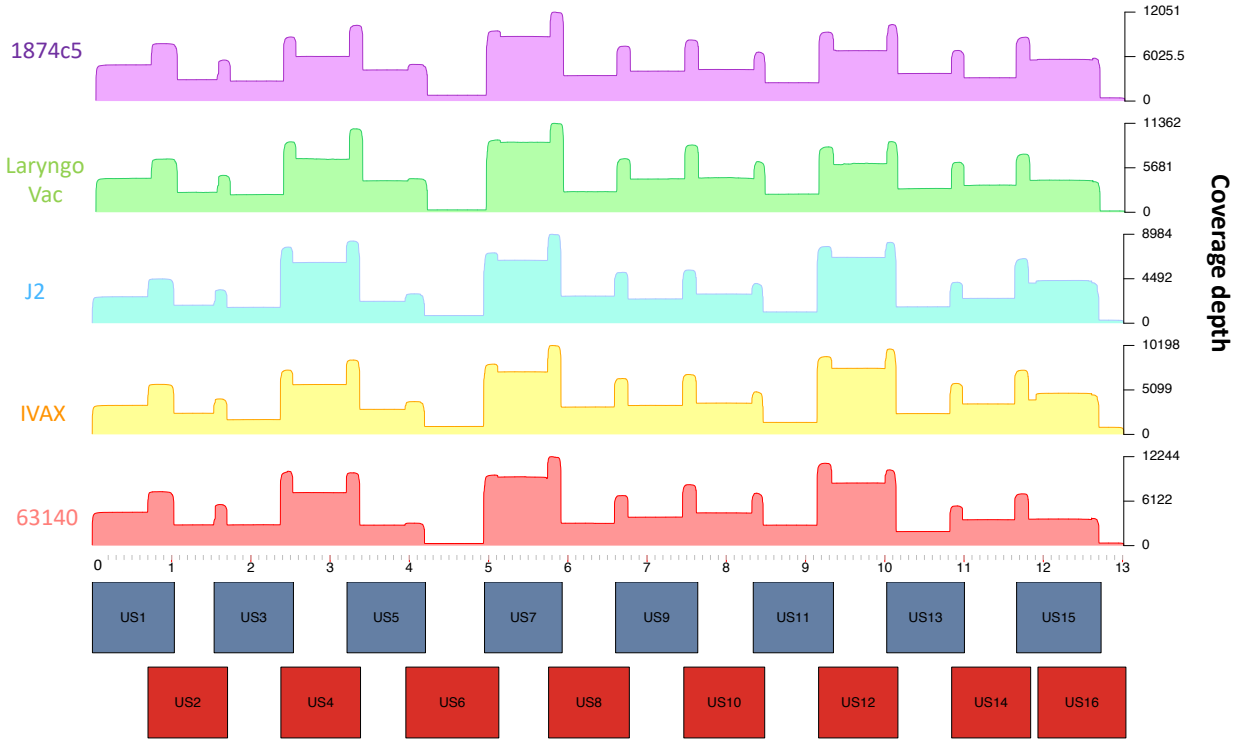


Figure 4. Target region coverage plot for five selected isolates. The plotted regions included the 13,019 bp long U_S region of ILTV genome displayed in the x-axis. The Ct values for the sequenced samples ranged from 28.2 to 30.1. Noted that the y-axis of the coverage depth for each sample was slightly different.

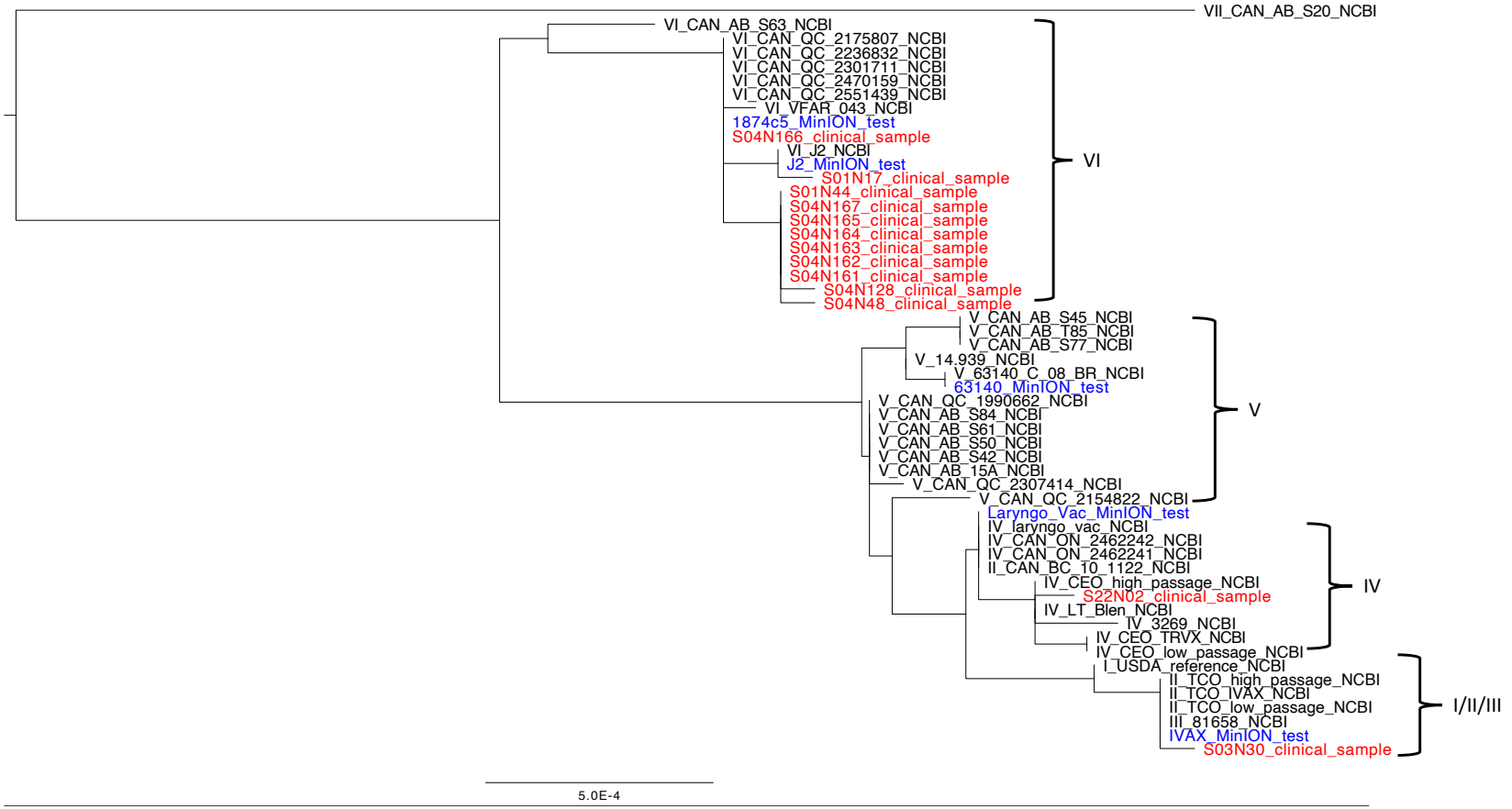
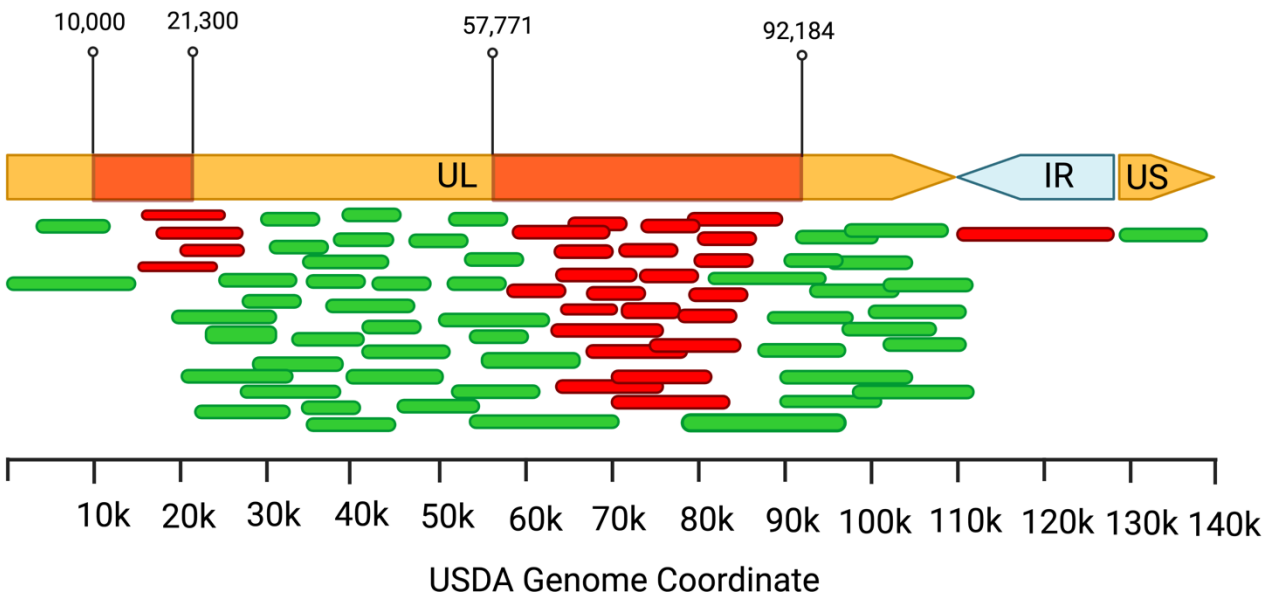


Figure 5. Phylogenetic tree of the NCBI selected North American sequences and sequencing result of clinical samples and viral isolates using U_S multiplex PCR MinION sequencing assay. The plotted regions included the 12,706 bp of U_S multiplex PCR ILTV genome. Viral isolates MinION sequenced samples were labeled in blue, NCBI database sequence labeled in black, whereas the clinical

samples were labeled in red. The trees were constructed with Neighbor Joining clustering algorithm using HKY substitution model using Geneious Prime (v.11.0.4).



Supplementary Figure 1. Schematic diagram on the qualified and unqualified partitions as expanded target for ILTV genotyping assay. The x-axis scale displayed the USDA genome coordinates and the arrow shaped block indicated the U_L, I_R and US genome regions of ILTV across the USDA genome. Within the U_L region, there are two distinct orange shaded rectangles highlighting regions with sparsely distributed SNPs (Single Nucleotide Polymorphisms) between genotype IV and V. Specifically, these regions span from 10,000 to 21,300 bp and from 57,771 to 92,184 bp. The red rod-shaped lines represent unqualified regions that are uncapable of clustering genotype IV and V apart on a phylogenetic tree, whereas green rod-shaped lines represent qualified regions that are capable of clustering genotype IV and V separately on a phylogenetic tree.